Nonparametric estimation of risk measures of collective risks

Dissertation zur Erlangung des Grades des
Doktors der Naturwissenschaften
der Fakultät für Mathematik und Informatik
der Universität des Saarlandes

eingereicht im August 2017
in Saarbrücken

von

Alexandra Lauer

Tag des Kolloquiums:	30.11.2017

Mitglieder des Prüfungsausschusses:

Vorsitzender:	Professor Dr. Mark Groves
1. Berichterstatter:	Professor Dr. Henryk Zähle
2. Berichterstatter:	Professor Dr. Eric Beutner
Protokollführer:	Dr. Tobias Mai
Dekan:	Professor Dr. Frank-Olaf Schreyer

“Prediction is very difficult,
especially if it’s about the future.”

Mark Twain (1835–1910)

Acknowledgements

First of all I want to express my gratitude towards my supervisor Professor Dr. Henryk Zähle, for giving me the opportunity to work with him. I deeply value his constant support and his motivation.

Second, I want to thank Professor Dr. Eric Beutner for being the co-referee to this thesis.

I also want to thank the present and former members of my working group, the Stochastics group at Saarland University. My special thanks go to Ulrike Mayer, Robert Knobloch, Philip Oberacker and Christian Gärtner for many interesting mathematical, as well as political discussions.

I also got to know many wonderful people from the department of mathematics. For their friendship, for good talks and many gym classes I want to thank Julia Seydel, Anne Wald, Jens Horn and Benedikt Hewer.

Finally I want to thank my beloved husband Michael Lembert for his never ending patience and his constant motivation. Without you all of this would not have been possible.

Contents

Abstract xiii

Zusammenfassung xiii

Introduction 1

1 Risk measures and risk functionals 7

1.1 Definitions 7

1.2 Examples of popular risk measures 8

1.3 Distortion risk measures and the Kusuoka representation 9

1.4 Risk functionals and regularity properties 10

I Estimation under a constant ratio of sampling size and collective size 15

2 Nonparametric estimation of risk measures in the individual model 17

2.1 Estimators for the individual premium in the individual model 18

2.2 Strong rates and asymptotic normality for the individual premium estimators in the individual model 22

2.3 Numerical simulations 28

2.4 Bootstrapping the individual premium in the individual model 32

2.4.1 Bootstrap-based bias correction 34

2.4.2 Bootstrap consistency for the nonparametric individual premium estimator 36

2.5 Numerical simulations: The effect of the bootstrap-based bias correction 46

2.5.1 Average Value at Risk 46

2.5.2 Value at Risk 51

2.5.3 Conclusion 52

3 Nonparametric estimation of risk measures in the collective model 53

3.1 Estimators for the individual premium in the compound Poisson model . . . 54

3.2 Strong rates and asymptotic normality for the individual premium estimators in the compound Poisson model 59

3.2.1 Proof of Theorems 3.2.2 and 3.2.3 63

3.2.2 Proof of Corollaries 3.2.4 and 3.2.6 70

II Estimation under constant collective sizes 79

4 Nonparametric estimation under constant collective sizes in the individual model (revisited) 81

4.1 An estimator for the premium in the individual model 82

4.2 The notion of uniform quasi-Hadamard differentiability 83

4.3 The asymptotic distribution and almost sure bootstrap consistency of the estimated premium in the individual model 84

5 Semiparametric estimation under constant collective sizes in the collective model 89

5.1 A semiparametric estimator for the premium in the compound Poisson model 91

5.2 Asymptotic distribution of the premium estimator in the compound Poisson model 92

5.3 An outlook on almost sure bootstrap consistency of the estimated premiums 94

5.4 Proofs 96

5.4.1 A functional delta-method for plug-in estimators of uniformly quasi-Hadamard differentiable statistical functionals 96

5.4.2 On the uniform quasi-Hadamard differentiability of the compound distribution functional 98

5.4.3 On the asymptotic distribution of the estimated compound distribution function 109

5.4.4 Proof of Theorem 5.2.1 115

Appendices **117**

A The Panjer recursion **119**

A.1 On the computation of $\widehat{\mu}_u^{*n}$ and $\mathcal{R}_\rho(\widehat{\mu}_u^{*n})$ 119

A.2 On the computation of $\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}$ and $\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})$ 120

B A proof of the $(\max\{d_{\mathrm{Wass}_p}, |\cdot|\}, d_{\mathrm{Wass}_p})$-continuity of the mapping $\mathcal{M}_1(L^p) \times (0,\infty) \to \mathcal{M}_1(L^p)$, $(\mu,\lambda) \mapsto \mu^{*\mathrm{Poiss}_\lambda}$ **121**

C A nonuniform Berry-Esséen inequality **123**

Abstract

This thesis is devoted to the nonparametric estimation of risk measures against the background of the determination of insurance premiums. We will discuss two approaches in this context. In the first part we will assume the ratio between the collective size and the observation size to be asymptotically constant, whereas in the second part we will assume the collective size to be constant and the observations size to tend to infinity.

The goal of this thesis is to determine strong rates and asymptotic distributions of the deviation of the estimated premiums from the true ones. Furthermore we will discuss bootstrap methods and their applicability to the premium estimation. Our particular attention will be paid to prove consistency, as well as almost sure bootstrap consistency for the sequence of estimated premiums. To this end, we will highlight several options how to choose suitable estimators in the individual model, as well as the collective model of insurance mathematics. The performance of these estimators will then be assessed with the help of a numerical simulation.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit der nichtparametrischen Schätzung von Risikomaßen vor dem Hintergrund der Bestimmung von Versicherungsprämien. Hierbei werden zwei Ansätze näher beleuchtet. Im ersten Teil wird angenommen, dass das Verhältnis der Beobachtungsgröße zur Kollektivgröße asymptotisch konstant ist, während im zweiten Teil der Arbeit die Kollektivgröße als konstant angenommen wird und die Beobachtungsgröße wächst.

Das Ziel dieser Arbeit besteht darin starke Raten und asymptotische Verteilungen der Abweichungen der geschätzten Prämie von der tatsächlichen herzuleiten. Des Weiteren beschäftigen wir uns mit Bootstrap-Methoden und deren Anwendbarkeit auf die Prämienschätzung. Ein besonderes Augenmerk liegt dabei darauf Konsistenz, sowie fast sichere Bootstrap-Konsistenz für die Folge der geschätzten Prämien zu zeigen. Wir werden hierzu in den beiden gängigen Modellen der Versicherungsmathematik, dem individuellen und dem kollektiven Modell, Möglichkeiten zur Wahl nichtparametrischer Schätzer angeben und deren Performanz anschließend anhand numerischer Simulationen überprüfen.

Introduction

The aim of this thesis is an investigation of several techniques to estimate premiums in an insurance collective. Given a collective consisting of a certain number of independent, homogeneous risks, the goal of an insurer is to determine a suitable premium to hedge the risk of a financial loss. That is, on the one hand the insurance company wants to impose a certain amount of money to its clients to be able to pay for future claims. On the other hand the premium imposed on the clients should not be too high in order to keep the price competitive. The basic idea in insurance mathematics is the so-called balancing of risks. Roughly speaking, this means that the expected individual risk in an insurance collective increases much slower than the number of clients in a collective. This thesis is therefore devoted to a characterization of the asymptotics of the exact premiums in relation to the estimated premiums if the number of clients in the insurance collective or the number of collected historical observations tends to infinity. We will approach this question in the first part of this thesis. The second part will then be devoted to a premium estimation based on a constant collective size, whereas the number of observations tends to infinity.

In this context different questions have to be dealt with. One obvious question is how to estimate the claim amount which will arise in the future. Based on formerly observed claims the insurer should think about an appropriate estimation of future claim amounts. In this context, the insurance company could be interested in the distribution of the deviation of the estimated premium from the true premium. Here it is important to know how the error in the estimation evolves in dependence on the underlying single claim distribution, the collective size or the number of observations taken into account, for instance. Another question is the choice of a suitable risk measure. Roughly speaking, the risk measure provides a tool to "map" the riskiness of the insurance collective to a suitable premium, which will then be imposed on the whole collective. However, the choice of the risk measure is a nontrivial task. Finally, the choice of the insurance model of course has a huge impact.

In actuarial practice there are two popular models. The first one is the so-called individual model of insurance mathematics and the second one is called the collective model of insurance mathematics. The idea behind the individual model is as follows. Assume an insurer is faced with an insurance collective consisting of $n \in \mathbb{N}$ independent, homogeneous risks. We will identify each of these risks with a real-valued nonnegative random variable. To this end, let $X_1, \ldots, X_n$ be nonnegative independent and identically distributed (i.i.d.) random variables on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with unknown distribution μ. That is, each of the X_i's corresponds to a claim amount being reported by the i-th client throughout the insurance period. These single claim amounts are either equal to zero, if the client does not report a claim at all, or strictly positive in case of a "true" claim. In this context, the total claim amount which will arise within the next insurance period is given by the random sum

$$S_n := \sum_{i=1}^{n} X_i \,.$$

The distribution of S_n is then given by the n-fold convolution of μ, which we will denote by μ^{*n}. Using this, an evaluation of an adequate risk measure $\mathcal{R}$ at μ^{*n} would provide a suitable premium for the whole collective. Examples 1.2.1–1.2.6 in Chapter 1 will give an overview over some popular risk measures. A "fair" premium for a single client in the collective would then be given by $\mathcal{R}(\mu^{*n})/n$. In this case the insurance company would equally distribute the total premium onto each of the n clients, such that every client would have to pay the same amount of money. In the following we will call this the individual premium. By so-called balancing of risks in large collectives, we observe that $\mathcal{R}(\mu^{*n})/n$ is often essentially smaller than $\mathcal{R}(\mu)$.

Of course in real actuarial practice a rather large number of clients would not report a claim at all within a certain insurance period. This is reflected by the fact that a rather large number of the random variables X_i would be 0, whereas the remaining "true" claims would take strictly positive values. That is, we can think of the distribution μ of a single claim as a compound distribution in the following way. Denoting by $p \in (0,1)$ the probability of an actually positive claim to arise, and by $\widetilde{\mu}$ a distribution possessing only mass on the positive real axis, the single claim distribution has a representation as

$$\mu := (1-p)\,\delta_0 + p\,\widetilde{\mu}. \tag{1}$$

We can think of $\widetilde{\mu}$ as a distribution of a claim conditional on its positiveness, that is $\widetilde{\mu}[\,\cdot\,] := \mu[\,\cdot \cap (0,\infty)]/\mu[(0,\infty)]$. An advantage of the individual model lies in the fact that the number of claims is equal to the collective size n. Given an estimator for the true single claim distribution μ, this makes the computability of the total claim distribution, which is given by the n-fold convolution of the single claim distribution, easier to handle. However, even in this case an exact computation of the convolution is more or less impossible.

The representation in (1) implies that the probability for a strictly positive claim to happen is constant. In contrast to this assertion, the collective model assumes the number of claims to be an integer-valued random variable, such that the whole collective is regarded as the "producer" of claims rather than the individual client itself. More explicitly, we let (X_i) be a sequence of positive i.i.d. random variables with distribution μ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Each of the X_i's represents a strictly positive single claim and corresponds to the so-called "true" single claims in the setting of the individual model. In this case the single claim distribution μ only possesses mass on the interval $(0,\infty)$ and does not have point mass in 0. Assuming that the number of positive claims is independent of the actual claim size, we let N be an integer-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ as well, being independent of the sequence (X_i). We identify with $X_1, \dots, X_N$ the claim sizes being reported throughout an insurance period. Here we stress the fact again, that both the claim sizes and the number of claims are random. Then the total claim size is given by

$$S_N := \sum_{i=1}^{N} X_i \,.$$

The distribution of the total claim size is then given by the convolution of the single claim distribution with respect to the distribution of the number of claims. We will call this a random convolution. A rigorous definition of the random convolution will follow in Chapter 3. Having noted that the n-fold convolution of a measure μ is usually very hard to compute, the computation of the random convolution is almost impossible. However, for suitable choices of risk measures, there are ways to compute or at least approximate the corresponding premium in finite time.

In the literature several approaches about the estimation of risk measures have already been discussed. Recent work concerning the estimation of asymptotic distributions of plug-in estimators of certain law-invariant risk measures has been done by [9], [51] and [60], for instance. Moreover, the functional delta-method in [10] provided a tool to directly derive the asymptotic distribution of statistical functional, such as risk functionals, to the weak convergence of the underlying empirical process. An refinement of this method can be found in [38].

Another problem the insurer is faced with, is the aforementioned problem of an estimation of the future claim size distribution based on historically collected data. Throughout this thesis we will mainly discuss approaches motivated by the Glivenko-Cantelli Theorem and the Central Limit Theorem. The Glivenko-Cantelli Theorem, for instance, proves the strong consistency of the empirical measure with respect to the unknown distribution μ. Thus, in our setting the empirical measure could be used to estimate the underlying single claim distribution μ on the basis of historically observed single claims. Based on this estimation one could then try to compute the convolution of the empirical measure, which is either the n-fold convolution in the individual model, or the random convolution in the collective model, to obtain an estimator for the total claim distribution. A suitable premium would then be given by the risk measure evaluated at the estimator for the total claim distribution. As mentioned before, the difficulty lies in the computability of the convolutions, which makes this approach often quite unhandy.

Another possibility to estimate the total claim distribution is to use the asymptotic normality of a suitably centered sum of random variables (always assuming that the second moments of the underlying single claim distribution exist). This approach is motivated by the Central Limit Theorem. The computation of this estimator simply boils down a computation of the empirical mean and the empirical variance of the historical observations and is thus quite simple. The corresponding premium is then given by an evaluation of the risk measure at the normal distribution with estimated parameters. A detailed discussion about the benefits of both estimators will follow in Chapter 2.

To carry out our estimations, we will discuss two different approaches throughout this thesis. The first part of this thesis is concerned with an estimation based on a varying collective size $n \in \mathbb{N}$, but only taking into account the last $u_n \in \mathbb{N}$ observations. Throughout the first part of this thesis we will therefore assume that the collective size n and the number of observations u_n used in our estimations fulfill

$$\lim_{n\to\infty} u_n/n = c, \tag{2}$$

for some integer-valued constant $c \in (0, \infty)$. This approach is motivated by the fact, that in many actuarial applications insurance companies only use collected observations based on the last few insurance periods. This setting was first considered by [39]. In actuarial practice, the insurance company only takes into account data from the last insurance period or data from the last three insurance periods. Consequently one would choose $c = 1$ or $c = 3$, respectively. This restriction of our estimation to observation sizes of the same "dimension" as the collective sizes makes our theory nonstandard. There are numerous examples giving a justification for this approach. One could, for instance, think of applications in the field of car insurances. Here, fast variations of car models or security systems for example make it important to use actual data for the estimation of future claim distributions. Again it is important to stress the fact that this restriction provides a new approach in contrast to the existing literature about constant collective sizes and increasing observation horizons, as in [52] for instance. We will focus on estimations of the individual premium in both the individual and the collective model. More explicitly, we are interested in the convergence of the deviation of the estimated individual premium from the true one. Our interest will lie on the determination of strong rates of the error in estimations and proving asymptotic normality of the error distribution in dependence on the collective size.

The second part of this thesis will then be concerned with new results for premium calculations based on constant collective sizes and increasing observation numbers in the context of the collective model. In the second part, we will therefore assume that the collective size $n \in \mathbb{N}$ is constant, whereas the number of observations $u \in \mathbb{N}$ tends to infinity. We will deal with this question in a semiparametric setting, which will be a generalization of the results in [12]. That is, we will assume knowledge about the class of distributions the distribution of the number of claims belongs to, such that the estimation of the distribution of the number of claims is a parametric one. At the same time we will use the standard nonparametric estimator to estimate the claim size distribution. Hence, the estimation of the random convolution is a combination of a parametric and a nonparametric approach.

The rest of this thesis is organized as follows. We will first provide some background in the field of risk measures. In Chapter 1 we will therefore give a rigorous definition of a risk measure and state some popular examples of risk measures from actuarial practice. In

Chapter 2 we will formulate the problem of the estimation of individual premiums in the individual model in the mathematical context and introduce two important estimators for the total claim size distribution. As we have mentioned before, we will approach this problem under the restriction of increasing collective sizes n and restricted observation sizes u_n, such that condition (2) holds true. In Section 2.2 we will then present results on the strong consistency for both the estimator based on the normal approximation and the convolution of the empirical measure. Moreover we will prove asymptotic normality of the error in the estimation for both estimators. The central tool which will be used to prove our assertions will be a nonuniform Berry-Essén inequality by [46]. In Section 2.3 we will then assess the performance of both estimators with the help of a numerical simulation. These simulations will show that both estimators are subject to a negative bias with respect to the true individual premium. The size of the bias is strongly affected by the heaviness of the tails of the underlying distribution $\widetilde{\mu}$ of strictly positive single claims.

Motivated by the results of Chapter 2, we will try to perform a bias-correction of the individual premium estimators using bootstrap methods. The goal of this chapter will be to establish a procedure to hopefully alleviate the bias in the former estimations. We will first give a definition of the bootstrap in our present setting and will then introduce the corresponding bootstrap estimators. In Section 2.4.2 we will then prove almost sure bootstrap consistency for our bootstrap estimators. Again, the proofs will strongly rely on the nonuniform Berry-Essén inequality by [46]. Subsequent to this Section we will carry out a numerical simulation to point out the performance of the bootstrap estimators in contrast to the original ones and discuss the benefits of this method. However, our investigations have shown that the benefits of the bootstrap-based bias correction are rather small compared to the increasing computation time and a higher mean squared error in the estimation.

Chapter 3 is devoted to the estimation of individual premiums against the background of the collective model. We will first discuss this issue in the setting of the compound Poisson model to serve a motivating example and then formulate our problem in a more general setting. Again, we are interested in strong consistency and asymptotic normality of the individual premium estimator. In Section 2.1 we will introduce our estimators. In Section 3.2 we will then prove asymptotic normality and strong consistency of the individual premium estimator in the compound Poisson model with only mild assumptions on the underlying distributions and a wide class of risk measures. For the proof we will rely on a new Berry-Essén inequality for non-randomly centered random sums by [20].

In Chapters 4 and 5 we will turn our focus to the estimation of premiums in the semiparametric setting for the case of constant collective sizes and increasing observation numbers. In contrast to our considerations in the first part, we will assume the collective size n to be constant and the observation size u to tend to infinity. We will consider the setting of the compound Poisson model again. However, many results are not restricted to the Poisson case and turn out to be valid in more general settings. In Chapter 4 we will present a summary of existing results about asymptotic normality and almost sure bootstrap consistency for estimated premiums in the individual model. These results are already known from [12]. Chapter 5 is then devoted to the derivation of the asymptotic distribution of estimated premiums in the collective model. The latter is the asymptotic distribution of the deviation of the total premium estimator from the true total premium when the observation size u tends to infinity. The goal of this chapter is to prove asymptotic normality of the total premium estimator in the compound Poisson model. The central tool which will be used to approach this problem is a special functional delta-method in the form of [12]. We will use this delta-method to derive the asymptotic distribution from a sequence of suitably chosen plug-in estimators with respect to the risk measure $\mathcal{R}$ from the asymptotic distribution of the underlying sequence of estimators. That is, we will introduce our estimators in the semiparametric setting, serving as our sequence of initial estimators and will determine the asymptotic distribution of this sequence with the help of the aforementioned delta-method. We will then present an example pointing out the practical use of our results. Finally we will give an outlook on the almost sure bootstrap consistency of the sequence of estimated premiums in Section 5.3.

The main results of the first part of this thesis are already published in two articles, jointly with Henryk Zähle. The results of Sections 2.1–2.3 are based on [41]:

Alexandra Lauer and Henryk Zähle (2015). Nonparametric estimation of risk measures of collective risks. *Statistics and Risk Modeling*, 32(2), 89–102.

Sections 2.4–2.5 are based on [42]:

Alexandra Lauer and Henryk Zähle (2017). Bootstrap consistency and bias correction in the nonparametric estimation of risk measures of collective risks. *Insurance: Mathematics and Economics*, 74, 99–108.

Chapter 1

Risk measures and risk functionals

This chapter gives a brief introduction into the theory of risk measures. The question about the determination of an adequate premium for a collective of individual risks is directly connected to the investigation of risk measures. In the following we will therefore introduce some popular risk measures which will be used throughout this thesis and summarize some basic properties.

1.1 Definitions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space. Let $\mathcal{X} \subset L^0$ be a vector space containing the constants, where $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ denotes the usual space of all finitely-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ modulo the equivalence of almost sure identity. An intrinsic example for $\mathcal{X}$ is the space $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ (consisting of all p-fold integrable random variables from L^0) for $p \geq 1$. We will say that a map $\rho : \mathcal{X} \to \mathbb{R}$ is

(1) *monotone* if $\rho(X_1) \leq \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$ with $X_1 \leq X_2$.

(2) *cash additive* if $\rho(X + m) = \rho(X) + m$ for all $X \in \mathcal{X}$ and $m \in \mathbb{R}$.

(3) *subadditive* if $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$.

(4) *positively homogeneous* if $\rho(\lambda X) = \lambda\rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \geq 0$.

In the sense of [27], we will call ρ a *monetary* risk measure if conditions (1) and (2) hold, and based on the ideas of [5] we will say that ρ is a *coherent* risk measure if conditions (1)–(4) are fulfilled. Furthermore we will call ρ a *law-invariant* risk measure, if $\rho(X) = \rho(Y)$ whenever X and Y have the same law. We will restrict ourselves to law-invariant maps $\rho : \mathcal{X} \to \mathbb{R}$. So we may and do associate with ρ a statistical functional $\mathcal{R}_\rho : \mathcal{M}(\mathcal{X}) \to \mathbb{R}$ via

$$\mathcal{R}_\rho(\mu) := \rho(X_\mu), \qquad \mu \in \mathcal{M}(\mathcal{X}), \tag{1.1}$$

where $\mathcal{M}(\mathcal{X})$ denotes the set of the distributions of the elements of $\mathcal{X}$, and $X_\mu \in \mathcal{X}$ has distribution μ.

1.2 Examples of popular risk measures

The first example introduces the risk measure based on the expected value of the future claims. The premium connected to this risk measure is also called the net premium.

Example 1.2.1 *The net premium is the premium derived from the risk measure $\rho : L^1 \to \mathbb{R}$, which is defined by*

$$\rho(X) := \mathbb{E}[X].$$

It is easily seen to be a law-invariant and coherent risk measure.

Example 1.2.2 *The premium based on the standard deviation principle is the premium, derived from the risk measure $\rho : L^2 \to \mathbb{R}$, which is defined by*

$$\rho(X) := \mathbb{E}[X] + c\sqrt{\mathbb{V}\mathrm{ar}[X]},$$

for any $c > 0$. One can easily check, that ρ is cash additive, subadditive and positively homogeneous but lacks monotonicity. In this context, the constant $c > 0$ is often referred to as the safety loading.

The next example introduces one of the most popular risk measures in practice, which is the *Value at Risk* at level $\alpha \in (0, 1)$. It is the lower α-quantile of the distribution function associated with the claims. In the practical context, the Value at Risk is the amount of money the insurer has to impose as a premium, which will guarantee for being spared a financial loss with probability α. More explicitly, the probability for the future claims to exceed the premium would be at most $1 - \alpha$.

Example 1.2.3 *The* Value at Risk *at level $\alpha \in (0, 1)$ is the map $\mathrm{V@R}_\alpha : L^0 \to \mathbb{R}$ defined by*

$$\mathrm{V@R}_\alpha(X) := F_X^{\leftarrow}(\alpha) := \inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}.$$

It is clearly law-invariant and it was shown in [5, Section 3] that it satisfies monotonicity, cash additivity, and positive homogeneity. However in general, the Value at Risk is not subadditive.

The next example introduces the so-called *Average Value at Risk* at level $\alpha \in (0, 1)$. It is often also referred to as the *Tail-conditional expectation* or the *Expected Shortfall*. In contrast to the Value at Risk of Example 1.2.3, the Average Value at Risk also takes into account the expected claim amount given the exceedence of the Value at Risk.

Example 1.2.4 *The* Average Value at Risk *at level* $\alpha \in (0,1)$ *is the map* $\mathrm{AV@R}_\alpha : L^1 \to \mathbb{R}$ *defined by*

$$\mathrm{AV@R}_\alpha(X) := \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{V@R}_s(X)\, ds.$$

Note, that if the distribution function F_X *of* X *is continuous at* $\mathrm{V@R}_\alpha(X)$*, then*

$$\mathrm{AV@R}_\alpha(X) = \mathbb{E}[X \,|\, X \geq \mathrm{V@R}_\alpha(X)].$$

It is easily seen to be law-invariant and it was shown in [1] that $\mathrm{AV@R}_\alpha$ *is a coherent risk-measure for every* $\alpha \in (0,1)$*.*

Example 1.2.5 *The* one-sided p-th moment-based risk measure *for* $p \in [1,\infty)$ *and* $a \in [0,1]$ *is the map* $\mathrm{OsM}_{p,a} : L^p \to \mathbb{R}$ *defined by*

$$\mathrm{OsM}_{p,a}(X) := \mathbb{E}[X] + a\, \mathbb{E}[((X - \mathbb{E}[X])^+)^p]^{1/p}.$$

It is clearly law-invariant. It was shown in Lemma 4.1 in [26] that the one-sided p-th moment-based risk measure provides a coherent risk measure.

Example 1.2.6 *The* expectiles-based risk measure *at level* $\alpha \in [1/2,1)$ *is the map* $\mathrm{Ept}_\alpha : L^2 \to \mathbb{R}$ *defined by*

$$\mathrm{Ept}_\alpha(X) := \mathrm{argmin}_{m\in\mathbb{R}} \{\alpha \|(X-m)^+\|_2^2 + (1-\alpha)\|(m-X)^+\|_2^2\}.$$

The expectiles-based risk measure is easily seen to be law-invariant. It has been shown in [8] that the expectiles-based risk measure provides a coherent risk measure.

1.3 Distortion risk measures and the Kusuoka representation

In this section we will recall the definition of a distortion risk measure. The following discussion is basically based on the ideas presented in [38].

Let $g : [0,1] \to [0,1]$ be a distortion function, that is a nondecreasing càdlàg function with $g(0) = 0$ and $g(1) = 1$. The *distortion risk measure* associated with g is then defined by

$$\rho_g(X) := -\int_{-\infty}^0 g(F_X(x))\, dx + \int_0^\infty (1 - g(F_X(x)))\, dx \qquad (1.2)$$

for every real-valued random variable X (on some given atomless probability space) satisfying $\int_0^\infty (1 - g(F_{|X|}(x))dx < \infty$, where F_X and $F_{|X|}$ denote the distribution functions of X and $|X|$, respectively. The set $\mathcal{X}_g \subset L^0$ of all such random variables forms a linear subspace of L^1; this follows from [19, Proposition 9.5] and [27, Proposition 4.75]. It is known that ρ_g is a law-invariant coherent risk measure if and only if the distortion function g is convex; see, for instance, [63].

If specifically $g(t) = \frac{1}{1-\alpha} \max\{1-\alpha, 0\}$ for some $\alpha \in (0,1)$, then we have $\mathcal{X} = L^1(\Omega, \mathcal{F}, \mathbb{P})$ and ρ_g is nothing but the Average Value at Risk at level α as introduced in Example 1.2.4. In [38] it was shown that any law-invariant coherent risk measure ρ admits the following so-called *Kusuoka representation*

$$\rho(X) = \sup_{g \in \mathcal{G}} \rho_g(X), \tag{1.3}$$

where $\mathcal{G}$ is a class of distortion functions and ρ_g denotes the distortion risk measure (1.2) for $g \in \mathcal{G}$. The one-sided p-th moment-based risk measure of Example 1.2.5 and the expectiles-based risk measure of Example 1.2.6 are examples of risk measures of the form (1.3) but not of the form (1.2). Indeed, Lemma A.5 in [39] has shown that the one-sided p-th moment-based risk measure does not provide a distortion risk measure. It follows from Lemma 8 in [17] that the expectiles-based risk measure is also not a distortion risk measure unless $\alpha = 1/2$. This points out that the distortion representation is not a necessary condition for a risk measure to be coherent.

1.4 Risk functionals and regularity properties

We will first focus on regularity properties based on the (L^1)-Wasserstein metric d_{Wass}. We will state assumptions under which certain risk measures are continuous or even β-Hölder continuous for some $\beta > 0$ w.r.t. d_{Wass}. In the second part of this section we will then derive corresponding properties w.r.t. the nonuniform Kolmogorov distance d_{ϕ_λ}, which will be introduced below.

Let $\mathcal{M}_1$ be the set of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and denote by F_μ the distribution function of $\mu \in \mathcal{M}_1$. For every $\lambda \geq 0$, let the function $\phi_\lambda : \mathbb{R} \to [1, \infty)$ be defined by $\phi_\lambda(x) := (1 + |x|^\lambda)$, $x \in \mathbb{R}$. For $\mu_1, \mu_2 \in \mathcal{M}_1$, we say that

$$d_{\phi_\lambda}(\mu_1, \mu_2) := \sup_{x \in \mathbb{R}} |F_{\mu_1}(x) - F_{\mu_2}(x)| \, \phi_\lambda(x) \tag{1.4}$$

is the nonuniform Kolmogorov distance of μ_1 and μ_2 w.r.t. the weight function ϕ_λ. It is easily seen that d_{ϕ_λ} provides a metric on the set $\mathcal{M}_1^\lambda$ of all $\mu \in \mathcal{M}_1$ satisfying $d_{\phi_\lambda}(\mu, \delta_0) < \infty$.

The (L^1)-Wasserstein distance on $\mathcal{M}_1^1$ is defined by

$$d_{\mathrm{Wass}}(\mu_1, \mu_2) := \int_{-\infty}^{\infty} |F_{\mu_1}(x) - F_{\mu_2}(x)| \, dx, \tag{1.5}$$

where F_{μ_1} and F_{μ_2} denote the distribution functions of μ_1 and μ_2, respectively. Lemma 8.1 in [13] has shown that d_{Wass} indeed defines a metric on $\mathcal{M}_1^1$. Moreover, it was shown in Proposition 4 in [29] that the metric d_{Wass} induces the L^1-weak topology. The latter is defined to be the coarsest topology on $\mathcal{M}(L^1)$ w.r.t. which each of the maps $\mu \mapsto \int f\, d\mu$, $f \in C_{\mathrm{b}}^1$, is continuous, where C_{b}^1 is the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which there exists a constant $C > 0$ such that $|f(x)| \leq C|x|$ for all $x \in \mathbb{R}$.

The following result is already known from Theorem 2.8 and Remark 2.9 in [37].

Theorem 1.4.1 *For any law-invariant coherent risk measure ρ on L^1, the associated risk functional $\mathcal{R}_\rho : \mathcal{M}(L^1) \to \mathbb{R}$ is continuous w.r.t. the Wasserstein metric d_{Wass}.*

The following theorem will state a sufficient condition under which a statistical functional $\mathcal{R}_{\rho_g}$ associated with a convex distortion function g is not only continuous, but Lipschitz continuous w.r.t. the Wasserstein metric d_{Wass}.

Theorem 1.4.2 *Let $\rho_g : \mathcal{X}_g \to \mathbb{R}$ be the distortion risk measure associated with a convex distortion function g. Assume that there exists a constant $L > 0$ such that*

$$1 - g(t) \leq L(1 - t) \quad \text{for all } t \in [0, 1]. \tag{1.6}$$

Then the statistical functional $\mathcal{R}_{\rho_g}$ associated with ρ_g is Lipschitz-continuous w.r.t. d_{Wass}. That is, for every $\mu_1, \mu_2 \in \mathcal{M}_1^1$ we have

$$|\mathcal{R}_{\rho_g}(\mu_1) - \mathcal{R}_{\rho_g}(\mu_2)| \leq L\, d_{\mathrm{Wass}}(\mu_1, \mu_2). \tag{1.7}$$

Proof Condition (1.6) and the convexity of the distortion function g together imply $|g(t) - g(t')| \leq L|t - t'|$ for all $t, t' \in [0, 1]$. Hence, using (1.2), we observe that

$$\begin{aligned} |\mathcal{R}_{\rho_g}(\mu_1) - \mathcal{R}_{\rho_g}(\mu_2)| &\leq \int_{-\infty}^{\infty} |g(F_{\mu_1}(x)) - g(F_{\mu_2}(x))|\, dx \\ &\leq L \int_{-\infty}^{\infty} |F_{\mu_1}(x) - F_{\mu_2}(x)|\, dx \\ &= L\, d_{\mathrm{Wass}}(\mu_1, \mu_2). \end{aligned} \tag{1.8}$$

This leads to the assertion. □

Choosing $g(t) := \mathbb{1}_{[\alpha,1]}(t)$ for any fixed $\alpha \in (0, 1)$, the corresponding distortion risk measure is nothing but the Value at Risk at level $\alpha \in (0, 1)$, as defined in Example 1.2.3. It is easily seen, that in this case g does not fulfill condition (1.6) in Theorem 1.4.2. However, the Value at Risk at level $\alpha \in (0, 1)$ is weakly continuous at every α such that the distribution function of the underlying random variable X takes the value α only once, see for instance [62], Lemma 21.2.

The following Theorem is basically already known from Lemma 2.14 in [39].

Theorem 1.4.3 *Let $p \geq 1$. Let $\rho : L^p \to \mathbb{R}$ be a law-invariant coherent risk measure and define a function $g_\rho : [0, 1] \to [0, 1]$ by $g_\rho(t) := 1 - \rho(B_{1-t})$, where B_{1-t} refers to any*

Bernoulli random variable with expectation $1-t$. *Assume that there exist constants* $L, \beta > 0$ *such that*

$$1 - g_\rho(t) \leq L(1-t)^\beta \quad \text{for all } t \in [0,1]. \tag{1.9}$$

Let $\lambda > 0$, *such that* $\lambda\beta > 1$. *Then the statistical functional* $\mathcal{R}_\rho$ *associated with* ρ *is* β*-Hölder continuous w.r.t.* d_{ϕ_λ}. *That is, for every* $\mu_1, \mu_2 \in \mathcal{M}_1^\lambda$ *there exists a constant* $C > 0$ *such that*

$$|\mathcal{R}_\rho(\mu_1) - \mathcal{R}_\rho(\mu_2)| \leq C d_{\phi_\lambda}(\mu_1, \mu_2)^\beta. \tag{1.10}$$

Proof Since ρ is defined on L^p, we can find a set $\mathcal{G}_\rho$ of continuous convex distortion functions such that $g_\rho = \inf_{g \in \mathcal{G}_\rho} g$ and

$$\rho(X) = \sup_{g \in \mathcal{G}_\rho} \rho_g(X) \quad \text{for all } X \in L^p. \tag{1.11}$$

This follows from Proposition 5.1 and Remark 3.2 in [9] (adapted to our definition of monotonicity and cash additivity); see also [39, 38]. Below we will show that (1.9) implies

$$|g(t) - g(t')| \leq L|t - t'|^\beta \quad \text{for all } t, t' \in [0,1] \text{ and } g \in \mathcal{G}_\rho. \tag{1.12}$$

With the help of (1.11) and (1.12) we then obtain

$$\begin{aligned} |\mathcal{R}_\rho(\mu_1) - \mathcal{R}_\rho(\mu_2)| &= \Big|\sup_{g \in \mathcal{G}_\rho} \mathcal{R}_{\rho_g}(\mu_1) - \sup_{g \in \mathcal{G}_\rho} \mathcal{R}_{\rho_g}(\mu_2)\Big| \\ &\leq \sup_{g \in \mathcal{G}_\rho} |\mathcal{R}_{\rho_g}(\mu_1) - \mathcal{R}_{\rho_g}(\mu_2)| \\ &\leq \sup_{g \in \mathcal{G}_\rho} \int_{-\infty}^{\infty} |g(F_{\mu_1}(x)) - g(F_{\mu_2}(x)| \, dx \\ &\leq \int_{-\infty}^{\infty} L|F_{\mu_1}(x) - F_{\mu_2}(x)|^\beta \, dx \\ &\leq C \, d_{\phi_\lambda}(\mu_1, \mu_2)^\beta \end{aligned}$$

for the constant $C := L \int_{-\infty}^{\infty} 1/\phi_\lambda(x)^\beta dx$ (which is finite due to the assumption $\lambda\beta > 1$). That is, the assertion of part (ii) holds true, too.

It remains to show (1.12), for which we will adapt the arguments of Section 4.3 in [39]. Let $0 \leq t < t' < 1$. Since the underlying probability space was assumed to be atomless, we may pick a measurable decomposition $A_1 \cup A_2 \cup A_3$ of the probability domain such that $\mathrm{P}[A_1] = 1 - t'$, $\mathrm{P}[A_2] = t' - t$ and $\mathrm{P}[A_3] = t$, where P refers to the corresponding probability measure. Define random variables $B_{1-t'} := \mathbb{1}_{A_1}$, $B_{1-t} := \mathbb{1}_{A_1 \cup A_2}$ and $B_{t'-t} := \mathbb{1}_{A_2}$, and note that they are distributed according to the Bernoulli distribution with parameters $1-t'$, $1-t$ and $t'-t$, respectively. Moreover we clearly have $B_{1-t} = B_{1-t'} + B_{t'-t}$. By the subadditivity of ρ_g we can conclude $\rho(B_{1-t}) \leq \rho(B_{1-t'}) + \rho(B_{t'-t})$, and so

$$\begin{aligned} g(t') - g(t) &= 1 - \rho_g(B_{1-t'}) - (1 - \rho_g(B_{1-t})) \\ &\leq \rho_g(B_{t'-t}) \end{aligned}$$

$$
\begin{aligned}
&\leq \rho(B_{t'-t}) \\
&\leq \sup_{u\in(0,1]} \frac{\rho(B_u)}{u^\beta}\,(t'-t)^\beta \\
&\leq \sup_{u\in(0,1]} \frac{1-g_\rho(1-u)}{u^\beta}\,(t'-t)^\beta \\
&\leq \sup_{v\in[0,1)} \frac{1-g_\rho(v)}{(1-v)^\beta}\,(t'-t)^\beta
\end{aligned}
$$

for every $g \in \mathcal{G}_\rho$, where the second "$\leq$" is ensured by (1.11). By (1.9) we observe that

$$\sup_{v\in[0,1)} \frac{1-g_\rho(B_v)}{(1-v)^\beta} < \infty.$$

Thus, since every $g \in \mathcal{G}_\rho$ is also continuous at 1, condition (1.9) indeed implies (1.12). □

If ρ is the one-sided p-th moment based risk measure for some $p \in [1,\infty)$ and $a \in [0,1]$, as introduced in Example 1.2.5, then condition (1.9) implies that ρ is β-Hölder continuous w.r.t. d_{ϕ_λ} for $\beta = 1/p$ and every $\lambda > p$.

Theorem 1.4.4 below is about the special case when ρ refers to a distortion risk measure. It is worth mentioning that if ρ is a distortion risk measure with distortion function g, then $g_\rho = g$ and condition (1.9) boils down to the assumption on the β-Hölder continuity of the distortion function g as in Theorem 1.4.4.

Theorem 1.4.4 *Let $\rho_g : \mathcal{X}_g \to \mathbb{R}$ be the distortion risk measure associated with a distortion function g. Moreover, assume that g is β-Hölder continuous for some $\beta > 0$.*

Let $\lambda > 0$, such that $\lambda\beta > 1$. Then the statistical functional $\mathcal{R}_{\rho_g}$ associated with ρ_g is β-Hölder continuous w.r.t. d_{ϕ_λ}. That is, for every $\mu_1, \mu_2 \in \mathcal{M}_1^\lambda$ there exists a constant $C > 0$ such that

$$|\mathcal{R}_{\rho_g}(\mu_1) - \mathcal{R}_{\rho_g}(\mu_2)| \leq C d_{\phi_\lambda}(\mu_1,\mu_2)^\beta. \tag{1.13}$$

Proof The β-Hölder continuity of the distortion function g implies that there exists some constant $L > 0$, such that $|g(t) - g(t')| \leq L|t-t'|^\beta$ for all $t, t' \in [0,1]$. Hence, using (1.2), we observe that

$$
\begin{aligned}
|\mathcal{R}_{\rho_g}(\mu_1) - \mathcal{R}_{\rho_g}(\mu_2)| &\leq \int_{-\infty}^{\infty} |g(F_{\mu_1}(x)) - g(F_{\mu_2}(x))|\,dx \\
&\leq L \int_{-\infty}^{\infty} |F_{\mu_1}(x) - F_{\mu_2}(x)|^\beta\,dx \\
&\leq C\, d_{\phi_\lambda}(\mu_1,\mu_2)^\beta, \qquad (1.14)
\end{aligned}
$$

with $C := L\int_{-\infty}^{\infty} 1/\phi_\lambda(x)^\beta\,dx$. In view of the assumption $\lambda\beta > 1$ it follows that the latter integral is finite. □

For the *Average Value at Risk* $\mathrm{AV@R}_\alpha$ at level α of Example 1.2.4 the assumption of Theorem 1.4.4 on the β-Hölder continuity of the distortion function g holds for $\beta = 1$. That is, for $\mathrm{AV@R}_\alpha$ the assertion of Theorem 1.4.4 hold true for every $\lambda > 1$.

Part I

Estimation under a constant ratio of sampling size and collective size

Chapter 2

Nonparametric estimation of risk measures in the individual model

This chapter is devoted to the derivation of premiums with increasing collective sizes. More explicitly, we focus on the rate of convergence of the error in estimations of risk measures in dependence on the collective size. In the setting regarded in this chapter the collective size will also coincide with the number of observations being used in the estimation, or will at least be proportional to the number of observations. A central task will be the estimation of the distribution of the total claim size, which will then be plugged into a certain risk measure to provide an estimator for the total premium. We will deal with this topic in the individual model of insurance mathematics.

In the setting throughout this chapter we will always consider a homogeneous insurance collective, consisting of $n \in \mathbb{N}$ clients. What is meant by a "homogeneous" insurance collective will become apparent in the next section. Throughout the terms of the insurance periods the insurance company is assumed to collect historical data based on the formerly observed claims which have been reported. On one hand, much is known about the statistical estimation of the underlying single claim distribution if the number of observations tends to infinity. However, in practice it might not be sensible to process all the data, which has been collected over a very long period of time. In the following we will therefore develop an approach for statistical estimations and numerical approximations of risk measures based on a certain number u_n of historically observed claims in a collective of n clients. Here (u_n) is a sequence of positive integers for which u_n/n converges to some integer $c \in (0, \infty)$. Roughly speaking, the number of observations should be of the "same dimension" as the number of clients. This makes the presented theory nonstandard.

A justification for this approach can be found in the field of car insurances for example. An insurer who has collected data during the last few decades might however not use all this information to estimate future claim sizes. Reasons for this might be technical advances in vehicle safety systems, such as security belts, airbags etc. or because car values as such have increased throughout this time. The insurer might therefore not want to include "too old" data in his estimations for future claim developments. In practice it is therefore often sensible to just process data from the last one to three years. This is why the presented theory is of great interest.

First approaches to this theory have been made by [39] for the case of the individual model. They used the normal approximation with estimated parameters to estimate the distribution of the total claim size and presented numerical simulations comparing the normal approximation to the convolution of the empirical measure. In Section 2.1 we will elaborate similar results for the normal approximation and the convolution of the empirical measure and present numerical results on our own. The proof of our main theorems, which derive strong rates of convergence for the individual premium, strongly relies on a Berry-Esséen inequality w.r.t. the nonuniform Kolmogorov distance d_{ϕ_λ}, which was introduced in (1.4). The proof of this inequality can be found in [50]. However in Appendix C we present an alternative proof, which is of an interest of its own.

2.1 Estimators for the individual premium in the individual model

In this section we deal with the individual model of actuarial theory. In this context let (X_i) be a sequence of i.i.d. random variables with distribution μ. For every $n \in \mathbb{N}$ let

$$S_n := \sum_{i=1}^{n} X_i.$$

In this case the distribution of S_n is given by the n-fold convolution μ^{*n} of μ. In our context S_n can be seen as the total claim in a homogeneous insurance collective consisting of n individual risks, such that μ^{*n} refers to the distribution of the total claim size. The premium w.r.t. a suitable law-invariant risk measure ρ in the sense of Chapter 1 imposed by the insurance company is then given by $\mathcal{R}_\rho(\mu^{*n})$. A suitable individual premium, that is the premium every single client has to pay, is then given by

$$\mathcal{R}_n := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}). \tag{2.1}$$

In this case the total premium gets equally distributed onto every client in the collective. Here it is important to note that $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$ is in most cases essentially smaller than $\mathcal{R}_\rho(\mu)$.

In the following we want to highlight two options to estimate future claim distributions on the basis of historically observed data. To this end, let $Y_1, \dots, Y_{u_n}$ be a sequence of historically observed claims based on a collective consisting of n clients. In real applications the collective size may of course vary over several insurance periods. In the mathematical context however, this does not impose a restriction to our theory. To this end, suppose that $Y_1, \dots, Y_{u_n}$ are i.i.d. random variables on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ. In this context let (u_n) be any sequence of positive integers for which u_n/n converges to some constant $c \in (0,\infty)$. Motivated by the Central Limit Theorem for instance, one could use the asymptotic normality of the total claim size S_n and use the normal distribution with estimated parameters $\mathcal{N}_{\widehat{m}_{u_n},\widehat{s}^2_{u_n}}$ to estimate the total claim distribution μ^{*n}. In this case the corresponding plug-in estimator

$$
{}^{\mathsf{NA}}\widehat{\mathcal{R}}_n := \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{\widehat{m}_{u_n},\widehat{s}^2_{u_n}}) \tag{2.2}
$$

provides a reasonable estimator for the individual premium $\mathcal{R}_n = \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$, where $\widehat{m}_{u_n}$ and $\widehat{s}^2_{u_n}$ refer to the sample mean and the sample variance of $Y_1,\ldots,Y_{u_n}$, respectively, assuming that $Y_1,\ldots,Y_{u_n}$ have finite second moments. This approach has already been discussed in [39]. In this article it was shown that for many law-invariant risk measures ρ we have

$$
n^r \left({}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n\right) \stackrel{\text{a.s.}}{\longrightarrow} 0, \qquad n \to \infty \tag{2.3}
$$

for every $r < 1/2$, and

$$
\text{law}\left\{n^{1/2}\left({}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n\right)\right\} \stackrel{\text{w}}{\longrightarrow} \mathcal{N}_{0,\,s^2}, \qquad n \to \infty \tag{2.4}
$$

with $s^2 := \mathbb{V}\text{ar}[X_1]$. From (2.4) we can also see that the convergence in (2.3) can not hold for $r \geq 1/2$. Again, this approach is nonstandard, because the parameters are estimated on the basis of a data set of the same “dimension” as the collective size. The assumption that u_n increases to infinity at the same speed as n, reflects the idea that in practical applications the parameters are typically estimated on the basis of data from the last year or the last few years. Moreover it was shown in [39] that for the exact mean m and the exact variance s^2 of μ, and for many law-invariant coherent risk measures ρ,

$$
\sup_{n\in\mathbb{N}} \left|\mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2}) - \mathcal{R}_\rho(\mu^{*n})\right| < \infty. \tag{2.5}
$$

Both (2.3)–(2.5) and the simulation study in [39] show that the overwhelming part of the error in the estimated normal approximation of the risk functional is due to the estimation of the unknown parameters rather than to the numerical approximation itself. Whereas in the case of known parameters the relative error converges to zero at rate (nearly) 1, in the case of estimated parameters the relative error converges to zero only at rate (nearly) 1/2. So it is very important to note that statistical aspects may not be neglected when investigating approximations of premiums for aggregate risks.

An advantage of the normal approximation with estimated parameters is the fact that the corresponding premium is very easy to compute. Indeed, whenever ρ refers to a law-invariant, cash-additive and positively homogeneous risk measure, the corresponding total premium has the following representation:

$$
{}^{\mathsf{NA}}\widehat{\mathcal{R}}_n = \frac{1}{\sqrt{n}}\widehat{s}_{u_n}\mathcal{R}_\rho(\mathcal{N}_{0,1}) + \widehat{m}_{u_n}. \tag{2.6}
$$

As the normal distribution is a symmetric distribution and in practical applications claim size distributions are often skewed to the right (see also Figure 2.1), this approach might only yield a moderate applicability to a large set of actuarial tasks. Against this background it might be more sensible to choose an estimator, which takes into account the natural skewness of the claim size distributions. Motivated by the Glivenko-Cantelli Theorem for example, one could choose the empirical measure $\widehat{\mu}_{u_n}$ based on $Y_1, \ldots, Y_{u_n}$, which is given by

$$\widehat{\mu}_{u_n} := \frac{1}{u_n} \sum_{i=1}^{u_n} \delta_{Y_i}, \tag{2.7}$$

to estimate the single claim distribution μ. This is the standard choice for an estimator of the unknown distribution μ in the nonparametric setting. Following this line of reasoning

$$\widehat{\mu}_{u_n}^{*n} := (\widehat{\mu}_{u_n})^{*n} \tag{2.8}$$

provides a reasonable estimator for the total claim size distribution μ^{*n}. Thus, we can use the corresponding plug-in estimator

$$^{\mathsf{CE}}\widehat{\mathcal{R}}_n := \frac{1}{n} \mathcal{R}_\rho(\widehat{\mu}_{u_n}^{*n}) \tag{2.9}$$

as an approximation for the true premium $\mathcal{R}_n$. In the following we will refer to $^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ as the *empirical plug-in estimator*.

In general the computation of the n-fold convolution $\widehat{\mu}_{u_n}^{*n}$ of $\widehat{\mu}_{u_n}$ is more or less impossible. However, in real applications the true μ has support in $h\mathbb{N}_0 := \{0, h, 2h, \ldots\}$ for some fixed $h > 0$, where h represents the smallest monetary unit. We stress the fact that continuous distributions are in fact approximations for the *equidistant discrete* true single claim distribution, and not vice versa. So the empirical probability measure $\widehat{\mu}_{u_n}$ is concentrated on an equidistant grid $h\mathbb{N}_0$, too. In this case the estimated total claim distribution $\widehat{\mu}_{u_n}^{*n}$ can be computed with the help of the recursive scheme

$$\widehat{\mu}_{u_n}^{*n}[\{0\}] = \widehat{\mu}_{u_n}[\{0\}]^n \tag{2.10}$$

$$\widehat{\mu}_{u_n}^{*n}[\{jh\}] = \frac{1}{j\,\widehat{\mu}_{u_n}[\{0\}]} \sum_{\ell=1}^{j} ((n+1)\ell - j)\, \widehat{\mu}_{u_n}[\{\ell h\}]\, \widehat{\mu}_{u_n}^{*n}[\{(j-\ell)h\}] \quad \text{for } j \in \mathbb{N}, \tag{2.11}$$

provided $\widehat{\mu}_{u_n}[\{0\}] > 0$. In the literature this scheme is often referred to as the *Panjer recursion*, see [49]. To observe that the upper scheme indeed coincides with the Panjer recursion for the convolution w.r.t. the binomial distribution with parameters n and p, where p is the probability of a strictly positive claim, see Appendix A.1. Note that $\widehat{\mu}_{u_n}$ is the empirical probability measure and therefore has bounded support. Consequently, in view of (2.10)–(2.11), the estimator $\mathcal{R}_\rho(\widehat{\mu}_{u_n}^{*n})$ can typically be computed in finite time, even for tail-dependent functionals $\mathcal{R}_\rho$ as, for instance, the one associated with the Average Value at Risk of Example 1.2.4.

Section 2.2 will show that for a large class of risk functionals ρ, any distribution μ with a finite λ-moment for some $\lambda > 2$ and any positive sequence of integers (u_n) for which u_n/n converges to some constant $c \in (0, \infty)$, we obtain similar results as in (2.3)–(2.4) for the convolution of the empirical measure $\widehat{\mu}_{u_n}^{*n}$. More precisely, Theorem 2.2.4 will show that

$$n^r \left({}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n \right) \xrightarrow{\text{a.s.}} 0, \qquad n \to \infty \tag{2.12}$$

for every $r < 1/2$, and

$$\text{law}\left\{ n^{1/2} \left({}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n \right) \right\} \xrightarrow{\text{w}} \mathcal{N}_{0,\, s^2}, \qquad n \to \infty. \tag{2.13}$$

The results of Theorems 2.2.2 and 2.2.4 will yield even more, namely

$$ {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n \;=\; (\widehat{m}_{u_n} - m) \,+\, o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}), \tag{2.14}$$

$$ {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n \;=\; (\widehat{m}_{u_n} - m) \,+\, o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}), \tag{2.15}$$

where $o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$ refers to any sequence of random variables (ξ_n) on $(\Omega, \mathcal{F}, \mathbb{P})$ for which $\sqrt{n}\xi_n$ converges $\mathbb{P}$-a.s. to zero. What strikes the most in formulae (2.14)–(2.15) is the fact that the asymptotics of both estimators are exactly the same and are independent of the concrete choice of the risk measure ρ. Both asymptotics are purely driven by the convergence of the sample mean to the true mean. With the help of (2.4) and (2.13), we can now derive asymptotic confidence intervals at level $(1 - \alpha)$ for the individual premium:

$$\left[{}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \frac{\widehat{s}_{u_n}}{\sqrt{n}}\,\Phi_{0,1}^{-1}\Big(1 - \frac{\alpha}{2}\Big)\,,\; {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \frac{\widehat{s}_{u_n}}{\sqrt{n}}\,\Phi_{0,1}^{-1}\Big(\frac{\alpha}{2}\Big) \right]$$

and

$$\left[{}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \frac{\widehat{s}_{u_n}}{\sqrt{n}}\,\Phi_{0,1}^{-1}\Big(1 - \frac{\alpha}{2}\Big)\,,\; {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \frac{\widehat{s}_{u_n}}{\sqrt{n}}\,\Phi_{0,1}^{-1}\Big(\frac{\alpha}{2}\Big) \right],$$

where $\Phi_{0,1}$ denotes the distribution function of $\mathcal{N}_{0,1}$.

As another consequence of Theorem 2.2.4 we observe that the true individual premium always has an asymptotic representation which is similar to the one of ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ in (2.6), namely

$$\mathcal{R}_n \;=\; m \,+\, \frac{\mathcal{R}_\rho(\mathcal{N}_{0,1})}{\sqrt{n}}\; s \,+ o(n^{-1/2}). \tag{2.16}$$

The representation in (2.16) has an astonishing meaning. No matter what the risk measure ρ looks like, the individual premium w.r.t. ρ asymptotically coincides with the premium derived from the standard deviation principle of Example 1.2.2 with safety loading $\frac{1}{\sqrt{n}}\mathcal{R}_\rho(\mathcal{N}_{0,1})$. Likewise we can obtain similar representations for the corresponding estimators:

$$ {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n \;=\; \widehat{m}_{u_n} \,+\, \frac{\mathcal{R}_\rho(\mathcal{N}_{0,1})}{\sqrt{n}}\,\widehat{s}_{u_n}, \tag{2.17}$$

$$ {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n \;=\; \widehat{m}_{u_n} \,+\, \frac{\mathcal{R}_\rho(\mathcal{N}_{0,1})}{\sqrt{n}}\,\widehat{s}_{u_n} \,+\, o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{2.18}$$

Formulae (2.17)–(2.18) can be interpreted as a justification for the use of the standard deviation principle, which is widely used in insurance practice. The representations do not only justify the use of the standard deviation principle, but also make a suggestion on how to sensibly choose the safety loading. The safety loading as such depends on the concrete choice of the risk measure evaluated at the standard normal distribution and the square root of the collective size. The division by $\sqrt{n}$ in the safety loading moreover reflects the so-called balancing of risks in large collectives.

In the following Section we will formulate assumptions under which the above results can be achieved and state our main theorems.

2.2 Strong rates and asymptotic normality for the individual premium estimators in the individual model

Let $\mathcal{M}_1$ again be the set of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and denote by F_μ the distribution function of $\mu \in \mathcal{M}_1$. For every $\lambda \geq 0$, let the function $\phi_\lambda : \mathbb{R} \to [1, \infty)$ be defined by $\phi_\lambda(x) := (1 + |x|^\lambda)$, $x \in \mathbb{R}$. Recall that the nonuniform Kolmogorov distance w.r.t. the weight function ϕ_λ was introduced in (1.4).

Assumption 2.2.1 *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, and $\mathcal{R}_\rho$ be the corresponding statistical functional introduced in (1.1). Let (u_n) be a sequence in $\mathbb{N}$, and assume that the following assertions hold for some $\lambda > 2$:*

(a) $\mu \in \mathcal{M}(L^\lambda)$, that is, $\mathbb{E}[|Y_1|^\lambda] < \infty$.

(b) u_n/n converges to some constant $c \in (0, \infty)$.

(c) ρ is cash additive and positively homogeneous, and $\mathcal{M}_1^\lambda \subset \mathcal{M}(\mathcal{X})$.

(d) The restriction of $\mathcal{R}_\rho$ to $\mathcal{M}_1^\lambda$ is $(d_{\phi_\lambda}, |\cdot|)$-continuous at $\mathcal{N}_{0,1}$.

Note that part (d) of Assumption 2.2.1 does not present a strong restriction. The results of Sections 1.3 and 1.4 have shown that a large variety of risk measures satisfies the imposed condition. For instance, the Value at Risk of Example 1.2.3, as well as the Average Value at Risk of Example 1.2.4 and the one-sided p-th moments based risk measure of Example 1.2.5 are amongst the most popular examples satisfying the condition in part (d).

We are now in a position to state our main theorems. Assertions (iv)–(v) in Theorem 2.2.2 describe the asymptotic behavior of the estimator ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n = \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}})$ for the individual premium $\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$. Note that $\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}^2_{u_n}})$ is always $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable due to the representation in (2.6).

Theorem 2.2.2 (Estimated normal approximation) *Suppose that Assumption 2.2.1 holds with $\lambda > 2$. Then the following assertions hold:*

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,\, ns^2}) = (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,\, ns^2}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = o(n^{-1/2})$.

(iii) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(iv) $n^r(\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})) \longrightarrow 0$ *$\mathbb{P}$-a.s. for every* $r < 1/2$.

(v) $\mathbb{P} \circ \{\sqrt{u_n}\,(\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}))\}^{-1} \xrightarrow{\text{w}} \mathcal{N}_{0,\, s^2}$.

The following corollary is a direct consequence of Theorem 2.2.2. It is devoted to the strong rates and asymptotic normality of the estimator ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$.

Corollary 2.2.3 *Suppose that the assumptions in 2.2.1 are fulfilled for some $\lambda > 2$. Then parts (iv) and (v) of Theorem 2.2.2 show that the convergences in (2.3) and (2.4) hold true.*

The following result provides the analogue of Theorem 2.2.2 for the empirical plug-in estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n = \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ for the individual premium $\mathcal{R}_n = \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$. Assertions (iii)–(iv) in Theorem 2.2.4 describe the asymptotic behavior of the estimator $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$.

Theorem 2.2.4 (Empirical plug-in estimator) *Suppose that Assumption 2.2.1 holds with $\lambda > 2$, and assume that $\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$. Then the following assertions hold:*

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) - \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) = o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}) = (\widehat{m}_{u_n} - m) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(iii) $n^r(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})) \longrightarrow 0$ *$\mathbb{P}$-a.s. for every* $r < 1/2$.

(iv) $\mathbb{P} \circ \{\sqrt{u_n}\,(\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*n}))\}^{-1} \xrightarrow{\text{w}} \mathcal{N}_{0,\, s^2}$.

The below corollary is a direct consequence of Theorem 2.2.4. It is devoted to the strong rates and asymptotic normality of the empirical plug-in estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$.

Corollary 2.2.5 *Suppose that Assumption 2.2.1 is fulfilled for some $\lambda > 2$, and assume that $\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$. Then parts (iii) and (iv) of Theorem 2.2.4 show that the convergences in (2.12) and (2.13) hold true.*

Before we present the proofs of the upper theorems, we first take our time to discuss some useful aspects related to these results. As a direct consequence of Theorems 2.2.2 and 2.2.4 we obtain the following representations for the estimated individual premiums:

$$\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\, n\widehat{s}^2_{u_n}}) \;=\; \widehat{m}_{u_n} + \frac{1}{\sqrt{n}}\,\widehat{s}_{u_n}\mathcal{R}_\rho(\mathcal{N}_{0,1}), \tag{2.19}$$

$$\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) \;=\; \widehat{m}_{u_n} + \frac{1}{\sqrt{n}}\,\widehat{s}_{u_n}\mathcal{R}_\rho(\mathcal{N}_{0,1}) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{2.20}$$

Equation (2.19) is a simple consequence of part (c) of Assumption 2.2.1, and (2.20) follows from (2.19) and part (i) of Theorem 2.2.4. Furthermore it is important to note that the measurability assumption in Theorem 2.2.4 on $\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ is not very restrictive. This is easily seen if the role of the risk measure ρ is played by the Value at Risk of Example 1.2.3 for example, or refers to some distortion risk measure in the sense of Section 1.3. The following remark, which we will also prove, will now guarantee measurability of the estimated premium based on the convolution of the empirical measure for a wider class of risk functionals.

Remark 2.2.6 *Let $\mathcal{X} = L^p$ for some $p \in [1,\infty)$. Then for every law-invariant coherent risk measure $\rho : L^p \to \mathbb{R}$ the estimator $\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})$ is $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$.*

Proof Let $\rho : L^p \to \mathbb{R}$ be a law-invariant coherent risk measure. First, Theorem 2.8 in [37] ensures that the corresponding risk functional $\mathcal{R}_\rho : \mathcal{M}(L^p) \to \mathbb{R}$ is continuous for the p-weak topology $\mathcal{O}_{p\text{-w}}$. The latter is defined to be the coarsest topology on $\mathcal{M}(L^p)$ w.r.t. which each of the maps $\mu \mapsto \int f\,d\mu$, $f \in C^p_{\mathrm{b}}$, is continuous, where C^p_{b} is the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which there exists a constant $C > 0$ such that $|f(x)| \le C(1+|x|^p)$ for all $x \in \mathbb{R}$. According to Corollary A.45 in [27] the topological space $(\mathcal{M}(L^p), \mathcal{O}_{p\text{-w}})$ is Polish. Second, the topology $\mathcal{O}_{p\text{-w}}$ is generated by the L^p-Wasserstein metric d_{Wass_p} which is defined by

$$d_{\mathrm{Wass}_p}(\mu,\nu) \;:=\; \Big(\int_0^1 |F_\mu^{-1}(x) - F_\nu^{-1}(x)|^p\,dx\Big)^{1/p},$$

for every $\mu,\nu \in \mathcal{M}(L^p)$. Here we used the notation $F^{-1}(x) := \inf\{y \in \mathbb{R} :\; F(y) \ge x\}$. The mapping $\mathcal{M}(L^p) \to \mathcal{M}(L^p)$, $\mu \mapsto \mu^{*n}$, is $(d_{\mathrm{Wass}_p}, d_{\mathrm{Wass}_p})$-continuous; see Lemma 8.6 in [13]. Third, the mapping $\omega \mapsto \widehat{\mu}_{u_n}(\omega,\cdot)$ is $(\mathcal{F},\sigma(\mathcal{O}_{p\text{-w}}))$ -measurable. Indeed, it is easily seen that the Borel σ-algebra $\sigma(\mathcal{O}_{p\text{-w}})$ on $\mathcal{M}(L^p)$ is generated by the maps $\mu \mapsto \int f d\mu$, $f \in C^p_{\mathrm{b}}$. So, for $(\mathcal{F},\sigma(\mathcal{O}_{p\text{-w}}))$-measurability of the mapping $\Omega \to \mathcal{M}(L^p)$, $\omega \mapsto \widehat{\mu}_{u_n}(\omega,\cdot)$, it suffices to show

$$\Big(\int f(x)\,\widehat{\mu}_{u_n}(\,\cdot\,,dx)\Big)^{-1}(A) \in \mathcal{F} \quad \text{for all } A \in \mathcal{B}(\mathbb{R}) \text{ and } f \in C^p_{\mathrm{b}}. \tag{2.21}$$

Since $\widehat{\mu}_{u_n}(\omega,\cdot)$ is a probability kernel from $(\Omega,\mathcal{F})$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, the mapping

$$\omega \mapsto \int f(x)\,\widehat{\mu}_{u_n}(\omega,dx)$$

is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable for every $f \in C_{\mathrm{b}}^{p}$; see e.g. Lemma 1.41 in [35]. This gives (2.21). Altogether, we have shown that the mapping $\omega \mapsto \mathcal{R}_\rho(\widehat{\mu}_{u_n}^{*n}(\omega, \cdot))$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable.

□

The proofs of Theorems 2.2.2 and 2.2.4 avail the following nonuniform Berry–Esséen inequality (2.22). The inequality provides an upper bound for the distance of the distribution of a suitably centered random sum to the standard normal distribution w.r.t. nonuniform Kolmogorov distance d_{ϕ_λ}.

Theorem 2.2.7 *Let (X_i) be a sequence of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{V}\mathrm{ar}[X_1] > 0$ and $\mathbb{E}[|X_1|^\lambda] < \infty$ for some $\lambda > 2$. For every $n \in \mathbb{N}$, let*

$$Z_n := \frac{\sum_{i=1}^{n}(X_i - \mathbb{E}[X_1])}{\sqrt{n\mathbb{V}\mathrm{ar}[X_1]}}\,.$$

Then there exists a universal constant $C_\lambda \in (0,\infty)$ such that

$$d_{\phi_\lambda}(\mathbb{P}_{Z_n}, \mathcal{N}_{0,1}) \;\le\; C_\lambda\, f(\mathbb{P}_{X_1})\, n^{-\gamma} \qquad \textit{for all } n \in \mathbb{N} \tag{2.22}$$

with $\gamma := \min\{1, \lambda - 2\}/2$, where

$$f(\mathbb{P}_{X_1}) \;:=\; \frac{\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]}{\mathbb{V}\mathrm{ar}[X_1]^{\lambda/2}}. \tag{2.23}$$

By "universal constant" we mean that the constant is independent of $\mathbb{P}_{X_1}$. Inequality (2.22) has been proven by Nagaev [47] and Bikelis [14] for $\lambda = 3$ and $\lambda \in (2,3]$, respectively. Meanwhile there exist several estimates for the constant C_λ for $\lambda \in (2,3]$; see [48] and references cited therein. For $\lambda > 3$ the inequality is a direct consequence of Theorem 13 of Chapter V in [50].

In Appendix C we will present a slightly different version of the nonuniform Berry–Esséen inequality, which we will also prove. The proof is based on the approach by [46]. However, as the proof presented in [46] did not make it clear how the constants in the upper inequality had to be chosen, or how these constants depended on the distribution of the underlying random variables, we take our time to carry out the proof in a more rigorous way.

Proof of Theorem 2.2.2:

(i): By part (c) of Assumption 2.2.1 and the representation (2.6) (and its analogue in the case of known parameters), we have

$$\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}, n\widehat{s}_{u_n}^2}) - \mathcal{R}_\rho(\mathcal{N}_{nm, ns^2}) \;=\; \sqrt{n}(\widehat{s}_{u_n} - s)\mathcal{R}_\rho(\mathcal{N}_{0,1}) + n(\widehat{m}_{u_n} - m). \tag{2.24}$$

Since the empirical standard deviation $\widehat{s}_{u_n}$ converges $\mathbb{P}$-a.s. to the true standard deviation s, the claim of part (i) follows through dividing Equation (2.24) by n.

(ii): Let S_n be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ^{*n}, set

$$Z_n := (S_n - nm)/(\sqrt{n}s),$$

and write $\mathfrak{m}_n$ for the law of Z_n. Note that

$$\mathrm{law}\{\sqrt{n}sZ_n + nm\} = \mu^{*n}.$$

Write N_n for any random variable distributed according to the normal distribution $\mathcal{N}_{nm,ns^2}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and note that $Z := (N_n - nm)/(\sqrt{n}s)$ is $\mathcal{N}_{0,1}$-distributed. Due to part (c) of Assumption 2.2.1, we obtain

$$\begin{aligned} \mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2}) - \mathcal{R}_\rho(\mu^{*n}) &= \rho(\sqrt{n}sZ + nm) - \rho(\sqrt{n}sZ_n + nm) \\ &= \sqrt{n}s(\rho(Z) - \rho(Z_n)) \\ &= \sqrt{n}s(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\mathfrak{m}_n)). \end{aligned} \tag{2.25}$$

The nonuniform Berry–Esséen inequality of Theorem 2.2.7 shows that there exists a constant $K_\lambda \in (0, \infty)$ such that $d_{\phi_\lambda}(\mathcal{N}_{0,1}, \mathfrak{m}_n) \leq K_\lambda n^{-\gamma}$ for all $n \in \mathbb{N}$. Along with (2.25) and the $(d_{\phi_\lambda}, |\cdot|)$-continuity of $\mathcal{R}_\rho$ at $\mathcal{N}_{0,1}$ part (d) of Assumption 2.2.1, this ensures that we have

$$n^{-1}|\mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2}) - \mathcal{R}_\rho(\mu^{*n})| = n^{-1/2}s|\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\mathfrak{m}_n)| = o(n^{-1/2})$$

for all $n \in \mathbb{N}$. This completes the proof of part (ii).

(iii): The assertion follows from (i)–(ii).

(iv): By the Marcinkiewicz–Zygmund strong law of large numbers, we have that $n^r(\widehat{m}_{u_n} - m)$ converges $\mathbb{P}$-a.s. to zero for every $r < 1/2$. So the assertion follows from part (iii).

(v): The classical Central Limit Theorem says that the law of $(u_n)^{1/2}(\widehat{m}_{u_n} - m)$ converges weakly to $\mathcal{N}_{0,\,s^2}$. So the assertion follows from Slutsky's lemma and part (iii). □

Proof of Theorem 2.2.4:

(i): Analogously to (2.25), we obtain

$$\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}(\omega),\, n\widehat{s}^2_{u_n}(\omega)}) - \mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}(\omega;\cdot)) \;=\; \sqrt{n}\widehat{s}_{u_n}(\omega)(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\widehat{\mathfrak{m}}_n(\omega;\cdot)) \tag{2.26}$$

for all $\omega \in \Omega$, where $\widehat{\mathfrak{m}}_n(\omega;\cdot)$ denotes the law of the random variable

$$\widehat{Z}^\omega_n(\cdot) := \frac{\widehat{S}^\omega_n(\cdot) - n\widehat{m}_{u_n}(\omega)}{\sqrt{n}\widehat{s}_{u_n}(\omega)}$$

for any random variable $\widehat{S}^\omega_n(\cdot)$ with distribution $\widehat{\mu}^{*n}_{u_n}(\omega;\cdot)$ and defined on some probability space $(\Omega^\omega, \mathcal{F}^\omega, \mathbb{P}^\omega)$. For (2.26) notice that $\widehat{\mu}_{u_n}(\omega;\cdot)$ has mean $\widehat{m}_{u_n}(\omega)$ and standard deviation $\widehat{s}_{u_n}(\omega)$ for every fixed ω.

By the nonuniform Berry–Esséen inequality of Theorem 2.2.7, we have

$$d_{\phi_\lambda}(\mathcal{N}_{0,1},\widehat{\mathfrak{m}}_n(\omega;\cdot)) \;\leq\; C_\lambda\,\frac{\int|x-\int y\,\widehat{\mu}_{u_n}(\omega;dy)|^\lambda\,\widehat{\mu}_{u_n}(\omega;dx)}{\{\int(x-\int y\,\widehat{\mu}_{u_n}(\omega;dy))^2\,\widehat{\mu}_{u_n}(\omega;dx)\}^{\lambda/2}}\,n^{-\gamma} \tag{2.27}$$

for all $n\in\mathbb{N}$, where $C_\lambda\in(0,\infty)$ is a universal constant depending only on λ and being independent of n and ω. As a consequence of part (a) of Assumption 2.2.2 we have that $\int|x|^\lambda\,\widehat{\mu}_{u_n}(\omega;dx)=\frac{1}{u_n}\sum_{i=1}^{u_n}|Y_i|^\lambda$ converges to $\mathbb{E}[|Y_1|^\lambda]$ for $\mathbb{P}$-a.e. ω. That is, the numerator of

$$\frac{\int|x-\int y\,\widehat{\mu}_{u_n}(\omega;dy)|^\lambda\,\widehat{\mu}_{u_n}(\omega;dx)}{\{\int(x-\int y\,\widehat{\mu}_{u_n}(\omega;dy))^2\,\widehat{\mu}_{u_n}(\omega;dx)\}^{\lambda/2}} \tag{2.28}$$

is bounded above by an expression that converges to $2^\lambda\mathbb{E}[|Y_1|^\lambda]$ for $\mathbb{P}$-a.e. ω. The denominator is nothing but $\widehat{s}_{u_n}(\omega)^\lambda$ and thus converges to s^λ for $\mathbb{P}$-a.e. ω. That is, the expression in (2.28) converges to a positive constant for $\mathbb{P}$-a.e. ω. Together with (2.26), part (d) of Assumption 2.2.1, (2.27), and the $\mathbb{P}$-a.s. convergence of $\widehat{s}_{u_n}$ to s, this implies

$$n^{-1}(\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}(\omega),\,n\widehat{s}^2_{u_n}(\omega)})-\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}(\omega;\cdot))\;=\;o(n^{-1/2}) \tag{2.29}$$

for $\mathbb{P}$-a.e. ω. This completes the proof of part (i).

(ii): The assertion follows from (i)–(ii) of Theorem 2.2.2 and part (i) of Theorem 2.2.4.

(iii)-(iv): The assertions can be proven in the same way as the assertions (iv)–(v) of Theorem 2.2.2; just replace part (iii) of Theorem 2.2.2 by part (ii) of Theorem 2.2.4. □

The following remark, which we will also prove, will show that we can obtain stronger rates of convergence as the ones in part (ii) of Theorem 2.2.2 and part (i) of Theorem 2.2.4 if the underlying risk measure is not only continuous at $\mathcal{N}_{0,1}$, but β-Hölder continuous for some $\beta\in(0,\infty)$.

Remark 2.2.8 *Note that we can achieve stronger results as the ones in part (ii) of Theorem 2.2.2 and part (i) of Theorem 2.2.4, if we replace part (d) of Assumption 2.2.1 by the following slightly stronger assumption:*

(d') For each sequence $(\mathfrak{m}_n)\subset\mathcal{M}_1^\lambda$ with $d_{\phi_\lambda}(\mathfrak{m}_n,\mathcal{N}_{0,1})\to 0$, there exist constants $C,\beta>0$ such that

$$|\mathcal{R}_\rho(\mathfrak{m}_n)-\mathcal{R}_\rho(\mathcal{N}_{0,1})|\leq C d_{\phi_\lambda}(\mathfrak{m}_n,\mathcal{N}_{0,1})^\beta$$

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

Let $\gamma:=\min\{\lambda-2,1\}/2$. Then the following assertions hold true:

(i) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2})-\frac{1}{n}\mathcal{R}_\rho(\mu^{*n})=\mathcal{O}(n^{-1/2-\beta\gamma})$

(ii) $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},\,n\widehat{s}^2_{u_n}})-\frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n})=(\widehat{m}_{u_n}-m)+\mathcal{O}_{\mathbb{P}\text{-a.s.}}(n^{-1/2-\beta\gamma})$.

Here $\mathcal{O}_{\mathbb{P}\text{-a.s.}}(n^{-1/2-\beta\gamma})$ refers to any sequence of random variables (η_n) on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for which the sequence $(n^{1/2+\beta\gamma}\eta_n)$ is bounded $\mathbb{P}$-a.s.

Proof of Remark 2.2.8 We will only show the first part. Part (ii) can be proven analogously.

(i) Following the same line of reasoning as in the proof of Theorem 2.2.2, we observe that (2.25) along with part (d') of the upper assumption ensures, that we can find some constants $K, \beta \in (0, , \infty)$, such that

$$n^{-1}|\mathcal{R}_\rho(\mathcal{N}_{nm,\,ns^2}) - \mathcal{R}_\rho(\mu^{*n})| \leq n^{-1/2} K \, d_{\phi_\lambda}(\mathcal{N}_{0,1}, \mathfrak{m}_n))^\beta \leq C K_\lambda \, n^{-1/2-\beta\gamma}$$

for all $n \in \mathbb{N}$. Here $K_\lambda \in (0, \infty)$ is the constant in the Berry-Esséen inequality of Theorem 2.2.7. This leads to the assertion. □

2.3 Numerical simulations

In this section we present some numerical examples to illustrate the results of Section 2.2. Our results show that both the estimated normal approximation and the empirical plug-in estimator lead to reasonable estimators for the premium of an individual risk within a homogeneous insurance collective. Our results also show that these two estimators are asymptotically equivalent. Nevertheless for small to moderate collective sizes n the goodness of the estimators can vary from case to case. For example, in the case where ρ is the Value at Risk at level α the results of Theorem 2.2.2 show that for both estimators the estimation error converges almost surely to zero at rate (nearly) 1/2 when $\mathbb{E}[|Y_1|^\lambda] < \infty$ for some $\lambda > 2$ (where Y_1 refers to any μ-distributed random variable). On the other hand, the latter condition does not exclude that $\mathbb{E}[|Y_1|^{2+\varepsilon}] = \infty$ for some small $\varepsilon > 0$. In this case the total claim distribution can be essentially skewed to the right when the number of individual risks n is small to moderate; cf. Figure 2.1. So one would expect that especially for heavy-tailed μ and small to moderate n, the estimators perform only moderately well. One would also expect that for heavy-tailed μ (and even for medium-tailed μ) and small to moderate n the empirical plug-in estimator should outperform the estimated normal approximation. Our goal in this section is to provide empirical evidence for our conjectures.

To this end let us consider a sequence (Y_i) of i.i.d. nonnegative random variables on a common probability space with distribution

$$\mu = (1-p)\,\delta_0 + p\,\mathrm{P}_{a,b}$$

for some $p \in (0, 1)$, where $\mathrm{P}_{a,b}$ is the Pareto distribution with parameters $a > 2$ and $b > 0$. The Pareto distribution $\mathrm{P}_{a,b}$ is determined by the Lebesgue density

$$f_{a,b}(x) := ab^{-1}(b^{-1}x + 1)^{-(a+1)} \, \mathbb{1}_{(0,\infty)}(x),$$

and the assumption $a > 2$ ensures that $\mathbb{E}[|Y_1|^\lambda] < \infty$ for all $\lambda \in (2, a)$. We regard $Y_1, \ldots, Y_n$ as a homogeneous insurance collective of size n, the number p as the probability for the event of a strictly positive individual claim amount, and $\mathrm{P}_{a,b}$ as the individual claim distribution conditioned on this event. Note that in our example the mean m and the variance s^2 of μ are given by

$$m = \frac{p\,b}{a-1} \quad \text{and} \quad s^2 = \frac{2b^2 p}{(a-1)(a-2)} - \frac{b^2 p^2}{(a-1)^2}\,. \tag{2.30}$$

In the first part of this section, we estimate the total claim distribution μ^{*n}, i.e. the distribution of $\sum_{i=1}^n Y_i$, by means of the empirical distribution based on a Monte-Carlo simulation. The plots in Figure 2.1 were derived from a simulation with 100,000 Monte-Carlo paths. We set $p = 0.1$ and chose the parameters a and b in such a way that the expected value of a single claim was normalized to 1. Each line shows the same set of parameters and each column shows the same collective size, starting with $n = 100$ on the left, $n = 150$ in the middle and $n = 200$ on the right. The first line shows the results for $a = 2.1$ and $b = 11$, the second line shows $a = 3$ and $b = 20$, the third line shows $a = 6$ and $b = 50$ and the fourth line shows $a = 10$ and $b = 90$. In each plot the continuous line represents the estimator for μ^{*n} and the dashed line the probability density of the normal distribution $\mathcal{N}_{nm,\,ns^2}$ with m and s^2 determined through (2.30). We emphasize that μ^{*n} has in fact point mass in zero. But the point mass is equal to $(1-p)^n$ and therefore extremely small. This is why the point mass of the empirical estimator is not visible in the plots.

One can see that the empirical total claim distributions in the first line of Figure 2.1 are strongly skewed to the right even for larger collective sizes. The density of the normal distribution is very flat and has much mass on the negative semiaxis. The reason for this shape is the high variance s^2, which increases rapidly as a gets closer to 2. In the case of $a = 2.1$ and $b = 11$ this rate is close to zero, saying that large collective sizes are needed to provide a suitable estimator.

In the second line of Figure 2.1 for $a = 3$ and $b = 20$ the empirical total claim distributions are still strongly skewed to the right. One can see that the normal approximation still does not resemble the empirical distribution. The deviation decreases visibly with increasing collective size due to the higher rate of convergence in the Berry–Esséen theorem. Compared to the first line with $a = 2.1$ and $b = 11$ the quality of the normal approximation was increased in the second line with $a = 3$ and $b = 20$, which can be explained by the increasing rate of convergence in the Berry–Esséen theorem. For $\lambda \in (2, 3]$ the convergence rate to the normal distribution is strictly increasing in λ. For $\lambda > 3$ the convergence rate can not be improved any more.

In the third and fourth line of Figure 2.1 for $a = 6$ and $b = 50$ and $a = 10$ and $b = 90$ the normal approximation provides a good approximation even for small collective sizes. The empirical total claim distributions are in both cases almost symmetric and the approximation leads to a good fit of both curves. The third moment of X_1 exists in both cases and due to the Berry–Esséen theorem the deviation of μ^{*n} from the normal distribution converges to

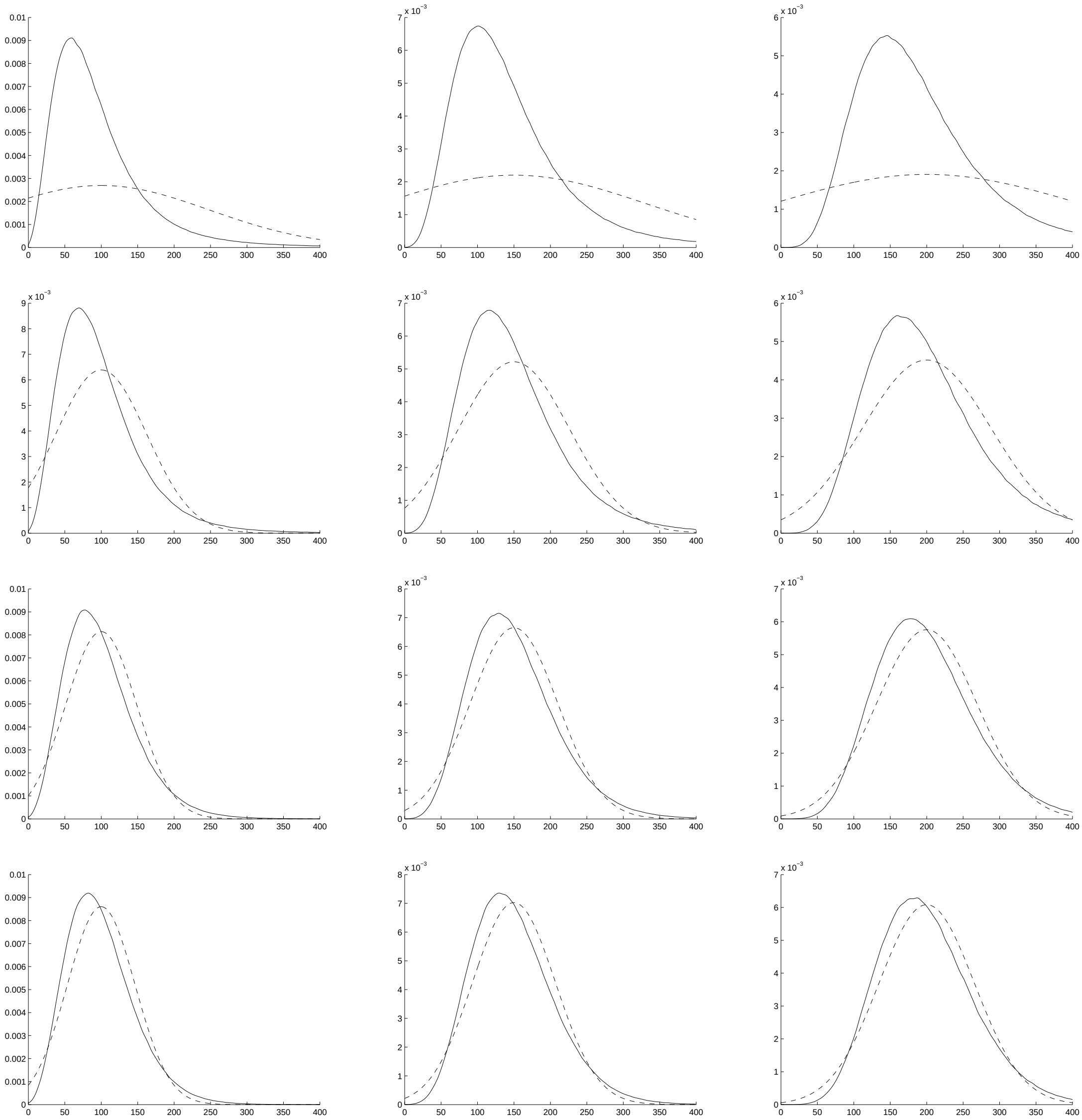

Figure 2.1: The continuous line shows the n-fold convolution μ^{*n} of $\mu = (1-p)\delta_0 + p\mathrm{P}_{a,b}$ for $p = 0.1$ and the Pareto distribution $\mathrm{P}_{a,b}$ with parameter $a = 2.1$ in the first line, $a = 3$ in the second line, $a = 6$ in the third line and $a = 10$ in the fourth line and collective sizes $n = 100$ in the first column, $n = 150$ in the second column and $n = 200$ in the third column. The dashed line shows the density of the respective normal distribution in each case.

zero with rate 1/2. We can see that there is no remarkable improve in the convergence rate once the existence of the third moment is guaranteed.

In the second part of this section we compare the estimated normal approximation with the empirical plug-in estimator where the role of the risk measure ρ is played by the Value at risk at level $\alpha = 0.99$. To save computing time we discretized the Pareto distribution $\mathrm{P}_{a,b}$ on the equidistant grid $10\mathbb{N}_0 = \{0, 10, 20, \ldots\}$. The plots in Figure 2.2 were derived by a Monte-Carlo method using 100 Monte-Carlo paths in each simulation. Once again we chose $p = 0.1$. In order to compare the estimators we first calculated the exact Value at Risks at level 0.99 of μ^{*n} (in fact we estimated it by means of a Monte-Carlo simulation based on 100.000 runs) in dependence on the collective size n. In each plot in Figure 2.2 the dotdashed line represents the relative Value at Risk $\mathcal{R}_\rho(\mu^{*n})/n$, which we take as a reference to illustrate the biases of the estimators. The dashed line shows the estimated normal approximation $\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_n,\, n\widehat{s}_n^2})/n$ for the Value at Risk relative to n. The continuous line shows the empirical plug-in estimator $\mathcal{R}_\rho(\widehat{\mu}_n^{*n})/n$ for the Value at Risk relative to n.

The first line shows the relative Value at Risks for the parameters $a = 2.1$ and $b = 11$ on the left and $a = 3$ and $b = 20$ on the right hand side. In the second line we have $a = 6$ and $b = 50$ on the left and $a = 10$ and $b = 90$ on the right hand side. Once again the parameters were chosen such that the expected value of a single claim was normalized to 1.

For $a = 2.1$ we can see that both estimators show a large negative bias. The slow convergence in the Berry–Esséen theorem transfers directly to the convergence of the relative Value at risk of the distributions (recall that the Value at Risk fulfills condition (d) of Assumption 2.2.1 for $\beta = 1$). Due to this slow convergence the collective size has to be chosen very large to provide a good estimation. What strikes the most is the large bias of the relative empirical plug-in estimator $\mathcal{R}_\rho(\mu^{*n})/n$. The heaviness of the tails causes the empirical distribution $\widehat{\mu}_n$ to converge very slowly to μ^{*n}. We can see that in the case $a = 3$ the bias of both estimators decreases visibly. However in both cases the empirical plug-in estimator yields a better estimation.

The plots for $a = 6$ and $a = 10$ resemble each other very much. In both cases the existence of the third moment of X_1 is guaranteed, yielding the same rate of convergence in the Berry–Esséen theorem. We can see that for small n, e.g. $n \leq 40$, both estimators show a large bias. However for $n \leq 100$ the empirical plug-in estimator provides a better estimation. For $n \geq 100$ the estimated normal approximation could be preferred over the empirical plug-in estimator, because the biases of both estimators are more or less the same and the estimated normal approximation consumes less computing time.

As a conclusion one can say that the estimated normal approximation is not suitable for heavy-tailed (to medium-tailed) distributions whenever small collective sizes are at hand. In this case it is sensible to apply the empirical plug-in estimator, which consumes more computing time compared to the estimated normal approximation. However, both estimators are subject to a negative bias w.r.t. the true individual premium. In the next chapter we

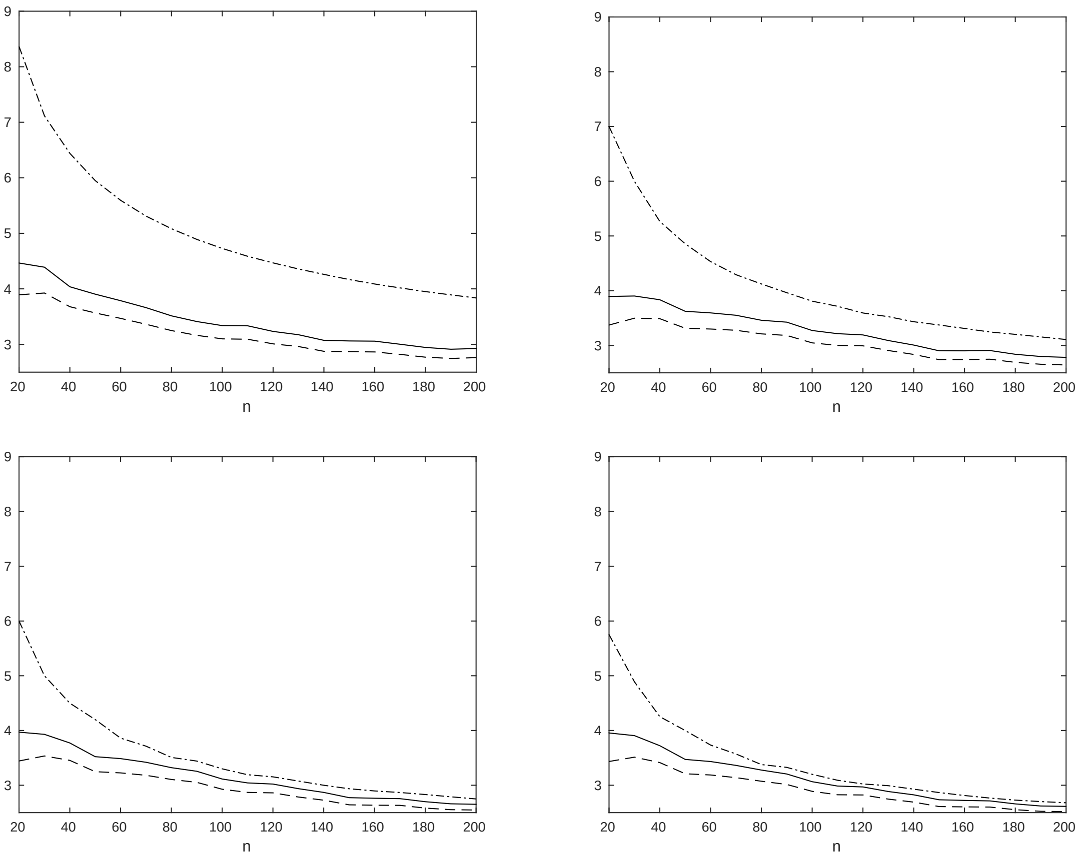

Figure 2.2: $\mathcal{R}_\rho(\mu^{*n})/n$ (dotdashed line) as well as the average of 100 Monte-Carlo paths of respectively $\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_n, n\widehat{s}_n^2})/n$ (dashed line) and $\mathcal{R}_\rho(\widehat{\mu}_n^{*n})/n$ (continuous line) for $\rho = \mathrm{V@R}_{0.99}$ in dependence on the collective size n, showing $a = 2.1$ on the left hand side and $a = 3$ on the right hand side of the first line and $a = 6$ on the left hand side and $a = 10$ on the right hand side of the second line.

will therefore develop a theory with the scope to alleviate the bias in our estimations.

2.4 Bootstrapping the individual premium in the individual model

As we have seen in Section 2.3, a nonparametric estimation of the individual premium is subject to a negative bias w.r.t. the true individual premium. Especially the cases with heavy-tailed single claim distributions led to rather large biases in our estimations. In order to hopefully improve the numerical results of Section 2.3, Section 2.4.1 introduces the so-called bootstrap-based bias correction. Roughly speaking, the idea behind this procedure is to “estimate“ the bias in an estimation by means of a suitable resampling of the original observations, and subtract the estimated bias from the original estimator.

More explicitly, in the former sections we considered the nonparametric estimators

$$ {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n := \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n},n\widehat{s}^2_{u_n}}) \qquad \text{and} \qquad {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}^{*n}_{u_n}) \tag{2.31} $$

for the individual premium $\mathcal{R}_n$ based on observed historical single claims $Y_1, \ldots, Y_{u_n}$, where the Y_i are assumed to be i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution μ. Then, assuming again that

$$ \lim_{n\to\infty} u_n/n = c \quad \text{for some constant } c \in (0, \infty) \tag{2.32} $$

and some additional mild assumptions on μ and the risk measure ρ, the results of Section 2.2 have shown that the estimators in (2.31) are strongly consistent in the sense that the deviation of the estimator from the true value converges to zero $\mathbb{P}$-almost surely, that is

$$ {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n \longrightarrow 0 \quad \mathbb{P}\text{-a.s.} \qquad \text{and} \qquad {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n \longrightarrow 0 \quad \mathbb{P}\text{-a.s.} \tag{2.33} $$

Furthermore we were able to proof asymptotic normality in the sense that

$$ \mathbb{P}\circ\{\sqrt{u_n}({}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n)\}^{-1} \overset{\mathsf{w}}{\longrightarrow} \mathcal{N}_{0,s^2} \qquad \text{and} \qquad \mathbb{P}\circ\{\sqrt{u_n}({}^{\mathsf{CE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n)\}^{-1} \overset{\mathsf{w}}{\longrightarrow} \mathcal{N}_{0,s^2}. \tag{2.34} $$

The latter means that for large n the deviation of the estimator from the true value is distributed according to the normal distribution with mean 0 and variance $u_n s^2$. Condition (2.32) is again motivated by the fact that the premium is typically estimated on the basis of the historical claims of the same collective from the last year or from the last few years. We stress the fact again, that this condition is somehow nonstandard, because in the literature on asymptotic statistical inference for convolutions it is usually assumed that the number of summands n is fixed and the number of observations u tends to infinity; see, for instance, [52].

On the other hand, the results of the simulation studies in Section 2.3 have shown that the estimators in (2.31) are subject to a negative bias for finite sample size n. In particular when the conditional single claim distribution $\mu_{>0}[\,\cdot\,] := \mu[\,\cdot\, \cap (0,\infty)]/\mu[(0,\infty)]$ is "heavy-tailed" the bias can be considerable. For a more detailed discussion and some further background in the field of bias correction, see for instance [23].

Throughout this section we address the question whether the biases of the estimators ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ for the individual premium $\mathcal{R}_n$ can be reduced by means of the bootstrap technique to be explained in Section 2.4.1. For the estimator $\mathcal{R}_\rho(\widehat{\mu}_n)$ of $\mathcal{R}_\rho(\mu)$ analogous investigations have been done by Kim and Hardy [33] for the Value at Risk and the Average Value at Risk, and by Kim [34] for more general distortion risk measures. Ahn and Shyamalkumar [2] provided some asymptotic analysis for the Average Value at Risk in this context. Part (iii) of Remark 2.2.6 below indicates that the bootstrap approach for reducing the bias is *not* expedient for the estimator ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$. On the other hand, the bootstrap approach can be (slightly) useful for ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$. In our numerical examples for ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ with ρ the Value at Risk and the Average Value at Risk of Example 1.2.3 and 1.2.4, respectively, we obtain results that are qualitatively comparable to the numerical results of [33, 34]. Whereas for the Value at Risk an application of the bootstrap-based method of Section 2.3 seems not useful, for the Average Value at Risk we can observe that on average a small to moderate reduction of the bias goes along with a small increase of the variance and thus, of the mean squared error.

In the framework of [2, 33, 34] the plug-in estimator $\mathcal{R}_\rho(\widehat{\mu}_n)$ for a distortion risk measure $\mathcal{R}_\rho(\mu)$ is an L-statistic, and thus bootstrap consistency is known from the literature. For the Average Value at Risk functional, see also Corollary 4.2 in [12]. Moreover, for L-statistics even the exact bootstrap mean can be calculated explicitly, see for instance [31]. In our setting, where the individual premium $\mathcal{R}_n = \mathcal{R}_\rho(\mu^{*n})/n$ is estimated by ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$, bootstrap results seem not to exist so far.

For this reason the results of Theorem 2.4.3 will yield bootstrap consistency for the bootstrap estimators to give a mathematical justification for the use of the bootstrap-based method of Section 2.4.1. Theorem 2.4.3 will show almost sure bootstrap consistency for the nonparametric estimators for the individual premium and thus provides the theoretical justification for the use of the bias correction. Although the method of Section 2.4.1 seems not to be appropriate for the estimator ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ (see part (iii) of Remark 2.2.6 below), in Theorem 2.4.3 we also establish bootstrap consistency for this estimator. In Subsection 2.5 we will present the results of some numerical simulation studies.

To this end, we will demonstrate a way to estimate the bias in our former estimation by means of the bootstrap, from which we will derive the bias correction. The bias correction is the central tool which will be used to construct estimators based on the original data, used in the former estimations, but with a smaller bias w.r.t. the true value. The bootstrap versions of our estimators in (2.31) will then be given by the bias-corrected original estimators.

2.4.1 Bootstrap-based bias correction

As we have already mentioned, the estimators defined in (2.31) have a negative bias w.r.t. $\mathcal{R}_n$. As a countermeasure one can try to "estimate" the bias and subtract it from the original estimator. The "estimation" of the bias can sometimes be done by means of bootstrap methods. The idea of the bootstrap was introduced by Efron in 1979 in his seminal paper [22]. Since then many variants of the bootstrap have been discussed in the literature; for background and details one may refer to [16, 23, 40, 57] among others.

To explain the bootstrap-based method for correcting the bias more precisely, let $\widehat{\mathcal{R}}_n$ be an estimator for a real-valued characteristic $\mathcal{R}_n$, $n \in \mathbb{N}$, where $\mathcal{R}_n$ may or may not be defined by (2.1). In any case assume that $\widehat{\mathcal{R}}_n$ is given by a statistical functional evaluated at a (random) probability measure which is uniquely determined by observed data $Y_1, \ldots, Y_{u_n}$, where the latter are given by the first u_n terms of a sequence (Y_i) of i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For illustrations of such estimators see (2.31). Assume that $\widehat{\mathcal{R}}_n$ is biased, i.e. that

$$\mathsf{Bias}(\widehat{\mathcal{R}}_n) := \mathbb{E}[\widehat{\mathcal{R}}_n - \mathcal{R}_n] \tag{2.35}$$

differs from 0 for finite sample size n. Further assume that

$$\mathbb{P} \circ \{\sqrt{u_n}(\widehat{\mathcal{R}}_n - \mathcal{R}_n)\}^{-1} \xrightarrow{\mathsf{w}} \mathcal{N}_{0,s^2} \tag{2.36}$$

holds for some $s^2 \in (0, \infty)$. See (2.34) for an illustration of condition (2.36). Now let $(\Omega', \mathcal{F}', \mathbb{P}')$ be a second probability space and extend the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the product space

$$(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}}) := (\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mathbb{P} \otimes \mathbb{P}'),$$

and assume that the result ω' of $(\Omega', \mathcal{F}', \mathbb{P}')$ and the original sample $Y_1(\omega), \ldots, Y_{u_n}(\omega)$ specify a new (random) probability measure. The latter is plugged in the underlying statistical functional to obtain a "bootstrap version" of $\widehat{\mathcal{R}}_n$, denoted by $\widehat{\mathcal{R}}_n^{\mathsf{B}}$. Note that $\widehat{\mathcal{R}}_n^{\mathsf{B}}$ depends on ω and ω', that is, it is defined on the probability space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$. Also note that, up to some measurability issues, the mapping $\omega' \mapsto \widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \omega')$ can be seen as a random variable on $(\Omega', \mathcal{F}', \mathbb{P}')$ for any fixed ω. For illustrations of $\widehat{\mathcal{R}}_n^{\mathsf{B}}$ see (2.40) and (2.42) below. In fact $\widehat{\mathcal{R}}_n^{\mathsf{B}}$ should be called (almost sure) bootstrap version of $\widehat{\mathcal{R}}_n$ only if

$$\mathbb{P}' \circ \{\sqrt{u_n}(\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \cdot\,) - \widehat{\mathcal{R}}_n(\omega))\}^{-1} \xrightarrow{\mathsf{w}} \mathcal{N}_{0,s^2} \qquad \mathbb{P}\text{-a.e. } \omega. \tag{2.37}$$

The left-hand side of (2.37) is often referred to as the conditional distribution of $\sqrt{u_n}(\widehat{\mathcal{R}}_n^{\mathsf{B}} - \widehat{\mathcal{R}}_n)$ given the observation $Y_1, \ldots, Y_{u_n}$. For a justification of this interpretation see, for instance, the discussion at the end of Section 2 in [12].

Whenever (2.36) and (2.37) can be shown, we have

$$\mathbb{P} \circ \{\widehat{\mathcal{R}}_n - \mathcal{R}_n\}^{-1} \approx \mathcal{N}_{0,s^2/u_n}$$

and

$$\mathbb{P}' \circ \{\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \cdot\,) - \widehat{\mathcal{R}}_n(\omega))\}^{-1} \approx \mathcal{N}_{0,s^2/u_n} \qquad \mathbb{P}\text{-a.e. } \omega$$

for "large" n. That is, informally,

$$\mathbb{P} \circ \{\widehat{\mathcal{R}}_n - \mathcal{R}_n\}^{-1} \approx \mathbb{P}' \circ \{\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \cdot\,) - \widehat{\mathcal{R}}_n(\omega))\}^{-1} \qquad \mathbb{P}\text{-a.e. } \omega \tag{2.38}$$

for "large" n. Sometimes it turns out that the two laws in (2.38) are not only "close" but even have a similar skewness so that the means of these two laws are close to each other. In this case the mean of the law on the right-hand side of (2.38) is a reasonable approximation of $\mathsf{Bias}(\widehat{\mathcal{R}}_n)$ defined in (2.35). Though the law on the right-hand side of (2.38) can be seldomly specified explicitly, it can be numerically approximated through

$$\frac{1}{L} \sum_{\ell=1}^{L} \delta_{\widehat{\mathcal{R}}_n^{\mathsf{B},\ell}(\omega,\,\cdot\,) - \widehat{\mathcal{R}}_n(\omega)} \qquad \text{with } L \gg n$$

for $\mathbb{P}$-a.e. ω due to the Glivenko–Cantelli theorem, where $\widehat{\mathcal{R}}_n^{\mathsf{B},1}(\omega,\,\cdot\,),\dots,\widehat{\mathcal{R}}_n^{\mathsf{B},L}(\omega,\,\cdot\,)$ are i.i.d. copies of $\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega,\,\cdot\,)$ for every fixed ω. In particular

$$\widehat{\mathsf{Bias}}_n^{\mathsf{B}} := \frac{1}{L}\sum_{\ell=1}^{L}(\widehat{\mathcal{R}}_n^{\mathsf{B},\ell}(\omega,\,\cdot\,) - \widehat{\mathcal{R}}_n(\omega))$$

is a reasonable approximation of $\mathsf{Bias}(\widehat{\mathcal{R}}_n)$ defined in (2.35) and thus

$$\widehat{\mathcal{R}}_n^{\mathsf{bc}} := \widehat{\mathcal{R}}_n - \widehat{\mathsf{Bias}}_n^{\mathsf{B}} \tag{2.39}$$

can provide an estimator for $\mathcal{R}_n$ with smaller bias than $\widehat{\mathcal{R}}_n$. At this point it is worth mentioning that $\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ often admits a larger mean squared error than the original estimator $\widehat{\mathcal{R}}_n$.

2.4.2 Bootstrap consistency for the nonparametric individual premium estimator

Analogously to Section 2.1 we write ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ for the estimator based on the normal approximation with estimated parameters and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ for the empirical plug-in estimator. Both estimators are used to estimate the individual premium $\mathcal{R}_n := \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$. Again, the estimators are based on a sequence (Y_i) of real-valued i.i.d. random variables on a common probability space $(\Omega,\mathcal{F},\mathbb{P})$ with distribution μ, and can be obtained by plugging the empirical probability measure $\widehat{\mu}_{u_n}$ of $Y_1,\dots,Y_{u_n}$ in the statistical functionals

$${}^{\mathsf{NA}}\mathcal{T}_n(\nu) := \frac{1}{n}\mathcal{R}_n(\mathcal{N}_{nm(\nu),ns^2(\nu)}) \qquad \text{and} \qquad {}^{\mathsf{CE}}\mathcal{T}_n(\nu) := \frac{1}{n}\mathcal{R}_n(\nu^{*n}) \tag{2.40}$$

respectively, i.e.

$${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n = {}^{\mathsf{NA}}\mathcal{T}_n(\widehat{\mu}_{u_n}) \qquad \text{and} \qquad {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n = {}^{\mathsf{CE}}\mathcal{T}_n(\widehat{\mu}_{u_n}).$$

Here $m(\nu)$ and $s^2(\nu)$ refer to the mean and the variance of a law ν respectively. We regard $\omega\in\Omega$ as a sample drawn from $\mathbb{P}$ and ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n(\omega)$ and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n(\omega)$ as statistics derived from ω, respectively from $(Y_1(\omega),\dots,Y_{u_n}(\omega))$. Let $(\Omega',\mathcal{F}',\mathbb{P}')$ be another probability space and set

$$(\overline{\Omega},\overline{\mathcal{F}},\overline{\mathbb{P}}) \;:=\; (\Omega\times\Omega',\mathcal{F}\otimes\mathcal{F}',\mathbb{P}\otimes\mathbb{P}').$$

The probability measure $\mathbb{P}'$ represents a random experiment which is run independently of the random mechanism $\mathbb{P}$. For every $n\in\mathbb{N}$ let

$$\widehat{\mu}_{u_n}^{\mathsf{B}}(\omega,\omega') := \frac{1}{u_n}\sum_{i=1}^{u_n} W_{u_n,i}(\omega')\delta_{Y_i(\omega)} \tag{2.41}$$

for some triangular array $(W_{u,i})$ of nonnegative real-valued random variables on the probability space $(\Omega',\mathcal{F}',\mathbb{P}')$. Note that the sequence (Y_i) and the triangular array $(W_{u,i})$ regarded as families of random variables on the product space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}}) = (\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mathbb{P} \otimes \mathbb{P}')$ are independent. Now we set

$$ {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n^{\mathsf{B}} := {}^{\mathsf{NA}}\mathcal{T}_n(\widehat{\mu}_n^{\mathsf{B}}) \qquad \text{and} \qquad {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{B}} := {}^{\mathsf{CE}}\mathcal{T}_n(\widehat{\mu}_n^{\mathsf{B}}) \tag{2.42} $$

with ${}^{\mathsf{NA}}\mathcal{T}_n$ and ${}^{\mathsf{CE}}\mathcal{T}_n$ as defined in (2.40). Theorem 2.2.2 ahead shows that under some mild assumptions these estimators can be seen as bootstrap versions of ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ respectively. In the theorem we will consider the so called *weighted exchangeable bootstrap* in the form of Assumption 2.4.1. Efron's bootstrap and the Bayesian bootstrap are special cases; see Example 2.4.2 ahead. For background on the weighted bootstrap see also [4, 18, 43, 45, 54, 56, 61] and references cited therein. Recall that exchangeable random variables are identically distributed.

Assumption 2.4.1 *The triangular array $(W_{u,i})$ of nonnegative random variables on the probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ satisfies the following conditions:*

A1. The random vector $(W_{u,1}, \dots, W_{u,u})$ is exchangeable for every $u \in \mathbb{N}$.

A2. $\sum_{i=1}^{u} W_{u,i} = u$ for every $u \in \mathbb{N}$.

A3. $W_{u,1} \in L^2(\Omega', \mathcal{F}', \mathbb{P}')$ for every $u \in \mathbb{N}$, and $\sup_{u \in \mathbb{N}} \mathbb{V}\mathrm{ar}'[W_{u,1}] < \infty$.

A4. $\frac{1}{\sqrt{u}} \max_{1 \le i \le u} |W_{u,i} - 1| \xrightarrow{\mathsf{p}} 0$ w.r.t. $\mathbb{P}'$.

A5. $\frac{1}{u} \sum_{i=1}^{u} (W_{u,i} - 1)^2 \xrightarrow{\mathsf{p}} 1$ w.r.t. $\mathbb{P}'$.

The following example will show that popular bootstrap schemes like Efron's bootstrap and the Bayesian bootstrap fulfill the conditions of Assumption 2.4.1.

Example 2.4.2 (i) *Efron's bootstrap* [22] is a special form of the weighted exchangeable bootstrap in the sense of Assumption 2.4.1. In this case the random vector $(W_{u,1}, \dots, W_{u,u})$ is multinomially distributed according to the parameters u and $p_1 = \cdots = p_u = \frac{1}{u}$ for every $u \in \mathbb{N}$. This choice of the weights $(W_{u,i})$ obviously fulfills conditions A1. through A3. of Assumption 2.4.1. Moreover it also satisfies conditions A4. and A5. This was already pointed out in Example 3.6.10 in [61], where one should note that by A2. we have $\overline{W}_u := \frac{1}{u} \sum_{i=1}^{u} W_{u,i} = 1$ and that, in view of Markov's inequality, the second condition in Display (3.6.8) of [61] implies condition A4.

(ii) Another version of the weighted exchangeable bootstrap in the sense of Assumption 2.4.1 can be obtained by choosing $W_{u,i} := Z_i / \overline{Z}_u$ for every $u \in \mathbb{N}$ and $i = 1, \dots, u$, where $\overline{Z}_u := \frac{1}{u} \sum_{j=1}^{u} Z_j$ and (Z_i) is any sequence of nonnegative i.i.d. random variables on $(\Omega', \mathcal{F}', \mathbb{P}')$ with $\mathbb{E}'[Z_1^2] < \infty$ and $\mathbb{E}'[Z_1] = \mathbb{V}\mathrm{ar}'[Z_1]^{1/2} > 0$. This choice of the weights $(W_{u,i})$ obviously fulfills conditions A1.–A2. of Assumption 2.4.1. It also satisfies conditions A4. and A5, as was already pointed out in Example 3.6.9 in [61] (again noting that $\overline{W}_u := \frac{1}{u} \sum_{i=1}^{u} W_{u,i} = 1$

and that the second condition in Display (3.6.8) of [61] implies condition A4). Moreover, we have

$$\begin{aligned}\mathbb{E}'[W_{u,1}^2] &= \mathbb{E}'[(Z_1/\overline{Z}_u)^2]\\ &= \int_0^\infty \mathbb{P}'[Z_1/\overline{Z}_u > t^{1/2}]\,dt\\ &\le \int_0^\infty \Big(\mathbb{P}'[Z_1 > t^{1/2}(1-\varepsilon)] + t^{-p/2}u^{p/2}\varrho(\varepsilon)^{u/2}\Big)\,dt\\ &= \frac{1}{(1-\varepsilon)^2}\,\mathbb{E}'[Z_1^2] + u^{p/2}\varrho(\varepsilon)^{u/2}\int_0^\infty t^{-p/2}\,dt\end{aligned}$$

for any $p > 0$, $\varepsilon \in (0,1)$, and some $\varrho(\varepsilon) \in (0,1)$, where the third step is ensured by Inequality (5.57) of [54] (assuming without loss of generality $\mathbb{E}'[Z_1] = 1$). Choosing $p > 2$ we can conclude that also conditions A3. of Assumption 2.4.1 is satisfied. In the special case where Z_1 is exponentially distributed to the parameter 1, the resulting scheme is the *Bayesian bootstrap* of Rubin [56]; see Example 3.1 of [54]. $\diamond$

The formulation of Theorem 2.4.3 will involve the weighted Kolmogorov distance w.r.t. the weight function ϕ_λ as introduced in (1.4). Let again $\mathcal{M}_1$ be the set of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and denote by F_μ the distribution function of $\mu \in \mathcal{M}_1$.

Theorem 2.4.3 *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, let $\mathcal{R}_\rho$ be the corresponding statistical functional as defined in (1.1), and assume that Assumptions 2.2.1 and 2.4.1 hold true. Then, if s^2 denotes the variance of μ, we have*

$$\mathbb{P}' \circ \{\sqrt{u_n}({}^{\mathsf{NA}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega,\,\cdot\,) - {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n(\omega))\}^{-1} \xrightarrow{\mathsf{w}} \mathcal{N}_{0,s^2} \qquad \mathbb{P}\text{-}a.e.\ \omega. \tag{2.43}$$

If the mapping $\omega' \mapsto {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega,\omega') - {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n(\omega)$ is $(\mathcal{F}', \mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$ and $\omega \in \Omega$, then we also have

$$\mathbb{P}' \circ \{\sqrt{u_n}\,({}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega,\,\cdot\,) - {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n(\omega))\}^{-1} \xrightarrow{\mathsf{w}} \mathcal{N}_{0,s^2} \qquad \mathbb{P}\text{-}a.e.\ \omega. \tag{2.44}$$

Note that it was shown in Section 2.2 that conditions (a)–(d) of Assumption 2.2.1 ensure (2.34). In this respect Theorem 2.4.3 complements these results. In fact, in Remark 2.2.8 condition (d) was replaced by a slightly stronger regularity condition (with the benefit of some additional results). However the proof there can be easily modified to obtain (2.34) under conditions (a)–(d) above. The assumptions of Theorem 2.4.3 will be discussed and illustrated in the following remarks and examples.

Remark 2.4.4 (i) The measurability assumption in Theorem 2.4.3 on ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega,\,\cdot\,) - {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n(\omega)$ is not very restrictive. For instance, when ρ is a distortion risk measure (see Section 1.3 for details), then we can show that ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega,\,\cdot\,) - {}^{\mathsf{CE}}\widehat{\mathcal{R}}_n(\omega)$ is $(\mathcal{F}', \mathcal{B}(\mathbb{R}))$-measurable for every fixed $\omega \in \Omega$. Furthermore the measurability holds when ρ is any law-invariant coherent risk measure on L^p for some $p \in [1, \infty)$.

(ii) If ρ is law-invariant, cash-additive, and positively homogeneous, then we obtain the representation

$$^{\mathsf{NA}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \omega') = \frac{\widehat{s}_{u_n}^{\mathsf{B}}(\omega, \omega')}{\sqrt{n}} \mathcal{R}_\rho(\mathcal{N}_{0,1}) + \widehat{m}_{u_n}^{\mathsf{B}}(\omega, \omega'), \tag{2.45}$$

where $\widehat{m}_{u_n}^{\mathsf{B}}$ and $\widehat{s}_{u_n}^{\mathsf{B}}$ refer to $\widehat{\mu}_{u_n}^{\mathsf{B}}$'s mean and standard deviation respectively; see (2.50) below. Due to the representation in (2.45), it is easily seen that $^{\mathsf{NA}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \,\cdot\,)$ is $(\mathcal{F}', \mathcal{B}(\mathbb{R}))$-measurable for every $\omega \in \Omega$ and every $n \in \mathbb{N}$.

(iii) In view of (2.45) and its analogue for $^{\mathsf{NA}}\widehat{\mathcal{R}}_n$, we have

$$^{\mathsf{NA}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \omega') - {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n(\omega) = n^{-1/2}\mathcal{R}_\rho(\mathcal{N}_{0,1})(\widehat{s}_{u_n}^{\mathsf{B}}(\omega, \omega') - \widehat{s}_{u_n}(\omega)) + (\widehat{m}_{u_n}^{\mathsf{B}}(\omega, \omega') - \widehat{m}_{u_n}(\omega)). \tag{2.46}$$

Theorem 2.2.2 indicates that, for fixed ω, the law of this expression in ω' can be seen as an approximation of the law of

$$^{\mathsf{NA}}\widehat{\mathcal{R}}_n(\cdot) - \mathcal{R}_n = n^{-1/2}\mathcal{R}_\rho(\mathcal{N}_{0,1})(\widehat{s}_{u_n}(\cdot) - s) + (\widehat{m}_{u_n}(\cdot) - m). \tag{2.47}$$

Since $\widehat{m}_{u_n}$ is an unbiased estimator for m, and $\widehat{s}_{u_n}$ is (nearly) an unbiased estimator for s, the mean of the expression in (2.47) nearly vanishes. So it may be expected that, for fixed ω, the mean of the expression in (2.46) (in ω') is close to 0, too. In particular one cannot expect that the mean of the expression in (2.46) (in ω') is a reasonable "estimator" for the bias of the estimator $^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ for $\widehat{\mathcal{R}}_n$.

◇

For the proof of the first part of Remark 2.4.4 we can use similar arguments as in the proof of 2.2.6.

Proof of part (i) of Remark 2.4.4. It suffices to prove the measurability of $^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{B}}(\omega, \,\cdot\,)$ for every $n \in \mathbb{N}$ and every fixed $\omega \in \Omega$. Let $\rho : L^p \to \mathbb{R}$ be a law-invariant coherent risk measure. First, Theorem 2.8 in [37] ensures that the corresponding risk functional $\mathcal{R}_\rho : \mathcal{M}(L^p) \to \mathbb{R}$ is continuous for the p-weak topology $\mathcal{O}_{p\text{-w}}$. The latter is defined to the the coarsest topology on $\mathcal{M}(L^p)$ w.r.t. which each of the maps $\mu \mapsto \int f\,d\mu$, $f \in C_{\mathrm{b}}^p$, is continuous, where C_{b}^p is the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which there exists a constant $C > 0$ such that $|f(x)| \leq C(1 + |x|^p)$ for all $x \in \mathbb{R}$. According to Corollary A.45 in [27] the topological space $(\mathcal{M}(L^p), \mathcal{O}_{p\text{-w}})$ is Polish. Second, the topology $\mathcal{O}_{p\text{-w}}$ is generated by the L^p-Wasserstein metric d_{Wass_p} and the mapping $\mathcal{M}(L^p) \to \mathcal{M}(L^p)$, $\mu \mapsto \mu^{*n}$, is $(d_{\mathrm{Wass}_p}, d_{\mathrm{Wass}_p})$-continuous; see Lemma 8.6 in [13]. Third, the mapping $(\omega, \omega') \mapsto \widehat{\mu}_{u_n}^{\mathsf{B}}(\omega, \omega'; \cdot)$ is $(\overline{\mathcal{F}}, \sigma(\mathcal{O}_{p\text{-w}}))$-measurable. Indeed, it is easily seen that the Borel σ-algebra $\sigma(\mathcal{O}_{p\text{-w}})$ on $\mathcal{M}(L^p)$ is generated by the maps $\mu \mapsto \int f d\mu$, $f \in C_{\mathrm{b}}^p$. To show the $(\mathcal{F}', \sigma(\mathcal{O}_{p\text{-w}}))$-measurability of the mapping $\Omega' \to \mathcal{M}(L^p), \omega' \mapsto \widehat{\mu}_{u_n}^{\mathsf{B}}(\omega, \omega')$ for every fixed $\omega \in \Omega$, it suffices to show that

$$\Big(\int f(x)\, \widehat{\mu}_{u_n}^{\mathsf{B}}((\omega, \,\cdot\,); dx)\Big)^{-1}(A) \in \mathcal{F}', \quad \text{for every } A \in \mathcal{B}(\mathbb{R}) \text{ and } f \in C_{\mathrm{b}}^p. \tag{2.48}$$

Since $\widehat{\mu}_{u_n}^{\mathsf{B}}(\omega,\omega')$ is a probability kernel from $(\overline{\Omega},\overline{\mathcal{F}})$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, we can conclude that the mapping $(\omega,\omega')\mapsto\int f(x)\,\widehat{\mu}_{u_n}^{\mathsf{B}}((\omega,\omega');dx)$ is $(\overline{\mathcal{F}},\mathcal{B}(\mathbb{R}))$-measurable for every $f\in C_{\mathrm{b}}^p$. This yields the $(\mathcal{F}',\mathcal{B}(\mathbb{R}))$-measurability of the mapping $\omega'\mapsto\int f(x)\,\widehat{\mu}_{u_n}^{\mathsf{B}}((\omega,\omega');dx)$ for every fixed $\omega\in\Omega$, because for every $A\in\mathcal{B}(\mathbb{R})$ and every $\omega\in\Omega$ we have

$$\Big(\int f(x)\widehat{\mu}_{u_n}^{\mathsf{B}}((\omega,\,\cdot\,);dx)\Big)^{-1}(A)=\Big\{\omega'\in\Omega':\int f(x)\,\widehat{\mu}_{u_n}^{\mathsf{B}}((\omega,\omega');dx)\in A\Big\}$$
$$=\Big\{\omega'\in\Omega':(\omega,\omega')\in\Big(\int f(x)\,\widehat{\mu}_{u_n}^{\mathsf{B}}(\,\cdot\,;dx)\Big)^{-1}(A)\Big\}\quad(2.49)$$

for every $n\in\mathbb{N}$. Now by the fact that $(\int f(x)\widehat{\mu}_{u_n}^{\mathsf{B}}(\,\cdot\,;dx))^{-1}(A)\in\overline{\mathcal{F}}$ for every $A\in\mathcal{B}(\mathbb{R})$, together with [15, Theorem 18.1] we conclude that the right-hand side of (2.49) lies in $\mathcal{F}'$.

Remark 2.4.5 Condition (a) and the second part of condition (c) of Theorem 2.4.3 are satisfied when the risk measure ρ is defined on L^p and the observations $Y_1,Y_2,\ldots$ lie in L^λ for some $\lambda>p\vee 2$. ◇

Remark 2.4.6 It was shown in Theorem 2.8 of [37] that the statistical functional $\mathcal{R}_\rho$ associated with any law-invariant coherent risk measure ρ on L^p with $p\in[1,\infty)$ is continuous for the so called $|\cdot|^p$-weak topology. Since for $\lambda>p$ the topology on $\mathcal{M}_1^\lambda$ generated by d_{ϕ_λ} is finer than the relative $|\cdot|^p$-weak topology on $\mathcal{M}_1^\lambda$, it follows that condition (d) is fulfilled for every law-invariant coherent risk measure on L^p and $\lambda>p$. ◇

To be able to present the proof of Theorem 2.4.3, we will first introduce some notation and state some useful results which will be needed throughout the proof. To this end, let

$$\begin{aligned}
\widehat{m}_u(\omega)&:=\int x\,\widehat{\mu}_u(\omega;dx)\\
\widehat{m}_u^{\mathsf{B}}(\omega,\omega')&:=\int x\,\widehat{\mu}_u^{\mathsf{B}}((\omega,\omega');dx)\\
\widehat{s}_u^{2}(\omega)&:=\int (x-\widehat{m}_u(\omega))^2\,\widehat{\mu}_u(\omega;dx)\\
\widehat{s}_u^{2,\mathsf{B}}(\omega,\omega')&:=\int (x-\widehat{m}_u^{\mathsf{B}}(\omega,\omega'))^2\,\widehat{\mu}_u^{\mathsf{B}}((\omega,\omega'),dx)\\
\widehat{m}_{\lambda,u}(\omega)&:=\int |x|^\lambda\,\widehat{\mu}_u(\omega;dx)\\
\widehat{m}_{\lambda,u}^{\mathsf{B}}(\omega,\omega')&:=\int |x|^\lambda\,\widehat{\mu}_u^{\mathsf{B}}((\omega,\omega');dx)
\end{aligned}\quad(2.50)$$

for every $(\omega,\omega')\in\overline{\Omega}$ denote the corresponding moments of $\widehat{\mu}_u(\omega)$ and $\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega')$. Furthermore, let $\widehat{s}_u^{\mathsf{B}}(\omega,\omega'):=(\widehat{s}_u^{2,\mathsf{B}}(\omega,\omega'))^{1/2}$ and $\widehat{s}_u(\omega):=(\widehat{s}_u^{2}(\omega))^{1/2}$ denote the standard deviations under $\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega')$ and $\widehat{\mu}_u(\omega)$, respectively.

The following Theorem 2.4.7 is known from Theorem 3.2 in [4]. In the special case of Efron's bootstrap it was proven much earlier in [13], Theorem 2.1 (a).

Theorem 2.4.7 *Assume that* $(W_{u,i})$ *satisfies Assumption 2.4.1. If* $\mu \in \mathcal{M}(L^2)$*, then*

$$\mathbb{P}' \circ \{\sqrt{u}\,(\widehat{m}_u^{\mathsf{B}}(\omega,\cdot) - \widehat{m}_u(\omega))\}^{-1} \xrightarrow{\mathsf{w}} \mathcal{N}_{0,s^2} \qquad \mathbb{P}\text{-}a.e.\ \omega.$$

Theorem 2.4.8 *Assume that* $(W_{u,i})$ *satisfies Assumption 2.4.1, and let* $\lambda \geq 0$*. If* $\mu \in \mathcal{M}(L^\lambda)$*, then*

$$\lim_{u\to\infty} \mathbb{P}'[\{\omega' \in \Omega' : d_{\phi_\lambda}(\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega'),\mu) \geq \eta\}] = 0 \qquad \text{for all } \eta > 0, \qquad \mathbb{P}\text{-}a.e.\ \omega. \tag{2.51}$$

For (2.51), note that the mapping $\omega' \mapsto d_{\phi_\lambda}(\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega'),\mu)$ is $(\mathcal{F}',\mathcal{B}(\mathbb{R}_+))$-measurable for any fixed ω. Indeed, denoting by $\widehat{F}_u^{\mathsf{B}}((\omega,\omega'),\,\cdot\,)$ and F the distribution functions of $\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega')$ and μ respectively, this is ensured by the representation

$$d_{\phi_\lambda}(\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega'),\mu) = \sup_{x\in\mathbb{Q}} |\widehat{F}_u^{\mathsf{B}}((\omega,\omega'),x) - F(x)|\phi_\lambda(x)$$

and the $(\mathcal{F}',\mathcal{B}(\mathbb{R}_+))$-measurability of the mapping $\omega' \mapsto \widehat{F}_u^{\mathsf{B}}((\omega,\omega'),x)$ for every $\omega \in \Omega$ and $x \in \mathbb{R}$.

Proof (of Theorem 2.4.8) By choosing $r = 0$ in Theorem 2.1 in [65], we deduce that $d_{\phi_\lambda}(\widehat{\mu}_u,\mu) \to 0$ $\mathbb{P}$-a.s. That is, the class $\{\phi_\lambda(x)\mathbb{1}_{(-\infty,x]} : x \in \mathbb{R}\}$ is a Glivenko–Cantelli class w.r.t. $\mathbb{P}$ in the sense of [61]; see p. 81 for a definition. Now the claim follows by an application of Lemma 3.6.16 in [61] (where $W_{u,i}/u$ plays the role of $W_{u,i}$ in [61]). Our assumptions A2. and A4. in Assumption 2.4.1 ensure that the weights $(W_{u,i})$ satisfy the assumptions of this lemma. For the application of the lemma in our specific setting, note that the star * can be skipped in the probability there, because we have seen above that the mapping $\omega' \mapsto d_{\phi_\lambda}(\widehat{\mu}_u^{\mathsf{B}}(\omega,\omega'),\mu)$ is $(\mathcal{F}',\mathcal{B}(\mathbb{R}_+))$-measurable for any fixed ω. Also note that outer almost sure convergence (as defined in part (iii) of Definition 1.9.1 in [61]) implies almost sure convergence (i.e. convergence almost everywhere) in the classical sense. The latter follows from Proposition 1.1 in [21]. □

The moment assumptions in the following two corollaries seem to be slightly too strong in the sense that it should be possible to replace $\mathcal{M}_1(L^\lambda)$ by $\mathcal{M}_1(L^{\lambda'})$ and $\mathcal{M}_1(L^2)$ respectively. However, since we will apply the nonuniform Berry–Esséen inequality (2.22) in the below proof of Theorem 2.4.3, we assumed $\lambda > 2$ in Theorem 2.2.2 anyway. That is, the two corollaries do not cause any additional assumption.

Corollary 2.4.9 *Assume that* $(W_{u,i})$ *satisfies Assumption 2.4.1, and let* $\lambda > \lambda' \geq 1$*. If* $\mu \in \mathcal{M}(L^\lambda)$*, then*

$$\lim_{u\to\infty} \mathbb{P}'[\{\omega' \in \Omega' : |\widehat{m}_{\lambda',u}^{\mathsf{B}}(\omega,\omega') - m_{\lambda'}| \geq \eta\}] = 0 \qquad \text{for all } \eta > 0, \qquad \mathbb{P}\text{-}a.e.\ \omega.$$

Proof Below we will show that the mapping $\nu \mapsto \int |x|^{\lambda'} \nu(dx)$ is Lipschitz continuous w.r.t. d_{ϕ_λ}. Using this, we obtain

$$\mathbb{P}'[\{\omega' \in \Omega' : |\widehat{m}^{\mathsf{B}}_{\lambda',u}(\omega,\omega') - m_{\lambda'}| \geq \eta\}] \leq \mathbb{P}'[\{\omega' \in \Omega' : L\, d_{\phi_\lambda}(\widehat{\mu}^{\mathsf{B}}_u(\omega,\omega'),\mu) \geq \eta\}]$$

for every $\omega \in \Omega$ and $\eta > 0$, where L denotes the corresponding Lipschitz constant. Now Theorem 2.4.8 implies the claim of the corollary.

It remains to show the mentioned Lipschitz continuity. The function $x \mapsto |x|^{\lambda'}$ generates the σ-finite Borel measure on $\mathbb{R}$ with Lebesgue density $x \mapsto \lambda'|x|^{\lambda'-1}$. Thus integration-by-parts yields

$$\begin{aligned}\int |x|^{\lambda'}\, \nu(dx) &= \lim_{b\to\infty} \int_{(-b,b]} |x|^{\lambda'}\, \nu(dx) \\ &= \lim_{b\to\infty} \Big(F_\nu(b)|b|^{\lambda'} - F_\nu(-b)|b|^{\lambda'} - \int_{(-b,b]} F_\nu(x-)\, \lambda'|x|^{\lambda'-1}\, dx \Big)\end{aligned}$$

for every Borel probability measure ν on $\mathbb{R}$. In particular, for any two such ν_1, ν_2,

$$\begin{aligned}&\Big| \int |x|^{\lambda'}\, \nu_1(dx) - \int |x|^{\lambda'}\, \nu_2(dx) \Big| \\ &\leq \lim_{b\to\infty} \Big(2d_{\phi_{\lambda'}}(\nu_1,\nu_2) + d_{\phi_\lambda}(\nu_1,\nu_2) \int_{(-b,b]} \phi_\lambda(x-)^{-1}\, \lambda'|x|^{\lambda'-1}\, dx \Big) \\ &\leq 2d_{\phi_\lambda}(\nu_1,\nu_2) + d_{\phi_\lambda}(\nu_1,\nu_2) \int \phi_\lambda(x)^{-1}\, \lambda'|x|^{\lambda'-1}\, dx \\ &= L\, d_{\phi_\lambda}(\nu_1,\nu_2)\end{aligned}$$

with $L := 2 + \int \phi_\lambda(x)^{-1} \lambda'|x|^{\lambda'-1}\, dx < \infty$. □

For Efron's bootstrap the following result is already known from part (b) of Theorem 2.1 in [13].

Corollary 2.4.10 *Assume that $(W_{u,i})$ satisfies Assumption 2.4.1, and let $\lambda > 2$. If $\mu \in \mathcal{M}(L^\lambda)$, then*

$$\lim_{u\to\infty} \mathbb{P}'[\{\omega' \in \Omega' : |\widehat{s}^{2,\mathsf{B}}_u(\omega,\omega') - s^2| \geq \eta\}] = 0 \qquad \text{for all } \eta > 0, \qquad \mathbb{P}\text{-a.e. } \omega \tag{2.52}$$

and

$$\lim_{u\to\infty} \mathbb{P}'[\{\omega' \in \Omega' : |\widehat{s}^{\mathsf{B}}_u(\omega,\omega') - s| \geq \eta\}] = 0 \qquad \text{for all } \eta > 0, \qquad \mathbb{P}\text{-a.e. } \omega. \tag{2.53}$$

Proof We clearly have

$$\begin{aligned}\mathbb{P}'[\{\omega' \in \Omega' : |\widehat{s}^{2,\mathsf{B}}_u(\omega,\omega') - s^2| \geq \eta\}] \;\leq\;& \mathbb{P}'[\{\omega' \in \Omega' : |\widehat{m}^{\mathsf{B}}_{2,u}(\omega,\omega') - m_2| \geq \eta/2\}] \\ &+ \mathbb{P}'[\{\omega' \in \Omega' : |(\widehat{m}^{\mathsf{B}}_u(\omega,\omega'))^2 - m^2| \geq \eta/2\}].\end{aligned}$$

By Corollary 2.4.9 the first summand converges to 0 for every $\eta > 0$, for $\mathbb{P}$-a.e. ω. The second summand converges to 0 for every $\eta > 0$, for $\mathbb{P}$-a.e. ω, by Theorem 2.4.7 and an ω-wise application of Slutsky's lemma. This proves (2.52). Moreover (2.53) follows by (2.52) and

$$|\widehat{s}_u^{\mathsf{B}}(\omega,\omega') - s| = |\widehat{s}_u^{2,\mathsf{B}}(\omega,\omega') - s^2|/(\widehat{s}_u^{\mathsf{B}}(\omega,\omega') + s) \le |\widehat{s}_u^{2,\mathsf{B}}(\omega,\omega') - s^2|/s.$$

This completes the proof. □

Proof of (2.44). For every $(\omega,\omega') \in \overline{\Omega}$ and $\omega \in \Omega$, let $\widehat{S}_{u_n}^{\mathrm{B},(\omega,\omega')}$ and $\widehat{S}_{u_n}^{\omega}$ be random variables on probability spaces $(\Omega^{(\omega,\omega')}, \mathcal{F}^{(\omega,\omega')}, \mathbb{P}^{(\omega,\omega')})$ and $(\Omega^{\omega}, \mathcal{F}^{\omega}, \mathbb{P}^{\omega})$, respectively. Assume that for every $(\omega,\omega') \in \overline{\Omega}$ the random variable $\widehat{S}_{u_n}^{\mathrm{B},(\omega,\omega')}$ has distribution $(\widehat{\mu}_{u_n}^{\mathsf{B}})^{*n}((\omega,\omega');\,\cdot\,)$ and that for every $\omega \in \Omega$, $\widehat{S}_{u_n}^{\omega}$ has distribution $(\widehat{\mu}_{u_n})^{*n}(\omega;\,\cdot\,)$. Furthermore, let $\nu_{u_n}^{\mathrm{B}}((\omega,\omega');\,\cdot\,)$ denote the law of the random variable

$$\widehat{Z}_{u_n}^{\mathrm{B},(\omega,\omega')}(\,\cdot\,) := \frac{\widehat{S}_{u_n}^{\mathrm{B},(\omega,\omega')}(\,\cdot\,) - n\,\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')}{\sqrt{n}\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')}$$

for every $(\omega,\omega') \in \overline{\Omega}$ and for every $\omega \in \Omega$, let $\nu_{u_n}(\omega;\,\cdot\,)$ denote the law of the random variable

$$\widehat{Z}_{u_n}^{\omega}(\,\cdot\,) := \frac{\widehat{S}_{u_n}^{\omega}(\,\cdot\,) - n\,\widehat{m}_{u_n}(\omega)}{\sqrt{n}\widehat{s}_{u_n}(\omega)}.$$

Then we observe that for every $(\omega,\omega') \in \overline{\Omega}$ we have

$$\mathrm{law}\{\sqrt{n}\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\widehat{Z}_{u_n}^{\mathrm{B},(\omega,\omega')} + n\,\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')\} \;=\; (\widehat{\mu}_{u_n}^{\mathsf{B}})^{*n}(\omega,\omega'),$$

and for every $\omega \in \Omega$

$$\mathrm{law}\{\sqrt{n}\widehat{s}_{u_n}(\omega)\,\widehat{Z}_{u_n}^{\omega} + n\,\widehat{m}_{u_n}(\omega)\} \;=\; (\widehat{\mu}_{u_n})^{*n}(\omega).$$

Using part (c) of Assumption 2.4.1 on the positive homogeneity and the cash additivity of ρ, we obtain

$$\begin{aligned}
&\sqrt{u_n}\Big(\frac{1}{n}\mathcal{R}_\rho((\widehat{\mu}_{u_n}^{\mathsf{B}})^{*n}(\omega,\omega')) - \frac{1}{n}\mathcal{R}_\rho((\widehat{\mu}_{u_n})^{*n}(\omega))\Big)\\
&= \sqrt{u_n}\Big(\frac{1}{n}\rho(\sqrt{n}\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\widehat{Z}_{u_n}^{\mathrm{B},(\omega,\omega')} + n\,\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')) - \frac{1}{n}\rho(\sqrt{n}\widehat{s}_{u_n}(\omega)\,\widehat{Z}_{u_n}^{\omega} + n\,\widehat{m}_{u_n}(\omega))\Big)\\
&= \sqrt{\frac{u_n}{n}}\,\Big(\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\rho(\widehat{Z}_{u_n}^{\mathrm{B},(\omega,\omega')}) - \widehat{s}_{u_n}(\omega)\,\rho(\widehat{Z}_{u_n}^{\omega})\Big) + \sqrt{u_n}\,(\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega') - \widehat{m}_{u_n}(\omega))\\
&= \sqrt{\frac{u_n}{n}}\,\Big(\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\mathcal{R}_\rho(\nu_{u_n}^{\mathrm{B}}(\omega,\omega')) - \widehat{s}_{u_n}(\omega)\,\mathcal{R}_\rho(\nu_{u_n}(\omega))\Big) + \sqrt{u_n}\,(\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega') - \widehat{m}_{u_n}(\omega))\\
&= \sqrt{\frac{u_n}{n}}\,\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\Big(\mathcal{R}_\rho(\nu_{u_n}^{\mathrm{B}}(\omega,\omega')) - \mathcal{R}_\rho(\mathcal{N}_{0,1})\Big) + \sqrt{\frac{u_n}{n}}\,\widehat{s}_{u_n}(\omega)\,\Big(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_{u_n}(\omega))\Big)\\
&\quad + \sqrt{\frac{u_n}{n}}\,\mathcal{R}_\rho(\mathcal{N}_{0,1})(\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega') - \widehat{s}_{u_n}(\omega)) + \sqrt{u_n}(\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega') - \widehat{m}_{u_n}(\omega))\\
&=: S_1(n;\omega,\omega') + S_2(n;\omega) + S_3(n;\omega,\omega') + S_4(n;\omega,\omega') \qquad (2.54)
\end{aligned}$$

for every $(\omega,\omega')\in\overline{\Omega}$. In the rest of the proof we will show that

$$\begin{aligned}
\mathbb{P}'[\{\omega'\in\Omega':|S_1(n;\omega,\omega')|\ge\eta\}] &\longrightarrow 0 \qquad \text{for all } \eta>0, \qquad \mathbb{P}\text{-a.e. } \omega, &(2.55)\\
S_2(n;\omega) &\longrightarrow 0 \qquad \mathbb{P}\text{-a.e. } \omega, &(2.56)\\
\mathbb{P}'[\{\omega'\in\Omega':|S_3(n;\omega,\omega')|\ge\eta\}] &\longrightarrow 0 \qquad \text{for all } \eta>0, \qquad \mathbb{P}\text{-a.e. } \omega, &(2.57)\\
\mathbb{P}'\circ S_4(n;\omega,\,\cdot\,)^{-1} &\overset{\mathsf{w}}{\longrightarrow} \mathcal{N}_{0,s^2} \qquad \mathbb{P}\text{-a.e. } \omega, &(2.58)
\end{aligned}$$

where $\mathbb{P}'\circ S_4(n;\omega,\,\cdot\,)^{-1}$ refers to the law of $\omega'\mapsto S_4(n;\omega,\omega')$ under $\mathbb{P}'$. Then (2.55)–(2.58) and an ω-wise application of Slutsky's lemma imply (2.44).

To show (2.55), note that by assumption (d) we can find (for given $\eta>0$) some $\eta'>0$ such that

$$\begin{aligned}
&\mathbb{P}'\Big[\Big\{\omega'\in\Omega':\Big|\mathcal{R}_\rho(\nu^{\mathsf{B}}_{u_n}((\omega,\omega')))-\mathcal{R}_\rho(\mathcal{N}_{0,1})\Big|\ge\eta\Big\}\Big] \qquad (2.59)\\
&\le\ \mathbb{P}'\Big[\Big\{\omega'\in\Omega':d_{\phi_\lambda}\Big(\nu^{\mathsf{B}}_{u_n}((\omega,\omega'))\,,\,\mathcal{N}_{0,1}\Big)\ge\eta'\Big\}\Big]\\
&\le\ \mathbb{P}'\Big[\Big\{\omega'\in\Omega':C_\lambda\,\frac{\int|x-\int y\,\widehat{\mu}^{\mathsf{B}}_{u_n}((\omega,\omega');dy)|^{\lambda'}\,\widehat{\mu}^{\mathsf{B}}_{u_n}((\omega,\omega');dx)}{(\int(x-\int y\,\widehat{\mu}^{\mathsf{B}}_{u_n}((\omega,\omega');dy))^2\,\widehat{\mu}^{\mathsf{B}}_{u_n}((\omega,\omega');dx))^{\lambda'/2}}\,n^{-\gamma}\ge\eta'\Big\}\Big]\\
&\le\ \mathbb{P}'\Big[\Big\{\omega'\in\Omega':C_\lambda\,2^{\lambda'-1}\,\frac{\widehat{m}^{\mathsf{B}}_{\lambda',u_n}(\omega,\omega')}{(\widehat{s}^{\mathsf{B}}_{u_n}(\omega,\omega'))^{\lambda'}}\,n^{-\gamma}\ge\eta'\Big\}\Big]\\
&\le\ \mathbb{P}'\Big[\Big\{\omega'\in\Omega':\Big|\frac{\widehat{m}^{\mathsf{B}}_{\lambda',u_n}(\omega,\omega')}{(\widehat{s}^{\mathsf{B}}_{u_n}(\omega,\omega'))^{\lambda'}}-\frac{m_{\lambda'}}{s^{\lambda'}}\Big|\,n^{-\gamma}\ge\eta'/(C_\lambda\,2^{\lambda'})\Big\}\Big]\\
&\quad\ +\mathbb{P}'\Big[\Big\{\omega'\in\Omega':\frac{m_\lambda}{s^{\lambda'}}\,n^{-\gamma}\ge\eta'/(C_\lambda\,2^{\lambda'})\Big\}\Big]\\
&=:\ S_{1,1}(n;\omega)+S_{1,2}(n).
\end{aligned}$$

where the second "$\le$" is justified by the nonuniform Berry–Essén inequality of Theorem 2.2.7 for arbitrary but fixed $\lambda'\in(2,\lambda)$ (and $\gamma:=\min\{1,\lambda'-2\}/2$). The summand $S_{1,2}(n)$ obviously converges to 0. Moreover, by Corollaries 2.4.9 and 2.4.10 along Slutsky's lemma (applied ω-wise) the summand $S_{1,1}(n;\omega)$ converges to 0 for $\mathbb{P}$-a.e. ω. That is, the left-hand side of (2.59) converges to 0 for $\mathbb{P}$-a.e. ω. Moreover by Corollary 2.4.10 we have that $\widehat{s}^{\mathsf{B}}_{u_n}(\omega,\,\cdot\,)$ converges in $\mathbb{P}'$-probability to s for $\mathbb{P}$-a.e. ω, and by assumption (b) we have $\sqrt{u_n/n}\to\sqrt{c}$. Then another (ω-wise) application of Slutsky's lemma leads to (2.55).

To show (2.56), we note that

$$\begin{aligned}
d_{\phi_\lambda}\Big(\nu_{u_n}(\omega)\,,\,\mathcal{N}_{0,1}\Big)\ &\le\ C_\lambda\,\frac{\int|x-\int y\,\widehat{\mu}_{u_n}(\omega;dy)|^\lambda\,\widehat{\mu}_{u_n}(\omega;dx)}{(\int(x-\int y\,\widehat{\mu}_{u_n}(\omega;dy))^2\,\widehat{\mu}_{u_n}(\omega;dx))^{\lambda/2}}\,n^{-\gamma}\\
&\le\ C_\lambda\,2^{\lambda-1}\,\frac{\widehat{m}_{\lambda,u_n}(\omega)}{(\widehat{s}_{u_n}(\omega))^\lambda}\,n^{-\gamma}
\end{aligned}$$

and that the ordinary strong law of large numbers ensures that the latter expression converges to 0 for $\mathbb{P}$-a.e. ω.

The convergence in (2.57) follows immediately from Corollary 2.4.10 and assumption (b). Finally (2.58) follows from Theorem 2.4.7. This completes the proof of (2.44). $\square$

Proof of (2.43). For every $(\omega,\omega')\in\overline{\Omega}$, let $\widehat{M}_{u_n}^{\mathrm{B},(\omega,\omega')}$ be a random variable on some probability space $(\Omega^{(\omega,\omega')},\mathcal{F}^{(\omega,\omega')},\mathbb{P}^{(\omega,\omega')})$ and assume that $\widehat{M}_{u_n}^{\mathrm{B},(\omega,\omega')}$ is distributed according to the normal distribution with mean $\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')$ and variance $\widehat{s}_{u_n}^{2,\mathsf{B}}(\omega,\omega')$. Let

$$\widehat{N}_{u_n}^{\mathrm{B},(\omega,\omega')}(\,\cdot\,):=\frac{\widehat{M}_{u_n}^{\mathrm{B},(\omega,\omega')}(\,\cdot\,)-n\,\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')}{\sqrt{n}\,\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')}.$$

Then we observe that $\widehat{N}_{u_n}^{\mathrm{B},(\omega,\omega')}$ has the standard normal distribution and for every $(\omega,\omega')\in\overline{\Omega}$ we have

$$\mathrm{law}\{\sqrt{n}\,\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\widehat{N}_{u_n}^{\mathrm{B},(\omega,\omega')}+n\,\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')\}\;=\;\mathcal{N}_{n\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega'),n\widehat{s}_{u_n}^{2,\mathsf{B}}(\omega,\omega')}.$$

Moreover, for every $\omega\in\Omega$, let $\widehat{Z}_{u_n}^{\omega}$ and $\nu_{u_n}(\omega)$ be defined as in the proof of (2.44).

Now, we again use part (c) of Assumption 2.4.1 on the positive homogeneity and the cash additivity of ρ to obtain

$$\begin{aligned}
&\sqrt{u_n}\Big(\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega'),n\widehat{s}_{u_n}^{2,\mathsf{B}}(\omega,\omega')})-\frac{1}{n}\mathcal{R}_\rho((\widehat{\mu}_{u_n})^{*n}(\omega))\Big)\\
&=\;\sqrt{u_n}\Big(\frac{1}{n}\rho(\sqrt{n}\,\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\widehat{N}_{u_n}^{\mathrm{B},(\omega,\omega')}+n\,\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega'))-\frac{1}{n}\,\rho(\sqrt{n}\widehat{s}_{u_n}(\omega)\,\widehat{Z}_{u_n}^{\omega}+n\,\widehat{m}_{u_n}(\omega))\Big)\\
&=\;\sqrt{\frac{u_n}{n}}\,\Big(\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\rho(\widehat{N}_{u_n}^{\mathrm{B},(\omega,\omega')})-\widehat{s}_{u_n}(\omega)\,\rho(\widehat{Z}_{u_n}^{\omega})\Big)+\sqrt{u_n}(\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')-\widehat{m}_{u_n}(\omega))\\
&=\;\sqrt{\frac{u_n}{n}}\,\Big(\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')\,\mathcal{R}_\rho(\mathcal{N}_{0,1})-\widehat{s}_{u_n}(\omega)\,\mathcal{R}_\rho(\nu_{u_n}(\omega))\Big)+\sqrt{u_n}(\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')-\widehat{m}_{u_n}(\omega))\\
&=\;\sqrt{\frac{u_n}{n}}(\widehat{s}_{u_n}^{\mathsf{B}}(\omega,\omega')-\widehat{s}_{u_n}(\omega))\mathcal{R}_\rho(\mathcal{N}_{0,1})+\sqrt{u_n}(\widehat{m}_{u_n}^{\mathsf{B}}(\omega,\omega')-\widehat{m}_{u_n}(\omega))\\
&\qquad+\sqrt{\frac{u_n}{n}}\,\widehat{s}_{u_n}(\omega)\,\Big(\mathcal{R}_\rho(\mathcal{N}_{0,1})-\mathcal{R}_\rho(\nu_{u_n}(\omega))\Big)\\
&=:\;S_1(n;\omega,\omega')+S_2(n;\omega)+S_3(n;\omega)
\end{aligned}$$

for every $(\omega,\omega')\in\overline{\Omega}$. Using arguments as in the proof of (2.55)–(2.58), we can conclude that

$$\mathbb{P}'[\{\omega'\in\Omega':\,|S_1(n;\omega,\omega')|\geq\eta\}]\;\longrightarrow\;0\qquad\text{for all }\eta>0,\quad\mathbb{P}\text{-a.e. }\omega,\tag{2.60}$$

$$\mathbb{P}'\circ S_2(n;\omega,\,\cdot\,)^{-1}\;\stackrel{\mathsf{w}}{\longrightarrow}\;\mathcal{N}_{0,s^2}\qquad\mathbb{P}\text{-a.e. }\omega,\tag{2.61}$$

$$S_3(n;\omega)\;\longrightarrow\;0\qquad\mathbb{P}\text{-a.e. }\omega.\tag{2.62}$$

Then (2.60)–(2.62) and an ω-wise application of Slutsky's lemma imply (2.43). This completes the proof. $\square$

2.5 Numerical simulations: The effect of the bootstrap-based bias correction

In this section we present some numerical examples for the method of correcting the bias discussed in Section 2.4.1. In our simulations we used Efron's bootstrap (see Example 2.4.2 (i)) to perform a bias correction of the original nonparametric estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ in the sense of (2.39). Other bootstrap methods are expected to lead to very similar results; see Table 2.5 and Figure 2.5. We fixed $n = 200$, chose $u_n = n$ and $u_n = 3n$, and estimated the individual premium, given by $\mathcal{R}_n = \frac{1}{n}\mathcal{R}_\rho(\mu^{*n})$. In our simulations the role of ρ is first played by the Average Value at Risk at level $\alpha = 0.95$ of Example 1.2.4, that is $\rho = \mathrm{AV@R}_\alpha$. Second, we chose $\rho = \mathrm{V@R}_\alpha$, which is the Value at Risk of Example 1.2.3 at level $\alpha = 0.95$. For the conditional single claim distribution

$$\mu_{>0}[\,\cdot\,] := \mu[\,\cdot\, \cap (0,\infty)]/\mu[(0,\infty)]$$

we considered the Pareto distribution and the log-normal distribution for different sets of parameters. For computational issues we discretized both distributions to the equidistant grid $0.1\mathbb{N}_0 := \{0, 0.1, 0.2, \ldots\}$. Note that $\mu_{>0}$ and $p := \mu[(0,\infty)]$ together determine the (unconditional) single claim distribution μ through the representation

$$\mu[\,\cdot\,] = (1-p)\,\delta_0[\,\cdot\,] + p\,\mu_{>0}[\,\cdot\,].$$

Here p is the probability of a strictly positive claim. We let $p = 0.1$ in all examples. In each setting, we simulated 1 000 independent observation vectors $(Y_1, \ldots, Y_{u_n})$. For each of these 1 000 vectors we computed ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$. The bias corrected estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ in the fashion of (2.39) were built upon $L = 500$ bootstrap paths conditional on each of the 1 000 "original" paths. The "exact" values were derived from a Monte Carlo simulation with 100 000 paths.

Our simulations show ambiguous results. Where on the one hand we experienced a reduction of the bias for $\mathrm{AV@R}_\alpha$, quite the contrary holds true for $\mathrm{V@R}_\alpha$. However, these results are similar to those of Kim and Hardy [33, 34]. Kim and Hardy too stated a good applicability of the bootstrap-based bias correction to $\mathrm{AV@R}_\alpha$, and observed that the same procedure might cause an increase of the estimated bias for $\mathrm{V@R}_\alpha$. A reason for this might be the fact that the Value at Risk lacks subadditivity and does therefore not provide a coherent risk measure. In both cases the use of the bootstrap-based bias correction has the effect to increase both the variance and the mean squared error (MSE).

2.5.1 Average Value at Risk

In this subsection we fix $\rho = \mathrm{AV@R}_\alpha$ and $\alpha = 0.95$. In the first example we let $\mu_{>0}$ be the Pareto distribution $\mathrm{Par}_{a,b}$ with parameters $a > 2$ and $b > 0$. The standard Lebesgue density of $\mathrm{Par}_{a,b}$ is

$$x \mapsto ab^{-1}(b^{-1}x + 1)^{-(a+1)}\, \mathbb{1}_{(0,\infty)}(x)$$

and the assumption $a > 2$ ensures again, that $\mathbb{E}[|Y_1|^\lambda] < \infty$ for all $\lambda \in (2, a)$. In our examples the parameters of the Pareto distribution were chosen such that the expected value of a single claim was normalised to 1, i.e. $\mathbb{E}[Y_1] = 1$. Tables 2.1–2.3 show the results of the simulation study in dependence on u_n and the choice of estimators. The tables show the (empirical) bias, standard deviation, and root mean squared error. Each value is shown in percentage of the true value.

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−18.38%	44.50%	49.88%	−12.86%	38.14%	40.89%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	−12.90%	50.41%	52.54%	−6.85%	45.46%	46.05%

Table 2.1: Estimators for $\rho = \mathrm{AV@R}_{0.95}$ and $\mu_{>0} = \mathrm{Par}_{3,20}$ (True value: 2.2585).

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−7.34%	36.60%	37.45%	−5.31%	20.63%	21.38%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	−2.71%	40.07%	40.17%	−0.25%	25.00%	25.00%

Table 2.2: Estimators for $\rho = \mathrm{AV@R}_{0.95}$ and $\mu_{>0} = \mathrm{Par}_{6,50}$ (True value: 1.8541).

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−4.68%	31.88%	32.26%	−3.23%	18.36%	18.66%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	1.10%	37.91%	37.92%	0.83%	22.47%	22.47%

Table 2.3: Estimators for $\rho = \mathrm{AV@R}_{0.95}$ and $\mu_{>0} = \mathrm{Par}_{10,90}$ (True value: 1.8114).

In the case $a = 3$ and $b = 20$ the conditional single claim distribution is a kind of "heavy-tailed". In view of this and the relatively small collective size, it is not surprising that the estimator shows a large negative bias and a large MSE. The application of the bootstrap-based method of Section 2.4.1 helps to reduce this bias by a third but has the effect to increase the MSE. The cases $a = 6$ and $b = 50$ and $a = 10$ and $b = 90$ refer to "medium-tailed" conditional single claim distributions. Again the estimator shows a negative bias. Especially in the case with $a = 10$ and $u_n = 3n$, the bootstrap-based method helps to get rid of the negative bias of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ and leads to a very small positive bias. Figure 2.3 shows the (empirical) law of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$. The vertical line in each plot represents the true value. One can see that in each case the bias correction has the effect to shift the mass of the distribution of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ to the right and thus makes the estimation more conservative. The negative bias

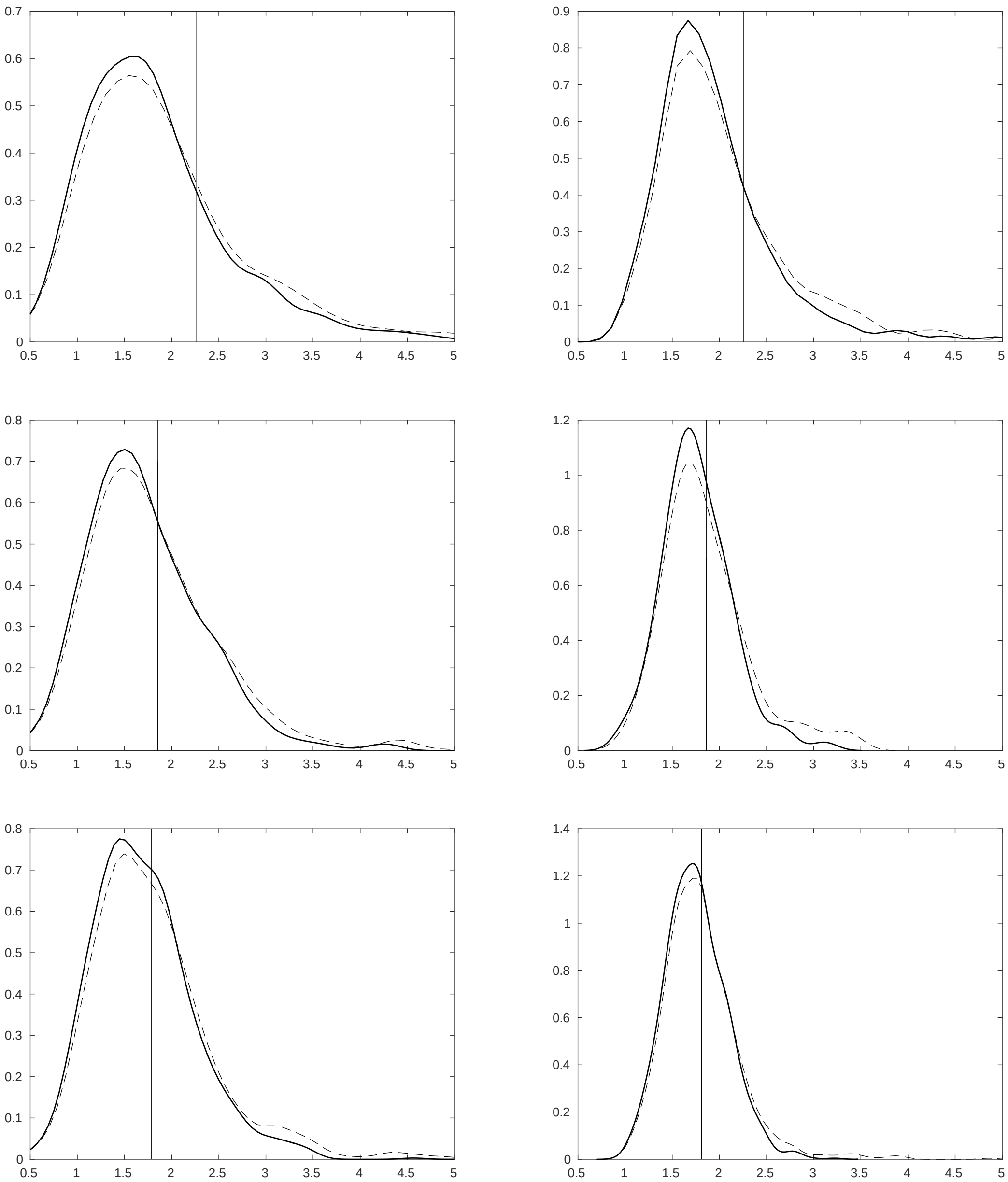

Figure 2.3: Empirical laws of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ (continuous line) and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ (dashed line) for $\rho = \mathrm{AV@R}_{95\%}$, $u_n = n$ (first column) as well as $u_n = 3n$ (second column), and $\mu_{>0} = \mathrm{Par}_{a,b}$ for the Pareto distribution $\mathrm{Par}_{a,b}$ with parameter $a = 3$ in the first line, $a = 6$ in the second line, and $a = 10$ in the third line.

is getting reduced. The law of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ possesses less mass on the left-hand side of the true value than the law of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$. However the law of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ possesses more mass on the outer right-hand side of the true value than the law of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$, even in such an extent that both the variance and the MSE of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ become larger than those of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$. Against the background of actuarial theory and the insurer's wish of a preferably conservative estimation, this effect is not so bad.

In the second example we considered $\mu_{>0} = \mathrm{LN}_{1.9,0.9}$, where LN_{c,σ^2} refers to the log-normal distribution with parameters $c \in \mathbb{R}$ and $\sigma^2 > 0$. The log-normal distribution possesses all moments and can thus be seen as a "light-tailed" distribution. The parameters c and σ^2 were chosen in such a way that the expected value and the variance of LN_{c,σ^2} coincided with the expected value and the variance under $\mathrm{Par}_{10,90}$ respectively. Again, the role of the risk measure ρ is played by $\mathrm{AV@R}_\alpha$ with $\alpha = 0.95$. Table 2.4 shows the results of our simulations. Just like in the Pareto case ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ shows a negative bias. The negative bias is fully eliminated by the bootstrap-based bias correction, yielding only a much smaller positive bias. Again, ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ tends to underestimate the risk in the collective quite strongly, whereas ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ only overestimates the risk slightly. Just like in the Pareto case we experience the usual increase of the variance and the MSE. However, the increase of the variance and the MSE is due to a right-shift of the mass of the law of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ (see Figure 2.4) which makes the estimation more conservative.

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−5.12%	34.80%	35.22%	−3.52%	23.36%	23.67%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	1.18%	40.29%	40.31%	1.47%	29.75%	29.79%

Table 2.4: Estimators for $\rho = \mathrm{AV@R}_{0.95}$ and $\mu_{>0} = \mathrm{LN}_{1.9,0.9}$ (True value: 1.8246).

We conclude this subsection with the following remark. We do not expect the particular choice of the bootstrap scheme to affect the outcome of our method essentially. In Table 2.5 and Figure 2.5 the results for Efron's bootstrap are compared to the analogous results for the Bayesian bootstrap (see Example 2.4.2 (ii)) for $\mu_{>0} = \mathrm{Par}_{10,90}$. Figure 2.5 shows the empirical laws of the original empirical plug-in estimator, the bias-corrected estimator based on Efron's bootstrap and the bias-corrected estimator based on the Bayesian bootstrap. One can see that the empirical laws of the bias-corrected estimators do not differ very much from each other. Both curves resemble each other strongly, such that the results for the Bayesian bootstrap are very much comparable to those for Efron's bootstrap.

Table 2.5 provides a comparison of the perfomance of both bootstrap techniques by means of the estimated bias, the standard deviation and the root mean squared error. Both procedures show similar results, that is, the formerly observed alleviation of the biases and the increase in the standard deviation. We do therefore not expect the choice of the bootstrap procedure to influence substantially.

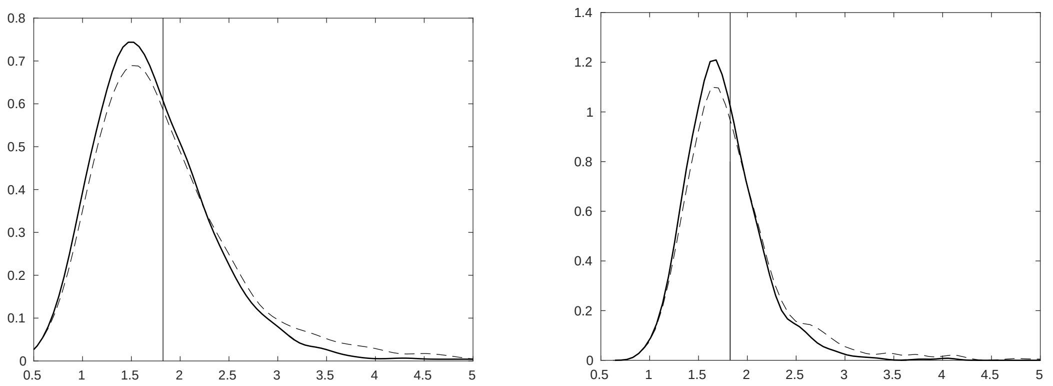

Figure 2.4: Empirical laws of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ (continuous line) and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ (dashed line) for $\rho = \mathrm{AV@R}_{95\%}$, $u_n = n$ (first column) as well as $u_n = 3n$ (second column), and $\mu_{>0} = \mathrm{LN}_{c,\sigma^2}$ for the log-normal distribution LN_{c,σ^2} with $c = 1.9$ and $\sigma^2 = 0.9$.

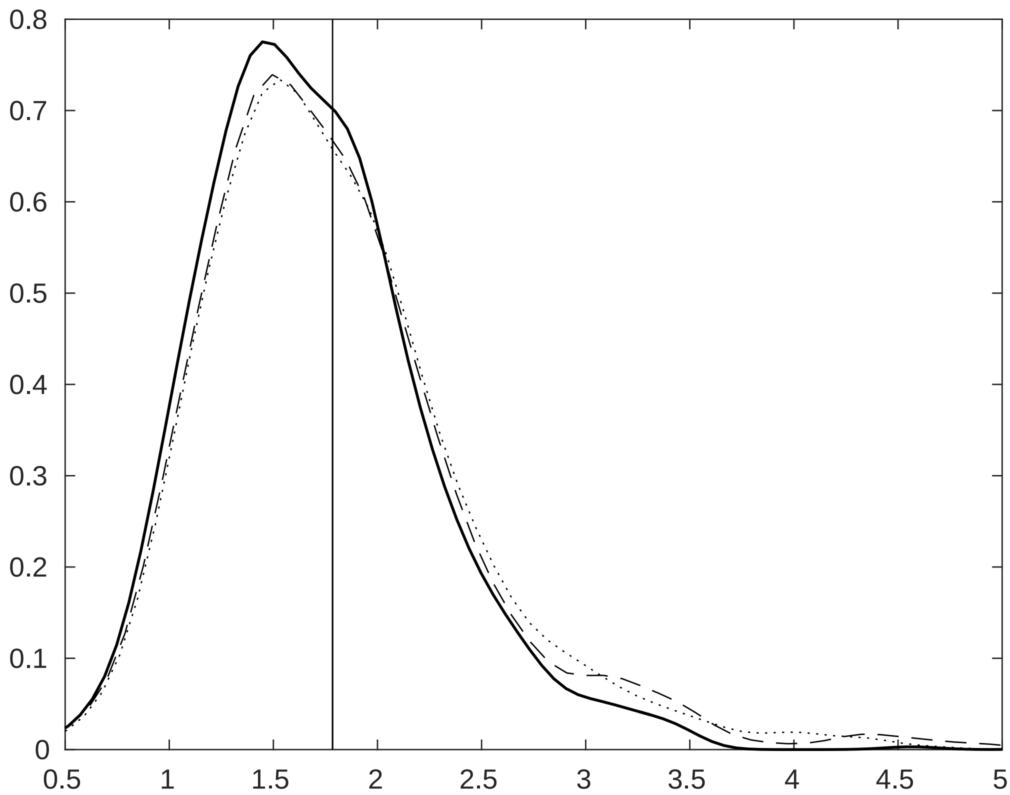

Figure 2.5: Empirical laws of ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ (continuous line), ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ derived from Efron's bootstrap (dashed line), and ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ derived from the Bayesian bootstrap (dotted line) for $\rho = \mathrm{AV@R}_{95\%}$, $u_n = n$, and $\mu_{>0} = \mathrm{Par}_{10,90}$.

	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−4.68%	31.88%	32.26%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ (Efron's bootstrap)	1.10%	37.91%	37.92%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ (Bayesian bootstrap)	1.66%	36.72%	36.76%

Table 2.5: Estimators for $\rho = \mathrm{AV@R}_{0.95}$, $u_n = n$, and $\mu_{>0} = \mathrm{Par}_{10,90}$ (True value: 1.8114).

2.5.2 Value at Risk

In this subsection we fix $\rho = \mathrm{V@R}_\alpha$ and $\alpha = 0.95$. In the first example we again consider $\mu_{>0} = \mathrm{Par}_{a,b}$ with $a > 2$ and $b > 0$. One can see that an application of the procedure of Section 2.4.1 can also have the effect to worsen the estimation. Again, ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ shows a negative bias, which increases with the heaviness of the tails of the underlying Pareto distribution. This effect has already been observed in our investigations in Section 2.3. The bootstrap-based method of Section 2.4.1 now increases both the bias and the MSE. Tables 2.6–2.8 show the results of our simulations.

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−8.21%	43.50%	44.41%	−4.11%	32.27%	32.56%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	−10.42%	51.04%	52.35%	−8.07%	43.14%	44.02%

Table 2.6: Estimators for $\rho = \mathrm{V@R}_{0.95}$ and $\mu_{>0} := \mathrm{Par}_{3,20}$ (True value: 1.7845).

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−4.62%	36.32%	36.64%	−2.92%	20.24%	20.47%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	−7.14%	42.88%	43.56%	−8.54%	34.72%	35.94%

Table 2.7: Estimators for $\rho = \mathrm{V@R}_{0.95}$ and $\mu_{>0} := \mathrm{Par}_{6,50}$ (True value: 1.6155).

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−2.68%	34.80%	35.22%	−1.75%	20.98%	21.05%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	−7.19%	41.44%	40.30%	−6.57%	31.24%	31.92%

Table 2.9: Estimators for $\rho = \mathrm{V@R}_{0.95}$ and $\mu_{>0} = \mathrm{LN}_{1.9,0.9}$ (True value: 1.5995).

In the second example we again consider $\mu_{>0} = \mathrm{LN}_{1.9,0.9}$, where again the parameters were chosen in such a way that the expected value and the variance coincided with the expected value and the variance of $\mathrm{Par}_{10,90}$. The results are consistent with those of the Pareto

	$u_n = n$			$u_n = 3n$		
	Bias	StD	rMSE	Bias	StD	rMSE
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$	−3.07%	28.94%	29.10%	−1.90%	17.83%	17.93%
${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$	−6.49%	35.54%	36.12%	−4.83%	25.68%	26.13%

Table 2.8: Estimators for $\rho = \mathrm{V@R}_{0.95}$ and $\mu_{>0} = \mathrm{Par}_{10,90}$ (True value: 1.5990).

example. For both ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ we observe a negative bias. This negative bias increases for ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$. Also, the method of Section 2.4.1 has the effect to increase the MSE. Table 2.9 shows the results of our simulations.

2.5.3 Conclusion

When the underlying risk measure ρ is the Average Value at Risk $\mathrm{AV@R}_\alpha$, the method of Section 2.4.1 provides a way to moderately improve the estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$, at least from the insurer's point of view. On the other hand, when the underlying risk measure ρ is the Value at Risk $\mathrm{V@R}_\alpha$, the method of Section 2.4.1 seems not to be useful for the estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$. For larger collective sizes the numerical specification (of $\widehat{\mu}_u^{*n}$ and thus) of the estimator ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n$ consumes some computing time. Due to the Monte Carlo simulation that comes along with the bootstrap, the computing time for the bias corrected version ${}^{\mathsf{CE}}\widehat{\mathcal{R}}_n^{\mathsf{bc}}$ is even much higher. So, when deciding whether or not to use the method of Section 2.4.1 for the Average Value at Risk, the required computing time should not be neglected. In our examples, the particular choice of the bootstrap scheme did not affect the outcome of the results.

Chapter 3

Nonparametric estimation of risk measures in the collective model

In the former chapter we have considered the so-called individual model of actuarial theory. In this model we assumed that every client in the insurance collective would produce a nonnegative claim during the next insurance period. Here we allowed the case, where several single claim amounts could be zero, whenever a client would not report a claim within the next insurance period. In the collective model of actuarial theory however, we assume that the whole collective produces a random number of strictly positive losses. In this context let (X_i) be a sequence of strictly positive i.i.d. random variables. Moreover let N be an $\mathbb{N}_0$-valued random variable on the same probability space, being independent of (X_i). The total claim amount in the collective model is then given by the so-called *random sum*

$$S_N := \sum_{i=1}^{N} X_i. \tag{3.1}$$

The investigation of asymptotic distributions of these random sums began with the work of Robbins ([55]) in 1948. A good summary about asymptotics of random sums can be found in the books by [25] and [30], for example.

Throughout this chapter we will consider the compound Poisson model, which in actuarial theory is also referred to as the *Cramér-Lundberg model* or *classical compound Poisson risk model*. It goes back to the work of [44]. The model assumes the times between two successive single claims to be exponentially distributed at a certain rate $\lambda > 0$. The stochastic process, modeling the number of claims occurring in dependence on the time is then a homogeneous Poisson process. This model is very popular in both non-life insurance mathematics and ruin theory.

Our aim in this chapter will be an estimation of individual premiums in the compound Poisson model similar to the individual model of Chapter 2. We will focus on the derivation of strong rates and asymptotic normality of the estimated individual premiums in dependence on the underlying collective size. In the former chapter the central tool, which was used to determine the strong rates was the nonuniform Berry-Esséen inequality in the form of Theorem 13 of Chapter V in [50]. In the setting of the compound Poisson model our proofs will strongly rely on a new Berry-Esséen type inequality for nonrandomly centered random sums in the form of [20]. This recently established inequality quantifies the rate of convergence of a suitably nonrandomly centered random sum to the standard normal distribution w.r.t. (L^1)-Wasserstein distance, which was introduced in (1.5).

The rest of this chapter is organized as follows. In Section 3.1 we will introduce the compound Poisson model in a mathematical way and introduce two estimators for the individual premium. In Section 3.2 we will formulate assumptions under which we will be able to prove our main theorems. Corollaries 3.2.4–3.2.6 will then state the strong rates and asymptotic normality of the estimated individual premiums for each choice of the estimators.

3.1 Estimators for the individual premium in the compound Poisson model

In this section we consider the so-called collective model of actuarial theory with respect to the Poisson distribution. In the literature this model is often referred to as the *Cramér-Lundberg model*. We suppose that the collective successively suffers losses at an exponential rate $\lambda_n > 0$, in dependence on the size $n \in \mathbb{N}$ of the underlying collective. We assume the single losses to be independent and identically distributed according to some distribution μ. In this case the total number of losses in a period of length $t \geq 0$ is given by $N_n(t) := \max\{k \in \mathbb{N} : \sum_{i=1}^{k} W_i^n \leq t\}$, where (W_i^n) is a sequence of Exp_{λ_n}-distributed random variables. That is the total claim amount until time t is given by $S_n(t) := \sum_{i=1}^{N_n(t)} X_i$, where (X_i) is a sequence of i.i.d. random variables with distribution μ, which is independent of (W_i^n). Note that $S_n := (S_n(t))_{t\geq 0}$ provides a compound Poisson process with rate λ_n and jump size distribution μ, and that $N_n := (N_n(t))_{t\geq 0}$ provides a Poisson process with intensity λ_n. In particular, the total claim distribution in a fixed insurance period of length $T > 0$, i.e. the distribution of $S_n(T)$, is given by the random convolution

$$\mu^{*\mathrm{Poiss}_{\lambda_n T}}[\,\cdot\,] := \sum_{k\in\mathbb{N}_0} \mu^{*k}[\,\cdot\,]\,\mathrm{Poiss}_{\lambda_n T}[\{k\}]$$

of μ with respect to the Poisson distribution $\mathrm{Poiss}_{\lambda_n T}$ with parameter $\lambda_n T$. An adequate individual premium w.r.t. a certain risk measure ρ is then given by

$$\mathcal{R}_n := \frac{1}{n}\,\mathcal{R}_\rho\Big(\mu^{*\mathrm{Poiss}_{\lambda_n T}}\Big), \tag{3.2}$$

where as before $\mathcal{R}_\rho$ refers to the statistical functional associated with the risk measure ρ as defined in (1.1). We identify with $X_1, \ldots, X_{N_n(T)}$ the single claims which will occur within the next insurance period. As we do not have information about these future claims, we will try to construct estimates of the distribution of the total claim amount $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$ on the basis of historically observed claims.

In the following we will introduce two possible estimators for the total claim size distribution $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$. We will first introduce an approach based on the convolution of the empirical measure w.r.t. the Poisson distribution with estimated parameter. The resulting estimator for the individual premium will be called the empirical plug-in estimator. Second, we will use the normal approximation with estimated parameters to estimate the total claim size distribution.

To this end, let (Y_i) be a sequence of i.i.d. random variables with distribution μ and $\widehat{N} = (\widehat{N}(t))_{t\geq 0}$ be a Poisson process with rate 1 on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and assume that (Y_i) and $\widehat{N}$ are independent. Then $\widehat{N}_n(t) := \widehat{N}(\lambda_n t)$, $t \geq 0$, defines a Poisson process with rate λ_n being independent of (Y_i). Let $\tau \in (0, \infty)$ be any (historical) time horizon. With the choice of τ we can model the fact that parameters and claim size distributions are usually estimated on the basis of claims from the last few insurance periods and not necessarily from the last period only. By choosing $\tau = 3T$ for instance, the insurance company would use data from the last three insurance periods of length T to estimate the future premium.

In this context the random variable $\widehat{N}_n(\tau)$ can be seen as the number of claims that occurred within a period of length $\tau > 0$, and the random variables $Y_i, \ldots, Y_{\widehat{N}_n(\tau)}$ can be seen as the corresponding claims. Now

$$\widehat{\lambda}_{n,\tau} := \frac{\widehat{N}_n(\tau)}{\tau} \tag{3.3}$$

provides an estimator for the exponential rate λ_n in dependence on the underlying collective size $n \in \mathbb{N}$ and time $\tau > 0$. Thus

$$\widehat{\mu}_{n,\tau} := \frac{1}{\widehat{N}_n(\tau)} \sum_{i=1}^{\widehat{N}_n(\tau)} \delta_{Y_i} \tag{3.4}$$

provides a reasonable estimator for the single claim distribution μ based on the time horizon $\tau > 0$ and underlying collective size $n \in \mathbb{N}$, whenever the observed number of losses $\widehat{N}_n(\tau)$ is strictly positive (otherwise we simply set $\widehat{\mu}_{n,\tau} := \delta_0$).

Based on (3.4), we can use

$$\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}} := (\widehat{\mu}_{n,\tau})^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}} \tag{3.5}$$

as an estimator for the total claim distribution $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$, and the corresponding plug-in estimator

$${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n := \frac{1}{n} \mathcal{R}_\rho\Big(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}}\Big) \tag{3.6}$$

to estimate the individual premium $\mathcal{R}_n$ as defined in (3.2). In the following we will refer to ${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n$ as the *empirical plug-in estimator*. Once again, the Panjer recursion of [49] provides a way to compute the right-hand side in (3.5) if the single claim distribution μ has support in

$h\mathbb{N}_0 := \{0, h, 2h, \ldots\}$ for some $h > 0$, see Appendix A.2 for a detailed discussion. Although $\widehat{\mu}_{n,\tau}$ has bounded support, the right-hand side in (3.5) has unbounded support. Therefore the estimator in (3.6) cannot be computed in finite time for tail-dependent risk functionals $\mathcal{R}_\rho$, such as the Average Value at Risk of Example 1.2.4, for instance. On the other hand, it can be computed in finite time for the Value at Risk of Example 1.2.3 for instance.

Similarly to the approach in the individual model, we can use the normal approximation with suitably estimated parameters to estimate the total claim size distribution $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$. The idea behind this choice is again the asymptotic normality of a suitably centered random sum. Indeed, Example 3 (i) in [55] has shown that for a Poisson random variable N with intensity $\lambda > 0$ and any sequence (ξ_i) of i.i.d. random variables with positive finite variance, being independent from N, we have

$$\mathrm{law}\left\{\frac{\sum_{i=1}^{N}\xi_i - \lambda\, m}{\sqrt{\lambda m^{(2)}}}\right\} \xrightarrow{\mathrm{w}} \mathcal{N}_{0,1} \qquad (\lambda \to \infty), \tag{3.7}$$

where m and $m^{(2)}$ denote the expectation and the second moment of ξ_1, respectively. That is, informally for "large" λ, we have

$$\mu^{*\mathrm{Poiss}_\lambda} \approx \mathcal{N}_{\lambda m, \lambda m^{(2)}}. \tag{3.8}$$

Note that, by Wald's formula, we observe that $\lambda\, m$ and $\lambda m^{(2)}$ are nothing but the mean and variance of the random variable $\sum_{i=1}^{N}\xi_i$, respectively. Using corresponding representations in our present setting, we let

$$\mathfrak{m}_n := \lambda_n T m, \tag{3.9}$$
$$\sigma_n^2 := \lambda_n T m^{(2)}, \tag{3.10}$$

denote the mean and variance of $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$. Motivated by (3.8), we can use

$$\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}, \widehat{\sigma}^2_{n,\tau}}$$

to estimate the total claim distribution based on an underlying collective of size $n \in \mathbb{N}$ and time horizon $\tau > 0$. Here $\widehat{\mathfrak{m}}_{n,\tau}$ and $\widehat{\sigma}^2_{n,\tau}$ are estimators for the true mean $\mathfrak{m}_n$ and the true variance σ_n^2 of $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$. Based on the representations in (3.9) and (3.10)

$$\widehat{\mathfrak{m}}_{n,\tau} := \widehat{\lambda}_{n,\tau}\, T\, \widehat{m}_{n,\tau} \tag{3.11}$$
$$\widehat{\sigma}^2_{n,\tau} := \widehat{\lambda}_{n,\tau}\, T\, \widehat{m}^{(2)}_{n,\tau}, \tag{3.12}$$

provide suitable estimators for $\mathfrak{m}_n$ and σ_n^2, respectively. Here $\widehat{m}_{n,\tau}$ and $\widehat{m}^{(2)}_{n,\tau}$ refer to the expected value and the second moment of $\widehat{\mu}_{n,\tau}$, respectively, that is

$$\widehat{m}_{n,\tau} := \frac{1}{\widehat{N}_n(\tau)} \sum_{i=1}^{\widehat{N}_n(\tau)} Y_i, \tag{3.13}$$

$$\widehat{m}^{(2)}_{n,\tau} := \frac{1}{\widehat{N}_n(\tau)} \sum_{i=1}^{\widehat{N}_n(\tau)} Y_i^2. \tag{3.14}$$

Note that $\widehat{\mathfrak{m}}_{n,\tau}$ and $\widehat{\sigma}^2_{n,\tau}$ are nothing but the mean and variance of $\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}$, respectively. Thus, the corresponding plug-in estimator

$$^{\mathsf{NA}}\widehat{\mathcal{R}}_n := \frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})$$

provides a second estimator for the individual premium in the collective model. In this context an estimation of the total claim distribution simply boils down to an estimation of the parameters in the sense of (3.11) and (3.12). Note that we will omit the dependence of $\mathfrak{m}_n$ and σ_n^2, as well as the dependence of all estimators, on T for the sake of a better reading.

For cash-additive and positively homogeneous risk measures ρ, the total premium derived from the normal approximation with estimated parameters has the following representation:

$$\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) \;=\; \widehat{\sigma}_{n,\tau}\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) \;+\; \widehat{\mathfrak{m}}_{n,\tau}. \tag{3.15}$$

Of course, one can obtain a similar representation for the case of the true parameters. In the following we will assume the exponential rate $\lambda_n > 0$ to be proportional to the collective size $n \in \mathbb{N}$. More explicitly, we will assume that λ_n/n converges to some constant $c \in (0,\infty)$ as $n \to \infty$. This notion is compatible with our approach for the estimation in the individual model and reflects the fact, that the number of expected claims during an insurance period should increase in the same way as the number of clients in the collective. Under the above assumptions, the results of Section 3.2 will show that under mild assumptions on the integrability of the underlying random variables and for a wide class of risk measures we have

$$n^r\left({}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n\right) \stackrel{\text{a.s.}}{\longrightarrow} 0, \tag{3.16}$$

$$n^r\left({}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n\right) \stackrel{\text{a.s.}}{\longrightarrow} 0, \tag{3.17}$$

for every $r < 1/2$ and

$$\mathbb{P}\circ\left\{\sqrt{\frac{n\tau}{cT^2}}\left({}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n\right)\right\}^{-1} \stackrel{\text{w}}{\longrightarrow} \mathcal{N}_{0,s^2+m^2}, \tag{3.18}$$

$$\mathbb{P}\circ\left\{\sqrt{\frac{n\tau}{cT^2}}\left({}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n\right)\right\}^{-1} \stackrel{\text{w}}{\longrightarrow} \mathcal{N}_{0,s^2+m^2}, \tag{3.19}$$

where m and s^2 refer to the mean and variance of μ, respectively. The results are comparable to those of Section 2.2. Again, formulae (3.18)–(3.19) imply that the convergences in (3.16) and (3.17) cannot hold for $n \geq 1/2$. For the convergence of the estimated premiums to the true ones, we will show even more, namely

$${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \mathcal{R}_n = \frac{1}{n}\left(\widehat{\mathfrak{m}}_{n,\tau} - \mathfrak{m}_n\right) \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}) \tag{3.20}$$

$${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n - \mathcal{R}_n = \frac{1}{n}\left(\widehat{\mathfrak{m}}_{n,\tau} - \mathfrak{m}_n\right) \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{3.21}$$

Again, the asymptotics for both estimators are exactly the same and are independent of the choice of the risk measure ρ. Keeping in mind that we have $\mathfrak{m}_n = \lambda_n T m$ by Wald's equation, and that λ_n/n was assumed to converge to some constant $c \in (0, \infty)$, the results are comparable formulae (2.14)–(2.15) in the individual model. If we do not focus on the individual premium as defined in (3.2), but on the total premium divided by the expected claim amount instead, we obtain the following somehow "nicer" representations:

$$\frac{1}{\widehat{\lambda}_{n,\tau} T} \mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}, \widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\lambda_n T} \mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}}) = \widehat{m}_{n,\tau} - m + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}) \tag{3.22}$$

$$\frac{1}{\widehat{\lambda}_{n,\tau} T} \mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}}) - \frac{1}{\lambda_n T} \mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}}) = \widehat{m}_{n,\tau} - m + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{3.23}$$

In this case the convergence of the estimated quantities to the true ones is purely driven by the convergence of the sample mean $\widehat{m}_{n,\tau} = \frac{1}{\widehat{N}_n(\tau)} \sum_{i=1}^{\widehat{N}_n(\tau)} Y_i$ to the true mean m. This effect has already been observed in formulae (2.14)–(2.15) in the individual model. Again, the asymptotics for both estimators are exactly the same and are not affected by the concrete choice of the risk measure ρ.

With the help of (3.18) and (3.19) we obtain the following asymptotic confidence intervals at level $(1-\alpha)$ for the individual premium:

$$\left[{}^{\mathsf{NA}}\widehat{\mathcal{R}}_n - \sqrt{\frac{\widehat{\lambda}_{n,\tau} T^2 \widehat{m}^{(2)}_{n,\tau}}{n^2 \tau}}\, \Phi^{-1}_{0,1}\Big(1 - \frac{\alpha}{2}\Big),\ {}^{\mathsf{NA}}\widehat{\mathcal{R}}_n + \sqrt{\frac{\widehat{\lambda}_{n,\tau} T^2 \widehat{m}^{(2)}_{n,\tau}}{n^2 \tau}}\, \Phi^{-1}_{0,1}\Big(\frac{\alpha}{2}\Big) \right]$$

and

$$\left[{}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n - \sqrt{\frac{\widehat{\lambda}_{n,\tau} T^2 \widehat{m}^{(2)}_{n,\tau}}{n^2 \tau}}\, \Phi^{-1}_{0,1}\Big(1 - \frac{\alpha}{2}\Big),\ {}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n + \sqrt{\frac{\widehat{\lambda}_{n,\tau} T^2 \widehat{m}^{(2)}_{n,\tau}}{n^2 \tau}}\, \Phi^{-1}_{0,1}\Big(\frac{\alpha}{2}\Big) \right]$$

where $\Phi_{0,1}$ denotes the distribution function of $\mathcal{N}_{0,1}$. Moreover, the results of Corollary 3.2.4 allow for the following asymptotic representation of the true individual premium in the collective model

$$\mathcal{R}_n = \frac{\mathfrak{m}_n}{n} + \frac{\sigma_n}{n} \mathcal{R}_\rho(\mathcal{N}_{0,1}) + o(n^{-1/2}). \tag{3.24}$$

Likewise we can obtain similar representations for the corresponding estimators:

$${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n = \frac{\widehat{\mathfrak{m}}_{n,\tau}}{n} + \frac{\widehat{\sigma}_{n,\tau}}{n} \mathcal{R}_\rho(\mathcal{N}_{0,1}), \tag{3.25}$$

$${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n = \frac{\widehat{\mathfrak{m}}_{n,\tau}}{n} + \frac{\widehat{\sigma}_{n,\tau}}{n} \mathcal{R}_\rho(\mathcal{N}_{0,1}) + o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{3.26}$$

In the following Section we will formulate assumptions under which the above results can be achieved and state our main theorems.

3.2 Strong rates and asymptotic normality for the individual premium estimators in the compound Poisson model

In this section we will present our main theorems. We will first summarize assumptions under which our results can be achieved. Then we will first formulate two theorems about the asymptotics of the total premium estimator divided by the expected number of claims. As corollaries we will then state the resulting asymptotics for the individual premium estimators. The formulation of the assumption involves the Wasserstein metric d_{Wass}, which was introduced in (1.5).

Assumption 3.2.1 *Let $\rho : \mathcal{X} \to \mathbb{R}$ be a law-invariant map, and $\mathcal{R}_\rho$ be the corresponding statistical functional introduced in (1.1). Suppose that the following assertions hold:*

(a) $\mu \in \mathcal{M}(L^3)$, *that is* $\mathbb{E}[|Y_1|^3] < \infty$.

(b) λ_n/n *converges to some constant* $c \in (0, \infty)$.

(c) ρ *is cash-additive and positively homogeneous, and* $\mathcal{M}_1^3 \subset \mathcal{M}(\mathcal{X})$.

(d) The restriction of $\mathcal{R}_\rho$ *to* $\mathcal{M}_1^3$ *is* $(d_{\text{Wass}}, |\cdot|)$*-continuous at* $\mathcal{N}_{0,1}$.

Note that it was shown in Theorem 1.4.1 that part (d) of Assumption 3.2.1 is always fulfilled, whenever ρ refers to a law-invariant and convex risk measure on L^1. The assumption does not impose a strong restriction.

Under the above assumptions we are now in a position to state our two main theorems. Theorems 3.2.2 and 3.2.3 yield strong rates and asymptotic normality for premiums derived from the normal approximation with both true and estimated parameters and the empirical plug-in estimator. However, in the formulations of the theorems we do not focus on the individual premium as defined in (3.2), but on the total premium divided by the expected value of the claim amount. We do this with the benefit of "nicer" representations. Corollaries 3.2.4 and 3.2.6 will then state the analogue for the individual premiums.

Theorem 3.2.2 (Estimated normal approximation) *Suppose that Assumption 3.2.1 is fulfilled. Then the following assertions hold:*

(i) $\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2}) \;=\; \widehat{m}_{n,\tau} - m \,+\, o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mu^{*\text{Poiss}_{\lambda_n T}}) \;=\; o(n^{-1/2})$.

(iii) $\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mu^{*\text{Poiss}_{\lambda_n T}}) \;=\; \widehat{m}_{n,\tau} - m \,+\, o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(iv) $(\lambda_n T)^r\,(\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mu^{*\text{Poiss}_{\lambda_n T}})) \longrightarrow 0$ $\mathbb{P}$*-a.s. for every* $r < 1/2$.

(v) $\mathbb{P}\circ\{\sqrt{\lambda_n\tau}\,(\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})-\frac{1}{\lambda_nT}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}}))\}^{-1}\stackrel{\mathrm{w}}{\longrightarrow}\mathcal{N}_{0,s^2}.$

Theorem 3.2.3 (Empirical plug-in estimator) *Suppose that Assumption 3.2.1 is fulfilled and assume that $\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})$ is $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-measurable for every $n\in\mathbb{N}$. Then the following assertions hold:*

(i) $\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})-\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})\;=\;o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}).$

(ii) $\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})-\frac{1}{\lambda_nT}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}})\;=\;\widehat{m}_{n,\tau}-m\;+\;o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}).$

(iii) $(\lambda_nT)^r\,(\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})-\frac{1}{\lambda_nT}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}}))\longrightarrow 0$ *$\mathbb{P}$-a.s. for every $r<1/2$.*

(iv) $\mathbb{P}\circ\{\sqrt{\lambda_n\tau}\,(\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})-\frac{1}{\lambda_nT}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}}))\}^{-1}\stackrel{\mathrm{w}}{\longrightarrow}\mathcal{N}_{0,s^2}.$

What strikes the most, is the fact, that in Theorem 3.2.2 (v) and Theorem 3.2.3 (iv) the asymptotic behavior of the estimators in this setting is not affected by the distribution of the number of claims $\mathrm{Poiss}_{\lambda_nT}$. Before we turn to the proofs of the above theorems, we first take our time to state two useful corollaries about the strong rates and asymptotic normality of the individual premiums.

The first corollary is concerned with the strong rates and asymptotic normality of the individual premiums derived from the normal approximation. Parts (iv) and (v) of Corollary 3.2.4 describe the asymptotic behavior of the sequence of estimators ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n:=\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})$ and their rate of convergence to the true premium $\mathcal{R}_n:=\frac{1}{n}\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}})$.

Corollary 3.2.4 (Estimated normal approximation) *Suppose that Assumption 3.2.1 is fulfilled. Then the following assertions hold:*

(i) $\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})-\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma^2_n})\;=\;\frac{1}{n}\,(\widehat{\mathfrak{m}}_{n,\tau}-\mathfrak{m}_n)\;+\;o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}).$

(ii) $\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma^2_n})-\frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}})\;=\;o(n^{-1/2}).$

(iii) $\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})-\frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}})\;=\;\frac{1}{n}\,(\widehat{\mathfrak{m}}_{n,\tau}-\mathfrak{m}_n)\;+\;o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}).$

(iv) $n^r\,(\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})-\frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}}))\longrightarrow 0$ *$\mathbb{P}$-a.s. for every $r<1/2$.*

(v) $\mathbb{P}\circ\{\sqrt{\frac{n\tau}{cT^2}}\,(\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}})-\frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_nT}}))\}^{-1}\stackrel{\mathrm{w}}{\longrightarrow}\mathcal{N}_{0,s^2+m^2}.$

The following corollary states the corresponding results for the estimator ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$. It is a direct consequence of parts (iv) and (v) of the above corollary and does therefore not need to be proved.

Corollary 3.2.5 *Under the assumptions in 3.2.1 parts (iv) and (v) of Corollary 3.2.4 show that the convergences in (3.16) and (3.18) hold true.*

Under the assumptions in 3.2.1, the following result provides the analogue to Corollary 3.2.4 for the empirical plug-in estimator in the collective model for the individual premium. It also gives an equivalent to Theorem 2.2.4 in the collective model. Assertions (iii) and (iv) in Corollary 3.2.6 describe the asymptotic behavior of the sequence of estimators ${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n := \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})$ for the individual premium $\mathcal{R}_n := \frac{1}{n}\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})$.

Corollary 3.2.6 (Empirical plug-in estimator) *Suppose that Assumption 3.2.1 is fulfilled and assume that $\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})$ is $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$. Then the following assertions hold:*

(i) $\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{n}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}) \;=\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(ii) $\frac{1}{n}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}) - \frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}}) \;=\; \frac{1}{n}\,(\widehat{\mathfrak{m}}_{n,\tau} - \mathfrak{m}_n) \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})$.

(iii) $n^r\,(\frac{1}{n}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}) - \frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})) \longrightarrow 0$ *$\mathbb{P}$-a.s. for every $r < 1/2$.*

(iv) $\mathbb{P}\circ\{\sqrt{\frac{n\tau}{cT^2}}\,(\frac{1}{n}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}) - \frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}}))\}^{-1} \xrightarrow{\mathrm{w}} \mathcal{N}_{0,s^2+m^2}$.

The following corollary states the corresponding results for the estimator ${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n$. It is a direct consequence of parts (iii) and (iv) of the above corollary and does therefore not need to be proved.

Corollary 3.2.7 *Suppose that the assumptions in 3.2.1 are fulfilled and that $\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})$ is $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$. Then parts (iii) and (iv) of Corollary 3.2.6 show that the convergences in (3.17) and (3.19) hold true.*

As a direct consequence of Assumption 3.2.1 and Corollaries 3.2.4 and 3.2.6 we obtain the following asymptotic representations of the estimated individual premiums:

$$
{}^{\mathsf{NA}}\widehat{\mathcal{R}}_n = \frac{\widehat{\mathfrak{m}}_{n,\tau}}{n} + \frac{\widehat{\sigma}_{n,\tau}}{n}\mathcal{R}_\rho(\mathcal{N}_{0,1}) \tag{3.27}
$$

$$
{}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n = \frac{\widehat{\mathfrak{m}}_{n,\tau}}{n} + \frac{\widehat{\sigma}_{n,\tau}}{n}\mathcal{R}_\rho(\mathcal{N}_{0,1}) \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{3.28}
$$

Equation (3.27) directly evolves from part (c) of Assumption 3.2.1, whereas equation (3.28) is a direct consequence of (3.27) in combination with part (i) of Theorem 3.2.3.

Remark 3.2.8 *(i) Note that the individual premium estimator ${}^{\mathsf{NA}}\widehat{\mathcal{R}}_n$ based on the normal approximation with estimated parameters of Theorem 3.2.2 and Corollary 3.2.4 is always $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-measurable due to the representation in (3.15).*

(ii) Let $\mathcal{X} = L^p$ for some $p \in [1,\infty)$. Then for every law-invariant coherent risk measure $\rho : L^p \to \mathbb{R}$ the estimator ${}^{\mathsf{PCE}}\widehat{\mathcal{R}}_n$ is $(\mathcal{F},\mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$.

Proof of part (ii) Let $\rho : L^p \to \mathbb{R}$ be a law-invariant coherent risk measure. First, Theorem 2.8 in [37] ensures that the corresponding risk functional $\mathcal{R}_\rho : \mathcal{M}(L^p) \to \mathbb{R}$ is continuous for the p-weak topology $\mathcal{O}_{p\text{-w}}$. According to Corollary A.45 in [27] the topological space $(\mathcal{M}(L^p), \mathcal{O}_{p\text{-w}})$ is Polish. Second, the topology $\mathcal{O}_{p\text{-w}}$ is generated by the L^p-Wasserstein metric d_{Wass_p}. The mapping $\mathcal{M}(L^p) \times (0,\infty) \to \mathcal{M}(L^p)$, $(\mu, \lambda) \mapsto \mu^{*\text{Poiss}_\lambda}$, is $(d_{\text{Wass}_p}, d_{\text{Wass}_p})$-continuous; see Appendix B for the proof. Third, the mappings $\omega \mapsto \widehat{\lambda}_{n,\tau}(\omega)$ and $\omega \mapsto \widehat{\mu}_{n,\tau}(\omega, \cdot)$ are $(\mathcal{F}, \sigma(\mathcal{O}_{p\text{-w}}))$-measurable. The first statement holds true due to the representation in (3.3). The latter is easily seen, because the Borel σ-algebra $\sigma(\mathcal{O}_{p\text{-w}})$ on $\mathcal{M}(L^p)$ is generated by the maps $\mu \mapsto \int f d\mu$, $f \in C_{\text{b}}^p$. Here C_{b}^p is again the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ for which there exists a constant $C > 0$ such that $|f(x)| \le C(1+|x|^p)$ for all $x \in \mathbb{R}$. So, for $(\mathcal{F}, \sigma(\mathcal{O}_{p\text{-w}}))$-measurability of the mapping $\Omega \to \mathcal{M}(L^p)$, $\omega \mapsto \widehat{\mu}_{n,\tau}(\omega, \cdot)$, it suffices to show

$$\Big(\int f(x)\, \widehat{\mu}_{n,\tau}(\,\cdot\,, dx) \Big)^{-1}(A) \in \mathcal{F} \quad \text{for all } A \in \mathcal{B}(\mathbb{R}) \text{ and } f \in C_{\text{b}}^p. \tag{3.29}$$

Since $\widehat{\mu}_{n,\tau}(\omega, \cdot)$ is a probability kernel from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the mapping

$$\omega \mapsto \int f(x)\, \widehat{\mu}_{n,\tau}(\omega, dx)$$

is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable for every $f \in C_{\text{b}}^p$; see e.g. Lemma 1.41 in [35]. This gives (3.29). Altogether, we have shown that the mapping $\omega \mapsto \mathcal{R}_\rho((\widehat{\mu}_{n,\tau}(\omega))^{*\text{Poiss}_{\widehat{\lambda}_{n,\tau}(\omega)T}})$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable. □

In the following remark we will discuss an option to obtain better rates of convergence in part (i) of Theorem 3.2.2 and part (ii) of Theorem 3.2.3 (and of course their analogues for the individual premiums in Corollaries 3.2.4 and 3.2.6) under a slightly stronger assumption as the one imposed by part (d) of Assumption 3.2.1. Similar to our investigations in Chapter 2 we can replace the assumption on the $(d_{\text{Wass}}, |\cdot|)$-continuity of $\mathcal{R}_\rho$ at $\mathcal{N}_{0,1}$ by the stronger notion of β-Hölder continuity for some $\beta > 0$ with the benefit of better rates of convergence. We will just state the results. The changes in the corresponding proofs are analogues to the proof of Remark 2.2.8.

Remark 3.2.9 *Note that we can achieve better rates of convergence in part (i) of Theorem 3.2.2 and part (ii) of Theorem 3.2.3 (and of course their analogues for the individual premiums in Corollaries 3.2.4 and 3.2.6) if we replace part (d) of Assumption 3.2.1 by the following slightly stronger assumption:*

(d') For each sequence $(\nu_n) \subset \mathcal{M}_1^3$ with $d_{\text{Wass}}(\nu_n, \mathcal{N}_{0,1}) \to 0$, there exist constants $L, \beta > 0$ such that

$$|\mathcal{R}_\rho(\nu_n) - \mathcal{R}_\rho(\mathcal{N}_{0,1})| \le L\, d_{\text{Wass}}(\nu_n, \mathcal{N}_{0,1})^\beta \tag{3.30}$$

for every $n \in \mathbb{N}$.

Then we have:

(i') $\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}) \;=\; \mathcal{O}_{\mathbb{P}\text{-a.s.}}(n^{-1/2(1+\beta)})$.

(ii') $\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}) \;=\; \mathcal{O}_{\mathbb{P}\text{-a.s.}}(n^{-1/2(1+\beta)})$.

3.2.1 Proof of Theorems 3.2.2 and 3.2.3

The proofs of Theorems 3.2.2 and 3.2.3 strongly rely on the following proposition, which gives a rate of growth of the Poisson process in relation to the underlying intensity. The proposition provides a strong law of a Marcinkiewicz-Zygmund type for the Poisson process.

Proposition 3.2.10 *Let $(\widehat{N}_n)_{n\in\mathbb{N}} := ((\widehat{N}_n(t))_{t\geq 0})_{n\in\mathbb{N}}$ be a sequence of Poisson processes, such that for every $n\in\mathbb{N}$, $\widehat{N}_n := (\widehat{N}_n(t))_{t\geq 0}$ is a Poisson process with rate $\lambda_n > 0$. Suppose that $(\lambda_n)\subset(0,\infty)$ and that λ_n/n converges to some strictly positive constant. Then for every $r < 1/2$ and every fixed $t > 0$ we have*

$$n^r\left|\frac{\widehat{N}_n(t)}{\lambda_n t} - 1\right| \longrightarrow 0 \quad (n\to\infty) \quad \mathbb{P}\text{-}a.s.$$

Proof Let $(\widetilde{N}(t))_{t\geq 0}$ be a Poisson process with rate 1. Then we observe that

$$\widehat{N}_n(t) \stackrel{d}{=} \widetilde{N}(\lambda_n t)$$

holds for every $n\in\mathbb{N}$ and $t\geq 0$. Now the claim is a direct consequence of Theorem 2.5.10 in [25] and the fact that λ_n/n converges to some $c\in(0,\infty)$. □

Moreover we will use the following Berry-Esséen inequality for nonrandomly centered random sums. The inequality provides a rate of convergence of the centered random sum to the standard normal distribution. The theorem and its proof can be found in [20] Corollary 2.12. It involves the Wasserstein metric d_{Wass}, which was introduced in (1.5).

Theorem 3.2.11 *Suppose (ξ_i) is a sequence of i.i.d. random variables on some probability space $(\Omega,\mathcal{F},\mathbb{P})$ satisfying $\mathbb{E}[|\xi_1|^3] < \infty$ and $\mathbb{E}[\xi_1^2] > 0$, and that N is a* Poiss$_\lambda$*-distributed random variable on the same probability space for some $\lambda > 0$, being independent from (ξ_i). Let $\mathfrak{m} := \lambda\,\mathbb{E}[\xi_1]$ and $\sigma^2 := \lambda\,\mathbb{E}[\xi_1^2]$ and set*

$$W \;:=\; \frac{\sum_{i=1}^N \xi_i - \mathfrak{m}}{\sigma}.$$

Then,

$$d_{\mathrm{Wass}}(\mathbb{P}_W,\mathcal{N}_{0,1}) \leq \frac{1}{\sqrt{\lambda}}\Big(\frac{2\,\mathrm{Var}[\xi_1]}{\mathbb{E}[\xi_1^2]} + \frac{3\,\mathbb{E}[|\xi_1-\mathbb{E}[\xi_1]|^3]}{\mathbb{E}[\xi_1^2]^{3/2}} + \frac{|\mathbb{E}[\xi_1]|}{\mathbb{E}[\xi_1^2]^{1/2}}\Big). \tag{3.31}$$

Note that the conditions in Theorem 3.2.11 are automatically fulfilled in our present setting, that is under Assumption 3.2.1.

Proof of Theorem 3.2.2 (i) For every $\omega \in \Omega$, let $\widehat{S}^{\omega}_{n,\tau}$ be a $\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}$-distributed random variable on some probability space $(\Omega^{\omega}, \mathcal{F}^{\omega}, \mathbb{P}^{\omega})$ and let R_n be a $\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2}$-distributed random variable. Let

$$\widehat{M}^{\omega}_{n,\tau}(\cdot) := \frac{\widehat{S}^{\omega}_{n,\tau}(\cdot) - \widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\sigma}_{n,\tau}(\omega)}$$

and

$$Z_n(\cdot) := \frac{R_n(\cdot) - \mathfrak{m}_n}{\sigma_n}.$$

Then we observe that both $\widehat{M}^{\omega}_{n,\tau}$ and Z_n have the standard normal distribution for every $\omega \in \Omega$. Using part (c) of Assumption 3.2.1 on the positive homogeneity and cash-invariance of ρ, we obtain for every $\omega \in \Omega$

$$\begin{aligned}
&\frac{1}{\widehat{\lambda}_{n,\tau}(\omega)T}\,\mathcal{R}_{\rho}(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}) - \frac{1}{\lambda_n T}\,\mathcal{R}_{\rho}(\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2})\\
&= \frac{1}{\widehat{\lambda}_{n,\tau}(\omega)T}\,\rho(\widehat{\sigma}_{n,\tau}(\omega)\,\widehat{M}^{\omega}_{n,\tau} + \widehat{\mathfrak{m}}_{n,\tau}(\omega)) - \frac{1}{\lambda_n T}\,\rho(\sigma_n\,Z_n + \mathfrak{m}_n)\\
&= \frac{\widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}(\omega)T} - \frac{\mathfrak{m}_n}{\lambda_n T} + \Big(\frac{\widehat{\sigma}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}T} - \frac{\sigma_n}{\lambda_n T}\Big)\,\mathcal{R}_{\rho}(\mathcal{N}_{0,1})\\
&= \widehat{m}_{n,\tau}(\omega) - m + \Big(\frac{\widehat{\sigma}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}(\omega)T} - \frac{\sigma_n}{\lambda_n T}\Big)\,\mathcal{R}_{\rho}(\mathcal{N}_{0,1}),
\end{aligned}\tag{3.32}$$

where we used the fact that $\widehat{\mathfrak{m}}_{n,\tau} = \widehat{\lambda}_{n,\tau}T\widehat{m}_{n,\tau}$ holds, and the equivalent for the case of known parameters. The latter holds true by Wald's equation. Now the claim would follow by showing that

$$\frac{\widehat{\sigma}_{n,\tau}}{\widehat{\lambda}_{n,\tau}T} - \frac{\sigma_n}{\lambda_n T} = o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}).$$

Denoting by $m^{(2)} := \int x^2\,\mu(dx)$ and $\widehat{m}^{(2)}_{n,\tau} := \int x^2\,\widehat{\mu}_{n,\tau}(dx)$, we can use Wald's equation again, yielding

$$\frac{\widehat{\sigma}_{n,\tau}}{\widehat{\lambda}_{n,\tau}T} - \frac{\sigma_n}{\lambda_n T} = \frac{\widehat{\sigma}_{n,\tau}}{\lambda_n T} - \frac{\sigma_n}{\lambda_n T} + \Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T} - 1\Big)\,\frac{\widehat{\sigma}_{n,\tau}}{\lambda_n T}\tag{3.33}$$

For the first summand we observe that

$$\begin{aligned}
\frac{\widehat{\sigma}_{n,\tau}}{\lambda_n T} - \frac{\sigma_n}{\lambda_n T} &= \frac{\widehat{\sigma}^2_{n,\tau} - \sigma_n^2}{\lambda_n T\,(\widehat{\sigma}_{n,\tau} + \sigma_n)}\\
&\le \frac{\widehat{\sigma}^2_{n,\tau} - \sigma_n^2}{\lambda_n T\,\sigma_n}\\
&= \frac{\widehat{\lambda}_{n,\tau}\,T\,\widehat{m}^{(2)}_{n,\tau} - \lambda_n T m^{(2)}}{(\lambda_n T)^{3/2}\,\sqrt{m^{(2)}}}\\
&= (m^{(2)}\,n)^{-1/2}\,\Big(\sqrt{\frac{n}{\lambda_n T}}\Big(\frac{\widehat{N}_n(\tau)}{\lambda_n \tau} - 1\Big)\,\widehat{m}^{(2)}_{n,\tau} + \sqrt{\frac{n}{\lambda_n T}}\,(\widehat{m}^{(2)}_{n,\tau} - m^{(2)})\Big).
\end{aligned}\tag{3.34}$$

From Theorem 2.5.5 in [25] we can derive the $\mathbb{P}$-a.s. convergence of $\widehat{m}_{n,\tau}^{(2)} - m^{(2)}$ to zero. Part (b) of Assumption 3.2.1 now guarantees that $(n/\lambda_n)^{1/2}$ converges to some constant $c^{-1/2} \in (0, \infty)$ for $n \to \infty$. Moreover Proposition 3.2.10 yields the $\mathbb{P}$-a.s. convergence of $\widehat{N}_n(\tau)/(\lambda_n \tau) - 1$ to zero. As $\widehat{m}_{n,\tau}^{(2)}$ converges to $m^{(2)}$ $\mathbb{P}$-a.s. by Theorem 2.5.5 of [25], we conclude that $\widehat{m}_{n,\tau}^{(2)}$ is also $\mathbb{P}$-a.s. bounded. Hence, we conclude that

$$\frac{\widehat{\sigma}_{n,\tau}}{\lambda_n T} - \frac{\sigma_n}{\lambda_n T} = o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{3.35}$$

For the second summand on the right-hand side of (3.33) we observe that

$$\frac{\widehat{\sigma}_{n,\tau}}{\lambda_n T} = \frac{(\widehat{\lambda}_{n,\tau} T \widehat{m}_{n,\tau}^{(2)})^{1/2}}{\lambda_n T} = \frac{1}{(\lambda_n T)^{1/2}} \sqrt{\frac{\widehat{N}_n(\tau)}{\lambda_n \tau}} \sqrt{\widehat{m}_{n,\tau}^{(2)}} = \mathcal{O}(n^{-1/2}) \tag{3.36}$$

holds $\mathbb{P}$-a.s. Indeed, part (b) of Assumption 3.2.1 yields that $(n/(\lambda_n T))^{1/2}$ converges to a positive constant again. Moreover we observe that $(\widehat{N}_n(\tau)/\lambda_n \tau)^{1/2}$ converges to 1 $\mathbb{P}$-a.s. by Proposition 3.2.10. As $\widehat{m}_{n,\tau}^{(2)}$ converges to $m^{(2)}$ $\mathbb{P}$-a.s., we conclude that $\widehat{m}_{n,\tau}^{(2)}$ is also $\mathbb{P}$-a.s. bounded. Together with Proposition 3.2.10, that is the fact that $(\widehat{N}_n(\tau)/(\lambda_n \tau) - 1)$ converges to zero $\mathbb{P}$-a.s., this yields

$$\Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau} T} - 1\Big) \frac{\widehat{\sigma}_{n,\tau}}{\lambda_n T} = o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}),$$

which completes the proof of part (i).

(ii) Let $\widetilde{S}_n$ be a $\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2}$-distributed random variable and set

$$M_n(\cdot) := \frac{\widetilde{S}_n(\cdot) - \mathfrak{m}_n}{\sigma_n}.$$

Then we observe that M_n has the standard normal distribution. Moreover, let R_n be a $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$-distributed random variable and set

$$Z_n(\cdot) := \frac{R_n(\cdot) - \mathfrak{m}_n}{\sigma_n}.$$

Let ν_n denote the distribution of Z_n. Then we observe that

$$\mathrm{law}\{\sigma_n Z_n + \mathfrak{m}_n\} = \mu^{*\mathrm{Poiss}_{\lambda_n T}}.$$

Hence, we can use part (c) of Assumption 3.2.1 on the positive homogeneity and cash-invariance of ρ to obtain

$$\begin{aligned} \frac{1}{\lambda_n T} \mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2}) - \frac{1}{\lambda_n T} \mathcal{R}_\rho(\mu^{\mathrm{Poiss}_{\lambda_n T}}) &= \frac{1}{\lambda_n T} (\rho(\sigma_n M_n + \mathfrak{m}_n) - \rho(\sigma_n Z_n + \mathfrak{m}_n)) \\ &= \frac{\sigma_n}{\lambda_n T} (\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_n)). \end{aligned} \tag{3.37}$$

An application of Wald's equation now yields

$$\frac{\sigma_n}{\lambda_n T} = \frac{(\lambda_n T m^{(2)})^{1/2}}{\lambda_n T} = \Big(\frac{m^{(2)}}{\lambda_n T}\Big)^{1/2} = \frac{1}{\sqrt{n}}\Big(\frac{n}{\lambda_n T}\, m^{(2)}\Big)^{1/2}. \tag{3.38}$$

By part (b) of Assumption 3.2.1, we conclude that

$$\frac{\sigma_n}{\lambda_n T} = \mathcal{O}(n^{-1/2}).$$

Now Döbler's Berry-Esséen inequality of Theorem 3.2.11 ensures that there exists some constant $K > 0$, such that $d_{\mathrm{Wass}}(\mathcal{N}_{0,1}, \nu_n) \leq K\,(\lambda_n T)^{-1/2}$ for all $n \in \mathbb{N}$. Now we keep in mind that λ_n tends to infinity as $n \to \infty$. Along with (3.37) and part (d) of Assumption 3.2.1 on the $(d_{\mathrm{Wass}}, |\cdot|)$-continuity of $\mathcal{R}_\rho$ at $\mathcal{N}_{0,1}$, this ensures that we have

$$\frac{\sigma_n}{\lambda_n T}\,(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_n)) = o(n^{-1/2})$$

for every $n \in \mathbb{N}$. This completes the proof of part (ii).

(iii) The assertion follows from parts (i) and (ii).

(v) To prove the assertion we will show that

$$\mathbb{P} \circ \Big\{\sqrt{\lambda_n T}\,\Big(\frac{1}{\widehat{\lambda}_{n,\tau} T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})\Big)\Big\}^{-1} \xrightarrow{\mathrm{w}} \mathcal{N}_{0,s^2T/\tau}. \tag{3.39}$$

The claim will then follow by an application of Slutskys Lemma. To this end, for every $\omega \in \Omega$, let $\widehat{S}^\omega_{n,\tau}$ be a $\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}$-distributed random variable on some probability space $(\Omega^\omega, \mathcal{F}^\omega, \mathbb{P}^\omega)$ and set

$$\widehat{M}^\omega_{n,\tau}(\cdot) := \frac{\widehat{S}^\omega_{n,\tau}(\cdot) - \widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\sigma}_{n,\tau}(\omega)}.$$

Then $\widehat{M}^\omega_{n,\tau}$ has the standard normal distribution for every ω. Furthermore let R_n be a random variable, being distributed according to $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$ and set

$$Z_n(\cdot) := \frac{R_n(\cdot) - \mathfrak{m}_n}{\sigma_n}.$$

Moreover let ν_n denote distribution of Z_n. We observe that

$$\mathrm{law}\{\sigma_n\, Z_n + \mathfrak{m}_n\} = \mu^{*\mathrm{Poiss}_{\lambda_n T}}.$$

Using part (c) of Assumption 3.2.1 on the positive homogeneity and the cash-invariance of ρ, we obtain for every ω

$$\begin{aligned}&\sqrt{\lambda_n T}\,\Big(\frac{1}{\widehat{\lambda}_{n,\tau}(\omega) T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})\Big)\\ &= \sqrt{\lambda_n T}\,\Big(\frac{1}{\widehat{\lambda}_{n,\tau}(\omega) T}\,\rho(\widehat{\sigma}_{n,\tau}(\omega)\,\widehat{M}^\omega_{n,\tau} + \widehat{\mathfrak{m}}_{n,\tau}(\omega)) - \frac{1}{\lambda_n T}\,\rho(\sigma_n\, Z_n + \mathfrak{m}_n)\Big)\end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\lambda_n T}\,\Big(\frac{\widehat{\sigma}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}(\omega)T}\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \frac{\sigma_n}{\lambda_n T}\,\mathcal{R}_\rho(\nu_n) + \frac{\widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}(\omega)T} - \frac{\mathfrak{m}_n}{\lambda_n T}\Big)\\
&= \frac{1}{\sqrt{\lambda_n T}}\Big(\sigma_n\,(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_n)) \;+\; \Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}(\omega)T}\,\widehat{\sigma}_{n,\tau}(\omega) - \sigma_n\Big)\,\mathcal{R}_\rho(\mathcal{N}_{0,1})\Big)\\
&\quad + \sqrt{\lambda_n T}\,\Big(\widehat{m}_{n,\tau}(\omega) - m\Big)\\
&= \frac{\sigma_n}{\sqrt{\lambda_n T}}\,(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_n)) \;+\; \frac{1}{\sqrt{\lambda_n T}}\,\Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}(\omega)T} - 1\Big)\,\widehat{\sigma}_{n,\tau}(\omega)\,\mathcal{R}_\rho(\mathcal{N}_{0,1})\\
&\quad + \frac{1}{\sqrt{\lambda_n T}}\,(\widehat{\sigma}_{n,\tau}(\omega) - \sigma_n)\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) \;+\; \sqrt{\lambda_n T}\,\Big(\widehat{m}_{n,\tau}(\omega) - m\Big)\\
&=: S_1(n) \;+\; S_2(n,\omega) \;+\; S_3(n,\omega) \;+\; S_4(n,\omega). \qquad (3.40)
\end{aligned}
$$

In Steps 1–4 below, we will show that

$$
\begin{aligned}
S_1(n) &= o(1) && (3.41)\\
S_2(n,\cdot\,) &= o_{\mathbb{P}\text{-a.s.}}(1) && (3.42)\\
S_3(n,\cdot\,) &= o_{\mathbb{P}\text{-a.s.}}(1) && (3.43)\\
\mathbb{P}\circ S_4(n,\cdot\,)^{-1} &\stackrel{\mathrm{w}}{\longrightarrow} \mathcal{N}_{0,s^2T/\tau}, \qquad n\to\infty. && (3.44)
\end{aligned}
$$

Step 1: This assertion has already been proven in part (ii).

Step 2: We observe that

$$
\begin{aligned}
\frac{1}{\sqrt{\lambda_n T}}\,\Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T} - 1\Big)\,\widehat{\sigma}_{n,\tau} &= \frac{1}{\sqrt{\lambda_n T}}\,\Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T} - 1\Big)\,(\widehat{\sigma}_{n,\tau} - \sigma_n) \;+\; \frac{1}{\sqrt{\lambda_n T}}\,\Big(\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T} - 1\Big)\,\sigma_n\\
&= \frac{1}{\sqrt{\lambda_n T}}\Big(\frac{\lambda_n \tau}{\widehat{N}_n(\tau)} - 1\Big)(\widehat{\sigma}_{n,\tau} - \sigma_n) + \frac{1}{\sqrt{\lambda_n T}}\,\Big(\frac{\lambda_n \tau}{\widehat{N}_n(\tau)} - 1\Big)\,\sigma_n, \qquad (3.45)
\end{aligned}
$$

for every $n\in\mathbb{N}$. By Proposition 3.2.10, we conclude that $\frac{\lambda_n\tau}{\widehat{N}_n(\tau)} - 1$ converges to zero $\mathbb{P}$-a.s. Moreover using Wald's equation again, we have

$$
\begin{aligned}
\widehat{\sigma}^2_{n,\tau} - \sigma^2_n &= \widehat{\lambda}_{n,\tau}T\widehat{m}^{(2)}_{n,\tau} - \lambda_n T m^{(2)}\\
&= (\widehat{\lambda}_{n,\tau}T - \lambda_n T)\,\widehat{m}^{(2)}_{n,\tau} + \lambda_n T\,(\widehat{m}^{(2)}_{n,\tau} - m^{(2)})\\
&= \lambda_n T\,\Big(\frac{\widehat{N}_n(\tau)}{\lambda_n\tau} - 1\Big) \;+\; \lambda_n T\,(\widehat{m}^{(2)}_{n,\tau} - m^{(2)}). \qquad (3.46)
\end{aligned}
$$

By Proposition 3.2.10 we conclude that $\widehat{N}_n(\tau)/(\lambda_n\tau) - 1$ converges to zero $\mathbb{P}$-a.s. Furthermore, by Theorem 2.5.5 in [25], we observe that $\widehat{m}^{(2)}_{n,\tau} - m^{(2)}$ converges to zero $\mathbb{P}$-a.s., where it is important to note that the integrability conditions are trivially satisfied due to part (a) of Assumption 3.2.1. Finally by part (b) of Assumption 3.2.1, we conclude that

$$
\widehat{\sigma}^2_{n,\tau} - \sigma^2_n \;=\; o_{\mathbb{P}\text{-a.s.}}(n). \qquad (3.47)
$$

Using (3.47), we observe that

$$\frac{1}{\sqrt{\lambda_n T}}(\widehat{\sigma}_{n,\tau} - \sigma_n) = o_{\mathbb{P}\text{-a.s.}}(1). \tag{3.48}$$

Hence, the first summand on the right-hand side of (3.45) converges to zero $\mathbb{P}$-a.s. To conclude on the convergence of the second summand in (3.45) we observe that

$$\frac{\sigma_n}{\sqrt{\lambda_n T}} = \Big(\frac{\lambda_n T m^{(2)}}{\lambda_n T}\Big)^{1/2} = \sqrt{m^{(2)}}. \tag{3.49}$$

This is a direct consequence of Wald's equation. Thus, the second summand on the right-hand side of (3.45) converges $\mathbb{P}$-a.s. to zero by Proposition 3.2.10. This proves (3.42).

Step 3: This assertion has already been proven in the second step.

Step 4: We observe that

$$\begin{aligned}
\sqrt{\lambda_n T}\,(\widehat{m}_{n,\tau} - m) &= \sqrt{\lambda_n T}\,\Big(\frac{1}{\widehat{N}_n(\tau)}\sum_{i=1}^{\widehat{N}_n(\tau)} Y_i - m\Big) \\
&= \Big(\frac{\lambda_n \tau}{\widehat{N}_n(\tau)} - 1\Big)\sqrt{\frac{T}{\tau}}\,(\lambda_n\tau)^{-1/2}\Big(\sum_{i=1}^{\widehat{N}_n(\tau)} Y_i - \widehat{N}_n(\tau)\,m\Big) \\
&\quad + \sqrt{\frac{T}{\tau}}\,(\lambda_n\tau)^{-1/2}\Big(\sum_{i=1}^{\widehat{N}_n(\tau)} Y_i - \widehat{N}_n(\tau)\,m\Big) \\
&=: S_{4,1}(n) + S_{4,2}(n).
\end{aligned} \tag{3.50}$$

Now part (a) of Theorem 2.5.15 in [25] yields that

$$\mathbb{P}\circ\Big\{(\lambda_n\tau)^{-1/2}\Big(\sum_{i=1}^{\widehat{N}_n(\tau)} Y_i - \widehat{N}_n(\tau)\,m\Big)\Big\}^{-1} \stackrel{\mathrm{w}}{\longrightarrow} \mathcal{N}_{0,s^2}, \quad n\to\infty,$$

such that

$$\mathbb{P}\circ S_{4,2}(n)^{-1} \stackrel{\mathrm{w}}{\longrightarrow} \mathcal{N}_{0,s^2T/\tau}, \quad n\to\infty$$

follows by Slutsky's Lemma. Moreover, by another application of Proposition 3.2.10, we can conclude that $S_{4,1}(n)$ converges to zero in probability as $n\to\infty$. Hence, the assertion follows again by Slutsky's Lemma, which completes the proof of part (v).

(iv) The assertion can be proven in the same way as part (v). Following the same line of reasoning as in (3.40), we observe that $S_1(n)$–$S_3(n)$ converge to zero $\mathbb{P}$-a.s. The claim now follows by an application of the Marcinkiewicz-Zygmund SLLN for random sums of Theorem 2.5.5 in [25]. □

Proof of Theorem 3.2.3 (i): Let $R^{\omega}_{n,\tau}$ be a random variable being distributed according to $\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}(\omega;\cdot)$, with

$$\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}(\omega;\,\cdot\,) := \big(\widehat{\mu}_{n,\tau}(\omega)\big)^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}(\omega)T}}[\,\cdot\,].$$

Here, for every $\omega \in \Omega$, $R^{\omega}_{n,\tau}$ is regarded as a random variable on a probability space $(\Omega^{\omega}, \mathcal{F}^{\omega}, \mathbb{P}^{\omega})$. Now set

$$Z^{\omega}_{n,\tau}(\cdot) := \frac{R^{\omega}_{n,\tau}(\cdot) \ - \ \widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\sigma}_{n,\tau}(\omega)},$$

with $\widehat{\mathfrak{m}}_{n,\tau}(\omega)$ and $\widehat{\sigma}^2_{n,\tau}(\omega)$ referring to the mean and the variance of $\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}(\omega;\cdot)$, respectively. Then we observe that

$$\mathrm{law}\{\widehat{\sigma}_{n,\tau}(\omega)\, Z^{\omega}_{n,\tau} + \widehat{\mathfrak{m}}_{n,\tau}(\omega)\} \ = \ \widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}(\omega;\,\cdot\,) \tag{3.51}$$

for every $n \in \mathbb{N}$ and $\omega \in \Omega$. Write $\nu_{n,\tau}(\omega;\cdot)$ for the distribution of $Z^{\omega}_{n,\tau}$. Let $S^{\omega}_{n,\tau}$ be a random variable distributed according to the normal distribution with mean $\widehat{\mathfrak{m}}_{n,\tau}(\omega)$ and variance $\widehat{\sigma}^2_{n,\tau}(\omega)$, and note that

$$M^{\omega}_{n,\tau}(\cdot) := \frac{S^{\omega}_{n,\tau}(\cdot) - \widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\sigma}_{n,\tau}(\omega)}$$

has the standard normal distribution. Due to part (c) of Assumption 3.2.1, this yields

$$\begin{aligned}
&\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}) - \mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}(\omega;\cdot))\\
&\quad = \ \rho(\widehat{\sigma}_{n,\tau}(\omega)\, M^{\omega}_{n,\tau} + \widehat{\mathfrak{m}}_{n,\tau}(\omega)) - \rho(\widehat{\sigma}_{n,\tau}(\omega)\, Z^{\omega}_{n,\tau} + \widehat{\mathfrak{m}}_{n,\tau}(\omega))\\
&\quad = \ \widehat{\sigma}_{n,\tau}(\omega)\,(\rho(M^{\omega}_{n,\tau}) - \rho(Z^{\omega}_{n,\tau}))\\
&\quad = \ \widehat{\sigma}_{n,\tau}(\omega)\,(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_{n,\tau}(\omega;\cdot)))
\end{aligned} \tag{3.52}$$

for every $n \in \mathbb{N}$ and $\omega \in \Omega$. Following the same line of reasoning as in the proof of part (ii) of Theorem 3.2.2, we intend to show that $\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_{n,\tau}(\omega;\cdot))$ converges to 0 for $\mathbb{P}$-a.e. ω by using part (d) of Assumption 3.2.1. To this end, we have to show, that the constant in the Berry-Esséen inequality of Theorem 3.2.11 is $\mathbb{P}$-a.s. bounded above. In particular, Theorem 3.2.11 yields

$$d_{\mathrm{Wass}}(\mathcal{N}_{0,1}, \nu_{n,\tau}(\omega;\cdot)) \leq \frac{1}{\sqrt{\widehat{\lambda}_{n,\tau}(\omega)T}}\Big(\frac{2\,\widehat{s}^2_{n,\tau}(\omega)}{\widehat{m}^{(2)}_{n,\tau}(\omega)} + \frac{12\,\widehat{m}^{(3)}_{n,\tau}(\omega)}{(\widehat{m}^{(2)}_{n,\tau}(\omega))^{3/2}} + \frac{|\widehat{m}_{n,\tau}(\omega)|}{(\widehat{m}^{(2)}_{n,\tau}(\omega))^{1/2}}\Big), \tag{3.53}$$

for every $n \in \mathbb{N}$ and $\mathbb{P}$-a.e. $\omega \in \Omega$. Here $\widehat{m}_{n,\tau}(\omega)$, $\widehat{s}^2_{n,\tau}(\omega)$, $\widehat{m}^{(2)}_{n,\tau}(\omega)$ and $\widehat{m}^{(3)}_{n,\tau}(\omega)$ denote the expected value, the variance, the second moment and the third moment of $\widehat{\mu}_{n,\tau}(\omega;\cdot)$, respectively. We will now show, that the second fraction in the bracket on the right-hand side of (3.53) converges to a constant for $\mathbb{P}$-a.e. ω. The convergence of the remaining summands can be proven in the same way. By Theorem 2.5.5 in [25] we conclude that

$$\begin{aligned}
\widehat{m}^{(3)}_{n,\tau}(\omega) - m^{(3)} \ &= \ \Big(\frac{1}{\widehat{N}_n(\tau;\omega)}\sum_{i=1}^{\widehat{N}_n(\tau;\omega)} Y_i(\omega)^3 - m^{(3)}\Big)\\
&= \ (\widehat{N}_n(\tau;\omega))^{-1}\Big(\sum_{i=1}^{\widehat{N}_n(\tau;\omega)} Y_i(\omega)^3 - \widehat{N}_n(\tau;\omega)\, m^{(3)}\Big)
\end{aligned} \tag{3.54}$$

converges to zero for $\mathbb{P}$-a.e. ω. Here part (a) in Assumption 3.2.1 about the existence of the third moments of Y_1 ensures the applicability of the theorem. Following the same line of reasoning, we conclude that $\widehat{m}^{(2)}_{n,\tau}(\omega) - m^{(2)}$ converges to zero for $\mathbb{P}$-a.e. ω. Thus, an application of Slutskys Lemma now yields the convergence of $\widehat{m}^{(3)}_{n,\tau}(\omega)/(\widehat{m}^{(2)}_{n,\tau})^{3/2}$ to $m^{(3)}/(m^{(2)})^{3/2}$ for $\mathbb{P}$-a.e. ω. Using the same arguments as above, we conclude that the bracket on the right-hand side of (3.53) converges to $2\,s^2/m^{(2)} + 12\,m^{(3)}/(m^{(2)})^{3/2} + |m|/(m^{(2)})^{1/2}$ for $\mathbb{P}$-a.e. ω. Moreover by Proposition 3.2.10 along with part (b) of Assumption 3.2.1, we observe that

$$\frac{1}{\sqrt{\widehat{\lambda}_{n,\tau}(\omega)T}} = \frac{1}{\sqrt{n}}\Big(\frac{n}{\lambda_n T}\Big)^{1/2}\Big(\frac{\lambda_n \tau}{N_n(\tau;\omega)}\Big)^{1/2} = \mathcal{O}(n^{-1/2}) \tag{3.55}$$

for $\mathbb{P}$-a.e. ω. By part (d) of Assumption 3.2.1 on the $(d_{\text{Wass}}, |\cdot|)$-continuity of $\mathcal{R}_\rho$ at $\mathcal{N}_{0,1}$, we can therefore conclude that $\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_{n,\tau}(\omega;\cdot))$ converges to 0 for $\mathbb{P}$-a.e. ω. Hence, the claim would follow by showing that

$$\frac{\widehat{\sigma}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}(\omega)T} = \mathcal{O}(n^{-1/2}) \tag{3.56}$$

for $\mathbb{P}$-a.e. ω. An application of Wald's formula yields

$$\frac{\widehat{\sigma}_{n,\tau}(\omega)}{\widehat{\lambda}_{n,\tau}T} = \frac{(\widehat{\lambda}_{n,\tau}T\,\widehat{m}^{(2)}_{n,\tau}(\omega))^{1/2}}{\widehat{\lambda}_{n,\tau}T} = \frac{1}{\sqrt{n}}\Big(\frac{\lambda_n\tau}{\widehat{N}_n(\tau;\omega)}\,\widehat{m}^{(2)}_{n,\tau}(\omega)\Big)^{1/2}\Big(\frac{n}{\lambda_n T}\Big)^{1/2}, \tag{3.57}$$

where $(n/(\lambda_n T))^{1/2}$ converges to $1/\sqrt{cT}$ by part (b) of Assumption 3.2.1. Proposition 3.2.10 yields the convergence of $\lambda_n\tau/\widehat{N}_n(\tau;\omega)$ to 1 for $\mathbb{P}$-a.e. ω. Moreover, we have already shown the convergence of $\widehat{m}^{(2)}_{n,\tau}(\omega)$ to $m^{(2)}$ for $\mathbb{P}$-a.e. ω. Thus, we observe that (3.56) holds true. This completes the proof of part (i).

(ii) The assertion follows from part (i), as well as part (i)–(ii) of Theorem 3.2.2.

(iii) and (iv): The assertions can be proven in the same way as the assertions (iv) and (v) in Theorem 3.2.2. □

3.2.2 Proof of Corollaries 3.2.4 and 3.2.6

For the proof of part (v) of Corollary 3.2.4 we will need the following Proposition about the joint asymptotic normality of the empirical mean in the collective model together with the estimator for the estimator for the parameter in the Poisson distribution.

Proposition 3.2.12 *Let (Y_i) be a sequence of i.i.d. random variables with finite variance $s^2 > 0$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and (N_{λ_n}) be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, being independent of (Y_i), such that for every $n \in \mathbb{N}$ N_{λ_n} is a* Poiss$_{\lambda_n}$*-distributed random variable. Furthermore assume that $\lambda_n \to \infty$. Then*

$$\sqrt{\lambda_n}\left(\begin{bmatrix} \frac{1}{\lambda_n}\sum_{j=1}^{N_{\lambda_n}} Y_j \\ \frac{N_{\lambda_n}}{\lambda_n}\end{bmatrix} - \begin{bmatrix} m \\ 1 \end{bmatrix}\right) \xrightarrow{\text{d}} Z\,, \tag{3.58}$$

as $n \to \infty$, where Z refers to some bivariate normally distributed random variable with mean $[0,0]'$ and covariance matrix $\Sigma := \begin{bmatrix} s^2 & 0 \\ 0 & 1 \end{bmatrix}$.

Proof The claim in (3.58) would follow, if we could prove the pointwise convergence of the characteristic function of the left-hand side in (3.58) to the characteristic function of Z. Thus, it suffices to show that

$$\lim_{n\to\infty} \mathbb{E}\Big[e^{i\sqrt{\lambda_n}\left(\alpha_1 \frac{1}{\lambda_n}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)+\alpha_2\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)\right)} \Big] = e^{-\frac{1}{2}(\alpha_1^2 s^2+\alpha_2^2)}, \tag{3.59}$$

holds for all $\alpha_1, \alpha_2 \in \mathbb{R}$. Now we can use the tower-property of the conditional expectation to derive

$$\begin{aligned}
&\mathbb{E}\Big[e^{i\sqrt{\lambda_n}\left(\alpha_1 \frac{1}{\lambda_n}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)+\alpha_2\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)\right)} \Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)+i\alpha_2\sqrt{\lambda_n}\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)} \Big| N_{\lambda_n}\Big]\Big] \\
&= \mathbb{E}\Big[\mathbb{E}\Big[e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)} \Big| N_{\lambda_n}\Big] e^{i\alpha_2\sqrt{\lambda_n}\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)}\Big] \\
&= \mathbb{E}\Big[\Big(\mathbb{E}\Big[e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)} \Big| N_{\lambda_n}\Big] - e^{-\frac{1}{2}\alpha_1^2 s^2}\Big) e^{i\alpha_2\sqrt{\lambda_n}\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)}\Big] \\
&\quad + e^{-\frac{1}{2}\alpha_1^2 s^2}\, \mathbb{E}\Big[e^{i\alpha_2\sqrt{\lambda_n}\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)}\Big] \\
&=: S_1(n;\alpha_1,\alpha_2) + S_2(n;\alpha_1,\alpha_2).
\end{aligned} \tag{3.60}$$

In the following we will show that

$$\lim_{n\to\infty} S_1(n;\alpha_1,\alpha_2) = 0 \tag{3.61}$$

$$\lim_{n\to\infty} S_2(n;\alpha_1,\alpha_2) = e^{-\frac{1}{2}(\alpha_1^2 s^2+\alpha_2^2)} \tag{3.62}$$

hold for all $\alpha_1, \alpha_2 \in \mathbb{R}$, which would then yield the claim.

Step 1: We will first show that (3.61) holds true. To this use, we observe that

$$\begin{aligned}
\mathbb{E}\Big[e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)} \Big| N_{\lambda_n}\Big] &= \mathbb{E}\Big[\prod_{j=1}^{N_{\lambda_n}} e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}(Y_j-m)} \Big| N_{\lambda_n}\Big] \\
&= \sum_{k=0}^{\infty} \mathbb{E}\Big[\prod_{j=1}^{N_{\lambda_n}} e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}(Y_j-m)} \Big| \{N_{\lambda_n}=k\}\Big] \mathbb{1}_{\{N_{\lambda_n}=k\}} \\
&= \sum_{k=0}^{\infty} \mathbb{E}\Big[\prod_{j=1}^{k} e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}(Y_j-m)} \Big| \{N_{\lambda_n}=k\}\Big] \mathbb{1}_{\{N_{\lambda_n}=k\}} \\
&= \sum_{k=0}^{\infty} \mathbb{E}\Big[\prod_{j=1}^{k} e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}(Y_j-m)} \Big] \mathbb{1}_{\{N_{\lambda_n}=k\}} \\
&= \sum_{k=0}^{\infty} \mathbb{E}\Big[e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}(Y_1-m)} \Big]^k \mathbb{1}_{\{N_{\lambda_n}=k\}} \\
&= \mathbb{E}\Big[e^{i\alpha_1 \frac{1}{\sqrt{\lambda_n}}(Y_1-m)} \Big]^{N_{\lambda_n}},
\end{aligned} \tag{3.63}$$

where for the fourth "=" we used the fact that the sequence (Y_i) is independent of N_{λ_n}. This was used to transform the expectation conditional on N_{λ_n} into a nonconditional one. For the fifth "=" we then used that (Y_i) is a sequence of i.i.d. random variables. To prove (3.61) we use the representation on the right-hand side of (3.63) to obtain

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\Big(\mathbb{E}\Big[e^{i\alpha_1\frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)}\,\Big|\,N_{\lambda_n}\Big]-e^{-\frac{1}{2}\alpha_1^2 s^2}\Big)\,e^{i\sqrt{\lambda_n}\alpha_2\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)}\Big]\Big|\\
\le\;&\mathbb{E}\Big[\Big|\mathbb{E}\Big[e^{i\alpha_1\frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)}\,\Big|\,N_{\lambda_n}\Big]-e^{-\frac{1}{2}\alpha_1^2 s^2}\Big|\cdot\Big|e^{i\sqrt{\lambda_n}\alpha_2\left(\frac{N_{\lambda_n}}{\lambda_n}-1\right)}\Big|\Big]\\
\le\;&\mathbb{E}\Big[\Big|\mathbb{E}\Big[e^{i\alpha_1\frac{1}{\sqrt{\lambda_n}}\sum_{j=1}^{N_{\lambda_n}}(Y_j-m)}\,\Big|\,N_{\lambda_n}\Big]-e^{-\frac{1}{2}\alpha_1^2 s^2}\Big|\Big]\\
=\;&\mathbb{E}\Big[\Big|\mathbb{E}\Big[\,e^{i\alpha_1\frac{Y_1-m}{\sqrt{\lambda_n}}}\,\Big]^{N_{\lambda_n}}-e^{-\frac{1}{2}\alpha_1^2 s^2}\Big|\Big].
\end{aligned}\tag{3.64}$$

Using a Taylor expansion (in α_1), yields

$$\begin{aligned}
\mathbb{E}\Big[\,e^{i\alpha_1\frac{Y_1-m}{\sqrt{\lambda_n}}}\,\Big]&=1+i\alpha_1\frac{1}{\sqrt{\lambda_n}}\,\mathbb{E}[Y_1-m]-\frac{1}{2}\,\alpha_1^2\,\frac{1}{\lambda_n}\,\mathbb{E}[(Y_1-m)^2]+o\Big(\frac{1}{\lambda_n}\Big)\\
&=1-\frac{1}{2}\,\alpha_1^2\,s^2\,\frac{1}{\lambda_n}+o\Big(\frac{1}{\lambda_n}\Big),
\end{aligned}\tag{3.65}$$

where we used the facts that $\mathbb{E}[Y_1-m]=0$ and $\mathbb{E}[(Y_1-m)^2]=s^2$ for the last step. Note that it is sufficient to consider the remainder $o\,(1/\lambda_n)$ in the above Taylor expansion, rather than "$o\,(-\frac{1}{2}\,\alpha_1^2 s^2/\lambda_n)$", because s^2 was supposed to be constant. Moreover, we might regard α_1 as a constant, too, because we only aim to prove the pointwise convergence of the characteristic functions. Now for every $n\in\mathbb{N}$ let $\gamma_n:\mathbb{R}\to\mathbb{R}$ be defined by

$$\gamma_n(\alpha_1):=-\frac{1}{2}\,\alpha_1^2\,s^2\,\frac{1}{\lambda_n}+o\Big(\frac{1}{\lambda_n}\Big).\tag{3.66}$$

Then the sequence

$$\begin{aligned}
N_{\lambda_n}\,\gamma_n(\alpha_1)&=-\frac{1}{2}\,\alpha_1^2\,s^2\,\frac{N_{\lambda_n}}{\lambda_n}+N_{\lambda_n}\,o\Big(\frac{1}{\lambda_n}\Big)\\
&=-\frac{1}{2}\,\alpha_1^2\,s^2\,\frac{N_{\lambda_n}}{\lambda_n}+\frac{N_{\lambda_n}}{\lambda_n}\,\frac{o\left(\frac{1}{\lambda_n}\right)}{1/\lambda_n}
\end{aligned}\tag{3.67}$$

converges $\mathbb{P}$-a.s. to $-\frac{1}{2}\alpha_1^2 s^2$ for every $\alpha_1\in\mathbb{R}$ as $n\to\infty$, because N_{λ_n}/λ_n converges to 1 $\mathbb{P}$-a.s. by Proposition 3.2.10 and $\lambda_n o\left(\frac{1}{\lambda_n}\right)$ converges to zero. Hence (3.65)–(3.67) yield

$$\lim_{n\to\infty}\mathbb{E}\Big[\,e^{i\alpha_1\frac{Y_1-m}{\sqrt{\lambda_n}}}\Big]^{N_{\lambda_n}}=\lim_{n\to\infty}\,(1+\gamma_n(\alpha_1))^{N_{\lambda_n}}=e^{-\frac{1}{2}\alpha_1^2 s^2},\qquad\mathbb{P}\text{-a.s.}\tag{3.68}$$

holds for every $\alpha_1\in\mathbb{R}$. Now dominated convergence applied to the right-hand side in (3.64) (because the integrand is obviously bounded), along with (3.68) lead to the assertion in (3.61).

Step 2: For the convergence in (3.62) we observe that

$$\mathbb{P} \circ \Big\{ \sqrt{\lambda_n} \Big(\frac{N_{\lambda_n}}{\lambda_n} - 1 \Big) \Big\}^{-1} \quad \xrightarrow{\mathrm{w}} \quad \mathcal{N}_{0,1}. \tag{3.69}$$

This is a direct consequence of Theorem 2.5.13 in [25], which leads to the assertion in (3.62) and completes the proof. □

We are now in a position to prove Corollaries 3.2.4 and 3.2.6.

Proof of Corollary 3.2.4 (i) First we observe that

$$\begin{aligned} \frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) \;-\; \frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma^2_n}) \;&=\; \frac{\lambda_n T}{n}\Big(\frac{1}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma^2_n})\Big) \\ &\quad + \frac{\lambda_n T}{n}\Big(\frac{1}{\lambda_n T} - \frac{1}{\widehat{\lambda}_{n,\tau}T}\Big)\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) \\ &=: S_1(n) \;+\; S_2(n). \end{aligned} \tag{3.70}$$

Using part (i) of Theorem 3.2.2 on the first summand, along with the fact that λ_n/n converges to some constant $c > 0$, we arrive at

$$\begin{aligned} S_1(n) &= \frac{\lambda_n T}{n}\Big(\widehat{m}_{n,\tau} - m \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2})\Big) \\ &= \frac{1}{n}\Big(\widehat{\lambda}_{n,\tau}T\widehat{m}_{n,\tau} - \lambda_n T m\Big) \;+\; \frac{\lambda_n T}{n}\Big(1 - \frac{\widehat{\lambda}_{n,\tau}T}{\lambda_n T}\Big)\,\widehat{m}_{n,\tau} \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}) \\ &= \frac{1}{n}\,(\widehat{\mathfrak{m}}_{n,\tau} - \mathfrak{m}_n) \;+\; \frac{\lambda_n T}{n}\Big(1 - \frac{\widehat{\lambda}_{n,\tau}T}{\lambda_n T}\Big)\,\widehat{m}_{n,\tau} \;+\; o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \end{aligned} \tag{3.71}$$

For the first summand, we will leave the representation in (3.71) as it is. Our investigations for the $S_2(n)$ will show that the second summand on the right-hand side of (3.71) will cancel itself with an expression in (3.72). For $S_2(n)$ we use the representation in (3.15) to derive

$$\begin{aligned} S_2(n) &= \frac{\lambda_n T}{n}\Big(\frac{\widehat{\lambda}_{n,\tau}T}{\lambda_n T} - 1\Big)\Big(\frac{\widehat{\sigma}_{n,\tau}}{\widehat{\lambda}_{n,\tau}T}\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) + \frac{\widehat{\mathfrak{m}}_{n,\tau}}{\widehat{\lambda}_{n,\tau}T}\Big) \\ &= \frac{\sqrt{\lambda_n T}}{n}\sqrt{\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T}}\Big(\frac{\widehat{N}_n(\tau)}{\lambda_n \tau} - 1\Big)\sqrt{\widehat{m}^{(2)}_{n,\tau}}\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \frac{\lambda_n T}{n}\Big(1 - \frac{\widehat{\lambda}_{n,\tau}T}{\lambda_n T}\Big)\,\widehat{m}_{n,\tau}. \end{aligned} \tag{3.72}$$

Now the claim would follow by showing that

$$\frac{\sqrt{\lambda_n T}}{n}\sqrt{\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T}}\Big(\frac{\widehat{N}_n(\tau)}{\lambda_n \tau} - 1\Big)\sqrt{\widehat{m}^{(2)}_{n,\tau}}\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) = o_{\mathbb{P}\text{-a.s.}}(n^{-1/2}). \tag{3.73}$$

By part (b) of Assumption 3.2.1, we conclude that $\sqrt{\lambda_n T}/n = \mathcal{O}(n^{-1/2})$. Furthermore, Proposition 3.2.10 yields

$$\sqrt{\frac{\lambda_n T}{\widehat{\lambda}_{n,\tau}T}} = \sqrt{\frac{\lambda_n \tau}{\widehat{N}_n(\tau)}} = \mathcal{O}_{\mathbb{P}\text{-a.s.}}(1),$$

as well as

$$\Big(\frac{\widehat{N}_n(\tau)}{\lambda_n \tau} - 1\Big) = o_{\mathbb{P}\text{-a.s.}}(1).$$

Finally, Theorem 2.5.5 in [25] yields the $\mathbb{P}$-a.s. convergence of $\widehat{m}^{(2)}_{n,\tau}$ to $m^{(2)}$, which leads to the assertion in (3.73) and completes the proof.

(ii) First, we observe that

$$\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2}) - \frac{1}{n}\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}}) = \frac{\lambda_n T}{n}\,\Big(\frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mathcal{N}_{\mathfrak{m}_n,\sigma_n^2})\; - \;\frac{1}{\lambda_n T}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})\Big). \quad (3.74)$$

Now the claim follows by part (ii) of Theorem 3.2.2 along with the fact that λ_n/n converges to some constant $c > 0$ (part (b) of Assumption 3.2.1).

(iii) The assertion follows from (i) and (ii).

(v) To prove the assertion, we will show that

$$\mathbb{P} \circ \Big\{\sqrt{n}\Big(\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau},\widehat{\sigma}^2_{n,\tau}}) - \frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})\Big)\Big\}^{-1} \xrightarrow{\mathrm{w}} \mathcal{N}_{0,c(s^2+m^2)T^2/\tau}. \quad (3.75)$$

The claim will then follow by an application of Slutskys Lemma. To this end, for every $\omega \in \Omega$, let $\widehat{S}^\omega_{n,\tau}$ be a $\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}$-distributed random variable on some probability space $(\Omega^\omega, \mathcal{F}^\omega, \mathbb{P}^\omega)$ and set

$$\widehat{M}^\omega_{n,\tau}(\cdot) := \frac{\widehat{S}^\omega_{n,\tau}(\cdot) - \widehat{\mathfrak{m}}_{n,\tau}(\omega)}{\widehat{\sigma}_{n,\tau}(\omega)}.$$

Then $\widehat{M}^\omega_{n,\tau}$ has the standard normal distribution for every $n \in \mathbb{N}$, $\tau > 0$ and $\omega \in \Omega$. Furthermore, let R_n be a $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$-distributed random variable and set

$$Z_n(\cdot) := \frac{R_n(\cdot) - \mathfrak{m}_n}{\sigma_n},$$

where again $\mathfrak{m}_n$ and σ_n denote the mean and standard deviation of $\mu^{*\mathrm{Poiss}_{\lambda_n T}}$, respectively. Moreover, let ν_n denote the distribution of Z_n. Then we observe that

$$\mathrm{law}\{\sigma_n\, Z_n + \mathfrak{m}_n\} = \mu^{*\mathrm{Poiss}_{\lambda_n T}}.$$

Using part (c) of Assumption 3.2.1 on the positive homogeneity and cash-invariance of ρ, we obtain for every ω

$$\begin{aligned}
&\sqrt{n}\Big(\frac{1}{n}\,\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}(\omega),\widehat{\sigma}^2_{n,\tau}(\omega)}) - \frac{1}{n}\,\mathcal{R}_\rho(\mu^{*\mathrm{Poiss}_{\lambda_n T}})\Big)\\
&= \frac{1}{\sqrt{n}}\,\Big(\rho(\widehat{\sigma}_{n,\tau}(\omega)\,\widehat{M}^\omega_{n,\tau} + \widehat{\mathfrak{m}}_{n,\tau}(\omega)) - \rho(\sigma_n\, Z_n + \mathfrak{m}_n)\Big)\\
&= \frac{1}{\sqrt{n}}\,\Big(\widehat{\sigma}_{n,\tau}(\omega)\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \sigma_n\,\mathcal{R}_\rho(\nu_n)\; + \;\widehat{\mathfrak{m}}_{n,\tau}(\omega) - \mathfrak{m}_n\Big)\\
&= \frac{1}{\sqrt{n}}\,\Big(\sigma_n\,(\mathcal{R}_\rho(\mathcal{N}_{0,1}) - \mathcal{R}_\rho(\nu_n))\; + \;(\widehat{\sigma}_{n,\tau}(\omega) - \sigma_n)\,\mathcal{R}_\rho(\mathcal{N}_{0,1})\; + \;\widehat{\mathfrak{m}}_{n,\tau}(\omega) - \mathfrak{m}_n\Big). \quad (3.76)
\end{aligned}$$

In the following we will show that

$$\frac{\sigma_n}{\sqrt{n}}\left(\mathcal{R}_\rho(\mathcal{N}_{0,1})-\mathcal{R}_\rho(\nu_n)\right) \;=\; o(1), \tag{3.77}$$

$$\frac{1}{\sqrt{n}}(\widehat{\sigma}_{n,\tau}-\sigma_n)\,\mathcal{R}_\rho(\mathcal{N}_{0,1}) \;=\; o_{\mathbb{P}\text{-a.s.}}(1), \tag{3.78}$$

$$\mathbb{P}\circ\{n^{-1/2}\,(\widehat{\mathfrak{m}}_{n,\tau}-\mathfrak{m}_n)\}^{-1} \xrightarrow{\text{w}} \mathcal{N}_{0,c(s^2+m^2)T^2/\tau}, \tag{3.79}$$

such that the assertion would follow by an application of Slutsky's Lemma.

Step 1: For every $n\in\mathbb{N}$ we can use Wald's equation along with part (b) of Assumption 3.2.1 to deduce

$$\frac{\sigma_n}{\sqrt{n}}=\Big(\frac{\lambda_n T m^{(2)}}{n}\Big)^{1/2}=\mathcal{O}(1).$$

Following the same line of reasoning as in the proof of part (ii) of Theorem 3.2.2, we conclude that (3.77) holds true.

Step 2: Using part (b) of Assumption 3.2.1 again, we observe that

$$\frac{1}{\sqrt{n}}(\widehat{\sigma}_{n,\tau}-\sigma_n)=\sqrt{\frac{\lambda_n T}{n}}\,\frac{1}{\sqrt{\lambda_n T}},(\widehat{\sigma}_{n,\tau}-\sigma_n),$$

such that the claim in (3.78) follows immediately from assertion (3.42) in the proof of part (v) of Theorem 3.2.2.

Step 3: Using Wald's equation again yields

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\,(\widehat{\mathfrak{m}}_{n,\tau}-\mathfrak{m}_n)\\
&=\frac{1}{\sqrt{n}}\,(\widehat{\lambda}_{n,\tau}T\widehat{m}_{n,\tau}-\lambda_n T m)\\
&=\frac{1}{\sqrt{n}}\Big((\widehat{\lambda}_{n,\tau}T-\lambda_n T)\,(\widehat{m}_{n,\tau}-m)\;+\;m\,(\widehat{\lambda}_{n,\tau}T-\lambda_n T)\;+\;\lambda_n T\,(\widehat{m}_{n,\tau}-m)\Big)\\
&=\sqrt{\frac{\lambda_n}{n\tau}}T\,\sqrt{\lambda_n\tau}\,\Big(\widehat{m}_{n,\tau}-m\;+\;m\,\Big(\frac{\widehat{\lambda}_{n,\tau}T}{\lambda_n T}-1\Big)\Big)\;+\;\frac{1}{\sqrt{n}}\,(\widehat{\lambda}_{n,\tau}T-\lambda_n T)\,(\widehat{m}_{n,\tau}-m)\\
&=:S_{3.1}(n)\;+\;S_{3.2}(n).
\end{aligned} \tag{3.80}$$

In Steps 3.1–3.2 below we will show that

$$\mathbb{P}\circ S_{3.1}(n)^{-1}\xrightarrow{\text{w}}\mathcal{N}_{0,c\,(s^2+m^2)T^2/\tau} \tag{3.81}$$

$$S_{3.2}(n)\xrightarrow{\text{p}}0, \qquad , \tag{3.82}$$

such that the claim in (3.79) follows by an application of Slutskys Lemma.

Step 3.1: It suffices to prove that the following statement holds true:

$$\sqrt{\lambda_n\tau}\left(\begin{bmatrix}\widehat{m}_{n,\tau}\\ \frac{\widehat{\lambda}_{n,\tau}T}{\lambda_n T}\end{bmatrix}-\begin{bmatrix}m\\1\end{bmatrix}\right)=\sqrt{\lambda_n\tau}\left(\begin{bmatrix}\frac{1}{\widehat{N}_n(\tau)}\sum_{j=1}^{\widehat{N}_n(\tau)}Y_j\\ \frac{\widehat{N}_n(\tau)}{\lambda_n\tau}\end{bmatrix}-\begin{bmatrix}m\\1\end{bmatrix}\right)\xrightarrow{\text{d}}Z', \tag{3.83}$$

as $n \to \infty$, where Z' refers to some bivariate normally distributed random variable with mean $[0,0]'$ and covariance matrix $\Sigma := \begin{bmatrix} s^2 & 0 \\ 0 & 1 \end{bmatrix}$. In this case the assertion in (3.81) follows directly by an application of the delta-method (see for instance [62], Section 3) w.r.t. the mapping $\phi : \mathbb{R}^2 \to \mathbb{R}$, $\phi(x,y) = x + m\, y$.

By

$$\sqrt{\lambda_n \tau}\,\left(\begin{bmatrix} \frac{1}{\widehat{N}_n(\tau)} \sum_{i=1}^{\widehat{N}_n(\tau)} Y_i \\ \frac{\widehat{N}_n(\tau)}{\lambda_n \tau} \end{bmatrix} - \begin{bmatrix} m \\ 1 \end{bmatrix}\right)$$

$$= \sqrt{\lambda_n \tau}\,\begin{bmatrix} \frac{1}{\lambda_n} \sum_{i=1}^{\widehat{N}_n(\tau)} (Y_i - m) \\ \frac{\widehat{N}_n(\tau)}{\lambda_n \tau} - 1 \end{bmatrix} + \sqrt{\lambda_n \tau}\,\begin{bmatrix} \left(\frac{\lambda_n \tau}{\widehat{N}_n(\tau)} - 1\right) \frac{1}{\lambda_n \tau} \sum_{i=1}^{\widehat{N}_n(\tau)} (Y_i - m) \\ 0 \end{bmatrix} \tag{3.84}$$

it suffices to show that

$$\sqrt{\lambda_n \tau}\,\Big(\frac{\lambda_n \tau}{\widehat{N}_n(\tau)} - 1\Big) \frac{1}{\lambda_n \tau} \sum_{i=1}^{\widehat{N}_n(\tau)} (Y_i - m) \xrightarrow{\ \mathrm{p}\ } 0, \quad (n \to \infty). \tag{3.85}$$

The convergence in distribution of the first summand on the right-hand side in (3.84) to Z' is a direct consequence of Proposition 3.2.12, where we observe that for every $n \in \mathbb{N}$ and $\tau > 0$ fixed, $\widehat{N}_n(\tau)$ is a $\mathrm{Poiss}_{\lambda_n \tau}$-distributed random variable. That is, the assertion in (3.83) would follow by an application of Slutskys Lemma again.

To prove (3.85), we can use Proposition 3.2.10 to conclude that $\lambda_n \tau / \widehat{N}_n(\tau)$ converges to 1 $\mathbb{P}$-a.s. Second, Lemma 2.5.6 in [25] along with an application of Slutskys Lemma yields

$$\mathbb{P} \circ \Big\{ \frac{1}{\sqrt{\lambda_n \tau}} \sum_{i=1}^{\widehat{N}_n(\tau)} (Y_i - m) \Big\}^{-1} \xrightarrow{\ \mathrm{w}\ } \mathcal{N}_{0,s^2}.$$

Hence, the left-hand side in (3.85) converges to zero in distribution by Slutskys Lemma, which leads to the convergence in probability. Thus, the assertion in (3.85) holds true. This completes the proof of part 3.1.

Step 3.2: Following the same line of reasoning as in the proof of (3.85) again, we conclude that (3.82) holds true. This completes the proof of part (v).

(iv) The assertion can be proven in same way as part (v). Following the same line of reasoning as in part (v), we can conclude that (3.77)–(3.78) hold true. Then, using arguments as in the proof of Step 3.2 above, the $\mathbb{P}$-a.s. convergence to 0 of the remaining summand follows by an application of the Marcinkiewicz-Zygmund SLLN of Theorem 2.5.5 in [25]. □

Proof of Corollary 3.2.6 (i) First, we observe that for every $n \in \mathbb{N}$ we have

$$\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}, \widehat{\sigma}^2_{n,\tau}}) - \frac{1}{n}\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{\,*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}})$$

$$= \frac{\widehat{\lambda}_{n,\tau} T}{n}\,\Big(\frac{1}{\widehat{\lambda}_{n,\tau} T}\mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}, \widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\widehat{\lambda}_{n,\tau} T}\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{\,*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}})\Big)$$

$$= \frac{\lambda_n T}{n} \frac{\widehat{N}_n(\tau)}{\lambda_n \tau} \Big(\frac{1}{\widehat{\lambda}_{n,\tau} T} \mathcal{R}_\rho(\mathcal{N}_{\widehat{\mathfrak{m}}_{n,\tau}, \widehat{\sigma}^2_{n,\tau}}) - \frac{1}{\widehat{\lambda}_{n,\tau} T} \mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau} T}}) \Big). \tag{3.86}$$

Hence, we can use part (b) of Assumption 3.2.1 to deduce that $\lambda_n T/n$ converges to some positive constant cT. Furthermore, Proposition 3.2.10 yields the $\mathbb{P}$-a.s. convergence of $\widehat{N}_n(\tau)/(\lambda_n \tau)$ to 1. Together with part (i) of Theorem 3.2.3 this yields the claim of part (i).

(ii) The assertion follows from part (i) together with part (iii) of Corollary 3.2.4.

(iii)–(iv): The assertions can be proven in the same way as assertions (iv)–(v) of Corollary 3.2.4. □

Part II

Estimation under constant collective sizes

Chapter 4

Nonparametric estimation under constant collective sizes in the individual model (revisited)

In the first part of this thesis we have developed a theory to estimate individual premiums against the background of increasing collective sizes, assuming that estimations were only based on the last "few" insurance periods rather than the whole observation history. In contrast to this theory, we now assume that the collective size is constant and the number of observations is increasing. In this chapter we will basically recall the ideas presented in [12] to derive the asymptotic distribution and almost sure bootstrap consistency for the estimated premiums in the individual model.

To this end, let (X_i) be a sequence of nonnegative i.i.d. random variables with distribution μ. Throughout this chapter we will refrain from our notation used in the former chapters. From now on we will therefore use the distribution function F associated with μ to develop our theory, instead of focusing on the distribution μ itself. For every $n \in \mathbb{N}$, let

$$S_n := \sum_{i=1}^{n} X_i.$$

In accordance with Chapter 2, we will think of S_n again as the total claim size in a homogeneous insurance collective with n individual risks. The distribution function of S_n is then given by F^{*n}, where F^{*n} refers to the n-fold convolution of F. That is, $F^{*0} := \mathbb{1}_{[0,\infty)}$ and

$$\begin{aligned} F^{*n}(x) &:= \int F(x - x_{n-1})\, dF^{*(n-1)}(x_{n-1}) \\ &= \int \cdots \int F(x - x_{n-1} - \cdots - x_1)\, dF(x_1) \cdots dF(x_{n-1}) \end{aligned} \tag{4.1}$$

for every $n \in \mathbb{N}$. We regard F^{*n} as the image of a mapping $\mathcal{C}_n$. To this end, let $\mathcal{C}_n : \mathbb{F} \to \mathbb{F}$ be the functional defined by

$$\mathcal{C}_n(F) := F^{*n}, \tag{4.2}$$

where $\mathbb{F}$ denotes the set of all distribution functions. Hence, an adequate total premium w.r.t. a risk measure ρ would then be given by the evaluation of the risk functional $\mathcal{R}_\rho$ at F^{*n}, that is by

$$\mathcal{R}_n := \mathcal{R}_\rho(\mathcal{C}_n(F)). \tag{4.3}$$

Note, that in this chapter we regard the risk functional as a functional of the distribution function F, rather than the measure μ. This does clearly not impose a restriction, whenever ρ refers to a law-invariant risk measure. In view of (4.3), we stress the fact again, that we assumed the collective size $n \in \mathbb{N}$ to be constant throughout the second part of this thesis. Of course, a suitable individual premium could then be obtained by dividing the total premium in (4.3) by the collective size n. However, as the collective size does not vary, this would not change the asymptotic behavior essentially. Throughout the rest of this chapter we will therefore refer to $\mathcal{R}_n$ as *the premium* rather than the individual premium.

The rest of this chapter is organized as follows. In Section 4.1 we will briefly introduce the estimator for the premium in the individual model. In Section 4.2 we will then introduce the notion of uniform quasi-Hadamard differentiability, which will be needed for the determination of the asymptotic distributions with the help of the delta-method in the form of [12]. Section 4.3 will then be devoted to deriving the asymptotic distribution and establishing almost sure bootstrap consistency for the sequence of estimated premiums. The representation in (4.3) already points out that this will be achieved by an application of the delta-method and the chain rule in the form of [12]. Again, we stress the fact that the presented theory throughout this chapter is a recapitulation of the results in [12].

4.1 An estimator for the premium in the individual model

This section is devoted to the estimation of the premium $\mathcal{R}_n$, as in (4.3). To this end, let $u \in \mathbb{N}$, and $Y_1, \dots, Y_u$ be nonnegative i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F. We will think of $Y_1, \dots, Y_u$ as single claims, which have been reported to the insurance company in the past. For every $u \in \mathbb{N}$, an estimator for the unknown distribution function F of the single claim distribution would then be given by the mapping $\widehat{F}_u : \Omega \to \mathbb{F}$, defined by

$$\widehat{F}_u := \frac{1}{u} \sum_{i=1}^{u} \mathbb{1}_{[Y_i,\infty)}, \tag{4.4}$$

where $\mathbb{F}$ denotes the set of all distribution functions. Hence, an estimator for the distribution function of the total claim amount would then be given by

$$\widehat{F}_u^{*n} := \mathcal{C}_n(\widehat{F}_u),$$

where $\mathcal{C}_n$ is as in (4.2). A reasonable estimator for the premium based on a collective of n clients and u observations, would then be given by the corresponding plug-in estimator

$$\widehat{\mathcal{R}}_{n,u} := \mathcal{R}_\rho(\mathcal{C}_n(\widehat{F}_u)). \tag{4.5}$$

In the following we will be interested in deriving the asymptotic distribution of the estimated premium in (4.5) as u tends to infinity and n remains constant. Furthermore we will establish the asymptotic distribution and almost sure bootstrap consistency for the estimator in (4.5).

4.2 The notion of uniform quasi-Hadamard differentiability

In this section we will introduce the notion of uniform quasi-Hadamard differentiability, which will be used to derive a delta-method and bootstrap results for plug-in estimators in our present setting. The uniform quasi-Hadamard differentiability, as introduced in [12], extends the notion of quasi-Hadamard differentiability, as introduced in [10]. To this end, let $\mathbf{V}$ and $\widetilde{\mathbf{V}}$ be vector spaces. Let $\mathbf{E} \subset \mathbf{V}$ and $\widetilde{\mathbf{E}} \subset \widetilde{\mathbf{V}}$ be subspaces equipped with norms $\|\cdot\|_{\mathbf{E}}$ and $\|\cdot\|_{\widetilde{\mathbf{E}}}$. Furthermore, let

$$G : V_G \longrightarrow \widetilde{\mathbf{V}}$$

be any map defined on some domain $V_G \subset \mathbf{V}$.

Definition 4.2.1 *Let $\mathbf{E}_0$ be a subset of $\mathbf{E}$, and $\mathcal{S}$ be a set of sequences in V_G.*

(i) The map G is said to be uniformly quasi-Hadamard differentiable w.r.t. $\mathcal{S}$ tangentially to $\mathbf{E}_0\langle\mathbf{E}\rangle$ with trace $\widetilde{\mathbf{E}}$, if $G(y_1) - G(y_2) \in \widetilde{\mathbf{E}}$ for all $y_1, y_2 \in V_G$, and there is some continuous map $\dot{G}_{\mathcal{S}} : \mathbf{E}_0 \to \widetilde{\mathbf{E}}$ such that

$$\lim_{u\to\infty} \left\| \dot{G}_{\mathcal{S}}(x) - \frac{G(z_u + \varepsilon_u x_u) - G(z_u)}{\varepsilon_u} \right\|_{\widetilde{\mathbf{E}}} = 0 \tag{4.6}$$

holds for each quadruple $((z_u), x, (x_u), (\varepsilon_u))$, with $(z_u) \in \mathcal{S}$, $x \in \mathbf{E}_0$, $(x_n) \subset \mathbf{E}$, satisfying $\|x - x_u\|_{\mathbf{E}} \to 0$, as well as $(z_u + \varepsilon_u x_u) \subset V_G$, and $(\varepsilon_u) \subset (0,\infty)$ satisfying $\varepsilon_u \to 0$. In this case the map $\dot{G}_{\mathcal{S}}$ is called uniform quasi-Hadamard derivative of G w.r.t. $\mathcal{S}$ tangentially to $\mathbf{E}_0\langle\mathbf{E}\rangle$.

(ii) If $\mathcal{S}$ consists of all sequences $(z_u) \subset V_G$ with $z_u - z \in \mathbf{E}$, $u \in \mathbb{N}$, and $\|z_u - z\|_{\mathbf{E}} \to 0$ for some fixed $z \in V_G$, then we replace the phrase "w.r.t. $\mathcal{S}$" by "at z" and "$\dot{G}_{\mathcal{S}}$" by "$\dot{G}_z$".

(iii) If $\mathcal{S}$ consists only of the constant sequence $z_u = z$, $u \in \mathbb{N}$, then we skip the phrase "uniformly" and replace the phrase "w.r.t. $\mathcal{S}$" by "at z" and "$\dot{G}_{\mathcal{S}}$" by "$\dot{G}_z$". In this case we may also replace "$G(y_1) - G(y_2) \in \widetilde{\mathbf{E}}$ for all $y_1, y_2 \in V_G$" by "$G(y) - G(z) \in \widetilde{\mathbf{E}}$ for all $y \in V_G$".

(iv) If $\mathbf{E} = \mathbf{V}$*, then we skip the phrase "quasi-".*

(v) If $\widetilde{\mathbf{E}} = \widetilde{\mathbf{V}}$*, then we skip the phrase "with trace* $\widetilde{\mathbf{E}}$*".*

The definition extends the classical notion of uniform Hadamard differentiability as introduced in Theorem 3.9.11 in [61] in the following sense. Using the differentiability concept in (i) with $\mathcal{S}$ as in (ii), leads to the classical uniform Hadamard differentiability. Proposition 4.1 in [12] shows that it might turn out to be beneficial to refrain from insisting on $\mathbf{E} = \mathbf{V}$ as in part (iv). Of course the condition of the uniform quasi-Hadamard differentiability gets weaker the smaller the set $\mathcal{S}$ gets.

4.3 The asymptotic distribution and almost sure bootstrap consistency of the estimated premium in the individual model

In this section we aim to derive results about the asymptotic distribution and almost sure bootstrap consistency of the estimated premium in the individual model. To this use, we will first introduce the bootstrap estimator for the premium in our present setting. The presented theory is based on Example 3.3 in [12]. Subsequent to the introduction of the bootstrap estimator, we will first state a general theorem about the asymptotic distribution and almost sure bootstrap consistency of the estimated premium for general risk measures ρ. To serve a concrete example, we will then consider the premium derived from the Average Value at Risk of Example 1.2.4 and state the corresponding asymptotic distribution.

In accordance with the former section, let (Y_i) be a sequence of i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F, and let $\widehat{F}_u$ be as in (4.4). Let $(\Omega', \mathcal{F}', \mathbb{P}')$ be a second probability space and set

$$(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}}) := (\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mathbb{P} \otimes \mathbb{P}').$$

Let $(W_{u,i})$ be a triangular array of nonnegative real-valued random variables on $(\Omega', \mathcal{F}', \mathbb{P}')$, such that $(W_{u,1}, \ldots, W_{u,u})$ is an exchangeable random vector for every $u \in \mathbb{N}$, and define the map $\widehat{F}_u^{\mathrm{B}} : \overline{\Omega} \to \mathbb{F}$ by

$$\widehat{F}_u^{\mathrm{B}}(\omega, \omega') := \frac{1}{u} \sum_{i=1}^{u} W_{u,i}(\omega') \, \mathbb{1}_{[Y_i(\omega), \infty)}. \tag{4.7}$$

Note that the triangular array $(W_{u,i})$ and the sequence (Y_i), regarded as random variables on the product space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$, are independent. Of course we will tactically assume that $(\Omega', \mathcal{F}', \mathbb{P}')$ is rich enough to host all random variables used below. Let $\overline{W}_u := \frac{1}{u} \sum_{i=1}^{u} W_{u,i}$. Now we assume that F satisfies $\int \phi_\lambda^2 dF < \infty$ and that the following assertions hold.

A1. $\sup_{u \in \mathbb{N}} \int_0^\infty \mathbb{P}'[|W_{u,1} - \overline{W}_u| > t]^{1/2} \, dt < \infty$.

A2. $\frac{1}{\sqrt{u}}\mathbb{E}'[\max_{1\le i\le u}|W_{u,i}-\overline{W}_u|]\to 0$.

A3. $\frac{1}{u}\sum_{i=1}^{u}(W_{u,1}-\overline{W}_u)^2\to 1$ in $\mathbb{P}'$-probability.

Examples 3.6.9 and 3.6.10 in [61] have shown that conditions A1.–A3. are met under the assumptions of parts (i) and (ii) of Example 2.4.2, that is, if the resampling scheme corresponds to Efron's bootstrap or the Bayesian bootstrap, for instance.

Note that we might regard (4.7) as a bootstrap version of the estimated distribution function $\widehat{F}_u$ of the single claim distribution. In this context the bootstrap estimator for the distribution function of the total claim distribution is given by

$$(\widehat{F}_u^{\mathrm{B}})^{*n} := \mathcal{C}_n(\widehat{F}_u^{\mathrm{B}}).$$

Hence, the bootstrap estimator for the premium in the individual model is given by

$$\widehat{\mathcal{R}}_{n,u}^{\mathrm{B}} := \mathcal{R}_\rho(\mathcal{C}_n(\widehat{F}_u^{\mathrm{B}})). \tag{4.8}$$

Let D be the space of all càdlàg functions. Moreover, let $\lambda\ge 0$ and $\phi_\lambda:\mathbb{R}\to[1,\infty)$ be a weight function given by $\phi_\lambda(x):=(1+|x|)^\lambda$. Let $\mathrm{D}_{\phi_\lambda}$ be the subspace of D consisting of all elements $v\in\mathrm{D}$, satisfying $\|v\|_{\phi_\lambda}:=\|v\phi_\lambda\|_\infty<\infty$ and $\lim_{x\to\pm\infty}|v(x)|=0$. Furthermore, let $\mathbb{F}_{\phi_\lambda}$ denote the subspace of $\mathbb{F}$ consisting of all elements $F\in\mathbb{F}$ satisfying $\int\phi_\lambda dF<\infty$. Note that the latter condition is equivalent to $\int|x|^\lambda dF(x)<\infty$. Moreover, let $\mathbb{F}_\lambda$ denote the subspace of all elements $F\in\mathbb{F}$ satisfying $\int|x|^\lambda\,dF(x)<\infty$.

To guarantee that the composition $\mathcal{R}_\rho\circ\mathcal{C}_n$ is well defined, we have to assume that the risk functional is a mapping defined on $\mathbb{F}_{\lambda'}$ for some $\lambda'\ge 0$. Choosing $F\in\mathbb{F}_{\phi_\lambda}$, Lemma 2.2 in [52] yields that $\mathcal{C}_n(\mathbb{F}_{\phi_\lambda})\subset\mathbb{F}_{\lambda'}$, for every $\lambda>\lambda'$, such that $\mathcal{R}_\rho\circ\mathcal{C}_n$ is well defined on $\mathbb{F}_{\phi_\lambda}$.

The formulation of the following theorem will require the definition of an F-Brownian bridge. The latter is defined as a centered Gaussian process with covariance function

$$\Gamma(t_0,t_1)=F(t_0\wedge t_1)(1-F(t_0\vee t_1)). \tag{4.9}$$

Theorem 4.3.1 *Let $\lambda>\lambda'>1$ and $F\in\mathbb{F}_{\phi_{2\lambda}}$, that is $\int\phi_\lambda^2\,dF<\infty$. Let $\mathcal{R},\widehat{\mathcal{R}}_{n,u},\widehat{\mathcal{R}}_{n,u}^{\mathrm{B}}$ be as in (4.3), (4.5) and (4.8), respectively. Moreover, let B_F be an F-Brownian bridge.*

Let $\mathcal{S}$ be the set of all sequences $(G_u)\subset\mathbb{F}$ satisfying $G_u\to\mathcal{C}_n(F)$ pointwise. Furthermore, assume that $\mathcal{R}_\rho$ is uniformly quasi-Hadamard differentiable w.r.t. $\mathcal{S}$ tangentially to $\mathrm{D}_{\phi_\lambda}\langle\mathrm{D}_{\phi_\lambda}\rangle$ with uniform quasi-Hadamard derivative $\dot{\mathcal{R}}_{\rho,\mathcal{S}}$. Then we have

$$\sqrt{u}\left(\widehat{\mathcal{R}}_{n,u}-\mathcal{R}_n\right)\xrightarrow{\mathrm{d}}\dot{\mathcal{R}}_{\rho,\mathcal{C}_n(F)}\circ\dot{\mathcal{C}}_{n,F}(B_F),\qquad \textit{in } (\mathbb{R},\mathcal{B}(\mathbb{R}))$$

and

$$\sqrt{u}\left(\widehat{\mathcal{R}}_{n,u}^{\mathrm{B}}(\omega,\cdot)-\widehat{\mathcal{R}}_{u,n}(\omega)\right)\xrightarrow{\mathrm{d}}\dot{\mathcal{R}}_{\rho,\mathcal{C}_n(F)}\circ\dot{\mathcal{C}}_{n,F}(B_F),\qquad \textit{in } (\mathbb{R},\mathcal{B}(\mathbb{R}))\quad \mathbb{P}\text{-}a.e.\ \omega,$$

where $\dot{\mathcal{C}}_{(n,F)} : \mathrm{D}_{\phi_\lambda} \to \mathrm{D}_{\phi_{\lambda'}}$ *is defined by*

$$\dot{\mathcal{C}}_{(n,F)}(v) := n\, v * F^{*(n-1)} = n \int v(\,\cdot\, - x)\, dF^{*(n-1)}(x). \tag{4.10}$$

The assertion in Theorem 4.3.1 is an obvious special case of the assertion in Theorem 5.2.1. We will therefore not prove Theorem 4.3.1 at this point.

As we have seen in Section 1.3, the Average Value at Risk at level $\alpha \in (0,1)$ is a distortion risk measure and thus, possesses an integral representation w.r.t. a distortion function. More explicitly, let $(\Omega, \mathcal{F}, \mathbb{P})$ be an atomless probability space and let $L^1(\Omega, \mathcal{F}, \mathbb{P})$ denote the usual L^1-space. Then the Average Value at Risk at level $\alpha \in (0,1)$ is the map $\mathrm{AV@R}_\alpha : L^1 \to \mathbb{R}$ defined by

$$\mathrm{AV@R}_\alpha(X) := -\int_{-\infty}^{0} g_\alpha(F_X(x))\, dx \, + \int_0^\infty (1 - g_\alpha(F_X(x)))\, dx,$$

where $g_\alpha(t) := \frac{1}{1-\alpha} \max\{0, t - \alpha\}$. In view of the upper identity, we might also think of $\mathrm{AV@R}_\alpha$ as a statistical functional $\mathcal{R}_\alpha : \mathbb{F}_1 \to \mathbb{R}$, defined by

$$\mathcal{R}_\alpha(F) \, := \, -\int_{-\infty}^{0} g_\alpha(F(x))\, dx \, + \int_0^\infty (1 - g_\alpha(F(x)))\, dx. \tag{4.11}$$

We will now consider the composition of the Average Value at Risk functional $\mathcal{R}_\alpha$, as defined in (4.11), and the compound distribution functional $\mathcal{C}_n$ as in (4.2). Note that for any $\lambda > 1$, Lemma 2.2 in [52] yields $\mathcal{C}_n(\mathbb{F}_{\phi_\lambda}) \subset \mathbb{F}_1$, such that the composition $\mathcal{R}_\alpha \circ \mathcal{C}_n$ is well defined on $\mathbb{F}_{\phi_\lambda}$. Let $\kappa_{\alpha,n} : \mathbb{F}_{\phi_\lambda} \to \mathbb{R}$ be defined by

$$\kappa_{\alpha,n} := \mathcal{R}_\alpha \circ \mathcal{C}_n. \tag{4.12}$$

Note that in this special case the premium $\mathcal{R}_n$, the premium estimator $\widehat{\mathcal{R}}_{n,u}$, as well as the bootstrap premium estimator $\widehat{\mathcal{R}}_{n,u}^{\mathrm{B}}$ are given by $\kappa_{\alpha,n}(F)$, $\kappa_{\alpha,n}(\widehat{F}_u)$ and $\kappa_{\alpha,n}(\widehat{F}_u^{\mathrm{B}})$, respectively.

Theorem 4.3.2 *Let* $\lambda > 1$ *and* $F \in \mathbb{F}_{\phi_\lambda}$. *Furthermore, assume that* $\mathcal{C}_n(F)$ *takes the value* $1 - \alpha$ *only once. Then the map* $\kappa_{\alpha,n} := \mathcal{R}_\alpha \circ \mathcal{C}_n : \mathbb{F}_{\phi_\lambda}(\subset \mathrm{D}) \to \mathbb{R}$ *is uniformly quasi-Hadamard differentiable at* F *tangentially to* $\mathrm{D}_{\phi_\lambda}\langle \mathrm{D}_{\phi_\lambda}\rangle$, *and the uniform quasi-Hadamard derivative* $\dot{\kappa}_{\alpha,n,F} : \mathrm{D}_{\phi_\lambda} \to \mathbb{R}$ *is given by* $\dot{\kappa}_{\alpha,n,F} := \dot{\mathcal{R}}_{\alpha,\mathcal{C}_n(F)} \circ \dot{\mathcal{C}}_{(n,F)}$, *i.e.*

$$\dot{\kappa}_{\alpha,n,F}(v) = \int g'_\alpha(\mathcal{C}_n(F)(x))\, \dot{\mathcal{C}}_{(n,F)}(v)(x)\, dx, \tag{4.13}$$

where $g'_\alpha(t) = \frac{1}{1-\alpha} \mathbb{1}_{(1-\alpha,1]}(t)$ *and* $\dot{\mathcal{C}}_{(n,F)}$ *is as in (4.10).*

Proof The proof of the assertion would be a direct consequence of Corollary 4.6 in [12] if we could show that the mapping $\mathcal{C}_n$ is uniformly quasi-Hadamard differentiable tangentially to $\mathrm{D}_{\phi_\lambda}\langle \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ and uniform quasi-Hadamard derivative $\dot{\mathcal{C}}_{(n,F)}$ given by (4.10).

The latter can be obtained as a special case of Proposition 4.3 in [12] by choosing $p_n = 1$ and $p_k = 0$ for every $k \neq n$. $\square$

As a direct consequence of Theorems 4.3.2 and 4.3.1 we obtain the following corollary on the asymptotic distribution and the almost sure bootstrap consistency of the premium in the individual model.

Corollary 4.3.3 *Let $\lambda > 1$ and let $F \in \mathbb{F}_{\phi_{2\lambda}}$. Moreover, let $\widehat{F}_u$ and $\widehat{F}_u^{\mathrm{B}}$ be as in (4.4) and (4.7) and let B_F be as in (4.9). Assume that $\mathcal{C}_n(F)$ takes the value $1-\alpha$ only once. Then we have*

$$\sqrt{u}\left(\kappa_{\alpha,n}(\widehat{F}_u) - \kappa_{\alpha,n}(F)\right) \stackrel{\mathrm{d}}{\longrightarrow} \dot{\kappa}_{\alpha,n,F}(B_F), \qquad \text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R})) \tag{4.14}$$

and

$$\sqrt{u}\left(\kappa_{\alpha,n}(\widehat{F}_u^{\mathrm{B}}(\omega,\,\cdot\,)) - \kappa_{\alpha,n}(\widehat{F}_u(\omega))\right) \stackrel{\mathrm{d}}{\longrightarrow} \dot{\kappa}_{\alpha,n,F}(B_F), \qquad \text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R})) \quad \mathbb{P}\text{-a.e. } \omega. \tag{4.15}$$

Chapter 5

Semiparametric estimation under constant collective sizes in the collective model

In this chapter we aim to derive similar results to those of the former chapter for the collective model. More explicitly, we will aim to derive the asymptotic distribution of the sequence of estimated premiums in the collective model. Again, we assume that the collective size is constant and the number of observations is increasing. In contrast to the former chapter, our theory has to be modified because the quantity to estimate is no longer just the empirical distribution function, but also the distribution of the number of claims. In the setting of the former chapter the latter distribution was nothing but the dirac measure at n. In the following we will therefore expand our theory to a wider class of claim number distributions.

To embed this notion in a mathematical context, let (X_i) be a sequence of nonnegative i.i.d. random variables with unknown distribution function F. Moreover let N be an $\mathbb{N}_0$-valued random variable being independent of (X_i). Furthermore, let

$$S_N := \sum_{i=1}^{N} X_i.$$

We will think of S_N again as the future total claim size of an insurance collective producing homogeneous claims. Note that we oppress the dependence of N on $n \in \mathbb{N}$ throughout this chapter, as we assume the collective size to be constant. Again, the distribution of N is unknown. However in many practical applications one does roughly know which parametrical class of distributions the distribution of N belongs to. In the Cramér-Lundberg model of Section 3.1 for instance, we know that this role is played by a Poisson distribution with unknown parameter $\theta \in (0, \infty)$. In particular, the distribution of the number of claims is then specified by a nonnegative sequence $p(\theta) := (p_k(\theta))_{k \in \mathbb{N}_0}$ in dependence on the claim intensity $\theta \in (0, \infty)$. In this particular example, for every $k \in \mathbb{N}_0$, $p_k : (0, \infty) \to \mathbb{R}_+$ is a mapping defined by

$$p_k(\theta) := \frac{\theta^k}{k!}\, e^{-\theta}.$$

In this context, the distribution function of the total claim distribution is given by

$$F^{*\mathrm{Poiss}_\theta} := \sum_{k=0}^{\infty} F^{*k} \frac{\theta^k}{k!}\, e^{-\theta}, \tag{5.1}$$

where F^{*k} refers to the k-fold convolution of F, as introduced in (4.1). In accordance with Chapter 4, we will regard $F^{*\mathrm{Poiss}_\theta}$ as the image of a mapping defined on the set of parameters and the underlying distribution functions. To this end, let $\mathcal{C} : (0, \infty) \times \mathbb{F} \to \mathbb{F}$ be defined by

$$\mathcal{C}(\theta, F) := \sum_{k=0}^{\infty} F^{*k} \frac{\theta^k}{k!}\, e^{-\theta}. \tag{5.2}$$

In the following we will refer to $\mathcal{C}$ as the *compound distribution functional*. The corresponding (total) premium w.r.t. a risk measure ρ is then given by

$$\mathcal{R} := \mathcal{R}_\rho(\mathcal{C}(\theta, F)). \tag{5.3}$$

Note that there is a hidden dependence of the premium on the collective size n, because the parameter modeling the claim intensity θ does depend on the size of the underlying collective. However, as we assumed the number of clients in the collective to be constant, we will omit this dependence because it does not affect the asymptotic behavior essentially. Once again, a suitable individual premium could be obtained by dividing the quantity in (5.3) by n.

Our goal in this chapter is to derive the asymptotic distribution of the estimated premium in this context. Furthermore, we will give an outlook on the almost sure bootstrap consistency of the estimated premiums. Section 5.1 will be devoted to the choice of estimators. To this end, let $\widehat{\mathcal{R}}_u$ be an estimator for the premium in (5.3), where u refers to the number of observations taken into account for the estimation. Based on $\widehat{\mathcal{R}}_u$, we will aim to derive the asymptotic distribution, that is, the weak limit of the laws of

$$\sqrt{u}\, \left(\widehat{\mathcal{R}}_u - \mathcal{R}\right) \tag{5.4}$$

as the number of observations u tends to infinity. In practical applications this error distribution can theoretically be used to derive asymptotic confidence intervals for the premium $\mathcal{R}$. However, in many applications a derivation of the exact asymptotic distribution of the sequence in (5.4) is more or less impossible. A widely used technique to handle this problem is again the bootstrap. To this end, let ζ denote the limit in distribution of the sequence in (5.4). Section 5.3 will give an outlook on the almost sure bootstrap consistency of the sequence of estimated premiums. The latter means, that

$$\sqrt{u}\, \left(\widehat{\mathcal{R}}_u^{\mathrm{B}}(\omega, \cdot\,) - \widehat{\mathcal{R}}_u(\omega)\right) \xrightarrow{\mathrm{d}} \zeta, \qquad \mathbb{P}\text{-a.e. } \omega, \tag{5.5}$$

holds, where $\widehat{\mathcal{R}}_u^{\mathrm{B}}$ is a suitable bootstrap version of $\widehat{\mathcal{R}}_u$. As the bootstrap version $\widehat{\mathcal{R}}_u^{\mathrm{B}}$ only depends on the bootstrap mechanism and the initial sample ω, one can at least numerically determine the asymptotic error distribution by means of a Monte-Carlo procedure. A central tool to derive the asymptotic distribution of the initial sequence of estimators will be the recently established functional delta-method for uniformly quasi-Hadamard differentiable functionals in the form of Corollary 3.1 in [12]. The representation in (5.3) points out that we will consider the premium $\mathcal{R}$ to be a composition of the mapping $\mathcal{C}$, mapping the claim intensity parameter and the distribution function onto the compound distribution function, together with the risk functional $\mathcal{R}_\rho$. To this end, we will also need the chain rule for uniformly quasi-Hadamard differentiable functionals in the form of Lemma A.5 in [12] to prove the uniform quasi-Hadamard differentiability of the composition.

The rest of this chapter is organized as follows. In Section 5.1 we will give a brief introduction to our considered estimators and introduce the estimator for the premium. In Section 5.2 we will present our main results about the asymptotic distribution of the estimated premiums. We will first formulate these results w.r.t. a general risk measure ρ and will then consider the premium w.r.t. the Average Value at Risk at level $\alpha \in (0, 1)$ to serve a concrete example. Section 5.3 will then give an outlook on the almost sure bootstrap consistency of the sequence of premium estimators if we could achieve almost sure boostrap consistency of the sequence of underlying estimators for the claim intensity θ and the single claim distribution function F. In Section 5.4 we will then prove our results of Section 5.2. To this use, we will first recall the recently established delta-method for uniform quasi-Hadamard differentiable functionals of Corollary 3.1 in [12], which will be the central tool to determine the asymptotic error distribution of our sequence of premium estimators. This will be done in Section 5.4.1. In Section 5.4.2 we will then determine the uniform quasi-Hadamard derivative of the compound distribution functional $\mathcal{C}$. Furthermore, we will derive the asymptotic distribution of the estimated compound distribution function, that is, the compound distribution functional $\mathcal{C}$ applied to our estimators for θ and F. This will be done in Section 5.4.3. The latter will be needed to derive results on the asymptotic distribution of the estimated premiums with the help of the delta-method.

5.1 A semiparametric estimator for the premium in the compound Poisson model

Our next goal is again to estimate a suitable premium for the insurance period to come based on historically observed claim amounts and claim numbers. To this end, let $N_1, \ldots, N_u$ be i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $N_1 \sim \mathrm{Poiss}_\theta$ for some $\theta \in (0, \infty)$. Each N_i represents the number of claims, which have been reported to the insurer during the i-th insurance period. Moreover, let (Y_i) be a sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution function F, being independent of $(N_1, \ldots, N_u)$. In this case

$Y_1, \ldots, Y_{N_1+\ldots+N_u}$ represent the claim sizes, which have been reported to the insurance company throughout the last $u \in \mathbb{N}$ insurance periods.

To estimate the unknown parameter θ of the Poisson-distribution, we can use the standard Maximum-Likelihood Estimator (MLE). In particular, for every $u \in \mathbb{N}$, let $\widehat{\theta}_u : \Omega \to (0, \infty)$ be the mapping defined by

$$\widehat{\theta}_u := \frac{1}{u} \sum_{i=1}^{u} N_i. \tag{5.6}$$

In this case $\widehat{\theta}_u$ provides an unbiased estimator for $\theta \in (0, \infty)$. To simplify the notation, let $N(u) := N_1 + \ldots + N_u$. Note that the total number of claims $N(u)$ might as well be regarded as the value of a homogeneous Poisson-process with rate 1 at time $u\theta$. Thus, the mapping $\widehat{F}_u : \Omega \to \mathbb{F}$, given by

$$\widehat{F}_u := \frac{1}{N(u)} \sum_{i=1}^{N(u)} \mathbb{1}_{[Y_i, \infty)} \tag{5.7}$$

is the standard nonparametric estimator for the unknown distribution function F, provided $N(u) > 0$. For $N(u) = 0$ we simply set $\widehat{F}_u = \mathbb{1}_{[0,\infty)}$. Here $\mathbb{F}$ denotes the space of all distribution functions. Following this line of reasoning, we can use

$$\widehat{F}_u^{*\mathrm{Poiss}_{\widehat{\theta}_u}} := \mathcal{C}(\widehat{\theta}_u, \widehat{F}_u) \tag{5.8}$$

to estimate the distribution of the total claim, where $\mathcal{C}$ is the compound distribution functional as in (5.2). Hence,

$$\widehat{\mathcal{R}}_u := \mathcal{R}_\rho(\mathcal{C}(\widehat{\theta}_u, \widehat{F}_u)) \tag{5.9}$$

provides a reasonable estimator for the total premium. Consequently, an estimator for the individual premium is then given by $\widehat{\mathcal{R}}_u / n$.

5.2 Asymptotic distribution of the premium estimator in the compound Poisson model

In this section we are going to derive results about the asymptotic distribution of the sequence of estimated premiums. We will first state a general result about the asymptotic distribution of the estimated premiums. To be able to formulate this result, we will assume the underlying risk functional to be uniformly quasi-Hadamard differentiable in the sense of Definition 4.2.1. To serve a concrete example, we will then determine the asymptotic distribution of a sequence of estimated premiums based on the Average Value at Risk of Example 1.2.4.

Theorem 5.2.1 will be the general formulation of our main theorem in Chapter 5. In the formulation of the theorem we will consider the premium w.r.t. a general risk measure ρ. To guarantee that the composition $\mathcal{R}_\rho \circ \mathcal{C}$ is well defined, we have to assume that the risk functional is a mapping defined on $\mathbb{F}_{\lambda'}$ for some $\lambda' \geq 0$. Choosing $F \in \mathbb{F}_{\phi_\lambda}$, Lemma 2.2 in

[52] yields that $\mathcal{C}((0,\infty)\times\mathbb{F}_{\phi_\lambda})\subset\mathbb{F}_{\lambda'}$, for every $\lambda>\lambda'$, such that $\mathcal{R}_\rho\circ\mathcal{C}$ is well defined on $(0,\infty)\times\mathbb{F}_{\phi_\lambda}$.

Theorem 5.2.1 *Let $\lambda>\lambda'>1$, $\theta\in(0,\infty)$ and let $F\in\mathbb{F}_{\phi_{2\lambda}}$, that is $\int\phi_\lambda^2\,dF<\infty$. Let $\mathcal{R}$ and $\widehat{\mathcal{R}}_u$ be as in (5.3) and (5.9), respectively. Moreover, let ξ be a $\mathcal{N}_{0,\theta}$-distributed random variable and B_F be an F-Brownian bridge, as in (4.9), being independent of ξ.*

Let $\mathcal{S}$ be the set of all sequences $(G_u)\subset\mathbb{F}_\lambda$, satisfying $G_u\to\mathcal{C}(\theta,F)$ pointwise. Assume that $\mathcal{R}_\rho$ is uniformly quasi-Hadamard differentiable w.r.t. $\mathcal{S}$ tangentially to $\mathrm{D}_{\phi_{\lambda'}}\langle\mathrm{D}_{\phi_{\lambda'}}\rangle$ with uniform quasi-Hadamard derivative $\dot{\mathcal{R}}_{\rho,\mathcal{S}}$. Then we have

$$\sqrt{u}\Big(\widehat{\mathcal{R}}_u-\mathcal{R}\Big)\stackrel{\mathrm{d}}{\longrightarrow}\dot{\mathcal{R}}_{\rho,\mathcal{C}(\theta,F)}\circ\dot{\mathcal{C}}_{(\theta,F)}\Big(\xi,\frac{1}{\sqrt{\theta}}B_F\Big),\qquad \text{in } (\mathbb{R},\mathcal{B}(\mathbb{R})),$$

where $\dot{\mathcal{C}}_{(\theta,F)}:(0,\infty)\times\mathrm{D}_{\phi_\lambda}\to\mathrm{D}_{\phi_{\lambda'}}$ is given by

$$\dot{\mathcal{C}}_{(\theta,F)}(w,v):=v*\sum_{k=1}^{\infty}kF^{*(k-1)}\frac{\theta^k}{k!}e^{-\theta}+w\,e^{-\theta}\sum_{k=0}^{\infty}F^{*k}\,\frac{1}{k!}\,(k\,\theta^{k-1}-\theta^k). \tag{5.10}$$

The proof of Theorem 5.2.1 can be found in Section 5.4.4.

We will now consider the composition of the Average Value at Risk functional with the compound distribution functional. More explicitly, we will prove an analogue to Theorem 5.2.1, where the role of ρ is played by the Average Value at Risk at level $\alpha\in(0,1)$. To this end, let again $\alpha\in(0,1)$ and $\mathcal{R}_\alpha:\mathbb{F}_1\to\mathbb{R}$ be as in (4.11). Let $\kappa_\alpha:(0,\infty)\times\mathbb{F}\to\mathbb{R}$ be defined by

$$\kappa_\alpha:=\mathcal{R}_\alpha\circ\mathcal{C}. \tag{5.11}$$

In this special case the premium $\mathcal{R}$ and the premium estimator $\widehat{\mathcal{R}}_u$ are given by $\kappa_\alpha(\theta,F)$ and $\kappa_\alpha(\widehat{\theta}_u,\widehat{F}_u)$, respectively. We will use the results of Proposition 4.1 in [12] on the uniform quasi-Hadamard differentiability of the Average Value at Risk functional. Let $\lambda>1$. Note that Lemma 2.2 in [52] yields $\mathcal{C}((0,\infty)\times\mathbb{F}_{\phi_\lambda})\subset\mathbb{F}_1$, such that the composition $\mathcal{R}_\alpha\circ\mathcal{C}$ is well defined on $(0,\infty)\times\mathbb{F}_{\phi_\lambda}$.

Theorem 5.2.2 *Let $\lambda>1$, $\theta\in(0,\infty)$ and, $F\in\mathbb{F}_{\phi_{2\lambda}}$. Assume that $\mathcal{C}(\theta,F)$ takes the value $1-\alpha$ only once. Then the map $\kappa_\alpha:=\mathcal{R}_\alpha\circ\mathcal{C}:(0,\infty)\times\mathbb{F}_{\phi_\lambda}(\subset\mathrm{D})\to\mathbb{R}$ is uniformly quasi-Hadamard differentiable at (θ,F) tangentially to $((0,\infty)\times\mathrm{D}_{\phi_\lambda})\langle(0,\infty)\times\mathrm{D}_{\phi_\lambda}\rangle$, and the uniform quasi-Hadamard derivative $\dot{\kappa}_{\alpha,\theta,F}:\mathbb{R}\times\mathrm{D}_{\phi_\lambda}\to\mathbb{R}$ is given by $\dot{\kappa}_{\alpha,\theta,F}:=\dot{\mathcal{R}}_{\alpha,\mathcal{C}(\theta,F)}\circ\dot{\mathcal{C}}_{(\theta,F)}$, i.e.*

$$\dot{\kappa}_{\alpha,\theta,F}(w,v):=\int g_\alpha'(\mathcal{C}(\theta,F)(x))\,\dot{\mathcal{C}}_{(\theta,F)}(w,v)(x)\,dx, \tag{5.12}$$

for every $(w,v)\in\mathbb{R}\times\mathrm{D}_{\phi_\lambda}$, where $g_\alpha'(t):=\frac{1}{1-\alpha}\mathbb{1}_{(1-\alpha,1]}(t)$ and $\dot{\mathcal{C}}_{(\theta,F)}$ is as in (5.10).

As a direct consequence of Theorem 5.2.2 and Corollary 5.4.8 we obtain the following corollary. Note that this is a special case of 5.2.1, where the role of ρ is played by the Average Value of Risk at level $\alpha\in(0,1)$.

Corollary 5.2.3 *Let $\lambda > 1$, $\theta \in (0,\infty)$ and let $F \in \mathbb{F}_{\phi_{2\lambda}}$, that is $\int \phi_\lambda^2 \, dF < \infty$. Let $\widehat{\theta}_u$ and $\widehat{F}_u$ be as in (5.6) and (5.7), respectively. Moreover, let ξ and B_F be as in Theorem 5.2.1. Furthermore, assume that the assumptions of Theorem 5.2.2 are fulfilled. Then we have*

$$\sqrt{u}\left(\kappa_\alpha(\widehat{\theta}_u, \widehat{F}_u) - \kappa_\alpha(\theta, F)\right) \xrightarrow{\mathrm{d}} \dot{\kappa}_{\alpha,\theta,F}\Big(\xi, \frac{1}{\sqrt{\theta}} B_F\Big), \qquad \textit{in } (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

Proof of Theorem 5.2.2 We intend to apply Lemma A.5 in [12] to $H := \mathcal{C} : \Theta \times \mathbb{F}_{\phi_\lambda} \to \mathbb{F}_1$ and $\widetilde{H} := \mathcal{R}_\alpha : \mathbb{F}_1 \to \mathbb{R}$. Note that in this context H refers to the notation used in the formulation of Lemma A.5 in [12] and is not to be confused with the compound tail functional of Section 5.4.2. We will show that the following conditions are fulfilled:

(a) For every sequence $(\theta_u, F_u) \subset (0,\infty) \times \mathbb{F}_{\phi_\lambda}$ satisfying $\max\{|\theta_u - \theta|, \|F_u - F\|_{\phi_\lambda}\} \to 0$, we have
$$\lim_{u\to\infty} \mathcal{C}(\theta_u, F_u)(t) \;=\; \mathcal{C}(\theta, F)(t), \qquad \text{for every } t \in \mathbb{R}.$$

(b) $\mathcal{C}$ is uniformly quasi-Hadamard differentiable at (θ, F) tangentially to $((0,\infty) \times \mathrm{D}_{\phi_\lambda})\langle(0,\infty) \times \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_\lambda}$ and uniform quasi-Hadamard derivative $\dot{\mathcal{C}}_{(\theta,F)}$ satisfying $\dot{\mathcal{C}}_{(\theta,F)}(\mathrm{D}_{\phi_\lambda}) \subset \mathrm{D}_{\phi_{\lambda'}}$.

(c) $\mathcal{R}_\alpha$ is uniformly quasi-Hadamard differentiable tangentially to $\mathrm{D}_{\phi_{\lambda'}}\langle \mathrm{D}_{\phi_{\lambda'}}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ at every distribution function of $\mathbb{F}_{\phi_{\lambda'}}$ taking the value $1-\alpha$ only once.

To verify that the assumptions of this lemma are fulfilled, we recall from the discussion above Corollary 5.2.2 that $\mathcal{C}((0,\infty) \times \mathbb{F}_{\phi_\lambda}) \subset \mathbb{F}_1$. Conditions (a) and (b) can be proven in exactly the same way as conditions (a) in (b) in the proof of Theorem 5.2.1. It was shown in Proposition 4.1 in [12] that $\mathcal{R}_\alpha$ is uniformly quasi-Hadamard differentiable tangentially to $\mathrm{D}_{\phi_{\lambda'}}\langle \mathrm{D}_{\phi_{\lambda'}}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ at every distribution function taking the value $1-\alpha$ only once. Hence, with the help of the chain rule of Lemma A.5 in [12] we conclude that the composition $\kappa_\alpha := \mathcal{R}_\alpha \circ \mathcal{C} : (0,\infty) \times \mathbb{F}_{\phi_\lambda} \to \mathbb{R}$ is uniformly quasi-Hadamard differentiable at (θ, F) tangentially to $((0,\infty) \times \mathrm{D}_{\phi_\lambda})\langle(0,\infty) \times \mathrm{D}_{\phi_\lambda}\rangle$ and the uniform quasi-Hadamard derivative is given by $\dot{\kappa}_{\alpha,\theta,F} := \dot{\mathcal{R}}_{\alpha,\mathcal{C}(\theta,F)} \circ \dot{\mathcal{C}}_{(\theta,F)}$. □

5.3 An outlook on almost sure bootstrap consistency of the estimated premiums

Before we present the proofs of the results of the former chapter we first take our time to give an outlook about almost sure bootstrap consistency for the estimated premiums. To this end, we will first introduce a bootstrap version for the premium estimator of Section 5.1.

The main tool to derive these results will be the functional delta method for the bootstrap of uniformly quasi-Hadamard differentiable functionals of Corollary 3.2 in [12].

Let again, $N_1, \dots N_u$ be i.i.d. random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $N_1 \sim \mathrm{Poiss}_\theta$ for some $\theta \in (0, \infty)$. Moreover, let (Y_i) be a sequence of i.i.d. random variables on the same probability space with distribution function F, being independent of $(N_1, \dots, N_u)$ and let $\widehat{\theta}_u$ and $\widehat{F}_u$ be as in (5.6) and (5.7), respectively. Let $(\Omega'_1 \times \Omega'_2, \mathcal{F}'_1 \otimes \mathcal{F}'_2, \mathbb{P}'_1 \otimes \mathbb{P}'_2)$ be another probability space and set

$$(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}}) := (\Omega \times (\Omega'_1 \times \Omega'_2), \mathcal{F} \otimes (\mathcal{F}'_1 \otimes \mathcal{F}'_2), \mathbb{P} \otimes (\mathbb{P}'_1 \otimes \mathbb{P}'_2)).$$

Now let $(W_{u,i}^{(1)})$ be a triangular array of nonnegative real-valued random variables on the probability space $(\Omega'_1, \mathcal{F}'_1, \mathbb{P}'_1)$ and assume that for every $u \in \mathbb{N}$ the random vector

$$(W_{u,1}^{(1)}, \dots, W_{u,u}^{(1)})$$

is multinomially distributed according to the parameters u and $p_1 = \dots = p_u = \frac{1}{u}$. Note that this setting is nothing but Efron's bootstrap of part (i) of Example 2.4.2. Now let $\widehat{\theta}_u^{\mathrm{B}} : \Omega \times \Omega'_1 \to (0, \infty)$ be the map defined by

$$\widehat{\theta}_u^{\mathrm{B}}(\omega, \omega'_1) \;:=\; \frac{1}{u} \sum_{i=1}^{u} W_{u,i}^{(1)}(\omega'_1)\, N_i(\omega). \tag{5.13}$$

Note, that (N_i) and $(W_{u,i}^{(1)})$ regarded as families of random variables on the probability space $(\Omega \times \Omega'_1, \mathcal{F} \otimes \mathcal{F}'_1, \mathbb{P} \otimes \mathbb{P}'_1)$ are independent. Given the value of $N(u;\omega)$, let $(W_{N(u;\omega),i}^{(2)})$ be a second triangular array of nonnegative real-valued random variables on $(\Omega'_2, \mathcal{F}'_2, \mathbb{P}'_2)$ and assume that for every $u \in \mathbb{N}$ the random vector

$$\Big(W_{N(u;\omega),1}^{(2)}, \dots, W_{N(u;\omega),N(u;\omega)}^{(2)}\Big)$$

is multinomially distributed according to the parameters $N(u;\omega)$ and $p_1 = \dots = p_{N(u;\omega)} = \frac{1}{N(u;\omega)}$, if $N(u;\omega) > 0$. In this setting we may regard (Y_i) and $(W_{N(u),i}^{(2)})$ as families of independent random variables on the whole product space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$. Furthermore, the bootstrap weights $(W_{u,i}^{(1)})$ and $(W_{N(u,i}^{(2)})$ are by construction independent on the whole product space $(\overline{\Omega}, \overline{\mathcal{F}}, \overline{\mathbb{P}})$. Then for every $u \in \mathbb{N}$, let $\widehat{F}_u^{\mathrm{B}} : \Omega \times \Omega'_2 \to \mathbb{F}$ be the map defined by

$$\widehat{F}_u^{\mathrm{B}}(\omega, \omega'_2) := \frac{1}{N(u;\omega)} \sum_{i=1}^{N(u;\omega)} W_{N(u;\omega),i}^{(2)}(\omega'_2)\, \mathbb{1}_{[Y_i(\omega), \infty)}, \tag{5.14}$$

if $N(u;\omega) > 0$ and $\widehat{F}_u^{\mathrm{B}}(\omega, \omega'_2) := \mathbb{1}_{[0,\infty)}$ else. Here we will tactically assume that the probability spaces are rich enough to host all the random variables used above.

With the help of the bootstrap versions of our original estimators, we can now define a bootstrap estimator for the premium $\mathcal{R}$. To this end, we use $\mathcal{C}(\widehat{\theta}_u^{\mathrm{B}}, \widehat{F}_u^{\mathrm{B}})$ as a bootstrap estimator for the total claim amount, and let $\widehat{\mathcal{R}}_u^{\mathrm{B}} : \overline{\Omega} \to \mathbb{R}$ be the mapping defined by

$$\widehat{\mathcal{R}}_u^{\mathrm{B}} := \mathcal{R}_\rho(\mathcal{C}(\widehat{\theta}_u^{\mathrm{B}}, \widehat{F}_u^{\mathrm{B}})). \tag{5.15}$$

As mentioned before, we aim to apply Corollary 3.2 in [12] to the sequences of underlying estimators $(\widehat{\theta}_u^{\mathrm{B}})$ and $\widehat{F}_u^{\mathrm{B}}$. More explicitly, if we could show that

$$\sqrt{u}\left(\begin{bmatrix}\widehat{\theta}_u^{\mathrm{B}}(\omega,\,\cdot\,)\\ \widehat{F}_u(\omega,\,\cdot\,)\end{bmatrix}-\begin{bmatrix}\widehat{\theta}_u(\omega)\\ \widehat{F}_u(\omega)\end{bmatrix}\right)\overset{\mathrm{d}}{\longrightarrow}\begin{bmatrix}\xi\\ \frac{1}{\sqrt{\theta}}B_F\end{bmatrix},\qquad \mathbb{P}\text{-a.e. }\omega \tag{5.16}$$

in $(\mathbb{R}\times\mathrm{D}_{\phi_\lambda},\mathcal{B}(\mathbb{R})\otimes\mathcal{D}_{\phi_\lambda},\max\{|\cdot|,\|\cdot\|_{\phi_\lambda}\})$, then we could use the functional delta-method for the bootstrap of uniformly quasi-Hadamard differentiable functionals in the form of Corollary 3.2 in [12] to derive almost sure bootstrap consistency for the sequence of estimated premiums. In this case one could show that under the assumptions of Theorem 5.2.1, the following assertion holds true

$$\sqrt{u}\left(\widehat{\mathcal{R}}_u^{\mathrm{B}}(\omega,\,\cdot\,)-\widehat{\mathcal{R}}_u(\omega)\right)\overset{\mathrm{d}}{\longrightarrow}\dot{\mathcal{R}}_{\rho,\mathcal{C}(\theta,F)}\circ\dot{\mathcal{C}}_{(\theta,F)}\Big(\xi,\frac{1}{\sqrt{\theta}}B_F\Big),\qquad \mathbb{P}\text{-a.e. }\omega,$$

in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, where $\dot{\mathcal{C}}_{(\theta,F)}:(0,\infty)\times\mathrm{D}_{\phi_\lambda}\to\mathrm{D}_{\phi_{\lambda'}}$ is as in Theorem 5.2.1. Of course one could obtain similar results under the assumptions of Corollary 5.2.3 for the special case when the role of $\mathcal{R}_\rho$ is played by the Average Value at Risk at level $\alpha\in(0,1)$.

5.4 Proofs

5.4.1 A functional delta-method for plug-in estimators of uniformly quasi-Hadamard differentiable statistical functionals

This section gives a brief summary of the techniques introduced in [12]. We will then concretize the delta-method and the bootstrap results from [12] to fit our present setting.

Based on the notion of differentiability in Definition 4.2.1 (as introduced in [12]), we now turn to a functional delta-method for plug-in estimators of statistical functionals. Due to this, we have to introduce some further notation. In accordance with the model discussed at the beginning of this chapter, we will consider a "two-dimensional" set of estimators to estimate the underlying parameter of the distribution of the number of claims and the distribution of the single claims. The following notation is an extension of the one used in Section 3 in [12].

To this end, let $d\in\mathbb{N}$ and let $\Theta\subset\mathbb{R}^d$ be an open set and write $\|\cdot\|$ for the euclidean metric on $\mathbb{R}^d$. Let D be the space of all càdlàg functions v on $\mathbb{R}$ with finite sup-norm $\|v\|_\infty:=\sup_{t\in\mathbb{R}}|v(t)|$. Let $\mathcal{D}$ be the σ-algebra on D generated by the one-dimensional coordinate projections $\pi_t(v):=v(t)$, $t\in\mathbb{R}$. Let $\phi:\mathbb{R}\to[1,\infty)$ be a weight-function. Here we refer to a weight-function as a continuous function, being non-increasing on $(-\infty,0]$ and non-decreasing on $[0,\infty)$. Let D_ϕ be a subspace of D, consisting of all elements $x\in\mathrm{D}$ satisfying $\|x\|_\phi:=\|x\phi\|_\infty<\infty$ and $\lim_{|t|\to\infty}|x(t)|=0$. The latter condition is automatically satisfied whenever $\lim_{|t|\to\infty}\phi(t)=\infty$. Let $\mathcal{D}_\phi:=\mathcal{D}\cap\mathrm{D}_\phi$ be the trace σ-algebra on D_ϕ. Furthermore we will write $\mathcal{B}_\phi^\circ$ for the σ-algebra on D_ϕ, generated by the $\|\cdot\|_\phi$-open balls.

It was shown in Lemma 4.1 in [11] that $\mathcal{B}_\phi^\circ$ coincides with $\mathcal{D}_\phi$. In the following we will write $\rightsquigarrow^\circ$ for the convergence in distribution w.r.t. an open-ball σ-algebra, such as $\mathcal{B}_\phi^\circ$.

Let C_ϕ be a $\|\cdot\|_\phi$-separable subspace of D_ϕ and assume that $\mathrm{C}_\phi \in \mathcal{D}_\phi$. Note that for any distribution function F the set C_ϕ can be chosen to be the set $\mathrm{C}_{\phi,F}$ of all $v \in \mathrm{D}_\phi$ whose discontinuities are also discontinuities of F. The separability of $C_{\phi,F}$ was shown in Corollary B.4 in [38].

We now equip the product space $\mathbb{R}^d \times \mathrm{D}_\phi$ with the metric $\widetilde{d}$, defined by $\widetilde{d}((x_1,x_2),(y_1,y_2)) := \max\{\|x_1-y_1\|, \|x_2-y_2\|_\phi\}$ for every $(x_1,x_2),(y_1,y_2) \in \mathbb{R}^d \times \mathrm{D}_\phi$. We equip the product space $\mathbb{R}^d \times \mathrm{D}_\phi$ with the σ-algebra $\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{D}_\phi$. Furthermore, let

$$G : D(G) \longrightarrow \widetilde{\mathbf{V}}$$

be a map defined on a set $D(G)$, such that the domain $D(G)$ is a product space of $\mathbb{R}^d$ with a set of distribution functions of finite (not necessarily probability) Borel measures on $\mathbb{R}$. That is, $D(G) \subset \mathbb{R}^d \times \mathrm{D}$. Moreover, let $\widetilde{\mathbf{V}}$ be any vector space.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let (N_i) be a sequence of i.i.d. integer-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\Theta \subset \mathbb{R}^d$. Furthermore, assume that the distribution of N_1 belongs to a certain class of distributions indexed by a parameter $\vartheta \in \Theta$. More explicitly, we can characterize the distribution of the count variables (N_i) by a sequence $p(\vartheta) := (p_k(\vartheta))_{k\in\mathbb{N}_0}$, satisfying

$$\sum_{k=0}^{\infty} p_k(\vartheta) \;=\; 1, \qquad \text{for every } \vartheta \in \Theta.$$

Let $\widehat{\vartheta}_u : \Omega \to \Theta$ be an estimator for the unknown parameter ϑ. Let (Y_i) be another sequence of i.i.d. random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ being independent of (N_i). Moreover, let $\widehat{F}_u : \Omega \to \mathrm{D}$ be the empirical distribution function based on a sample of size $N_1 + \ldots + N_u$ and observations $Y_1, \ldots, Y_{N_1+\ldots+N_u}$. In this context, the empirical distribution function is given by

$$\widehat{F}_u \;:=\; \frac{1}{N_1 + \ldots + N_u} \sum_{i=1}^{N_1+\ldots+N_u} \mathbb{1}_{[Y_i,\infty)}. \tag{5.17}$$

Assume that $[\widehat{\vartheta}_u, \widehat{F}_u]'$ takes values only in $D(G)$. We are now able to formulate a delta-method for the upper setting.

Theorem 5.4.1 *Let $([\vartheta_u, F_u]')$ be a sequence in $D(G)$ and $\mathcal{S} := \{([\vartheta_u, F_u]')\}$. Let (a_u) be a sequence of positive real numbers with $a_u \to \infty$, and assume that the following assertions hold:*

(a) $a_u\,([\widehat{\vartheta}_u, \widehat{F}_u]' - [\vartheta_u, F_u]')$ takes values only in $\mathbb{R}^d \times \mathrm{D}_\phi$ and satisfies

$$a_u \left(\begin{bmatrix} \widehat{\vartheta}_u \\ \widehat{F}_u \end{bmatrix} - \begin{bmatrix} \vartheta_u \\ F_u \end{bmatrix} \right) \rightsquigarrow^\circ \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tag{5.18}$$

in $(\mathbb{R}^d \times \mathrm{D}_\phi, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}^\circ_\phi, \max\{\|\cdot\|, \|\cdot\|_\phi\})$ *for some* $(\mathbb{R}^d \times D_\phi, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}^\circ_\phi)$*-valued random variable* $[B_1, B_2]'$ *on some probability space* $(\Omega_0, \mathcal{F}_0, \mathbb{P}_0)$ *with* $[B_1, B_2]'(\Omega_0) \subset \mathbb{R}^d \times \mathrm{C}_\phi$.

(b) $a_u\,(G(\widehat{\vartheta}_u, \widehat{F}_u) - G(\vartheta_u, F_u))$ *takes values only in* $\widetilde{\mathbf{E}}$ *and is* $(\mathcal{F}, \widetilde{\mathcal{B}}^\circ)$*-measurable.*

(c) G *is uniformly quasi-Hadamard differentiable w.r.t.* $\mathcal{S}$ *tangentially to* $(\Theta \times \mathrm{C}_\phi)\langle\Theta \times \mathrm{D}_\phi\rangle$ *with trace* $\widetilde{\mathbf{E}}$ *and uniform quasi-Hadamard derivative* $\dot{G}_{\mathcal{S}}$.

Then $\dot{G}_{\mathcal{S}}(B_1, B_2)$ *is* $(\mathcal{F}_0, \widetilde{\mathcal{B}}^\circ)$*-measurable and*

$$a_u\,(G(\widehat{\vartheta}_u, \widehat{F}_u) - G(\vartheta_u, F_u)) \rightsquigarrow^\circ \dot{G}_{\mathcal{S}}(B_1, B_2) \quad in\ (\widetilde{\mathbf{E}}, \widetilde{\mathcal{B}}^\circ, \|\cdot\|_{\widetilde{\mathbf{E}}}). \tag{5.19}$$

The upper Theorem is a special case of Corollary 3.1 in [12] and does therefore not need to be proved.

5.4.2 On the uniform quasi-Hadamard differentiability of the compound distribution functional

In this section we are going to derive the uniform quasi-Hadamard derivative of the compound distribution functional. For the case of the compound Poisson model, the corresponding compound distribution functional was introduced in (5.2). However, in this section we are going to differentiate the compound distribution functional w.r.t. a more general class of distributions of count variables. The Poisson case will then be a special case.

Let $\mathbb{F}$ denote the set of all distribution functions on $\mathbb{R}$. Let $d \in \mathbb{N}$ and $\Theta \subset \mathbb{R}^d$ be an open set. Write $\|\cdot\|$ for the euclidean metric on $\mathbb{R}^d$. Let $\mathcal{C} : \Theta \times \mathbb{F} \to \mathbb{F}$ be a mapping with

$$\mathcal{C}(\theta, F) := \sum_{k=0}^{\infty} F^{*k}\, p_k(\theta), \tag{5.20}$$

where $p_k : \Theta \to \mathbb{R}_+$ is a continuously differentiable map for every $k \in \mathbb{N}_0$, and

$$\sum_{k \in \mathbb{N}_0} p_k(\theta) \;=\; 1 \quad \text{for every } \theta \in \Theta.$$

Note that in this case $(p_k(\theta))_{k\in\mathbb{N}_0}$ specifies the distribution of an integer-valued random variable in dependence on the parameter $\theta \in \Theta$. If specifically $\Theta = (0, \infty)$ and $p_k(\theta) = e^{-\theta}\,\theta^k/k!$ for every $k \in \mathbb{N}_0$ and $\theta \in (0, \infty)$, we observe that $\mathcal{C}(\theta, F)$ as introduced in (5.20) is the compound distribution function w.r.t. the Poisson distribution of formula (5.2). Moreover, the setting of the individual model of Chapter 4 can be obtained by choosing $p_n = 1$ and $p_k = 0$ for every $k \neq n$. Note that in this case the distribution of the count variable is not dependent on a parameter.

The following theorem states the uniform quasi-Hadamard derivative of $\mathcal{C}$. To this end, for every $k \in \mathbb{N}_0$, write ∇p_k for the gradient of p_k, that is

$$\nabla p_k := \left[\frac{\partial p_k}{\partial \theta_1}, \ldots, \frac{\partial p_k}{\partial \theta_d} \right]'.$$

For any $\lambda \geq 0$, let again $\phi_\lambda : \mathbb{R} \to [1, \infty)$ be defined by $\phi_\lambda(x) := (1 + |x|)^\lambda$ and let

$$\mathbb{F}_{\phi_\lambda} := \Big\{ F \in \mathbb{F} : \int \phi_\lambda \, dF < \infty \Big\}. \tag{5.21}$$

To be able to prove the assertion we will need two auxiliary results, which will be proved first.

Theorem 5.4.2 *Let $\lambda > \lambda' \geq 0$, $\theta \in (0, \infty)$ and $F \in \mathbb{F}_{\phi_\lambda}$. Assume that for every $k \in \mathbb{N}_0$, p_k is twofold differentiable and that $\sum_{k \in \mathbb{N}} k^{(1+\lambda)\vee 2} p_k(\theta) < \infty$ and assume that there exists an $r \in (0, \infty)$ such that*

$$\sum_{k \in \mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\theta' \in B_r(\theta)} \|\nabla p_k(\theta')\| < \infty \quad \text{and} \tag{5.22}$$

$$\sum_{k \in \mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\theta' \in B_r(\theta)} \Big(\sum_{i,j=1}^{d} \Big| \frac{\partial^2 p_k}{\partial \theta_i \partial \theta_j}(\theta') \Big|^2 \Big)^{1/2} < \infty, \tag{5.23}$$

where $B_r(\theta)$ denotes the open ball with radius r around θ. Then the map $\mathcal{C} : \Theta \times \mathbb{F}_{\phi_\lambda} \to \mathbb{F}$ is uniformly quasi-Hadamard differentiable in (θ, F) tangentially to $(\Theta \times \mathbf{D}_{\phi_\lambda})\langle \Theta \times \mathbf{D}_{\phi_\lambda} \rangle$ with trace $\mathbf{D}_{\phi_{\lambda'}}$. Moreover the uniform quasi-Hadamard derivative $\dot{\mathcal{C}}_{(\theta,F)} : \mathbb{R}^d \times \mathbf{D}_{\phi_\lambda} \to \mathbf{D}_{\phi_{\lambda'}}$ is given by

$$\dot{\mathcal{C}}_{(\theta,F)}(w, v) := v * \sum_{k=1}^{\infty} k \, F^{*(k-1)} \, p_k(\theta) \; + \sum_{k=0}^{\infty} F^{*k} \, \langle w, \nabla p_k(\theta) \rangle, \tag{5.24}$$

where

$$\Big(v * \sum_{k=1}^{\infty} k \, F^{*(k-1)} \, p_k(\theta) \Big)(\cdot) := \int v(\cdot - x) \, d\Big(\sum_{k=1}^{\infty} k \, F^{*(k-1)} \, p_k(\theta) \Big)(x). \tag{5.25}$$

Note that we had to restrict $\mathcal{C}$ to the set $\Theta \times \mathbb{F}_{\phi_\lambda}$ instead of $\Theta \times \mathbb{F}$, to obtain $\mathbf{D}_{\phi_{\lambda'}}$ as the trace. Furthermore, note that if we choose $p_n = 1$ for some $n \in \mathbb{N}$ and $p_k = 0$ for every $k \neq n$, then the uniform quasi-Hadamard derivative $\dot{\mathcal{C}}_{(\theta,F)}$ in Theorem 5.4.3 is nothing but the uniform quasi-Hadamard derivative in the individual model with n clients as in (4.10).

Furthermore, note that if $\Theta := \{\theta\}$ for some $\theta \in \mathbb{R}^d$, then the assertion of Theorem 5.4.2 boils down to the assertion of Proposition 4.3 in [12]. Theorem 5.4.3 therefore provides an extension to the existing theory.

To be able to derive the uniform quasi-Hadamard derivate of $\mathcal{C}$ in the sense of Definition 4.2.1 and prove Theorem 5.4.2, we will write $\mathcal{C}$ as a composition of two auxiliary mappings. The use of this approach will become apparent later. To this end, let $\mathcal{T} := \{\mathbb{1}_{[0,\infty)} - F : F \in \mathbb{F}\}$ denote the set of all (two-sided) tail functions associated with elements of $\mathbb{F}$ and let $H : \Theta \times \mathbb{F} \to \mathcal{T}$ be a mapping, defined by

$$H(\theta, F) := \mathbb{1}_{[0,\infty)} - \sum_{k=0}^{\infty} F^{*k}\, p_k(\theta). \tag{5.26}$$

In the following we will refer to H as the (two-sided) compound tail functional. Note that H is not a tail functional in the classical sense, but coincides with the tail functional associated with the compound distribution function on the nonnegative semiaxis and equals $-\mathcal{C}$ on the negative semiaxis. This approach has been introduced in [53].

Furthermore, let $\Lambda : \mathcal{T} \to \mathbb{F}$ be a second mapping, defined by

$$\Lambda(T) := \mathbb{1}_{[0,\infty)} - T. \tag{5.27}$$

Note that Λ maps a (two-sided) tail function onto the corresponding distribution function. Then we observe that

$$\mathcal{C} = \Lambda \circ H. \tag{5.28}$$

In the next steps we will determine the uniform quasi-Hadamard derivatives of both H and Λ. To obtain the uniform quasi-Hadamard derivative of $\mathcal{C}$ we will apply the chain rule for uniform quasi-Hadamard differentiable maps in the form of Lemma A.5 in [12]. Theorem 5.4.3 will state the uniform quasi-Hadamard derivative of H.

Then we stress the fact again, that for every $F \in \mathbb{F}$ and $\theta \in \Theta$, we have $H(\theta, F) \in \mathcal{T}$, because H is the tail function associated with the random convolution of F w.r.t. the random measure characterized by $(p_k(\theta))_{k\in\mathbb{N}_0}$.

Lemma 5.4.3 *Let $\lambda > \lambda' \geq 0$, $\theta \in (0,\infty)$ and $F \in \mathbb{F}_{\phi_\lambda}$. Assume that for every $k \in \mathbb{N}_0$, p_k is twofold differentiable and that $\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2}\, p_k(\theta) < \infty$ and assume that there exists an $r \in (0,\infty)$ such that*

$$\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\theta' \in B_r(\theta)} \|\nabla p_k(\theta')\| < \infty \quad \text{and} \tag{5.29}$$

$$\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\theta' \in B_r(\theta)} \Big(\sum_{i,j=1}^{d} \Big| \frac{\partial^2 p_k}{\partial\theta_i \partial\theta_j}(\theta') \Big|^2 \Big)^{1/2} < \infty, \tag{5.30}$$

where $B_r(\theta)$ denotes the open ball with radius r around θ. Then the map $H : \Theta \times \mathbb{F}_{\phi_\lambda} \to \mathcal{T}$ is uniformly quasi-Hadamard differentiable at (θ, F) tangentially to $(\Theta \times \mathrm{C}_{\phi_\lambda})\langle \Theta \times \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$. Moreover the uniform quasi-Hadamard derivative $\dot{H}_{(\theta,F)} : \mathbb{R}^d \times \mathrm{C}_{\phi_\lambda} \to \mathrm{D}_{\phi_{\lambda'}}$ is given by

$$\dot{H}_{(\theta,F)}(w,v) := -v * \sum_{k=1}^{\infty} k\, F^{*(k-1)}\, p_k(\theta) \; + \; \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k})\, \langle w, \nabla p_k(\theta)\rangle. \tag{5.31}$$

Proof First we note that for $\theta_1, \theta_2 \in \Theta$ and $F_1, F_2 \in \mathbb{F}_{\phi_\lambda}$ we have

$$\begin{aligned}\|H(\theta_1,F_1)-H(\theta_2,F_2)\|_{\phi_{\lambda'}} &\le \|H(\theta_1,F_1)\|_{\phi_{\lambda'}}+\|H(\theta_2,F_2)\|_{\phi_{\lambda'}}\\ &\le \int \phi_{\lambda'}(x)\,d\Big(\sum_{k=0}^{\infty} F_1^{*k}\,p_k(\theta_1)\Big)(x)\\ &\quad+\int \phi_{\lambda'}(x)\,d\Big(\sum_{k=0}^{\infty} F_2^{*k}\,p_k(\theta_2)\Big)(x), \qquad (5.32)\end{aligned}$$

by Equation (2.1) in [52]. According to Lemma 2.2 in [52] we can conclude that both integrals on the right-hand side of (5.32) are finite under the assumptions of the lemma. Hence, the set $\mathrm{D}_{\phi_{\lambda'}}$ can be seen as the trace.

Second, we have to show that $\dot{H}_{(\theta,F)}$ is $(\max\{\|\cdot\|,\|\cdot\|_{\phi_\lambda}\},\|\cdot\|_{\phi_{\lambda'}})$-continuous and that

$$\lim_{u\to\infty}\Big\|\frac{H(\theta_u+\varepsilon_u\,w_u,F_u+\varepsilon_u\,v_u)-H(\theta_u,F_u)}{\varepsilon_u}-\dot{H}_{(\theta,F)}(w,v)\Big\|_{\phi_{\lambda'}}=0 \qquad (5.33)$$

holds for every quadrupel $((\theta_u,F_u),(w,v),(w_u,v_u),(\varepsilon_u))$, with $(\theta_u,F_u)\in\Theta\times\mathbb{F}_{\phi_\lambda}$ satisfying $\max\{\|\theta_u-\theta\|,\|F_u-F\|_{\phi_\lambda}\}\to 0$, $(w,v)\in\Theta\times\mathrm{D}_{\phi_\lambda}$, $(w_u,v_u)\subset\Theta\times\mathrm{D}_{\phi_\lambda}$, satisfying $\max\{\|w_u-w\|,\|v_u-v\|_{\phi_\lambda}\}\to 0$ and $(\theta_u+\varepsilon_u\,w_u,F_u+\varepsilon_u\,v_u)\subset\Theta\times\mathbb{F}_{\phi_\lambda}$, and $(\varepsilon_u)\subset(0,\infty)$ satisfying $\varepsilon_u\to 0$.

Note that for every $u\in\mathbb{N}$ we have

$$\begin{aligned}&\frac{H(\theta_u+\varepsilon_u\,w_u,F_u+\varepsilon_u\,v_u)-H(\theta_u,F_u)}{\varepsilon_u}\\ &=\frac{1}{\varepsilon_u}\sum_{k=0}^{\infty}(\mathbb{1}_{[0,\infty)}-(F_u+\varepsilon_u v_u)^{*k})\,p_k(\theta_u+\varepsilon_u w_u)-\frac{1}{\varepsilon_u}\sum_{k=0}^{\infty}(\mathbb{1}_{[0,\infty)}-F_u^{*k})\,p_k(\theta_u)\\ &=\frac{1}{\varepsilon_u}\sum_{k=0}^{\infty}(F_u^{*k}-(F_u+\varepsilon_u v_u)^{*k})\,p_k(\theta_u)+\sum_{k=0}^{\infty}(F_u^{*k}-(F_u+\varepsilon_u v_u)^{*k})\frac{p_k(\theta_u+\varepsilon_u w_u)-p_k(\theta_u)}{\varepsilon_u}\\ &\quad+\sum_{k=0}^{\infty}(\mathbb{1}_{[0,\infty)}-F_u^{*k})\,\frac{p_k(\theta_u+\varepsilon_u w_u)-p_k(\theta_u)}{\varepsilon_u}. \qquad (5.34)\end{aligned}$$

Hence

$$\begin{aligned}&\Big\|\frac{H(\theta_u+\varepsilon_u\,w_u,F_u+\varepsilon_u\,v_u)-H(\theta_u,F_u)}{\varepsilon_u}-\dot{H}_{(\theta,F)}(w,v)\Big\|_{\phi_{\lambda'}}\\ &\le \Big\|\frac{1}{\varepsilon_u}\sum_{k=0}^{\infty}(F_u^{*k}-(F_u+\varepsilon_u v_u)^{*k})\,p_k(\theta_u)-\Big(-v*\sum_{k=1}^{\infty}k\,F^{*(k-1)}\,p_k(\theta)\Big)\Big\|_{\phi_{\lambda'}}\\ &\quad+\Big\|\sum_{k=0}^{\infty}\frac{1}{\varepsilon_u}(F_u^{*k}-(F_u+\varepsilon_u v_u)^{*k})\,(p_k(\theta_u+\varepsilon_u w_u)-p_k(\theta_u))\Big\|_{\phi_{\lambda'}}\\ &\quad+\Big\|\sum_{k=0}^{\infty}(\mathbb{1}_{[0,\infty)}-F_u^{*k})\,\frac{p_k(\theta_u+\varepsilon_u w_u)-p_k(\theta_u)}{\varepsilon_u}-\sum_{k=0}^{\infty}(\mathbb{1}_{[0,\infty)}-F^{*k})\,\langle w,\nabla p_k(\theta)\rangle\Big\|_{\phi_{\lambda'}}\\ &=:\; S_1(u)+S_2(u)+S_3(u). \qquad (5.35)\end{aligned}$$

In the following we will show, that $S_1(u)$–$S_3(u)$ converge to zero for $u \to \infty$. For the convergence of $S_1(u)$ we observe that

$$
\begin{aligned}
&\left\| \frac{1}{\varepsilon_u} \sum_{k=0}^{\infty} (F_u^{*k} - (F_u + \varepsilon_u v_u)^{*k})\, p_k(\theta_u) + v * \sum_{k=1}^{\infty} k\, F^{*(k-1)}\, p_k(\theta) \right\|_{\phi_{\lambda'}} \\
&\le \left\| \frac{1}{\varepsilon_u} \sum_{k=0}^{\infty} ((F_u + \varepsilon_u v_u)^{*k} - F_u^{*k})\, p_k(\theta) - v * \sum_{k=1}^{\infty} k\, F^{*(k-1)}\, p_k(\theta) \right\|_{\phi_{\lambda'}} \\
&\quad + \left\| \sum_{k=0}^{\infty} \frac{1}{\varepsilon_u} (F_u^{*k} - (F_u + \varepsilon_u v_u)^{*k})\, (p_k(\theta_u) - p_k(\theta)) \right\|_{\phi_{\lambda'}} . \qquad (5.36)
\end{aligned}
$$

Now the first summand on the right-hand side of (5.36) converges to 0 by Proposition 4.3 in [12]. For the convergence of the second summand on the right-hand side of (5.36), let for every $k \in \mathbb{N}_0$ $H_k : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ be a map defined by

$$
H_k(G_1, G_2) \;:=\; \sum_{j=0}^{k-1} G_1^{*(k-1-j)} * G_2^{*j}, \qquad (5.37)
$$

with the usual convention that the sum over the empty set equals zero. Then simple algebra yields

$$
(G_1 - G_2) * H_k(G_1, G_2) \;=\; G_1^{*k} - G_2^{*k} \qquad (5.38)
$$

for every $G_1, G_2 \in \mathbb{F}$ and every $k \in \mathbb{N}_0$. Thus, we conclude that

$$
\begin{aligned}
&\left\| \sum_{k=0}^{\infty} \frac{1}{\varepsilon_u} (F_u^{*k} - (F_u + \varepsilon_u v_u)^{*k})\, (p_k(\theta_u) - p_k(\theta)) \right\|_{\phi_{\lambda'}} \\
&= \left\| \sum_{k=0}^{\infty} \frac{1}{\varepsilon_u} (F_u + \varepsilon_u v_u - F_u) * H_k(F_u + \varepsilon_u v_u, F_u)\, (p_k(\theta_u) - p_k(\theta)) \right\|_{\phi_{\lambda'}} \\
&= \left\| \sum_{k=0}^{\infty} v_u * H_k(F_u + \varepsilon_u v_u, F_u)\, (p_k(\theta_u) - p_k(\theta)) \right\|_{\phi_{\lambda'}} \\
&\le \|v_u\|_{\phi_{\lambda'}} \sum_{k=0}^{\infty} |p_k(\theta_u) - p_k(\theta)|\, 2^{\lambda'} k \left(1 + 2^{\lambda'}(2^{\lambda'} \vee 1)(2 + (k-1)^{\lambda' \vee 1} C_1)\right), \qquad (5.39)
\end{aligned}
$$

for some $C_1 > 0$. The last inequality in a consequence of part (ii) of Lemma 4.5 in [12]. The lemma can be applied because $\|F_u + \varepsilon_u v_u - F\|_{\phi_\lambda} \to 0$, by

$$
\begin{aligned}
\|F_u + \varepsilon_u v_u - F\|_{\phi_\lambda} &\le \|F_u + \varepsilon_u v_u - F_u\|_{\phi_\lambda} + \|F_u - F\|_{\phi_\lambda} \\
&= \varepsilon_u \|v_u\|_{\phi_\lambda} + \|F_u - F\|_{\phi_\lambda}. \qquad (5.40)
\end{aligned}
$$

Now we assumed that $\|F_u - F\|_{\phi_\lambda} \to 0$ and $\varepsilon_u \to 0$ for $u \to \infty$. Since $\|v_u - v\|_{\phi_\lambda} \to 0$, we observe that $\|v_u\|_{\phi_\lambda}$ is finite for every $u \in \mathbb{N}$. Thus, we conclude that $\|F_u + \varepsilon_u v_u - F\|_{\phi_\lambda}$

indeed converges to zero and Lemma 4.5 in [12] is applicable. By $\lambda' < \lambda$ and $\|v_u - v\|_{\phi_\lambda} \to 0$, we observe that $\|v_u - v\|_{\phi_{\lambda'}} \to 0$, such that $\|v_u\|_{\phi_{\lambda'}}$ is also finite. Using the finiteness of $\|v_u\|_{\phi_{\lambda'}}$, it suffices to show that the sum on the right-hand side of (5.39) converges to zero. Applying the Mean Value theorem to $|p_k(\theta_u) - p_k(\theta)|$, ones finds that there exists some $h_{k,u} \in (0,1)$, such that

$$\begin{aligned} |p_k(\theta_u) - p_k(\theta)| &= |\langle \theta_u - \theta, \nabla p_k(\theta - h_{k,u}(\theta_u - \theta)) \rangle| \\ &\leq \|\theta_u - \theta\| \, \|\nabla p_k(\theta - h_{k,u}(\theta_u - \theta))\| \end{aligned} \tag{5.41}$$

for every $k \in \mathbb{N}_0$. Hence

$$\begin{aligned} &\sum_{k=0}^{\infty} |p_k(\theta_u) - p_k(\theta)| \, 2^{\lambda' k} \left(1 + 2^{\lambda'}(2^{\lambda'} \vee 1)(2 + (k-1)^{\lambda' \vee 1} C_1)\right) \\ &\leq \|\theta_u - \theta\| \sum_{k=0}^{\infty} \|\nabla p_k(\theta - h_{k,u}(\theta_u - \theta))\| \, 2^{\lambda' k} \left(1 + 2^{\lambda'}(2^{\lambda'} \vee 1)(2 + (k-1)^{\lambda' \vee 1} C_1)\right), \end{aligned} \tag{5.42}$$

where the sum converges for every $u \in \mathbb{N}$ sufficiently large, such that $\theta - h_{k,u}(\theta_u - \theta) \in B_r(\theta)$ for every $k \in \mathbb{N}_0$. The latter is due to the assumption (5.29). Thus, $S_1(u)$ converges to zero.

Next, we consider the second summand on the right-hand side of (5.35). Using the same arguments as in (5.39), we arrive at

$$\begin{aligned} &\left\| \sum_{k=0}^{\infty} \frac{1}{\varepsilon_u} (F_u^{*k} - (F_u + \varepsilon_u v_u)^{*k}) \, (p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)) \right\|_{\phi_{\lambda'}} \\ &\leq \|v_u\|_{\phi_{\lambda'}} \sum_{k=0}^{\infty} |p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)| \, 2^{\lambda'} \left(1 + 2^{\lambda'}(2^{\lambda'} \vee 1)(2 + (k-2)^{\lambda' \vee 1} C_1)\right). \end{aligned} \tag{5.43}$$

By

$$\|\theta_u + \varepsilon_u w_u - \theta_u\| \leq \varepsilon_u \, \|w\| \, + \, \varepsilon_u \|w_u - w\| \tag{5.44}$$

we conclude that $\|\theta_u + \varepsilon_u w_u - \theta_u\| \to 0$ as $u \to \infty$. Thus, the convergence of the sum on the right-hand side of (5.43) follows again by an application of the Mean Value theorem. Thus, $S_2(u)$ converges to zero.

It remains to be shown that

$$\lim_{u \to \infty} \left\| \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F_u^{*k}) \, \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \, \langle w, \nabla p_k(\theta) \rangle \right\|_{\phi_{\lambda'}} = 0, \tag{5.45}$$

First, we observe that

$$\left\| \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F_u^{*k}) \, \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \, \langle w, \nabla p_k(\theta) \rangle \right\|_{\phi_{\lambda'}}$$

$$
\begin{aligned}
&\leq \left\| \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \langle w, \nabla p_k(\theta) \rangle \right\|_{\phi_{\lambda'}} \\
&\quad + \left\| \sum_{k=0}^{\infty} (F^{*k} - F_u^{*k}) \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} \right\|_{\phi_{\lambda'}} \\
&= \left\| \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \left(\frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \langle w, \nabla p_k(\theta) \rangle \right) \right\|_{\phi_{\lambda'}} \\
&\quad + \left\| \sum_{k=0}^{\infty} (F^{*k} - F_u^{*k}) \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} \right\|_{\phi_{\lambda'}} . \qquad (5.46)
\end{aligned}
$$

For the first summand on the right-hand side of (5.46) we observe that

$$
\begin{aligned}
&\left\| \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \left(\frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \langle w, \nabla p_k(\theta) \rangle \right) \right\|_{\phi_{\lambda'}} \\
&\quad \leq \sum_{k=0}^{\infty} \left\| \mathbb{1}_{[0,\infty)} - F^{*k} \right\|_{\phi_{\lambda'}} \left| \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \langle w, \nabla p_k(\theta) \rangle \right| . \qquad (5.47)
\end{aligned}
$$

According to Equation (2.4) of [52], we have

$$
\left\| \mathbb{1}_{[0,\infty)} - F^{*k} \right\|_{\phi_{\lambda'}} \leq (2^{\lambda'-1} \vee 1) \Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} dF(x) \Big) \qquad (5.48)
$$

for every $k \in \mathbb{N}_0$. Thus, $\|\mathbb{1}_{[0,\infty)} - F^{*k}\|_{\phi_{\lambda'}}$ is bounded above. By the Mean Value Theorem, we furthermore conclude that for every $k \in \mathbb{N}_0$ and every $u \in \mathbb{N}$ sufficiently large, there exists some $h'_{k,u} \in (0,1)$, such that

$$
\begin{aligned}
&\left| \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \langle w, \nabla p_k(\theta) \rangle \right| \\
&\quad = \left| \frac{1}{\varepsilon_u} \langle \theta_u + \varepsilon_u w_u - \theta_u, \nabla p_k(\theta_u + h'_{k,u} (\theta_u + \varepsilon_u w_u - \theta_u)) \rangle - \langle w, \nabla p_k(\theta) \rangle \right| \\
&\quad = \left| \langle w_u, \nabla p_k(\theta_u + h'_{k,u} \varepsilon_u w_u) \rangle - \langle w, \nabla p_k(\theta) \rangle \right| \\
&\quad \leq \left| \langle w_u - w, \nabla p_k(\theta_u + h'_{k,u} \varepsilon_u w_u) \rangle \right| + \left| \langle w, \nabla p_k(\theta_u + h'_{k,u} \varepsilon_u w_u) - \nabla p_k(\theta) \rangle \right| \\
&\quad \leq \|w_u - w\| \, \|\nabla p_k(\theta_u + h_k \varepsilon_u w_u)\| + \|w\| \, \|\nabla p_k(\theta_u + h'_{k,u} \varepsilon_u w_u) - \nabla p_k(\theta)\|, \quad (5.49)
\end{aligned}
$$

where the last inequality holds true due to the Cauchy-Schwarz inequality. Note that we have to choose $u \in \mathbb{N}$ sufficiently large to be able to apply the Mean Value Theorem. In particular, we have to make sure, that $\{\theta_u + t\,\varepsilon_u w_u : t \in [0,1]\} \subset \Theta$, which is always possible with our choices of θ_u, w_u and ε_u, because $\Theta \subset \mathbb{R}^d$ was assumed to be open. Now we observe that

$$
\|w_u - w\| \sum_{k=0}^{\infty} (2^{\lambda'-1} \vee 1) \Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} dF(x) \Big) \|\nabla p_k(\theta_u + h_k \varepsilon_u w_u)\| \qquad (5.50)
$$

converges to zero for $u \to \infty$ because $\|w_u - w\| \to 0$ and the sum in (5.50) is finite due to assumption (5.29). Second, we can apply the Mean Value theorem to $\|\nabla p_k(\theta_u + h'_{k,u}\,\varepsilon_u w_u) - \nabla p_k(\theta)\|$ to conclude that

$$
\begin{aligned}
&\|\nabla p_k(\theta_u + h'_{k,u}\,\varepsilon_u w_u) - \nabla p_k(\theta)\| \\
&= \Big(\sum_{i=1}^{d} \Big| \frac{\partial p_k}{\partial \theta_i}(\theta_u + h'_{k,u}\,\varepsilon_u w_u) - \frac{\partial p_k}{\partial \theta_i}(\theta) \Big|^2 \Big)^{1/2} \\
&= \Big(\sum_{i=1}^{d} \Big| \langle \theta - \theta_u - h'_{k,u}\,\varepsilon_u w_u, \nabla \frac{\partial p_k}{\partial \theta_i}(\theta + h_{k,u,i}(\theta_u + h'_{k,u}\,\varepsilon_u w_u)) \rangle \Big|^2 \Big)^{1/2} \\
&\leq \Big(\sum_{i=1}^{d} \|\theta - \theta_u - h'_{k,u}\,\varepsilon_u w_u\|^2 \Big\| \nabla \frac{\partial p_k}{\partial \theta_i}(\theta + h_{k,u,i}(\theta_u + h'_{k,u}\,\varepsilon_u w_u)) \Big\|^2 \Big)^{1/2} \\
&= \|\theta - \theta_u - h'_{k,u}\,\varepsilon_u w_u\| \Big(\sum_{i,j=1}^{d} \Big| \frac{\partial^2 p_k}{\partial \theta_i \partial \theta_j}(\theta + h_{k,u,i}(\theta_u + h'_{k,u}\,\varepsilon_u w_u)) \Big|^2 \Big)^{1/2} \\
&\leq (\|\theta - \theta_u\| + h'_{k,u}\,\varepsilon_u \|w_u\|) \Big(\sum_{i,j=1}^{d} \Big| \frac{\partial^2 p_k}{\partial \theta_i \partial \theta_j}(\theta + h_{k,u,i}(\theta_u + h'_{k,u}\,\varepsilon_u w_u)) \Big|^2 \Big)^{1/2}. \qquad (5.51)
\end{aligned}
$$

Now let $u \in \mathbb{N}$ be sufficiently large, such that $\theta + h_{k,u,i}(\theta_u + h'_{k,u}\,\varepsilon_u w_u) \in B_r(\theta)$. Then we observe that

$$
\|\theta - \theta_u\| \sum_{k=0}^{\infty} (2^{\lambda'-1} \vee 1) \Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} dF(x)\Big) \Big(\sum_{i,j=1}^{d} \Big| \frac{\partial^2 p_k}{\partial \theta_i \partial \theta_j}(\theta + h_{k,u,i}(\theta_u + h'_{k,u}\varepsilon_u w_u)) \Big|^2 \Big)^{1/2} \tag{5.52}
$$

converges to zero as $u \to \infty$, because $\|\theta - \theta_u\| \to 0$ and the sum is finite due to assumption (5.30). Following the same line of reasoning, we conclude that

$$
\varepsilon_u \|w_u\| \sum_{k=0}^{\infty} h'_{k,u} (2^{\lambda'-1} \vee 1) \Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} dF(x)\Big) \Big(\sum_{i,j=1}^{d} \Big| \frac{\partial^2 p_k}{\partial \theta_i \partial \theta_j}(\theta + h_{k,u,i}(\theta_u + h'_{k,u}\varepsilon_u w_u)) \Big|^2 \Big)^{1/2} \tag{5.53}
$$

converges to zero. Hence, the right-hand side of (5.47) converges to zero.

For the second summand, we can use (5.38) again, to obtain

$$
\begin{aligned}
&\Big\| \sum_{k=0}^{\infty} (F^{*k} - F_u^{*k}) \, \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} \Big\|_{\phi_{\lambda'}} \\
&\leq \sum_{k=0}^{\infty} \|(F - F_u) * H_k(F, F_u)\|_{\phi_{\lambda'}} \Big| \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} \Big| \\
&= \sum_{k=0}^{\infty} \|(F - F_u) * H_k(F, F_u)\|_{\phi_{\lambda'}} \big| \langle w_u, \nabla p_k(\theta_u + h'_{k,u}\,\varepsilon_u w_u) \rangle \big| \\
&\leq \|w_u\| \sum_{k=0}^{\infty} \|(F - F_u) * H_k(F, F_u)\|_{\phi_{\lambda'}} \Big\| \nabla p_k(\theta_u + h'_{k,u}\,\varepsilon_u w_u) \Big\|, \qquad (5.54)
\end{aligned}
$$

for some $h'_{k,u} \in (0,1)$, where we applied the Mean Value theorem to the difference quotient, similar to the approach in (5.49). Now we can apply part (ii) of Lemma 4.5 in [12] to obtain

$$\|(F - F_u) * H_k(F, F_u)\|_{\phi_{\lambda'}} \leq \|F_u - F\|_{\phi_{\lambda'}} \, 2^{\lambda'} \, k \, (1 + 2^{\lambda'} \, (2^{\lambda'-1} \vee 1)(2 + (k-1)^{\lambda' \vee 1} C_2))$$

for every $k \in \mathbb{N}_0$. By $\lambda > \lambda'$ and $\|F_u - F\|_{\phi_\lambda} \to 0$, we can conclude that $\|F_u - F\|_{\phi_{\lambda'}} \to 0$ for $u \to \infty$. Now assume again, that $u \in \mathbb{N}$ is sufficiently large, such that $\theta_u + h'_{k,u} \, \varepsilon_u w_u \in B_r(\theta)$. In this case the sum on the right-hand side of (5.54) is finite due to assumption (5.29). Hence, the right-hand side of (5.54) converges to zero. Note that the calculations in (5.49)–(5.53) also show that

$$\lim_{u \to \infty} \left| \sum_{k=0}^{\infty} \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} - \sum_{k=0}^{\infty} \langle w, \nabla p_k(\theta) \rangle \right| = 0,$$

under the assumptions of the lemma. Furthermore, we observe that

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{1}_{[0,\infty)} \frac{p_k(\theta_u + \varepsilon_u w_u) - p_k(\theta_u)}{\varepsilon_u} &= \frac{1}{\varepsilon_u} \mathbb{1}_{[0,\infty)} \Big(\sum_{k=0}^{\infty} p_k(\theta_u + \varepsilon_u w_u) - \sum_{k=0}^{\infty} p_k(\theta_u) \Big) \\ &= 0, \end{aligned} \tag{5.55}$$

for every $u \in \mathbb{N}$, because $(\theta_u), (\theta_u + \varepsilon_u w_u) \subset \Theta$ and we assumed that $\sum_{k=0}^{\infty} p_k(\theta) = 1$ for every $\theta \in \Theta$. Hence, one finds that

$$\sum_{k=0}^{\infty} \mathbb{1}_{[0,\infty)} \, \langle w, \nabla p_k(\theta) \rangle = 0$$

for every $w \in \Theta$, such that the expression in (5.31) is equivalent to

$$\dot{H}_{(\theta,F)}(w, v) := -v * \sum_{k=1}^{\infty} k \, F^{*(k-1)} \, p_k(\theta) \; - \; \sum_{k=0}^{\infty} F^{*k} \, \langle w, \nabla p_k(\theta) \rangle. \tag{5.56}$$

What remains to be shown is the $(\max\{\| \cdot \|, \| \cdot \|_{\phi_\lambda}\}, \| \cdot \|_{\phi_{\lambda'}})$-continuity of $\dot{H}_{(\theta,F)}$. To this aim, let $(w, v) \in \mathbb{R}^d \times \mathrm{D}_{\phi_\lambda}$ and (w_u, v_u) such that $\max\{\|w_u - w\|, \|v_u - v\|_{\phi_\lambda}\} \to 0$. Then we have

$$\begin{aligned} &\Big\| \dot{H}_{(\theta,F)}(w_u, v_u) - \dot{H}_{(\theta,F)}(w, v) \Big\|_{\phi_{\lambda'}} \\ &= \Big\| - \Big(v_u * \sum_{k=1}^{\infty} k \, F^{*(k-1)} \, p_k(\theta) \; - v * \sum_{k=1}^{\infty} k \, F^{*(k-1)} \, p_k(\theta) \Big) \\ &\quad + \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \, \langle w_u, \nabla p_k(\theta) \rangle - \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) \, \langle w, \nabla p_k(\theta) \rangle \Big\|_{\phi_{\lambda'}} \\ &\leq \sum_{k=1}^{\infty} \|(v_u - v) * F^{*(k-1)}\|_{\phi_{\lambda'}} \, k \, p_k(\theta) \; + \sum_{k=0}^{\infty} \|\mathbb{1}_{[0,\infty)} - F^{*k}\|_{\phi_{\lambda'}} \, |\langle w_u - w, \nabla p_k(\theta) \rangle|. \end{aligned} \tag{5.57}$$

Now for every $k \in \mathbb{N}$ we can apply Lemma 2.3 and Equation (2.4) in [52], yielding

$$\begin{aligned}
&\|(v_u - v) * F^{*(k-1)}\|_{\phi_{\lambda'}} \\
&\leq \; 2^{\lambda'} \|v_u - v\|_{\phi_{\lambda'}} \left(\|\mathbb{1}_{[0,\infty)} \|F^{*(k-1)}\|_\infty - F^{*(k-1)}\|_{\phi_{\lambda'}} + \|F^{*(k-1)}\|_\infty\right) \\
&= \; 2^{\lambda'} \|v_u - v\|_{\phi_{\lambda'}} \left(\|\mathbb{1}_{[0,\infty)} - F^{*(k-1)}\|_{\phi_{\lambda'}} + 1\right) \\
&\leq \; 2^{\lambda'} \|v_u - v\|_{\phi_{\lambda'}} \left((2^{\lambda'-1} \vee 1) \left(1 + (k-1)^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\right) + 1\right). \qquad (5.58)
\end{aligned}$$

Hence, the first sum on the right-hand side of (5.57) can be bounded as follows

$$\begin{aligned}
&\sum_{k=1}^{\infty} \|(v_u - v) * F^{*(k-1)}\|_{\phi_{\lambda'}} \, k \, p_k(\theta) \\
&\leq 2^{\lambda'} \|v_u - v\|_{\phi_{\lambda'}} \sum_{k=1}^{\infty} k \, p_k(\theta) \left((2^{\lambda'-1} \vee 1) \left(1 + (k-1)^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\right) + 1\right). \qquad (5.59)
\end{aligned}$$

Now the sum converges due to the assumption $\sum_{k \in \mathbb{N}} k^{(1+\lambda)\vee 2} \, p_k(\theta) < \infty$, such that the series converges to zero as $\|v_u - v\|_{\phi_{\lambda'}} \to 0$, which is implied by $\|v_u - v\|_{\phi_\lambda} \to 0$. For the second sum on the right-hand side of (5.57) we can use part (i) of Lemma 4.5 in [12] together with the Cauchy-Schwarz inequality to obtain

$$\|\mathbb{1}_{[0,\infty)} - F^{*k}\|_{\phi_{\lambda'}} \, |\langle w_u - w, \nabla p_k(\theta)\rangle| \;\; \leq \;\; \|w_u - w\| \, \|\nabla p_k(\theta)\| \, \left(2^{\lambda'-1} \vee 1\right) \left(1 + k^{\lambda' \vee 1} C_2\right)$$

for every $k \in \mathbb{N}_0$. Hence

$$\begin{aligned}
&\sum_{k=0}^{\infty} \|\mathbb{1}_{[0,\infty)} - F^{*k}\|_{\phi_{\lambda'}} |\langle w_u - w, \nabla p_k(\theta)\rangle| \\
&\leq \|w_u - w\| \sum_{k=1}^{\infty} \|\nabla p_k(\theta)\| \, \left(2^{\lambda'-1} \vee 1\right) \left(1 + k^{\lambda' \vee 1} C_2\right), \qquad (5.60)
\end{aligned}$$

where we observe that the series on the right-hand side of (5.60) converges due to the assumption $\sum_{k=1}^{\infty} \|\nabla p_k(\theta)\| k^{\lambda' \vee 1} < \infty$, which is an immediate consequence of assumption (5.29). That is, the second sum on the right-hand side of (5.57) converges to zero as $\|w_u - w\| \to 0$, such that

$$\left\|\dot{H}_{(\theta,F)}(w_u, v_u) - \dot{H}_{(\theta,F)}(w, v)\right\|_{\phi_{\lambda'}} \longrightarrow 0, \qquad (5.61)$$

as $\max\{\|w_u - w\|, \|v_u - v\|_{\phi_\lambda}\} \to 0$. This completes the proof. □

Note that we had to consider the tail functional associated with the compound distribution function instead of the distribution function itself, to guarantee the finiteness w.r.t. $\|\cdot\|_{\phi_{\lambda'}}$ of the third summand on the right-hand side of (5.34).

To be able to derive the uniform quasi-Hadamard derivative of the compound distribution functional, we will now determine the uniform quasi-Hadamard derivate of Λ. The proof of Theorem 5.4.2 will then be a direct consequence of the chain rule for uniformly quasi-Hadamard differentiable functionals in the form of Lemma A.5 in [12]. The following lemma will be concerned with the uniform quasi-Hadamard derivative of the map Λ. To this end, let for any $\lambda \geq 0$

$$\mathcal{T}_{\phi_\lambda} := \left\{ \mathbb{1}_{[0,\infty)} - F\,;\, F \in \mathbb{F}_{\phi_\lambda} \right\}.$$

Then $\mathcal{T}_{\phi_\lambda}$ is nothing but the set of all (two-sided) tail functions associated with distribution functions in $\mathbb{F}_{\phi_\lambda}$.

Lemma 5.4.4 *Let $\lambda' \geq 0$ and $T \in \mathcal{T}_{\phi_{\lambda'}}$. The map $\Lambda : \mathcal{T}_{\phi_{\lambda'}} \to \mathbb{F}$ is uniformly quasi-Hadamard differentiable at T tangentially to $\mathrm{C}_{\phi_{\lambda'}}\langle \mathrm{D}_{\phi_{\lambda'}} \rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$. Moreover, the uniform quasi-Hadamard derivative $\dot{\Lambda}_T : \mathrm{D}_{\phi_{\lambda'}} \to \mathrm{D}_{\phi_{\lambda'}}$ is given by*

$$\dot{\Lambda}_T(v) := -v.$$

Proof First we observe that for every $T_1, T_2 \in \mathcal{T}_{\phi_{\lambda'}}$ we have

$$\begin{aligned} \|\Lambda(T_1) - \Lambda(T_2)\|_{\phi_{\lambda'}} &= \|T_2 - T_1\|_{\phi_{\lambda'}} \\ &\leq \|T_1\|_{\phi_{\lambda'}} + \|T_2\|_{\phi_{\lambda'}} \\ &\leq \int (1+|x|)^{\lambda'} dF_1(x) + \int (1+|x|)^{\lambda'} dF_2(x), \end{aligned} \tag{5.62}$$

where for every $i = 1, 2$ we denoted by $F_i = \mathbb{1}_{[0,\infty)} - T_i$ the corresponding distribution function. The last inequality is due to Equation (2.1) in [52]. According to Lemma 2.2 in [52] we can conclude that both integrals on the right-hand side of (5.62) are finite under the assumptions of the theorem. Thus, $\mathrm{D}_{\phi_{\lambda'}}$ can be seen as the trace.

We will now show, that $\dot{\Lambda}_T$ is indeed the uniform quasi-Hadamard derivative of Λ. What remains to be shown is that

$$\lim_{u\to\infty} \left\| \frac{\Lambda(T_u + \varepsilon_u v_u) - \Lambda(T_u)}{\varepsilon_u} + v \right\|_{\phi_{\lambda'}} = 0, \tag{5.63}$$

holds for every quadrupel $((T_u), v, (v_u), (\varepsilon_u))$ with $(T_u) \subset \mathcal{T}_{\phi_{\lambda'}}$, satisfying $\|T_u - T\|_{\phi_{\lambda'}} \to 0$, $v \in \mathrm{D}_{\phi_{\lambda'}}$, $(v_u) \subset \mathrm{D}_{\phi_{\lambda'}}$, $\|v_u - v\|_{\phi_{\lambda'}} \to 0$, $(T_u + \varepsilon_u v_u) \subset \mathcal{T}_{\phi_{\lambda'}}$, and $(\varepsilon_u) \subset (0, \infty)$, with $\varepsilon_u \to 0$. Indeed, we observe that for every $u \in \mathbb{N}$, we have

$$\begin{aligned} \left\| \frac{\Lambda(T_u + \varepsilon_u v_u) - \Lambda(T_u)}{\varepsilon_u} + v \right\|_{\phi_{\lambda'}} &= \left\| \frac{\mathbb{1}_{[0,\infty)} - (T_u + \varepsilon_u v_u) - (\mathbb{1}_{[0,\infty)} - T_u)}{\varepsilon_u} + v \right\|_{\phi_{\lambda'}} \\ &= \|v - v_u\|_{\phi_{\lambda'}}. \end{aligned} \tag{5.64}$$

Hence, the assertion in (5.63) holds true, because we assumed that $\|v - v_u\|_{\phi_{\lambda'}} \to 0$. Moreover, by the representation $-\dot{\Lambda}_T = \mathrm{id}_{\mathrm{D}_{\phi_{\lambda'}}}$, we can easily conclude on the $(\|\cdot\|_{\phi_{\lambda'}}, \|\cdot\|_{\phi_{\lambda'}})$-continuity of the uniform quasi-Hadamard derivative. This completes the proof. $\square$

We are now in a position to prove Theorem 5.4.2.

Proof of Theorem 5.4.2 Our goal is to apply the chain rule for uniform quasi-Hadamard differentiable functionals in the form of Lemma A.5 in [12] to the mappings $H : \Theta \times \mathbb{F}_{\phi_\lambda} \to \mathcal{T}$ and $\widetilde{H} := \Lambda : \mathcal{T}_{\phi_{\lambda'}} \to \mathbb{F}$. By Lemma 2.2 in [52] we observe that $H(\Theta \times \mathbb{F}_{\phi_\lambda}) \subset \mathcal{T}_{\phi_{\lambda'}}$, such that the composition $\Lambda \circ H$ is well defined on $\Theta \times \mathbb{F}_{\phi_\lambda}$.

Now Lemmas 5.4.3 and 5.4.4 imply that the assumptions of Lemma A.5 in [12] are fulfilled. With the help of the latter lemma we can now conclude that $\mathcal{C} = \Lambda \circ H : \Theta \times \mathbb{F}_{\phi_\lambda} \to \mathbb{F}$ is uniformly quasi-Hadamard differentiable in (θ, F) tangentially to $(\Theta \times \mathrm{C}_{\phi_\lambda})\langle\Theta \times \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ and the uniform quasi-Hadamard derivative $\dot{\mathcal{C}}_{(\theta,F)} : \mathbb{R}^d \times \mathrm{D}_{\phi_\lambda} \to \mathrm{D}_{\phi_{\lambda'}}$ is given by

$$\dot{\mathcal{C}}_{(\theta,F)} = \dot{\Lambda}_{H(\theta,F)} \circ \dot{H}_{(\theta,F)}.$$

This yields the claim and completes the proof. □

5.4.3 On the asymptotic distribution of the estimated compound distribution function

In this section we are going to determine the asymptotic error distribution of the sequence of plug-in estimators $(\mathcal{C}(\widehat{\theta}_u, \widehat{F}_u))$. The central tool to be used to prove the assertion will be the uniform delta-method of Theorem 5.4.1. More explicitly, we aim to derive the asymptotic error distribution of $(\mathcal{C}(\widehat{\theta}_u, \widehat{F}_u))$ from the asymptotic error distributions of the sequences of underlying estimators $(\widehat{\theta}_u)$ and $(\widehat{F}_u)$ by an application of the delta-method. We therefore have to check that all assumptions of the theorems are fulfilled.

In the following the roles of ϑ_u, $\widehat{\vartheta}_u$, F_u and $\widehat{F}_u$ in the sense of Theorem 5.4.1 will be played by θ, $\widehat{\theta}_u$, F and $\widehat{F}_u$. Here $\widehat{\theta}_u$ and $\widehat{F}_u$ are as in (5.6) and (5.7), respectively. Moreover, the roles of (a_u), G, B_1 and B_2 will be played by $(\sqrt{u})$, $\mathcal{C}$, ξ and B_F, respectively. Here ξ and B_F are as in Theorem 5.2.1.

The following lemma will show measurability of the estimators. In particular, Lemma 5.4.5 will show, that part (a) of Theorem 5.4.1 is fulfilled.

Lemma 5.4.5 *Let $\widehat{\theta}_u$ and $\widehat{F}_u$ be as in (5.6) and (5.7), respectively, and let $F \in \mathbb{F}_{\phi_\lambda}$ and $\theta \in (0, \infty)$. For every $u \in \mathbb{N}$*

$$\sqrt{u}\left(\begin{bmatrix} \widehat{\theta}_u \\ \widehat{F}_u \end{bmatrix} - \begin{bmatrix} \theta \\ F \end{bmatrix}\right) \tag{5.65}$$

takes values only in $\mathbb{R} \times \mathrm{D}_{\phi_\lambda}$ and is $(\mathcal{F}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{D}_{\phi_\lambda})$-measurable. That is, the first part of condition (a) in Theorem 5.4.1 holds true.

Proof Let $u \in \mathbb{N}$ and let $\Psi_u : \Omega \times \mathbb{R} \to \mathbb{R}^2$ be defined by

$$\Psi_u(\omega; t) := \sqrt{u}\begin{bmatrix} \widehat{\theta}_u(\omega) - \theta \\ \widehat{F}_u(\omega; t) - F(t) \end{bmatrix}.$$

Now for every $t \in \mathbb{R}$, the mapping $\omega \mapsto \Psi_u(\omega; t)$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}^2))$-measurable. Moreover for every $\omega \in \Omega$ the mapping $t \mapsto \Psi_u(\omega; t)$ is right-continuous, such that the mapping $\omega \mapsto \Psi_u(\omega; \cdot)$ from Ω to D is $(\mathcal{F}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{D})$-measurable and takes values only in $\mathbb{R} \times \mathrm{D}_{\phi_\lambda}$. Thus, the quantity in (5.65) can be seen as a $(\mathcal{F}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{D}_{\phi_\lambda})$-measurable mapping. This leads to the assertion. □

The following theorem will now prove that the second part of part (a) of Theorem 5.4.1 is fulfilled.

Theorem 5.4.6 *Let (N_i) be a sequence of Poiss_θ-distributed random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for some $\theta \in (0, \infty)$, and let (Y_i) be a sequence of i.i.d. random variables on the same probability space with distribution function F, satisfying $\int \phi_\lambda^2 \, dF < \infty$ and being independent of (N_i). Let $N(u) := \sum_{i=1}^{u} N_i$. Moreover, let ξ be a $\mathcal{N}_{0,\theta}$-distributed random variable and B_F be a F-Brownian bridge, as in (4.9), being independent of ξ.*

Let $\widehat{\theta}_u$ and $\widehat{F}_u$ be as in (5.6) and (5.7), respectively. Then we have

$$\sqrt{u}\left(\begin{bmatrix} \widehat{\theta}_u \\ \widehat{F}_u \end{bmatrix} - \begin{bmatrix} \theta \\ F \end{bmatrix}\right) \xrightarrow{\mathrm{d}} \begin{bmatrix} \xi \\ \frac{1}{\sqrt{\theta}} B_F \end{bmatrix}, \tag{5.66}$$

in $(\mathbb{R} \times \mathrm{D}_{\phi_\lambda}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{D}_{\phi_\lambda}, \max\{|\cdot|, \|\cdot\|_{\phi_\lambda}\})$.

Proof To proof the assertion, we will first show that two auxiliary statements hold true. More explicitly, we will show that

$$\sqrt{u}\begin{bmatrix} \frac{1}{u}N(u) - \theta \\ \frac{1}{u\theta}\sum_{i=1}^{\lceil u\theta \rceil}(\mathbb{1}_{[Y_i,\infty)} - F) \end{bmatrix} \xrightarrow{\mathrm{d}} \begin{bmatrix} \xi \\ \frac{1}{\sqrt{\theta}} B_F \end{bmatrix} \tag{5.67}$$

in $(\mathbb{R} \times \mathrm{D}_{\phi_\lambda}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{D}_{\phi_\lambda}, \max\{|\cdot|, \|\cdot\|_{\phi_\lambda}\})$ and

$$\frac{\sqrt{u}}{N(u)}\sum_{i=1}^{N(u)}(\mathbb{1}_{[Y_i,\infty)} - F) - \frac{1}{\sqrt{u}\theta}\sum_{i=1}^{\lceil u\theta \rceil}(\mathbb{1}_{[Y_i,\infty)} - F) \xrightarrow{\mathrm{p}} 0 \tag{5.68}$$

w.r.t. $\mathbb{P}$ in $(\mathrm{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, \|\cdot\|_{\phi_\lambda})$. By

$$\begin{aligned}\sqrt{u}\begin{bmatrix} \frac{1}{u}N(u) - \theta \\ \frac{1}{N(u)}\sum_{i=1}^{N(u)}\mathbb{1}_{[Y_i,\infty)} - F \end{bmatrix} &= \sqrt{u}\begin{bmatrix} \frac{1}{u}N(u) - \theta \\ \frac{1}{u\theta}\sum_{i=1}^{\lceil u\theta \rceil}(\mathbb{1}_{[Y_i,\infty)} - F) \end{bmatrix} \\ &\quad + \sqrt{u}\begin{bmatrix} 0 \\ \frac{1}{N(u)}\sum_{i=1}^{N(u)}(\mathbb{1}_{[Y_i,\infty)} - F) - \frac{1}{u\theta}\sum_{i=1}^{\lceil u\theta \rceil}(\mathbb{1}_{[Y_i,\infty)} - F) \end{bmatrix}\end{aligned}$$

the claim would follow by an application of Slutskys Lemma. To show that (5.67) holds true, we observe that we have

$$\sqrt{u}\left(\frac{1}{u}N(u) - \theta\right) \xrightarrow{\mathrm{d}} \xi, \quad \text{in } (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

by the ordinary Central Limit Theorem, where we keep in mind that $N(u) = \sum_{i=1}^{u} N_i$ holds. With the help of Donskers' invariance principle, we conclude that

$$\frac{1}{\sqrt{u\theta}} \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F) \;=\; \frac{1}{\sqrt{\theta}} \sqrt{\frac{\lceil u\theta \rceil}{u\theta}} \sqrt{\lceil u\theta \rceil}(\widehat{F}_{\lceil u\theta \rceil} - F) \xrightarrow{\mathrm{d}} \frac{1}{\sqrt{\theta}} B_F, \quad \text{in } (\mathrm{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, \|\cdot\|_{\phi_\lambda}).$$

The latter is a direct consequence of Theorem 6.2.1 in [58] along with Slutskys Lemma and the fact that $\lceil u\theta \rceil/(u\theta)$ converges to 1. For (5.68), we can see that

$$\begin{aligned}
&\frac{\sqrt{u}}{N(u)} \sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \frac{1}{\sqrt{u\theta}} \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F) \\
&= \;\Big(\frac{u\theta}{N(u)} - 1\Big) \frac{1}{\sqrt{u\theta}} \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F) \;+\; \frac{\sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)}{\sqrt{u\theta}} \\
&\quad + \Big(\frac{u\theta}{N(u)} - 1\Big) \left(\frac{\sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)}{\sqrt{u\theta}} \right) \\
&=: \;S_1(u) + S_2(u) + S_3(u). \qquad (5.69)
\end{aligned}$$

Hence, it suffices to prove that $S_1(u)$, $S_2(u)$ and $S_3(u)$ converge to zero in probability in $(\mathrm{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, \|\cdot\|_{\phi_\lambda})$. The assertion in (5.68) would then follow by an application of Slutskys Lemma again. Here we stress the fact, that we can indeed use Slutskys Lemma to conclude on the convergence in probability, because convergence in probability to a constant is equivalent to the convergence in distribution to this constant.

Step 1: By the ordinary strong law of large numbers, we conclude that $u\theta/N(u)$ converges to 1 $\mathbb{P}$-a.s. Furthermore, Theorem 6.2.1 in [58] yields the convergence in distribution of $\frac{1}{\sqrt{u\theta}} \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)$ to $\theta^{-1/2} B_F$ in $(\mathrm{D}_{\phi_\lambda}, \mathcal{D}_{\phi_\lambda}, \|\cdot\|_{\phi_\lambda})$, such that $S_1(u)$ converges to zero in probability w.r.t. $\mathbb{P}$.

Step 2: We have to show, that

$$\lim_{u\to\infty} \mathbb{P}\Big[\Big\{\Big\| \frac{1}{\sqrt{u\theta}} \Big(\sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)\Big) \Big\|_{\phi_\lambda} > \delta \Big\}\Big] = 0, \qquad (5.70)$$

for every $\delta > 0$. To this end, let $\alpha \in (0, 1/2)$ and $\delta > 0$. Then we conclude that

$$\begin{aligned}
&\mathbb{P}\Big[\Big\{\Big\| \frac{1}{\sqrt{u\theta}} \Big(\sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)\Big) \Big\|_{\phi_\lambda} > \delta \Big\}\Big] \\
&= \mathbb{P}\Big[\Big\{\Big\| \frac{1}{\sqrt{u\theta}} \Big(\sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)\Big) \Big\|_{\phi_\lambda} > \delta,\ \Big|\frac{N(u)}{u} - \theta\Big| \geq \frac{1}{u^\alpha} \Big\}\Big] \\
&\quad + \mathbb{P}\Big[\Big\{\Big\| \frac{1}{\sqrt{u\theta}} \Big(\sum_{i=1}^{N(u)} (\mathbb{1}_{[Y_i,\infty)} - F) - \sum_{i=1}^{\lceil u\theta \rceil} (\mathbb{1}_{[Y_i,\infty)} - F)\Big) \Big\|_{\phi_\lambda} > \delta,\ \Big|\frac{N(u)}{u} - \theta\Big| < \frac{1}{u^\alpha} \Big\}\Big]
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}\Big[\Big\{\Big\|\frac{1}{\sqrt{u\theta}}\Big(\sum_{i=1}^{N(u)}(\mathbb{1}_{[Y_i,\infty)}-F)-\sum_{i=1}^{\lceil u\theta\rceil}(\mathbb{1}_{[Y_i,\infty)}-F)\Big)\Big\|_{\phi_\lambda}>\delta,\ \Big|\frac{N(u)}{u}-\theta\Big|<\frac{1}{u^\alpha}\Big\}\Big]\\
&\quad+\mathbb{P}\Big[\Big\{\Big|\frac{N(u)}{u}-\theta\Big|\geq\frac{1}{u^\alpha}\Big\}\Big]\\
&=: S_{2,1}(u,\delta,\alpha)\ +\ S_{2,2}(u,\delta,\alpha).
\end{aligned}
\tag{5.71}
$$

Using Proposition 3.2.10 again, together with

$$
\mathbb{P}\Big[\Big\{\Big|\frac{N(u)}{u}-\theta\Big|\geq\frac{1}{u^\alpha}\Big\}\Big]\ =\ \mathbb{P}\Big[\Big\{u^\alpha\Big|\frac{N(u)}{u}-\theta\Big|\geq 1\Big\}\Big]
$$

we observe that $S_{2,2}(u,\delta,\alpha)$ converges to zero for every $\alpha\in(0,1/2)$ and $\delta>0$. For the convergence of $S_{1,2}(u,\delta,\alpha)$ we conclude that

$$
\begin{aligned}
&\mathbb{P}\Big[\Big\{\Big\|\frac{1}{\sqrt{u\theta}}\Big(\sum_{i=1}^{N(u)}(\mathbb{1}_{[Y_i,\infty)}-F)-\sum_{i=1}^{\lceil u\theta\rceil}(\mathbb{1}_{[Y_i,\infty)}-F)\Big)\Big\|_{\phi_\lambda}>\delta,\ \Big|\frac{N(u)}{u}-\theta\Big|<\frac{1}{u^\alpha}\Big\}\Big]\\
&=\mathbb{P}\Big[\Big\{\Big\|\Big(\sum_{i=1}^{N(u)}(\mathbb{1}_{[Y_i,\infty)}-F)-\sum_{i=1}^{\lceil u\theta\rceil}(\mathbb{1}_{[Y_i,\infty)}-F)\Big)\Big\|_{\phi_\lambda}>\sqrt{u}\theta\delta,\ \Big|\frac{N(u)}{u}-\theta\Big|<\frac{1}{u^\alpha}\Big\}\Big]\\
&\leq\mathbb{P}\Big[\Big\{\Big\|\sum_{i=1}^{u^{1-\alpha}}(\mathbb{1}_{[Y_i,\infty)}-F)\Big\|_{\phi_\lambda}>\sqrt{u}\theta\delta\Big\}\Big]\\
&=\mathbb{P}\Big[\Big\{u^{1/2-\alpha}\Big\|\frac{1}{u^{1-\alpha}}\sum_{i=1}^{u^{1-\alpha}}(\mathbb{1}_{[Y_i,\infty)}-F)\Big\|_{\phi_\lambda}>\theta\delta\Big\}\Big]\\
&=\mathbb{P}\Big[\Big\{u^{1/2-\alpha}\Big\|\frac{1}{u^{1-\alpha}}\sum_{i=1}^{u^{1-\alpha}}\mathbb{1}_{[Y_i,\infty)}-F\Big\|_{\phi_\lambda}>\theta\delta\Big\}\Big].
\end{aligned}
\tag{5.72}
$$

Letting $m=u^{1-\alpha}$, the right-hand side of (5.72) is nothing but

$$
\mathbb{P}\Big[\Big\{m^{\frac{1-2\alpha}{2(1-\alpha)}}\Big\|\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}_{[Y_i,\infty)}-F\Big\|_{\phi_\lambda}>\theta\delta\Big\}\Big].
\tag{5.73}
$$

By $\alpha\in(0,1/2)$ we conclude that $\frac{1-2\alpha}{2(1-\alpha)}\in(0,1/2)$. Hence, the right-hand side of (5.73) converges to zero as $m\to\infty$ for every $\delta>0$ by Theorem 2.1 in [65]. This shows (5.70).

Step 3: By the Marcinkiewicz-Zygmund SLLN of Proposition 3.2.10, we deduce that $u\theta/N(u)-1$ converges to zero $\mathbb{P}$-a.s. Now, following the same line of reasoning as in Step 2 the fraction in $S_3(u)$ converges to zero in probability w.r.t. $\mathbb{P}$ in $(\mathrm{D}_{\phi_\lambda},\mathcal{D}_{\phi_\lambda},\|\cdot\|_{\phi_\lambda})$. Hence, the assertion follows by an application of Slutskys Lemma again. □

Lemma 5.4.5 and Theorem 5.4.6 have shown that condition (a) of the uniform delta-method of Theorem 5.4.1 is fulfilled under the particular assumptions. Moreover the examinations subsequent to (5.32) have shown that

$$
\sqrt{u}(\mathcal{C}(\widehat{\theta}_u.\widehat{F}_u)-\mathcal{C}(\theta,F))=-\sqrt{u}(H(\widehat{\theta}_u.\widehat{F}_u)-H(\theta,F))
$$

takes values only in $\mathrm{D}_{\phi_{\lambda'}}$. That is, the second part of condition (b) in 5.4.1 is also satisfied. Furthermore, Theorem 5.4.2 has shown the uniform quasi-Hadamard differentiability of the compound distribution function $\mathcal{C}$ at (θ, F) tangentially to $(\Theta \times \mathrm{D}_{\phi_\lambda})\langle \Theta \times \mathrm{D}_{\phi_\lambda} \rangle$. However, to be able to apply the results on the differentiability of the compound distribution functional $\mathcal{C}$, we have to check that the assumptions of Theorem 5.4.2 are fulfilled in the setting of the compound Poisson model. More explicitly, we have to guarantee that $\sum_{k\in\mathbb{N}} k^{1+\lambda}\, p_k(\theta) < \infty$ and that there exists some $r \in (0, \infty)$, such that the following two assertions hold true for $\lambda > 0$:

$$\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\widetilde{\theta}\in(\theta-r,\theta+r)} |p_k'(\widetilde{\theta})| < \infty, \tag{5.74}$$

$$\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\widetilde{\theta}\in(\theta-r,\theta+r)} |p_k''(\widetilde{\theta})| < \infty. \tag{5.75}$$

The convergence of the first series is due to the fact that the Poisson distribution possesses all moments. For the convergence in (5.74), let $r \in (0, \theta)$. Then we observe that $(\theta - r, \theta + r) \subset (0, \infty)(= \Theta)$, such that

$$\begin{aligned}
\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\widetilde{\theta}\in(\theta-r,\theta+r)} |p_k'(\widetilde{\theta})| &= \sum_{k=1}^{\infty} k^{(1+\lambda)\vee 2} \sup_{\widetilde{\theta}\in(\theta-r,\theta+r)} \left|\frac{1}{k!}(k\,\widetilde{\theta}^{k-1} - \widetilde{\theta}^k)\, e^{-\widetilde{\theta}}\right| \\
&\leq \sum_{k\in\mathbb{N}} \frac{k^{(1+\lambda)\vee 2}}{k!} \Big(k(\theta + r)^{k-1} + (\theta + r)^k\Big) e^{-(\theta-r)} \\
&= e^{2r} \sum_{k\in\mathbb{N}_0} ((k+1)^{(1+\lambda)\vee 2} + k^{(1+\lambda)\vee 2}) \frac{(\theta+r)^k}{k!} e^{-(\theta+r)}\Big) \\
&= e^{2r} \sum_{k\in\mathbb{N}_0} ((k+1)^{(1+\lambda)\vee 2} + k^{(1+\lambda)\vee 2})\, p_k(\theta + r),
\end{aligned}$$

such that the convergence of the series on the right-hand side follows by the fact that the Poisson distribution with parameter $\theta + r \in (0, \infty)$ possesses all moments. This proves (5.74) for any $\lambda > 0$ and $r \in (0, \theta)$. To show (5.75), similar arguments lead to

$$\begin{aligned}
\sum_{k\in\mathbb{N}} k^{(1+\lambda)\vee 2} \sup_{\widetilde{\theta}\in(\theta-r,\theta+r)} |p_k''(\widetilde{\theta})| &= \sum_{k=1}^{\infty} k^{(1+\lambda)\vee 2} \sup_{\widetilde{\theta}\in(\theta-r,\theta+r)} \left|\frac{1}{k!}(k(k-1)\widetilde{\theta}^{k-2} - 2k\widetilde{\theta}^{k-1} + \theta^k)\, e^{-\widetilde{\theta}}\right| \\
&\leq e^{2r} \sum_{k\in\mathbb{N}_0} ((k+2)^{(1+\lambda)\vee 2} + 2(k+1)^{(1+\lambda)\vee 2} + k^{(1+\lambda)\vee 2}) p_k(\theta + r).
\end{aligned}$$

Now the series is finite again because the Poisson distribution with parameter $\theta + r \in (0, \infty)$ possesses all moments. This proves (5.75) for any $\lambda > 0$ and $r \in (0, \theta)$ and shows that the assumptions of Theorem 5.4.2 are fulfilled in the setting of the compound Poisson model.

Hence, to be able to apply the functional delta-method of Theorem 5.4.1, we still have to show the $(\mathcal{F}, \mathcal{D}_{\phi_{\lambda'}})$ measurability of $\sqrt{u}(\mathcal{C}(\widehat{\theta}_u.\widehat{F}_u) - \mathcal{C}(\theta, F))$. This will be done in the following Lemma.

Lemma 5.4.7 *Assume that $\theta \in (0, \infty)$ and $F \in \mathbb{F}_{\phi_\lambda}$. Let $\mathcal{C}$ be the compound distribution functional as defined in (5.20) and let $\widehat{\theta}_u$ and $\widehat{F}_u$ be as in (5.6) and (5.7), respectively. Then we always have that*

$$\sqrt{u}(\mathcal{C}(\widehat{\theta}_u.\widehat{F}_u) - \mathcal{C}(\theta, F))$$

is $(\mathcal{F}, \mathcal{D}_{\phi_{\lambda'}})$-measurable for every $u \in \mathbb{N}$.

Proof To show the $(\mathcal{F}, \mathcal{D}_{\phi_{\lambda'}})$-measurability of $\sqrt{u}(\mathcal{C}(\widehat{\theta}_u.\widehat{F}_u) - \mathcal{C}(\theta, F))$, write

$$\begin{aligned}\sqrt{u}(\mathcal{C}(\widehat{\theta}_u(\omega), \widehat{F}_u(\omega))(t) - \mathcal{C}(\theta, F)(t)) \;&=\; \sqrt{u}\left(\sum_{k=0}^{\infty} \widehat{F}_u^{*k}(\omega; t)\, p_k(\widehat{\theta}_u(\omega)) - \sum_{k=0}^{\infty} F^{*k}(t)\, p_k(\theta)\right) \\ &=: \Psi(\omega; t),\end{aligned} \tag{5.76}$$

for every $t \in \mathbb{R}$ and $\omega \in \Omega$. Then we observe that the mapping $\omega \mapsto \Psi(\omega; t)$ is $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable for every $t \in \mathbb{R}$. Moreover, the mapping $t \mapsto \Psi(\omega; t)$ is right-continuous for every $\omega \in \Omega$. Thus, the mapping is $(\mathcal{F}, \mathcal{D})$-measurable and takes values only in $\mathrm{D}_{\phi_{\lambda'}}$. This has been shown in the examinations subsequent to (5.32). Thus, the mapping $\omega \mapsto \Psi(\omega; \cdot)$ is a $(\mathcal{F}, \mathcal{D}_{\phi_{\lambda'}})$-measurable map. This proves the assertion. □

The below Corollary 5.4.8 will now state the asymptotic error distribution of the compound distribution function. The assertion of the corollary follows directly by an application of the uniform delta-method of Theorem 5.4.1.

The previous examinations have shown that the conditions of Theorem 5.4.1 are fulfilled in our present setting, where the mapping G is given by the compound distribution function defined in (5.20). Corollary 5.4.8 will now yield the asymptotic distribution of the sequence of estimated compound distribution functions.

Note that by

$$\sum_{k=0}^{\infty} \frac{1}{k!}\,(k\,\theta^{k-1} - \theta^k)\, e^{-\theta} \;=\; \sum_{k=1}^{\infty} \frac{1}{(k-1)!}\,\theta^{k-1}\, e^{-\theta} - \sum_{k=0}^{\infty} \frac{1}{k!}\,\theta^k\, e^{-\theta} \;=\; 0, \tag{5.77}$$

the uniform quasi Hadamard derivative of the compound distribution functional at (θ, F) $\dot{\mathcal{C}}_{(\theta,F)} : \mathbb{R} \times \mathrm{D}_{\phi_\lambda} \to \mathrm{D}_{\phi_{\lambda'}}$ has the following representation

$$\dot{\mathcal{C}}_{(\theta,F)}(w, v) \;=\; v * \sum_{k=1}^{\infty} kF^{*(k-1)}\frac{\theta^k}{k!}e^{-\theta} + w\, e^{-\theta} \sum_{k=0}^{\infty} F^{*k}\, \frac{1}{k!}\,(k\,\theta^{k-1} - \theta^k). \tag{5.78}$$

Corollary 5.4.8 *Let $\lambda > 1$ and let $\mathcal{C}$ be the compound distribution functional as defined in (5.20). Moreover, let $F \in \mathbb{F}_{\phi_{2\lambda}}$, that is $\int \phi_\lambda^2\, dF < \infty$, and let $\theta \in (0, \infty)$. Let $\widehat{\theta}_u$ and $\widehat{F}_u$ be as in (5.6) and (5.7), respectively. Then for every $\lambda' \in (0, \lambda)$*

$$\sqrt{u}\, \Big(\mathcal{C}(\widehat{\theta}_u, \widehat{F}_u) - \mathcal{C}(\theta, F)\Big) \;\;\rightsquigarrow^{\circ}\;\; \frac{1}{\sqrt{\theta}}B_F * \sum_{k=1}^{\infty} kF^{*(k-1)}\frac{\theta^k}{k!}e^{-\theta} + \xi\, e^{-\theta}\sum_{k=0}^{\infty} F^{*k}\,\frac{1}{k!}\,(k\,\theta^{k-1} - \theta^k),$$

in $(\mathrm{D}_{\phi_{\lambda'}}, \mathcal{D}_{\phi_{\lambda'}}, \|\cdot\|_{\phi_{\lambda'}})$, where ξ refers to a $\mathcal{N}_{0,\theta}$-distributed random variable and B_F is a F-Brownian bridge, as in (4.9), being independent of ξ.

5.4.4 Proof of Theorem 5.2.1

Proof of Theorem 5.2.1 Note that Theorem 5.2.1 is a direct consequence of the chain rule of Lemma A.5 in [12] along with Theorem 5.4.1.

To apply Lemma A.5 in [12] we have to check that the assumptions of this lemma are fulfilled. To this use, we have to show that

(a) For every sequence $(\theta_u, F_u) \subset (0,\infty) \times \mathbb{F}_{\phi_\lambda}$ satisfying $\max\{|\theta_u - \theta|, \|F_u - F\|_{\phi_\lambda}\} \to 0$, we have
$$\lim_{u\to\infty} \mathcal{C}(\theta_u, F_u)(t) \;=\; \mathcal{C}(\theta, F)(t), \qquad \text{for every } t \in \mathbb{R}.$$

(b) $\mathcal{C}$ is uniformly quasi-Hadamard differentiable at (θ, F) tangentially to $((0,\infty) \times \mathrm{D}_{\phi_\lambda})\langle(0,\infty) \times \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ and the uniform quasi-Hadamard derivative $\dot{\mathcal{C}}_{(\theta,F)}$ satisfies $\dot{\mathcal{C}}_{(\theta,F)}(\mathrm{D}_{\phi_\lambda}) \subset \mathrm{D}_{\phi_{\lambda'}}$.

(c) $\mathcal{R}_\rho$ is uniformly quasi-Hadamard differentiable at $\mathcal{S}$ tangentially to $\mathrm{D}_{\phi_{\lambda'}}\langle \mathrm{D}_{\phi_{\lambda'}}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ and uniform quasi-Hadamard derivative $\dot{\mathcal{R}}_{\rho,\mathcal{S}}$.

To show that condition (a) holds true, we have to show the pointwise convergence of $\mathcal{C}(\theta_u, F_u)$ to $\mathcal{C}(\theta, F)$ for every sequence $(\theta_u, F_u) \subset (0,\infty)\times\mathbb{F}_{\phi_\lambda}$ satisfying $\max\{|\theta_u-\theta|, \|F_u-F\|_{\phi_\lambda}\} \to 0$. We will show even more, namely $\|\mathcal{C}(\theta_u, F_u) - \mathcal{C}(\theta, F)\|_{\phi_{\lambda'}} \to 0$. To this end, let $(\theta_u, F_u) \subset (0,\infty) \times \mathbb{F}_{\phi_\lambda}$. For every $k \in \mathbb{N}_0$, let $H_k : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ be as in (5.37). Then, using the fact that $\sum_{k=0}^\infty p_k(\theta) = 1$ for every $\theta \in (0,\infty)$, we have

$$\begin{aligned}
&\|\mathcal{C}(\theta_u, F_u) - \mathcal{C}(\theta, F)\|_{\phi_{\lambda'}}\\
&= \Big\| \sum_{k=0}^\infty F_u^{*k}\, p_k(\theta_u) \;-\; \sum_{k=0}^\infty F^{*k}\, p_k(\theta)\Big\|_{\phi_{\lambda'}}\\
&= \Big\| \sum_{k=0}^\infty (\mathbb{1}_{[0,\infty)} - F_u^{*k})\, p_k(\theta_u) \;-\; \sum_{k=0}^\infty (\mathbb{1}_{[0,\infty)} - F^{*k})\, p_k(\theta)\Big\|_{\phi_{\lambda'}}\\
&\le \Big\| \sum_{k=0}^\infty (F_u^{*k} - F^{*k})\, p_k(\theta)\Big\|_{\phi_{\lambda'}} + \Big\| \sum_{k=0}^\infty (\mathbb{1}_{[0,\infty)} - F_u^{*k})\, (p_k(\theta_u) \;-\; p_k(\theta))\Big\|_{\phi_{\lambda'}}\\
&\le \Big\| \sum_{k=0}^\infty (F_u^{*k} - F^{*k})\, p_k(\theta)\Big\|_{\phi_{\lambda'}} + \Big\| \sum_{k=0}^\infty (F^{*k} - F_u^{*k})\, (p_k(\theta_u) \;-\; p_k(\theta))\Big\|_{\phi_{\lambda'}}\\
&\quad + \Big\| \sum_{k=0}^\infty (\mathbb{1}_{[0,\infty)} - F^{*k})\, (p_k(\theta_u) \;-\; p_k(\theta))\Big\|_{\phi_{\lambda'}}\\
&= \Big\| (F_u - F) * \sum_{k=0}^\infty H_k(F_u, F)\, p_k(\theta)\Big\|_{\phi_{\lambda'}} + \Big\| (F_u - F) * \sum_{k=0}^\infty H_k(F_u, F)\, (p_k(\theta_u) \;-\; p_k(\theta))\Big\|_{\phi_{\lambda'}}\\
&\quad + \Big\| \sum_{k=0}^\infty (\mathbb{1}_{[0,\infty)} - F^{*k})\, (p_k(\theta_u) \;-\; p_k(\theta))\Big\|_{\phi_{\lambda'}}\\
&=: S_1(u) \;+\; S_2(u) \;+\; S_3(u),
\end{aligned} \tag{5.79}$$

where we used (5.38) for the second "=". For the first summand, we apply part (ii) of Lemma 4.5 in [12] to each of the summands of H_k to derive

$$S_1(u) \le \|F_u - F\|_{\phi_{\lambda'}} \sum_{k=0}^{\infty} p_k(\theta) k\Big(1 + 2^{\lambda'}(2^{\lambda'-1} \vee 1)(2 + (k-1)^{\lambda' \vee 1} C_1)\Big). \tag{5.80}$$

As the series converges, we conclude that $S_1(u) \to 0$ as $\|F_u - F\|_{\phi_\lambda} \to 0$. Using similar arguments, we can conclude that $S_2(u)$ converges to zero as $\|F_u - F\|_{\phi_\lambda} \to 0$. For the last summand we use Inequality (2.4) of [52] to derive

$$\begin{aligned} S_3(u) &\le \sum_{k=0}^{\infty} \|\mathbb{1}_{[0,\infty)} - F^{*k}\|_{\phi_{\lambda'}} \, |p_k(\theta_u) - p_k(\theta)| \\ &\le \sum_{k=0}^{\infty} |p_k(\theta_u) - p_k(\theta)| \, (2^{\lambda'-1} \vee 1)\Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\Big). \end{aligned} \tag{5.81}$$

By the Mean Value Theorem, we conclude that for every $k \in \mathbb{N}_0$ there exists some $\widetilde{\theta}_{u,k}$ between θ and θ_u such that

$$\begin{aligned} &\sum_{k=0}^{\infty} |p_k(\theta_u) - p_k(\theta)| \, (2^{\lambda'-1} \vee 1)\Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\Big) \\ &\le |\theta_u - \theta| \sum_{k=0}^{\infty} |p_k'(\widetilde{\theta}_{u,k})| \, (2^{\lambda'-1} \vee 1)\Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\Big). \end{aligned} \tag{5.82}$$

Since the series on the right-hand side of (5.82) converges, we conclude that $S_3(u)$ converges to zero as $|\theta_u - \theta| \to 0$. This shows that part (a) of the upper conditions is fulfilled.

According to Theorem 5.4.2, the map $\mathcal{C}$ is quasi-Hadamard differentiable at (θ, F) tangentially to $((0,\infty) \times \mathrm{D}_{\phi_\lambda})\langle(0,\infty) \times \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$, which is the first part of condition (b). Then the second part follows from

$$\begin{aligned} \|\dot{\mathcal{C}}_{(\theta,F)}(w,v)\|_{\phi_{\lambda'}} &\le \Big\| v * \sum_{k=1}^{\infty} k \, F^{*(k-1)} p_k(\theta) \Big\|_{\phi_{\lambda'}} + \Big\| w \sum_{k=0}^{\infty} (\mathbb{1}_{[0,\infty)} - F^{*k}) p_k'(\theta) \Big\|_{\phi_{\lambda'}} \\ &\le \Big\| v * \sum_{k=1}^{\infty} k \, F^{*(k-1)} p_k(\theta) \Big\|_{\phi_{\lambda'}} + |w| \sum_{k=0}^{\infty} \|\mathbb{1}_{[0,\infty)} - F^{*k}\|_{\phi_{\lambda'}} \, |p_k'(\theta)| \\ &\le 2^{\lambda'} \, \|v\|_{\phi_{\lambda'}} \sum_{k=1}^{\infty} p_k(\theta) \, k \, \Big(1 + (2^{\lambda'-1} \vee 1)\Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\Big)\Big) \\ &\quad + |w| \sum_{k=1}^{\infty} |p_k'(\theta)| \, (2^{\lambda'-1} \vee 1)\Big(1 + k^{\lambda' \vee 1} \int |x|^{\lambda'} \, dF(x)\Big), \end{aligned} \tag{5.83}$$

for which we applied Lemma 2.3 in [12] and Inequality (2.4) in [52]. Now the claim follows by $\|v\|_{\phi_\lambda} \le \|v\|_{\phi_{\lambda'}} < \infty$ and the fact that both sequences converge. Moreover we assumed that $\mathcal{R}_\rho$ is uniformly quasi-Hadamard differentiable at $\mathcal{S}$ tangentially to $\mathrm{D}_{\phi_{\lambda'}}\langle\mathrm{D}_{\phi_{\lambda'}}\rangle$ with trace

$\mathrm{D}_{\phi_{\lambda'}}$. This yields assumption (c). Hence, we observe that the composition $T_\rho := \mathcal{R}_\rho \circ \mathcal{C}$ is uniformly quasi-Hadamard differentiable at (θ, F) tangentially to $((0,\infty) \times \mathrm{D}_{\phi_\lambda})\langle(0,\infty) \times \mathrm{D}_{\phi_\lambda}\rangle$ with trace $\mathrm{D}_{\phi_{\lambda'}}$ and the uniform quasi-Hadamard derivative is given by $\dot{T}_{\rho,\theta,F} := \dot{\mathcal{R}}_{\rho,\mathcal{C}(\theta,F)} \circ \dot{\mathcal{C}}_{(\theta,F)}$. By Corollary 5.4.8 we also know that

$$\sqrt{u}\,\Big(\mathcal{C}(\widehat{\theta}_u, \widehat{F}_u) - \mathcal{C}(\theta, F)\Big) \rightsquigarrow^\circ \frac{1}{\sqrt{\theta}} B_F * \sum_{k=1}^{\infty} k F^{*(k-1)} \frac{\theta^k}{k!} e^{-\theta} + \xi\, e^{-\theta} \sum_{k=0}^{\infty} F^{*k}\, \frac{1}{k!}\,(k\,\theta^{k-1} - \theta^k),$$

in $(\mathrm{D}_{\phi_{\lambda'}}, \mathcal{D}_{\phi_{\lambda'}}, \|\cdot\|_{\phi_{\lambda'}})$, such that the assertion follows by an application of Corollary 3.1 in [12]. Here we denote by $\rightsquigarrow^\circ$ the convergence in distribution w.r.t. the σ-algebra on $\mathrm{D}_{\phi_{\lambda'}}$, generated by the $\|\cdot\|_{\phi_{\lambda'}}$-open balls.

□

Appendix A

The Panjer recursion

A.1 On the computation of $\widehat{\mu}_u^{*n}$ and $\mathcal{R}_\rho(\widehat{\mu}_u^{*n})$

In general the computation of the n-fold convolution $\widehat{\mu}_u^{*n}$ of $\widehat{\mu}_u$ is more or less impossible. However, in real applications the true μ has support in $h\mathbb{N}_0 := \{0, h, 2h, \ldots\}$ for some fixed $h > 0$, where h represents the smallest monetary unit. We stress the fact that continuous distributions are in fact approximations for the *equidistant discrete* true single claim distribution, and not vice versa. So the empirical probability measure $\widehat{\mu}_u$ is concentrated on the equidistant grid $h\mathbb{N}_0$, too. In this case the estimated total claim distribution $\widehat{\mu}_u^{*n}$ can be computed with the help of the recursive scheme

$$\widehat{\mu}_u^{*n}[\{0\}] \;=\; \widehat{\mu}_u[\{0\}]^n \tag{A.1}$$

$$\widehat{\mu}_u^{*n}[\{jh\}] \;=\; \frac{1}{j\,\widehat{\mu}_u[\{0\}]}\sum_{\ell=1}^{j}((n+1)\ell - j)\,\widehat{\mu}_u[\{\ell h\}]\,\widehat{\mu}_u^{*n}[\{(j-\ell)h\}] \quad \text{for } j \in \mathbb{N}, \tag{A.2}$$

provided $\widehat{\mu}_u[\{0\}] > 0$; see the discussion below. Note that $\widehat{\mu}_u$ as an empirical probability measure has bounded support. Therefore, the whole distribution $\widehat{\mu}_u^{*n}$ can be computed by the scheme (2.10)–(2.11) in finitely many steps. In particular, the estimator $\mathcal{R}_\rho(\widehat{\mu}_u^{*n})$ can be computed in finitely many steps even for tail-dependent functionals $\mathcal{R}_\rho$ as, for instance, the one associated with the Average Value at Risk of Example 1.2.4.

To justify the scheme (A.1)–(A.2) note that the empirical probability probability measure $\widehat{\mu}_u$ defined in (2.7) has the representation

$$\widehat{\mu}_u[\,\cdot\,] \;=\; \widehat{p}_u\,\widehat{\nu}_u[\,\cdot\,] + (1-\widehat{p}_u)\,\delta_0[\,\cdot\,],$$

where $\widehat{p}_u := \widehat{\mu}_u[(0,\infty)]$ is the mass of $\widehat{\mu}_u$ on $(0,\infty)$, and $\widehat{\nu}_u[\,\cdot\,] := \widehat{\mu}_u[\,\cdot \cap (0,\infty)]/\widehat{\mu}_u[(0,\infty)]$ is the probability measure $\widehat{\mu}_u$ conditioned on $(0,\infty)$. It is easily seen that the n-fold convolution $\widehat{\mu}_u^{*n}$ coincides with the random convolution

$$\widehat{\nu}_u^{*\mathrm{B}_{n,\widehat{p}_u}}[\,\cdot\,] \;:=\; \sum_{k=0}^{n}\widehat{\nu}_u^{*k}[\,\cdot\,]\,\mathrm{B}_{n,\widehat{p}_u}[\{k\}]$$

of $\widehat{\nu}_u$ w.r.t. the binomial distribution $\mathrm{B}_{n,\widehat{p}_u}$ with parameters n and $\widehat{p}_u$, i.e.

$$\widehat{\mu}_u^{*n} = \widehat{\nu}_u^{*\mathrm{B}_{n,\widehat{p}_u}}. \tag{A.3}$$

When $\widehat{p}_u < 1$ and $\widehat{\nu}_u$ has support in $h\mathbb{N} := \{h, 2h, \ldots\}$ for some $h > 0$, the random convolution $\widehat{\nu}_u^{*\mathrm{B}_{n,\widehat{p}_u}}$ can be computed with the help of the Panjer recursion [49]:

$$\widehat{\nu}_u^{*\mathrm{B}_{n,\widehat{p}_u}}[\{0\}] = \mathrm{B}_{n,\widehat{p}_u}[\{0\}] \tag{A.4}$$

$$\widehat{\nu}_u^{*\mathrm{B}_{n,\widehat{p}_u}}[\{jh\}] = \frac{\widehat{p}_u/j}{1-\widehat{p}_u}\sum_{\ell=1}^{j}[(n+1)\ell - j]\,\widehat{\nu}_u[\{\ell h\}]\,\widehat{\nu}_u^{*\mathrm{B}_{n,\widehat{p}_u}}[\{(j-\ell)h\}] \ \text{ for } j \in \mathbb{N}. \tag{A.5}$$

Since $1-\widehat{p}_u = \widehat{\mu}_u[\{0\}]$ and $\widehat{p}_u\widehat{\nu}_u[\{\ell h\}] = \widehat{\mu}_u[\{\ell h\}]$ for $\ell \in \mathbb{N} = \{1, 2, \ldots\}$, the recursive scheme (A.1)–(A.2) follows from (A.3)–(A.5).

A.2 On the computation of $\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}$ and $\mathcal{R}_\rho(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}})$

A computation of the random convolution of a measure μ w.r.t. a Poisson distribution is, just like in the case of the n-fold convolution, more or less impossible. However, if the single claim distribution μ has support in $h\mathbb{N}_0 := \{0, h, 2h, \ldots\}$ for some $h > 0$. In this case the convolution of the empirical measure w.r.t. the Poisson distribution with estimated parameter, which is given by

$$\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}} := (\widehat{\mu}_{n,\tau})^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}} \tag{A.6}$$

is nothing but a random convolution of an equidistant discrete distribution w.r.t. a Poisson distribution. Here $\widehat{\mu}_{n,\tau}$ and $\widehat{\lambda}_{n,\tau}$ are as in (3.4) and (3.3), respectively. The right-hand side in (A.6) can then be computed via

$$\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}[\{0\}] = \mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}[\{0\}] \tag{A.7}$$

$$\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}[\{jh\}] = \frac{\widehat{\lambda}_{n,\tau}T}{j}\sum_{m=1}^{j} m\,\widehat{\mu}_{n,\tau}[\{mh\}]\,\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}[\{(j-m)h\}], \tag{A.8}$$

for every $j \in \mathbb{N}$. Although $\widehat{\mu}_{n,\tau}$ has bounded support, the right-hand side in (A.6) has unbounded support. Therefore the corresponing plug-in estimator

$$\mathcal{R}_\rho\Big(\widehat{\mu}_{n,\tau}^{*\mathrm{Poiss}_{\widehat{\lambda}_{n,\tau}T}}\Big)$$

cannot be computed in finite time for tail-dependent functionals $\mathcal{R}_\rho$, such as the Average Value at Risk of Example 1.2.4, for instance. On the other hand, it can be computed in finite time for the Value at Risk of Example 1.2.3 for instance.

Appendix B

A proof of the $(\max\{d_{\mathrm{Wass}_p}, |\cdot|\}, d_{\mathrm{Wass}_p})$-continuity of the mapping $\mathcal{M}_1(L^p) \times (0,\infty) \to \mathcal{M}_1(L^p)$, $(\mu,\lambda) \mapsto \mu^{*\mathrm{Poiss}_\lambda}$

Let $p \in [1,\infty)$. Recall that the L^p-Wasserstein metric on $\mathcal{M}_1(L^p)$ is defined by

$$d_{\mathrm{Wass}_p}(\mu,\nu) := \Big(\int_0^1 |F_\mu^{-1}(t) - F_\nu^{-1}(t)|^p \, dt\Big)^{1/p}, \tag{B.1}$$

where for every $\mu \in \mathcal{M}_1$ and $t \in [0,1]$ $F_\mu^{-1}(t) := \inf\{y \in \mathbb{R}, F_\mu(y) \geq t\}$ denotes the generalized inverse of the distribution function F_μ associated with μ. It was shown in Lemma 8.1 in [13] that d_{Wass_p} defines a metric on $\mathcal{M}_1(L^p)$.

Theorem B.0.1 *Let $p \in [1,\infty)$. The mapping $\mathcal{M}_1(L^p) \times (0,\infty) \to \mathcal{M}_1(L^p)$, $(\mu,\lambda) \mapsto \mu^{*\mathrm{Poiss}_\lambda}$ is $(d_{\mathrm{Wass}_p}, d_{\mathrm{Wass}_p})$-continuous.*

Proof Let $\mu,\nu \in \mathcal{M}_1(L^p)$ and $\lambda,\beta \in (0,\infty)$. With the help of the triangle inequality we clearly have

$$d_{\mathrm{Wass}_p}(\mu^{*\mathrm{Poiss}_\lambda}, \nu^{*\mathrm{Poiss}_\beta}) \leq d_{\mathrm{Wass}_p}(\mu^{*\mathrm{Poiss}_\beta}, \nu^{*\mathrm{Poiss}_\beta}) + d_{\mathrm{Wass}_p}(\mu^{*\mathrm{Poiss}_\lambda}, \mu^{*\mathrm{Poiss}_\beta}). \tag{B.2}$$

For the first summand on the right-hand side of (B.2), we can use Lemma 8.5 in [13] to observe that

$$\begin{aligned} d_{\mathrm{Wass}_p}(\mu^{*\mathrm{Poiss}_\beta}, \nu^{*\mathrm{Poiss}_\beta}) &= \sum_{k=0}^{\infty} \mathrm{Poiss}_\beta[\{k\}] \, d_{\mathrm{Wass}_p}(\mu^{*k}, \nu^{*k}) \\ &\leq \sum_{k=0}^{\infty} k \, \mathrm{Poiss}_\beta[\{k\}] \, d_{\mathrm{Wass}_p}(\mu,\nu) \\ &= \beta \, d_{\mathrm{Wass}_p}(\mu,\nu). \end{aligned} \tag{B.3}$$

Hence, the first summand on the right-hand side of (B.2) converges to zero as $d_{\mathrm{Wass}_p}(\mu,\nu) \to 0$. Now the claim would follow by showing that the second summand on the right-hand side of (B.2) also converges to 0 as $|\lambda-\beta| \to 0$. To this end, we will use Lemma 8.3 in [13] to prove that $d_{\mathrm{Wass}_p}(\mu^{*\mathrm{Poiss}_\lambda}, \mu^{*\mathrm{Poiss}_\beta})$ converges to zero as $|\lambda-\beta| \to 0$. By Lemma 8.3 in [13] the claim would follow by showing that the following two assertions hold true:

(i) $|\int f(x)\mu^{*\mathrm{Poiss}_\lambda}(dx) - \int f(x)\mu^{*\mathrm{Poiss}_\beta}(x)| \longrightarrow 0$ for every continuous and bounded real-valued function f and $|\lambda-\beta| \to 0$.

(ii) $|\int |x|^p\mu^{*\mathrm{Poiss}_\lambda}(dx) - \int |x|^p\mu^{*\mathrm{Poiss}_\beta}(dx)| \longrightarrow 0$ for $|\lambda-\beta| \to 0$.

To prove part (i), let f be a bounded and continuous real-valued function. Then we can use the Mean Value Theorem to derive

$$
\begin{aligned}
&\Big|\int f(x)\mu^{*\mathrm{Poiss}_\lambda}(dx) - \int f(x)\mu^{*\mathrm{Poiss}_\beta}(dx)\Big| \\
&= \Big|\int f(x)\Big(\sum_{k=0}^{\infty}\mathrm{Poiss}_\lambda[\{k\}]\mu^{*k}\Big)(dx) - \int f(x)\Big(\sum_{k=0}^{\infty}\mathrm{Poiss}_\beta[\{k\}]\mu^{*k}\Big)(dx)\Big| \\
&\leq \sum_{k=0}^{\infty}|\mathrm{Poiss}_\lambda[\{k\}] - \mathrm{Poiss}_\beta[\{k\}]| \int f(x)\mu^{*k}(dx) \\
&\leq \|f\|_\infty \sum_{k=0}^{\infty}\left|\frac{\lambda^k}{k!}e^{-\lambda} - \frac{\beta^k}{k!}e^{-\beta}\right| \\
&= |\lambda-\beta|\,\|f\|_\infty \sum_{k=0}^{\infty}\frac{1}{k!}|k\,\alpha^{k-1} - \alpha^k|\,e^{-\alpha} \qquad \text{(B.4)}
\end{aligned}
$$

for some α between λ and β. As the series on the right-hand side of (B.4) converges, this shows that part (i) is fulfilled. For the second part, we use similar arguments as in (B.4) along with Minkowski's inequality to derive

$$
\begin{aligned}
\Big|\int |x|^p\mu^{*\mathrm{Poiss}_\lambda}(dx) - \int |x|^p\mu^{*\mathrm{Poiss}_\beta}(x)\Big| &\leq \sum_{k=0}^{\infty}\left|\frac{\lambda^k}{k!}e^{-\lambda} - \frac{\beta^k}{k!}e^{-\beta}\right| \int |x|^p\mu^{*k}(dx) \\
&\leq \int |x|^p\mu(dx)\,\sum_{k=0}^{\infty}k^p\,\left|\frac{\lambda^k}{k!}e^{-\lambda} - \frac{\beta^k}{k!}e^{-\beta}\right| \\
&= |\lambda-\beta|\int |x|^p\mu(dx)\,\sum_{k=0}^{\infty}\frac{k^p}{k!}|k\,\alpha^{k-1} - \alpha^k|\,e^{-\alpha}. \qquad \text{(B.5)}
\end{aligned}
$$

As the series on the right-hand side of (B.5) converges, we can conclude that part (ii) is also fulfilled. This yields the convergence of $d_{\mathrm{Wass}_p}(\mu^{*\mathrm{Poiss}_\lambda}, \mu^{*\mathrm{Poiss}_\beta})$ to zero as $|\lambda-\beta| \to 0$ and completes the proof. □

Appendix C

A nonuniform Berry-Esséen inequality

The following nonuniform Berry–Esséen inequality is already known from [46]. However, as the proof presented in [46] did not make it clear how the constants in the upper inequality had to be chosen, or how these constants depended on the distribution of the underlying random variables, we take our time to carry out the proof in a more rigorous way.

Theorem C.0.2 *Let* (X_i) *be a sequence of i.i.d. random variables on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *such that* $\mathbb{V}\mathrm{ar}[X_1] > 0$ *and* $\mathbb{E}[|X_1|^\lambda] < \infty$ *for some* $\lambda > 2$. *For every* $n \in \mathbb{N}$, *let*

$$Z_n := \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_1])}{\sqrt{n \mathbb{V}\mathrm{ar}[X_1]}}.$$

Then there exists a universal constant $C_\lambda \in (0, \infty)$ *such that*

$$d_{\phi_\lambda}(\mathbb{P}_{Z_n}, \mathcal{N}_{0,1}) \leq C_\lambda \, f(\mathbb{P}_{X_1}) \, n^{-\gamma} \quad \text{for all } n \in \mathbb{N} \tag{C.1}$$

with $\gamma := \min\{1, \lambda - 2\}/2$, *where for some universal constant* $D_\lambda > 0$,

$$f(\mathbb{P}_{X_1}) := \begin{cases} \frac{\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]}{\mathbb{V}\mathrm{ar}[X_1]^{\lambda/2}} & , \; 2 < \lambda \leq 3 \\ \exp\left(D_\lambda \frac{\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]^2}{\mathbb{V}\mathrm{ar}[X_1]^\lambda}\right) & , \; \lambda > 3 \end{cases}. \tag{C.2}$$

By "universal constant" we mean that the constant is independent of $\mathbb{P}_{X_1}$. Note that the constant $f(\mathbb{P}_{X_1})$ in Theorem (C.0.2) can still be improved. The formulation of Theorem 2.2.7 in the form of Petrov [50], for instance, allows for a better estimation of $f(\mathbb{P}_{X_1})$.

Proof (of Theorem C.0.2) As discussed above, the case $2 < \lambda \leq 3$ is already known. So we may and do assume $\lambda > 3$. In particular, for (C.1) it suffices to show

$$d_{\phi_\lambda}(\mathbb{P}_{Z_n'}, \mathcal{N}_{0,1}) \leq C_\lambda \exp(D_\lambda \mathbb{E}[|X_1'|^\lambda]^2) \, n^{-1/2} \quad \text{for all } n \in \mathbb{N} \tag{C.3}$$

for $Z_n' := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i'$ and any sequence (X_i') of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}[X_1'] = 0$, $\mathbb{V}\mathrm{ar}[X_1'] = 1$, and $\mathbb{E}[|X_1'|^\lambda] < \infty$. Indeed, if specifically

$X_i' := (X_i - \mathbb{E}[X_1])/\sqrt{\mathbb{V}\mathrm{ar}[X_1]}$ in the setting of Theorem C.0.2, then we have $Z_n = Z_n'$ and $\mathbb{E}[|X_1'|^\lambda] = \mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]/\mathbb{V}\mathrm{ar}[X_1]^{\lambda/2}$.

To verify (C.3), let F_n and $\Phi_{0,1}$ denote the distribution functions of Z_n' and the standard normal distribution, respectively. Below we will show in three steps that the inequalities

$$
\begin{aligned}
|F_n(x) - \Phi_{0,1}(x)| &\le c_\lambda\, \mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda/2-1)}(1+|x|^\lambda)^{-1} \quad \text{for } |x| < 1, && \text{(C.4)}\\
|F_n(x) - \Phi_{0,1}(x)| &\le c_\lambda\, e^{d_\lambda \mathbb{E}[|X_1'|^\lambda]^2}\, n^{-1/2}|x|^{-\lambda} \quad \text{for } 1 \le |x| \le \sqrt{(\lambda-1)\log n}, && \text{(C.5)}\\
|F_n(x) - \Phi_{0,1}(x)| &\le c_\lambda\, e^{d_\lambda \mathbb{E}[|X_1'|^\lambda]}\, n^{-(\lambda/2-1)}|x|^{-\lambda} \quad \text{for } |x| > \max\{1; \sqrt{(\lambda-1)\log n}\} && \text{(C.6)}
\end{aligned}
$$

hold for all $n \in \mathbb{N}$, where $c_\lambda, d_\lambda > 0$ refer to any constants depending only on λ and being independent of the distribution of X_1'. Inequalities (C.4)–(C.6) clearly imply (C.3).

Step 1. Inequality (C.4) follows from Katz' generalization of the classical Berry–Esséen inequality. In [36], Katz showed the following result. Let $g : \mathbb{R} \to (0,\infty)$ be any function that is even (i.e. $g(-x) = g(x)$ for all $x \in \mathbb{R}$), nondecreasing on $\mathbb{R}_+$ and satisfies $\lim_{x\to\infty} g(x) = \infty$ as well as $x/g(x) \le y/g(y)$ for all $0 \le x \le y$. Then for any sequence (Y_i) of of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}[Y_1] = 0$, $\mathbb{E}[Y_1^2] < \infty$ and $\mathbb{E}[Y_1^2 g(Y)] < \infty$ there exists an universal constant $C_g \in (0,\infty)$ (i.e. independent of $\mathbb{P}_{Y_1}$) such that

$$
d_{\phi_0}(\mathbb{P}_{W_n}, \mathcal{N}_{0,1}) \le (C_g \mathbb{E}[Y_1^2 g(Y_1)])\, g(\sqrt{n})^{-1} \qquad \text{for all } n \in \mathbb{N},
$$

where $W_n := \sum_{i=1}^n Y_i/\sqrt{n}$. Choosing specifically $g(x) := |x|^{\lambda-2}$ and $Y_i := X_i'$ for $i \in \mathbb{N}$, in particular $W_n = Z_n'$ for $n \in \mathbb{N}$, we easily obtain (C.4).

Step 2. We now prove (C.5). It suffices to show that there exists some constant $\widetilde{c}_\lambda > 0$ depending only on λ and being independent of the distribution of X_1' such that (C.5) holds for all $n \ge n_0 := \lceil \widetilde{c}_\lambda \mathbb{E}[|X_1'|^\lambda]^8 \rceil$ (this observation will be relevant in Steps 2.2.2 and 2.2.3 below). Indeed, for $n < n_0$ we get (C.5) from Katz' generalization of the classical Berry–Esséen inequality (cf. Step 1) as follows:

$$
\begin{aligned}
&\sup_{x \in [-\sqrt{(\lambda-1)\log n}, \sqrt{(\lambda-1)\log n}]} |F_n(x) - \Phi_{0,1}(x)|(1+|x|^\lambda)\\
&\le \sup_{x \in [-\sqrt{(\lambda-1)\log n_0}, \sqrt{(\lambda-1)\log n_0}]} |F_n(x) - \Phi_{0,1}(x)|(1+|x|^\lambda)\\
&\le \|F_n - \Phi_{0,1}\|_\infty\, (1 + ((\lambda-1)\log n_0)^\lambda)\\
&\le c_{\lambda,1}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda/2-1)}\, (1 + ((\lambda-1)\log(\lceil \widetilde{c}_\lambda \mathbb{E}[|X_1'|^\lambda]^8 \rceil))^\lambda)\\
&\le c_{\lambda,2}\, \mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2}.
\end{aligned}
$$

Without loss of generality we restrict ourselves to $1 \le x \le \max\{1; \sqrt{(\lambda-1)\log n}\}$. Let $r_\lambda \in (0, \min\{1; \lambda-3\}/(2(\lambda-1))]$, consider the truncations

$$
X_i^{n,x} := X_i'\, \mathbb{1}_{\{|X_i'| \le r_\lambda n^{1/2} x\}}, \qquad 1 \le i \le n, \quad n \in \mathbb{N},
$$

and set $\widetilde{Z}_n^{n,x} := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^{n,x}$.

We have

$$\begin{aligned}|F_n(x)-\Phi_{0,1}(x)| &= |(1-F_n(x))-(1-\Phi_{0,1}(x))|\\ &\le |\mathbb{P}[Z_n'>x]-\mathbb{P}[\widetilde{Z}_n^{n,x}>x]| + |\mathbb{P}[\widetilde{Z}_n^{n,x}>x]-(1-\Phi_{0,1}(x))|\\ &= |\mathbb{P}[Z_n'>x]-\mathbb{P}[\widetilde{Z}_n^{n,x}>x]| + |\mathbb{P}[\widetilde{Z}_n^{n,x}>x]-\Phi_{0,1}(-x))|. \qquad \text{(C.7)}\end{aligned}$$

In Steps 2.1–2.2 below we will show that

$$|\mathbb{P}[Z_n'>x]-\mathbb{P}[\widetilde{Z}_n^{n,x}>x]| \le c_{\lambda,1}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-(\lambda/2-1)}\,x^{-\lambda} \qquad \text{(C.8)}$$

and

$$|\mathbb{P}[\widetilde{Z}_n^{n,x}>x]-\Phi_{0,1}(-x))| \le c_{\lambda,2}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1/2}\,x^{-\lambda}. \qquad \text{(C.9)}$$

Then, (C.7)–(C.9) imply (C.5).

Step 2.1. To prove (C.8), note that

$$\begin{aligned}\mathbb{P}[Z_n'>x] &= \mathbb{P}[\{Z_n'>x\}\cap\{X_1'=X_1^{n,x},\dots,X_n'=X_n^{n,x}\}]\\ &\quad +\mathbb{P}[\{Z_n'>x\}\cap\{\text{there exists } 1\le i\le n \text{ with } X_i'\ne X_i^{n,x}\}]\\ &\le \mathbb{P}[\{\widetilde{Z}_n^{n,x}>x\}\cap\{X_1'=X_1^{n,x},\dots,X_n'=X_n^{n,x}\}] + n\,\mathbb{P}[X_1'\ne X_1^{n,x}]\\ &\le \mathbb{P}[\widetilde{Z}_n^{n,x}>x] + n\,\mathbb{P}[|X_1'|>r_\lambda n^{1/2}x] \qquad \text{(C.10)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}[\widetilde{Z}_n^{n,x}>x] &= \mathbb{P}[\{\widetilde{Z}_n^{n,x}>x\}\cap\{X_1'=X_1^{n,x},\dots,X_n'=X_n^{n,x}\}]\\ &\quad +\mathbb{P}[\{\widetilde{Z}_n^{n,x}>x\}\cap\{\text{there exists } 1\le i\le n \text{ with } X_i'\ne X_i^{n,x}\}]\\ &\le \mathbb{P}[\{\widetilde{Z}_n^{n,x}>x\}\cap\{X_1'=X_1^{n,x},\dots,X_n'=X_n^{n,x}\}] + n\,\mathbb{P}[X_1'\ne X_1^{n,x}]\\ &\le \mathbb{P}[Z_n'>x] + n\,\mathbb{P}[|X_1'|>r_\lambda n^{1/2}x]. \qquad \text{(C.11)}\end{aligned}$$

Then (C.10)–(C.11) and an application of Markov's inequality give

$$\begin{aligned}|\mathbb{P}[Z_n'>x]-\mathbb{P}[\widetilde{Z}_n^{n,x}>x]| &\le n\,\mathbb{P}[|X_1'|>r_\lambda n^{1/2}x]\\ &\le n\,\frac{\mathbb{E}[|X_1'|^\lambda]}{(r_\lambda n^{1/2}x)^\lambda}\\ &\le r_\lambda^{-\lambda}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-(\lambda/2-1)}\,x^{-\lambda}.\end{aligned}$$

That is, (C.8) holds for $c_{\lambda,1} := r_\lambda^{-\lambda}$.

Step 2.2. To verify (C.9) we consider the probability measure $\mathbb{Q}_{n,x}$ on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ defined by

$$\mathbb{Q}_{n,x}[A] := \frac{1}{\beta_{n,x}}\int_A e^{xn^{-1/2}x_1}\,\mathbb{P}_{X_1^{n,x}}(dx_1), \qquad A\in\mathcal{B}(\mathbb{R}),$$

where $\beta_{n,x} := \int e^{xn^{-1/2}x_1}\,\mathbb{P}_{X_1^{n,x}}(dx_1)$. In particular,

$$\frac{d\mathbb{P}_{X_1^{n,x}}}{d\mathbb{Q}_{n,x}}(x_1) = \beta_{n,x}\,e^{-xn^{-1/2}x_1} \qquad \text{for all } x_1\in\mathbb{R}.$$

It follows that the n-fold product measure $\mathbb{Q}_{n,x}^{\otimes n}$ of $\mathbb{Q}_{n,x}$ satisfies

$$\mathbb{Q}_{n,x}^{\otimes n}[A] \;=\; \frac{1}{\beta_{n,x}^n}\int_A e^{xn^{-1/2}\sum_{i=1}^n x_i}\,\mathbb{P}_{X_1^{n,x}}^{\otimes n}(d(x_1,\ldots,x_n)) \quad \text{for all } A\in\mathcal{B}(\mathbb{R}^n).$$

In particular,

$$\frac{d\mathbb{P}_{X_1^{n,x}}^{\otimes n}}{d\mathbb{Q}_{n,x}^{\otimes n}}(x_1,\ldots,x_n) \;=\; \beta_{n,x}^n\,e^{-xn^{-1/2}\sum_{i=1}^n x_i} \quad \text{for all } (x_1,\ldots,x_n)\in\mathbb{R}^n.$$

Using the notation

$$m_{n,x} \;:=\; \mathbb{E}_{\mathbb{Q}_{n,x}}[X_1^{n,x}] \;=\; \int x_1\,\mathbb{Q}_{n,x}(dx_1)$$

we obtain

$$\begin{aligned}
&\mathbb{P}[\widetilde{Z}_n^{n,x}>x]\\
&= \mathbb{P}\Big[n^{-1/2}\sum_{i=1}^n (X_i^{n,x}-m_{n,x}) > x-n^{1/2}m_{n,x}\Big]\\
&= \int \mathbb{1}_{(x-n^{1/2}m_{n,x},\infty)}\Big(n^{-1/2}\sum_{i=1}^n(x_i-m_{n,x})\Big)\,\mathbb{P}_{X_1^{n,x}}^{\otimes n}(d(x_1,\ldots,x_n))\\
&= \int \mathbb{1}_{(x-n^{1/2}m_{n,x},\infty)}\Big(n^{-1/2}\sum_{i=1}^n(x_i-m_{n,x})\Big)\,\beta_{n,x}^n\,e^{-xn^{-1/2}\sum_{i=1}^n x_i}\,\mathbb{Q}_{n,x}^{\otimes n}(d(x_1,\ldots,x_n))\\
&= \beta_{n,x}^n\,e^{-xn^{1/2}m_{n,x}}\times\\
&\qquad \int \mathbb{1}_{(x-n^{1/2}m_{n,x},\infty)}\Big(n^{-1/2}\sum_{i=1}^n(x_i-m_{n,x})\Big)\,e^{-xn^{-1/2}\sum_{i=1}^n(x_i-m_{n,x})}\,\mathbb{Q}_{n,x}^{\otimes n}(d(x_1,\ldots,x_n))\\
&= \beta_{n,x}^n\,e^{-xn^{1/2}m_{n,x}}\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\Pi_{n,x}(dz),
\end{aligned}$$

where $\Pi_{n,x}$ refers to the image (probability) measure of the probability measure $\mathbb{Q}_{n,x}^{\otimes n}$ w.r.t. the mapping $(x_1,\ldots,x_n)\mapsto n^{-1/2}\sum_{i=1}^n(x_i-m_{n,x})$. Hence, for the left-hand side in (C.9) we obtain

$$\begin{aligned}
&|\mathbb{P}[\widetilde{Z}_n^{n,x}>x]-\Phi_{0,1}(-x))|\\
&\quad= \Big|\beta_{n,x}^n\,e^{-xn^{1/2}m_{n,x}}\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\Pi_{n,x}(dz)-\Phi_{0,1}(-x))\Big|\\
&\quad\le |\beta_{n,x}^n\,e^{-xn^{1/2}m_{n,x}}-e^{-x^2/2}|\\
&\qquad\quad +e^{-x^2/2}\Big|\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\Pi_{n,x}(dz)-\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\mathcal{N}_{0,s_{n,x}^2}(dz)\Big|\\
&\qquad\quad +\Big|e^{-x^2/2}\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\mathcal{N}_{0,s_{n,x}^2}(dz)-\Phi_{0,1}(-x)\Big|\\
&\quad=: S_{2,2,1}(\lambda,n,x)+S_{2,2,2}(\lambda,n,x)+S_{2,2,3}(\lambda,n,x),
\end{aligned} \tag{C.12}$$

where

$$s_{n,x} := \mathbb{V}\mathrm{ar}_{\mathbb{Q}_{n,x}}[X_1^{n,x}]^{1/2} = \Big(\int x_1^2 \, \mathbb{Q}_{n,x}(dx_1) - m_{n,x}^2 \Big)^{1/2}.$$

In Steps 2.2.1–2.2.3 below we will show that

$$\begin{aligned} S_{2,2,1}(\lambda,n,x) &\le c_{\lambda,3}\, e^{c_{\lambda,4}\mathbb{E}[|X_1'|^\lambda]}\, n^{-1/2}\, x^{10}\, e^{-x^2/2}, && \text{(C.13)} \\ S_{2,2,2}(\lambda,n,x) &\le c_{\lambda,5}\, e^{c_{\lambda,6}\mathbb{E}[|X_1'|^\lambda]}\, n^{-1/2}\, x^{2}\, e^{-x^2/2}, && \text{(C.14)} \\ S_{2,2,3}(\lambda,n,x) &\le c_{\lambda,7}\, e^{c_{\lambda,8}\mathbb{E}[|X_1'|^\lambda]^2}\, n^{-1/2}\, x^{9}\, e^{-x^2/2}, && \text{(C.15)} \end{aligned}$$

which gives (C.9).

Step 2.2.0.a. First of all we observe that

$$\begin{aligned} |\mathbb{E}[X_1^{n,x}]| &\le c_{\lambda,9}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda-1)/2}\, x^{-(\lambda-1)}, && \text{(C.16)} \\ \mathbb{E}[(X_1^{n,x})^2] &\le 1, && \text{(C.17)} \\ \mathbb{E}[(X_1^{n,x})^2] &\ge 1 - c_{\lambda,10}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda-2)/2}\, x^{-(\lambda-2)}, && \text{(C.18)} \\ \mathbb{E}[|X_1'|^r] &\le \mathbb{E}[|X_1'|^\lambda] \quad \text{for } 2 \le r \le \lambda, && \text{(C.19)} \\ \mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}] &\le c_{\lambda,11}\, \mathbb{E}[|X_1'|^\lambda]\, n^{r_\lambda(\lambda-1)+(4-\lambda)/2}\, x^{4-\lambda} \quad \text{for } \lambda \in (3,4), && \text{(C.20)} \\ \mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}] &\le \mathbb{E}[|X_1'|^\lambda]\, n^{r_\lambda(\lambda-1)} \quad \text{for } \lambda \ge 4. && \text{(C.21)} \end{aligned}$$

Indeed: In view of $X_1^{n,x} = X_1' - X_1' \mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}$ and $\mathbb{E}[X_1'] = 0$ we have

$$\begin{aligned} |\mathbb{E}[X_1^{n,x}]| &= |\mathbb{E}[X_1' \mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}]| \\ &\le \mathbb{E}[|X_1'|\, \mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}] \\ &\le \mathbb{E}\Big[\frac{|X_1'|^\lambda}{(r_\lambda n^{1/2}x)^{\lambda-1}}\, \mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}\Big] \\ &\le r_\lambda^{1-\lambda}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda-1)/2}\, x^{-(\lambda-1)}, \end{aligned}$$

which proves (C.16) with $c_{\lambda,9} = r_\lambda^{1-\lambda}$. Inequality (C.17) is justified by

$$\mathbb{E}[(X_1^{n,x})^2] \le \mathbb{E}[(X_1')^2] = 1,$$

and inequality (C.18) can be obtained as follows:

$$\begin{aligned} \mathbb{E}[(X_1^{n,x})^2] - 1 &= \mathbb{E}[(X_1^{n,x})^2 - (X_1')^2] \\ &= -\mathbb{E}[(X_1')^2 \mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}] \\ &\ge -\mathbb{E}\Big[\frac{|X_1'|^\lambda}{(r_\lambda n^{1/2}x)^{\lambda-2}}\, \mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}\Big] \\ &\ge -r_\lambda^{2-\lambda}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda-2)/2}\, x^{-(\lambda-2)}. \end{aligned}$$

Due to the assumption $\mathbb{E}[|X_1'|^2] = 1$ and Jensen's inequality we obtain

$$1 = \mathbb{E}[|X_1'|^2]^{1/2} \le \mathbb{E}[|X_1'|^r]^{1/r} \le \mathbb{E}[|X_1'|^\lambda]^{1/\lambda},$$

which leads to (C.19) for $2 \le r \le \lambda$. Since $|X_1^{n,x}| \le r_\lambda n^{1/2} x$ and $x^2 \le (\lambda-1)\log n$, we obtain for $\lambda \in (3,4)$ that

$$\begin{aligned}
\mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}] &\le \mathbb{E}[(X_1^{n,x})^4]\, e^{r_\lambda x^2} \\
&\le \mathbb{E}[|X_1'|^\lambda]\,(r_\lambda n^{1/2} x)^{4-\lambda}\, n^{r_\lambda(\lambda-1)} \\
&= r_\lambda^{4-\lambda}\, x^{4-\lambda}\, n^{r_\lambda(\lambda-1)+(4-\lambda)/2}\, \mathbb{E}[|X_1'|^\lambda].
\end{aligned}$$

This proves (C.20) with $c_{\lambda,11} = r_\lambda^{4-\lambda}$. Finally, (C.21) follows by (C.19), $|X_1^{n,x}| \le r_\lambda n^{1/2} x$, and $x^2 \le (\lambda-1)\log n$.

Step 2.2.0.b. Next we will prove that the following auxiliary inequalities hold:

$$|\beta_{n,x} - 1 - x^2/(2n)| \le c_{\lambda,12}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-3/2} x^5, \tag{C.22}$$

$$|m_{n,x} - xn^{-1/2}| \le c_{\lambda,13}\, \mathbb{E}[|X_1'|^\lambda]\, n^{-1} x^6, \tag{C.23}$$

$$|s_{n,x}^2 - 1| \le c_{\lambda,14}\, \mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2} x^{12}. \tag{C.24}$$

We first show (C.22). Using (C.17), we obtain

$$\begin{aligned}
&|\beta_{n,x} - 1 - x^2/(2n)| \\
&= \Big|\mathbb{E}\Big[\sum_{i=0}^{\infty} \frac{(xn^{-1/2}X_1^{n,x})^i}{i!}\Big] - 1 - \frac{x^2}{2n}\Big| \\
&\le xn^{-1/2}|\mathbb{E}[X_1^{n,x}]| + \frac{x^2}{2n}(1 - \mathbb{E}[(X_1^{n,x})^2]) + \frac{1}{3!}(xn^{-1/2})^3 \mathbb{E}[|X_1^{n,x}|^3] \\
&\quad + \frac{1}{4!}(xn^{-1/2})^4\, \mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}].
\end{aligned}$$

On the one hand, for $\lambda \in (3,4)$ we can use (C.16), (C.18), (C.19), and (C.20) to conclude

$$\begin{aligned}
&|\beta_{n,x} - 1 - x^2/(2n)| \\
&\le c_{\lambda,9}\, x^{2-\lambda} n^{-\lambda/2} \mathbb{E}[|X_1'|^\lambda] + c_{\lambda,15}\, x^{4-\lambda} n^{-\lambda/2} \mathbb{E}[|X_1'|^\lambda] + \frac{1}{3!}\, x^3 n^{-3/2} \mathbb{E}[|X_1'|^\lambda] \\
&\quad + \frac{1}{4!}\, x^4 n^{-2}\, c_{\lambda,11} \mathbb{E}[|X_1'|^\lambda] n^{r_\lambda(\lambda-1)+(4-\lambda)/2} x^{4-\lambda} \\
&\le c_{\lambda,16} \mathbb{E}[|X_1'|^\lambda] n^{-3/2} x^3 + c_{\lambda,11} \frac{1}{4!}\, x^{8-\lambda}\, n^{-\lambda/2 + r_\lambda(\lambda-1)}\, \mathbb{E}[|X_1'|^\lambda] \\
&\le c_{\lambda,17} \mathbb{E}[|X_1'|^\lambda] n^{-3/2} x^5,
\end{aligned}$$

where for the last step we used $n^{-\lambda/2 + r_\lambda(\lambda-1)} \le n^{-3/2}$ (which follows from the assumption $r_\lambda \le (\lambda-3)/(2(\lambda-1))$). On the other hand, for $\lambda \ge 4$ we can use (C.16), (C.18), (C.19), and (C.21) to conclude

$$\begin{aligned}
&|\beta_{n,x} - 1 - x^2/(2n)| \\
&\le c_{\lambda,9} x^{2-\lambda} n^{-\lambda/2} \mathbb{E}[|X_1'|^\lambda] + c_{\lambda,15} x^{4-\lambda} n^{-\lambda/2} \mathbb{E}[|X_1'|^\lambda] + \frac{1}{3!}\, x^3 n^{-3/2} \mathbb{E}[|X_1'|^\lambda] \\
&\quad + \frac{1}{4!}\, x^4 n^{-2} \mathbb{E}[|X_1'|^\lambda]\, n^{r_\lambda(\lambda-1)}
\end{aligned}$$

$$
\begin{aligned}
&\le c_{\lambda,18}\mathbb{E}[|X_1'|^\lambda]n^{-3/2}x^3 + \frac{1}{4!}\,x^4 n^{-2}\mathbb{E}[|X_1'|^\lambda]\,n^{r_\lambda(\lambda-1)} \\
&\le c_{\lambda,19}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-3/2}x^5,
\end{aligned}
$$

where for the last step we used $n^{r_\lambda(\lambda-1)-2} \le n^{-3/2}$ (which follows from the assumption $r_\lambda \le 1/(2(\lambda-1))$). This completes the proof of (C.22).

To prove (C.23), we will show that the following inequalities hold:

$$
m_{n,x} - xn^{-1/2} \le c_{\lambda,20}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1}x^4, \tag{C.25}
$$

$$
xn^{-1/2} - m_{n,x} \le c_{\lambda,21}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1}x^6. \tag{C.26}
$$

For (C.25) we observe that since $\beta_{n,x} \ge 1$,

$$
\begin{aligned}
m_{n,x} - xn^{-1/2} &= \beta_{n,x}^{-1}\,\mathbb{E}[X_1^{n,x}e^{xn^{-1/2}X_1^{n,x}}] - xn^{-1/2} \\
&\le \mathbb{E}[X_1^{n,x}e^{xn^{-1/2}X_1^{n,x}}] - xn^{-1/2} \\
&\le |\mathbb{E}[X_1^{n,x}]| + xn^{-1/2}(\mathbb{E}[(X_1^{n,x})^2]-1) + \frac{1}{2}\,x^2n^{-1}\mathbb{E}[|X_1^{n,x}|^3] \\
&\quad + \frac{1}{3!}\,x^3n^{-3/2}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}].
\end{aligned}
$$

On the one hand, for $\lambda \in (3,4)$ we can use (C.16), (C.17), (C.19), and (C.20) to conclude

$$
\begin{aligned}
&m_{n,x} - xn^{-1/2} \\
&\le c_{\lambda,9}x^{1-\lambda}n^{(1-\lambda)/2}\mathbb{E}[|X_1'|^\lambda] + \frac{1}{2}\,x^2n^{-1}\mathbb{E}[|X_1'|^\lambda] + c_{\lambda,22}\,x^{7-\lambda}n^{(1-\lambda)/2+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda] \\
&\le c_{\lambda,23}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1}x^4,
\end{aligned}
$$

where for the last step we used $n^{(1-\lambda)/2+r_\lambda(\lambda-1)} \le n^{-1}$ (which follows from the assumption $r_\lambda \le (\lambda-3)/(2(\lambda-1))$). On the other hand, for $\lambda \ge 4$ we can use (C.16), (C.17), (C.19), and (C.21) to conclude

$$
\begin{aligned}
&m_{n,x} - xn^{-1/2} \\
&\le c_{\lambda,9}x^{1-\lambda}n^{(1-\lambda)/2}\mathbb{E}[|X_1'|^\lambda] + \frac{1}{2}\,x^2n^{-1}\mathbb{E}[|X_1'|^\lambda] + x^3n^{-3/2+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda] \\
&\le c_{\lambda,24}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1}x^3,
\end{aligned}
$$

where for the last step we used $n^{r_\lambda(\lambda-1)-2} \le n^{-3/2}$ (which follows from the assumption $r_\lambda \le 1/(2(\lambda-1))$). This proves (C.25). We will now prove (C.26). In view of (C.16), (C.18), (C.19), $\beta_{n,x} \ge 1$, $x^2 \le (\lambda-1)\log n$, and (C.22) we obtain

$$
\begin{aligned}
&xn^{-1/2} - m_{n,x} \\
&= xn^{-1/2} - \beta_{n,x}^{-1}\,\mathbb{E}[X_1^{n,x}e^{xn^{-1/2}X_1^{n,x}}] \\
&= xn^{-1/2} - \beta_{n,x}^{-1}\Big(\mathbb{E}[X_1^{n,x}] + xn^{-1/2}\mathbb{E}[(X_1^{n,x})^2] + x^2n^{-1}\mathbb{E}[(X_1^{n,x})^3]
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \mathbb{E}\Big[\sum_{i=3}^{\infty} \frac{(x\,n^{-1/2})^i}{i!}(X_1^{n,x})^{i+1}\Big]\Big) \\
&\le\; xn^{-1/2} \\
&\quad -\beta_{n,x}^{-1}\Big(- c_{\lambda,9}\,\mathbb{E}[|X_1'|^\lambda]n^{-(\lambda-1)/2}x^{-(\lambda-1)} \;+\; xn^{-1/2}(1 - c_{\lambda,10}\mathbb{E}[|X_1'|^\lambda]n^{-(\lambda-2)/2}x^{-(\lambda-2)}) \\
&\quad - x^2n^{-1}\mathbb{E}[|X_1'|^\lambda] \;-\; x^3n^{-3/2}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}]\Big) \\
&\le\; c_{\lambda,25}\,x^2n^{-1}\mathbb{E}[|X_1'|^\lambda] \;+\; x^3n^{-3/2}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}] \;+\; xn^{-1/2}(1-\beta_{n,x}^{-1}) \\
&\le\; c_{\lambda,25}\,x^2n^{-1}\mathbb{E}[|X_1'|^\lambda] \;+\; x^3n^{-3/2}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}] + xn^{-1/2}(\beta_{n,x}-1) \\
&\le\; c_{\lambda,25}\,x^2n^{-1}\mathbb{E}[|X_1'|^\lambda] + x^3n^{-3/2}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}] + xn^{-1/2}\Big(c_{\lambda,12}\,x^5n^{-3/2}\mathbb{E}[|X_1'|^\lambda] \\
&\quad + \frac{x^2}{2n}\Big) \\
&\le\; c_{\lambda,26}\,x^6n^{-1}\mathbb{E}[|X_1'|^\lambda] + x^3n^{-3/2}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}].
\end{aligned}
$$

For $\lambda \in (3,4)$ we can use (C.20) to deduce

$$
\begin{aligned}
xn^{-1/2} - m_{n,x} \;&\le\; c_{\lambda,26}\,x^6n^{-1}\mathbb{E}[|X_1'|^\lambda] + c_{\lambda,11}\,x^{7-\lambda}n^{1/2-\lambda/2+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda] \\
&\le\; c_{\lambda,27}\,x^6n^{-1}\mathbb{E}[|X_1'|^\lambda], \qquad\qquad (C.27)
\end{aligned}
$$

where for the last step we used $n^{1/2-\lambda/2+r_\lambda(\lambda-1)} \le n^{-1}$ (which follows from the assumption $r_\lambda \le (\lambda-3)/(2(\lambda-1))$). On the other hand for $\lambda \ge 4$ we can use (C.21) to obtain

$$
\begin{aligned}
xn^{-1/2} - m_{n,x} \;&\le\; c_{\lambda,26}\,x^6\,n^{-1}\mathbb{E}[|X_1'|^\lambda] + \; x^3n^{-3/2+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda] \\
&\le\; c_{\lambda,28}\,x^6n^{-1}\mathbb{E}[|X_1'|^\lambda], \qquad\qquad (C.28)
\end{aligned}
$$

where for the last step we used $n^{-3/2+r_\lambda(\lambda-1)} \le n^{-1}$ (which follows from the assumption $r_\lambda \le 1/(2(\lambda-1))$). Now (C.27) and (C.28) lead to (C.26).

To prove (C.24) we will show that the following inequalities hold:

$$
\begin{aligned}
s_{n,x}^2 - 1 \;&\le\; c_{\lambda,29}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1/2}x^3 \qquad\qquad (C.29) \\
1 - s_{n,x}^2 \;&\le\; c_{\lambda,30}\,\mathbb{E}[|X_1'|^\lambda]^2\,n^{-1/2}x^{12}. \qquad\qquad (C.30)
\end{aligned}
$$

First we will prove (C.29). By virtue of $\beta_{n,x} \ge 1$, (C.17), and (C.19), we obtain

$$
\begin{aligned}
s_{n,x}^2 - 1 \;&=\; (\beta_{n,x}^{-1}\mathbb{E}[(X_1^{n,x})^2e^{xn^{-1/2}X_1^{n,x}}] - m_{n,x}^2) - 1 \\
&\le\; \mathbb{E}[(X_1^{n,x})^2\,e^{xn^{-1/2}X_1^{n,x}}] - 1 \\
&\le\; \mathbb{E}[(X_1^{n,x})^2] - 1 \;+\; xn^{-1/2}\mathbb{E}[|X_1^{n,x}|^3] \;+\; x^2n^{-1}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}] \\
&\le\; xn^{-1/2}\,\mathbb{E}[|X_1'|^\lambda] \;+\; x^2n^{-1}\mathbb{E}[(X_1^{n,x})^4e^{xn^{-1/2}|X_1^{n,x}|}].
\end{aligned}
$$

For $\lambda \in (3,4)$ we can use (C.20) to deduce

$$
\begin{aligned}
s_{n,x}^2 - 1 \;&\le\; xn^{-1/2}\mathbb{E}[|X_1'|^\lambda] \;+\; c_{\lambda,11}\,x^3n^{1-\lambda/2+r_\lambda(\lambda-1)}\,\mathbb{E}[|X_1'|^\lambda] \\
&\le\; c_{\lambda,31}\,x^3n^{-1/2}\mathbb{E}[|X_1'|^\lambda],
\end{aligned}
$$

where for the last step we used $n^{1-\lambda/2+r_\lambda(\lambda-1)} \le n^{-1/2}$ (which follows from the assumption $r_\lambda \le (\lambda-3)/(2(\lambda-1))$). For $\lambda \ge 4$ we can use (C.21) to obtain

$$\begin{aligned} s_{n,x}^2 - 1 &\le xn^{-1/2}\mathbb{E}[|X_1'|^\lambda] + x^2 n^{-1+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda] \\ &\le c_{\lambda,32}\, x^2 n^{-1/2}\mathbb{E}[|X_1'|^\lambda], \end{aligned}$$

where for the last step we used $n^{-1+r_\lambda(\lambda-1)} \le n^{-1/2}$ (which follows from the assumption $r_\lambda \le 1/2(\lambda-1)$). This proves (C.29). We next prove (C.30). Using $\beta_{n,x} \ge 1$ and (C.23), we obtain

$$\begin{aligned} 1 - s_{n,x}^2 &= 1 - (\beta_{n,x}^{-1}\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] - m_{n,x}^2) \\ &= 1 + m_{n,x}^2 - \beta_{n,x}^{-1}\,\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] \\ &\le 1 + (c_{\lambda,13}\, n^{-1}x^6\mathbb{E}[|X_1'|^\lambda] + x^2 n^{-1/2})^2 - \beta_{n,x}^{-1}\,\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] \\ &\le 1 + c_{\lambda,33}\, n^{-1}x^{12}\mathbb{E}[|X_1'|^\lambda]^2 - \beta_{n,x}^{-1}\,\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}]. \end{aligned} \tag{C.31}$$

Now, (C.24) would follow if we can show that

$$\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] \ge 1 - c_{\lambda,34}\, n^{-1/2}x^3\mathbb{E}[|X_1'|^\lambda], \tag{C.32}$$

because (C.31)–(C.32) together with $\beta_{n,x} \ge 1$ and (C.22) imply

$$\begin{aligned} &1 - s_{n,x}^2 \\ &\quad\le 1 + c_{\lambda,33}\, n^{-1}x^{12}\,\mathbb{E}[|X_1'|^\lambda]^2 - \beta_{n,x}^{-1}(1 - c_{\lambda,34}\, n^{-1/2}x^3\mathbb{E}[|X_1'|^\lambda]) \\ &\quad\le c_{\lambda,33}\, n^{-1}x^{12}\,\mathbb{E}[|X_1'|^\lambda]^2 + c_{\lambda,34}\, n^{-1/2}x^3\mathbb{E}[|X_1'|^\lambda] + (1 - \beta_{n,x}^{-1}) \\ &\quad\le c_{\lambda,33}\, n^{-1}x^{12}\,\mathbb{E}[|X_1'|^\lambda]^2 + c_{\lambda,34}\, n^{-1/2}x^3\mathbb{E}[|X_1'|^\lambda] + (\beta_{n,x} - 1) \\ &\quad\le c_{\lambda,33}\, n^{-1}x^{12}\,\mathbb{E}[|X_1'|^\lambda]^2 + c_{\lambda,34}\, n^{-1/2}x^3\mathbb{E}[|X_1'|^\lambda] + c_{\lambda,12}\, n^{-3/2}x^5\mathbb{E}[|X_1'|^\lambda] + x^2/(2n) \\ &\quad\le c_{\lambda,35}\,\mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2}x^{12}. \end{aligned}$$

To prove (C.32) we use (C.18) and (C.19) to obtain

$$\begin{aligned} &\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] \\ &\quad= \mathbb{E}[(X_1^{n,x})^2] + xn^{-1/2}\mathbb{E}[(X_1^{n,x})^3] + \mathbb{E}\Big[\sum_{i=2}^\infty \frac{(xn^{-1/2})^i}{i!}(X_1^{n,x})^{i+2}\Big] \\ &\quad\ge 1 - c_{\lambda,10}\, n^{-(\lambda-2)/2}x^{-(\lambda-2)}\mathbb{E}[|X_1'|^\lambda] - xn^{-1/2}\mathbb{E}[|X_1'|^\lambda] - x^2 n^{-1}\mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}] \\ &\quad\ge 1 - c_{\lambda,36}\, xn^{-1/2}\mathbb{E}[|X_1'|^\lambda] - x^2 n^{-1}\mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}]. \end{aligned}$$

If $\lambda \in (3,4)$ we can use (C.20) to deduce

$$\begin{aligned} \mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] &\ge 1 - c_{\lambda,36}\, xn^{-1/2}\mathbb{E}[|X_1'|^\lambda] - c_{\lambda,11}\, x^{6-\lambda}\, n^{-1+r_\lambda(\lambda-1)-(\lambda-4)/2}\mathbb{E}[|X_1'|^\lambda] \\ &\ge 1 - c_{\lambda,37}\, x^3 n^{-1/2}\mathbb{E}[|X_1'|^\lambda], \end{aligned}$$

where for the last step we used $n^{1-\lambda/2+r_\lambda(\lambda-1)} \leq n^{-1/2}$ (which follows from the assumption $r_\lambda \leq (\lambda-3)/(2(\lambda-1))$). For $\lambda \geq 4$ we can use (C.21), yielding

$$\begin{aligned}\mathbb{E}[(X_1^{n,x})^2 e^{xn^{-1/2}X_1^{n,x}}] &\geq 1 - c_{\lambda,36}\, xn^{-1/2}\mathbb{E}[|X_1'|^\lambda] - x^2 n^{r_\lambda(\lambda-1)-1}\mathbb{E}[|X_1'|^\lambda] \\ &\geq 1 - c_{\lambda,38}\, x^2 n^{-1/2}\mathbb{E}[|X_1'|^\lambda],\end{aligned}$$

where for the last step we used $n^{-1+r_\lambda(\lambda-1)} \leq n^{-1/2}$ (which follows from the assumption $r_\lambda \leq 1/(2(\lambda-1))$). This proves (C.32).

Step 2.2.1. In this part we will verify the inequalities

$$\beta_{n,x}^n\, e^{-xn^{1/2}m_{n,x}} - e^{-x^2/2} \leq c_{\lambda,39}\, n^{-1/2}x^7 e^{-x^2/2} e^{c_{\lambda,40}\mathbb{E}[|X_1'|^\lambda]}, \tag{C.33}$$

$$e^{-x^2/2} - \beta_{n,x}^n\, e^{-xn^{1/2}m_{n,x}} \leq c_{\lambda,41}\, n^{-1/2}x^{10} e^{-x^2/2} e^{c_{\lambda,42}\mathbb{E}[|X_1'|^\lambda]}, \tag{C.34}$$

which imply (C.13). First we will show (C.33). Using the inequality $\log(\beta_{n,x}) \leq \beta_{n,x} - 1$ (which is valid in our case as we have $\beta_{n,x} \geq 1$), the Mean Value theorem, (C.22), (C.23), and the assumption $x^2 \leq (1-\lambda)\log n$ we obtain

$$\begin{aligned}&\beta_{n,x}^n\, e^{-xn^{1/2}m_{n,x}} - e^{-x^2/2} \\
&\leq e^{n(\beta_{n,x}-1)-xn^{1/2}m_{n,x}} - e^{-x^2/2} \\
&= e^{-(xn^{1/2}m_{n,x}-n(\beta_{n,x}-1))} - e^{-x^2/2} \\
&\leq (x^2/2 - (xn^{1/2}m_{n,x} - n(\beta_{n,x}-1)))\, e^{-(xn^{1/2}m_{n,x}-n(\beta_{n,x}-1))} \\
&= (x^2/2 - xn^{1/2}m_{n,x} + n(\beta_{n,x}-1))\, e^{n(\beta_{n,x}-1)-xn^{1/2}m_{n,x}} \\
&\leq \Big(x^2/2 - xn^{1/2}(-c_{\lambda,13}\, n^{-1}x^6\mathbb{E}[|X_1'|^\lambda] + xn^{-1/2}) + n(c_{\lambda,12}\, n^{-3/2}x^5\mathbb{E}[|X_1'|^\lambda] + x^2/(2n))\Big) \\
&\quad \cdot e^{n(c_{\lambda,12}\, n^{-3/2}x^5\mathbb{E}[|X_1'|^\lambda]+x^2/(2n))-xn^{1/2}(xn^{-1/2}-c_{\lambda,13}\, n^{-1}x^6\mathbb{E}[|X_1'|^\lambda])} \\
&\leq \Big(x^2/2 + c_{\lambda,13}\, n^{-1/2}x^7\mathbb{E}[|X_1'|^\lambda] - x^2 + c_{\lambda,12}\, n^{-1/2}x^5\mathbb{E}[|X_1'|^\lambda] + x^2/2\Big) \\
&\quad \cdot e^{c_{\lambda,12}\, n^{-1/2}x^5\mathbb{E}[|X_1'|^\lambda]+x^2/2-x^2+c_{\lambda,13}\, n^{-1/2}x^7\mathbb{E}[|X_1'|^\lambda])} \\
&\leq c_{\lambda,43}\, n^{-1/2}x^7\mathbb{E}[|X_1'|^\lambda]\, e^{-x^2/2}\, e^{c_{\lambda,44}\, n^{-1/2}((1-\lambda)\log n)^{7/2}\mathbb{E}[|X_1'|^\lambda]} \\
&\leq c_{\lambda,43}\, n^{-1/2}x^7\mathbb{E}[|X_1'|^\lambda]\, e^{-x^2/2}\, e^{c_{\lambda,44}\mathbb{E}[|X_1'|^\lambda]},\end{aligned}$$

where we assumed without loss of generality $xn^{1/2}m_{n,x} - n(\beta_{n,x}-1) \leq x^2/2$ (otherwise we obtain the trivial upper bound 0). Next we will show that (C.34) holds true. Using the inequality $1 - e^{-z} \leq z$ for $z \geq 0$, the inequality $\log(\beta_{n,x}) \geq \beta_{n,x} - 1 - \frac{1}{2}(\beta_{n,x}-1)^2$ as well as (C.22), (C.23), and $\beta_{n,x} \geq 1$ we obtain

$$\begin{aligned}&e^{-x^2/2} - \beta_{n,x}^n\, e^{-xn^{1/2}m_{n,x}} \\
&= e^{-x^2/2}(1 - e^{x^2/2+n\log\beta_{n,x}-x\,n^{1/2}\,m_{n,x}}) \\
&= e^{-x^2/2}(1 - e^{-(xn^{1/2}m_{n,x}-x^2/2-n\log\beta_{n,x})}) \\
&\leq e^{-x^2/2}(xn^{1/2}m_{n,x} - x^2/2 - n\log\beta_{n,x}) \\
&\leq e^{-x^2/2}\Big(xn^{1/2}(c_{\lambda,13}\, n^{-1}x^6\mathbb{E}[|X_1'|^\lambda] + xn^{-1/2}) - x^2/2 - n(\beta_{n,x} - 1 - (\beta_{n,x}-1)^2/2\Big)\end{aligned}$$

$$
\begin{aligned}
&\le\; e^{-x^2/2}\Big(c_{\lambda,13}\,n^{-1/2}x^7\mathbb{E}[|X_1'|^\lambda]+x^2/2+n\{c_{\lambda,12}\,n^{-3/2}x^5\mathbb{E}[|X_1'|^\lambda]-x^2/(2n))\}\\
&\qquad +n\{c_{\lambda,12}\,n^{-3/2}x^5\mathbb{E}[|X_1'|^\lambda]+x^2/(2n))\}^2/2\Big)\\
&\le\; e^{-x^2/2}\Big(c_{\lambda,13}\,n^{-1/2}x^7\mathbb{E}[|X_1'|^\lambda]+c_{\lambda,12}\,n^{-1/2}x^5\mathbb{E}[|X_1'|^\lambda]+c_{\lambda,12}^2\,n^{-2}x^{10}\mathbb{E}[|X_1'|^\lambda]^2+\frac{x^4}{4n}\Big)\\
&\le\; c_{\lambda,45}\,e^{-x^2/2}\,n^{-1/2}x^{10}e^{c_{\lambda,46}\mathbb{E}[|X_1'|^\lambda]},
\end{aligned}
$$

where we assumed without loss of generality $xn^{1/2}m_{n,x}-n\log\beta_{n,x}\ge x^2/2$ (otherwise we obtain the trivial upper bound 0). Note that this assumption allows us to apply the inequality $1-e^{-z}\le z$ for $z\ge 0$, to the third line in the upper calculations. This proves (C.34).

Step 2.2.2. Next we will prove (C.14). Let $F_{\Pi_{n,x}}$ denote the distribution function of $\Pi_{n,x}$, and note that $F_{\Pi_{n,x}}$, $\Phi_{0,s_{n,x}^2}$, and Ψ_x are of bounded variation on every right-sided half-line, where $\Psi_x(z):=e^{-xz}$. So integration-by-parts yields

$$
\begin{aligned}
&\Big|\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\Pi_{n,x}(dz)-\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\mathcal{N}_{0,s_{n,x}^2}(dz)\Big|\\
&=\;\Big|\int_{x-n^{1/2}m_{n,x}}^{\infty}\Psi_x(z)\,dF_{\Pi_{n,x}}(z)-\int_{x-n^{1/2}m_{n,x}}^{\infty}\Psi_x(z)\,d\Phi_{0,s_{n,x}^2}(z)\Big|\\
&=\;\Big|\lim_{b\to\infty}\Big(e^{-xb}\,F_{\Pi_{n,x}}(b)-e^{-x(x-n^{1/2}m_{n,x})}\,F_{\Pi_{n,x}}(x-n^{1/2}m_{n,x})\\
&\qquad\qquad -\int_{x-n^{1/2}m_{n,x}}^{b}F_{\Pi_{n,x}}(z)\,d\Psi_x(z)\Big)\\
&\qquad -\lim_{b\to\infty}\Big(e^{-xb}\,\Phi_{0,s_{n,x}^2}(b)+e^{-x(x-n^{1/2}m_{n,x})}\,\Phi_{0,s_{n,x}^2}(x-n^{1/2}m_{n,x})\\
&\qquad\qquad -\int_{x-n^{1/2}m_{n,x}}^{b}\Phi_{0,s_{n,x}^2}(z)\,d\Psi_x(z)\Big)\Big|\\
&\le\; e^{-x(x-n^{1/2}m_{n,x})}\,|F_{\Pi_{n,x}}(x-n^{1/2}m_{n,x})-\Phi_{0,s_{n,x}^2}(x-n^{1/2}m_{n,x})|\\
&\qquad +\int_{x-n^{1/2}m_{n,x}}^{\infty}|F_{\Pi_{n,x}}(z)-\Phi_{0,s_{n,x}^2}(z)|\,d\Psi_x(z)\\
&\le\; e^{-x(x-n^{1/2}m_{n,x})}\,\|F_{\Pi_{n,x}}-\Phi_{0,s_{n,x}^2}\|_\infty+\|F_{\Pi_{n,x}}-\Phi_{0,s_{n,x}^2}\|_\infty\Big|\int_{x-n^{1/2}m_{n,x}}^{\infty}d\Psi_x(z)\Big|\\
&=\; 2e^{-x(x-n^{1/2}m_{n,x})}\,\|F_{\Pi_{n,x}}-\Phi_{0,s_{n,x}^2}\|_\infty. \qquad\qquad (C.35)
\end{aligned}
$$

Furthermore we observe that

$$
\|F_{\Pi_{n,x}}-\Phi_{0,s_{n,x}^2}\|_\infty\;=\;\sup_{y\in\mathbb{R}}|F_{\Pi_{n,x}}(s_{n,x}\,y)-\Phi_{0,s_{n,x}^2}(s_{n,x}\,y)|\;=\;\|F_{\widetilde{\Pi}_{n,x}}-\Phi_{0,1}\|_\infty,
$$

where $\widetilde{\Pi}_{n,x}$ refers to the image measure of of the probability measure $\mathbb{Q}_{n,x}^{\otimes n}$ w.r.t. the mapping $(x_1,\dots,x_n)\mapsto n^{-1/2}\sum_{i=1}^n(x_i-m_{n,x})/s_{n,x}$. Hence by the classical Berry–Esséen theorem we have

$$
\|F_{\widetilde{\Pi}_{n,x}}-\Phi_{0,1}\|_\infty\;\le\;\frac{\int|x_1-m_{n,x}|^3\,\mathbb{Q}_{n,x}(dx_1)}{n^{1/2}s_{n,x}^3}
$$

$$\leq\ 4\,\frac{\int |x_1|^3\,\mathbb{Q}_{n,x}(dx_1) - m_{n,x}^3}{n^{1/2} s_{n,x}^3}$$

$$\leq\ 8\,\frac{\beta_{n,x}^{-1}\,\mathbb{E}[|X_1^{n,x}|^3 e^{xn^{-1/2}X_1^{n,x}}]}{n^{1/2}\,s_{n,x}^3}$$

$$\leq\ 8\,\frac{\mathbb{E}[|X_1^{n,x}|^3 e^{xn^{-1/2}X_1^{n,x}}]}{n^{1/2}\,s_{n,x}^3}. \tag{C.36}$$

Now for the numerator in (C.36) we can use (C.19) to obtain

$$\begin{aligned}\mathbb{E}[|X_1^{n,x}|^3\,e^{xn^{-1/2}X_1^{n,x}}] &= \mathbb{E}\Big[\sum_{i=0}^{\infty}\frac{(xn^{-1/2})^i}{i!}\,|X_1^{n,x}|^{i+3}\Big]\\ &\leq \mathbb{E}[|X_1^{n,x}|^3] + xn^{-1/2}\mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}]\\ &\leq \mathbb{E}[|X_1'|^\lambda] + xn^{-1/2}\mathbb{E}[(X_1^{n,x})^4 e^{xn^{-1/2}|X_1^{n,x}|}].\end{aligned}$$

Now for $\lambda\in(3,4)$ we can use (C.20) to deduce

$$\begin{aligned}\mathbb{E}[|X_1^{n,x}|^3 e^{xn^{-1/2}X_1^{n,x}}] &\leq \mathbb{E}[|X_1'|^\lambda] + c_{\lambda,11}\,x^{5-\lambda}n^{3/2-\lambda/2+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda]\\ &\leq c_{\lambda,47}\,x^2\mathbb{E}[|X_1'|^\lambda],\end{aligned}$$

where for the last step we used $n^{3/2-\lambda/2+r_\lambda(\lambda-1)}\leq 1$ (which follows from the assumption $r_\lambda\leq(\lambda-3)/(2(\lambda-1))$). For $\lambda\geq 4$ we can use (C.21) to obtain

$$\begin{aligned}\mathbb{E}[|X_1^{n,x}|^3 e^{xn^{-1/2}X_1^{n,x}}] &\leq \mathbb{E}[|X_1'|^\lambda] + xn^{-1/2+r_\lambda(\lambda-1)}\mathbb{E}[|X_1'|^\lambda]\\ &\leq c_{\lambda,48}\,x\,\mathbb{E}[|X_1'|^\lambda],\end{aligned}$$

where for the last step we used $n^{-1/2+r_\lambda(\lambda-1)}\leq 1$ (which follows from the assumption $r_\lambda\leq 1/2(\lambda-1)$). This proves

$$\mathbb{E}[|X_1^{n,x}|^3\,e^{xn^{-1/2}\,X_1^{n,x}}]\ \leq\ c_{\lambda,49}\,x^2\,\mathbb{E}[|X_1'|^\lambda]. \tag{C.37}$$

Furthermore we will assume without loss of generality that n is chosen sufficiently large such that $s_{n,x}^2\geq 1/2$. By (C.24) we have $|s_{n,x}^2-1|\leq c_{\lambda,14}\,\mathbb{E}[|X_1'|^\lambda]^2\,n^{-1/2}\,x^{12}$, i.e. it suffices to assume that $n\geq 4\,c_{\lambda,14}^2\,\mathbb{E}[|X_1'|^\lambda]^4\,x^{24}$. In view of $x^2\leq(\lambda-1)\log(n)$ this assumption holds if there is some constant $c_\lambda>0$ such that $n\geq n_0$ for $n_0:=\lceil c_\lambda\mathbb{E}[|X_1'|^\lambda]^8\rceil$. Recall from the discussion at the beginning of Step 2 that the assumption $n\geq n_0$ (for this specific choice of n_0) does not lead to any loss of generality. This and (C.35)–(C.37) lead to

$$\begin{aligned}&e^{-x^2/2}\Big|\int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\Pi_{n,x}(dz) - \int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\,\mathcal{N}_{0,s_{n,x}^2}(dz)\Big|\\ \leq\ & e^{-x^2/2}\,e^{-x(x-n^{1/2}m_{n,x})}\,c_{\lambda,49}\,x^2\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1/2}s_{n,x}^{-3}\\ \leq\ & e^{-x^2/2}\,e^{c_{\lambda,13}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1/2}\,x^7}\,c_{\lambda,49}\,2\sqrt{2}\,x^2\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1/2}\\ \leq\ & c_{\lambda,50}\,e^{c_{\lambda,13}\,\mathbb{E}[|X_1'|^\lambda]\,n^{-1/2}\,((\lambda-1)\log n)^{7/2}}\,n^{-1/2}\,x^2e^{-x^2/2}\,\mathbb{E}[|X_1'|^\lambda]\\ \leq\ & c_{\lambda,50}\,n^{-1/2}\,x^2e^{-x^2/2}\,e^{c_{\lambda,51}\,\mathbb{E}[|X_1'|^\lambda]}.\end{aligned}$$

This proves (C.14).

Step 2.2.3. Finally we will show (C.15). With the transformation $a := zs_{n,x}^{-1} + xs_{n,x}$ we have

$$\begin{aligned}
& e^{-x^2/2} \int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\, \mathcal{N}_{0,s_{n,x}^2}(dz) \\
&= e^{-x^2/2}\,(2\pi s_{n,x}^2)^{-1/2} \int_{x-n^{1/2}m_{n,x}}^{\infty} e^{-xz-z^2/(2s_{n,x}^2)}\, dz \\
&= e^{-x^2/2}\,(2\pi s_{n,x}^2)^{-1/2}\, e^{x^2 s_{n,x}^2/2} \int_{x-n^{1/2}m_{n,x}}^{\infty} e^{-(z/s_{n,x}+xs_{n,x})^2/2}\, dz \\
&= (2\pi s_{n,x}^2)^{-1/2}\, e^{-x^2(1-s_{n,x}^2)/2} \int_{(x-n^{1/2}m_{n,x})/s_{n,x}+xs_{n,x}}^{\infty} e^{-a^2/2}\, s_{n,x}\, da \\
&= e^{-x^2(1-s_{n,x}^2)/2} \int_{(x-n^{1/2}m_{n,x})/s_{n,x}+xs_{n,x}}^{\infty} (2\pi)^{-1/2} e^{-a^2/2}\, da \\
&= e^{-x^2(1-s_{n,x}^2)/2}\, \{1-\Phi_{0,1}(s_{n,x}^{-1} n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x})\} \\
&= e^{-x^2(1-s_{n,x}^2)/2}\, \Phi_{0,1}(-\{s_{n,x}^{-1} n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x}\}).
\end{aligned}$$

This leads to

$$\begin{aligned}
& \left| e^{-x^2/2} \int_{(x-n^{1/2}m_{n,x},\infty)} e^{-xz}\, \mathcal{N}_{0,s_{n,x}^2}(dz) - \Phi_{0,1}(-x) \right| \\
&= |e^{-x^2(1-s_{n,x}^2)/2} \Phi_{0,1}(-\{s_{n,x}^{-1} n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x}\}) - \Phi_{0,1}(-x)| \\
&\leq e^{-x^2(1-s_{n,x}^2)/2}\, |\Phi_{0,1}(-\{s_{n,x}^{-1} n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x}\}) - \Phi_{0,1}(-xs_{n,x})| \\
&\quad + |e^{-x^2(1-s_{n,x}^2)/2} - 1|\, \Phi_{0,1}(-xs_{n,x}) \\
&\quad + |\Phi_{0,1}(-xs_{n,x}) - \Phi_{0,1}(-x)| \\
&=: S_{2,2,3,1}(\lambda,n,x) + S_{2,2,3,2}(\lambda,n,x) + S_{2,2,3,3}(\lambda,n,x).
\end{aligned}$$

We will now show that the following inequalities are valid:

$$\begin{aligned}
S_{2,2,3,1}(\lambda,n,x) &\leq c_{\lambda,51}\, n^{-1/2}\, x^6 e^{-x^2/4}\, e^{c_{\lambda,52}\mathbb{E}[|X_1'|^\lambda]^2}, && \text{(C.38)} \\
S_{2,2,3,2}(\lambda,n,x) &\leq c_{\lambda,53}\, n^{-1/2}\, x^{14} e^{-x^2/4}\, e^{c_{\lambda,54}\mathbb{E}[|X_1'|^\lambda]^2}, && \text{(C.39)} \\
S_{2,2,3,3}(\lambda,n,x) &\leq c_{\lambda,55}\, n^{-1/2}\, x^{13} e^{-x^2/4}\, e^{c_{\lambda,56}\mathbb{E}[|X_1'|^\lambda]^2}, && \text{(C.40)}
\end{aligned}$$

which will lead to (C.15) immediately. We will start with the proof of (C.39). First, by the Mean Value theorem, (C.24), and $x^2 \leq (\lambda-1)\log n$ we have

$$\begin{aligned}
|e^{-x^2(1-s_{n,x}^2)/2} - 1| &\leq e^{x^2|1-s_{n,x}^2|/2} - 1 \\
&\leq (x^2|1-s_{n,x}^2|/2)\, e^{x^2|1-s_{n,x}^2|/2} \\
&\leq c_{\lambda,57}\mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2} x^{14}\, e^{c_{\lambda,14}\mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2} x^{14}/2} \\
&\leq c_{\lambda,58}\, n^{-1/2} x^{14}\, e^{c_{\lambda,59}\mathbb{E}[|X_1'|^\lambda]^2}. && \text{(C.41)}
\end{aligned}$$

In particular, using $x^2 \le (\lambda-1)\log n$ again,

$$e^{x^2|1-s_{n,x}^2|/2} \le 1+|e^{x^2|1-s_{n,x}^2|/2}-1| \le 1+c_{\lambda,58}e^{c_{\lambda,59}\mathbb{E}[|X_1'|^\lambda]^2} \le c_{\lambda,60}e^{c_{\lambda,59}\mathbb{E}[|X_1'|^\lambda]^2}. \quad \text{(C.42)}$$

First, (C.39) is a consequence of (C.41) and

$$\begin{aligned}
\Phi_{0,1}(-xs_{n,x}) &= \int_{-\infty}^{-xs_{n,x}} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, dy \\
&\le e^{-x^2 s_{n,x}^2/4}\, \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/4}\, dy \\
&= e^{-(s_{n,x}^2-1+1)x^2/4}\,\sqrt{2} \\
&\le \sqrt{2}\, e^{|s_{n,x}^2-1|x^2/4} e^{-x^2/4} \\
&\le \sqrt{2}\, c_{\lambda,60}e^{c_{\lambda,59}\mathbb{E}[|X_1'|^\lambda]^2}\, e^{-x^2/4},
\end{aligned}$$

where we used (C.42) for the latter step. We next prove (C.38). We will assume without loss of generality that n is chosen sufficiently large such that $n^{1/2}|m_{n,x}-xn^{-1/2}| \le 1/4$ and $s_{n,x}^2 \ge 1/2$. By (C.23) and (C.24) we have $|m_{n,x}-xn^{-1/2}| \le c_{\lambda,13}\mathbb{E}[|X_1'|^\lambda]n^{-1}x^6$ and $|s_{n,x}^2-1| \le c_{\lambda,14}\,\mathbb{E}[|X_1'|^\lambda]^2 n^{-1/2}x^{12}$, i.e. it suffices to assume that $n \ge (16c_{\lambda,13}^2)\mathbb{E}[|X_1'|^\lambda]^2 x^{12}$ and $n \ge (4c_{\lambda,14}^2)\mathbb{E}[|X_1'|^\lambda]^4 x^{24}$. In view of $x^2 \le (\lambda-1)\log n$, these assumptions holds if $n \ge ((16c_{\lambda,12}^2)\vee(4c_{\lambda,13}^2))\mathbb{E}[|X_1'|^\lambda]^4((\lambda-1)\log n))^{12}$. That is, there is some constant $\widetilde{c}_\lambda > 0$ such that $n \ge n_0$ for $n_0 := \lceil \widetilde{c}_\lambda \mathbb{E}[|X_1'|^\lambda]^8\rceil$ implies $n^{1/2}|m_{n,x}-xn^{-1/2}| \le 1/4$ and $s_{n,x}^2 \ge 1/2$ for all $1 \le |x| \le \sqrt{(\lambda-1)\log n}$. Recall from the discussion at the beginning of Step 2 that the assumption $n \ge n_0$ (for this specific choice of n_0) indeed does not lead to any loss of generality. Now, using (C.23) and (C.42) we obtain

$$\begin{aligned}
S_{2,2,3,1}(\lambda,n,x) &= e^{-x^2(1-s_{n,x}^2)/2}|\Phi_{0,1}(s_{n,x}^{-1}n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x})-\Phi_{0,1}(xs_{n,x})| \\
&= e^{-x^2(1-s_{n,x}^2)/2}\,\frac{1}{\sqrt{2\pi}}\int_{a_{n,x}}^{b_{n,x}} e^{-y^2/2}\,dy \\
&\le e^{-x^2(1-s_{n,x}^2)/2}\,\frac{1}{\sqrt{2\pi}}\,(b_{n,x}-a_{n,x}) \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&= e^{x^2|1-s_{n,x}^2|/2}\,\frac{1}{\sqrt{2\pi}}\, s_{n,x}^{-1}n^{1/2}|xn^{-1/2}-m_{n,x}| \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&\le e^{x^2|1-s_{n,x}^2|/2}\,\frac{1}{\sqrt{2\pi}}\,\sqrt{2}\,n^{1/2}c_{\lambda,13}\mathbb{E}[|X_1'|^\lambda]\,n^{-1}x^6 \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&= c_{\lambda,60}e^{c_{\lambda,59}\mathbb{E}[|X_1'|^\lambda]^2}\,\frac{1}{\sqrt{\pi}}\, n^{-1/2}c_{\lambda,13}\mathbb{E}[|X_1'|^\lambda]\,x^6 \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2}, \quad \text{(C.43)}
\end{aligned}$$

where $a_{n,x}$ and $b_{n,x}$ refer to respectively the minimum and the maximum of the real numbers $s_{n,x}^{-1}n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x}$ and $xs_{n,x}$. By assumption we have $x \ge 1$ as well as $n^{1/2}|m_{n,x}-xn^{-1/2}| \le 1/4$ and $s_{n,x}^2 \ge 1/2$, and therefore $s_{n,x}^{-1}n^{1/2}(xn^{-1/2}-m_{n,x})+xs_{n,x} \ge -\sqrt{2}(1/4)+1/\sqrt{2} > 0$. The implications are twofold. First, $a_{n,x}$ is nonnegative so that $\max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} = e^{-a_{n,x}^2/2}$. Second, $a_{n,x}^2 \ge (xs_{n,x})^2-(s_{n,x}^{-1}n^{1/2}|xn^{-1/2}-m_{n,x}|)^2$. Hence,

$$\max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \le e^{-(xs_{n,x})^2/2}\, e^{(s_{n,x}^{-1}n^{1/2}|xn^{-1/2}-m_{n,x}|)^2/2} \le e^{-x^2/4}\, e^{(\sqrt{2}(1/4))^2/2}.$$

Together with (C.43) this implies (C.38). Finally, we will prove (C.40). By (C.24) we obtain

$$
\begin{aligned}
S_{2,2,3,3}(\lambda,n,x) &= |\Phi_{0,1}(xs_{n,x}) - \Phi_{0,1}(x)| \\
&= \frac{1}{\sqrt{2\pi}} \int_{a_{n,x}}^{b_{n,x}} e^{-y^2/2}\,dy \\
&\le \frac{1}{\sqrt{2\pi}} (b_{n,x} - a_{n,x}) \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&= \frac{1}{\sqrt{2\pi}} x|s_{n,x} - 1| \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&\le \frac{1}{\sqrt{2\pi}} x|s_{n,x}^2 - 1| \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&\le \frac{1}{\sqrt{2\pi}} x\, c_{\lambda,14}\, \mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2} x^{12} \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
&\le c_{\lambda,61}\, \mathbb{E}[|X_1'|^\lambda]^2\, n^{-1/2} x^{13} \max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2}. \qquad \text{(C.44)}
\end{aligned}
$$

where $a_{n,x}$ and $b_{n,x}$ refer to respectively the minimum and the maximum of the real numbers $xs_{n,x}$ and x. As before we may assume without loss of generality that n is chosen sufficiently large such that $s_{n,x}^2 \ge 1/2$. This implies $\max_{\xi\in[a_{n,x},b_{n,x}]} e^{-\xi^2/2} \le e^{-x^2/4}$, which together with (C.44) implies (C.40).

Step 3. Finally, we will prove (C.6). Without loss of generality we restrict ourselves to $x > \max\{1; \sqrt{(\lambda-1)\log n}\}$. Let $r_\lambda := 1/(2\lambda(\lambda-1))$. As before consider the truncations

$$
X_i^{n,x} := X_i'\, \mathbb{1}_{\{|X_i'| \le r_\lambda n^{1/2} x\}}, \qquad 1 \le i \le n, \quad n \in \mathbb{N},
$$

and set $\widetilde{Z}_n^{n,x} := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i^{n,x}$. The specific choice of the constant r_λ will be needed in (C.54) below. On the one hand, as in (C.10) we obtain

$$
1 - F_n(x) = \mathbb{P}[Z_n' > x] \le \mathbb{P}[\widetilde{Z}_n^{n,x} > x] + n\,\mathbb{P}[|X_1'| > r_\lambda n^{1/2} x]. \qquad \text{(C.45)}
$$

On the other hand, we can use the transformation $a := \sqrt{z^2 - x^2}$ to obtain

$$
\begin{aligned}
1 - \Phi_{0,1}(x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz \\
&= e^{-\frac{x^2}{2}} \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{\frac{x^2 - z^2}{2}}\,dz \\
&= e^{-\frac{x^2}{2}} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} \frac{a}{\sqrt{a^2 + x^2}}\,da \\
&\le e^{-\frac{x^2}{2}} \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}}\,da \\
&= e^{-\frac{x^2}{2}}/2 \\
&= e^{-\frac{x^2}{2(\lambda-1)}} e^{-\frac{x^2(\lambda-2)}{2(\lambda-1)}}/2 \\
&\le (c_{\lambda,1}\, x^{-\lambda})\, e^{-\frac{(\lambda-1)\log n(\lambda-2)}{2(\lambda-1)}}/2 \\
&= c_{\lambda,2}\, x^{-\lambda}\, n^{-(\lambda/2-1)}, \qquad \text{(C.46)}
\end{aligned}
$$

where we used $x^2 \geq (\lambda-1)\log n$. From (C.45)–(C.46) we can deduce

$$|F_n(x) - \Phi_{0,1}(x)| \;\leq\; \mathbb{P}[\widetilde{Z}_n^{n,x} > x] \;+\; n\,\mathbb{P}[|X_1'| > r_\lambda n^{1/2} x] \;+\; c_{\lambda,2}\, n^{-(\lambda/2-1)}\, x^{-\lambda}. \tag{C.47}$$

By Markov's inequality we have

$$n\,\mathbb{P}[|X_1'| > r_\lambda n^{1/2} x] \;\leq\; n\,\frac{\mathbb{E}[|X_1'|^\lambda]}{(r_\lambda n^{1/2} x)^\lambda} \;\leq\; c_{\lambda,3}\,\mathbb{E}[|X_1'|^\lambda]\, n^{-(\lambda/2-1)}\, x^{-\lambda}. \tag{C.48}$$

Finally, we will show that

$$\mathbb{P}[\widetilde{Z}_n^{n,x} > x] \;\leq\; c_{\lambda,4}\, e^{c_{\lambda,5}\mathbb{E}[|X_1'|^\lambda]}\, n^{-(\lambda/2-1)}\, x^{-\lambda}. \tag{C.49}$$

Then, (C.47)–(C.49) and the assumption $\lambda > 3$ imply (C.6).

To prove (C.49), let

$$k_n(x) \;:=\; \frac{1}{x}\frac{1}{\sqrt{n}}((\lambda-2)\log n + 2\lambda(\lambda-1)\log x)$$

(note that $k_n(x)$ is nonnegative since $x \geq 1$) and use Markov's inequality to obtain

$$\begin{aligned}
\mathbb{P}[\widetilde{Z}_n^{n,x} > x] &= \mathbb{P}\Big[\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i^{n,x} > x\Big] \\
&= \mathbb{P}[e^{k_n(x)\sum_{i=1}^n X_i^{n,x}} > e^{\sqrt{n}\,x\,k_n(x)}] \\
&\leq \frac{\mathbb{E}[e^{k_n(x)\sum_{i=1}^n X_i^{n,x}}]}{e^{(\lambda-2)\log n + 2\lambda(\lambda-1)\log x}} \\
&= \frac{\mathbb{E}[e^{k_n(x)X_1^{n,x}}]^n}{n^{\lambda-2}\,x^{2\lambda(\lambda-1)}}.
\end{aligned}$$

So (C.49) would follow if we can show that

$$\mathbb{E}[e^{k_n(x)X_1^{n,x}}]^n \;\leq\; c_{\lambda,4}\, e^{c_{\lambda,5}\,\mathbb{E}[|X_1'|^\lambda]}\, n^{(\lambda-2)/2}\, x^{2\lambda(\lambda-2)}. \tag{C.50}$$

In the rest of the proof we will show (C.50). We clearly have

$$\begin{aligned}
&|\mathbb{E}[e^{k_n(x)X_1^{n,x}}]| \\
&\quad= \Big|\mathbb{E}\Big[\sum_{i=0}^\infty \frac{(k_n(x)X_1^{n,x})^i}{i!}\Big]\Big| \\
&\quad\leq 1 + k_n(x)|\mathbb{E}[X_1^{n,x}]| + \frac{k_n^2(x)}{2}\,\mathbb{E}[(X_1^{n,x})^2] + \mathbb{E}\Big[\Big|\sum_{i=3}^\infty \frac{(k_n(x)X_1^{n,x})^i}{i!}\Big|\Big] \\
&\quad\leq 1 + k_n(x)|\mathbb{E}[X_1^{n,x}]| + \frac{k_n^2(x)}{2}\,\mathbb{E}[(X_1^{n,x})^2] + \mathbb{E}\Big[\Big|(k_n(x)\,X_1^{n,x})^3\sum_{i=0}^\infty \frac{(k_n(x)\,X_1^{n,x})^i}{i!}\Big|\Big] \\
&\quad\leq 1 + k_n(x)|\mathbb{E}[X_1^{n,x}]| + \frac{k_n^2(x)}{2}\,\mathbb{E}[(X_1^{n,x})^2] \;+\; \mathbb{E}[|k_n(x)\,X_1^{n,x}|^3\, e^{k_n(x)\,X_1^{n,x}}].
\end{aligned} \tag{C.51}$$

We have $|\mathbb{E}[X_1^{n,x}]| \le \mathbb{E}[|X_1'|\mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}]$, because $X_1^{n,x} = X_1' - X_1'\mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}$ and $\mathbb{E}[X_1'] = 0$. Thus, the second summand on the right-hand side in (C.51) can be bounded above by

$$\begin{aligned} k_n(x)|\mathbb{E}[X_1^{n,x}]| &\le k_n(x)\,\mathbb{E}[|X_1'|\,\mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}] \\ &\le k_n(x)\,\mathbb{E}\Big[\frac{|X_1'|^\lambda}{(r_\lambda n^{1/2}x)^{\lambda-1}}\,\mathbb{1}_{\{|X_1'|>r_\lambda n^{1/2}x\}}\Big] \\ &\le \frac{1}{x}\frac{1}{\sqrt{n}}((\lambda-2)\log n + 2\lambda(\lambda-1)\log x)\,(r_\lambda n^{1/2}x)^{1-\lambda}\,\mathbb{E}[|X_1'|^\lambda] \\ &\le c_{\lambda,6}\,n^{-1}\,\mathbb{E}[|X_1'|^\lambda]. \end{aligned} \tag{C.52}$$

Since $\mathbb{E}[(X_1^{n,x})^2] \le \mathbb{E}[(X_1')^2] = 1$, the third summand on the right-hand side in (C.51) is bounded above by

$$\frac{k_n(x)^2}{2}\,\mathbb{E}[(X_1^{n,x})^2] \le \frac{k_n(x)^2}{2}. \tag{C.53}$$

Using the same arguments as in (C.19) we observe that $\mathbb{E}[|X_1^{n,x}|^3] \le \mathbb{E}[|X_1'|^\lambda]$. Thus for the fourth summand on the right-hand side in (C.51) we have

$$\begin{aligned} &\mathbb{E}[|k_n(x)\,X_1^{n,x}|^3\,e^{k_n(x)\,X_1^{n,x}}] \\ &\le k_n(x)^3\,e^{k_n(x)\,r_\lambda n^{1/2}x}\,\mathbb{E}[|X_1^{n,x}|^3] \\ &\le k_n(x)^3(n^{\frac{\lambda-2}{2\lambda(\lambda-1)}}\,x)\,\mathbb{E}[|X_1'|^\lambda] \\ &= ((\lambda-2)\log n + 2\lambda(\lambda-1)\log x)^3\,n^{\frac{\lambda-2}{2\lambda(\lambda-1)}-\frac{3}{2}}x^{-2}\,\mathbb{E}[|X_1'|^\lambda] \\ &= c_{\lambda,7}\,n^{-1}\mathbb{E}[|X_1'|^\lambda], \end{aligned} \tag{C.54}$$

where we used the definition of r_λ and the fact that $(\lambda-2)/(2\lambda(\lambda-1)) - 3/2 \le -1$ for $\lambda > 3$.

Now, (C.51)–(C.54) yield

$$\begin{aligned} |\mathbb{E}[e^{k_n(x)X_1^{n,x}}]| &\le 1 + c_{\lambda,6}\,n^{-1}\,\mathbb{E}[|X_1'|^\lambda] + \frac{k_n^2(x)}{2} + c_{\lambda,7}\,n^{-1}\mathbb{E}[|X_1'|^\lambda] \\ &\le e^{k_n^2(x)/2 + c_{\lambda,8}\mathbb{E}[|X_1'|^\lambda]n^{-1}}. \end{aligned} \tag{C.55}$$

Thus, for every $n \in \mathbb{N}$ we obtain

$$\mathbb{E}[e^{k_n(x)X_1^{n,x}}]^n \le e^{nk_n^2(x)/2 + c_{\lambda,8}\mathbb{E}[|X_1'|^\lambda]}. \tag{C.56}$$

Since $x \ge ((\lambda-1)\log n)^{1/2}$, we have

$$\begin{aligned} &\frac{n}{2}\,k_n^2(x) \\ &= \frac{n}{2}\,\frac{1}{x^2}\,\frac{1}{n}((\lambda-2)^2(\log n)^2 + 2\lambda(\lambda-1)(\lambda-2)\log x\log n + 4\lambda^2(\lambda-1)^2(\log x)^2) \\ &\le \frac{\lambda-2}{2}\log n\frac{\lambda-2}{\lambda-1}(\lambda-1)\log n\,\frac{1}{x^2} + 2\lambda(\lambda-2)\log x + 2\lambda^2(\lambda-1)^2(\log x)^2\frac{1}{x^2} \\ &\le \frac{\lambda-2}{2}\log n + 2\lambda(\lambda-2)\log x + c_{\lambda,9}, \end{aligned} \tag{C.57}$$

with $c_{\lambda,9} = 2\lambda^2(\lambda-1)^2$. Now (C.56)–(C.57) imply

$$\begin{aligned}\mathbb{E}\big[e^{k_n(x)X_1^{n,x}}\big]^n &\leq e^{\frac{1}{2}(\lambda-2)\log n + 2\lambda(\lambda-2)\log x + c_{\lambda,9} + c_{\lambda,8}\mathbb{E}[|X_1'|^\lambda]} \\ &\leq c_{\lambda,10}\, n^{(\lambda-2)/2}\, x^{2\lambda(\lambda-2)} e^{c_{\lambda,8}\mathbb{E}[|X_1'|^\lambda]}. \qquad \text{(C.58)}\end{aligned}$$

This shows (C.50) with $c_{\lambda,4} := c_{\lambda,10}$ and $c_{\lambda,5} := c_{\lambda,8}$, and the proof is complete. □

Bibliography

[1] Acerbi, C. and Tasche, D. (2001). On the coherence of Expected Shortfall. *Working Paper* http://www.gloriamundi.org/var/wps.html.

[2] Ahn, J.Y. and Shyamalkumar, N.D. (2010). An asymptotic analysis of the bootstrap bias correction for the empirical CTE. *North American Actuarian Journal*, 14, 217–234.

[3] Andersen, N.T., Giné, E. and Zinn, J. (1988). The central limit theorem for empirical processes under local conditions: The case of Radon infinitely divisible limits without Gaussian component. *Transactions of the American Mathematical Society*, 308, 303–335.

[4] Arenal-Gutiérrez, E. and Matrán, C. (1996). A zero-one law approach to the central limit theorem for the weighted bootstrap mean. *Annals of Probability*, 24, 532–540.

[5] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203–228.

[6] Araujo, A. and Giné, E. (1980). *The central limit theorem for real and Banach valued random variables*. Wiley, New York.

[7] Bauer, H. (2001). *Measure and integration theory*. de Gruyter, Berlin.

[8] Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54, 41–48.

[9] Belomestny, D. and Krätschmer, V. (2012). Central limit theorems for law-invariant coherent risk measures. *Journal of Applied Probability*, 49, 1–21.

[10] Beutner, E. and Zähle, H. (2010). A modified functional delta-method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis*, 101, 2452-2463.

[11] Beutner, E. and Zähle, H. (2016). Functional delta-method for the bootstrap of quasi-Hadamard differentiable functionals. *Electronic Journal of Statistics*, 10, 1181–1222.

[12] Beutner, E. and Zähle, H. (2016). Functional delta-method for the bootstrap of uniformly quasi-Hadamard differentiable functionals. *Submitted for publication* (arXiv:1609.05803).

[13] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9, 1196-1217.

[14] Bikelis, A. (1965). Estimates of the remainder in the central limit theorem. *Lithuanian Mathematical Journal*, 6, 323–346.

[15] Billingsley, P. (1999). *Convergence of probability measures*, second edition. Wiley, New York.

[16] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap methods and their application.* Cambridge University Press, Cambridge.

[17] Delbaen, F. (2013). A remark on the structure of expectiles. *Preprint* (ArXiv: 1307.5881).

[18] Del Barrio, E. and Matrán, C. (2000). The weighted bootstrap mean for heavy-tailed distributions. *Journal of Theoretical Probability*, 13, 547–569.

[19] Denneberg, D. (1994). *Non-additive measure and integral.* Kluwer, Dordrecht.

[20] Döbler, C. (2015). New Berry-Esseen and Wasserstein bounds in the CLT for non-randomly centered random sums by probabilistic methods. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 2, 863-902.

[21] Dudley, R.M. (2010). An extended Wichura theorem, definitions of Donsker class, and weighted empirical distributions, *In*: E. Giné, V. Koltchinskii, R. Norvaisa (Eds.), *Selected Works of R.M. Dudley*, Springer, New York, 38–75.

[22] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.

[23] Efron, B. and Tibshirani, R. (1994). *An introduction to the bootstrap.* Chapman & Hall, New York.

[24] Ekisheva, S. and Houdré, C. (2007). Transportation Distance and the Central Limit Theorem. Preprint *(ArXiv: 0607089).*

[25] Embrechts, P., Klüppelberg, C. and Mikosch T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer, Heidelberg.

[26] Fischer, T. (2003). Risk capital allocation by coherent risk measures based on one-sided moments. *Insurance: Mathematics and Economics*, 32(1), 135–146.

[27] Föllmer, H. and Schied, A. (2011). *Stochastic finance. An introduction in discrete time.* de Gruyter, Berlin.

[28] Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Annals of Probability*, 18, 851–869.

[29] Givens, C.R. and Shortt, R.M. (1984). A class of Wasserstein metrics for probability distributions. *The Michigan Mathematical Journal*, 31(2), 231–240.

[30] Gnedenko, B.V. and Korolev, V.Y. (1996). *Random summation*. CRC Press, Boca Raton, FL.

[31] Hutson, A.D. and Ernst, M.D. (2000). The exact bootstrap mean and variance of an L-estimator. *Journal of Royal Statistical Society: Series B*, 62, 89–94.

[32] Jeske, D.R. and Sampath A. (2003). A real example that illustrates interesting properties of bootstrap bias correction. *The American Statistician*, 57(1), 62–65.

[33] Kim, J.H.T. and Hardy, M.R. (2007). Quantifying and correcting the bias in estimated risk measures. *ASTIN Bulletin*, 37(2), , 365–386.

[34] Kim, J.H.T. (2010). Bias correction for estimated distortion risk measure using the bootstrap. *Insurance: Mathematics and Economics*, 47 (2010), 198–205.

[35] Kallenberg, O. (2002). *Foundations of modern probability*. Springer-Verlag, New York.

[36] Katz, M.L. (1963). Note on the Berry–Esséen theorem. *Annals of Mathematical Statistics*, 34, 1107–1108.

[37] Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, 18, 271–295.

[38] Krätschmer, V., Schied, A. and Zähle, H. (2015). Quasi-Hadamard differentiability of general risk functionals and its application. *Statistics and Risk Modeling*, 32, 25–47.

[39] Krätschmer, V. and Zähle, H. (2011). Sensitivity of risk measures with respect to the normal approximation of total claim distributions. *Insurance: Mathematics and Economics*, 49, 335–344.

[40] Lahiri, S.N. (2003). *Resampling methods for dependent data*. Springer-Verlag, New York.

[41] Lauer, A. and Zähle, H. (2015). Nonparametric estimation of risk measures of collective risks. *Statistics and Risk Modeling*, 32(2), 89–102.

[42] Lauer, A. and Zähle, H. (2017). Bootstrap consistency and bias correction in the nonparametric estimation of risk measures of collective risks. *Insurance: Mathematics and Economics*, 74, 99–108.

[43] Lo, A.Y. (1987). A Large Sample Study of the Bayesian Bootstrap. *The Annals of Statistics*, 15(1), 360–375.

[44] Lundberg, F. (1903). Approximerad framställning av sannolikhetsfunktionen, Aterförsäkring av kollektivrisker. *Akad. Afhandling. Almqvist och Wiksell*, Uppsala, 8, 22.

[45] Mason, D. and Newton, M. (1992). A rank statistic approach to the consistency of a general bootstrap. *The Annals of Statistics*, 20, 1611–1624.

[46] Michel, R. (1976). Nonuniform central limit bounds with applications to probabilities of deviations. *Annals of Probability*, 4, 102–106.

[47] Nagaev, S.V. (1965). Some limit theorems for large deviations. *Theory of Probability and its Applications*, 10, 214–235.

[48] Nefedova, Yu.S. and Shevtsova, I.G. (2013). On nonuniform convergence rate estimates in the central limit theorem. *Theory of Probability and its Applications*, 57, 28–59.

[49] Panjer, H.H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12, 22–26.

[50] Petrov, V.V. (1995). *Limit theorems of probability theory.* Oxford University Press, Oxford.

[51] Pflug, G. and Wozabal N. (2007). Asymptotic distribution of law-invariant risk functionals. *Finance and Stochastics*, 14, 397–418.

[52] Pitts, S.M. (1994). Nonparametric estimation of compound distributions with applications in insurance. *Annals of the Institute of Mathematical Statistics*, 46, 537–555.

[53] Pitts, S.M. and Qin, L. (2010). Nonparametric Estimation of the Finite-Time Survival Probability with Zero Initial Capital in the Classical Risk Model. *Methodology and Computing in Applied Probability*, 14.4, 919–936.

[54] Praestgaard, J. and Wellner, J.A. (1993). Exchangeably Weighted Bootstraps of the General Empirical Process. *Annals of Probability*, 21(4), 2053–2086.

[55] Robbins, H. (1948). The asymptotic distribution of the random sum of a random number of random variables. *Bulletin of the American Mathematical Society*, 54, 1151–1161.

[56] Rubin, D. (1981). The Bayesian bootstrap.*The Annals of Statistics*, 9, 130–134.

[57] Shao, J. and Tu, D. (1995). *The jackknife and bootstrap.* Springer-Verlag, New York.

[58] Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

[59] Su, C., Zhao, L. and Wang, Y. (1997). Moment inequalities and weak convergence for negatively associated sequences. *Science in China Series*, 40(2), 172–182.

[60] Tsukahara, H. (2014). Estimation of distortion risk measures. *Journal of Financial Econometrics*, 12(1), 213–235.

[61] Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.

[62] van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

[63] Wang, S. and Dhaene, J. (1998). Comontonicity, correlation order and premium principles. *Insurance: Mathematics and Economics* 22, 235–242.

[64] Zähle, H. (2011). Rates of almost sure convergence of plug-in estimates for distrortion risk measures. *Metrika* 74, 267–285.

[65] Zähle, H. (2014). Marcinkiewicz-Zygmund and ordinary strong laws for empirical distribution functions and plug-in estimators. *Statistics*, 48, 951-964.