Nonparametric estimation of risk measures of collective risks and a nonuniform Berry–Esséen inequality

Alexandra Lauer∗ Henryk Zähle†

Abstract

We consider two nonparametric estimators for the risk measure of the sum of \( n \) i.i.d. individual insurance risks where the number of historical single claims that are used for the statistical estimation is of order \( n \). This framework matches the situation that nonlife insurance companies are faced with within the scope of premium calculation. Indeed, the risk measure of the aggregate risk divided by \( n \) can be seen as a suitable premium for each of the individual risks. For both estimators divided by \( n \) we derive a sort of Marcinkiewicz–Zygmund strong law as well as a weak limit theorem. The behavior of the estimators for small to moderate \( n \) is studied by means of Monte-Carlo simulations. The proof of our main result relies on a new Berry–Esséen result, which is of independent interest.

Keywords: Aggregate risk, Total claim distribution, Convolution, Law-invariant risk measure, Nonuniform Berry–Esséen inequality, Marcinkiewicz–Zygmund strong law, Weak limit theorem, Panjer recursion

∗Department of Mathematics, Saarland University, lauer@math.uni-sb.de
†Department of Mathematics, Saarland University, zaehle@math.uni-sb.de
1. Introduction

Let \((X_i)\) be a sequence of nonnegative i.i.d. random variables on a common probability space with distribution \(\mu\). In the context of actuarial theory, the random variable \(S_n := \sum_{i=1}^n X_i\) can be seen as the total claim of a homogeneous insurance collective consisting of \(n\) risks. The distribution of \(S_n\) is given by the \(n\)-fold convolution \(\mu^{*n}\) of \(\mu\). A central task in insurance practice is the specification of the premium \(R_\rho(\mu^{*n})\) for the aggregate risk \(S_n\), where \(R_\rho\) is the statistical functional associated with any suitable law-invariant risk measure \(\rho\) (henceforth referred to as risk functional associated with \(\rho\)). Note that \(R_\rho(\mu^{*n})/n\) can be seen as a suitable premium for each of the individual risks \(X_1, \ldots, X_n\), where it is important to note that \(R_\rho(\mu^{*n})/n\) is typically essentially smaller than \(R_\rho(\mu)\).

On the one hand, much is known about the statistical estimation of the single claim distribution \(\mu\) and about the numerical approximation of the convolution \(\mu^{*n}\) with known \(\mu\). On the other hand, an analysis that combines both statistical aspects and the numerical approximation of \(\mu^{*n}\) seems to be rare. In [5], this question was approached through an estimation of \(\mu^{*n}\) by the normal distribution \(N_{\hat{m}_{un}, \hat{s}^2_{un}}\) with estimated parameters based on a sample of size \(u_n \in \mathbb{N}\). Here \(\hat{m}_{un}\) and \(\hat{s}^2_{un}\) refer to respectively the empirical mean and the empirical variance of a sequence of i.i.d. random variables with distribution \(\mu\) having a finite second moment. It was shown in [5] that for many law-invariant coherent risk measures \(\rho\) and any sequence \((u_n)\) of positive integers for which \(u_n/n\) converges to some constant \(c \in (0, \infty)\) we have

\[
n^r \frac{R_\rho(N_{\hat{m}_{un}, \hat{s}^2_{un}}) - R_\rho(\mu^{*n})}{n} \xrightarrow{a.s.} 0, \quad n \to \infty \tag{1}
\]

for every \(r < 1/2\), and

\[
\text{law}\left\{n^{1/2} \frac{R_\rho(N_{\hat{m}_{un}, \hat{s}^2_{un}}) - R_\rho(\mu^{*n})}{n}\right\} \xrightarrow{w} \mathcal{N}_{0, s^2}, \quad n \to \infty \tag{2}
\]

with \(s^2 := \text{Var}[X_1]\). Of course, (2) implies in particular that the convergence in (1) cannot hold for \(r \geq 1/2\). The assumption that \(u_n\) increases to infinity at the same speed as \(n\) increases to infinity is motivated by the fact that the parameters are typically estimated on the basis of the historical claims of the same collective from the last year or from the last few years. This is also why the presented theory is nonstandard. In the existing literature on the statistical estimation of convolutions the number of summands is typically fixed or increases essentially slower to infinity than \(u_n\) does; see, for instance, [11] for the nonparametric estimation of a (compound) convolution where the (distribution of the) number of summands is fixed and known. It was also shown in [5] that for the exact mean \(m\) and the exact variance \(s^2\) of \(\mu\), and for many law-invariant coherent risk measures \(\rho\),

\[
\sup_{n \in \mathbb{N}} |R_\rho(N_{nm, ns^2}) - R_\rho(\mu^{*n})| < \infty. \tag{3}
\]
Both (1)–(3) and the simulation study in [5] 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.

The estimated normal approximation $R_\rho(N_{\hat{m}_{n}, n\hat{s}_{n}})$ of $R_\rho(\mu^*)$ is very simple and saves computing time in great measure. Indeed, we have

$$R_\rho(N_{\hat{m}_{n}, n\hat{s}_{n}}) = \sqrt{n\hat{s}_{n}} R_\rho(N_{0,1}) + n\hat{m}_{n}$$

whenever $R_\rho$ corresponds to a cash additive and positively homogeneous risk measure $\rho$. On the other hand, in real applications the total claim distribution $\mu^*$ is typically skewed to the right, whereas the normal distribution is symmetric; see also Figure 1. So it is natural to study methods which better fit skewed total claim distributions. In this article, we will therefore replace $N_{n\hat{m}_{n}, n\hat{s}_{n}^2}$ by the $n$-fold convolution $\hat{\mu}^{*n}_{n}$ of the empirical estimator $\hat{\mu}_{n}$ of $\mu$. The corresponding estimator $R_\rho(\hat{\mu}^{*n}_{n})$ will be referred to as empirical plug-in estimator. The calculation of the empirical plug-in estimator will be more computing time consuming than the calculation of the estimated normal approximation, nevertheless the needed computing time is still satisfying for actuarial applications. It is quite clear, and can also be seen from Figure 1 that $\mu^*$ gets increasingly skewed as the tail of $\mu$ gets heavier. So it is not surprising that the estimated normal approximation works well for light-tailed $\mu$ and gets worse for medium-tailed and heavy-tailed $\mu$. A simulation study for the Value at Risk functional in Section 3 indicates that the empirical plug-in estimator is only slightly better than the estimated normal approximation for light-tailed $\mu$ but is essentially better for medium-tailed $\mu$. For heavy-tailed $\mu$ both estimators work well only for rather large $n$. Throughout this article we will use the terms “light-tailed”, “medium-tailed” and “heavy-tailed” in a quite sloppy way. By definition “heavy-tailed” refers to distributions without a finite second moment. However our theory is only applicable to distributions with a finite $\lambda$-moment for some $\lambda > 2$. In this context we refer to heavy-tailed distributions whenever $\lambda$ is close to 2 and will use the terms “medium-tailed” and “light-tailed” for larger $\lambda$.

To introduce the empirical plug-in estimator rigorously, let $(Y_i)$ be a sequence of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution $\mu$. The random variables $Y_i$ can be seen as observed historical single claims. The empirical probability measure of the first $u \in \mathbb{N}$ observations,

$$\hat{\mu}_u := \frac{1}{u} \sum_{i=1}^{u} \delta_{Y_i},$$

whenever $R_\rho$ corresponds to a cash additive and positively homogeneous risk measure $\rho$. On the other hand, in real applications the total claim distribution $\mu^*$ is typically skewed to the right, whereas the normal distribution is symmetric; see also Figure 1. So it is natural to study methods which better fit skewed total claim distributions. In this article, we will therefore replace $N_{n\hat{m}_{n}, n\hat{s}_{n}^2}$ by the $n$-fold convolution $\hat{\mu}^{*n}_{n}$ of the empirical estimator $\hat{\mu}_{n}$ of $\mu$. The corresponding estimator $R_\rho(\hat{\mu}^{*n}_{n})$ will be referred to as empirical plug-in estimator. The calculation of the empirical plug-in estimator will be more computing time consuming than the calculation of the estimated normal approximation, nevertheless the needed computing time is still satisfying for actuarial applications. It is quite clear, and can also be seen from Figure 1 that $\mu^*$ gets increasingly skewed as the tail of $\mu$ gets heavier. So it is not surprising that the estimated normal approximation works well for light-tailed $\mu$ and gets worse for medium-tailed and heavy-tailed $\mu$. A simulation study for the Value at Risk functional in Section 3 indicates that the empirical plug-in estimator is only slightly better than the estimated normal approximation for light-tailed $\mu$ but is essentially better for medium-tailed $\mu$. For heavy-tailed $\mu$ both estimators work well only for rather large $n$. Throughout this article we will use the terms “light-tailed”, “medium-tailed” and “heavy-tailed” in a quite sloppy way. By definition “heavy-tailed” refers to distributions without a finite second moment. However our theory is only applicable to distributions with a finite $\lambda$-moment for some $\lambda > 2$. In this context we refer to heavy-tailed distributions whenever $\lambda$ is close to 2 and will use the terms “medium-tailed” and “light-tailed” for larger $\lambda$.

To introduce the empirical plug-in estimator rigorously, let $(Y_i)$ be a sequence of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution $\mu$. The random variables $Y_i$ can be seen as observed historical single claims. The empirical probability measure of the first $u \in \mathbb{N}$ observations,
is the standard nonparametric estimator for $\mu$, and therefore
\[ \hat{\mu}^*_{un} := (\hat{\mu}_u)^n \] (6)
provides a reasonable estimator for $\mu^*_{un}$. Then it is natural to use the plug-in estimator
\[ \mathcal{R}_\rho(\hat{\mu}^*_{un}) \] (7)
for the estimation of $\mathcal{R}_\rho(\mu^*_{un})$. In general the computation of the $n$-fold convolution $\hat{\mu}^*_{un}$ of $\hat{\mu}_u$ is more or less impossible. However, in real applications the true $\mu$ has support in $hN_0 := \{0, h, 2h, \ldots\}$ 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 $\hat{\mu}_u$ is concentrated on the equidistant grid $hN_0$, too. In this case the estimated total claim distribution $\hat{\mu}^*_{un}$ can be computed with the help of the recursive scheme
\[ \hat{\mu}^*_{un}[^{\{0\}} ] = \hat{\mu}_u[^{\{0\}} ]^n \] (8)
\[ \hat{\mu}^*_{un}[^{\{jh\}} ] = \frac{1}{j \hat{\mu}_u[^{\{0\}} ]} \sum_{\ell=1}^{j} ((n+1)\ell - j) \hat{\mu}_u[^{\{\ell h\}} ] \hat{\mu}^*_{un}[^{\{(j-\ell)h\}} ] \quad \text{for } j \in \mathbb{N}, \] (9)
provided $\hat{\mu}_u[^{\{0\}} ] > 0$; cf. the Appendix [A]. Note that $\hat{\mu}_u$ as an empirical probability measure has bounded support. Therefore, in view of (8)–(9), the estimator $\mathcal{R}_\rho(\hat{\mu}^*_{un})$ can typically be computed in finite time, even for tail-dependent functionals $\mathcal{R}_\rho$ as, for instance, the one associated with the Expected Shortfall.

We will see in Section 2 that for a very large class of law-invariant risk measures $\rho$, any distribution $\mu$ with a finite $\lambda$-moment for some $\lambda > 2$, and any sequence $(u_n)$ of positive integers for which $u_n/n$ converges to some constant $c \in (0, \infty)$, we also have (1)–(2) with $\mathcal{N}_{n\hat{m}_{un}, n\hat{s}_{un}^2}$ replaced by $\hat{\mu}^*_{un}$. We will prove even more, namely
\[ \frac{\mathcal{R}_\rho(\mathcal{N}_{n\hat{m}_{un}, n\hat{s}_{un}^2}) - \mathcal{R}_\rho(\mu^*)}{\sqrt{n}} = (\hat{m}_{un} - m) + o_{\mathbb{P} \text{-a.s.}}(n^{-1/2}), \] (10)
\[ \frac{\mathcal{R}_\rho(\hat{\mu}^*_{un}) - \mathcal{R}_\rho(\mu^*)}{\sqrt{n}} = (\hat{m}_{un} - m) + o_{\mathbb{P} \text{-a.s.}}(n^{-1/2}), \] (11)
where $o_{\mathbb{P} \text{-a.s.}}(n^{-1/2})$ refers to any sequence of random variables $(\xi_n)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ for which $\sqrt{n}\xi_n$ converges $\mathbb{P}$-a.s. to zero as $n \rightarrow \infty$. Assertions (10)–(11) have an astonishing consequence. No matter what the particular risk measure $\rho$ looks like, the asymptotics of the estimators $\frac{1}{n}\mathcal{R}_\rho(\mathcal{N}_{n\hat{m}_{un}, n\hat{s}_{un}^2})$ and $\frac{1}{n}\mathcal{R}_\rho(\hat{\mu}^*_{un})$ for the individual premium $\frac{1}{n}\mathcal{R}_\rho(\mu^*)$ are exactly the same as for the empirical mean regarded as an estimator for the mean; see Corollary 2.5 below for details. By the classical Central Limit Theorem, we can derive from (10) and (11) the following asymptotic confidence intervals at level $1 - \alpha$ for
the individual premium \( \frac{1}{n} \mathcal{R}_p(\mu^n) \):

\[
\left[ \frac{\mathcal{R}_p(N_{m\tilde{u}_n}, \tilde{s}_{\tilde{u}_n}^2)}{n} - \frac{\tilde{s}_{\tilde{u}_n}}{\sqrt{n}} \Phi_{0.1}^{-1}(1 - \frac{\alpha}{2}), \frac{\mathcal{R}_p(N_{n\tilde{u}_n}, \tilde{s}_{\tilde{u}_n}^2)}{n} - \frac{\tilde{s}_{\tilde{u}_n}}{\sqrt{n}} \Phi_{0.1}^{-1}(\frac{\alpha}{2}) \right]
\]

and

\[
\left[ \frac{\mathcal{R}_p(\hat{\mu}_n^n)}{n} - \frac{\tilde{s}_{\tilde{u}_n}}{\sqrt{n}} \Phi_{0.1}^{-1}(1 - \frac{\alpha}{2}), \frac{\mathcal{R}_p(\hat{\mu}_n^n)}{n} - \frac{\tilde{s}_{\tilde{u}_n}}{\sqrt{n}} \Phi_{0.1}^{-1}(\frac{\alpha}{2}) \right],
\]

where \( \Phi_{0.1} \) denotes the distribution function of \( N_{0.1} \).

Further, it is a simple consequence of part (ii) of Theorem 2.4 below that

\[
\frac{\mathcal{R}_p(\mu^n)}{n} = m + \frac{\mathcal{R}_p(N_{0.1})}{\sqrt{n}} s + O(n^{-1/2 - \gamma}) \tag{12}
\]

with \( \gamma := \min\{\lambda - 2; 1\}/2 \). The identity (12) shows that for large \( n \) (and \( \gamma \) away from 0) the individual premium \( \frac{1}{n} \mathcal{R}_p(\mu^n) \) can be seen as an approximation of the premium which is determined according to the standard deviation principle with safety loading \( \frac{1}{\sqrt{n}} \mathcal{R}_p(N_{0.1}) \). For the corresponding estimators we will obtain from parts (iv) and (v) of Theorem 2.4 below the following empirical analogues of (12):

\[
\frac{\mathcal{R}_p(N_{m\tilde{u}_n}, \tilde{s}_{\tilde{u}_n}^2)}{n} = \hat{m}_{\tilde{u}_n} + \frac{\mathcal{R}_p(N_{0.1})}{\sqrt{n}} \tilde{s}_{\tilde{u}_n}, \tag{13}
\]

where \( \mathcal{R}_p(N_{0.1}) \) refers to any sequence of random variables \((\xi_n)\) on \((\Omega, \mathcal{F}, \mathbb{P})\) for which the sequence \((n^{1/2 + \gamma} \xi_n)\) is bounded \( \mathbb{P} \)-a.s. To some extent, (12) and (13) justify the use of the standard deviation principle (with \( m \) and \( s \) estimated by \( \hat{m}_{\tilde{u}_n} \) and \( \tilde{s}_{\tilde{u}_n} \), respectively), which many insurance companies use to determine individual premiums in large collectives. In practice the specific choice of the safety loading in the context of the standard deviation principle is often somewhat arbitrary. Formulae (12) and (13) now give a deeper insight into the practical choice of the safety loading. It should be chosen as the product of a suitable risk functional (which one has actually in mind) evaluated at the standard normal distribution and the factor \( 1/\sqrt{n} \) (where \( n \) is the size of the collective). The factor \( 1/\sqrt{n} \) reflects the balancing of risks in large collectives.

It is quite clear that the goodness of the estimator in (7) can be improved through replacing the nonparametric estimator \( \hat{\mu}_n \) in (5)–(6) by a suitable estimator that is based on a parametric statistical model. However, this requires preliminary considerations w.r.t. a proper choice of the parametric model. Such considerations are feasible and common. Nevertheless we leave the parametric approach for future work.

The rest of the article is organized as follows. In Section 2 we will present our main results, and in Section 3 these results will be illustrated by means of numerical examples. The proof of Theorem 2.4 relies on a new nonuniform Berry–Esséen inequality.
This inequality will be presented in Section 4 and is of independent interest. From a mathematical point of view the proof of the nonuniform Berry–Esséen inequality is the hard part about this article.

2. Main results

Let $L^0$ denote the usual set of all finitely-valued random variables on an atomless probability space modulo the equivalence relation of almost sure identity. Let $\mathcal{X} \subset L^0$ be a vector space containing the constants. We will say that a map $\rho : \mathcal{X} \rightarrow \mathbb{R}$ is

- **monotone** if $\rho(X_1) \leq \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$ with $X_1 \leq X_2$.
- **cash additive** if $\rho(X + m) = \rho(X) + m$ for all $X \in \mathcal{X}$ and $m \in \mathbb{R}$.
- **subadditive** if $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$ for all $X_1, X_2 \in \mathcal{X}$.
- **positively homogenous** if $\rho(\lambda X) = \lambda \rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \geq 0$.

As usual, we will say that $\rho$ is **coherent** if it satisfies all of these four conditions, and that $\rho$ is **law-invariant** if $\rho(X) = \rho(Y)$ whenever $X$ and $Y$ have the same law. We will restrict ourselves to law-invariant maps $\rho : \mathcal{X} \rightarrow \mathbb{R}$. So we may and do associate with $\rho$ a statistical functional $\mathcal{R}_\rho : \mathcal{M}(\mathcal{X}) \rightarrow \mathbb{R}$ via

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

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 $\mu$.

Let $\mathcal{M}_1$ be the set of all probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and denote by $F_\mu$ the distribution function of $\mu \in \mathcal{M}_1$. For every $\lambda \geq 0$, let the function $\phi_\lambda : \mathbb{R} \rightarrow [1, \infty)$ 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

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

is the nonuniform Kolmogorov distance of $\mu_1$ and $\mu_2$ w.r.t. the weight function $\phi_\lambda$. It is easily seen that $d_{\phi_\lambda}$ provides a metric on the set $\mathcal{M}_1^\lambda$ of all $\mu \in \mathcal{M}_1$ satisfying $d_{\phi_\lambda}(\mu, \delta_0) < \infty$.

Recall that $(Y_i)$ is a sequence of i.i.d. real-valued random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution $\mu$ having a finite second moment, and that the estimators $\hat{\mu}_u$, $\hat{\mu}_u^n$, and $\mathcal{R}_\rho(\hat{\mu}_u^n)$ are given by (3), (3), and (7), respectively. We set $m := \mathbb{E}[Y_1]$ and $s := \mathbb{V}ar[Y_1]^{1/2}$, and let $\hat{m}_u := \frac{1}{n} \sum_{i=1}^n Y_i$ and $\hat{s}_u := \left(\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{m}_u)^2\right)^{1/2}$ be the corresponding standard nonparametric estimators.
Assumption 2.1 Let \( \rho : \mathcal{X} \rightarrow \mathbb{R} \) be a law-invariant map, and \( \mathcal{R}_\rho \) be the corresponding statistical functional introduced in \((14)\). Let \((u_n)\) be a sequence in \(\mathbb{N}\), and assume that the following assertions hold for some \(\lambda > 2\):

(a) \( \mu \) lies in \( \mathcal{M}_1^\lambda \), that is, \( \mathbb{E}[|Y_1|^\lambda] < \infty \).

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

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

(d) For each sequence \((m_n) \subset \mathcal{M}_1^\lambda\) with \( d_{\phi_\lambda}(m_n, N_{0,1}) \rightarrow 0 \), there exists a constant \( C > 0 \) such that \( |\mathcal{R}_\rho(m_n) - \mathcal{R}_\rho(N_{0,1})| \leq C d_{\phi_\lambda}(m_n, N_{0,1}) \) for all \( n \in \mathbb{N} \).

The probably most popular risk measure in practice, the Value at Risk, is known to be monotone, cash additive, and positively homogeneous on \( \mathcal{X} = L^0 \). In particular, it satisfies condition (c). It follows from Theorem 2 in [12] that the Value at Risk also satisfies condition (d) for any \( \lambda \geq 0 \). The following Remark 2.2 and Example 2.3 show that there are also a lot of coherent risk measures that satisfy condition (d). So it is fair to say that condition (d) is quite weak.

Remark 2.2 Let \( H^\Psi \subset L^0 \) be the Orlicz heart associated with some continuous Young function \( \Psi \), and assume that the standard normal distribution \( N_{0,1} \) lies in \( \mathcal{M}(H^\Psi) \). Let \( \rho : H^\Psi \rightarrow \mathbb{R} \) be a law-invariant coherent risk measure. Define a function \( g_\rho \) by \( g_\rho(t) := \rho(B_{1,t}) \), \( t \in [0,1] \), where \( B_{1,t} \) is a Bernoulli random variable with expectation \( t \). Then it follows from Theorem 2.4, Lemma A.5, and Remark 2.8 in [6] that condition (d) of Assumption 2.1 is satisfied for \( \lambda \geq 0 \) when \( \hat{g}_\rho(z\Phi_{0,1}(x))\Phi_{0,1}(x)dx < \infty \) for some \( z \in (0,1) \). (16)

It is worth mentioning that the Orlicz heart \( H^\Psi_p \) for \( \Psi_p(x) := x^p \) (with \( p \in [1,\infty) \)) is just the usual \( L^p \)-space. It is also worth mentioning that \( g_\rho = g \) when \( \rho \) is a distortion risk measure with distortion function \( g \). For background see [6].

Example 2.3 In Section 2.2 in [6], the integrability condition (16) has been further investigated for different risk measures. It is easily seen that condition (16) is satisfied for any \( \lambda > p \) when \( \rho \) is the risk measure based on one-sided \( p \)th moments for \( p \in [1,\infty) \), and for any \( \lambda > 1 \) when \( \rho \) is either the expectiles-based risk measure (recently introduced in [1]) or the Average Value at Risk (also known as Expected Shortfall).

We now turn to our main result. We emphasize that the measurability condition on \( \mathcal{R}_\rho(N_{n\hat{m}_{u_n},n\hat{s}_{u_n}}) \) and \( \mathcal{R}_\rho(\hat{\mu}_{u_n}^{*n}) \) in the following theorem is fulfilled when \( \rho \) is the Value at Risk or a law-invariant coherent risk measure on an Orlicz heart with continuous Young function.
Theorem 2.4 Suppose that Assumption [2.1] holds with $\lambda > 2$, let $\gamma := \min\{\lambda - 2; 1\}/2$, and assume that $\mathcal{R}_\rho(N_{nmu_n,n^2u_n})$ and $\mathcal{R}_\rho(\hat{\mu}_{un}^{*n})$ are $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$-measurable for every $n \in \mathbb{N}$. Then the following assertions hold:

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

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

(iii) $\frac{1}{n}(\mathcal{R}_\rho(N_{nmu_n,n^2u_n}) - \mathcal{R}_\rho(\hat{\mu}_{un}^{*n})) = O_{\mathbb{P}\text{-a.s.}}(n^{-1/2-\gamma})$.

(iv) $\frac{1}{n}\mathcal{R}_\rho(N_{nmu_n,n^2u_n}) = \hat{m}_{un} + \frac{1}{\sqrt{n}}\bar{s}_{un}\mathcal{R}_\rho(N_{0,1})$.

(v) $\frac{1}{n}\mathcal{R}_\rho(\hat{\mu}_{un}^{*n}) = \hat{m}_{un} + \frac{1}{\sqrt{n}}\bar{s}_{un}\mathcal{R}_\rho(N_{0,1}) + O_{\mathbb{P}\text{-a.s.}}(n^{-1/2-\gamma})$.

Proof (i): By part (c) of Assumption [2.1] and the representation [4] (and its analogue in the case of known parameters), we have

$$\mathcal{R}_\rho(N_{nmu_n,n^2u_n}) - \mathcal{R}_\rho(N_{nm,n^2}) = \sqrt{n}(\bar{s}_{un} - s)\mathcal{R}_\rho(N_{0,1}) + n(\hat{m}_{un} - m).$$  \hspace{1cm} (17)

Since the empirical standard deviation $\bar{s}_{un}$ converges $\mathbb{P}$-a.s. to the true standard deviation $s$, the claim of part (i) follows through dividing Equation (17) by $n$.

(ii): Let $S_n$ be a random variable with distribution $\mu^{*n}$, set $Z_n := (S_n - nm)/(\sqrt{ns})$, and note that law$\{\sqrt{ns}Z_n + nm\} = \mu^{*n}$. Write $N_n$ for any random variable distributed according to the normal distribution $N_{nm,n^2}$, and note that $Z := (N_n - nm)/(\sqrt{ns})$ is $N_{0,1}$-distributed. Due to part (c) of Assumption [2.1] we obtain

$$\mathcal{R}_\rho(N_{nm,n^2}) - \mathcal{R}_\rho(\mu^{*n}) = \rho(\sqrt{ns}Z + nm) - \rho(\sqrt{ns}Z + nm)$$

$$= \sqrt{ns}(\rho(Z) - \rho(Z_n))$$

$$= \sqrt{ns}(\mathcal{R}_\rho(N_{0,1}) - \mathcal{R}_\rho(\mu_{un})), \hspace{1cm} (18)$$

where $\mu_{un}$ denotes the law of $Z_n$. The nonuniform Berry–Essèen inequality of Theorem [4.1] shows that there exists a constant $K_\lambda \in (0, \infty)$ such that $d_{\phi_\lambda}(N_{0,1}, \mu_{un}) \leq K_\lambda n^{-\gamma}$ for all $n \in \mathbb{N}$. Along with (18) and part (d) of Assumption [2.1] this ensures that we can find a constant $K \in (0, \infty)$ such that $n^{-1/2}\mathcal{R}_\rho(N_{nm,n^2}) - \mathcal{R}_\rho(\mu^{*n})| \leq n^{-1/2}Kd_{\phi_\lambda}(N_{0,1}, \mu_{un}) \leq KK_\lambda n^{-1/2-\gamma}$ for all $n \in \mathbb{N}$. This completes the proof of part (ii).

(iii): Analogously to (18), we obtain

$$\mathcal{R}_\rho(N_{nm\hat{m}_{un}(\omega),n\hat{s}_{u_n}(\omega)}) - \mathcal{R}_\rho(\hat{\mu}_{un}^{*n}(\omega; \cdot)) = \sqrt{n}\bar{s}_{un}(\omega)(\mathcal{R}_\rho(N_{0,1}) - \mathcal{R}_\rho(\hat{m}_{un}(\omega; \cdot)))$$  \hspace{1cm} (19)

for all $\omega \in \Omega$, where $\hat{m}_{un}(\omega; \cdot)$ denotes the law of the random variable $\hat{Z}_n^\omega(\cdot) := (\hat{S}_n^\omega(\cdot) - n\hat{m}_{un}(\omega))/(\sqrt{n}\bar{s}_{un}(\omega))$ for any random variable $\hat{S}_n^\omega(\cdot)$ with distribution $\hat{\mu}_{un}^{*n}(\omega; \cdot)$ and defined on some probability space $(\Omega^\omega, \mathcal{F}^\omega, \mathbb{P}^\omega)$. For (19) notice that $\hat{m}_{un}(\omega; \cdot)$ has mean
This completes the proof of part (iii).

Part (iv) is an immediate consequence of part (c) of Assumption 2.1 and part (v) follows from (iii)–(iv).

Parts (i)–(ii) of Theorem 2.4 and part (i) of the following corollary are already known from [5]. On the other hand, part (iii) of Theorem 2.4 and part (ii) of the following corollary are new and strongly rely on the nonuniform Berry–Esséen inequality of Theorem 4.1.

**Corollary 2.5** Under the assumptions of Theorem 2.4 the following assertions hold:

(i) We have
\[
\frac{\mathcal{R}_p(N_{\tilde{m}_{un}},\tilde{s}_{un}^2) - \mathcal{R}_p(\mu^{sn})}{n} = (\hat{m}_{un} - m) + o_{P.a.s.}(n^{-1/2}).
\] (21)

In particular, for every \( r < 1/2 \),
\[
n^{-r} \frac{\mathcal{R}_p(N_{\tilde{m}_{un}},\tilde{s}_{un}^2) - \mathcal{R}_p(\mu^{sn})}{n} \to 0 \quad \text{P-a.s.,}
\] (22)

\[
l_{[1/2]n^{-r} \frac{\mathcal{R}_p(N_{\tilde{m}_{un}},\tilde{s}_{un}^2) - \mathcal{R}_p(\mu^{sn})}{n}} \to \mathcal{N}_0,s^2.
\] (23)

(ii) We have
\[
\frac{\mathcal{R}_p(\tilde{\mu}_{un}^{sn}) - \mathcal{R}_p(\mu^{sn})}{n} = (\hat{m}_{un} - m) + o_{P.a.s.}(n^{-1/2}).
\] (24)

In particular, for every \( r < 1/2 \),
\[
n^{-r} \frac{\mathcal{R}_p(\tilde{\mu}_{un}^{sn}) - \mathcal{R}_p(\mu^{sn})}{n} \to 0 \quad \text{P-a.s.,}
\] (25)

\[
l_{[1/2]n^{-r} \frac{\mathcal{R}_p(\tilde{\mu}_{un}^{sn}) - \mathcal{R}_p(\mu^{sn})}{n}} \to \mathcal{N}_0,s^2.
\] (26)
Proof Assertions (21) and (24) are immediate consequences of respectively (i)–(ii) and (i)–(iii) of Theorem 2.4. By the Marcinkiewicz–Zygmund strong law of large numbers, we have that \( n^r(\hat{m}_n - m) \) converges \( \mathbb{P} \)-a.s. to zero for every \( r < 1/2 \). So (22) and (25) follow from (21) and (24), respectively. Finally, the classical Central Limit Theorem says that the law of \( n^{1/2}(\hat{m}_n - m) \) converges weakly to \( \mathcal{N}_0, s^2 \). Thus, (23) and (26) follow from Slutzky’s lemma and respectively (21) and (24).

\[ \blacksquare \]

3. Numerical examples

In this section we present some numerical examples to illustrate the results of Section 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 \( \rho \) is the Value at Risk at level \( \alpha \) the results of Corollary 2.5 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 \( \mu \)-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 1. So one would expect that especially for heavy-tailed \( \mu \) and small to moderate \( n \) the estimators perform only moderately well. One would also expect that for heavy-tailed \( \mu \) (and even for medium-tailed \( \mu \)) 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

\[ \mu = (1 - p) \delta_0 + p P_{a,b} \]

for some \( p \in (0, 1) \), where \( P_{a,b} \) is the Pareto distribution with parameters \( a > 2 \) and \( b > 0 \). The Pareto distribution \( P_{a,b} \) 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 \( P_{a,b} \) as the
individual claim distribution conditioned on this event. Note that in our example the mean $m$ and the variance $s^2$ of $\mu$ are given by

$$ m = \frac{pb}{a-1} \quad \text{and} \quad s^2 = \frac{2b^2p}{(a-1)(a-2)} - \frac{b^2p^2}{(a-1)^2}. $$ (27)

In the first part of this section, we estimate the total claim distribution $\mu^*\nu$, 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 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 normalised to 1. Each line shows the same set of parameters and each row 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 $\mu^*\nu$ and the dashed line the probability density of the normal distribution $N_{mn,ns^2}$ with $m$ and $s^2$ determined through (27). We emphasize that $\mu^*\nu$ 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 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 context of the Berry–Esséen theorem one can say that the highest order of existing moments $\lambda$ is very close to 2, which leads to a low convergence rate, namely $(\lambda - 2)/2$. 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 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 $\lambda$. For $\lambda > 3$ the convergence rate can not be improved any more.

In the third and fourth line of Figure 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
Figure 1: The continuous line shows the $n$-fold convolution $\mu^m$ of $\mu = (1 - p)\delta_0 + pP_{a,b}$ for $p = 0.1$ and the Pareto distribution $P_{a,b}$ 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 row, $n = 150$ in the second row and $n = 200$ in the third row. The dashed line shows the density of the respective normal distribution in each case.
both cases and due to the Berry–Esséen theorem the deviation of $\mu^{*n}$ from the normal distribution converges to zero with rate $1/2$. We can see that there is no remarkable improvement 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 $\rho$ is played by the Value at Risk at level $\alpha = 0.99$. To save computing time we discretized the Pareto distribution $P_{a,b}$ on the equidistant grid $10N_0 = \{0, 10, 20, \ldots \}$. The plots in Figure 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 Risk of $\mu^{*n}$ (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 the dotdashed line represents the relative Value at Risk $R_\rho(\mu^{*n})/n$, which we take as a reference to illustrate the biases of the estimators. The dashed line shows the estimated normal approximation $R_\rho(N_{\hat{\mu}^{*n}, \hat{\mu}_s^2})/n$ for the Value at Risk relative to $n$. The continuous line shows the empirical plug-in estimator $R_\rho(\hat{\mu}_n^{*n})/n$ 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 normalised 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. 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 $R_\rho(\mu^{*n})/n$. The heaviness of the tails causes the empirical distribution $\hat{\mu}_n$ to converge very slowly to $\mu^{*n}$. 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.
Figure 2: $R_\rho(\mu^*)/n$ (dotdashed line) as well as the average of 100 Monte-Carlo paths of respectively $R_\rho(\mathcal{N}(\hat{m}_n, n\hat{\sigma}_n^2))/n$ (dashed line) and $R_\rho(\hat{\mu}_n^*)/n$ (continuous line) for $\rho = \text{VaR}_{0.99}$ 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.

4. A nonuniform Berry–Esséen inequality

The proof of Theorem 2.4 avails the following nonuniform Berry–Esséen inequality (28). The inequality involves the nonuniform Kolmogorov distance $d_{\phi,\lambda}$, which was introduced in (15).

**Theorem 4.1** Let $(X_i)$ be a sequence of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, P)$ such that $\text{Var}[X_1] > 0$ and $E[|X_1|^{\lambda}] < \infty$ for some $\lambda > 2$. For every $n \in \mathbb{N}$, let

$$Z_n := \frac{\sum_{i=1}^{n} (X_i - E[X_1])}{\sqrt{n \text{Var}[X_1]}}.$$
Then there exists a universal constant $C_\lambda \in (0, \infty)$ such that

$$d_{\phi,\lambda}(P_{Z_n}, N_{0,1}) \leq C_\lambda f(P_{X_1}) n^{-\gamma} \quad \text{for all } n \in \mathbb{N} \quad (28)$$

with $\gamma := \min\{1, \lambda - 2\}/2$, where for some universal constant $D_\lambda > 0$,

$$f(P_{X_1}) := \begin{cases} 
\frac{\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]}{\text{Var}[X_1]^{\lambda/2}}, & 2 < \lambda \leq 3 \\
\exp\left(D_\lambda \frac{\mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]}{\text{Var}[X_1]^{\lambda/2}}\right), & \lambda > 3
\end{cases} \quad (29)$$

By “universal constant” we mean that the constant is independent of $P_{X_1}$. Inequality (28) has been proven by Nagaev [3] and Bikelis [2] for $\lambda = 3$ and $\lambda \in (2, 3]$, respectively. Meanwhile there exist several estimates for the constant $C_\lambda$ for $\lambda \in (2, 3]$; see [4] and references cited therein. For $\lambda > 3$ the inequality has been stated by Michel [7] but he did not specify the function $f$ (and he left large parts of the proof to the reader). On the other hand, for the proof of part (iii) of Theorem 2.4 it is essential that $f(P_{X_1})$ is a continuous transformation of moments of $P_{X_1}$. We will now elaborate Michel’s approach in order to obtain Inequality (28) with $f(P_{X_1})$ as in (29).

**Proof** (of Theorem 4.1) As discussed above, the case $2 < \lambda \leq 3$ is already known. So we may and do assume $\lambda > 3$. In particular, for (28) it suffices to show

$$d_{\phi,\lambda}(P_{Z_n}, N_{0,1}) \leq C_\lambda \exp(D_\lambda \mathbb{E}[|X'_1|^\lambda]) n^{-1/2} \quad \text{for all } n \in \mathbb{N} \quad (30)$$

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 $(\Omega, \mathcal{F}, P)$ such that $\mathbb{E}[X'_1] = 0$, $\text{Var}[X'_1] = 1$, and $\mathbb{E}[|X'_1|^\lambda] < \infty$. Indeed, if specifically $X'_i := (X_i - \mathbb{E}[X_1])/\sqrt{\text{Var}[X_1]}$ in the setting of Theorem 4.1, then we have $Z_n = Z'_n$ and $\mathbb{E}[|X'_1|^\lambda] = \mathbb{E}[|X_1 - \mathbb{E}[X_1]|^\lambda]/\text{Var}[X_1]^{\lambda/2}$.

To verify (30), 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

$$|F_n(x) - \Phi_{0,1}(x)| \leq c_\lambda \mathbb{E}[|X'_1|^\lambda] n^{-(\lambda/2-1)}(1 + |x|^\lambda)^{-1} \quad \text{for } |x| < 1, \quad (31)$$

$$|F_n(x) - \Phi_{0,1}(x)| \leq c_\lambda e^{d_\lambda \mathbb{E}[|X'_1|^\lambda]/n} n^{-1/2}|x|^{-\lambda} \quad \text{for } 1 \leq |x| \leq \sqrt{(\lambda - 1) \log n}, \quad (32)$$

$$|F_n(x) - \Phi_{0,1}(x)| \leq c_\lambda e^{d_\lambda \mathbb{E}[|X'_1|^\lambda] n^{-(\lambda/2-1)}|x|^{-\lambda}} \quad \text{for } |x| > \max\{1, \sqrt{(\lambda - 1) \log n}\} \quad (33)$$

hold for all $n \in \mathbb{N}$, where $c_\lambda, d_\lambda > 0$ refer to any constants depending only on $\lambda$ and being independent of the distribution of $X'_1$. Inequalities (31)–(33) clearly imply (30).

**Step 1.** Inequality (31) follows from Katz’ generalization of the classical Berry–Esséen inequality. In [1], 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 $x/g(x) \leq g'(y)$ for all $0 \leq x \leq y$. Then for any sequence $(Y_i)$ of i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}[Y_1] = 0$,
$\mathbb{E}[Y_i^2] < \infty$ and $\mathbb{E}[Y_i^2 g(Y)] < \infty$ there exists an universal constant $C_g \in (0, \infty)$ (i.e. independent of $\mathbb{P}_Y$) such that

$$d_{\mu_0}(\mathbb{P}_{W_n}, \mathcal{N}_{0,1}) \leq (C_g \mathbb{E}[Y_i^2 g(Y_i)]) g(\sqrt{n})^{-1}$$

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 (31).

Step 2. We now prove (32). It suffices to show that there exists some constant $\tilde{c}_\lambda > 0$ depending only on $\lambda$ and being independent of the distribution of $X_1'$ such that (32) holds for all $n \geq n_0 := [\tilde{c}_\lambda \mathbb{E}[][X_1'|]^8]^{1/2}$ (this observation will be relevant in Steps 2.2.2 and 2.2.3 below). Indeed, for $n < n_0$ we get (32) from Katz’ generalization of the classical Berry–Esséen inequality (cf. Step 1) as follows:

$$\sup_{x \in [-\sqrt{(\lambda-1) \log n}, \sqrt{(\lambda-1) \log n}]} |F_n(x) - \Phi_{0,1}(x)| (1 + |x|^\lambda)$$

\[\leq \sup_{x \in [-\sqrt{(\lambda-1) \log n_0}, \sqrt{(\lambda-1) \log n_0}]} |F_n(x) - \Phi_{0,1}(x)| (1 + |x|^\lambda)\]

\[\leq \|F_n - \Phi_{0,1}\|_{\infty} (1 + ((\lambda - 1) \log n_0)^\lambda)\]

\[\leq c_{\lambda,1} \mathbb{E}[][X_1'|]^{-\lambda/2} (1 + ((\lambda - 1) \log(\mathbb{E}[][X_1'|]^8)))^\lambda\]

\[\leq c_{\lambda,2} \mathbb{E}[][X_1'|]^2 n^{-1/2}.

Without loss of generality we restrict ourselves to $1 \leq x \leq \max\{1; \sqrt{(\lambda-1) \log n}\}$. Let $r_\lambda \in (0, \min\{1; \lambda - 3\}/(2(\lambda - 1)))$, consider the truncations

$$X_n^{i,x} := X_1' \mathbb{1}_{\{|X_1'| \leq r_\lambda n^{1/2}x\}}, \quad 1 \leq i \leq n, \quad n \in \mathbb{N},$$

and set $\tilde{Z}_n^{i,x} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_n^{i,x}$. We have

$$|F_n(x) - \Phi_{0,1}(x)| = |(1 - F_n(x)) - (1 - \Phi_{0,1}(x))|$$

\[\leq |\mathbb{P}[Z_n' > x] - \mathbb{P}[\tilde{Z}_n^{i,x} > x]| + |\mathbb{P}[\tilde{Z}_n^{i,x} > x] - (1 - \Phi_{0,1}(x))|\]

\[= |\mathbb{P}[Z_n' > x] - \mathbb{P}[\tilde{Z}_n^{i,x} > x]| + |\mathbb{P}[\tilde{Z}_n^{i,x} > x] - \Phi_{0,1}(-x)|. \quad (34)\]

In Steps 2.1–2.2 below we will show that

$$|\mathbb{P}[Z_n' > x] - \mathbb{P}[\tilde{Z}_n^{i,x} > x]| \leq c_{\lambda,1} \mathbb{E}[][X_1'|]^{-\lambda/2} x^{-\lambda} \quad (35)$$

and

$$|\mathbb{P}[\tilde{Z}_n^{i,x} > x] - \Phi_{0,1}(-x)| \leq c_{\lambda,2} \mathbb{E}[][X_1'|]^2 n^{-1/2} x^{-\lambda}. \quad (36)$$

Then, (34)–(36) imply (32).

Step 2.1. To prove (35), note that

$$\mathbb{P}[Z_n' > x] = \mathbb{P}[\{Z_n' > x\} \cap \{X_1' = X_n^{1,x}, \ldots, X_n' = X_n^{n,x}\}]$$

\[+ \mathbb{P}[\{Z_n' > x\} \cap \{\text{there exists } 1 \leq i \leq n \text{ with } X_i' \neq X_n^{i,x}\}]

\[\leq \mathbb{P}[\{Z_n^{i,x} > x\} \cap \{X_1' = X_n^{1,x}, \ldots, X_n' = X_n^{n,x}\}] + n \mathbb{P}[X_1' \neq X_n^{1,x}]

\[\leq \mathbb{P}[\tilde{Z}_n^{i,x} > x] + n \mathbb{P}[|X_1'| > r_\lambda n^{1/2}x]. \quad (37)$$
and
\[
\mathbb{P}[\tilde{Z}^{n,x}_n > x] = \mathbb{P}\{\tilde{Z}^{n,x}_n > x\} \cap \{X'_1 = X'^{n,x}_1, \ldots, X'_n = X'^{n,x}_n\} \\
+ \mathbb{P}\{{\tilde{Z}^{n,x}_n > x}\} \cap \{\text{there exists } 1 \leq i \leq n \text{ with } X'_i \neq X'^{n,x}_i\}
\leq \mathbb{P}\{\tilde{Z}^{n,x}_n > x\} \cap \{X'_1 = X'^{n,x}_1, \ldots, X'_n = X'^{n,x}_n\} + n \mathbb{P}[X'_1 \neq X'^{n,x}_1]
\leq \mathbb{P}[Z'_n > x] + n \mathbb{P}[|X'_1| > r\lambda n^{1/2}x].
\]

Then (37)–(38) and an application of Markov’s inequality give
\[
|\mathbb{P}[Z'_n > x] - \mathbb{P}[\tilde{Z}^{n,x}_n > x]| \leq n \mathbb{P}[|X'_1| > r\lambda n^{1/2}x]
\leq n E[|X'_1|^\lambda]/(r\lambda n^{1/2}x)^\lambda
\leq r^{-\lambda} E[|X'_1|^\lambda] n^{-(\lambda/2)-1} x^{-\lambda}.
\]

That is, (35) holds for \(c_{\lambda,1} := r^{-\lambda}\).

Step 2.2. To verify (36) we consider the probability measure \(Q_{n,x}\) on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) defined by
\[
Q_{n,x}[A] := \frac{1}{\beta_{n,x}} \int_A e^{x^{-1/2}x_1} \mathbb{P}_X^{n,x}(dx_1), \quad A \in \mathcal{B}(\mathbb{R}),
\]
where \(\beta_{n,x} := \int e^{x^{-1/2}x_1} \mathbb{P}_X^{n,x}(dx_1)\). In particular,
\[
\frac{d\mathbb{P}_X^{n,x}}{dQ_{n,x}}(x_1) = \beta_{n,x} e^{-x^{-1/2}x_1} \quad \text{for all } x_1 \in \mathbb{R}.
\]

It follows that the \(n\)-fold product measure \(Q_{n,x}^\otimes \) of \(Q_{n,x}\) satisfies
\[
Q_{n,x}^\otimes [A] = \frac{1}{\beta_{n,x}^n} \int_A e^{x^{-1/2}X_1} \sum_{i=1}^n x_i \mathbb{P}_X^{n,x}(dx_1, \ldots, dx_n) \quad \text{for all } A \in \mathcal{B}(\mathbb{R}^n).
\]

In particular,
\[
\frac{d\mathbb{P}_X^{n,x}}{dQ_{n,x}^\otimes}(x_1, \ldots, x_n) = \beta_{n,x}^n e^{-x^{-1/2}X_1} \sum_{i=1}^n x_i \quad \text{for all } (x_1, \ldots, x_n) \in \mathbb{R}^n.
\]

Using the notation
\[
m_{n,x} := \mathbb{E}_{Q_{n,x}}[X_1^{n,x}] = \int x_1 Q_{n,x}(dx_1)
\]
we obtain
\[
\mathbb{P}[\tilde{Z}^{n,x}_n > x] = \mathbb{P}\left[n^{-1/2} \sum_{i=1}^n (X_i^{n,x} - m_{n,x}) > x - n^{1/2}m_{n,x}\right]
\]
17
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 (39) we obtain

\[
\|\mathbb{P}[Z_{n,x} > x] - \Phi_{0,1}(-x)]\| \\
\leq |\beta_{n,x}^n e^{-x n^{1/2} m_{n,x}} \int_{(x-n^{1/2} m_{n,x}, \infty)} e^{-xz} \Pi_{n,x}(dz) - \Phi_{0,1}(-x)| \\
+ e^{-x^2/2} \int_{(x-n^{1/2} m_{n,x}, \infty)} e^{-xz} \Pi_{n,x}(dz) - \int_{(x-n^{1/2} m_{n,x}, \infty)} e^{-xz} N_{0,s_{n,x}^2}(dz) \\
+ |e^{-x^2/2} \int_{(x-n^{1/2} m_{n,x}, \infty)} e^{-xz} N_{0,s_{n,x}^2}(dz) - \Phi_{0,1}(-x)| \\
=: S_{2,2,1}(\lambda, n, x) + S_{2,2,2}(\lambda, n, x) + S_{2,2,3}(\lambda, n, x),
\]  

(39)

where

\[
s_{n,x} := \var_{\mathbb{Q}_{n,x}}[X_1^{n,x}]^{1/2} = \left( \int x_1^2 \mathbb{Q}_{n,x}(dx_1) - m_{n,x}^2 \right)^{1/2}.
\]

In Steps 2.2.1–2.2.3 below we will show that

\[
S_{2,2,1}(\lambda, n, x) \leq c_{\lambda,3} e^{\lambda,4 E[|X_1|^{\lambda}]} n^{-1/2} x^{10} e^{-x^2/2},
\]

(40)

\[
S_{2,2,2}(\lambda, n, x) \leq c_{\lambda,5} e^{\lambda,6 E[|X_1|^{\lambda}]} n^{-1/2} x^2 e^{-x^2/2},
\]

(41)

\[
S_{2,2,3}(\lambda, n, x) \leq c_{\lambda,7} e^{\lambda,8 E[|X_1|^{\lambda}]} n^{-1/2} x^9 e^{-x^2/2},
\]

(42)

which gives (36).

Step 2.2.0.a. First of all we observe that

\[
|\mathbb{E}[X_1^{n,x}]} | \leq c_{\lambda,9} E[|X_1|^{\lambda}] n^{-(\lambda-1)/2} x^{-(\lambda-1)},
\]

(43)

\[
\mathbb{E}[(X_1^{n,x})^2] \leq 1,
\]

(44)

\[
\mathbb{E}[(X_1^{n,x})^2] \geq 1 - c_{\lambda,10} E[|X_1|^{\lambda}] n^{-(\lambda-2)/2} x^{-(\lambda-2)},
\]

(45)

\[
\mathbb{E}[|X_1|^{r}] \leq E[|X_1|^{\lambda}] \text{ for } 2 \leq r \leq \lambda.
\]

(46)
\[
\mathbb{E}[(X_1^{n,x})^4 e^{x \cdot n^{-1/2} |X_1^{n,x}|}] \leq c_{\lambda,11} \mathbb{E}[|X_1'|^\lambda] n^{r_x(\lambda-1) + (4-\lambda)/2 \cdot x} \quad \text{for } \lambda \in (3, 4),
\]
\[
\mathbb{E}[(X_1^{n,x})^4 e^{x \cdot n^{-1/2} |X_1^{n,x}|}] \leq \mathbb{E}[|X_1'|^\lambda] n^{r_x(\lambda-1)} \quad \text{for } \lambda \geq 4.
\] (47) (48)

Indeed: In view of \(X_1^{n,x} = X_1' - X_1' \mathbb{I}_{\{|X_1'| > r_x n^{1/2} x\}}\) and \(\mathbb{E}[X_1'] = 0\) we have
\[
|\mathbb{E}[X_1^{n,x}]| = |\mathbb{E}[X_1' \mathbb{I}_{\{|X_1'| > r_x n^{1/2} x\}}]| \\
\leq \mathbb{E}[|X_1'| \mathbb{I}_{\{|X_1'| > r_x n^{1/2} x\}}] \\
\leq \mathbb{E} \left[ \frac{|X_1'|^\lambda}{(r_x n^{1/2} x)^{-1-\lambda}} \mathbb{I}_{\{|X_1'| > r_x n^{1/2} x\}} \right] \\
\leq r^{-\lambda}_{\lambda} \mathbb{E}[|X_1'|^\lambda] n^{-(\lambda-1)/2 \cdot x} x^{-(\lambda-1)},
\]
which proves (43) with \(c_{\lambda,9} = r^{-\lambda}_{\lambda}\). Inequality (41) is justified by \(\mathbb{E}[(X_1^{n,x})^2] \leq \mathbb{E}[(X_1')^2] = 1\), and inequality (45) can be obtained as follows:
\[
\mathbb{E}[(X_1^{n,x})^2] - 1 = \mathbb{E}[(X_1')^2 - (X_1')^2] \\
= -\mathbb{E}[(X_1')^2 \mathbb{I}_{\{|X_1'| > r_x n^{1/2} x\}}] \\
\geq -\mathbb{E} \left[ \frac{|X_1'|^\lambda}{(r_x n^{1/2} x)^{-1-\lambda}} \mathbb{I}_{\{|X_1'| > r_x n^{1/2} x\}} \right] \\
\geq -r^{-\lambda}_{\lambda} \mathbb{E}[|X_1'|^\lambda] n^{-(\lambda-2)/2 \cdot x} x^{-(\lambda-2)}.
\]
Due to the assumption \(\mathbb{E}[|X_1'|^2] = 1\) and Jensen’s inequality we obtain
\[
1 = \mathbb{E}[|X_1'|^2]^{1/2} \leq \mathbb{E}[|X_1'|^r]^{1/r} \leq \mathbb{E}[|X_1'|^\lambda]^{1/\lambda},
\]
which leads to (46) for \(2 \leq r \leq \lambda\). Since \(|X_1^{n,x}| \leq r_x n^{1/2} x\) and \(x^2 \leq (\lambda - 1) \log n\), we obtain for \(\lambda \in (3, 4)\) that
\[
\mathbb{E}[(X_1^{n,x})^4 e^{x \cdot n^{-1/2} |X_1^{n,x}|}] \leq \mathbb{E}[(X_1^{n,x})^4] e^{x^2} \\
\leq \mathbb{E}[|X_1'|^\lambda] (r_x n^{1/2} x)^{4-\lambda} n^{r_x(\lambda-1)} \\
= r^{-\lambda}_{\lambda} x^{4-\lambda} n^{r_x(\lambda-1) + (4-\lambda)/2} \mathbb{E}[|X_1'|^\lambda].
\]
This proves (47) with \(c_{\lambda,11} = r^{-\lambda}_{\lambda}\). Finally, (48) follows by (46), \(|X_1^{n,x}| \leq r_x n^{1/2} x\), and \(x^2 \leq (\lambda - 1) \log n\).

\textit{Step 2.2.0.b.} Next we will prove that the following auxiliary inequalities hold:
\[
|\beta_{n,x} - 1 - x^2/(2n)| \leq c_{\lambda,12} \mathbb{E}[|X_1'|^\lambda] n^{-3/2} x^5, 
\]
\[
|m_{n,x} - xn^{-1/2}| \leq c_{\lambda,13} \mathbb{E}[|X_1'|^\lambda] n^{-1} x^6, 
\]
\[
|s_{n,x}^2 - 1| \leq c_{\lambda,14} \mathbb{E}[|X_1'|^\lambda]^2 n^{-1/2} x^{12}.
\]
(49) (50) (51)

We first show (49). Using (44), we obtain
\[
|\beta_{n,x} - 1 - x^2/(2n)|
\]
where for the last step we used $\beta$.

For (52) we observe that since $\lambda \leq (3, 4)$ we can use (43), (45), (46), and (47) to conclude

$$|\beta_{n,x} - 1 - x^2/(2n)| \leq c_{\lambda,9} x^2 - \lambda - \lambda/2 \mathbb{E}[|X|^4] + c_{\lambda,15} x^4 - \lambda - \lambda/2 \mathbb{E}[|X|^4] + \frac{1}{3!} x^3 n^{-3/2} \mathbb{E}[|X|^3]$$

$$+ \frac{1}{4!} x^4 n^{-2} \mathbb{E}[|X|^4] n^{-\lambda+5}\lambda^{-\lambda}(\lambda - 1)/2 x^{4-\lambda} \leq c_{\lambda,16} \mathbb{E}[|X|^4] n^{-3/2} x^3 + c_{\lambda,11} \frac{1}{4!} x^4 n^{-2} \mathbb{E}[|X|^4] n^{-\lambda+2}\lambda^{-\lambda}(\lambda - 1) \mathbb{E}[|X|^3] \leq c_{\lambda,17} \mathbb{E}[|X|^4] n^{-3/2} x^5,$$

where for the last step we used $n^{-\lambda+2}\lambda^{-\lambda}(\lambda - 1) \leq n^{-3/2}$ (which follows from the assumption $\lambda \leq (3, 4)/(2\lambda - 1)$). On the other hand, for $\lambda \geq 4$ we can use (43), (45), (46), and (48) to conclude

$$|\beta_{n,x} - 1 - x^2/(2n)| \leq c_{\lambda,9} x^2 - \lambda - \lambda/2 \mathbb{E}[|X|^4] + c_{\lambda,15} x^4 - \lambda - \lambda/2 \mathbb{E}[|X|^4] + \frac{1}{3!} x^3 n^{-3/2} \mathbb{E}[|X|^3]$$

$$+ \frac{1}{4!} x^4 n^{-2} \mathbb{E}[|X|^4] n^{-\lambda+5}\lambda^{-\lambda}(\lambda - 1)/2 x^{4-\lambda} \leq c_{\lambda,18} \mathbb{E}[|X|^4] n^{-3/2} x^3 + c_{\lambda,11} \frac{1}{4!} x^4 n^{-2} \mathbb{E}[|X|^4] n^{-\lambda+2}\lambda^{-\lambda}(\lambda - 1) \mathbb{E}[|X|^3] \leq c_{\lambda,19} \mathbb{E}[|X|^4] n^{-3/2} x^5,$$

where for the last step we used $n^{-\lambda+2}\lambda^{-\lambda}(\lambda - 1) \leq n^{-3/2}$ (which follows from the assumption $\lambda \leq 1/(2\lambda - 1)$). This completes the proof of (49).

To prove (50), we will show that the following inequalities hold:

$$m_{n,x} - x n^{-1/2} \leq c_{\lambda,20} \mathbb{E}[|X|^4] n^{-1} x^4,$$  \hspace{1cm} (52)

$$x n^{-1/2} - m_{n,x} \leq c_{\lambda,21} \mathbb{E}[|X|^4] n^{-1} x^6.$$  \hspace{1cm} (53)

For (52) we observe that since $\beta_{n,x} \geq 1$,

$$m_{n,x} - x n^{-1/2} = \beta_{n,x} - 1 \mathbb{E}[X_{n,x}^2 e^{x n^{-1/2} X_{n,x}}] - x n^{-1/2} \leq \mathbb{E}[X_{n,x} e^{x n^{-1/2} X_{n,x}}] - x n^{-1/2} \leq |\mathbb{E}[X_{n,x}^2]| + x n^{-1/2} (\mathbb{E}[X_{n,x}^2] - 1) + \frac{1}{2} x^2 n^{-1} \mathbb{E}[|X_{n,x}^3|] + \frac{1}{3!} x^3 n^{-3/2} \mathbb{E}[|X_{n,x}^4| e^{x n^{-1/2} X_{n,x}}].$$
On the one hand, for \( \lambda \in (3, 4) \) we can use (43), (44), (46), and (47) to conclude
\[
m_{n,x} - xn^{-1/2} \leq c_{\lambda,9} x^{1-\lambda} n^{(1-\lambda)/2} E[|X'_1|^\lambda] + \frac{1}{2} x^2 n^{-1} E[|X_1'|^\lambda] + c_{\lambda,22} x^{7-\lambda} n^{(1-\lambda)/2 + r_{\lambda}(\lambda-1)} E[|X_1'|^\lambda]
\]
where for the last step we used \( n^{(1-\lambda)/2 + r_{\lambda}(\lambda-1)} \leq n^{-1} \) (which follows from the assumption \( r_{\lambda} \leq (\lambda - 3)/(2(\lambda - 1)) \)). On the other hand, for \( \lambda \geq 4 \) we can use (43), (44), (46), and (48) to conclude
\[
m_{n,x} - xn^{-1/2} \leq c_{\lambda,23} E[|X'_1|^\lambda] n^{-1/4},
\]
where for the last step we used \( n^{r_{\lambda}(\lambda-1)-2} \leq n^{-3/2} \) (which follows from the assumption \( r_{\lambda} \leq 1/(2(\lambda - 1)) \)). This proves (52). We will now prove (53). In view of (43), (45), (46), \( \beta_{n,x} \geq 1 \), \( x^2 \leq (\lambda - 1) \log n \), and (49) we obtain
\[
xn^{-1/2} - m_{n,x} = x n^{-1/2} - \beta_{n,x}^{-1} E[X_1^{n,x} e^{xn^{-1/2}X_1^{n,x}}] = x n^{-1/2} - \beta_{n,x}^{-1} \left( E[X_1^{n,x}] + xn^{-1/2} E[(X_1^{n,x})^2] + x^2 n^{-1} E[(X_1^{n,x})^3] + \sum_{i=3}^{\infty} \frac{(xn^{-1/2})^i}{i!} (X_1^{n,x})^{i+1} \right) \leq x n^{-1/2} - \beta_{n,x}^{-1} \left( -c_{\lambda,9} E[|X'_1|^\lambda] n^{-(\lambda-1)/2} e^{-x^{-1/2}X_1^{n,x}} + xn^{-1/2} (1 - c_{\lambda,10} E[|X'_1|^\lambda] n^{-(\lambda-2)/2} e^{-x^{-1/2}X_1^{n,x}}) - x^2 n^{-1} E[|X'_1|^\lambda] - x^3 n^{-3/2} E[(X_1^{n,x})^4 e^{xn^{-1/2}X_1^{n,x}}] \right) \leq c_{\lambda,25} x^2 n^{-1} E[|X'_1|^\lambda] + x^3 n^{-3/2} E[(X_1^{n,x})^4 e^{xn^{-1/2}X_1^{n,x}}] + xn^{-1/2} (1 - \beta_{n,x}^{-1}) \leq c_{\lambda,25} x^2 n^{-1} E[|X'_1|^\lambda] + x^3 n^{-3/2} E[(X_1^{n,x})^4 e^{xn^{-1/2}X_1^{n,x}}] + xn^{-1/2} (\beta_{n,x} - 1) \leq c_{\lambda,25} x^2 n^{-1} E[|X'_1|^\lambda] + x^3 n^{-3/2} E[(X_1^{n,x})^4 e^{xn^{-1/2}X_1^{n,x}}] + xn^{-1/2} \left( c_{\lambda,12} x^5 n^{-3/2} E[|X'_1|^\lambda] \right) + \frac{x^2}{2n} \leq c_{\lambda,26} x^6 n^{-1} E[|X'_1|^\lambda] + x^3 n^{-3/2} E[(X_1^{n,x})^4 e^{xn^{-1/2}X_1^{n,x}}].
\]
For \( \lambda \in (3, 4) \) we can use (47) to deduce
\[
xn^{-1/2} - m_{n,x} \leq c_{\lambda,26} x^6 n^{-1} E[|X'_1|^\lambda] + c_{\lambda,11} x^{7-\lambda} n^{1/2-\lambda/2 + r_{\lambda}(\lambda-1)} E[|X'_1|^\lambda] \leq c_{\lambda,27} x^6 n^{-1} E[|X'_1|^\lambda],
\]
(54)
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 (48) to obtain

$$xn^{-1/2} - m_{n,x} \le c_{n,28} n^6 n^{-1} E[|X|^{\lambda}] + x^3 n^{-3/2+r,\lambda(\lambda-1)} E[|X|^{\lambda}]$$

$$\le c_{n,28} n^6 n^{-1} E[|X|^{\lambda}],$$

(55)

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 (54) and (55) lead to (53).

To prove (51) we will show that the following inequalities hold:

$$s_{n,x}^2 - 1 \le c_{n,29} E[|X|^{\lambda}] n^{-1/2} x^3$$

$$1 - s_{n,x}^2 \le c_{n,30} E[|X|^{\lambda}]^2 n^{-1/2} x^2,$$

(56)

(57)

First we will prove (56). By virtue of $\beta_{n,x} \ge 1$, (44), and (46), we obtain

$$s_{n,x}^2 - 1 = (\beta_{n,x}^{-1} E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}] - m_{n,x}^2 - 1$$

$$\le E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}] - 1$$

$$\le E[(X_{1,x}^n)^2] - 1 + x n^{-1/2} E[|X_{1,x}^n|^3] + x^2 n^{-1} E[(X_{1,x}^n)^4 e^{x n^{-1/2} X_{1,x}^n}]$$

$$\le x n^{-1/2} E[|X_{1,x}^n|^3] + x^2 n^{-1} E[(X_{1,x}^n)^4 e^{x n^{-1/2} X_{1,x}^n}],$$

For $\lambda \in (3, 4)$ we can use (47) to deduce

$$s_{n,x}^2 - 1 \le x n^{-1/2} E[|X_{1,x}^n|^\lambda] + c_{n,11} x^3 n^{-1/2+\lambda} E[|X_{1,x}^n|^\lambda]$$

$$\le c_{n,31} x^3 n^{-1/2} E[|X_{1,x}^n|^\lambda],$$

where for the last step we used $n^{1/2-\lambda/2+r,\lambda(\lambda-1)} \le n^{-1/2}$ (which follows from the assumption $r, \lambda \le (\lambda-3)/(2(\lambda-1))$). For $\lambda \ge 4$ we can use (48) to obtain

$$s_{n,x}^2 - 1 \le x n^{-1/2} E[|X_{1,x}^n|^\lambda] + x^2 n^{-1+\lambda,\lambda(\lambda-1)} E[|X_{1,x}^n|^\lambda]$$

$$\le c_{n,32} x^2 n^{-1/2} E[|X_{1,x}^n|^\lambda],$$

where for the last step we used $n^{-1+\lambda,\lambda(\lambda-1)} \le n^{-1/2}$ (which follows from the assumption $r, \lambda \le 1/(2(\lambda-1))$). This proves (56). We next prove (57). Using $\beta_{n,x} \ge 1$ and (50), we obtain

$$1 - s_{n,x}^2 = 1 - (\beta_{n,x}^{-1} E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}] - m_{n,x}^2 - 1)$$

$$= 1 + m_{n,x}^2 - \beta_{n,x}^{-1} E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}]$$

$$\le 1 + (c_{n,13} x^3 n^{-1} E[|X_{1,x}^n|^\lambda] + x^2 n^{-1/2})^2 - \beta_{n,x}^{-1} E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}]$$

$$\le 1 + c_{n,33} x^2 n^{-1} E[|X_{1,x}^n|^\lambda] - \beta_{n,x}^{-1} E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}],$$

(58)

Now, (51) would follow if we can show that

$$E[(X_{1,x}^n)^2 e^{x n^{-1/2} X_{1,x}^n}] \ge 1 - c_{n,34} n^{-1/2} x^2 E[|X_{1,x}^n|^\lambda],$$

(59)
\[ 1 - s_n^2 \leq 1 + c_{\lambda,33} n^{-1.12} \mathbb{E}[X_1^2|-\lambda|] - \beta_{n,x}^{-1} (1 - c_{\lambda,34} n^{-1/2} x^3 \mathbb{E}[X_1^3|-\lambda|]) \leq c_{\lambda,33} n^{-1.12} \mathbb{E}[X_1^2|-\lambda|] + c_{\lambda,34} n^{-1/2} x^3 \mathbb{E}[X_1^3|-\lambda|] + (1 - \beta_{n,x}^{-1}) \leq c_{\lambda,33} n^{-1.12} \mathbb{E}[X_1^2|-\lambda|] + c_{\lambda,34} n^{-1/2} x^3 \mathbb{E}[X_1^3|-\lambda|] + (\beta_{n,x} - 1) \leq c_{\lambda,33} n^{-1.12} \mathbb{E}[X_1^2|-\lambda|] + c_{\lambda,34} n^{-1/2} x^3 \mathbb{E}[X_1^3|-\lambda|] + c_{\lambda,32} n^{-3/2} x^5 \mathbb{E}[X_1^5|-\lambda|] + x^2/(2n) \leq c_{\lambda,35} \mathbb{E}[X_1^2|-\lambda|] n^{-1/2} x^{12}. \]

To prove (59) we use (45) and (46) to obtain
\[
\mathbb{E}[(X_1^n x)^2 e^{x n^{-1/2} X_1^n}] = \mathbb{E}[(X_1^n x)^2] + x n^{-1/2} \mathbb{E}[(X_1^n x)^3] + \mathbb{E} \left[ \sum_{i=2}^{\infty} \frac{(x n^{-1/2})^i}{i!} (X_1^n x)^{i+2} \right] \geq 1 - c_{\lambda,10} n^{-(\lambda-2)/2} x^{-2 \lambda / 2} \mathbb{E}[X_1^\lambda] - x n^{-1/2} \mathbb{E}[X_1^\lambda] - x^2 n^{-1} \mathbb{E}[(X_1^n x)^4 e^{x n^{-1/2} X_1^n}] \geq 1 - c_{\lambda,30} x n^{-1/2} \mathbb{E}[X_1^\lambda] - x^2 n^{-1} \mathbb{E}[(X_1^n x)^4 e^{x n^{-1/2} X_1^n}].
\]

If \( \lambda \in (3, 4) \) we can use (47) to deduce
\[
\mathbb{E}[(X_1^n x)^2 e^{x n^{-1/2} X_1^n}] \geq 1 - c_{\lambda,36} x n^{-1/2} \mathbb{E}[X_1^\lambda] - c_{\lambda,11} x^6 \lambda n^{-1 + r_\lambda (\lambda - 1) - (\lambda - 4)/2} \mathbb{E}[X_1^\lambda] \geq 1 - c_{\lambda,37} x^3 n^{-1/2} \mathbb{E}[X_1^\lambda],
\]

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 (48), yielding
\[
\mathbb{E}[(X_1^n x)^2 e^{x n^{-1/2} X_1^n}] \geq 1 - c_{\lambda,36} x n^{-1/2} \mathbb{E}[X_1^\lambda] - x^2 n^{r_\lambda (\lambda - 1) - 1} \mathbb{E}[X_1^\lambda] \geq 1 - c_{\lambda,38} x^2 n^{-1/2} \mathbb{E}[X_1^\lambda],
\]

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 (59).

**Step 2.2.1.** In this part we will verify the inequalities
\[
\beta_{n,x}^n e^{-x n^{1/2} m_{n,x}} - e^{-x^2/2} \leq c_{\lambda,39} n^{-1/2} x^7 e^{-x^2/2} e^{\lambda,40} \mathbb{E}[X_1^\lambda],
\]

\[
e^{-x^2/2} - \beta_{n,x}^n e^{-x n^{1/2} m_{n,x}} \leq c_{\lambda,41} n^{-1/2} x^{10} e^{-x^2/2} e^{\lambda,42} \mathbb{E}[X_1^\lambda],
\]

which imply (60). First we will show (61). 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, (49), (50), and the assumption \( x^2 \leq (1 - \lambda) \log n \) we obtain
\[
\beta_{n,x}^n e^{-x n^{1/2} m_{n,x}} - e^{-x^2/2} \leq e^n(\beta_{n,x} - 1) x n^{1/2} m_{n,x} - e^{-x^2/2} \leq e^{-x^2/2}/2.
\]

23
\[
\text{where we assumed without loss of generality } \; xn^{1/2}m_{n,x} - n(\beta_{n,x} - 1) \leq x^2/2 \; \text{(otherwise we obtain the trivial upper bound 0)}. \quad \text{Next we will show that (61) holds true.}
\]

\[ e^{-x^2/2} - \beta_{n,x} e^{-x n^{1/2} m_{n,x}} = \]
\[ = e^{-x^2/2} (1 - e^{-x^2/2 + n \log \beta_{n,x} - x n^{1/2} m_{n,x}}) \]
\[ = e^{-x^2/2} (1 - e^{-x n^{1/2} m_{n,x} - x^2/2 - n \log \beta_{n,x}}) \]
\[ \leq e^{-x^2/2} (x n^{1/2} m_{n,x} - x^2/2 - n \log \beta_{n,x}) \]
\[ \leq e^{-x^2/2} \left( c_{\lambda,13} n^{-1} x^6 E[|X'|^4] + x n^{-1/2} \right) - x^2/2 - n(\beta_{n,x} - 1 - (\beta_{n,x} - 1)^2/2) \]
\[ \leq e^{-x^2/2} \left( c_{\lambda,13} n^{-1/2} x^7 E[|X'|^4] + x^2/2 + n \left\{ c_{\lambda,12} n^{-3/2} x^5 E[|X'|^4] - x^2/(2n) \right\} \right) \]
\[ + n \left\{ c_{\lambda,12} n^{-3/2} x^5 E[|X'|^4] + x^2/(2n) \right\}^2/2 \]
\[ \leq e^{-x^2/2} \left( c_{\lambda,13} n^{-1/2} x^7 E[|X'|^4] + c_{\lambda,12} n^{-1/2} x^5 E[|X'|^4] + c_{\lambda,12} n^{-2} x^{10} E[|X'|^4]^2 + x^4/4n \right) \]
\[ \leq c_{\lambda,45} e^{-x^2/2} n^{-1/2} x^{10} e^{c_{\lambda,46} E[|X'|^4]} \],

where we assumed without loss of generality \( xn^{1/2} m_{n,x} - n \log \beta_{n,x} \geq x^2/2 \) (otherwise we obtain the trivial upper bound 0). Note that this assumption allows us to apply the inequality \( 1 - e^{-z} \leq z \) for \( z \geq 0 \), to the third line in the upper calculations. This proves (61).

**Step 2.2.2.** Next we will prove (11). Let \( F_{n,x} \) denote the distribution function of \( \Pi_{n,x} \), and note that \( \Phi_{0,s^2_n,x} \), and \( \Psi_x \) are of bounded variation on every right-sided half-line, where \( \Psi_x(z) := e^{-xz} \). So integration-by-parts yields

\[
\left| \int_{(x - n^{1/2} m_{n,x}, \infty)} e^{-xz} \Pi_{n,x}(dz) - \int_{(x - n^{1/2} m_{n,x}, \infty)} e^{-xz} \mathcal{N}_0, s^2_n, x (dz) \right| = \left| \int_{x - n^{1/2} m_{n,x}}^{\infty} \Psi_x(z) dF_{n,x}(z) - \int_{x - n^{1/2} m_{n,x}}^{\infty} \Psi_x(z) d\Phi_{0,s^2_n,x}(z) \right|
\]

24
Now for the numerator in (63) we can use (46) to obtain

\[
\begin{align*}
&\lim_{b \to \infty} \left( e^{-xb} F_{\Pi_n,x}(b) - e^{-x(x-n^{1/2}m_{n,x})} F_{\Pi_n,x}(x - n^{1/2}m_{n,x}) - \int_{x-n^{1/2}m_{n,x}}^{b} F_{\Pi_n,x}(z) \, d\Psi_x(z) \right) \\
&\quad\quad\quad - \lim_{b \to \infty} \left( e^{-xb} \Phi_{0,s_{n,x}^2}(b) + e^{-x(x-n^{1/2}m_{n,x})} \Phi_{0,s_{n,x}^2}(x - n^{1/2}m_{n,x}) - \int_{x-n^{1/2}m_{n,x}}^{b} \Phi_{0,s_{n,x}^2}(z) \, d\Psi_x(z) \right) \\
&\quad\leq e^{-x(x-n^{1/2}m_{n,x})} \left| F_{\Pi_n,x}(x - n^{1/2}m_{n,x}) - \Phi_{0,s_{n,x}^2}(x - n^{1/2}m_{n,x}) \right| \\
&\quad\quad\quad + \int_{x-n^{1/2}m_{n,x}}^{\infty} \left| F_{\Pi_n,x}(z) - \Phi_{0,s_{n,x}^2}(z) \right| \, d\Psi_x(z) \\
&\quad\leq e^{-x(x-n^{1/2}m_{n,x})} \left\| F_{\Pi_n,x} - \Phi_{0,s_{n,x}^2} \right\|_\infty + \left\| F_{\Pi_n,x} - \Phi_{0,s_{n,x}^2} \right\|_\infty \int_{x-n^{1/2}m_{n,x}}^{\infty} d\Psi_x(z) \\
&\quad= 2e^{-x(x-n^{1/2}m_{n,x})} \left\| F_{\Pi_n,x} - \Phi_{0,s_{n,x}^2} \right\|_\infty.
\end{align*}
\] (62)

Furthermore we observe that

\[
\left\| F_{\Pi_n,x} - \Phi_{0,s_{n,x}^2} \right\|_\infty = \sup_{y \in \mathbb{R}} \left| F_{\Pi_n,x}(s_{n,x} y) - \Phi_{0,s_{n,x}^2}(s_{n,x} y) \right| = \left\| F_{\Pi_n,x} - \Phi_{0,1} \right\|_\infty,
\]

where \( \Pi_{n,x} \) refers to the image measure of of the probability measure \( \mathbb{Q}^{\otimes n}_{\Pi_{n,x}} \) w.r.t. the mapping \((x_1, \ldots, x_n) \mapsto n^{-1/2} \sum_{i=1}^{n} (x_i - m_{n,x}) / s_{n,x} \). Hence by the classical Berry–Esséen theorem we have

\[
\left\| F_{\Pi_{n,x}} - \Phi_{0,1} \right\|_\infty \leq \frac{\int |x_1 - m_{n,x}|^3 \mathbb{Q}_{\Pi_{n,x}}(dx_1)}{n^{1/2}s_{n,x}^3} \\
\leq 4 \frac{\int |x_1|^3 \mathbb{Q}_{\Pi_{n,x}}(dx_1) - m_{n,x}^3}{n^{1/2}s_{n,x}^3} \\
\leq 8 \frac{\beta_{n,x}^{-1} E[|X_1^{n,x}|^3 e^{x_{n^{-1/2}} X_1^{n,x}}]}{n^{1/2}s_{n,x}^3} \\
\leq \frac{8 E[|X_1^{n,x}|^3 e^{x_{n^{-1/2}} X_1^{n,x}}]}{n^{1/2}s_{n,x}^3}. \tag{63}
\]

Now for the numerator in (63) we can use (46) to obtain

\[
\mathbb{E}[|X_1^{n,x}|^3 e^{x_{n^{-1/2}} X_1^{n,x}}] = \mathbb{E} \left[ \sum_{i=0}^{\infty} \frac{(xn^{-1/2})^i}{i!} |X_1^{n,x}|^{i+3} \right] \\
\leq \mathbb{E}[|X_1^{n,x}|^3] + xn^{-1/2} \mathbb{E}[(X_1^{n,x})^4 e^{x_{n^{-1/2}} X_1^{n,x}}] \\
\leq \mathbb{E}[|X_1^\lambda|] + xn^{-1/2} \mathbb{E}[(X_1^\lambda)^4 e^{x_{n^{-1/2}} X_1^\lambda}].
\]

Now for \( \lambda \in (3, 4) \) we can use (47) to deduce

\[
\mathbb{E}[|X_1^{n,x}|^3 e^{x_{n^{-1/2}} X_1^{n,x}}] \leq \mathbb{E}[|X_1^\lambda|] + c_{\lambda, 11} x^{-\lambda} n^{3-2\lambda/2 - \lambda/2} E[|X_1^\lambda|] \\
\leq c_{\lambda, 47} x^2 E[|X_1^\lambda|],
\]
where for the last step we used $n^{3/2-\lambda/2+\lambda} (\lambda-1) \leq 1$ (which follows from the assumption $r_\lambda \leq (\lambda - 3)/(2(\lambda - 1))$). For $\lambda \geq 4$ we can use (15) to obtain

$$E[|X_1^{n,x}|^3 e^{x n^{-1/2} X_1^{n,x}}] \leq E[|X_1'|^\lambda] + x n^{-1/2 + r_\lambda (\lambda - 1)} E[|X_1'|^\lambda] \leq c_{\lambda,49} x E[|X_1'|^\lambda],$$

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

$$E[|X_1^{n,x}|^3 e^{x n^{-1/2} X_1^{n,x}}] \leq c_{\lambda,49} x^2 E[|X_1'|^\lambda]. \quad (64)$$

Furthermore we will assume without loss of generality that $n$ is chosen sufficiently large such that $s^2_{n,x} \geq 1/2$. By (51) we have $|s^2_{n,x} - 1| \leq c_{\lambda,14} E[|X_1'|^\lambda] n^{-1/2} x^{12}$, i.e. it suffices to assume that $n \geq 4 c_{\lambda,14} E[|X_1'|^\lambda] x^{24}$. 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 $n_0 := [c_\lambda E[|X_1'|^\lambda]^{8}]$. 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 (52) – (54) lead to

$$e^{-x^2/2} \int_{(x-n^{1/2} s_{n,x} \to \infty)} e^{-x z} \Pi_{n,x}(dz) - \int_{(x-n^{1/2} s_{n,x} \to \infty)} e^{-x z} N_{0,s_{n,x}^2}(dz) \leq e^{-x^2/2} e^{-x(x-n^{1/2} s_{n,x})} c_{\lambda,49} x^2 E[|X_1'|^\lambda] n^{-1/2} s_{n,x}^{-3} \leq e^{-x^2/2} e^{-x(x-n^{1/2} s_{n,x})} c_{\lambda,49} 2 \sqrt{2} x^2 E[|X_1'|^\lambda] n^{-1/2} \leq c_{\lambda,50} e^{-x(x-n^{1/2} s_{n,x})} E[|X_1'|^\lambda] n^{-1/2} ((\lambda - 1) \log n)^7/2 n^{-1/2} x^2 e^{-x^2/2} E[|X_1'|^\lambda] \leq c_{\lambda,50} n^{-1/2} x^2 e^{-x^2/2} e^{-x(x-n^{1/2} s_{n,x})} E[|X_1'|^\lambda].$$

This proves (51).

Step 2.2.3. Finally we will show (52). With the transformation $a := x s_{n,x}^{-1} + x s_{n,x}$ we have

$$e^{-x^2/2} \int_{(x-n^{1/2} s_{n,x} \to \infty)} e^{-x z} N_{0,s_{n,x}^2}(dz) = e^{-x^2/2} (2\pi s_{n,x}^2)^{-1/2} \int_{x-n^{1/2} s_{n,x}}^\infty e^{-x 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} s_{n,x}}^\infty e^{-(z/s_{n,x} + x s_{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} s_{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} s_{n,x})/(s_{n,x} + x s_{n,x})}^\infty (2\pi)^{-1/2} e^{-a^2/2} \ da = e^{-x^2 (1-s_{n,x}^2) /2} \left\{ 1 - \Phi_{0,1}(s_{n,x}^{-1} n^{1/2} (x n^{-1/2} - m_{n,x}) + x s_{n,x}) \right\} = e^{-x^2 (1-s_{n,x}^2) /2} \Phi_{0,1}(\left\{ 1 - \Phi_{0,1}(s_{n,x}^{-1} n^{1/2} (x n^{-1/2} - m_{n,x}) + x s_{n,x}) \right\}.$
This leads to
\[
\left| e^{-x^2/2} \int_{(x-n^{-1/2}\ln x, \infty)} e^{-x^2/2}N_{0, s_{n,x}^2} (dz) - \Phi_{0,1}(-x) \right|
\]
\[
= |e^{-x^2(1-s_{n,x}^2)/2} \Phi_{0,1}(-\{s_{n,x}^{-1/2}(xn^{-1/2} - mn) + xs\}) - \Phi_{0,1}(-x)|
\]
\[
\leq e^{-x^2(1-s_{n,x}^2)/2} |\Phi_{0,1}(-\{s_{n,x}^{-1/2}(xn^{-1/2} - mn) + xs\}) - \Phi_{0,1}(-x)|
\]
\[
+ \left| e^{-x^2(1-s_{n,x}^2)/2} - 1 \right| |\Phi_{0,1}(-x)|
\]
\[
= S_{2,2,3,1}(\lambda, n, x) + S_{2,2,3,2}(\lambda, n, x) + S_{2,2,3,3}(\lambda, n, x).
\]

We will now show that the following inequalities are valid:

\[
S_{2,2,3,1}(\lambda, n, x) \leq c_{\lambda,51} n^{1/2} x^{6} e^{-x^2/4} e^{c_{\lambda,52} \mathbb{E}[|X_{1}^\lambda|^2]}, \tag{65}
\]
\[
S_{2,2,3,2}(\lambda, n, x) \leq c_{\lambda,53} n^{1/2} x^{14} e^{-x^2/4} e^{c_{\lambda,54} \mathbb{E}[|X_{1}^\lambda|^2]}, \tag{66}
\]
\[
S_{2,2,3,3}(\lambda, n, x) \leq c_{\lambda,55} n^{1/2} x^{13} e^{-x^2/4} e^{c_{\lambda,56} \mathbb{E}[|X_{1}^\lambda|^2]}, \tag{67}
\]

which will lead to (42) immediately. We will start with the proof of (66). First, by the Mean Value theorem, (51), and $x^2 \leq (\lambda - 1) \log n$ we have

\[
|e^{-x^2(1-s_{n,x}^2)/2} - 1| \leq e^{-x^2(1-s_{n,x}^2)/2} - 1
\]
\[
\leq (x^2 |1 - s_{n,x}^2|/2) e^{x^2(1-s_{n,x}^2)/2}
\]
\[
\leq c_{\lambda,57} \mathbb{E}[|X_{1}^\lambda|^2] n^{-1/2} x^{14} e^{c_{\lambda,14} \mathbb{E}[|X_{1}^\lambda|^2] n^{-1} x^{14}/2}
\]
\[
\leq c_{\lambda,58} n^{-1/2} x^{14} e^{c_{\lambda,59} \mathbb{E}[|X_{1}^\lambda|^2]}.
\tag{68}
\]

In particular, using $x^2 \leq (\lambda - 1) \log n$ again,

\[
e^{x^2(1-s_{n,x}^2)/2} \leq 1 + |e^{x^2(1-s_{n,x}^2)/2} - 1| \leq 1 + c_{\lambda,58} e^{c_{\lambda,50} \mathbb{E}[|X_{1}^\lambda|^2]} \leq c_{\lambda,60} e^{c_{\lambda,50} \mathbb{E}[|X_{1}^\lambda|^2]}.
\tag{69}
\]

First, (66) is a consequence of (68) and

\[
\Phi_{0,1}(-x) = \int_{-\infty}^{-xs_{n,x}} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy
\]
\[
\leq e^{-x^2s_{n,x}^2/4} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-y^2/4} dy
\]
\[
= e^{-(s_{n,x}^2 - 1)x^2/4} \sqrt{2}
\]
\[
\leq \sqrt{2} e^{s_{n,x}^2 - 1|x|^2/4} e^{-x^2/4}
\]
\[
\leq \sqrt{2} c_{\lambda,60} e^{c_{\lambda,50} \mathbb{E}[|X_{1}^\lambda|^2]} e^{-x^2/4},
\]

where we used (69) for the latter step. We next prove (65). We will assume without loss of generality that $n$ is chosen sufficiently large such that $n^{1/2} |m_{n,x} - xn^{-1/2}| \leq 1/4$
and \( s_{n,x}^2 \geq 1/2 \). By (50) and (51) we have \(|m_{n,x} - xn^{-1/2}| \leq c_{\lambda,13} \mathbb{E}[|X_1^1|]n^{-1/2}x^6 \) and \(|s_{n,x}^2 - 1| \leq c_{\lambda,14} \mathbb{E}[|X_1^1|]n^{-1/2}x^{12} \), i.e. it suffices to assume that \( n \geq (16c_{\lambda,13}^2) \mathbb{E}[|X_1^1|]^2x^{12} \) and \( n \geq (4c_{\lambda,14}^2) \mathbb{E}[|X_1^1|]^4x^{24} \). In view of \( x^2 \leq (\lambda - 1) \log n \), these assumptions holds if \( n \geq ((16c_{\lambda,13}^2) \vee (4c_{\lambda,14}^2)) \mathbb{E}[|X_1^1|]^4((\lambda - 1) \log n)^{12} \). That is, there is some constant \( \bar{c}_\lambda > 0 \) such that \( n \geq n_0 \) for \( n_0 := \left[ \bar{c}_\lambda \mathbb{E}[|X_1^1|]^8 \right] \) implies \( n^{1/2}|m_{n,x} - xn^{-1/2}| \leq 1/4 \) and \( s_{n,x}^2 \geq 1/2 \) for all \( 1 \leq |x| \leq \sqrt{(\lambda - 1) \log n} \). Recall from the discussion at the beginning of Step 2 that the assumption \( n \geq n_0 \) for this specific choice of \( n_0 \) indeed does not lead to any loss of generality. Now, using (50) and (69) we obtain

\[
S_{2,2,3,1}(\lambda, n, x) = e^{-x^2(1-s_{n,x}^2)/2} |\Phi_{0,1}(s_{n,x}^{-1}(xn^{-1/2} - m_{n,x}) + xs_{n,x}) - \Phi_{0,1}(xs_{n,x})| \\
= e^{-x^2(1-s_{n,x}^2)/2} \frac{1}{\sqrt{2\pi}} \int_{a_{n,x}}^{b_{n,x}} e^{-y^2/2} dy \\
\leq e^{-x^2(1-s_{n,x}^2)/2} \frac{1}{\sqrt{2\pi}} \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
= e^{-x^2(1-s_{n,x}^2)/2} \frac{1}{\sqrt{2\pi}} \max_{\xi \in [a_{n,x},b_{n,x}]} n^{-1/2} \mathbb{E}[|X_1^1|]n^{-1/2}|x|n^{-1/2} - m_{n,x}| \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
= e^{-x^2(1-s_{n,x}^2)/2} \frac{1}{\sqrt{2\pi}} \max_{\xi \in [a_{n,x},b_{n,x}]} n^{-1/2} \mathbb{E}[|X_1^1|]n^{-1/2} \mathbb{E}[|X_1^1|]n^{-1/2} \mathbb{E}[|X_1^1|]n^{-1/2} \mathbb{E}[|X_1^1|]n^{-1/2} \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2},
\]

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 \( x \geq 1 \) as well as \( n^{1/2}|m_{n,x} - xn^{-1/2}| \leq 1/4 \) and \( s_{n,x}^2 \geq 1/2 \), and therefore \( s_{n,x}^{-1}n^{1/2}(xn^{-1/2} - m_{n,x}) + xs_{n,x} \geq -\sqrt{2}(1/4) + 1/\sqrt{2} > 0 \). The implications are twofold. First, \( a_{n,x} \) is nonnegative so that \( \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2} = e^{-a_{n,x}^2/2} = e^{-s_{n,x}^2/2} \). Second, \( a_{n,x}^2 \geq (xs_{n,x})^2 - (s_{n,x}^{-1}n^{1/2}(xn^{-1/2} - m_{n,x}))^2 \). Hence,

\[
\max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2} \leq e^{-(xs_{n,x})^2/2} e^{(s_{n,x}^{-1}n^{1/2}|xn^{-1/2} - m_{n,x}|)^2/2} \leq e^{-x^2/4} e^{(\sqrt{2}/4)^2/2}.
\]

Together with (70) this implies (63). Finally, we will prove (67). By (51) we obtain

\[
S_{2,2,3,3}(\lambda, n, x) = |\Phi_{0,1}(xs_{n,x}) - \Phi_{0,1}(x)| \\
= \frac{1}{\sqrt{2\pi}} \int_{a_{n,x}}^{b_{n,x}} e^{-y^2/2} dy \\
\leq \frac{1}{\sqrt{2\pi}} \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
= \frac{1}{\sqrt{2\pi}} \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2} \\
\leq \frac{1}{\sqrt{2\pi}} \max_{\xi \in [a_{n,x},b_{n,x}]} e^{-\xi^2/2}.
\]

28
chosen sufficiently large such that $s$ which together with (71) implies (67).

$x > \max_{n, x}$ where we used $\tilde{a}$ and set (81) below. On the one hand, as in (37) we obtain

Finally, we will prove (33). Without loss of generality we restrict ourselves to $x > \max\{1; \sqrt{(\lambda - 1) \log n}\}$. Let $\lambda := 1/(2\lambda(\lambda - 1))$. As before consider the truncations

$$X_{i,n,x} := X_i 1_{[|X_i| \leq r_{\lambda} n^{1/2} x]}, \quad 1 \leq i \leq n, \quad n \in \mathbb{N},$$

and set $\tilde{Z}_{n,x} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{i,n,x}$. The specific choice of the constant $r_{\lambda}$ will be needed in (81) below. On the one hand, as in (87) we obtain

$$1 - F_n(x) = \mathbb{P}[Z_n' > x] \leq \mathbb{P}[\tilde{Z}_{n,x} > x] + n \mathbb{P}[|X_i'| > r_{\lambda} n^{1/2} x]. \quad (72)$$

On the other hand, we can use the transformation $a := \sqrt{z^2 - x^2}$ to obtain

$$1 - \Phi_{0,1}(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2 - x^2}{2}} dz = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{a}{\sqrt{a^2 + x^2}} da \leq \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{x^2}{2} da = e^{-\frac{x^2}{2}}/2 = e^{-\frac{x^2}{2(\lambda - 1)}} e^{-\frac{x^2(\lambda - 2)}{2(\lambda - 1)^2}}/2 \leq (c_{\lambda,1} x^{-\lambda}) e^{-\frac{(\lambda - 1) \log n(\lambda - 2)}{2(\lambda - 1)^2}}/2 = c_{\lambda,2} x^{-\lambda} n^{-(\lambda - 2)}, \quad (73)$$

where we used $x^2 \geq (\lambda - 1) \log n$. From (72)–(73) we can deduce

$$|F_n(x) - \Phi_{0,1}(x)| \leq \mathbb{P}[\tilde{Z}_{n,x} > x] + n \mathbb{P}[|X_i'| > r_{\lambda} n^{1/2} x] + c_{\lambda,2} n^{-(\lambda - 2)} x^{-\lambda}. \quad (74)$$

By Markov’s inequality we have

$$n \mathbb{P}[|X_i'| > r_{\lambda} n^{1/2} x] \leq \frac{\mathbb{E}[|X_i'| \lambda]}{(r_{\lambda} n^{1/2} x)^{\lambda}} \leq c_{\lambda,3} \mathbb{E}[|X_i'| \lambda] n^{-(\lambda - 2)} x^{-\lambda}. \quad (75)$$

29
Finally, we will show that
\[ \mathbb{P}[\tilde{Z}_n^{n,x} > x] \leq c_{\lambda,4} e^{c_{\lambda,5} E[|X_1'|]} n^{-(\lambda/2 - 1)} x^{-\lambda}. \]  
(76)

Then, (74)–(76) and the assumption \( \lambda > 3 \) imply (33).

To prove (76), 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
\[
\mathbb{P}[\tilde{Z}_n^{n,x} > x] = \mathbb{P}\left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_1^{n,x} > x \right] = \mathbb{P}\left[ e^{k_n(x) \sum_{i=1}^{n} X_1^{n,x}} > e^{\sqrt{n}k_n(x)} \right] \leq \frac{E[e^{k_n(x) \sum_{i=1}^{n} X_1^{n,x}}]}{e^{(\lambda - 2) \log n + 2\lambda(\lambda - 1) \log x}} = c_{\lambda,4} e^{c_{\lambda,5} E[|X_1'|]} n^{(\lambda - 2)/2} x^{2\lambda(\lambda - 2)}. \]

So (76) would follow if we can show that
\[
E[e^{k_n(x) X_1^{n,x}}] \leq c_{\lambda,4} e^{c_{\lambda,5} E[|X_1'|]} n^{(\lambda - 2)/2} x^{2\lambda(\lambda - 2)}. \]  
(77)

In the rest of the proof we will show (77). We clearly have
\[
|E[e^{k_n(x) X_1^{n,x}}]| = \left| \mathbb{E}\left[ \sum_{i=0}^{\infty} \frac{(k_n(x) X_1^{n,x})^i}{i!} \right] \right| \leq 1 + k_n(x) |\mathbb{E}[X_1^{n,x}]| + k_n^2(x) \mathbb{E}[(X_1^{n,x})^2] + \mathbb{E}\left[ \left| \sum_{i=3}^{\infty} \frac{(k_n(x) X_1^{n,x})^i}{i!} \right| \right] \leq 1 + k_n(x) |\mathbb{E}[X_1^{n,x}]| + k_n^2(x) \mathbb{E}[(X_1^{n,x})^2] + \mathbb{E}\left[ (k_n(x) X_1^{n,x})^3 \sum_{i=0}^{\infty} \frac{(k_n(x) X_1^{n,x})^i}{i!} \right] \leq 1 + k_n(x) |\mathbb{E}[X_1^{n,x}]| + k_n^2(x) \mathbb{E}[(X_1^{n,x})^2] + \mathbb{E}[k_n(x) X_1^{n,x}] e^{k_n(x) X_1^{n,x}}. \]

(78)

We have \( |E[X_1^{n,x}]| \leq E[|X_1'| 1_{\{|X_1'| > r_\lambda n^{1/2}x\}}] \), because \( X_1^{n,x} = X_1' - X_1' 1_{\{|X_1'| > r_\lambda n^{1/2}x\}} \) and \( \mathbb{E}[X_1'] = 0 \). Thus, the second summand on the right-hand side in (78) can be bounded above by
\[
k_n(x) |\mathbb{E}[X_1^{n,x}]| \leq k_n(x) \mathbb{E}[|X_1'| 1_{\{|X_1'| > r_\lambda n^{1/2}x\}}] \leq k_n(x) \mathbb{E}\left[ \frac{|X_1'|}{(r_\lambda n^{1/2}x)^{\lambda - 1}} 1_{\{|X_1'| > r_\lambda n^{1/2}x\}} \right] \leq \frac{1}{x} \frac{1}{\sqrt{n}} ((\lambda - 2) \log n + 2\lambda(\lambda - 1) \log x) (r_\lambda n^{1/2}x)^{1 - \lambda} \mathbb{E}[|X_1'|^\lambda] \leq c_{\lambda,6} n^{-1} \mathbb{E}[|X_1'|^\lambda]. \]

(79)
Since \( \mathbb{E}[(X_1^{n,x})^2] \leq \mathbb{E}[(X_1^x)^2] = 1 \), the third summand on the right-hand side in (78) is bounded above by
\[
\frac{k_n(x)^2}{2} \mathbb{E}[(X_1^{n,x})^2] \leq \frac{k_n(x)^2}{2}.
\]
Using the same arguments as in (46) we observe that \( \mathbb{E}[(X_1^{n,x})^3] \leq \mathbb{E}[(X_1^x)^3] \). Thus the fourth summand on the right-hand side in (78) we have
\[
\mathbb{E}[k_n(x)X_1^{n,x} e^{k_n(x)X_1^{n,x}}]
\leq k_n(x)^3 e^{k_n(x)x} \mathbb{E}[(X_1^{n,x})^3]
\leq k_n(x)^3(n^2x(n-1)x) \mathbb{E}[(X_1^x)^3]
= ((\lambda - 2) \log n + 2\lambda(\lambda - 1) \log x)^3 n^2 \frac{x^2}{2} \mathbb{E}[(X_1^x)^3]
= c_{\lambda,7} n^{-1} \mathbb{E}[(X_1^x)^3],
\]
where we used the definition of \( r_\lambda \) and the fact that \((\lambda - 2)/(2\lambda(\lambda - 1)) - 3/2 \leq -1 \) for \( \lambda > 3 \).

Now, (78)–(84) yield
\[
|\mathbb{E}[e^{k_n(x)X_1^{n,x}}]| \leq 1 + c_{\lambda,6} n^{-1} \mathbb{E}[(X_1^x)^3] + \frac{k_n^2(x)}{2} + c_{\lambda,7} n^{-1} \mathbb{E}[(X_1^x)^3]
\leq e^{k_n^2(x)/2 + c_{\lambda,8} \mathbb{E}[(X_1^x)^3]}. \tag{82}
\]
Thus, for every \( n \in \mathbb{N} \) we obtain
\[
\mathbb{E}[e^{k_n(x)X_1^{n,x}}] \leq e^{n k_n^2(x)/2 + c_{\lambda,8} \mathbb{E}[(X_1^x)^3]}. \tag{83}
\]
Since \( x \geq ((\lambda - 1) \log n)^{1/2} \), we have
\[
\frac{n}{2} k_n^2(x)
= \frac{n}{2} \frac{1}{x^2} \frac{1}{n} ((\lambda - 2)^2 (\log n)^2 + 2\lambda(\lambda - 1)(\lambda - 2) \log x \log n + 4\lambda^2(\lambda - 1)^2(\log x)^2)
\leq \frac{\lambda - 2}{2} \log n \frac{1}{\lambda - 1} (\lambda - 1) \log n \frac{1}{x^2} + 2\lambda(\lambda - 2) \log x + 2\lambda^2(\lambda - 1)^2(\log x)^2 \frac{1}{x^2}
\leq \frac{\lambda - 2}{2} \log n + 2\lambda(\lambda - 2) \log x + c_{\lambda,9}, \tag{84}
\]
with \( c_{\lambda,9} = 2\lambda^2(\lambda - 1)^2 \). Now (83)–(84) imply
\[
\mathbb{E}[e^{k_n(x)X_1^{n,x}}] \leq e^{\frac{1}{2}(\lambda - 2) \log n + 2\lambda(\lambda - 2) \log x + c_{\lambda,9} + c_{\lambda,8} \mathbb{E}[(X_1^x)^3]}
\leq c_{\lambda,10} n^{(\lambda - 2)/2} e^{2\lambda(\lambda - 2) + c_{\lambda,8} \mathbb{E}[(X_1^x)^3]}. \tag{85}
\]
This shows (77) with \( c_{\lambda,4} := c_{\lambda,10} \) and \( c_{\lambda,5} := c_{\lambda,8} \), and the proof is complete. \( \square \)
A. A remark on the recursive scheme (8)–(9)

The empirical probability measure \( \hat{\mu}_u \) defined in (5) has the representation

\[
\hat{\mu}_u[\cdot] = \hat{p}_u \hat{\nu}_u[\cdot] + (1 - \hat{p}_u) \delta_0[\cdot],
\]

where \( \hat{p}_u := \hat{\mu}_u[(0, \infty)] \) is the mass of \( \hat{\mu}_u \) on \((0, \infty)\), and \( \hat{\nu}_u[\cdot] := \hat{\mu}_u[\cdot \cap (0, \infty)] / \hat{\mu}_u[(0, \infty)] \) is the probability measure \( \hat{\mu}_u \) conditioned on \((0, \infty)\). It is easily seen that the \( n \)-fold convolution \( \hat{\mu}_u^* \) coincides with the random convolution

\[
\hat{\nu}_u^* B_{n, \hat{p}_u}[\cdot] := \sum_{k=0}^n \hat{\nu}_u^k[\cdot] B_{n, \hat{p}_u}[\{k\}]
\]

of \( \hat{\nu}_u \) w.r.t. the binomial distribution \( B_{n, \hat{p}_u} \) with parameters \( n \) and \( \hat{p}_u \), i.e.

\[
\hat{\mu}_u^* = \hat{\nu}_u^* B_{n, \hat{p}_u}.
\]  

When \( \hat{p}_u < 1 \) and \( \hat{\nu}_u \) has support in \( hN := \{h, 2h, \ldots\} \) for some \( h > 0 \), the random convolution \( \hat{\nu}_u^* B_{n, \hat{p}_u} \) can be computed with the help of the Panjer recursion [10]:

\[
\begin{align*}
\hat{\nu}_u^* B_{n, \hat{p}_u}[\{0\}] &= B_{n, \hat{p}_u}[\{0\}] \\
\hat{\nu}_u^* B_{n, \hat{p}_u}[\{jh\}] &= \frac{\hat{p}_u^j}{1 - \hat{p}_u} \sum_{\ell=1}^j [(n+1)\ell - j] \hat{\nu}_u[\{\ell h\}] \hat{\nu}_u^* B_{n, \hat{p}_u}[\{(j - \ell) h\}] \quad \text{for } j \in \mathbb{N}.
\end{align*}
\]  

Since \( 1 - \hat{p}_u = \hat{\mu}_u[\{0\}] \) and \( \hat{p}_u \hat{\nu}_u[\{\ell h\}] = \hat{\mu}_u[\{\ell h\}] \) for \( \ell \in \mathbb{N} = \{1, 2, \ldots\} \), the recursive scheme (8)–(9) follows from (86)–(88).

References

[1] Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. Insurance: Mathematics and Economics, 54, 41–48.

[2] Bikelis, A. (1965). Estimates of the remainder in the central limit theorem. Lithuanian Mathematical Journal, 6, 323–346.

[3] Föllmer, H. and Schied, A. (2004). Stochastic finance. An introduction in discrete time. de Gruyter, Berlin.

[4] Katz, M.L. (1963). Note on the Berry–Esséen theorem. Annals of Mathematical Statistics, 34, 1107–1108.

[5] 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.
[6] Krätzschmer, V., Schied, A. and Zähle, H. (2015). Quasi-Hadamard differentiability of general risk functionals and its application. Forthcoming in *Statistics and Risk Modeling*, online first.

[7] Michel, R. (1976). Nonuniform central limit bounds with applications to probabilities of deviations. *Annals of Probability*, 4, 102–106.

[8] Nagaev, S.V. (1965). Some limit theorems for large deviations. *Theory of Probability and its Applications*, 10, 214–235.

[9] 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.

[10] Panjer, H.H. (1981). Recursive evaluation of a family of compound distributions. *ASTIN Bulletin*, 12, 22–26.

[11] Pitts, S.M. (1994). Nonparametric estimation of compound distributions with applications in insurance. *Annals of the Institute of Mathematical Statistics*, 46, 537–555.

[12] Zähle, H. (2011). Rates of almost sure convergence of plug-in estimates for distortion risk measures. *Metrika*, 74, 267–285.