# Robustness issues in risk estimation

Alexander Schied
University of Waterloo

**Workshop on risk measurement and regulatory issues in business**
**Centre de recherches mathématiques, Montreal**

**September 14, 2017**

Based on joint work with
Volker Krätschmer (Duisburg) and Henryk Zähle (Saarbrücken)

# 1 Introduction

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be atomless and let $\mathscr{X} \subset L^0 := L^0(\Omega, \mathscr{F}, \mathbb{P})$ be a vector space containing the constants. A map $\rho : \mathscr{X} \to \mathbb{R}$ is a convex risk measure if the following conditions are satisfied:

(i) monotonicity: $\rho(X) \geq \rho(Y)$ for $X, Y \in \mathscr{X}$ with $X \leq Y$;

(ii) convexity: $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y)$ for all $X, Y \in \mathscr{X}$ and $\lambda \in [0,1]$;

(iii) cash additivity: $\rho(X + m) = \rho(X) - m$ for $X \in \mathscr{X}$ and $m \in \mathbb{R}$.

*Note:* it is possible here to replace axiom (iii) by the following weaker notion:

(iii') cash coercivity: $\rho(-m) \longrightarrow +\infty$ if $m \in \mathbb{R}$ tends to $+\infty$.

If $\rho$ is law-invariant,

$$\rho(X) = \rho(\widetilde{X}) \text{ whenever } X \text{ and } \widetilde{X} \text{ have the same law under } \mathbb{P},$$

it makes sense to estimate $\rho(X)$ by means of a Monte Carlo procedure or from a sequence of historical data.

Let

$$\mathscr{M}(\mathscr{X}) := \{\mathbb{P} \circ X^{-1} \,|\, X \in \mathscr{X}\}$$

Law invariance of a risk measure $\rho : \mathscr{X} \to \mathbb{R}$ is equivalent to the existence of a map

$$\mathscr{R}_\rho : \mathscr{M}(\mathscr{X}) \to \mathbb{R}$$

such that

$$\rho(X) = \mathscr{R}_\rho(\mathbb{P} \circ X^{-1}), \qquad X \in \mathscr{X}. \tag{1}$$

This map $\mathscr{R}_\rho$ will be called the risk functional associated with $\rho$.

If $\widehat{\mu}_n$ is an estimator for the law $\mu = \mathbb{P} \circ X^{-1}$ of $X$ then

$$\widehat{\rho}_n := \mathscr{R}_\rho(\widehat{\mu}_n) \tag{2}$$

is an estimator for $\rho(X)$. A typical choice is the empirical distribution of a sequence $X_1, \dots, X_n$ of historical observations or Monte Carlo simulations

$$\widehat{\mu}_n = \widehat{m}_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$

If $\widehat{\mu}_n$ is an estimator for the law $\mu = \mathbb{P} \circ X^{-1}$ of $X$ then

$$\widehat{\rho}_n := \mathscr{R}_\rho(\widehat{\mu}_n) \tag{2}$$

is an estimator for $\rho(X)$. A typical choice is the empirical distribution of a sequence $X_1, \ldots, X_n$ of historical observations or Monte Carlo simulations

$$\widehat{\mu}_n = \widehat{m}_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$

**Questions:**

- **Consistency:** do we have $\widehat{\rho}_n \to \rho(X)$ as $n \uparrow \infty$?
- **Continuity:** is $\mu \mapsto \mathscr{R}_\rho(\mu)$ continuous?
- **Asymptotic analysis:** what can be said about the asymptotic distribution of the estimation error $\widehat{\rho}_n - \rho(X)$?
- **Robustness:** is the law of $\widehat{\rho}_n$ stable with respect to small perturbations of the law generating the $X_1, \ldots, X_n$?

If $\widehat{\mu}_n$ is an estimator for the law $\mu = \mathbb{P} \circ X^{-1}$ of $X$ then

$$\widehat{\rho}_n := \mathscr{R}_\rho(\widehat{\mu}_n) \tag{2}$$

is an estimator for $\rho(X)$. Typical choice is the empirical distribution of a sequence $X_1, \ldots, X_n$ of historical observations or Monte Carlo simulations

$$\widehat{\mu}_n = \widehat{m}_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$

**Questions:**

- **Robustness:** is the law of $\widehat{\rho}_n$ stable with respect to small perturbations of the law generating the $X_1, \ldots, X_n$?

  Cont, Deguest, and Scandolo (2010):
  If $\rho$ is **coherent**, the risk functional $\mathscr{R}_\rho$ cannot be qualitatively robust in the sense of Hampel (1971). However, $\mathscr{R}_{V@R}$ is essentially qualitatively robust.

Cont, Deguest, and Scandolo (2010):
If $\rho$ is **coherent**, the risk functional $\mathscr{R}_\rho$ cannot be qualitatively robust in the sense of Hampel (1971). However, $\mathscr{R}_{V@R}$ is essentially qualitatively robust.

Sounds like a strong argument in favor of $V@R$ and against coherent risk measures.

Cont, Deguest, and Scandolo (2010):
If $\rho$ is **coherent**, the risk functional $\mathscr{R}_\rho$ cannot be qualitatively robust in the sense of Hampel (1971). However, $\mathscr{R}_{V@R}$ is essentially qualitatively robust.

Sounds like a strong argument in favor of $V@R$ and against coherent risk measures.

But let's review the argument.

- It does not address the main objection against $V@R$:
  $V@R$ creates toxic incentives.

Cont, Deguest, and Scandolo (2010):
If $\rho$ is **coherent**, the risk functional $\mathscr{R}_\rho$ cannot be qualitatively robust in the sense of Hampel (1971). However, $\mathscr{R}_{V@R}$ is essentially qualitatively robust.

Sounds like a strong argument in favor of $V@R$ and against coherent risk measures.

But let's review the argument.

- It does not address the main objection against $V@R$:
  $V@R$ creates toxic incentives.

- By Hampel's theorem, qualitative robustness of $\mathscr{R}_\rho$ implies that $\mathscr{R}_\rho$ is continuous with respect to weak convergence in $\mathscr{M}_1(\mathbb{R})$.

  Since the compactly supported probability measures are dense in $\mathscr{M}_1(\mathbb{R})$ with respect to weak convergence, Hampel's theorem thus implies that $\mathscr{R}_\rho(\mu)$ must be insensitive to the tail behavior of $\mu$

Cont, Deguest, and Scandolo (2010):
If $\rho$ is **coherent**, the risk functional $\mathscr{R}_\rho$ cannot be qualitatively robust in the sense of Hampel (1971). However, $\mathscr{R}_{V@R}$ is essentially qualitatively robust.

Sounds like a strong argument in favor of $V@R$ and against coherent risk measures.

But let's review the argument.

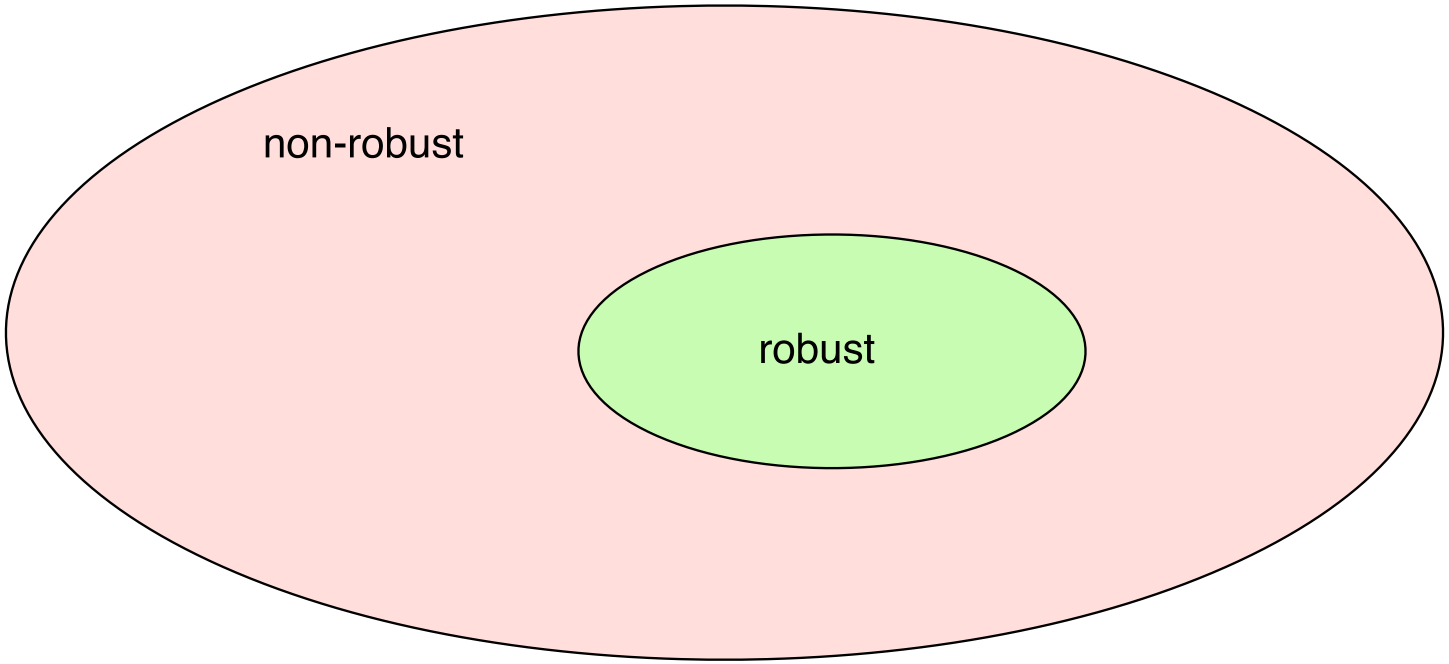
- It does not address the main objection against $V@R$:
  $V@R$ creates toxic incentives.

- By Hampel's theorem, qualitative robustness of $\mathscr{R}_\rho$ implies that $\mathscr{R}_\rho$ is continuous with respect to weak convergence in $\mathscr{M}_1(\mathbb{R})$.

  Since the compactly supported probability measures are dense in $\mathscr{M}_1(\mathbb{R})$ with respect to weak convergence, Hampel's theorem thus implies that $\mathscr{R}_\rho(\mu)$ must be insensitive to the tail behavior of $\mu$.

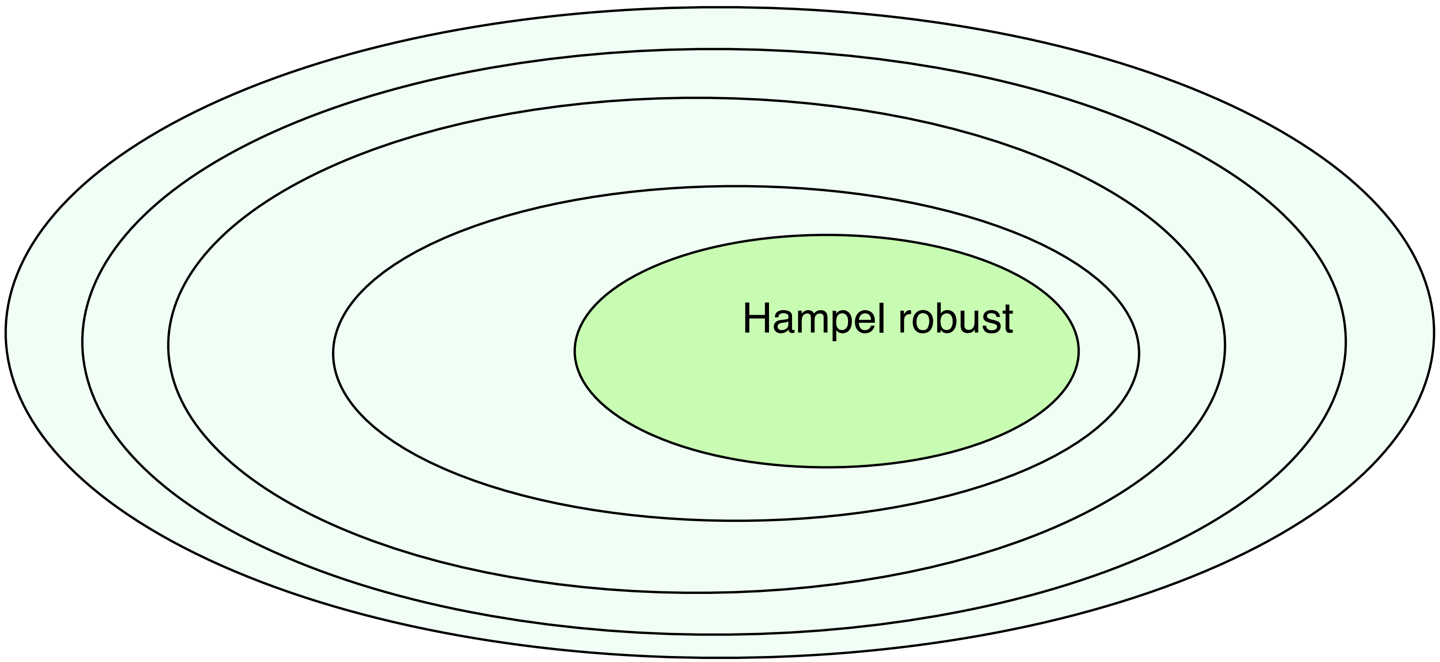
- $\rho(X) = -\mathbb{E}[\, X \,]$ is a coherent risk measure and thus corresponds to a non-robust risk functional. So already estimating the expectation is not robust.

- Hampel's terminology of qualitative robustness generates a sharp division of risk functionals into those that are called "robust" and others that are called "not robust".




But, e.g., estimating the expected value should be "more robust" than estimating variance.

Can one thus define a **refined** notion of robustness that produces a picture like the following one?




Such a refined notion of robustness will help us to bring the argument against coherent risk measures back into perspective: robustness is not lost entirely but only to some degree when Value at Risk is replaced by a coherent risk measure. It might also capture the natural tradeoff between robustness and tail-sensitivity.

Such refined versions of robustness was introduced in Krätschmer, A.S., and Zähle (2012, 2017), and here we apply it to, and explain it at the hand of, law-invariant convex risk measures and other statistical functionals (Krätschmer, A.S., and Zähle, 2014)

Such refined versions of robustness was introduced in Krätschmer, A.S., and Zähle (2012, 2017), and here we apply it to, and explain it at the hand of, law-invariant convex risk measures and other statistical functionals (Krätschmer, A.S., and Zähle, 2014)

The key is to replace metrics for the weak topology in Hampel's robustness with metrics for the $\psi$-weak topology on

$$\mathscr{M}_1^{\psi} := \mathscr{M}_1^{\psi}(\mathbb{R}) := \Big\{ \mu \in \mathscr{M}_1(\mathbb{R}) \,\Big|\, \int \psi \, d\mu < \infty \Big\}$$

where $\psi : \mathbb{R} \to [0, \infty)$ is a continuous weight function satisfying $\psi \geq 1$ outside some compact set.

Typical example: $\psi(x) = |x|^p$

We have

$$\begin{aligned}\mu_n \longrightarrow \mu \ \psi\text{-weakly} \quad &:\Longleftrightarrow \quad \int f\, d\mu_n \longrightarrow \int f\, d\mu \ \forall \text{ continuous } f \text{ with } |f| \le c(1+\psi) \\ &\Longleftrightarrow \quad \mu_n \longrightarrow \mu \text{ weakly and } \int \psi\, d\mu_n \longrightarrow \int \psi\, d\mu\end{aligned}$$

A suitable metric is

$$d_\psi(\mu,\nu) := d_{\mathrm{Proh}}(\mu,\nu) + \left| \int \psi\, d\mu - \int \psi\, d\nu \right|$$

The $\psi$-weak topology coincides with the weak topology iff $\psi$ is bounded.

# 2 Preliminaries

The choice $\mathscr{X} := L^\infty := L^\infty(\Omega, \mathscr{F}, \mathbb{P})$ is not suitable when dealing with possibly unbounded risks. Better: Orlicz spaces or Orlicz hearts (S. Biagini and Frittelli (2008), Cheridito and Li (2009)).

A Young function will be a left-continuous, nondecreasing convex function $\Psi : \mathbb{R}_+ \to [0, \infty]$ such that $0 = \Psi(0) = \lim_{x\downarrow 0} \Psi(x)$ and $\lim_{x\uparrow\infty} \Psi(x) = \infty$.

The Orlicz space associated with $\Psi$ is

$$L^\Psi := L^\Psi(\Omega, \mathscr{F}, \mathbb{P}) = \big\{X \in L^0 \,|\, \mathbb{E}[\,\Psi(c|X|)\,] < \infty \text{ for some } c > 0\big\}.$$

It is a Banach space if endowed with the Luxemburg norm,

$$\|X\|_\Psi := \inf\left\{\lambda > 0 \,|\, \mathbb{E}[\,\Psi(|X|/\lambda)\,] \leq 1\right\}.$$

The Orlicz heart is defined as

$$H^\Psi := H^\Psi(\Omega, \mathscr{F}, \mathbb{P}) = \big\{X \in L^0 \,|\, \mathbb{E}[\,\Psi(c|X|)\,] < \infty \textbf{ for all } c > 0\big\}$$

Cheridito and Li (2009): finite risk measures on $H^\Psi$ are continuous for $\|\cdot\|_\Psi$

For a finite Young function $\Psi$,

$$L^\infty \subset H^\Psi \subset L^\Psi \subset L^1$$

and these inclusions may all be strict. In fact, the identity $H^\Psi = L^\Psi$ holds if and only if $\Psi$ satisfies the so-called $\Delta_2$-condition

(3) there are $C$, $x_0 > 0$ such that $\Psi(2x) \leq C\Psi(x)$ for all $x \geq x_0$.

The $\Delta_2$-condition is clearly satisfied if specifically $\Psi(x) = x^p/p$ for some $p \in [1, \infty)$. In this case, $H^\Psi = L^\Psi = L^p$ and $\|Y\|_\Psi = p^{-1/p}\|Y\|_p$.

In the sequel, $\Psi$ will always denote a finite Young function

# 3 Consistency

For a distortion risk measure $\rho$, the estimator $\widehat{\rho}_n = \mathscr{R}_\rho(\widehat{m}_n)$ has the form of an *L-statistic* and results by van Zwet (1980), Gilat and Helmers (1997), and Tsukahara (2013) can be applied.

Our following result works for general law-invariant convex risk measures:

**Theorem 1.** *Suppose that $\rho$ is a law-invariant convex risk measure on $H^\Psi$ and $X_1, X_2, \ldots$ is a stationary and ergodic sequence of random variables with the same law as $X \in H^\Psi$. Then $\widehat{\rho}_n$ is a strongly consistent estimator in the sense that*

$$\widehat{\rho}_n = \mathscr{R}_\rho(\widehat{m}_n) = \mathscr{R}_\rho\Big(\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}\Big) \longrightarrow \rho(X) \qquad \mathbb{P}\text{-}a.s.$$

It follows from Birkhoff's ergodic theorem that, $\mathbb{P}$-a.s.,

$$\widehat{m}_n \longrightarrow \mu := \mathbb{P} \circ X^{-1} \qquad \Psi(|\cdot|)\text{-weakly}$$

So Theorem 1 would have followed if it was possible to establish the continuity of $\nu \mapsto \mathscr{R}_\rho(\nu)$ with respect to the $\Psi(|\cdot|)$-weak topology. But this is not possible unless $\Psi$ satisfies the $\Delta_2$-condition:

# 4 Continuity

If $\Psi$ is a Young function, then $\Psi(|\cdot|)$ is a weight function, and we will simply write $\mathscr{M}_1^{\Psi}$ in place of $\mathscr{M}_1^{\Psi(|\cdot|)}$. We will also use the term $\Psi$-weak convergence instead of $\Psi(|\cdot|)$-weak convergence etc. We recall the notation

$$\mathscr{M}(H^{\Psi}) = \{\mathbb{P} \circ X^{-1} \,|\, X \in H^{\Psi}\}$$

for the class of all laws of random variables $X \in H^{\Psi}$.

**Remark 1.** The identity $\mathscr{M}(H^{\Psi}) = \mathscr{M}_1^{\Psi}$ holds if and only if $\Psi$ satisfies the $\Delta_2$-condition (3).

**Theorem 2.** *For a finite Young function* $\Psi$ *the following conditions are equivalent.*

(a) *For every law-invariant convex risk measure* $\rho$ *on* $H^\Psi$*, the map* $\mathscr{R}_\rho : \mathscr{M}(H^\Psi) \to \mathbb{R}$ *is continuous for the* $\Psi$*-weak topology.*

(b) $\Psi$ *satisfies the* $\Delta_2$*-condition* (3)*.*

# 5 Qualitative and comparative robustness

$\Omega = \mathbb{R}^{\mathbb{N}}$, $X_i(\omega) = \omega(i)$ for $\omega \in \Omega$ and $i \in \mathbb{N}$, and $\mathcal{F} := \sigma(X_1, X_2, \dots)$. For any Borel probability measure $\mu$ on $\mathbb{R}$, we will denote

$$\mathbb{P}_\mu := \mu^{\otimes \mathbb{N}}$$

**Definition 1** (Qualitative robustness)**.** Suppose $\mathscr{N} \subset \mathscr{M}_1$ is a set, $d_A$ is a metric on $\mathscr{N}$, and $d_B$ is a metric on $\mathscr{M}_1$. Then $\mathscr{R}_\rho$ is called robust on $\mathscr{N}$ with respect to $d_A$ and $d_B$ if for all $\mu \in \mathscr{N}$ and $\varepsilon > 0$ there exists $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$(4) \qquad \nu \in \mathscr{N},\ d_A(\mu, \nu) \le \delta \quad \Longrightarrow \quad d_B\big(\mathbb{P}_\mu \circ \widehat{\rho}_n^{\,-1}, \mathbb{P}_\nu \circ \widehat{\rho}_n^{\,-1}\big) \le \varepsilon \quad \text{for } n \ge n_0.$$

Interpretation: $\nu$ “contamination” of “true” distribution $\mu$

In Hampel’s classical notion of qualitative robustness: $\mathscr{N} = \mathscr{M}_1$ and $d_A$ and $d_B$ = Lévy metric or Prohorov metric

Here we take for $d_A$:

$$d_\psi(\mu,\nu) \;=\; d_{\mathrm{Proh}}(\mu,\nu) + \Big|\int \psi\, d\mu - \int \psi\, d\nu\Big|$$

and

$$d_B = d_{\mathrm{Proh}}$$

**Definition 2.** Let $\psi$ be a weight function. A set $\mathscr{N} \subset \mathscr{M}_1^\psi$ is called uniformly $\psi$-integrating if

$$\lim_{M\to\infty} \sup_{\nu\in\mathscr{N}} \int_{\{\psi\geq M\}} \psi\, d\nu \;=\; 0. \tag{5}$$

If $\psi$ is bounded, every set $\mathscr{N} \subset \mathscr{M}_1$ is uniformly $\psi$-integrating.

First attempt to extend the notion of robustness:

**Definition 3.** Let $\psi$ be a weight function and $\mathscr{M} \subset \mathscr{M}_1^{\psi}$. The risk functional $\mathscr{R}_\rho$ is called $\psi$-robust on $\mathscr{M}$ if $\mathscr{R}_\rho$ is robust with respect to $d_\psi$ and $d_{\mathrm{Proh}}$ on every uniformly $\psi$-integrating set $\mathscr{N} \subset \mathscr{M}$.

**Theorem 3.** *For a finite Young function $\Psi$, the following conditions are equivalent.*

(a) *For every law-invariant convex risk measure $\rho$ on $H^\Psi$, $\mathscr{R}_\rho$ is $\Psi$-robust on $\mathscr{M}(H^\Psi)$.*

(b) *$\Psi$ satisfies the $\Delta_2$-condition (3).*

# 6 Questions, criticism, and improvement

(A) The theory presented so far only applies if the Young function $\Psi$ satisfies the $\Delta_2$-condition. What happens, e.g., in the case of the entropic risk measure

$$\rho(X) = \frac{1}{\beta} \log \mathbb{E}[\, e^{-\beta X}\, ]$$

for which we could use $\Psi(x) = e^x - 1$, which does not satisfy the $\Delta_2$-condition?

This issue could be addressed by using instead a Young function $\widetilde{\Psi}$ for which $\widetilde{\Psi}(x)/\Psi(x) \to 0$ as $x \uparrow \infty$. But is there perhaps an optimal choice?

(B) The use of the metric $d_\psi$ in our definition of $\psi$-robustness was criticized by some statisticians who feel that the Prohorov metric is more natural than our metric $d_\psi$.

Both issues addressed in Krätschmer, A.S., and Zähle (2017).

As for issue (A), for a general Young function $\Psi$, the convergence $\|X_n - X\|_\Psi \to 0$ is equivalent to

$$\mathbb{E}[\,\Psi(k|X_n - X|)\,] \longrightarrow 0 \qquad \text{for all } k \in \mathbb{N}.$$

This fact suggests not to work with the single weight function $\psi(x) = \Psi(|x|)$ but with the countable collection

$$\psi_k(x) = \Psi(k|x|), \qquad k \in \mathbb{N}$$

Thus, for any sequence $(\psi_k)$ of weight functions, we define

$$\mathscr{M}_1^{(\psi_k)} := \mathscr{M}_1^{(\psi_k)}(\mathbb{R}) := \left\{ \mu \in \mathscr{M}_1(\mathbb{R}) \,\middle|\, \int \psi_k \, d\mu < \infty \text{ for all } k \right\}$$

and

$$d_{(\psi_k)}(\mu, \nu) \;:=\; d_{\mathrm{Proh}}(\mu, \nu) + \sum_{k=1}^{\infty} 2^{-k} \left| \int \psi_k \, d\mu - \int \psi_k \, d\nu \right| \wedge 1$$

which gives rise to the notion of $(\psi_k)$-weak convergence.

The results on the continuity and robustness of risk functionals extend without much difficulty to this more general setup, and they allow us to drop the $\Delta_2$-condition.

As for issue (B): Recall

**Definition 3:** Let $\psi$ be a weight function and $\mathscr{M} \subset \mathscr{M}_1^\psi$. The risk functional $\mathscr{R}_\rho$ is called $\psi$-robust on $\mathscr{M}$ if $\mathscr{R}_\rho$ is robust with respect to $d_\psi$ and $d_{\mathrm{Proh}}$ on every uniformly $\psi$-integrating set $\mathscr{N} \subset \mathscr{M}$.

This definition

- extends immediately to $(\psi_k)$
- requires the stronger metric $d_\psi$ or $(d_{(\psi_k)})$
- requires the restriction of "contaminatined" probabilities to uniformly $\psi$-integrating sets $\mathscr{N}$

But, actually, the stronger metric is not needed:

**Theorem 4.** *The following conditions are equivalent for a set $\mathscr{N} \subseteq \mathscr{M}_1^{(\psi_k)}$:*

(a) *The standard weak topology and the $(\psi_k)$-weak topologies coincide on $\mathscr{N}$.*

(b) *$\mathscr{N}$ is locally uniformly $(\psi_k)$-integrating: For every $\mu \in \mathscr{N}$, $\varepsilon > 0$, and $k \in \mathbb{N}$ there exist $c > 0$ and a weakly open neighborhood $U$ of $\mu$ such that*

$$\nu \in \mathscr{N} \cap U \quad \Longrightarrow \quad \int \psi_k \mathrm{I}_{\{\psi_k \geq c\}} \, d\nu \leq \varepsilon.$$

Call such sets $\mathscr{N}$ from now on w-sets.

**Definition 4.** The risk functional $\mathscr{R}_\rho$ is called $(\psi_k)$-robust on $\mathscr{M}$ if $\mathscr{R}_\rho$ is robust with respect to $d_{\mathrm{Proh}}$ and $d_{\mathrm{Proh}}$ on every w-set $\mathscr{N} \subset \mathscr{M}$.

## Examples of w-sets

- The class of all $d$-dimensional normal distributions $N(m, \Sigma)$ is a w-set for

$$\psi_k(x) = \exp(\lambda_k |x|^{\alpha_k}), \qquad \text{where } \lambda_k \uparrow \infty \text{ and } \alpha_k \uparrow 2$$

- The class of all Gamma distributions is a w-set for $\psi_k(x) = |x|^k$ or

$$\psi_k(x) = \exp(\lambda_k x^{\beta_k}), \qquad \text{where } \lambda_k \uparrow \infty \text{ and } \beta_k \uparrow 1$$

- Many other parametric classes of distributions (Pareto, Gumbel etc)
- Images of Fréchet classes under certain contractions, which yields aggregation robustness in the sense of Embrechts, Wang & Wang (2015)

**Theorem 5.** *Let*

$$\rho : H^{\Psi} \longrightarrow \mathbb{R}$$

*be a law-invariant risk measure and*

$$\psi_k(x) := \Psi(k|x|).$$

*Then*

(a) $\mathscr{R}_\rho$ *is* $(\psi_k)$*-weakly continuous on* $\mathscr{M}_1^{(\psi_k)}$

(b) $\mathscr{R}_\rho$ *is* $(\psi_k)$*-robust on* $\mathscr{M}_1^{(\psi_k)}$

# 7 Other statistical functionals

Consider a general statistical functional $T$ that gives rise to the estimator

$$\widehat{T}_n = T(m_n)$$

Examples:

- mean, variance,...
- maximum likelihood estimators
- ...

Robustness, continuity etc. can be defined as above.

**Theorem 6. (Hampel-type theorem)**
*Suppose that* $T : \mathscr{M} \to \mathbb{R}$, *where* $\mathscr{M} \subset \mathscr{M}_1^{(\psi_k)}$. *Then:*

(i) *If* $T$ *is* $(\psi_k)$*-weakly continuous on* $\mathscr{M}$, *then it is* $(\psi_k)$*-robust on every w-set* $M \subseteq \mathcal{M}$.

(ii) *If* $(\widehat{T}_n)$ *is weakly consistent under every* $\mathbb{P}_\mu$ *with* $\mu \in \mathscr{M}$ *and robust on every w-set* $M \subseteq \mathscr{M}$, *then* $T$ *is* $(\psi_k)$*-weakly continuous on* $\mathscr{M}$.

## Special case: Maximum likelihood estimators

Parametric family $\mathscr{M}_\Theta$ of distinct laws $\mu_\theta$ with density $f_\theta$ for $\theta \in \Theta$ with $\Theta \subset \mathbb{R}^d$ open and convex

Statistical functional $T : \mathscr{M} \to \Theta$ such that

$$T(\mu) \in \arg\max_{\theta \in \Theta} \int \log f_\theta \, d\mu$$

Define a sequence of weight functions by taking $\{\theta_1, \theta_2, \ldots\} = \Theta \cap \mathbb{Q}^d$ and

$$\psi_k(x) := |\log f_{\theta k}(x)|, \qquad k = 1, 2, \ldots$$

**Theorem 7.** *Suppose that*

(a) $f_\theta(x) > 0$ *for all* $x$ *and all* $\theta \in \Theta$

(b) $x \mapsto \log f_\theta(x)$ *is continuous for all* $\theta \in \Theta$

(c) $\theta \mapsto \log f_\theta(x)$ *is concave for all* $x$

If $\mu \in \mathscr{M}$ is such that $\theta \mapsto \int \log f_\theta, d\mu$ has a unique maximizer, then $T$ is continuous at $\mu$ for the $(\psi_k)$-weak topology.

**Example: Gumbel distributions**

For $\theta \in \Theta := (0, \infty)$ let $\mu_\theta$ be the Gumbel distribution with location parameter 0 and scale parameter $1/\theta$, which has Lebesgue density

$$f_\theta(x) = \theta e^{-\theta x - e^{-\theta x}}, \qquad x \in \mathbb{R}$$

Then conditions (a), (b), (c) hold. One can show that

$$f_\theta(x)\,dx \in \mathscr{M}_1^{(\psi_k)} \qquad \text{for all } \theta$$

Let

$$\mathscr{M} := \mathscr{M}_1^{(\psi_k)} \setminus \{\delta_0\}$$

For every $\mu \in \mathscr{M}$, the map

$$\mathscr{L}_\mu(\theta) := \int f_\theta \, d\mu = \int (\log\theta - \theta x - e^{-\theta x}) \, \mu(dx)$$

has a unique maximizer. This follows from the strict concavity of $\mathscr{L}_\mu$ and the easily obtained fact that

$$\lim_{\theta \downarrow 0} \mathscr{L}_\mu(\theta) = -\infty \quad \text{and} \quad \lim_{\theta \uparrow \infty} \mathscr{L}_\mu(\theta) = -\infty$$

We thus obtain:

**Proposition 1.** *The unique maximum likelihood functional is $(\psi_k)$-weakly continuous and hence $(\psi_k)$-robust on $\mathscr{M}$.*

**Example: Exponential distributions**

$\mu_\theta =$ exponential distribution with parameter $\theta \in \Theta = (0, \infty)$

$$\log f_\theta = -\log\theta - \frac{x}{\theta}, \qquad x > 0$$

So the $(\psi_k)$-weak topology is equivalent to the $\psi$-weak topology for

$$\psi(x) = x, \qquad x > 0$$

Next, for $\mu \in \mathscr{M}_1^\psi$,

$$\theta \longmapsto \int \log f_\theta \, d\mu = -\log\theta - \frac{1}{\theta}\int x\,\mu(dx)$$

has a unique maximizer, namely $\theta = \int x\,\mu(dx)$, and so the maximum likelihood functional is

$$T(\mu) = \int x\,\mu(dx)$$

# References

Cont, R., Deguest, R. & Scandolo, G. (2010), 'Robustness and sensitivity analysis of risk measurement procedures', *Quant. Finance* **10**(6), 593–606.

Embrechts, P., Wang, B. & Wang, R. (2015), 'Aggregation-robustness and model uncertainty of regulatory risk measures', *Finance and Stochastics* **19**(4), 763–790.

Filipović, D. & Svindland, G. (2012), 'The canonical model space for law-invariant convex risk measures is $L^1$', *Math. Finance* **22**(3), 585–589.

Hampel, F. R. (1971), 'A general qualitative definition of robustness', *Ann. Math. Statist.* **42**, 1887–1896.

Krätschmer, V., Schied, A. & Zähle, H. (2012), 'Qualitative and infinitesimal robustness of tail-dependent statistical functionals', *J. Multivariate Anal.* **103**, 35–47.

Krätschmer, V., Schied, A. & Zähle, H. (2014), 'Comparative and qualitative robustness for law-invariant risk measures', *Finance and Stochastics* **18**, 271–295.

Krätschmer, V., Schied, A. & Zähle, H. (2015), 'Quasi-Hadamard differentiability of general risk functionals and its application', *Stat. Risk Model.* **32**(1), 25–47.

Krätschmer, V., Schied, A. & Zähle, H. (2017), 'Domains of weak continuity of statistical functionals with a view toward robust statistics', *Journal of Multivariate Analysis* **158**, 1–19.

Mizera, I. (2010), Qualitative robustness and weak continuity: the extreme unction?, *in* 'Nonparametrics and robustness in modern statistical inference and time series analysis: a Festschrift in honor of Professor Jana Jurečková', Vol. 7 of *Inst. Math. Stat. Collect.*, Inst. Math. Statist., Beachwood, OH, pp. 169–181.