

**Rates of almost sure convergence of
plug-in estimates for distortion risk
measures**

Henryk Zähle

Preprint 2010-01

Januar 2010

Rates of almost sure convergence of plug-in estimates for distortion risk measures

Henryk Zähle*

Abstract

In this article, we consider plug-in estimates for distortion risk measures as for instance the Value-at-Risk, the Expected Shortfall or the Wang transform. We allow for fairly general estimates of the underlying unknown distribution function (beyond the classical empirical distribution function) to be plugged in the risk measure. We establish strong consistency of the estimates, we investigate the rate of almost sure convergence, and we study the small sample behavior by means of simulations.

Key words and phrases. Risk measure, plug-in estimation, empirical distribution function, smoothing, censoring, Glivenko-Cantelli theorem for weighted errors

MSC 2000 subject classification. 62G05, 91B30

*Postal Address: Technische Universität Dortmund, Fakultät für Mathematik, Vogelpothsweg 87, D-44227 Dortmund, Germany; Email: henryk.zaehle@math.tu-dortmund.de

1 Introduction

Risk measures are the basic tools for the determination of insurance premiums. In actuarial practice several risk measures are used, and there are different concepts of how to specify a “good” risk measure. In the last decade the concept of monetary risk measures and especially the concept of coherent risk measures received common credit (cf. [4, 5, 9, 11, 12] and others). This article is concerned with the nonparametric estimation of a certain class of monetary risk measures. More precisely, we are going to investigate nonparametric estimates for so called distortion risk measures.

Suppose X is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which should be seen as the outcome of a financial position. For instance, X may model the claim of an insurance contract from the point of view of an insurer. To some extent the negative mean

$$\rho_{\mathbb{I}}(X) := - \int_{\mathbb{R}} x dF_X(x) \quad (1)$$

provides the simplest distribution invariant monetary risk measure, where F_X denotes the distribution function (df) of X . A \mathbb{P} -integrable position X will be accepted by $\rho_{\mathbb{I}}$ if and only if $\mathbb{E}[X] \geq 0$. For many purposes, in particular for the determination of premiums of insurance policies, this rule is “too indulgent” insofar as it accepts “too many” positions. For instance, it is well known that the “net risk premium” $\mathbb{E}[-X]$ entails technical ruin of an insurer in finite time with probability one. One can overcome this problem by manipulating F_X by a risk-averse distortion function (cf. [24]). A càdlàg function $g : [0, 1] \rightarrow [0, 1]$ is called distortion function if it is nondecreasing and satisfies $g(0) = 0$ and $g(1) = 1$. A distortion function is called risk-averse if it is concave. The basic idea is to replace in (1) the df F_X by the distorted df $g(F_X(\cdot))$. This leads to the risk measure

$$\rho_g(X) := - \int_{\mathbb{R}} x dg(F_X(x)) \quad (2)$$

on the set of all random variables X for which the integral in (2) exists. It is called distortion risk measure with distortion function g and has been studied several times in the literature, cf. [1, 13, 14, 15, 16, 24, 25, 26] and references cited therein. We emphasize that most of the popular risk measures in practice can be represented as in (2), cf. the Appendix B. It is known that ρ_g provides a positively homogeneous monetary risk measure, which is subadditive (hence coherent) if and only if g is concave, cf. [26]. Moreover it was shown in [13] that $\rho_g \geq \rho_{\mathbb{I}}$ if and only if $g(x) \geq x$ for all $x \in (0, 1)$. That is, ρ_g is indeed more conservative than $\rho_{\mathbb{I}}$ if g is a risk-averse distortion function.

The monetary risk measure ρ_g is obviously distribution invariant, i.e. $\rho_g(X) = \rho_g(Y)$ if X and Y coincide in law. Therefore we may and do write without ambiguity $\rho_g(F_X)$ in place of $\rho_g(X)$. In particular we regard ρ_g as a mapping from \mathbb{F}_g to \mathbb{R} ,

$$F \mapsto \rho_g(F) := - \int_{\mathbb{R}} x dg(F(x))$$

$$= \int_{(-\infty, 0)} g(F(t))dt - \int_{[0, \infty)} (1 - g(F(t)))dt, \quad (3)$$

where \mathbb{F}_g is the class of all df F satisfying $\int_{\mathbb{R}} |x| dg(F(x)) < \infty$.

In actuarial practice the df F of a certain claim is typically unknown and has to be estimated. If one can observe (possibly censored) i.i.d. replications $\tilde{X}_1, \dots, \tilde{X}_n$ corresponding to F then one can at first determine a suitable empirical estimate F_n of F based on $\tilde{X}_1, \dots, \tilde{X}_n$. The estimate F_n can then be plugged in ρ_g in order to obtain an estimate $\rho_g(F_n)$ of $\rho_g(F)$. In the case where F_n is the classical empirical df based on uncensored data and where g is continuous, strong consistency and the asymptotic distribution of $\rho_g(F_n)$ were established in [15, 16]. The results there rely on the facts that in this case $\rho_g(F_n)$ is an L -statistic (cf. the representation (17) in the Appendix B) and that L -statistics are well studied (see, e.g., [20, 22, 23]). If however F_n differs from the classical empirical df, e.g., if F_n is a smooth empirical df (see, e.g., [28]), or if F_n is based on censored data, then the arguments of [15, 16] do not apply anymore.

In this article, we will extend the consistency results in [15, 16] to more general estimates F_n of F , and to *discontinuous* distortion functions g . Moreover we will establish satisfactory statements on the rates of almost sure convergence under fairly transparent conditions on g and F . A partial generalization of the consistency results of [15, 16] in a different direction has been recently considered in [27]. There, strong consistency of the empirical estimators of distribution-invariant monetary risk measures having a certain robust representation (cf. (18) in the Appendix B) is established. Under mild assumptions, each distribution invariant coherent risk measure – in particular each distortion risk measure with concave (and therefore continuous) distortion function – admits such a representation. An extension of the results on the asymptotic distribution of $\rho_g(F_n)$ in [15, 16] to more general estimates F_n of F can be found in [6]. The analysis there requires a sort of functional delta method based on a generalized notion of Hadamard differentiability, which is also introduced in [6].

The rest of this article is organized as follows. In Section 2, we show that $\rho_g(F)$ depends continuously on F with respect to a certain distance between distribution functions on \mathbb{R} . In Section 3, we demonstrate that both a smoothed empirical df and an estimate of F based on multiplicatively censored data converge almost surely to the true underlying df F with respect to the mentioned distance. Along with the results of Section 2 this implies strong consistency of $\rho_g(F_n)$ for $\rho_g(F)$ as well as satisfactory statements on the rates of almost sure convergence of $\rho_g(F_n)$ to $\rho_g(F)$. In Section 4, we study the small sample behavior of the plug-in estimates by means of simulations. Section 5 provides the proofs of the main results. The Appendix A provides a generalized Glivenko-Cantelli theorem which will be the crux of our analysis, and the Appendix B calls to mind some popular examples for distortion risk measures.

2 Regularity of ρ_g

In this section, we give two regularity results concerning ρ_g (Theorems 2.1 and 2.2) which immediately provide tools for establishing strong consistency and the rates of almost sure convergence of plug-in estimates for $\rho_g(F)$ (cf. Remarks 2.4 and 2.5 below). For any $\lambda > 0$, set $\phi_\lambda(x) := (1 + |x|)^\lambda$ for all $x \in \mathbb{R}$. Notice that the reciprocal function ϕ_λ^{-1} is Lebesgue integrable on \mathbb{R} if $\lambda > 1$. Let \mathbb{D} be the space of all càdlàg functions on \mathbb{R} , and \mathbb{D}_λ be the subspace of all functions $\psi \in \mathbb{D}$ with $\|\psi\|_\lambda := \|\psi\phi_\lambda\|_\infty < \infty$. Notice that $\|\cdot\|_\lambda$ provides a norm on \mathbb{D}_λ . For any function f on (a subset of) \mathbb{R} taking values in $[0, 1]$, we set $\bar{f} := 1 - f$. For a df F , we define the left-continuous and the right-continuous inverse by $F^\leftarrow(x) := \inf\{t \in \mathbb{R} : F(t) \geq x\}$ and $F^\rightarrow(x) := \inf\{t \in \mathbb{R} : F(t) > x\}$, $x \in (0, 1)$, respectively. Also notice that g has at most countably many discontinuities since it is a bounded càdlàg function, and that \mathbb{F}_g was introduced subsequent to (3).

Theorem 2.1 *Suppose that*

- (i) *there are constants $\beta, C > 0$ such that $g(x) \leq Cx^\beta$ and $\bar{g}(1-x) \leq Cx^\beta$, $x \in [0, 1]$,*
- (ii) *$F, F_1, F_2, \dots \in \mathbb{F}_g$,*
- (iii) *$F^\leftarrow(d) = F^\rightarrow(d)$ for every discontinuity d of g ,*
- (iv) *$F_n(t) \rightarrow F(t)$ for dt -almost every $t \in \mathbb{R}$,*
- (v) *$\sup_{n \in \mathbb{N}} (\|F_n \mathbb{1}_{(-\infty, 0)}\|_\lambda + \|\bar{F}_n \mathbb{1}_{[0, \infty)}\|_\lambda) < \infty$ for some $\lambda > 1/\beta$.*

Then $\rho_g(F_n) \rightarrow \rho_g(F)$.

The proof of Theorem 2.1 is relegated to Section 5.1. The next result shows that if g and F are sufficiently regular then the distance of $\rho_g(F_n)$ and $\rho_g(F)$ can be bounded above by a suitable distance of F_n and F .

Theorem 2.2 *Suppose that*

- (i) *there are some constants $\beta, C > 0$, $k \in \mathbb{N}_0$, and $0 = d_0 < d_1 < \dots < d_{k+1} = 1$, such that g is Hölder- β -continuous with Hölder constant C on each of the intervals (d_i, d_{i+1}) , $i = 0, \dots, k$,*
- (ii) *$F, F_1, F_2, \dots \in \mathbb{F}_g$,*
- (iii) *F is differentiable at $F^\leftarrow(d_i)$, and $F'(F^\leftarrow(d_i)) > 0$, for $i = 1, \dots, k$,*
- (iv) *$\|F_n - F\|_\infty \rightarrow 0$.*

Then for every $\lambda \geq 1$ satisfying $\lambda\beta > 1$ there is a finite constant $K = K_{k,F,C,\lambda\beta} > 0$, and some $n_0 \in \mathbb{N}$ (such that $\|F_n - F\|_\infty$ is sufficiently small for all $n \geq n_0$), such that

$$|\rho_g(F_n) - \rho_g(F)| \leq K \|F_n - F\|_\lambda^{\beta \wedge 1}, \quad n \geq n_0. \quad (4)$$

If g is Hölder- β -continuous on all of $[0, 1]$, i.e. if $k = 0$, then condition (iv) can be skipped, the assumption $\lambda \geq 1$ can be replaced by $\lambda > 0$, in (4) the exponent $\beta \wedge 1$ can be replaced by β , and inequality (4) holds for all $n \in \mathbb{N}$.

The proof of Theorem 2.2 is relegated to Section 5.2. Notice that on the right-hand side of (4) the expression $\|F_n - F\|_\lambda$ might be infinite, and that $\|F_n - F\|_\lambda \rightarrow 0$ implies condition (iv) whenever $\lambda \geq 1$. Also notice that condition (iii) implies in particular $F^{\leftarrow}(d_i) = F^{\rightarrow}(d_i)$ for all $i = 1, \dots, k$. Therefore we may and do also write $F^{-1}(d_i)$ in place of $F^{\leftarrow}(d_i)$.

Example 2.3 Notice that the condition on g in Theorem 2.1 is fulfilled for, e.g., $g(x) = \mathbb{1}_{[\alpha, 1]}(x)$, $g(x) = (x/\alpha) \wedge 1$ and $g(x) = \Phi(\Phi^{-1}(x) - \theta)$, which correspond to the Value-at-Risk at level $\alpha \in (0, 1)$, the Average Value-at-Risk at level $\alpha \in (0, 1)$ and the Wang transform WT_θ with parameter $\theta \in \mathbb{R}$, respectively. See the Appendix B for the definitions of these risk measures. These examples also satisfy the condition on g in Theorem 2.2 with any $\beta > 0$, $\beta = 1$ and $\beta \in (0, 1)$, respectively. \diamond

Remark 2.4 As an immediate consequence of Theorem 2.2 we obtain that if $\|F_n - F\|_\lambda$ converges to 0 at rate r , with $\lambda \geq 1$ (resp. $\lambda > 0$) satisfying $\lambda\beta > 1$, then $|\rho_g(F_n) - \rho_g(F)|$ converges to 0 at least at rate $(\beta \wedge 1)r$ (resp. βr). \diamond

Remark 2.5 Let \mathcal{D} be the σ -algebra on \mathbb{D} generated by the usual coordinate projections $\pi_t : \mathbb{D} \rightarrow \mathbb{R}$. Further, let $F \in \mathbb{F}_g$, $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and (F_n) be a sequence of \mathbb{F}_g -valued estimates $F_n : (\Omega, \mathcal{F}) \rightarrow (\mathbb{D}, \mathcal{D})$ for F . If the conditions of Theorem 2.1 are fulfilled \mathbb{P} -almost surely, then we obtain strong consistency of $\rho_g(F_n)$ for $\rho_g(F)$, i.e. $\rho_g(F_n) \rightarrow \rho_g(F)$ \mathbb{P} -almost surely. If the conditions of Theorem 2.2 are fulfilled \mathbb{P} -almost surely, then we obtain that (4) holds \mathbb{P} -almost surely. \diamond

3 Plug-in estimation of $\rho_g(F)$

In this section, we are going to illustrate the benefit of Theorems 2.1 and 2.2 for the plug-in estimation of $\rho_g(F)$ by means of two examples. First, we let F_n be a smoothed version of the empirical df corresponding to F (setting $\varepsilon_n = 0$, we get back the empirical df itself). Second, we let F_n be an estimate of F based on multiplicatively censored data.

3.1 Uncensored data, smoothed empirical df

Suppose X_1, X_2, \dots are i.i.d. random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with df F . We denote by $\hat{F}_n := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[X_i, \infty)}$ the corresponding empirical df at stage n . For some purposes it might be gainful to consider a smoothed version $P_\varepsilon \hat{F}_n$ of \hat{F}_n , where ε is a certain smoothing parameter. Here, we consider a smoothing by the heat kernel (also called Gaussian kernel). We set $p_\varepsilon(y) := (2\pi\varepsilon)^{-1/2} \exp(-y^2/(2\varepsilon))$ ($\varepsilon > 0, y \in \mathbb{R}$), and we denote by $(P_\varepsilon)_{\varepsilon \geq 0}$ the corresponding (heat) semigroup, i.e., $P_\varepsilon \psi(\cdot) := \int_{\mathbb{R}} \psi(y) p_\varepsilon(\cdot - y) dy$ for $\varepsilon > 0$, and $P_0 := \mathbb{I}$. We focus on the following estimate of F

$$F_n(t) := P_{\varepsilon_n} \hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n P_{\varepsilon_n} \mathbb{1}_{[X_i, \infty)}(t), \quad t \in \mathbb{R}. \quad (5)$$

For $\varepsilon_n = 0$ we obviously have $F_n = \hat{F}_n$.

A smoothing of \hat{F}_n may be also beneficial (in terms of the mean square error) for the plug-in estimation of risk measures. For particular distortion risk measures, namely the Value-at-Risk (quantile) and the Average Value-at-Risk, the benefit of a smoothing is studied in, for instance, [8, 19]. Here we intend to establish strong consistency of $\rho_g(F_n)$ for $\rho_g(F)$, provided $\varepsilon_n \downarrow 0$. To this end we will assume that for some suitable $\gamma > 0$,

$$\limsup_{t \rightarrow -\infty} F(t)|t|^\gamma < \infty \quad \text{and} \quad \limsup_{t \rightarrow \infty} \bar{F}(t)t^\gamma < \infty. \quad (6)$$

The key for our consistency results (Theorems 3.4 and 3.5) will be Lemma 3.2 which is a sort of Glivenko-Cantelli theorem. Its proof relies on the following lemma, where we use the notation $\phi_{-\lambda}(t) := (1 + |t|)^{-\lambda}$.

Lemma 3.1 *For every $\lambda \geq 0$ there are some constants $C_1(\lambda), C_2(\lambda) > 0$ such that for every $t, x \in \mathbb{R}$ and $\varepsilon \in (0, 1]$*

$$\int_{\mathbb{R}} |x - y|^\lambda p_\varepsilon(x - y) dy \leq C_1(\lambda) \varepsilon^{\lambda/2}, \quad (7)$$

$$\int_{\mathbb{R}} \phi_{-\lambda}(y) p_\varepsilon(t - y) dy \leq C_2(\lambda) \phi_{-\lambda}(t). \quad (8)$$

Proof Inequality (7) is a standard estimate for the heat kernel, see, e.g., [7]. Inequality (8) follows from

$$\int_{\mathbb{R}} \phi_{-\lambda}(y) p_\varepsilon(t - y) dy = \int_{\mathbb{R}} \frac{\phi_\lambda(t)}{\phi_\lambda(y)} e^{-\frac{(t-y)^2}{4\varepsilon}} \sqrt{2} p_{2\varepsilon}(t - y) dy \phi_{-\lambda}(t)$$

and the fact that the mapping $(t, y) \mapsto \frac{\phi_\lambda(t)}{\phi_\lambda(y)} e^{-\frac{(t-y)^2}{4\varepsilon}}$ is bounded above. \square

Lemma 3.2 *Let $\lambda \geq 0$. Suppose that F is Lipschitz continuous and satisfies (6) for some $\gamma > \lambda$. Then there is some constant $K_{\lambda,\gamma} > 0$ such that for every $n \in \mathbb{N}$ and $\varepsilon \in (0, 1]$*

$$\|P_\varepsilon \hat{F}_n - F\|_\lambda \leq K_{\lambda,\gamma} \left(\|\hat{F}_n - F\|_\lambda + \varepsilon^{(\gamma-\lambda)/(2\gamma)} \right). \quad (9)$$

Proof We clearly have

$$|P_\varepsilon \hat{F}_n(t) - F(t)| \phi_\lambda(t) \leq |P_\varepsilon(\hat{F}_n - F)(t)| \phi_\lambda(t) + |P_\varepsilon F(t) - F(t)| \phi_\lambda(t) =: S_1(t) + S_2(t).$$

On the one hand we obtain by the use of (8)

$$\begin{aligned} S_1(t) &\leq \int_{\mathbb{R}} \|\hat{F}_n - F\|_\lambda \phi_{-\lambda}(y) p_\varepsilon(t-y) dy \phi_\lambda(t) \\ &\leq \|\hat{F}_n - F\|_\lambda \int_{\mathbb{R}} \phi_{-\lambda}(y) p_\varepsilon(t-y) dy \phi_\lambda(t) \\ &= \|\hat{F}_n - F\|_\lambda C_2(\lambda). \end{aligned}$$

On the other hand, for $t \geq 0$ we obtain with the help of the Lipschitz continuity of F (we denote the Lipschitz constant by L), condition (6), and the inequalities (7) and (8),

$$\begin{aligned} S_2(t) &\leq \int_{\mathbb{R}} |F(y) - F(t)| p_\varepsilon(t-y) dy \phi_\lambda(t) \\ &\leq \left(\int_{(0,\infty)} (\bar{F}(y) + \bar{F}(t)) p_\varepsilon(t-y) dy \right)^\delta \left(\int_{(0,\infty)} L|y-t| p_\varepsilon(t-y) dy \right)^{1-\delta} \phi_\lambda(t) \\ &\quad + \int_{[-1,0]} |y-t| p_\varepsilon(t-y) dy \phi_\lambda(1) \\ &\quad + \int_{(-\infty,-1)} \sqrt{3} p_{3\varepsilon}(t-y) dy \left(e^{-t^2/6} \phi_\lambda(t) \right) e^{-1/(6\varepsilon)} \\ &\leq \left(\int_{\mathbb{R}} (C_F \phi_{-\gamma}(y) + C_F \phi_{-\gamma}(t)) p_\varepsilon(t-y) dy \right)^\delta C_1(1)^{1-\delta} \varepsilon^{(1-\delta)/2} \phi_\lambda(t) \\ &\quad + 2^\lambda C_1(1) \varepsilon^{1/2} + \sqrt{3} C_\lambda e^{-1/(6\varepsilon)} \\ &\leq C_F^\delta C_1(\lambda)^{1-\delta} \left(\int_{\mathbb{R}} \phi_{-\gamma}(y) p_\varepsilon(t-y) dy + \phi_{-\gamma}(t) \right)^\delta \varepsilon^{(1-\delta)/2} \phi_\lambda(t) \\ &\quad + \tilde{C}_\lambda \varepsilon^{(\gamma-\lambda)/(2\gamma)} + \tilde{C}_\lambda \varepsilon^{(\gamma-\lambda)/(2\gamma)} \\ &\leq C_F^\delta C_1(\lambda)^{1-\delta} \left(C_2(\gamma) \phi_{-\gamma}(t) + \phi_{-\gamma}(t) \right)^\delta \varepsilon^{(1-\delta)/2} \phi_\lambda(t) + 2\tilde{C}_\lambda \varepsilon^{(\gamma-\lambda)/(2\gamma)} \\ &\leq C_F^\delta C_1(\lambda)^{1-\delta} (C_2(\gamma) + 1)^\delta \phi_{-\gamma\delta}(t) \phi_\lambda(t) \varepsilon^{(1-\delta)/2} + 2\tilde{C}_\lambda \varepsilon^{(\gamma-\lambda)/(2\gamma)} \\ &\leq \left(C_F^{\lambda/\gamma} C_1(\lambda)^{(\gamma-\delta)/\gamma} (C_2(\gamma) + 1)^{\lambda/\gamma} + 2\tilde{C}_\lambda \right) \varepsilon^{(\gamma-\lambda)/(2\gamma)} \\ &=: K_{\lambda,\gamma} \varepsilon^{(\gamma-\lambda)/(2\gamma)} \end{aligned}$$

for $\delta := \lambda/\gamma$ and some suitable constants $C_F, C_\lambda, \tilde{C}_\lambda > 0$. In the same way we also obtain $S_2(t) \leq K_{\lambda,\gamma} \varepsilon^{(\gamma-\lambda)/(2\gamma)}$ for $t < 0$. This proves (9). \square

Lemma 3.3 *Suppose that g satisfies assumption (i) of Theorem 2.1 or assumption (i) of Theorem 2.2. Then we have $P_\varepsilon \hat{F}_n \in \mathbb{F}_g$ \mathbb{P} -almost surely for all $n \in \mathbb{N}$ and $\varepsilon \in (0, 1]$.*

Proof We may and do assume without loss of generality $\beta \in (0, 1]$. By assumption (i) of Theorem 2.1 or (i) of Theorem 2.2 we can find some constant $\tilde{C} > 0$ (in the setting of Theorem 2.1 we may choose $\tilde{C} := C$) such that $g(x) \leq \tilde{C}x^\beta$. Further, since \hat{F}_n has compact support, there is some constant $C_{\hat{F}_n} > 0$ such that $\hat{F}_n(t) \leq C_{\hat{F}_n} \phi_{-\lambda}(t)$, $t < 0$, for some $\lambda > 1/\beta$. Therefore,

$$\begin{aligned} \int_{-\infty}^0 g(P_\varepsilon \hat{F}_n(t)) dt &\leq \int_{-\infty}^0 \tilde{C} (P_\varepsilon \hat{F}_n(t))^\beta dt \\ &= \int_{-\infty}^0 \tilde{C} \left(\int_{\mathbb{R}} \hat{F}_n(y) p_\varepsilon(t-y) dy \right)^\beta dt \\ &\leq \int_{-\infty}^0 \tilde{C} \left(\int_{\mathbb{R}} C_{\hat{F}_n} \phi_{-\lambda}(y) p_\varepsilon(t-y) dy \right)^\beta dt \\ &\leq \int_{-\infty}^0 \tilde{C} C_{\hat{F}_n}^\beta (C_2(\lambda) \phi_{-\lambda}(t))^\beta dt \\ &\leq \tilde{C} C_{\hat{F}_n}^\beta C_2(\lambda)^\beta \int_{-\infty}^0 \phi_{-\lambda\beta}(t) dt, \end{aligned}$$

where we used (8). The latter bound is finite since $\lambda\beta > 1$. In the same way one shows that $\int_0^\infty (1 - g(P_\varepsilon \hat{F}_n(t))) dt < \infty$. Therefore we obtain $P_\varepsilon \hat{F}_n \in \mathbb{F}_g$. \square

Theorem 3.4 (Strong consistency) *Suppose that*

- (a) g satisfies assumption (i) of Theorem 2.1 for some $\beta > 0$,
- (b) $F \in \mathbb{F}_g$ satisfies assumption (iii) of Theorem 2.1, and (6) for some $\gamma > 1/\beta$,
- (c) F is Lipschitz continuous.

Then, if $\varepsilon_n \downarrow 0$, we have $\rho_g(F_n) \rightarrow \rho_g(F)$ \mathbb{P} -almost surely. In the case $\varepsilon_n = 0$, $n \in \mathbb{N}$, the assumption (c) is superfluous.

Proof According to Remark 2.5, it suffices to verify that the assumptions (i)–(v) of Theorem 2.1 hold (\mathbb{P} -almost surely). Conditions (i) and (iii) hold by assumption, and condition (ii) is ensured by Lemma 3.3. If we choose $\lambda \in (1/\beta, \gamma)$ then conditions (iv)–(v) of Theorem 2.1 follow from Lemma 3.2 and Corollary A.2. In the case $\varepsilon_n = 0$, $n \in \mathbb{N}$, we do not need Lemma 3.2. \square

Theorem 3.5 (Rate of convergence) *Suppose that*

- (a) g satisfies assumption (i) of Theorem 2.2 for some $\beta > 0$,

- (b) $F \in \mathbb{F}_g$ satisfies assumption (iii) of Theorem 2.2, and (6) for some $\gamma > \max\{1; 1/\beta\}$,
- (c) F is Lipschitz continuous.

Then, for every $r \in (0, \min\{1/2; 1 - \max\{1; 1/\beta\}/\gamma\})$ and $q \in (0, 1/2 - \max\{1; 1/\beta\}/(2\gamma))$, we can find some \mathbb{R}_+ -valued random variable $K_{r,\gamma}$ and some finite constant $K_{q,\gamma} > 0$ such that \mathbb{P} -almost surely

$$|\rho_g(F_n) - \rho_g(F)| \leq K_{r,\gamma} n^{-(\beta \wedge 1)r} + K_{q,\gamma} \varepsilon_n^{(\beta \wedge 1)q}, \quad n \in \mathbb{N} \quad (10)$$

(provided the sequence $(\varepsilon_n) \subset (0, \infty)$ is bounded). If g is Hölder- β -continuous on all of $[0, 1]$, then in the exponents on the right-hand side of (10) the expression $\beta \wedge 1$ can be replaced by β , we may everywhere replace $\max\{1; 1/\beta\}$ by $1/\beta$, and $K_{r,\gamma}$ can be chosen to be deterministic. In the case $\varepsilon_n = 0$, $n \in \mathbb{N}$, the assumption (c) is superfluous.

Proof As in the proof of Theorem 3.4, one can verify that the assumptions (i)–(iii) of Theorem 2.2 and $\|F_n - F\|_\lambda \rightarrow 0$ hold (\mathbb{P} -almost surely) for any $\lambda \in [1, \gamma)$ (respectively $\lambda \in (0, \gamma)$ in the case where g is Hölder continuous on all of $[0, 1]$) satisfying $\lambda\beta > 1$. Condition (6) ensures that the assumptions of Lemma 3.2 are fulfilled. Thus, Theorem 2.2 (along with Remark 2.5), Lemma 3.2 and Corollary A.2 imply (10) for $r \in (0, \min\{1/2, 1 - \lambda/\gamma\})$ and $q = (\gamma - \lambda)/(2\gamma)$. (In the case $\varepsilon_n = 0$, $n \in \mathbb{N}$, we do not need Lemma 3.2.) Since λ can be chosen arbitrarily close to $\max\{1; 1/\beta\}$, we obtain (10) for every $r \in (0, \min\{1/2; 1 - \max\{1; 1/\beta\}/\gamma\})$ and $q \in (0, (\gamma - \max\{1; 1/\beta\})/(2\gamma))$. If g is Hölder- β -continuous on all of $[0, 1]$, then the same arguments apply if we replace $\max\{1; 1/\beta\}$ by $1/\beta$. \square

Examples for a distortion function g satisfying the assumptions of Theorems 3.4 and 3.5 can be found in Example 2.3. Notice that the assumption $F^{\leftarrow}(d) = F^{\rightarrow}(d)$, for every discontinuity d of g , seems to be necessary. For instance, in the case $g(x) = \mathbb{1}_{[\alpha, 1]}(x)$, i.e. in the case of the Value-at-Risk at level $\alpha \in (0, 1)$, it was shown in [10] that $F^{\leftarrow}(\alpha) < F^{\rightarrow}(\alpha)$ implies that $\rho_g(\hat{F}_n)$ is divergent \mathbb{P} -almost surely.

Remark 3.6 If $\gamma \geq 2 \max\{1; 1/\beta\}$ in the setting of Theorem 3.5 then (10) holds for any $r \in (0, 1/2)$. For all distortion functions g presented in Example 2.3 we may choose $\beta = 1$. Thus, for these g we have $|\rho_g(F_n) - \rho_g(F)| \leq K_{r,\gamma} n^{-r}$ for every $r \in (0, 1/2)$, provided the tails of F decay at least at rate 2 and ε_n tends to 0 at least at rate $(2\gamma r)/(\gamma - 1)$. On the other hand, if ε_n tends to 0 at a rate being strictly larger than $\gamma/(\gamma - 1)$ and g is “regular”, we have a related CLT (cf. [6]). So in this case the rate of convergence of $\rho_g(F_n)$ to $\rho_g(F)$ cannot be improved to $r = 1/2$. We emphasize that the mentioned CLT relies on the assumption $\beta = 1$. A related CLT for the case $\beta \in (0, 1)$ seems to be an open problem, so that in this case it is not immediately obvious whether or not the rate of almost sure convergence specified in Theorem 3.5 can be improved. \diamond

3.2 Censored data

In insurance practice one often encounters the problem that the data is censored. For instance, one might have the following relation between the actual (but unobservable) claim X and the observable fraction \tilde{X} :

$$X = \frac{\tilde{X}}{C} \quad (11)$$

with C some random variable taking values in the interval $(0, 1]$. Here X and \tilde{X} are nonpositive random variables (a negative value of X corresponds to a payout to the client), and \tilde{X} and C are assumed to be independent. We denote by F , \tilde{F} and H the df of X , \tilde{X} and C (respectively), and we assume that H is known.

The goal is the estimation of

$$\rho_g(F) = - \int_{\mathbb{R}} x dg(F(x)) = \int_{(-\infty, 0)} g(F(t)) dt$$

based on i.i.d. copies $\tilde{X}_1, \dots, \tilde{X}_n$ of \tilde{X} . From (11) we have the following representation for the df F

$$F(t) = \int_0^1 \tilde{F}(tz) dH(z), \quad t \leq 0.$$

Thus a natural estimator F_n for F based on the censored observations $\tilde{X}_1, \dots, \tilde{X}_n$ is

$$F_n(t) = \int_0^1 \tilde{F}_n(tz) dH(z) = \frac{1}{n} \sum_{i=1}^n H((\tilde{X}_i/t) \wedge 1), \quad t < 0, \quad (12)$$

with $\tilde{F}_n := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[\tilde{X}_i, 0]}$ the empirical df of $\tilde{X}_1, \dots, \tilde{X}_n$. Theorems 3.9 and 3.10 below show that, under some conditions on g , F and H , the plug-in estimate $\rho_g(F_n)$ for $\rho_g(F)$ is strongly consistent. For their proofs we need two lemmas.

Lemma 3.7 *Let $\lambda > 0$, and suppose that $F \in \mathbb{F}_g$ satisfies (6) for some $\gamma > \lambda$. Suppose further that H has a Lebesgue density h , and there are some constants $K > 0$ and $\delta \in (0, \lambda)$ such that $h(x) \leq Kx^{\delta-1}$ for all $x \in (0, 1]$. Then there is some constant $K_{\lambda, \delta} > 0$ such that*

$$\|F_n - F\|_{\lambda} \leq K_{\lambda, \delta} \|\tilde{F}_n - \tilde{F}\|_{\lambda}, \quad n \in \mathbb{N}. \quad (13)$$

Moreover the right-hand side of (13) converges to 0 \mathbb{P} -almost surely as $n \rightarrow \infty$.

Proof For every $t < 0$ we have

$$\begin{aligned} |F_n(t) - F(t)| \phi_{\lambda}(t) &\leq \int_0^1 |\tilde{F}_n(tz) - \tilde{F}(tz)| \phi_{\lambda}(t) dH(z) \\ &= \int_0^1 |\tilde{F}_n(tz) - \tilde{F}(tz)| \phi_{\lambda}(tz) \frac{\phi_{\lambda}(t)}{\phi_{\lambda}(tz)} h(z) dz \\ &\leq \|\tilde{F}_n - \tilde{F}\|_{\lambda} \int_0^1 \frac{(1+|t|)^{\lambda}}{(1+|tz|)^{\lambda}} h(z) dz \end{aligned}$$

$$\begin{aligned}
&\leq \|\tilde{F}_n - \tilde{F}\|_\lambda (1 + |t|)^\lambda \int_0^1 \frac{1}{(1 + |tz|)^\lambda} K z^{\delta-1} dz \\
&= \|\tilde{F}_n - \tilde{F}\|_\lambda K \delta \frac{(1 + |t|)^\lambda}{|t|^\delta} \int_0^{|t|} \frac{u^{\delta-1}}{(1 + u)^\lambda} du \\
&=: \|\tilde{F}_n - \tilde{F}\|_\lambda K_{\lambda, \delta},
\end{aligned}$$

where we substituted $u := |tz|$. The assumption $\delta < \lambda$ implies $K_{\lambda, \delta} < \infty$, so that (13) holds true. The second statement is a consequence of Corollary A.2 applied to \tilde{F}_n and \tilde{F} ; notice that if F satisfies (6) for some $\gamma > \lambda$ then the same is true for \tilde{F} with the same γ . \square

Lemma 3.8 *Suppose that g satisfies assumption (i) of Theorem 2.1 or assumption (i) of Theorem 2.2 with $\beta > 0$. Suppose further that there are some constants $K > 0$ and $\delta > 1/\beta$ such that $H(x) \leq Kx^\delta$ for all $x \in (0, 1]$. Then we have $F_n \in \mathbb{F}_g$ \mathbb{P} -almost surely for all $n \in \mathbb{N}$.*

Proof By assumption (i) of Theorem 2.1 or (i) of Theorem 2.2 we can find some constant $\tilde{C} > 0$ (in the setting of Theorem 2.1 we may choose $\tilde{C} := C$) such that $g(x) \leq \tilde{C}x^\beta$. Further, it is well known that $|\sum_{i=1}^n b_i|^\beta \leq n^{(\beta-1)\vee 0} \sum_{i=1}^n |b_i|^\beta$ for every $b_1, \dots, b_n \in \mathbb{R}$ and $n \in \mathbb{N}$. With the help of this inequality we obtain by (12)

$$\begin{aligned}
\int_{-\infty}^0 g(F_n(t)) dt &\leq \int_{-\infty}^0 \tilde{C} F_n(t)^\beta dt \\
&= \int_{-\infty}^0 \tilde{C} \left(\frac{1}{n} \sum_{i=1}^n H((\tilde{X}_i/t) \wedge 1) \right)^\beta dt \\
&\leq \int_{-\infty}^0 \tilde{C} \sum_{i=1}^n H((\tilde{X}_i/t) \wedge 1)^\beta dt \\
&\leq \tilde{C} \sum_{i=1}^n \int_{-\infty}^0 K((\tilde{X}_i/t) \wedge 1)^{\delta\beta} dt \\
&= \tilde{C} K \sum_{i=1}^n \left(\int_{\tilde{X}_i}^0 1 dt + \int_{-\infty}^{\tilde{X}_i} (\tilde{X}_i/t)^{\delta\beta} dt \right) \\
&= \tilde{C} K \sum_{i=1}^n \left((-\tilde{X}_i) + \int_0^1 u^{\delta\beta-2} (-\tilde{X}_i) du \right) \\
&= \tilde{C} K \left(1 + \int_0^1 u^{\delta\beta-2} du \right) \sum_{i=1}^n (-\tilde{X}_i),
\end{aligned}$$

where we substituted $u := \tilde{X}_i/t$. Since we assumed $\beta\delta > 1$, this term is finite. Hence, $F_n \in \mathbb{F}_g$ \mathbb{P} -almost surely for all $n \in \mathbb{N}$. \square

Theorem 3.9 (Strong consistency) *Suppose that*

- (a) g satisfies condition (i) of Theorem 2.1 for some $\beta > 0$,
- (b) H has a Lebesgue density h , and there are some constants $K > 0$ and $\delta > 1/\beta$ such that $h(x) \leq Kx^{\delta-1}$, $x \in (0, 1]$,
- (c) $F \in \mathbb{F}_g$ satisfies condition (iii) of Theorem 2.1, and (6) for some $\gamma > \delta$.

Then $\rho_g(F_n) \rightarrow \rho_g(F)$ \mathbb{P} -almost surely.

Proof According to Remark 2.5, it suffices to verify that the assumptions (i)–(v) of Theorem 2.1 hold (\mathbb{P} -almost surely). Conditions (i) and (iii) hold by assumption, and condition (ii) is a consequence of Lemma 3.8. Conditions (iv)–(v) finally follow from Lemma 3.7 with any $\lambda \in (\delta, \gamma)$. \square

Theorem 3.10 (Rate of convergence) *Suppose that*

- (a) g satisfies condition (i) of Theorem 2.2 for some $\beta > 0$,
- (b) H has a Lebesgue density h , and there are some constants $K > 0$ and $\delta > 1/\beta$ such that $h(x) \leq Kx^{\delta-1}$, $x \in (0, 1]$,
- (c) $F \in \mathbb{F}_g$ satisfies condition (iii) of Theorem 2.2, and (6) for some $\gamma > \max\{1; \delta\}$.

Then we can find for every $r \in (0, \min\{1/2; 1 - \max\{1; \delta\}/\gamma\})$ some \mathbb{R}_+ -valued random variable $K_{r,\gamma}$ such that \mathbb{P} -almost surely

$$|\rho_g(F_n(\omega)) - \rho_g(F)| \leq K_{r,\gamma} n^{-(\beta \wedge 1)r}, \quad n \in \mathbb{N}. \quad (14)$$

If g is Hölder- β -continuous on all of $[0, 1]$, then on the right-hand side of (14) the exponent $-(\beta \wedge 1)r$ can be replaced by $-\beta r$, we may everywhere replace $\max\{1; \delta\}$ by δ , and $K_{r,\gamma}$ can be chosen to be deterministic.

Proof As in the proof of Theorem 3.9, one can verify that the assumptions (i)–(iii) of Theorem 2.2 and $\|F_n - F\|_\lambda \rightarrow 0$ hold (\mathbb{P} -almost surely) for any $\lambda \in [1, \gamma)$ (respectively $\lambda \in (0, \gamma)$ in the case where g is Hölder continuous on all of $[0, 1]$) satisfying $\lambda > \delta$. Condition (6) ensures that the assumptions of Lemma 3.7 are fulfilled. Thus, Theorem 2.2 (along with Remark 2.5), Lemma 3.7 and Corollary A.2 (applied to \tilde{F}_n and \tilde{F} ; notice that if F satisfies (6) for some $\gamma > \lambda$ then the same is true for \tilde{F} with the same γ) imply (14) for $r \in (0, \min\{1/2, 1 - \lambda/\gamma\})$. Since λ can be chosen arbitrarily close to $\max\{1; \delta\}$, we obtain (14) for every $r \in (0, \min\{1/2; 1 - \max\{1; \delta\}/\gamma\})$. If g is Hölder- β -continuous on all of $[0, 1]$, then the same arguments apply if we replace $\max\{1; \delta\}$ by δ . \square

Examples for a distortion function g and a df H satisfying the assumptions of Theorems 3.9 and 3.10 can be found in Example 2.3 and Example 3.11, respectively. Notice that

if $\gamma \geq 2 \max\{1; \delta\}$ in the setting of Theorem 3.10 then (14) holds for any $r \in (0, 1/2)$. Moreover notice that in [6] a related CLT for $\rho_g(F_n)$ is given. Therefore the essence of Remark 3.6 applies also to the present setting.

Example 3.11 Let H be the df of the beta distribution on $(0, 1]$ with parameters $a, b > 0$, whose Lebesgue density is given by

$$h(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, \quad x \in (0, 1), \quad (15)$$

where $B(a, b)$ is the beta function. If $a > 1/\beta$, then H satisfies the assumptions (b) of Theorems 3.9 and 3.10 for $\delta := a$. \diamond

4 Simulations

Theorem 3.5 specifies the precision of the estimator $\rho_g(F_n)$ for “large” sample sizes n . For “small” sample sizes the goodness of the estimator $\rho_g(F_n)$ is hard to determine. For this reason we are going to study the estimation error for smaller sample sizes by means of simulations. Each single Monte Carlo simulation will be based on $N = 5'000$ i.i.d. replications X_1, \dots, X_N with df $F^\theta(t) = e^{-\theta t} \mathbb{1}_{(-\infty, 0)}(t) + \mathbb{1}_{[0, \infty)}(t)$ for $\theta = 0.1$. That is, $-X_1$ is exponentially distributed with respect to the parameter θ . In particular, $\mathbb{E}[-X_1] = m_\theta := 1/\theta$.

Theorem 3.5 indicates that the Hölder exponent β of g plays an essential role for the goodness of the estimate $\rho_g(F_n)$ for $\rho_g(F)$. One would expect that a larger Hölder exponent β entails a better approximation of $\rho_g(F)$ by $\rho_g(F_n)$. This guess is backed by our simulation results, cf. Figure 1. We let F_n be the classical empirical df of X_1, \dots, X_n . Moreover we set $g_1(x) = \mathbb{1}_{[\alpha, 1]}(x)$, $g_2(x) = (x/\alpha) \wedge 1$ and $g_3(x) = (x^\beta/\alpha^\beta) \wedge 1$ with $\alpha = 0.25$ and $\beta = 0.5$. Figure 1 displays the path $n \mapsto (\rho_{g_i}(F_n) - \rho_{g_i}(F^\theta))/m_\theta$ for $i = 1, 2, 3$ (solid, dashed, dotted). The plotted paths show indeed that in the case $i = 3$, where the Hölder exponent is smallest, the approximation of $\rho_g(F)$ by $\rho_g(F_n)$ is worst.

5 Proof of main theorems

5.1 Proof of Theorem 2.1

The proof of Theorem 2.1 relies on the following lemma.

Lemma 5.1 *Let $F, F_1, F_2, \dots \in \mathbb{F}_g$, and suppose that*

- (a) $\lim_{K \rightarrow \infty} \sup_{n \in \mathbb{N}} \int_{(-K, K)^c} |x| dg(F_n(x)) = 0$,
- (b) $\lim_{n \rightarrow \infty} g(F_n(t)) = g(F(t))$ for Lebesgue almost every $t \in \mathbb{R}$.

Then $\rho_g(F_n) \rightarrow \rho_g(F)$.

Proof Using the triangle inequality we obtain for every $K > 0$,

$$\begin{aligned} |\rho_g(F_n) - \rho_g(F)| &\leq \int_{(-K,K)^c} |x| dg(F_n(x)) + \int_{(-K,K)^c} |x| dg(F(x)) \\ &\quad + \left| \int_{(-K,0)} x dg(F_n(x)) - \int_{(-K,0)} x dg(F(x)) \right| \\ &\quad + \left| \int_{[0,K)} x dg(F_n(x)) - \int_{[0,K)} x dg(F(x)) \right| \\ &=: S_1(n, K) + \dots + S_4(n, K). \end{aligned}$$

Now, let $\varepsilon > 0$. Because of $F \in \mathbb{F}_g$ and assumption (a), we may and do pick $K_\varepsilon > 0$ such that $S_i(n, K_\varepsilon) < \varepsilon/4$ for $i = 1, 2$ and all $n \in \mathbb{N}$. For every df H on the real line we have $\int_0^K x dH(x) = \int_0^K (H(K) - H(t)) dt$. For the fourth summand we thus obtain

$$\begin{aligned} S_4(n, K_\varepsilon) &= \left| \int_{[0, K_\varepsilon)} (g(F_n(K_\varepsilon)) - g(F_n(t))) dt - \int_{[0, K_\varepsilon)} (g(F(K_\varepsilon)) - g(F(t))) dt \right| \\ &\leq \int_{[0, K_\varepsilon)} |g(F_n(K_\varepsilon)) - g(F(K_\varepsilon))| dt + \int_{[0, K_\varepsilon)} |g(F(t)) - g(F_n(t))| dt. \end{aligned}$$

If we assume without loss of generality that K_ε is *not* an exceptional t in condition (b), then the latter bound converges to 0 as $n \rightarrow \infty$ by assumption (b). Analogously we obtain $S_3(n, K_\varepsilon) \rightarrow 0$ as $n \rightarrow \infty$. That is, we may and do pick some $n_\varepsilon \in \mathbb{N}$ such that $S_i(n, K_\varepsilon) < \varepsilon/4$ for $i = 3, 4$ and all $n \geq n_\varepsilon$. By all account, we can find for every $\varepsilon > 0$ some $n_\varepsilon \in \mathbb{N}$ such that $|\rho_g(F_n) - \rho_g(F)| < \varepsilon$ for all $n \geq n_\varepsilon$. \square

Remark 5.2 It is well known that L^p -boundedness for some $p > 1$ is sufficient for uniform integrability. That is, conditions (a) of Lemma 5.1 is implied by

$$(a)' \exists p > 1: \sup_{n \in \mathbb{N}} \int_{\mathbb{R}} |x|^p dg(F_n(x)) < \infty.$$

Also, since the nondecreasing g is dt -almost everywhere continuous, one easily verifies that condition (b) of Lemma 5.1 is satisfied if

$$(b)' F_n(t) \rightarrow F(t) \text{ for } dt\text{-a.e. } t \in \mathbb{R}, \text{ and } F^{\leftarrow}(d) = F^{\rightarrow}(d) \text{ for every discontinuity } d \text{ of } g,$$

where F^{\leftarrow} and F^{\rightarrow} refer to the left- and the right-continuous inverse of F , respectively. \diamond

In view of Lemma 5.1 and Remark 5.2, the claim of Theorem 2.1 holds if we can show that the conditions (a)'–(b)' of Remark 5.2 hold. By assumption (i) of Theorem 2.1 we obtain for any $p \in (1, \lambda\beta)$

$$\int_0^\infty x^p dg(F_n(x)) = \int_0^\infty (1 - g(F_n(t^{1/p}))) dt$$

$$\begin{aligned}
&\leq \int_0^\infty C\bar{F}_n(t^{1/p})^\beta dt \\
&\leq C\left(1 + \int_1^\infty \bar{F}_n(s)^\beta p s^{p-1} dt\right) \\
&\leq C\left(1 + p(\|\bar{F}_n \mathbb{1}_{[0,\infty)}\|_\lambda)^\beta \int_1^\infty (1+|s|)^{-\lambda\beta} s^{p-1} dt\right),
\end{aligned}$$

where we substituted $s := t^{1/p}$. Since $-\lambda\beta + p - 1 < -1$ and $\sup_{n \in \mathbb{N}} \|\bar{F}_n \mathbb{1}_{[0,\infty)}\|_\lambda < \infty$ by assumption (v), the latter expression is bounded above uniformly in $n \in \mathbb{N}$. Completely analogously we can show $\sup_{n \in \mathbb{N}} \int_{-\infty}^0 |x|^p dg(F_n(x)) < \infty$. Thus, assertion (a)' holds true. Assertion (b)' is ensured by assumptions (iii)–(iv). This finishes the proof of Theorem 2.1.

5.2 Proof of Theorem 2.2

At first we assume $k \geq 1$. By assumption (iii) we have $m := \min_{i=1,\dots,k} F'(F^{-1}(d_i)) > 0$. We set $\varepsilon_n := \|F_n - F\|_\infty$, $\delta_n := 2\varepsilon_n/m$, and we define the intervals

$$\begin{aligned}
I_{0,n} &:= (-\infty, F^{-1}(d_1) - \delta_n) & J_{0,n} &:= [0, d_1 - \varepsilon_n) \\
I_{i,n} &:= (F^{-1}(d_i) + \delta_n, F^{-1}(d_{i+1}) - \delta_n) & J_{i,n} &:= (d_i + \varepsilon_n, d_{i+1} - \varepsilon_n) \\
I_{k,n} &:= (F^{-1}(d_k) + \delta_n, \infty) & J_{k,n} &:= (d_k + \varepsilon_n, 1]
\end{aligned}$$

($i = 1, \dots, k-1$). For $\delta_n = 2\varepsilon_n/m$ sufficiently small, i.e. for n sufficiently large (notice that assumption (iv) is $\varepsilon_n \rightarrow 0$), we obtain $F(I_{i,n}) \subset J_{i,n}$ for every $i = 0, \dots, k$. In this case, $t \in I_{i,n}$ implies $F(t) \in J_{i,n}$. By the definitions of ε_n and $J_{i,n}$, we conclude that $F_n(t)$ lies in (d_i, d_{i+1}) . That is, for n sufficiently large, we have for all $t \in I_{i,n}$ that $F(t), F_n(t) \in (d_i, d_{i+1})$. Thus, for n sufficiently large, we have for all $t \in I_n := I_{0,n} \cup \dots \cup I_{k,n}$ that $|g(F_n(t)) - g(F(t))| \leq C|F_n(t) - F(t)|^\beta$ by assumption (i). With the representation (3) in mind, we thus obtain for n sufficiently large

$$\begin{aligned}
|\rho_g(F_n) - \rho_g(F)| &\leq \int_{I_n^c} (g(F_n(t)) + g(F(t))) dt + \int_{I_n} |g(F_n(t)) - g(F(t))| dt \\
&\leq k4\delta_n + \int_{I_n} C|F_n(t) - F(t)|^\beta dt \\
&\leq \frac{8k}{m} \|F_n - F\|_\infty + \int_{\mathbb{R}} C|(F_n(t) - F(t))\phi_\lambda(t)|^\beta \phi_{\lambda\beta}(t)^{-1} dt \\
&\leq \frac{8k}{m} \|F_n - F\|_\infty + \left(C \int_{\mathbb{R}} \phi_{\lambda\beta}(t)^{-1} dt\right) \|F_n - F\|_\lambda^\beta.
\end{aligned}$$

Since we assumed $\lambda \geq 1$, this yields (4) with $K := (8k/m) + (C \int_{\mathbb{R}} \phi_{\lambda\beta}(t)^{-1} dt)$.

If g is Hölder- β -continuous on all of $[0, 1]$, i.e. if $k = 0$, then we easily obtain

$$|\rho_g(F_n) - \rho_g(F)| \leq \left(C \int_{\mathbb{R}} \phi_{\lambda\beta}(t)^{-1} dt\right) \|F_n - F\|_\lambda^\beta$$

without assuming $\lambda \geq 1$. This yields (4) with $\beta \wedge 1$ replaced by β and $K := C \int_{\mathbb{R}} \phi_{\lambda\beta}(t)^{-1} dt$.

A Glivenko-Cantelli theorem for weighted errors

Suppose F is a df on the real line. Further suppose X_1, X_2, \dots are i.i.d. random variables with df F on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $F_n := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[X_i, \infty)}$ denote the corresponding empirical df at stage n . The classical Glivenko-Cantelli theorem states that $\|F_n - F\|_\infty \rightarrow 0$ \mathbb{P} -almost surely. For our purposes we need a stronger convergence result. In fact we need the convergence with respect to the more stringent norm $\|\cdot\|_\lambda$ defined at the beginning of Section 2. Corollary A.2 below provides a respective result. It relies on the following theorem, which is an immediate consequence of [3, Theorem 7.3 (3)].

Theorem A.1 *Let G_n be the empirical df at stage n of a sequence of independent and identically $U[0, 1]$ -distributed random variables on some probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$. If $r \in [0, 1/2)$ and $\theta \in [0, 1 - r)$, then $\bar{\mathbb{P}}$ -almost surely*

$$\lim_{n \rightarrow \infty} n^r \sup_{x \in (0, 1)} \frac{|G_n(x) - x|}{x^\theta (1 - x)^\theta} = 0.$$

Corollary A.2 *Let $\lambda > 0$, and suppose there is some constant $\gamma > \lambda$ such that*

$$\limsup_{t \rightarrow -\infty} F(t)|t|^\gamma < \infty \quad \text{and} \quad \limsup_{t \rightarrow \infty} \bar{F}(t)t^\gamma < \infty. \quad (16)$$

Then we have for every $r \in [0, \min\{1/2, 1 - \lambda/\gamma\})$ \mathbb{P} -almost surely

$$\lim_{n \rightarrow \infty} n^r \|F_n - F\|_\lambda = 0.$$

Proof We may and do choose a sequence of independent and identically $U[0, 1]$ -distributed random variables, possibly on an extension of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that the corresponding empirical df G_n satisfies $F_n = G_n(F)$ \mathbb{P} -almost surely (cf. [21] or [20, p.103]). Further, assumption (16) implies the existence of some constant $C > 0$ such that $|F^\leftarrow(x)| \leq Cx^{-1/\gamma}$ for all $x \in (0, x_0)$, and $F^\rightarrow(x) \leq C(1 - x)^{-1/\gamma}$ for all $x \in (x_0, 1)$, where x_0 is chosen such that $F^\leftarrow(x_0) \leq 0 \leq F^\rightarrow(x_0)$. Thus

$$\begin{aligned} n^r \|F_n - F\|_\lambda &= n^r \|G_n(F(\cdot)) - F\|_\lambda \\ &= n^r \sup_{t \in \mathbb{R}} |G_n(F(t)) - F(t)| (1 + |t|)^\lambda \\ &\leq n^r \sup_{x \in (0, 1)} |G_n(x) - x| \left(1 + \frac{C}{x^{1/\gamma}(1 - x)^{1/\gamma}}\right)^\lambda. \end{aligned}$$

Theorem A.1 with $\theta = \lambda/\gamma$ implies that the latter expression converges to 0 as $n \rightarrow \infty$. \square

B Examples for distribution-invariant risk measures

The possibly most popular risk measure in practice, the Value-at-Risk VaR_α at level $\alpha \in (0, 1)$, is defined by

$$\text{VaR}_\alpha^\mathbb{P}(X) := \text{VaR}_\alpha(F_X) := -F_X^\rightarrow(\alpha) = F_X^-(1 - \alpha)$$

for any random variable X on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is a distortion risk measure with respect to $g(x) = \mathbb{1}_{[\alpha, 1]}(x)$. The drawback of VaR_α is the well-known fact that it is not subadditive and therefore not coherent (note that g is not concave). The most popular coherent risk measure dominating VaR_α is the Average Value-at-Risk (also called Expected Shortfall) at level $\alpha \in (0, 1)$:

$$\text{AVaR}_\alpha^\mathbb{P}(X) := \text{AVaR}_\alpha(F_X) := \frac{1}{\alpha} \int_0^\alpha \text{VaR}_a(F_X) da$$

for $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. It provides a distortion risk measure with respect to $g(x) = (x/\alpha) \wedge 1$. Note that if $F_X(F_X^\rightarrow(\alpha)) = \alpha$ then $\text{AVaR}_\alpha^\mathbb{P}(X) = \mathbb{E}[-X | -X \geq \text{VaR}_\alpha(F_X)]$, cf. [2]. Another common distortion risk measure is the Wang transform WT_θ with parameter $\theta \in \mathbb{R}$ which is defined by

$$\text{WT}_\theta^\mathbb{P}(X) := \text{WT}_\theta(F_X) := \int_0^1 \text{VaR}_a(F_X) \psi_\theta(a) da$$

for all random variables X for which the integral exists, where $\psi_\theta(a) := e^{\theta\Phi^{-1}(a) - \theta^2/2}$ and Φ^{-1} denotes the inverse of the standard normal df Φ . It corresponds to the distortion function $g(x) = \Phi(\Phi^{-1}(x) - \theta)$. The Wang transform WT_θ is coherent if and only if $\theta \leq 0$.

Notice that the Value-at-Risks are not only distortion risk measures for themselves but they even are the building blocks of every distortion risk measure. Indeed, it can straightforwardly be shown with the help of standard arguments of integration theory that

$$\rho_g(X) = \int_0^1 \text{VaR}_a(X) dg(a) \tag{17}$$

for every random variable X for which the integral in (2) exists. If g is concave then $\rho_g(X)$ can also be represented as $\rho_g(X) = \int_0^1 \text{AVaR}_a(X) d\tilde{g}(a)$, where g and \tilde{g} correspond one-to-one via $g'_+(x) = \int_x^1 s^{-1} d\tilde{g}(s)$, cf. Lemma 4.63 and Corollary 4.71 in [12]. Moreover, it was shown in [17, 18] that, under some mild assumptions, every distribution-invariant coherent risk measure ρ can be represented as

$$\rho(X) = \inf_{h \in \mathcal{G}} \int_0^1 \text{AVaR}_a(F) dh(a) \tag{18}$$

for some set \mathcal{G} of distortion functions. For details and examples the reader is kindly referred to [18, 27]. Notice that the distortion risk measure ρ_g has the representation (18) with $\mathcal{G} = \{\tilde{g}\}$.

References

- [1] Acerbi, C. (2002) Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26, 1505-1518.
- [2] Acerbi, C. and Tasche, D. (2002) Expected Shortfall: a natural coherent alternative to Value at Risk. *Economic Notes*, 31(2), 379–388.
- [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(2), 303–335.
- [4] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1997) Thinking coherently. *Risk*, 10(11), 68–71.
- [5] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999) Coherent measures of risk. *Mathematical Finance*, 9, 203–228.
- [6] Beutner, E. and Zähle, H. (2009) A modified functional delta method and its application to the estimation of risk measures. *Submitted*.
- [7] Bronstein, I.N., Semendyayev, K.A., Musiol, G. and Mühlig, H. (2002) *Handbook of Mathematics*. Springer-Verlag, Berlin.
- [8] Chen, S.X. (2008) Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1), 87–107.
- [9] Delbaen F. (2000), *Coherent Risk Measures*, Lecture Notes, Pisa.
- [10] Feldman, D. and Tucker, H.G. (1966) Estimation of non-unique quantiles. *Annals of Mathematical Statistics*, 37, 451–457.
- [11] Föllmer, H. and Schied, A. (2002) Convex measures of risk and trading constraints. *Finance & Stochastics*, 6(4), 429–447.
- [12] Föllmer, H. and Schied, A. (2004) *Stochastic finance. An introduction in discrete time*. de Gruyter, Berlin.
- [13] Hardy, M.R. and Wirth, J. (2001) Distortion risk measures: coherence and stochastic dominance. *Working Paper, available in www.gloriamundi.com*.
- [14] Jones, B.L., Puri, M.L. and Zitikis, R. (2006) Testing hypotheses about the equality of several risk measure values with applications in insurance. *Insurance: Mathematics and Economics*, 38, 253–270

- [15] Jones, B.L. and Zitikis, R. (2003) Empirical estimation of risk measures and related quantities. *The North American Actuarial Journal*, 7(4), 44–54.
- [16] Jones, B.L. and Zitikis, R. (2007) Risk measures, distortion parameters, and their empirical estimation. *Insurance: Mathematics and Economics*, 41(2), 279–297.
- [17] Kusuoka, S. (2001) On law invariant risk measures. *Advances in Mathematical Economics*, 3, 83–95.
- [18] Pflug, G.Ch. and Römisch, W. (2007) *Modelling, managing and measuring risk*. World Scientific Publishing, Singapore.
- [19] Sheather, S.J. and Marron, J.S. (1990) Kernel quantile estimators. *Journal of American Statistical Association*, 85, 410–416.
- [20] Shorack, G.R. and Wellner, J.A. (1986) *Empirical processes with applications to statistics*. Wiley, New York.
- [21] Shorack, G.R. (1979) The weighted empirical process of row independent random variables with arbitrary d.f.'s. *Statistica Neerlandica*, 33(1), 169–189.
- [22] Stigler, S. (1974) Linear functions of order statistics with smooth weight functions, *Annals of Statistics*, 2, 676–693.
- [23] van der Vaart, A.W. (1998) *Asymptotic statistics*. Cambridge University Press, Cambridge.
- [24] Wang, S.S. (1996) Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26, 71–92.
- [25] Wang, S.S. (2000) A class of distortion operators for pricing financial and insurance risks. *Actuarial Research Clearing House* Vol. 1.
- [26] Wang, S.S. and Dhaene, J. (1998) Comonotonicity, correlation order and premium principles. *Insurance: Mathematics and Economics*, 22, 235–242.
- [27] Wozabal, D. and Wozabal, N. (2009) Asymptotic consistency of risk functionals. *Journal of Nonparametric Statistics*, 21(8), 977–990.
- [28] Yamato, H. (1972). Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics*, 15, 69–78.

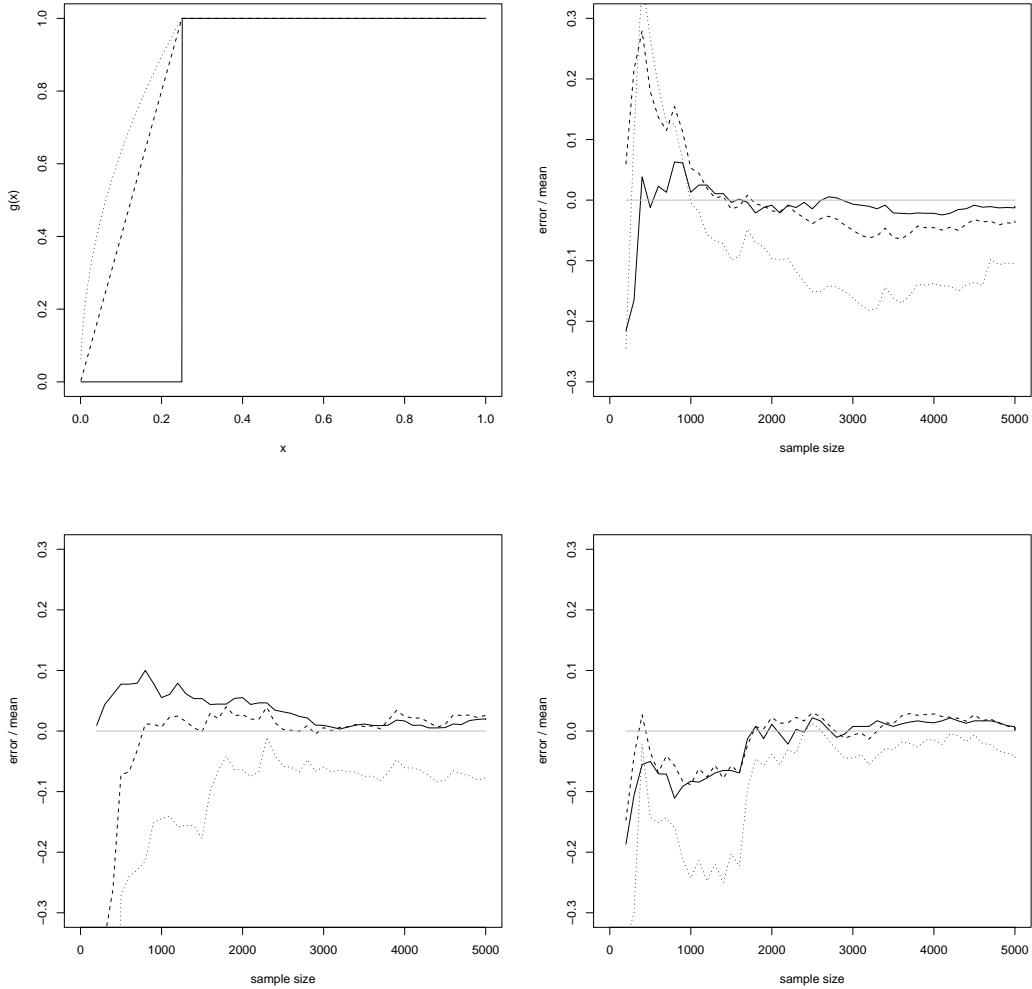


Figure 1: Top left: distortion functions g_1, g_2, g_3 (solid, dashed, dotted) defined in Section 4. From the top right to the bottom right: Paths of $n \mapsto (\rho_{g_i}(F_n) - \rho_{g_i}(F^\theta))/m_\theta$ for $i = 1, 2, 3$ (solid, dashed, dotted), where F_n is the empirical df at stage n based on i.i.d. replications X_1, X_2, \dots with $X_1 \sim \text{Exp}_\theta$, $\theta := 0.1$, $m_\theta := \mathbb{E}[-X_1] = 1/\theta$, and F^θ is the common df of the X_i .

- 2010-01 **Henryk Zähle**
Rates of almost sure convergence of plug-in estimates for distortion risk measures
- 2009-16 **Lorenz J. Schwachhöfer**
Nonnegative curvature on disk bundles
- 2009-15 **Iuliu Pop and Ben Schweizer**
Regularization schemes for degenerate Richards equations and outflow conditions
- 2009-14 **Guy Bouchitté and Ben Schweizer**
Cloaking of small objects by anomalous localized resonance
- 2009-13 **Tom Krantz, Lorenz J. Schwachhöfer**
Extrinsically Immersed Symplectic Symmetric Spaces
- 2009-12 **Alexander Kaplun**
Continuous time Ehrenfest process in term structure modelling
- 2009-11 **Henryk Zähle**
Ein aktuarielles Modell für die Portabilität der Alterungsrückstellungen in der PKV
- 2009-10 **Andreas Neuenkirch and Henryk Zähle**
Asymptotic error distribution of the Euler method for SDEs with non-Lipschitz coefficients
- 2009-09 **Karl Friedrich Siburg, Pavel A. Stoimenov**
Regression dependence
- 2009-08 **Wilfried Hazod**
Continuous convolution hemigroups integrating a sub-multiplicative function
- 2009-07 **Sergio Conti and Ben Schweizer**
On optimal metrics preventing mass transfer
- 2009-06 **Simon Castle, Norbert Peyerimhoff, Karl Friedrich Siburg**
Billiards in ideal hyperbolic polygons
- 2009-05 **Ludwig Danzer**
Quasiperiodic Tilings - Substitution Versus Inflation
- 2009-04 **Flavius Guias**
Direct simulation of the infinitesimal dynamics of semi-discrete approximations for convection-diffusion-reaction problems
- 2009-03 **Franz Kalhoff and Victor Pambuccian**
Existential definability of parallelism in terms of betweenness in Archimedean ordered affine geometry
- 2009-02 **Fulvia Buzzi, Michael Lenzinger and Ben Schweizer**
Interface conditions for degenerate two-phase flow equations in one space dimension
- 2009-01 **Henryk Zähle**
Approximation of SDEs by population-size-dependent Galton-Watson processes

- 2008-25 **Winfried Hazod**
Mehler semigroups, Ornstein-Uhlenbeck processes and background driving Lévy processes on locally compact groups and on hypergroups
- 2008-24 **Karl Friedrich Siburg, Pavel A. Stoimenov**
Symmetry of functions and exchangeability of random variables
- 2008-23 **Ina Kirsten Voigt**
Voronoi Cells of Discrete Point Sets
- 2008-22 **Michael Lenzinger and Ben Schweizer**
Effective reaction rates of a thin catalyst layer
- 2008-21 **Michael Voit**
Bessel convolutions on matrix cones: Algebraic properties and random walks
- 2008-20 **Margit Rösler and Michael Voit**
Limit theorems for radial random walks on $p \times q$ -matrices as p tends to infinity
- 2008-19 **Michael Voit**
Central Limit Theorems for Radial Random Walks on $p \times q$ Matrices for $p \rightarrow \infty$
- 2008-18 **Michael Voit**
Limit theorems for radial random walks on homogeneous spaces with growing dimensions
- 2008-17 **Ansgar Steland and Henryk Zähle**
Sampling inspection by variables: nonparametric setting
- 2008-16 **Guy Bouchitté and Ben Schweizer**
Homogenization of Maxwell's equations with split rings
- 2008-15 **Winfried Hazod**
Multiple selfdecomposable laws on vector spaces and on groups: The existence of background driving processes
- 2008-14 **Winfried Hazod**
Mixing of generating functionals and applications to (semi-)stability of probabilities on groups
- 2008-13 **Winfried Hazod**
Probability on Matrix-Cone Hypergroups: Limit Theorems and Structural Properties
- 2008-12 **Michael Lenzinger and Ben Schweizer**
Two-phase flow equations with outflow boundary conditions in the hydrophobic-hydrophilic case
- 2008-11 **Karl Friedrich Siburg**
Geometric proofs of the two-dimensional Borsuk-Ulam theorem