Qualitative and infinitesimal robustness of tail-dependent statistical functionals

Volker Krätschmer [a] Alexander Schied [b]
Henryk Zähle [c,*]

[a] *University of Duisburg-Essen, Faculty of Mathematics, Forsthausweg 2, D–47057 Duisburg, Germany*

[b] *University of Mannheim, Department of Mathematics, A5, 6, D–68131 Mannheim, Germany*

[c] *Saarland University, Department of Mathematics, Postfach 151150, D–66041 Saarbrücken, Germany*

Abstract

The main goal of this article is to introduce a new notion of qualitative robustness that applies also to tail-dependent statistical functionals and that allows us to compare statistical functionals in regards to their degree of robustness. By means of new versions of the celebrated Hampel theorem, we show that this degree of robustness can be characterized in terms of certain continuity properties of the statistical functional. The proofs of these results rely on strong uniform Glivenko-Cantelli theorems in fine topologies, which are of independent interest. We also investigate the sensitivity of tail-dependent statistical functionals w.r.t. infinitesimal contaminations, and we introduce a new notion of infinitesimal robustness. The theoretical results are illustrated by means of several examples including general L- and V-functionals.

Key words: qualitative robustness, Hampel's theorem, uniform Glivenko-Cantelli theorem, weighted Kolmogorov metric, ψ-weak topology, generalized Birnbaum-Marshall inequality, infinitesimal robustness, quasi-Hadamard differentiability, L- and V-functionals

---

Preprint submitted to Journal of Multivariate Analysis *June 22, 2011*

1 Introduction

Let $\mathcal{M}$ be a class of probability measures on $\mathbb{R}^d$, and let T be a statistical functional from $\mathcal{M}$ into a measurable space $\mathbf{T}$. Next, let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with common distribution $\mu \in \mathcal{M}$. If $\hat{m}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ denotes the empirical distribution of $X_1, \ldots, X_n$, then $T(\hat{m}_n)$ can provide a reasonable estimator for $T(\mu)$. Apart from issues such as consistency and error asymptotics, a central question concerns the *qualitative robustness* of the estimation against contaminations of the underlying model μ. Informally, qualitative robustness holds when for large n a small change in μ results only in a small change of the law of the estimator $T(\hat{m}_n)$. More precisely, a statistical functional T is said to be qualitatively robust at μ if for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for every $\nu \in \mathcal{M}$ and every $n \geq n_0$

$$d(\mu,\nu) \leq \delta \qquad \Longrightarrow \qquad d'(\mathrm{law}\{T(\hat{m}_n)|\mu\}, \mathrm{law}\{T(\hat{m}_n)|\nu\}) \leq \varepsilon, \tag{1}$$

where d and d' are probability metrics on $\mathcal{M}$ and on the class of all probability measures on $\mathbf{T}$, respectively.

In the classical literature on qualitative robustness [7,15,16,17,21,23], the distances d and d' in (1) are typically chosen so that they generate the respective weak topologies of measures. For instance, d and d' are often taken as the respective Prohorov metrics. In dimension 1 the Lévy distance is also a common choice for d. It is a consequence of Hampel's celebrated theorem that such a choice essentially limits the concept of qualitative robustness to functionals T that are continuous for the weak topology at μ; see, e.g., [17, Theorem 2.21]. Statistical functionals such as the mean or the (co-)variance, which are tail-dependent and hence not weakly continuous, are consequently classified as nonrobust.

This division of the class of all statistical functionals into qualitatively robust and nonrobust ones appears to be somewhat coarse. For instance, it is intu-

* Corresponding author
Email addresses: `volker.kraetschmer@uni-due.de` (Volker Krätschmer), `schied@uni-mannheim.de` (Alexander Schied), `zaehle@math.uni-sb.de` (Henryk Zähle).

itively clear that a statistical functional such as the mean should be deemed more robust than a statistical functional like the (co-)variance, which involves higher moments of the underlying distribution.

The main purpose of this article is to introduce a new concept of qualitative robustness that applies to a very large class of tail-dependent statistical functionals T. The key to our approach lies in specifying a metric d on the left-hand side of (1) for which T becomes a continuous functional at μ. This can for instance be a weighted Kolmogorov-type distance or the sum of the Prohorov metric and a moment distance. Then we establish novel extensions of Hampel's theorem essentially stating that when T is continuous with respect to d then it is also *qualitatively robust* in the sense that (1) holds if we choose the Prohorov metric for d'.

In many situations, our results allow for a direct comparison of two statistical functionals T_0 and T_1 in regards to their degree of robustness, a concept that has not been possible within the classical framework of robust statistics. In our setup, T_0 will have a higher degree of robustness than T_1 when T_0 is qualitatively robust for any choice of d for which T_1 is qualitatively robust. It will follow from our extensions of Hampel's theorem that such a comparison of the degree of robustness reduces to the much simpler analysis of the continuity properties of T_0 and T_1. In the simple case of the mean and the (co-)variance, we thus see immediately that the latter has a lower degree of robustness than the former, which in turn has a lower degree of robustness than the median. Less obvious and more interesting examples will be discussed in Section 4.

Our specific extensions of Hampel's theorem derive from a general extension that, besides the continuity of T, requires to limit the probability measure ν in (1) to a subset $\mathcal{M}_0 \subset \mathcal{M}$ that satisfies what we call the *UGC property* with respect to the metric d. This UGC property refers to the validity of a uniform Glivenko-Cantelli theorem with respect to d for i.i.d. random variables with distribution $\mu \in \mathcal{M}_0$. Establishing the UGC property of a specific choice $(\mathcal{M}_0, d)$ is thus the key ingredient of our robustness analysis. In Section 3, the UGC property is first established for a weighted version of the Kolmogorov metric. By comparison arguments, the UGC property is then extended to the L^1-Wasserstein metric and to probability distances that arise as sums of the Prohorov metric and moment distances and may even be defined on a general

Polish space. These results are not only at the core of our robustness analysis but also of independent interest.

We expect that our results will have several possible application, one for instance in the ongoing discussion on the appropriate choice of a financial risk measure. In finance, risk measures are used to determine the economic capital required for holding a risky position; see, e.g., [13]. The industry standard, Value at Risk, is a simple quantile functional, which has several drawbacks. These drawbacks have led to the introduction of the alternative concept of a coherent risk measure [1]. Recently, however, it was stressed in Cont et al. [6] that Value at Risk is qualitatively robust in the sense of Hampel, while most coherent alternatives are not. Since nonrobustness can supposedly lead to a frequent and expensive readjustment of the economic capital, the results from [6] seem to weigh heavily in favor of Value at Risk. Our refined notion of qualitative robustness, however, might bring this argument back into perspective since it is now clear that robustness is not lost entirely but only to some degree when Value at Risk is replaced by a coherent risk measure such as Average Value at Risk (also called Expected Shortfall). In fact, our results also show that Average Value at Risk is still more robust than the mean-standard deviation risk measure, which is also commonly used in practice.

Up to this point, we have discussed *qualitative* robustness. A related issue is *infinitesimal* robustness. There are different definitions of infinitesimal robustness in the literature, all relying on a special type of *differentiability* of the functional T (in contrast to qualitative robustness which is related to *continuity* of T through Hampel's theorem). For a discussion see, e.g., [16,18]. Typically, the mode of infinitesimal robustness involves only the directional derivative of T at μ in direction $\delta_x - \mu$, $x \in \mathbb{R}^d$, and this directional derivative, regarded as a function of x, the so-called influence function, is used to quantify the sensitivity of T at μ w.r.t. a contamination caused by an outlier x. Sometimes, however, a stronger notion of differentiability as Gâteaux or Hadamard differentiability is used. In this case one can also measure the sensitivity of T w.r.t. a general model contamination. On the other hand, this entails a similar problem as in the case of qualitative robustness. In fact, while most of the popular tail-*in*dependent statistical functionals are known to be Hadamard differentiable, it is commonly acknowledged that tail-dependent functionals are typically *not* Hadamard differentiable w.r.t. uniform norms; see also [3].

That is, we again have a somewhat coarse division of the class of all statistical functionals into robust and nonrobust ones.

However, recently in [3] the notion of *quasi*-Hadamard differentiability was introduced, which ensures that also tail-dependent functionals can be Hadamard differentiable in a suitably modified sense. So, in Section 5 we will propose quasi-Hadamard differentiability as an alternative concept for infinitesimal robustness, and we will give some illustrating examples.

2 Qualitative robustness and a Hampel theorem

Let $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\mathbb{N}}, \mu^{\mathbb{N}} : \mu \in \mathcal{M})$ be a statistical product model for i.i.d. observations on $\mathbb{R}^d$, where $\mathcal{M}$ is some subset of the set $\mathcal{M}_1(\mathbb{R}^d)$ of all Borel probability measures on $\mathbb{R}^d$. Further, let $(\mathbf{T}, \mathcal{T})$ be a measurable space, and $T : \mathcal{M} \to \mathbf{T}$ be a mapping. Since $\mathcal{M}$ is a class of probability measures, T is also called *statistical functional*. For every $n \in \mathbb{N}$, we assume that the mapping

$$\hat{T}_n(x_1, x_2, \ldots) = \hat{T}_n(x^{(n)}) := T(\hat{m}_{x^{(n)}}), \qquad (x_1, x_2, \ldots) \in (\mathbb{R}^d)^{\mathbb{N}} \tag{2}$$

is $(\mathcal{B}(\mathbb{R}^d)^{\mathbb{N}}, \mathcal{T})$-measurable, where $\hat{m}_{x^{(n)}} := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ denotes the empirical probability measures associated with $x^{(n)} := (x_1, \ldots, x_n)$. For (2) to be well defined, we assume that the set of all such empirical probability measures is contained in $\mathcal{M}$. Notice that $\hat{T}_n$ provides an estimator for $T(\mu)$. We let d' be some metric on the set $\mathcal{M}_1(\mathbf{T})$ of all probability measures on $(\mathbf{T}, \mathcal{T})$, and d be some metric on $\mathcal{M}$.

Definition 2.1 (Qualitative $\mathcal{M}_0$-robustness) *Let $\mathcal{M}_0$ be some subset of $\mathcal{M}$, and let $\mu \in \mathcal{M}_0$. Then the sequence $(\hat{T}_n)$ of estimators is said to be* qualitatively $\mathcal{M}_0$-robust *at μ w.r.t. (d, d') if for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for all $\nu \in \mathcal{M}_0$ and $n \geq n_0$*

$$d(\mu, \nu) \leq \delta \qquad \Longrightarrow \qquad d'(\mu^{\mathbb{N}} \circ \hat{T}_n^{-1}, \nu^{\mathbb{N}} \circ \hat{T}_n^{-1}) \leq \varepsilon.$$

If in addition $(\hat{T}_n)$ arises as in (2) from a statistical functional T, then T is called qualitatively $\mathcal{M}_0$-robust *at μ w.r.t. (d, d').*

If $\mathcal{M} = \mathcal{M}_0 = \mathcal{M}_1(\mathbb{R})$, then qualitative $\mathcal{M}_0$-robustness coincides with Hampel's notion of qualitative robustness as defined in [17, p. 41]. In the current literature on robust statistics, the metrics d and d' are usually chosen in such a way that they generate the respective weak topologies. Nevertheless, we emphasize that the concept of qualitative robustness depends on the specific choice of the metrics d and d' and not just on the topologies generated by them; see also [17, p. 42].

In applications, the validity of qualitative $\mathcal{M}_0$-robustness in the sense of Definition 2.1 is typically hard to check "directly". So it is natural to ask for transparent sufficient conditions. The celebrated Hampel theorem (see, e.g., [17, Theorem 2.21]) provides a sufficient condition for qualitative robustness when $\mathcal{M} = \mathcal{M}_0 = \mathcal{M}_1(\mathbb{R})$, d is the Lévy metric, and d' is the Prohorov metric. Here we are interested in extending Hampel's theorem to tail-dependent functionals T such as the mean or the variance. Such functionals may not be defined on all of $\mathcal{M}_1(\mathbb{R}^d)$ and they are typically not continuous w.r.t. the weak topology. One therefore has to think of $\mathcal{M}_0$ as a subset of the domain $\mathcal{M}$ of T and of d as a metric that generates a *finer* topology than the weak topology. Several possible choices for d will be discussed in Section 3. For d', we will retain the classical choice of the Prohorov metric. To this end, we assume that $\mathbf{T}$ is equipped with a complete and separable metric $d_{\mathbf{T}}$ and that $\mathcal{T}$ is the corresponding Borel σ-field. Recall that the Prohorov distance is given by

$$d'_{\mathrm{Proh}}(\mu,\nu) := \inf\{\varepsilon > 0 : \mu(A) \le \nu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{T}\},$$

where $A^\varepsilon := \{t \in T : \inf_{a\in A} d_{\mathbf{T}}(t,a) \le \varepsilon\}$ is the ε-hull of A.

Definition 2.2 ($\mathcal{N}$-continuity) *Let $\mu \in \mathcal{M}$, and $\mathcal{N}$ be some subset of $\mathcal{M}$. Then T is called* $\mathcal{N}$-continuous *at μ w.r.t. $(d, d_{\mathbf{T}})$ if for every $\varepsilon > 0$ there is some $\delta > 0$ such that for all $\nu \in \mathcal{N}$*

$$d(\mu,\nu) \le \delta \qquad \Longrightarrow \qquad d_{\mathbf{T}}(T(\mu), T(\nu)) \le \varepsilon.$$

Definition 2.3 (UGC property) *Let $\mathcal{M}_0$ be some subset of $\mathcal{M}$. Then we say that the metric space $(\mathcal{M}_0, d)$* has the UGC property *if one can find for every $\varepsilon > 0$ and $\delta > 0$ some $n_0 \in \mathbb{N}$ such that for all $\mu \in \mathcal{M}_0$ and $n \ge n_0$*

$$\mu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} : d(\mu, \hat{m}_{x^{(n)}}) \ge \delta\Big\}\Big] \le \varepsilon. \tag{3}$$

Here the acronym UGC stands for "uniform Glivenko-Cantelli". Examples for spaces having the UGC property will be given in Section 3. The following result extends Hampel's theorem and we will see in Sections 3 and 4 that it applies to a wide range of statistical estimators, including the sample mean or the empirical (co-)variance, when the metric d is chosen appropriately. Let $\mathcal{E}$ be the space of all empirical probability measures $\hat{m}_{x(n)}$ with $x \in (\mathbb{R}^d)^{\mathbb{N}}$ and $n \in \mathbb{N}$, and recall that we assumed $\mathcal{E} \subset \mathcal{M}$.

Theorem 2.4 (Hampel's theorem) *Let $\mathcal{M}_0$ be some subset of $\mathcal{M}$, and let $\mu \in \mathcal{M}_0$. Further assume that $(\mathcal{M}_0, d)$ has the UGC property. Then, if the mapping T is $\mathcal{E}$-continuous at μ w.r.t. $(d, d_{\mathbf{T}})$, the sequence $(\hat{T}_n)$ is qualitatively $\mathcal{M}_0$-robust at μ w.r.t. (d, d'_{Proh}).*

The proof of this theorem can be found in Section A.1. The following Theorem 2.6 generalizes the converse of Hampel's theorem as given in [17, Theorem 2.21].

Definition 2.5 (Weak consistency) *Let $\mu \in \mathcal{M}$. The sequence $(\hat{T}_n)$ of estimators is said to be* weakly consistent *at μ w.r.t. $d_{\mathbf{T}}$ if $(\hat{T}_n)$ converges in $\mu^{\mathbb{N}}$-probability to $T(\mu)$, i.e., if for every $\delta > 0$*

$$\lim_{n\to\infty} \mu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} : d_{\mathbf{T}}(\hat{T}_n(x), T(\mu)) \geq \delta\Big\}\Big] = 0.$$

Theorem 2.6 (Converse of Hampel's theorem) *Let $\mathcal{M}_0$ be some subset of $\mathcal{M}$, and let $\mu \in \mathcal{M}_0$. Further assume that $(\hat{T}_n)$ is weakly consistent w.r.t. $d_{\mathbf{T}}$ at every element ν of some neighborhood of μ w.r.t. d. Then, if the sequence $(\hat{T}_n)$ is qualitatively $\mathcal{M}_0$-robust at μ w.r.t. (d, d'_{Proh}), the mapping T is $\mathcal{M}_0$-continuous at μ w.r.t. $(d, d_{\mathbf{T}})$.*

The proof of this theorem can be found in Section A.2.

3 Examples for metrics with the UGC property

In this section, we present several choices for $\mathcal{M}$, $\mathcal{M}_0$, and d for which the UGC property can be verified. By Theorem 2.4, any statistical functional T that is $\mathcal{E}$-continuous w.r.t. $(d, d_{\mathbf{T}})$ will then be qualitatively $\mathcal{M}_0$-robust.

3.1 Known results

One of the most general versions of Hampel's theorem in the current literature is due to Mizera [21]. It works in the situation in which $\mathcal{M}_0 = \mathcal{M}$ is the set of all Borel probability measures on a Polish space and d is the corresponding Prohorov metric d_{Proh}. As ingredient in the proof, the UGC property is established for d_{Proh} (see [21, Lemma 4]). A similar statement involving slightly stronger assumptions was obtained earlier by Cuevas [7, Theorem 2]. *Dudley's Lipschitz metric* is defined as

$$d_{\mathrm{Lip}}(\mu,\nu) \;:=\; \sup\Big\{\int f\,d\mu - \int f\,d\nu \,:\, \|f\|_\infty + \mathrm{Lip}(f) \le 1\Big\},$$

where $\|f\|_\infty$ denotes the usual sup-norm and $\mathrm{Lip}(f)$ is the Lipschitz constant of the Lipschitz function f. Since $\frac{2}{3}d_{\mathrm{Proh}}(\mu,\nu)^2 \le d_{\mathrm{Lip}}(\mu,\nu) \le 2d_{\mathrm{Proh}}(\mu,\nu)$ (see, e.g., Eq. (4) in [21]), the UGC property and a Hampel theorem for d_{Lip} follow.

For one-dimensional observations and $\mathcal{M} = \mathcal{M}_0 = \mathcal{M}_1(\mathbb{R})$, it is an immediate consequence of the Dvoretsky-Kiefer-Wolfowitz inequality (see, e.g., [11,20] or [30, p. 268]) that the UGC property holds w.r.t. the *Kolmogorov metric*

$$d_{\mathrm{Kolm}}(\mu,\nu) \;:=\; \|F_\mu - F_\nu\|_\infty.$$

Here, F_μ denotes the distribution function (or: df) of $\mu \in \mathcal{M}_1(\mathbb{R})$. Note that the Kolmogorov metric does not generate the weak topology but that it dominates the *Lévy distance*,

$$d_{\mathrm{Lévy}}(\mu,\nu) \;:=\; \inf\{\varepsilon > 0 \,:\, F_\mu(x) \le F_\nu(x+\varepsilon) + \varepsilon \text{ for all } x \in \mathbb{R}\}, \tag{4}$$

which does (cf. [17, p. 36]). Qualitative robustness of continuous statistical functionals T w.r.t. $d_{\mathrm{Lévy}}$ is the content of the classical Hampel theorem as in [17, Theorem 2.21] and [15, Theorem 1].

Here we present a uniform Glivenko-Cantelli theorem w.r.t. a *weighted* version of the Kolmogorov metric. This metric is based on distribution functions and hence requires a univariate setting. But, as we will see in the subsequent sections, the corresponding Glivenko-Cantelli theorem will also yield multivariate corollaries via a comparison argument.

Let ϕ be a *u-shaped function*, i.e., a continuous function $\phi : \mathbb{R} \to [1, \infty)$ that is nonincreasing on $(-\infty, 0)$ and nondecreasing on $(0, \infty)$. Define $\mathcal{M}_1^{(\phi)}(\mathbb{R})$ as the set of all $\mu \in \mathcal{M}_1(\mathbb{R})$ for which $\sup_{x\le 0} |F_\mu(x)\phi(x)| + \sup_{x>0} |(1-F_\mu(x))\phi(x)| < \infty$. Denoting by $\|f\|_\phi := \sup_{x\in\mathbb{R}} |f(x)\phi(x)|$ the sup-norm of a function f weighted by ϕ, we can define a metric on $\mathcal{M}_1^{(\phi)}(\mathbb{R})$ by

$$d_{(\phi)}(\mu, \nu) \;:=\; \|F_\mu - F_\nu\|_\phi, \qquad \mu, \nu \in \mathcal{M}_1^{(\phi)}(\mathbb{R}). \tag{5}$$

This metric is a weighted version of the Kolmogorov metric, which corresponds to the special case $\phi := \mathbb{1}$. The following theorem shows that, for all $p > 1$ and $\kappa > 0$, the metric space $(\mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R}), d_{(\phi)})$ has the UGC property, where $\mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R})$ is the class of all $\mu \in \mathcal{M}_1(\mathbb{R})$ satisfying $\int \phi^p \, d\mu \le \kappa$. Note that $\mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R}) \subset \mathcal{M}_1^{(\phi)}(\mathbb{R})$ if $p \ge 1$.

Theorem 3.1 (Uniform Glivenko-Cantelli w.r.t. $d_{(\phi)}$) *Let ϕ be a u-shaped function and fix $\kappa > 0$. Then for all $p > 1$, $\delta > 0$ and $\varepsilon > 0$, there is some $n_0 \in \mathbb{N}$ such that for all $\mu \in \mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R})$ and $n \ge n_0$*

$$\mathbb{P}[d_{(\phi)}(\hat{m}_{n;\mu}, \mu) \ge \delta] \;\le\; \varepsilon,$$

where $\hat{m}_{n;\mu} = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical probability measure at stage n of an i.i.d. sequence of random variables (on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$) with distribution μ. The choice of n_0 is independent of the choice of $(\Omega, \mathcal{F}, \mathbb{P})$ and (X_i).

The proof of this theorem can be found in Section A.3. In the case $\phi = \mathbb{1}$ and $\kappa = 1$, we have $\mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R}) = \mathcal{M}_1(\mathbb{R})$. Thus, since the Kolmogorov metric $d_{(\mathbb{1})}$ dominates the Lévy distance $d_{\text{Lévy}}$, Theorem 3.1 shows in particular that $(\mathcal{M}_1(\mathbb{R}), d_{\text{Lévy}})$ has the UGC property. This was already seen in Section 3.1.

By combining Theorem 2.4 and Theorem 3.1 we get the following extension of Hampel's theorem.

Corollary 3.2 *Let ϕ be any u-shaped function, and suppose that T is $\mathcal{E}$-continuous w.r.t. $d_{(\phi)}$ at some μ belonging to $\mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R})$ for some $p > 1$ and $\kappa > 0$. Then the sequence $(\hat{T}_n)$ is qualitatively $\mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R})$-robust at μ w.r.t. $(d_{(\phi)}, d'_{\mathrm{Proh}})$.*

3.3 The L^1-Wasserstein metric

For every $\lambda \geq 0$, define a u-shaped function ϕ_λ by $\phi_\lambda(x) := (1 + |x|)^\lambda$, $x \in \mathbb{R}$. The L^1-*Wasserstein metric* on $\mathcal{M}_1^{\phi_1}(\mathbb{R}) := \{\mu \in \mathcal{M}_1 : \int \phi_1 \, d\mu < \infty\}$ is given by

$$d_{W_1}(\mu, \nu) := \int |F_\mu(x) - F_\nu(x)| \, dx, \qquad \mu, \nu \in \mathcal{M}_1^{\phi_1}(\mathbb{R}). \tag{6}$$

Many alternative representations of d_{W_1} are known; see, e.g., [10]. Clearly, for every $\kappa > 0$ and $\lambda > 1$ the L^1-Wasserstein metric on $\mathcal{M}_{1,\kappa}^{\phi_\lambda}(\mathbb{R})$ is dominated by a multiple of the metric $d_{(\phi_\lambda)}$. Thus, an application of Theorem 3.1 (with some $\lambda' \in (1, \lambda)$) yields:

Corollary 3.3 (Uniform Glivenko-Cantelli w.r.t. d_{W_1}) *For all $\kappa > 0$ and $\lambda > 1$, the space $(\mathcal{M}_{1,\kappa}^{\phi_\lambda}(\mathbb{R}), d_{W_1})$ has the UGC property.*

Notice that Theorem 11.1.6 in [22] and Proposition 3.4 in [9] together imply a Glivenko-Cantelli result but not a UGC result for d_{W_1}. Further notice that Theorem 2.4 and Corollary 3.3 together yield a version of Hampel's theorem similar to Corollary 3.2.

3.4 A UGC metric for the ψ-weak topology

We now continue in a multivariate setting. A *weight function* will be a continuous function $\psi : \mathbb{R}^d \to [1, \infty)$, and we denote by $\mathcal{M}_1^\psi(\mathbb{R}^d)$ the set of all $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ such that $\int \psi \, d\mu < \infty$. Furthermore, $C_\psi(\mathbb{R}^d)$ is the space of all continuous functions on $\mathbb{R}^d$ for which $\|f\|_\psi := \|f/\psi\|_\infty < \infty$. The ψ-*weak topology* on $\mathcal{M}_1^\psi(\mathbb{R}^d)$ is the coarsest topology for which all mappings $\mu \mapsto \int f\,d\mu$ with $f \in C_\psi(\mathbb{R}^d)$ are continuous (cf. Section A.6 in [13]). Clearly, the ψ-weak topology is finer than the weak topology, and the two topologies coincide if and only if ψ is bounded. The following lemma provides some useful characterizations of the ψ-weak topology.

Lemma 3.4 *Let ψ be a weight function. Then the ψ-weak topology is always metrizable, and for $\mu, \mu_1, \mu_2, \ldots \in \mathcal{M}_1^\psi(\mathbb{R}^d)$ the following statements are equivalent:*

(i) $\mu_n \to \mu$ *ψ-weakly.*
(ii) $\int f\,d\mu_n \to \int f\,d\mu$ *for every* $f \in C_\psi(\mathbb{R}^d)$.
(iii) $\int f\,d\mu_n \to \int f\,d\mu$ *for every continuous f with compact support and for* $f = \psi$.
(iv) $\mu_n \to \mu$ *weakly and* $\int \psi\,d\mu_n \to \int \psi\,d\mu$.

Proof It is already known that the ψ-weak topology is metrizable (cf. [13], Corollary A.45). Furthermore, the equivalence of (i) and (ii) follows by definition of the ψ-weak topology, whereas the chain of implications (i)⇒(iv)⇒(iii) is obvious. In order to show (iii)⇒(i), let $\mathcal{M}(\mathbb{R}^d)$ be the set of all finite Borel measures on $\mathbb{R}^d$. We may associate any $\nu \in \mathcal{M}_1^\psi(\mathbb{R}^d)$ with $\Psi(\nu) \in \mathcal{M}(\mathbb{R}^d)$ defined via $d\Psi(\nu) = \psi\,d\nu$. It is known that $\mu_n \to \mu$ ψ-weakly if and only $\Psi(\mu_n) \to \Psi(\mu)$ weakly; cf. [13, p. 502]. Since statement (iii) implies that $(\Psi(\mu_n))$ converges weakly to $\Psi(\mu)$ by Theorem 30.8 in [2] we have shown the desired implication (iii)⇒(i). The proof is now complete. □

In the sequel, we will focus on the following metric which generates the ψ-weak topology for any weight function ψ due to Lemma 3.4:

$$d_\psi(\mu,\nu) := d_{\mathrm{Proh}}(\mu,\nu) + \left|\int \psi\,d\mu - \int \psi\,d\nu\right|, \qquad \mu,\nu \in \mathcal{M}_1^\psi(\mathbb{R}^d). \tag{7}$$

Let $\mathcal{M}_{1,\kappa}^{\psi^p}(\mathbb{R}^d)$ be the class of all $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ satisfying $\int \psi^p\,d\mu \le \kappa$.

Corollary 3.5 (Uniform Glivenko-Cantelli w.r.t. d_ψ) *Let ψ be a weight function. Then the space $(\mathcal{M}_{1,\kappa}^{\psi^p}(\mathbb{R}^d), d_\psi)$ has the UGC property for all $\kappa > 0$ and $p > 1$.*

Proof Fix $\kappa > 0$ and $p > 1$, and let μ_ψ denote the law of ψ under $\mu \in \mathcal{M}_1^\psi(\mathbb{R}^d)$. Then

$$\Big|\int \psi\, d\mu - \int \psi\, d\nu\Big| \;=\; \Big|\int (F_{\mu_\psi}(x) - F_{\nu_\psi}(x))\, dx\Big| \;\le\; d_{W_1}(\mu_\psi, \nu_\psi),$$

where d_{W_1} is as in (6). Thus, Corollary 3.3 (with $\lambda = p$and $\kappa' := 2^p\kappa$) implies that for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for every $\mu \in \mathcal{M}_{1,\kappa}^{\psi^p}(\mathbb{R}^d)$ and $n \ge n_0$

$$\begin{aligned}
&\mu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} :\, \Big|\int \psi\; d\mu - \int \psi\; d\hat{m}_{x^{(n)}}\Big| \ge \delta/2\Big\}\Big]\\
&\le \mu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} :\, d_{W_1}(\mu_\psi, (\hat{m}_{x^{(n)}})_\psi) \ge \delta/2\Big\}\Big]\\
&= (\mu_\psi)^{\mathbb{N}}\Big[\Big\{y \in \mathbb{R}^{\mathbb{N}} :\, d_{W_1}(\mu_\psi, \hat{m}_{y^{(n)}}) \ge \delta/2\Big\}\Big]\\
&\le \varepsilon/2.
\end{aligned}$$

Combining this with the UGC property of $(\mathcal{M}_1(\mathbb{R}^d), d_{\mathrm{Proh}})$ due to [21, Lemma 4], we obtain the claim of Corollary 3.5. □

By applying Theorem 2.4 we get the following extension of Hampel's theorem.

Corollary 3.6 *Let ψ be any weight function, and suppose that T is continuous w.r.t. the ψ-weak topology at some μ belonging to $\mathcal{M}_{1,\kappa}^{\psi^p}(\mathbb{R}^d)$ for any $p > 1$ and $\kappa > 0$. Then the sequence $(\hat{T}_n)$ is qualitatively $\mathcal{M}_{1,\kappa}^{\psi^p}(\mathbb{R}^d)$-robust at μ w.r.t. $(d_\psi, d'_{\mathrm{Proh}})$.*

Remark 3.7 We point out that Corollaries 3.5 and 3.6 are not limited to measures in $\mathcal{M}_\psi(\mathbb{R}^d)$ but that they also hold when $\mathbb{R}^d$ is replaced by any Polish space S. To get this, one has to note first that we may retain the equivalence of the respective statements (i), (ii), (iv) in Lemma 3.4 using as the same argument from [13, Section A.6]. In particular, d_ψ defined in (7) also metrizes the ψ-weak topology on $\mathcal{M}_\psi(S)$. Next, the UGC property for the Prohorov metric on $\mathcal{M}_1(S)$ was proved in [21, Lemma 4]. Now the result follows as before. ◇

Remark 3.8 *(Index of qualitative robustness)* As indicated in the Introduction, our concept allows for a direct comparison of two statistical functionals in regards to their degree of robustness. A statistical functional T_0 can be considered to have a higher degree of qualitative robustness than another statistical functional T_1 when T_0 is qualitatively robust for any choice of d for which T_1 is qualitatively robust. If we restrict our attention to the class of metrics d_{ψ_λ} with weight functions $\psi_\lambda(\cdot) := (1 + |\cdot|)^\lambda$, $\lambda \geq 0$, where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$, then we can for instance define an *index of qualitative robustness* of a statistical functional T as

$$\mathrm{iqr}(T) := \Big(\inf\Big\{\lambda \in [0,\infty) \,:\, T \text{ is qualitatively robust w.r.t. } d_{\psi_\lambda}\Big\}\Big)^{-1}.$$

This index can be a quantitative measure for the degree of robustness of a statistical functional, and a higher index should intuitively reflect a higher degree of robustness. It is also clear, and will be verified formally in Section 4 (Examples 4.3, 4.2 and 4.5), that the median functional T_0, the mean functional T_1 and the variance functional T_2 have the indices $\mathrm{iqr}(T_0) = \infty$, $\mathrm{iqr}(T_1) = 1$, and $\mathrm{iqr}(T_2) = \frac{1}{2}$. $\Diamond$

4 Examples for qualitatively robust statistical functionals

In this section, we illustrate our theoretical results by means of general L- and V-functionals as well as the covariance functional. We will frequently use the u-shaped function $\phi_\lambda(x) := (1 + |x|)^\lambda$, $x \in \mathbb{R}$, and the weight function $\psi_\lambda(x) := (1 + |x|)^\lambda$, $x \in \mathbb{R}^d$, with $\lambda \geq 0$ and $|\cdot|$ denoting the Euclidean norm on $\mathbb{R}^d$. We will also work with the function ϕ_λ for $\lambda < 0$.

4.1 L-functionals

Let K be the df of a probability measure on $([0,1], \mathcal{B}([0,1]))$, and $\mathcal{M}_K$ be the class of all $\mu \in \mathcal{M}_1(\mathbb{R})$ for which $\int |x|\, dK(F_\mu(x)) < \infty$, where F_μ refers to the df of μ. The functional $\mathcal{L}_K$, defined by

$$\mathcal{L}_K(\mu) := \int x\, dK(F_\mu(x)), \qquad \mu \in \mathcal{M}_K,$$

is called *L-functional* associated with K; cf., e.g., [24, p.265]. If $K(x) = \mathbb{I}(x) := x$, then $\mathcal{L}_K$ is the mean functional, i.e. $\mathcal{L}_{\mathbb{I}}(\mu) = \int x\,\mu(dx)$. If $K(x) = \mathbb{1}_{[\alpha,1]}$ for some $\alpha \in (0,1)$, then $\mathcal{L}_K$ is the lower α-quantile functional, i.e. $\mathcal{L}_{\mathbb{1}_{[\alpha,1]}}(\mu) = F_\mu^{\leftarrow}(\alpha)$. These two examples are covered by the following Example 4.1, where qualitative robustness w.r.t. the weighted Kolmogorov metric is studied. In Examples 4.2–4.3 below, we will study qualitative robustness of the sample mean and the sample lower α-quantile w.r.t. the metric d_ψ (defined in (7)) generating the ψ-weak topology.

Example 4.1 Let $\mu \in \mathcal{M}_K$, and assume that the following two assertions hold:

(a) There are constants $\beta, C > 0$, $m \in \mathbb{N}_0$, and $0 = d_0 < d_1 < \ldots < d_{m+1} = 1$, such that K is Hölder-β-continuous on each of the intervals (d_i, d_{i+1}), $i = 0, \ldots, m$.

(b) F_μ is differentiable at $F_\mu^{\leftarrow}(d_i)$, and $F_\mu'(F_\mu^{\leftarrow}(d_i)) > 0$, for $i = 1, \ldots, m$.

It follows from results in [32] that in this case, and for every $\lambda > 0$ satisfying $\lambda\beta > 1$, the functional $\mathcal{L}_K$ is $\mathcal{M}_1^{(\phi_\lambda)}(\mathbb{R})$-continuous (and so $\mathcal{E}$-continuous) at μ w.r.t. $(d_{(\phi_\lambda)}, |\cdot|)$. Thus the Hampel-type result of Corollary 3.2 implies that in this case the sequence $(\mathcal{L}_K(\hat{m}_{n;\mu}))$ is qualitatively $\mathcal{M}_{1,\kappa}^{\phi_{\lambda''}}(\mathbb{R})$-robust at $\mu \in \mathcal{M}_{1,\kappa}^{\phi_{\lambda''}}(\mathbb{R})$ w.r.t. $(d_{(\phi_\lambda)}, |\cdot|)$ for every $\lambda'' > \lambda$ and $\kappa \geq \int \phi_{\lambda''}\,d\mu$. $\diamond$

Example 4.2 *(Sample mean)* It follows directly from Lemma 3.4 (i)$\Rightarrow$(ii) that the mean functional $\mathcal{L}_{\mathbb{I}}$ is continuous w.r.t. the ψ_1-weak topology. Therefore the Hampel-type result of Corollary 3.6 shows that the sample mean is qualitatively $\mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R})$-robust at $\mu \in \mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R})$ w.r.t. $(d_{\psi_1}, d'_{\mathrm{Proh}})$ for every $\lambda'' > 1$ and $\kappa \geq \int \psi_{\lambda''}\,d\mu$. On the other hand, the sample mean is weakly consistent at every $\nu \in \mathcal{M}_1^{\psi_1}(\mathbb{R})$, and it is easy to see that $\mathcal{L}_{\mathbb{I}}$ is *not* continuous w.r.t. the ψ_λ-weak topology at μ for every $0 \leq \lambda < 1$. Therefore the converse of Hampel's criterion (Theorem 2.6) implies that the sample mean is *not* qualitatively $\mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R})$-robust at μ w.r.t. $(d_{\psi_\lambda}, d'_{\mathrm{Proh}})$ if $0 \leq \lambda < \lambda'' \leq 1$. $\diamond$

Example 4.3 *(Sample quantile)* It is well-known that the lower α-quantile functional $\mathcal{L}_{\mathbb{1}_{[\alpha,1]}}$ is continuous w.r.t. the weak topology at $\mu \in \mathcal{M}_1(\mathbb{R})$ if the leftcontinuous quantile function $F_\mu^{\leftarrow}$ is continuous at α (see, e.g., [30, Lemma 21.2]). The classical Hampel theorem as given in [17, Theorem 2.21] (cf. Section 3.1) then shows that the sample lower α-quantile is qualitatively $\mathcal{M}_1(\mathbb{R})$-robust at $\mu \in \mathcal{M}_1(\mathbb{R})$ w.r.t. $(d_{\mathrm{Proh}}, d'_{\mathrm{Proh}})$. Notice that $d_{\mathrm{Proh}} = d_{\psi_0}$. $\diamond$

4.2 V-functionals

Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function, and $\mathcal{M}_g$ be the class of all $\mu \in \mathcal{M}_1(\mathbb{R})$ for which $\int\int |g(x_1, x_2)|\, \mu(dx_1)\mu(dx_2) < \infty$. The functional $\mathcal{V}_g$, defined by

$$\mathcal{V}_g(\mu) := \int\int g(x_1, x_2)\, \mu(dx_1)\mu(dx_2), \qquad \mu \in \mathcal{M}_g,$$

is called *V-functional* (or *von Mises-functional*) associated with g. For background see, e.g., [8,19,24]. If $g(x_1, x_2) = \widetilde{g}(x_1, x_2) := \frac{1}{2}(x_1 - x_2)^2$, then $\mathcal{V}_{\widetilde{g}}$ is the variance functional, i.e. $\mathcal{V}_{\widetilde{g}}(\mu) = \int (x - \mathcal{L}_{\mathbb{I}}(\mu))^2\, \mu(dx)$, where $\mathcal{L}_{\mathbb{I}}$ is the mean functional (cf. Section 4.1). This example is covered by the following Example 4.4, where qualitative robustness w.r.t. the weighted Kolmogorov metric is studied. In Example 4.5 below, we will study qualitative robustness of the sample variance w.r.t. the metric d_ψ (defined in (7)) generating the ψ-weak topology.

Example 4.4 Let $BV_{\mathrm{loc}}(\mathbb{R})$ be the space of all functions $f : \mathbb{R} \to \mathbb{R}$ of locally bounded variation, and $D_{(\phi_{-\lambda'})}(\mathbb{R})$ be the space of all càdlàg functions $f : \mathbb{R} \to \mathbb{R}$ with $\|f\|_{\phi_{-\lambda'}} < \infty$. Let $\mu \in \mathcal{M}_g$, and assume that for some $\lambda > \lambda' \geq 0$ the integral $\int \phi_{\lambda'}\, d\mu$ is finite and the following two assertions hold:

(a) For every $x_2 \in \mathbb{R}$ fixed, the function $g_{x_2}(\cdot) := g(\,\cdot\,, x_2)$ lies in $BV_{\mathrm{loc}}(\mathbb{R}) \cap D_{(\phi_{-\lambda'})}(\mathbb{R})$. Moreover, the function $x_2 \mapsto \int \phi_{-\lambda}(x_1)|dg_{x_2}|(x_1)$ lies in $D_{(\phi_{-\lambda'})}(\mathbb{R})$.
(b) The functions $g_{1,\mu}(\cdot) := \int g(\,\cdot\,, x_2)\mu(dx_2)$ and $g_{2,\mu}(\cdot) := \int g(x_1, \,\cdot\,)\mu(dx_1)$ lie in $BV_{\mathrm{loc}}(\mathbb{R})$, and we have $\int \phi_{-\lambda}(x)\, |dg_{i,\mu}|(x) < \infty$ for $i = 1, 2$. Moreover, the functions $\overline{g_{1,\mu}}(\cdot) := \int |g(\,\cdot\,, x_2)|\, \mu(dx_2)$ and $\overline{g_{2,\mu}}(\cdot) := \int |g(x_1, \,\cdot\,)|\, \mu(dx_1)$ lie in $D_{(\phi_{-\lambda'})}(\mathbb{R})$.

As pointed out in [33], it follows directly from results in [4] that under assumptions (a)–(b) the functional $\mathcal{V}_g$ is $\mathcal{M}_1^{(\phi_\lambda)}(\mathbb{R})$-continuous (and so $\mathcal{E}$-continuous) at μ w.r.t. $(d_{(\phi_\lambda)}, |\cdot|)$. Thus the Hampel-type result of Corollary 3.2 implies that in this case the sequence $(\mathcal{V}_g(\hat{m}_{n;\mu}))$ is qualitatively $\mathcal{M}_{1,\kappa}^{\phi_{\lambda''}}(\mathbb{R})$-robust at $\mu \in \mathcal{M}_{1,\kappa}^{\phi_{\lambda''}}(\mathbb{R})$ w.r.t. $(d_{(\phi_\lambda)}, |\cdot|)$ for every $\lambda'' > \lambda$ and $\kappa \geq \int \phi_{\lambda''}\, d\mu$. $\diamond$

Example 4.5 *(Sample variance)* It follows easily from Lemma 3.4 (i)⇒(ii) that the variance functional $\mathcal{V}_{\widetilde{g}}$ is continuous w.r.t. the ψ_2-weak topology. Therefore the Hampel-type result of Corollary 3.6 shows that the sample variance is qualitatively $\mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R})$-robust at $\mu \in \mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R})$ w.r.t. $(d_{\psi_2}, d'_{\text{Proh}})$ for every $\lambda'' > 2$ and $\kappa \geq \int \psi_{\lambda''}\, d\mu$. On the other hand, the sample variance is weakly consistent at every $\nu \in \mathcal{M}_1^{\psi_2}(\mathbb{R})$, and it is easy to see that $\mathcal{V}_{\widetilde{g}}$ is *not* continuous w.r.t. the ψ_λ-weak topology at μ for every $0 \leq \lambda < 2$. Therefore the converse of Hampel's criterion (Theorem 2.6) implies that the sample mean is *not* qualitatively $\mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R})$-robust at μ w.r.t. $(d_{\psi_\lambda}, d'_{\text{Proh}})$ if $0 \leq \lambda < \lambda'' \leq 2$. $\diamond$

4.3 Covariance

The *covariance functional* $\mathcal{C}$ on $\mathcal{M}_1^{\psi_2}(\mathbb{R}^2)$ is given by

$$\mathcal{C}(\mu) := \int (x_1 - \mathcal{L}_{\mathbb{I}}(\mu_{\pi_1}))(x_2 - \mathcal{L}_{\mathbb{I}}(\mu_{\pi_2}))\, \mu(d(x_1, x_2)), \qquad \mu \in \mathcal{M}_1^{\psi_2}(\mathbb{R}^2),$$

where $\mathcal{L}_{\mathbb{I}}$ is the mean functional (cf. Section 4.1), and μ_{π_1}, μ_{π_2} denote the marginal distributions of μ. In view of Lemma 3.4 (i)⇒(ii), it is easily seen that $\mathcal{C}$ is continuous w.r.t. the ψ_2-weak topology. Therefore the Hampel-type result of Corollary 3.6 implies that the sample covariance, i.e. the sequence $(\mathcal{C}(\hat{m}_{n;\mu}))$, is qualitatively $\mathcal{M}_{1,\kappa}^{\psi_\lambda}(\mathbb{R}^2)$-robust at $\mu \in \mathcal{M}_{1,\kappa}^{\psi_\lambda}(\mathbb{R}^2)$ w.r.t. $(d_{\psi_2}, d'_{\text{Proh}})$ for every $\lambda > 2$ and $\kappa \geq \int \psi_\lambda\, d\mu$. On the other hand, the sample covariance is weakly consistent at every $\nu \in \mathcal{M}_1^{\psi_2}(\mathbb{R}^2)$, and it can be shown that $\mathcal{C}$ is *not* continuous w.r.t. the ψ_λ-weak topology at μ for every $0 \leq \lambda < 2$. Therefore the converse of Hampel's criterion (Theorem 2.6) implies that the sample covariance is *not* qualitatively $\mathcal{M}_{1,\kappa}^{\psi_{\lambda''}}(\mathbb{R}^2)$-robust at μ w.r.t. $(d_{\psi_\lambda}, d'_{\text{Proh}})$ if $0 \leq \lambda < \lambda'' \leq 2$.

5 Infinitesimal robustness and error sensitivity

We continue in the setting of Section 2. However, we will focus on the case of one-dimensional observations and —by an abuse of notation— we will identify a probability measure μ with its df F_μ. In particular, the empirical measure $\hat{m}_{n;F}$ will be identified with its df $\hat{F}_n$. Apart from qualitative robustness, the question of error sensitivity is of large interest. In this context, the influence function of T at F, defined by

$$\begin{aligned}\mathrm{IF}(x;T,F) &:= \lim_{h\downarrow 0}\frac{1}{h}\Big(T\Big(F+h(\mathbb{1}_{[x,\infty)}-F)\Big)-T(F)\Big)\\ &= \lim_{h\downarrow 0}\frac{1}{h}\Big(T\Big((1-h)F+h\mathbb{1}_{[x,\infty)}\Big)-T(\mu)\Big), \qquad x\in\mathbb{R},\end{aligned}$$

is highly relevant; cf. [16]. Notice that, on the one hand, $\mathrm{IF}(x;T,F)$ can be seen as the directional derivative of T at F in direction of $\mathbb{1}_{[x,\infty)}-F$. On the other hand, it can also be seen as a measure for the sensitivity of $T(\hat{F}_n)$ w.r.t. a change of the underlying model F to the mixture $F_h=(1-h)F+h\mathbb{1}_{[x,\infty)}$ of F and the Dirac measure $\mathbb{1}_{[x,\infty)}$ at x (with $h>0$ small). Indeed: For small fixed $h>0$ we have

$$\begin{aligned}T(\hat{F}_{h,n})-T(\hat{F}_n) &= (T(\hat{F}_{h,n})-T(F_h))+(T(F_h)-T(F))+(T(F)-T(\hat{F}_n))\\ &\approx (T(\hat{F}_{h,n})-T(F_h))+\mathrm{IF}(x;T,F)\,h+(T(F)-T(\hat{F}_n))\end{aligned}$$

which tends to $\mathrm{IF}(x;T,F)\,h$ as $n\to\infty$ whenever $T(\hat{F}_{h,n})$ and $T(\hat{F}_n)$ are consistent for $T(F_h)$ and $T(F)$, respectively.

Now, on the one hand, the influence function is interesting on its own, and it is the basis for the notion of robustness against outliers; recall that a functional T is called *robust against outliers* if the gross-error sensitivity $\gamma(T,F):=\sup_{x\in\mathbb{R}}|\mathrm{IF}(x;T,F)|$ is finite. On the other hand, the influence function suffers a lack of information because it might be also interesting to know how the functional T reacts on a change of the underlying model F to the mixture $F_h=(1-h)F+hG$ of F and a more general distribution G (with $h>0$ small), i.e. on a change of F in direction $G-F$. So it is natural to ask for the directional derivative for "general" directions. Actually, the existence of the directional derivative for "all" directions $G-F$ is not the broadest requirement.

Typically one is even interested in a regular behavior of the directional derivative as a function of the direction (this is the mode of Gâteaux differentiability) and, in addition, in a uniform convergence of the differential quotient within compact sets of directions (this is the mode of Hadamard differentiability). Hadamard differentiability obviously implies a stronger regularity of T than just the existence of all directional derivatives. Hence, Hadamard differentiability can, to some extent, be regarded as a sort of infinitesimal (order-one) robustness (cf. Definition 5.2 below), whereas the directional derivative of T at F in direction $G-F$ can be seen as a measure for the sensitivity of $T(\hat{F}_n)$ w.r.t. to a contamination $F_h = (1-h)F + hG$ of F. We emphasize that Hadamard differentiability implies in particular that the sensitivity of $T(\hat{F}_n)$ w.r.t. a contamination $F_h = (1-h)F + hG$ of F is hardly distinguishable for any two G being close to each other.

If T is a tail-dependent functional, then one has to be careful about the choice of G. If G is allowed to be an arbitrary df on $\mathbb{R}$, then the contamination $F_h = (1-h)F+hG$ may lie outside the domain of T. That is, G has to be chosen in such a way that F_h is contained in the domain of T. But this is typically not enough. If the functional T is tail-dependent, then one has to be careful about the exact notion of "Hadamard differentiability" to be used. It was discussed in [3] that for tail-dependent functionals the classical concept of tangential Hadamard differentiability (where the tangential space is equipped with the same norm as the space in which F lies; cf. [12,14,31]) is often not suitable, because many popular tail-dependent functionals are *not* tangentially Hadamard differentiable. Along with this discussion, in [3] a refined concept of tangential Hadamard differentiability —called *quasi*-Hadamard differentiability— was introduced. This concept ensures that much more functionals are "Hadamard differentiable". For the reader's convenience we recall the definition.

Definition 5.1 (Quasi-Hadamard differentiability) *Let* $\mathbf{V}$ *be a vector space,* $(\mathbf{V}', \|\cdot\|_{\mathbf{V}'})$ *be a normed vector space, and* $\tau : \mathbf{V}_\tau \to \mathbf{V}'$ *be a mapping defined on a subset* $\mathbf{V}_\tau$ *of* $\mathbf{V}$. *Further, let* $\mathbf{V}_0$ *be a subspace of* $\mathbf{V}$ *equipped with a norm* $\|\cdot\|_{\mathbf{V}_0}$, *and* $\mathbb{C}_0$ *be a subset of* $\mathbf{V}_0$. *Then* τ *is said to be* quasi-Hadamard differentiable *at* $\theta \in \mathbf{V}_\tau$ *tangentially to* $\mathbb{C}_0\langle\mathbf{V}_0\rangle$ *if there is some continuous mapping* $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}\tau : \mathbb{C}_0 \to \mathbf{V}'$ *such that*

$$\lim_{n\to\infty} \left\| D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}\tau\,(v) - \frac{\tau(\theta+h_nv_n)-\tau(\theta)}{h_n}\right\|_{\mathbf{V}'} = 0 \tag{8}$$

holds for each triplet $(v,(v_n),(h_n))$*, with* $v\in\mathbb{C}_0$*,* $(h_n)\subset(0,\infty)$ *satisfying* $h_n\to 0$*, and* $(v_n)\subset\mathbf{V}_0$ *satisfying* $\|v_n-v\|_{\mathbf{V}_0}\to 0$ *as well as* $\theta+h_nv_n\in\mathbf{V}_\tau$ *for every* $n\in\mathbb{N}$*. In this case the mapping* $D^{\mathrm{Had}}_{\theta;\mathbb{C}_0\langle\mathbf{V}_0\rangle}\tau$ *is called* quasi-Hadamard derivative *of* τ *at* θ *tangentially to* $\mathbb{C}_0\langle\mathbf{V}_0\rangle$*. If* $\mathbb{C}_0=\mathbf{V}_0$*, then we replace "*$\mathbb{C}_0\langle\mathbf{V}_0\rangle$*" by "*$\mathbf{V}_0$*".*

Now, let $D(\mathbb{R})$ be the space of all bounded càdlàg functions on the real line. The domain $\mathcal{M}$ of T will be regarded as a subset of the set of all df on the real line, in particular $\mathcal{M}\subset D(\mathbb{R})$. Further, let D_0 be a subspace of $D(\mathbb{R})$, C_0 be a subset of D_0, and $\|\cdot\|_{D_0}$ be a norm on D_0. Finally, assume that $(\mathbf{T},\|\cdot\|_{\mathbf{T}})$ is a normed vector space.

Definition 5.2 (Infinitesimal $C_0\langle D_0\rangle$-robustness) *Let* $F\in\mathcal{M}$*. Then the sequence* $(\hat{T}_n)$ *of estimators is said to be* infinitesimally $C_0\langle D_0\rangle$-robust *(of order one) at* F *w.r.t.* $(\|\cdot\|_{D_0},\|\cdot\|_{\mathbf{T}})$ *if* $T:\mathcal{M}\to\mathbf{T}$ *is quasi-Hadamard differentiable at* F *tangentially to* $C_0\langle D_0\rangle$*. Moreover, for every df* G *on the real line with* $G-F\in C_0$ *the quasi-Hadamard derivative evaluated at* $G-F$*,*

$$D^{\mathrm{Had}}_{F;C_0\langle D_0\rangle}T\,(G-F),$$

is called infinitesimal $C_0\langle D_0\rangle$-sensitivity *of* $T(\hat{F}_n)$ *w.r.t. to a contamination* $F_h=(1-h)F+hG$ *of* F *with infinitesimally small* $h>0$*. In the case* $C_0=D_0$ *we replace "*$C_0\langle D_0\rangle$*" by "*D_0*".*

Remark 5.3 As demonstrated in [33], in many relevant situations quasi-Hadamard differentiability implies continuity w.r.t. $(\|\cdot\|_{D_0},\|\cdot\|_{\mathbf{T}})$. That is, in view of the Hampel Theorem 2.4, infinitesimal robustness is typically a stronger notion of "robustness" than qualitative robustness. $\diamond$

In the following examples we have $C_0=D_0=D_\lambda:=D_{(\phi_\lambda)}(\mathbb{R})$, $\|\cdot\|_{D_0}=\|\cdot\|_{\phi_\lambda}$ and $(\mathbf{T},\|\cdot\|_{\mathbf{T}})=(\mathbb{R},|\cdot|)$, where $D_{(\phi_\lambda)}(\mathbb{R})$ is the space of all $f\in D(\mathbb{R})$ with $\|f\|_{\phi_\lambda}<\infty$; recall $\phi_\lambda(x):=(1+|x|)^\lambda$, $x\in\mathbb{R}$.

Example 5.4 *(L-functionals)* In Section 4.1 we introduced the L-functional $\mathcal{L}_K$ on $\mathcal{M}_K$. Let $\lambda>1$ and $F\in\mathcal{M}_K$. It was shown in [3] that if K is continuous and piecewise differentiable, the (piecewise) derivative K' is bounded above and F takes the value $x \in (0,1)$ at most once if K is not differentiable at x, then the functional $\mathcal{L}_K : \mathcal{M}_K \to \mathbb{R}$ is quasi-Hadamard differentiable at F tangentially to D_λ with quasi-Hadamard derivative $D^{\rm Had}_{F;D_\lambda}\mathcal{L}_K : D_\lambda \to \mathbb{R}$ given by

$$D^{\rm Had}_{F;D_\lambda}\mathcal{L}_K(v) \;=\; \int K'(F(x))\, v(x)\, dx, \qquad v \in D_\lambda.$$

That is, in this case the sequence $(\mathcal{L}_K(\hat{F}_n))$ is infinitesimally D_λ-robust w.r.t. $(\|\cdot\|_{\phi_\lambda}, |\cdot|)$. Moreover, for every df G on the real line with $G - F \in D_\lambda$ the value $D^{\rm Had}_{F;D_\lambda}\mathcal{L}_K(G-F)$ can be seen as the infinitesimal D_λ-sensitivity of $\mathcal{L}_K(\hat{F}_n)$ w.r.t. to a contamination $F_h = (1-h)F + hG$ of F with infinitesimally small $h > 0$. $\diamond$

Example 5.5 *(V-functionals)* In Section 4.2 we introduced the V-functional $\mathcal{V}_g$ on $\mathcal{M}_g$. Let $\lambda > \lambda' \geq 0$, suppose that $F \in \mathcal{M}_g$ satisfies $\int \phi_{\lambda'}\, dF < \infty$ and that the assumptions (a)–(b) of Example 4.4 are fulfilled. Then, as it was shown in [4], the functional $\mathcal{V}_g : \mathcal{M}_g \to \mathbb{R}$ is quasi-Hadamard differentiable at F tangentially to D_λ with quasi-Hadamard derivative $D^{\rm Had}_{F;D_\lambda}\mathcal{V}_g : D_\lambda \to \mathbb{R}$ given by

$$D^{\rm Had}_{F;D_\lambda}\mathcal{V}_g(v) \;=\; -\int v\, dg_{1,F} - \int v\, dg_{2,F}, \qquad v \in D_\lambda \tag{9}$$

with $g_{1,F}, g_{2,F}$ defined as in (b) of Example 4.4. That is, in this case the sequence $(\mathcal{V}_g(\hat{F}_n))$ is infinitesimally D_λ-robust w.r.t. $(\|\cdot\|_{\phi_\lambda}, |\cdot|)$. Moreover, for every df G on the real line with $G - F \in D_\lambda$ the value $D^{\rm Had}_{F;D_\lambda}\mathcal{V}_g(G-F)$ can be seen as the infinitesimal D_λ-sensitivity of $\mathcal{V}_g(\hat{F}_n)$ w.r.t. to a contamination $F_h = (1-h)F + hG$ of F with infinitesimally small $h > 0$. $\diamond$

Remark 5.6 *(Degenerate V-functionals)* Among V-functionals the functionals with a degenerate kernel have attracted special interest; see, e.g., [24, Section 5.5], [30, Section 12.3] or [5, Section 3.3]. Recall that a kernel g is called *degenerate* w.r.t. $F \in \mathcal{M}_g$ if the functions $g_{1,F}$ and $g_{2,F}$ defined in (b) of Example 4.4 are identically zero. In this case, $\mathcal{V}_g$ is referred to as *degenerate* V-functional w.r.t. F. Thus, the quasi-Hadamard derivative (given by (9)) of a degenerate V-functional $\mathcal{V}_g$ at F tangentially to D_λ vanishes. Hence, degenerate V-functionals w.r.t. F are very insensitive w.r.t. a small contamination of F. $\diamond$

A Remaining proofs

A.1 Proof of Theorem 2.4

We adapt the proof of the classical Hampel theorem as given in [17]. We have to show that for every $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for all $\nu \in \mathcal{M}_0$ and $n \geq n_0$, the inequality $d(\mu, \nu) \leq \delta$ implies $d'_{\mathrm{Proh}}(\mu^{\mathbb{N}} \circ \hat{T}_n^{-1}, \nu^{\mathbb{N}} \circ \hat{T}_n^{-1}) \leq \varepsilon$. So, let $\varepsilon > 0$ be fixed. By the triangular inequality, it suffices to find some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for all $\nu \in \mathcal{M}_0$ and $n \geq n_0$

$$d(\mu, \nu) \leq \delta \quad \Longrightarrow \quad d'_{\mathrm{Proh}}(\delta_{T(\mu)}, \nu^{\mathbb{N}} \circ \hat{T}_n^{-1}) \leq \frac{\varepsilon}{2},$$

where $\delta_{T(\mu)}$ is the Dirac-measure at $T(\mu)$ on the measurable space $(\mathbf{T}, \mathcal{T})$. As a consequence of Strassen's theorem (cf. [29], or [17, Theorem 2.13]), we have

$$\nu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} : d_{\mathbf{T}}(T(\mu), T(\hat{m}_{x^{(n)}})) \leq \frac{\varepsilon}{2}\Big\}\Big] \geq 1 - \frac{\varepsilon}{2} \quad \Longrightarrow \quad d'_{\mathrm{Proh}}(\delta_{T(\mu)}, \nu^{\mathbb{N}} \circ \hat{T}_n^{-1}) \leq \frac{\varepsilon}{2},$$

where $x^{(n)} := (x_1, \ldots, x_n)$ for $x = (x_1, x_2, \ldots) \in (\mathbb{R}^d)^{\mathbb{N}}$. Thus, it suffices to find some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for all $\nu \in \mathcal{M}_0$ and $n \geq n_0$

$$d(\mu, \nu) \leq \delta \quad \Longrightarrow \quad \nu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} : d_{\mathbf{T}}(T(\mu), T(\hat{m}_{x^{(n)}})) \leq \frac{\varepsilon}{2}\Big\}\Big] \geq 1 - \frac{\varepsilon}{2}. \tag{A.1}$$

Since T is $\mathcal{E}$-continuous at μ w.r.t. $(d, d_{\mathbf{T}})$, we can find some $\delta > 0$ such that for all $n \in \mathbb{N}$ and $x^{(n)} \in \mathbb{R}^n$, the inequality $d(\mu, \hat{m}_{x^{(n)}}) \leq 2\delta$ implies $d_{\mathbf{T}}(T(\mu), T(\hat{m}_{x^{(n)}})) \leq \frac{\varepsilon}{2}$. Thus, in order to obtain (A.1), let us fix any $\nu \in \mathcal{M}_0$ satisfying $d(\mu, \nu) \leq \delta$. In view of the triangular inequality $d(\mu, \hat{m}_{x^{(n)}}) \leq d(\nu, \hat{m}_{x^{(n)}}) + d(\mu, \nu)$ we have

$$\nu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} : d(\nu, \hat{m}_{x^{(n)}}) \leq \delta\Big\}\Big]$$

$$
\begin{aligned}
&\leq \nu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} \,:\, d(\mu, \hat{m}_{x(n)}) \leq \delta + d(\mu,\nu)\Big\}\Big] \\
&\leq \nu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} \,:\, d(\mu, \hat{m}_{x(n)}) \leq 2\delta\Big\}\Big] \\
&\leq \nu^{\mathbb{N}}\Big[\Big\{x \in (\mathbb{R}^d)^{\mathbb{N}} : d_{\mathbf{T}}(T(\mu), T(\hat{m}_{x(n)})) \leq \frac{\varepsilon}{2}\Big\}\Big].
\end{aligned}
$$

Now, (A.1) is an immediate consequence of the UGC property of $(\mathcal{M}_0, d)$.

A.2 Proof of Theorem 2.6

By the triangular inequality, we have for every $\nu \in \mathcal{M}_0$

$$
\begin{aligned}
&d_{\mathbf{T}}(T(\mu), T(\nu)) \\
&= d'_{\text{Proh}}(\delta_{T(\mu)}, \delta_{T(\nu)}) \\
&\leq d'_{\text{Proh}}(\delta_{T(\mu)}, \mu^{\mathbb{N}} \circ \hat{T}_n^{-1}) + d'_{\text{Proh}}(\mu^{\mathbb{N}} \circ \hat{T}_n^{-1}, \nu^{\mathbb{N}} \circ \hat{T}_n^{-1}) + d'_{\text{Proh}}(\nu^{\mathbb{N}} \circ \hat{T}_n^{-1}, \delta_{T(\nu)}).
\end{aligned}
$$

For the claim of Theorem 2.6 it suffices to show that for every $\varepsilon > 0$ one can find some $\delta > 0$ such that $d(\mu, \nu) \leq \delta$ implies that the sum in the inequality above converges to ε as $n \to \infty$. Now, let $\delta > 0$ so small so that $d(\mu, \nu) \leq \delta$ implies that ν lies in the neighborhood of μ which appears in the statement of Theorem 2.6. Since $(\hat{T}_n)$ is weakly consistent at every such ν, the first and the third summand converge to zero as $n \to \infty$. Moreover, by the qualitative $\mathcal{M}_0$-robustness of $(\hat{T}_n)$ at μ and a suitable choice of δ, the second summand is bounded above by ε. This implies the claim.

A.3 Proof of Theorem 3.1

The proof of Theorem 3.1 relies on the following version of the Birnbaum-Marshall inequality, which is due to Slud [28].

Theorem A.1 *On some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(M(t) : t \in [0, T))$ be a square-integrable càdlàg martingale with $M(0) = 0$. Furthermore, let $h : [0, T) \to [0, \infty]$ be a nonincreasing and rightcontinuous function which is* *finite on* $(0,T)$. *Then, for every* $\lambda > 0$,

$$\mathbb{P}\Big[\sup_{t\in(0,T)} |M(t)|\, h(t) > \lambda\Big] \;\le\; \frac{1}{\lambda}\int_{(0,T)} h(t)\, dv(t),$$

where $v(t) := \mathbb{E}[\,M^+(t)\,]$.

Proof Take a sequence of nonincreasing rightcontinuous functions $h_k : [0,T) \to \mathbb{R}_+$ such that $h_k \nearrow h$ pointwise. Clearly,

$$\mathbb{P}\Big[\sup_{t\in(0,T)} |M(t)|\, h(t) > \lambda\Big] \;=\; \lim_{k\to\infty}\lim_{m\to\infty} \mathbb{P}\Big[\sup_{t\in(0,T-T/m)} |M(t)|\, h_k(t) > \lambda\Big].$$

Applying Theorem 2.1 from [28] to the probability on the right yields

$$\mathbb{P}\Big[\sup_{t\in(0,T-T/m)} |M(t)|\, h_k(t) > \lambda\Big] \;\le\; \frac{1}{\lambda}\int_0^{T-T/m} h_k(t)\, dv(t) \;=\; \frac{1}{\lambda}\int_{(0,T-T/m]} h_k(t)\, dv(t).$$

Hence, the assertion follows by sending m and k to infinity. $\square$

To prove Theorem 3.1, let $\hat{F}_n$ and F denote the df of $\hat{m}_{n;\mu}$ and μ, respectively. We can choose a sequence $U_1, U_2, \ldots$ of i.i.d. $U[0,1]$-random variables, possibly on an extension of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that the corresponding empirical df $\hat{G}_n$ satisfies $\hat{F}_n = \hat{G}_n(F(\cdot))$ $\mathbb{P}$-almost surely; cf. [25] or [27, p.103]. Then, for every $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}[d_{(\phi)}(\hat{m}_{n;\mu}, \mu) \ge \delta]\\
&= \mathbb{P}[\|\hat{F}_n - F\|_\phi \ge \delta]\\
&= \mathbb{P}[\|\hat{G}_n(F(\cdot)) - F\|_\phi \ge \delta]\\
&= \mathbb{P}\Big[\sup_{x\in\mathbb{R}} |\hat{G}_n(F(x)) - F(x)|\, \phi(x) \;\ge\; \delta\Big]\\
&\le \mathbb{P}\Big[\sup_{x>0} |\hat{G}_n(F(x)) - F(x)|\, \phi(x) \;\ge\; \delta\Big] + \mathbb{P}\Big[\sup_{x<0} |\hat{G}_n(F(x)) - F(x)|\, \phi(x) \;\ge\; \delta\Big] \qquad (A.2)
\end{aligned}$$

Denoting the left- and rightcontinuous inverse functions of F by $F^{\leftarrow}$ and $F^{\rightarrow}$, we have

$$
\begin{aligned}
\sup_{x<0} |\hat{G}_n(F(x)) - F(x)|\, \phi(x) &\leq \sup_{t\in(0,F(0-))} |\hat{G}_n(t) - t| \phi(F^{\leftarrow}(t)) \\
&= \sup_{t\in(0,F(0-))} \left| \frac{\hat{G}_n(t) - t}{1-t} \right| \phi(F^{\rightarrow}(t))(1-t)
\end{aligned}
$$

where the last equality holds $\mathbb{P}$-almost surely due to the facts that $F^{\rightarrow}(t) = F^{\leftarrow}(t+)$ and that, with probability one, $\hat{G}_n$ does not jump at the countable set of $t \in (0,1)$ for which $F^{\rightarrow}(t) \neq F^{\leftarrow}(t)$. Since

$$
M_n(t) := \frac{\hat{G}_n(t) - t}{1-t}, \qquad t \in [0,1),
$$

is a càdlàg martingale by [26, Proposition 12.11.1] and has $\mathbb{E}[M_n(t)^2] = t/(n(1-t))$ we may apply Theorem A.1 with $h(t) := \phi(F^{\rightarrow}(t))(1-t)$ and obtain

$$
\begin{aligned}
\mathbb{P}\Big[\sup_{x<0} |\hat{G}_n(F(x)) - F(x)|\, \phi(x) \ \geq \ \delta\Big] &\leq \mathbb{P}\Big[\sup_{t\in(0,F(0-))} |M_n(t)| h(t) \ \geq \ \delta\Big] \\
&\leq \frac{1}{\delta} \int\limits_{(0,F(0-))} h(t)\, dv_n(t),
\end{aligned}
$$

where $v_n(t) := \mathbb{E}[M_n^+(t)]$.

To deal with the remaining term in (A.2), note first that

$$
\check{M}_n(t) := \frac{1 - t - \hat{G}_n((1-t)-)}{1-t}, \qquad t \in [0,1)
$$

is also a càdlàg martingale with the same law as M_n, because $1-U_1, 1-U_2, \ldots$ is again an i.i.d. sequence of uniform random variables. Letting $\check{h}(t) := \phi(F^{\leftarrow}(1-t))(1-t)$ and using a similar argument as for the first term, we have

$$
\begin{aligned}
\mathbb{P}\Big[\sup_{x>0} |\hat{G}_n(F(x)) - F(x)|\, \phi(x) \ \geq \ \delta\Big] &\leq \mathbb{P}\Big[\sup_{t\in(0,1-F(0))} |\check{M}_n(t)|\, \check{h}(t) \ \geq \ \delta\Big] \\
&\leq \frac{1}{\delta} \int\limits_{(0,1-F(0))} \check{h}(t)\, d\check{v}_n(t),
\end{aligned}
$$

where

$$\check{v}_n(t) \;:=\; \mathbb{E}[\check{M}_n^+(t)] \;=\; v_n(t) \;=\; \frac{1}{n(1-t)} \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} (k-nt)^+ .$$

Since v_n is absolutely continuous on every interval $[a,b] \subset (0,1)$, we have

$$\delta \cdot \mathbb{P}[d_{(\phi)}(\hat{m}_{n;\mu}, \mu) \geq \delta] \;\leq\; \int\limits_{(0,F(0-))} \phi(F^{\rightarrow}(t))(1-t)v_n'(t)\,dt + \int\limits_{(F(0),1)} \phi(F^{\leftarrow}(t))tv_n'(1-t)\,dt.$$

It follows that for $p > 1$ and $q = p/(1-p)$,

$$\begin{aligned}\delta \cdot \mathbb{P}[d_{(\phi)}(\hat{m}_{n;\mu}, \mu) \geq \delta] &\leq 2\Big(\int \phi^p \, d\mu\Big)^{1/p} \Big(\int\limits_0^1 \big[(1-t)v_n'(t)\big]^q \, dt\Big)^{1/q} \\ &\leq 2\kappa^{1/p} \Big(\int\limits_0^1 \big[(1-t)v_n'(t)\big]^q \, dt\Big)^{1/q},\end{aligned}$$

when $\mu \in \mathcal{M}_{1,\kappa}^{\phi^p}(\mathbb{R})$. The assertion of Theorem 3.1 will now follow from the next lemma.

Lemma A.2 *For any $q \in [1,\infty)$, we have $(1-t)v_n'(t) \to 0$ in $L^q(0,1)$.*

Proof Let $w_n(t) := (1-t)v_n(t)$. Then

$$(1-t)v_n'(t) \;=\; w_n'(t) + \frac{w_n(t)}{1-t}.$$

We have

$$w_n(t) \;=\; \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \Big(\frac{k}{n} - t\Big)^+ \longrightarrow 0$$

for each $t \in (0,1)$. Moreover,

$$0 \;\leq\; \frac{w_n(t)}{1-t} \;=\; \sum_{k=0}^{n} \binom{n}{k} t^n (1-t)^{n-k} \frac{(\frac{k}{n} - t)^+}{1-t} \;\leq\; \sum_{k=0}^{n} \binom{n}{k} t^n (1-t)^{n-k} \mathbb{1}_{\{\frac{k}{n} > t\}} \;\leq\; 1,$$

and hence $w_n(t)/(1-t) \to 0$ in $L^q(0,1)$.

To show that also $w_n' \to 0$ in $L^q(0,1)$, we fix $t \in (0,1) \setminus \mathbb{Q}$. Then

$$
\begin{aligned}
w_n'(t) = &\sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \frac{k}{t} \Big(\frac{k}{n} - t\Big)^+ - \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \frac{n-k}{1-t} \Big(\frac{k}{n} - t\Big)^+ \\
&- \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} \mathbb{1}_{\{\frac{k}{n} - t > 0\}} \\
= &\, n\, \mathbb{E}\Big[\frac{\hat{G}_n(t)}{t} (\hat{G}_n(t) - t)^+ - \frac{1 - \hat{G}_n(t)}{1-t} (\hat{G}_n(t) - t)^+ - \frac{1}{n} \mathbb{1}_{\{\hat{G}_n(t) - t > 0\}} \Big] \\
= &\, \mathbb{E}\Big[Z_n Z_n^+ \Big] - \mathbb{P}[Z_n > 0],
\end{aligned} \tag{A.3}
$$

where

$$
Z_n := \frac{\sqrt{n}(\hat{G}_n(t) - t)}{\sqrt{t(1-t)}}.
$$

Since Z_n converges in law to a standard normal random variable Z, the probability in (A.3) converges to 1/2. Furthermore, $\mathbb{E}\big[Z_n^2\big] = 1 = \mathbb{E}\big[Z^2\big]$. So Lemma 3.4 yields that the laws of Z_n converge even $(1 + x^2)$-weakly to the law of Z. Hence,

$$
\lim_{n \to \infty} \mathbb{E}\Big[Z_n Z_n^+\Big] = \mathbb{E}\Big[Z Z^+\Big] = \mathbb{E}\Big[\mathbb{1}_{\{Z>0\}} Z^2\Big] = 1/2.
$$

Therefore $w_n'(t) \to 0$ for any $t \in (0,1) \setminus \mathbb{Q}$. Since in addition $-1 \leq w_n'(t) \leq \mathbb{E}\Big[Z_n(t)^2\Big] = 1$ we obtain $w_n' \to 0$ in $L^q(0,1)$. $\square$

References

[1] Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9, 203–228.

[2] Bauer, H. (2001) *Measure and integration theory.* de Gruyter, Berlin.

[3] Beutner, E. and Zähle, H. (2010). A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis*, 101, 2452–2463.

[4] Beutner, E. and Zähle, H. (2010). Deriving the asymptotic distribution of U- and V-statistics of dependent data using weighted empirical processes. *Bernoulli*, forthcoming.

[5] Beutner, E., Wu, W.B. and Zähle, H. (2010). Asymptotics for statistical functionals of long-memory sequences. *Submitted.*

[6] Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*, 10, 5931–606.

[7] Cuevas, A. (1988). Qualitative robustness in abstract inference. *Journal of Statistical Planning and Inference*, 18, 277–289.

[8] Denker, M. (1985). *Asymptotic distribution theory in nonparametric statistics.* Fr. Vieweg & Sohn, Braunschweig.

[9] Dudley, R.M. (1969). The speed of mean Glivenko-Cantelli convergence. *Annals of Mathematical Statistics* 40, 40–50.

[10] Dudley, R.M. (2004). Real analysis and probability. Cambridge University Press, Cambridge.

[11] Dvoretzky, A., Kiefer, J., Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics*, 27, 642–669.

[12] Fernholz, L.T. (1983). Von Mises calculus for statistical functionals, *Lecture Notes in Statistics*, 19, Springer-Verlag, New York.

[13] Föllmer, H. and Schied, A. (2011). *Stochastic finance. An introduction in discrete time.* Third edition, de Gruyter, Berlin.

[14] Gill, R.D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I). *Scandinavian Journal of Statistics*, 16, 97–128.

[15] Hampel, F.R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, 42, 1887–1896

[16] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust statistics - the approach based on influence functions*, Wiley, New York.

[17] Huber, P.J. and Ronchetti E.M. (2009). *Robust statistics.* Wiley, New York.

[18] La Vecchia D., Ronchetti E.M. and Trojani F. (2009). Higher-order robustness. *Working Paper*, http://www.icors2009.unipr.it/presentations/la-vecchia.pdf.

[19] Lee, A.J. (1990). *U-Statistics.* Marcel Dekker Inc., New York.

[20] Massart, P. (1992). The tight constant in the Dvoretsky-Kiefer-Wolfowitz inequality. *Annals of Probability*, 18, 1269–1283.

[21] Mizera, I. (2010). Qualitative robustness and weak continuity: the extreme unction. *IMS Collections Festschrift*, Institute of Mathematical Statistics.

[22] Rachev, S.T. (1991). *Probability metrics and the stability of stochastic models.* Wiley, New York.

[23] Rieder, H. (1994). *Robust asymptotic statistics.* Springer-Verlag, New York.

[24] Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*, Wiley, New York.

[25] Shorack, G.R. (1979) The weighted empirical process of row independent random variables with arbitrary d.f.'s. *Statistica Neerlandica*, 33, 169–189.

[26] Shorack, G.R (2000). *Probability for statisticians.* Springer-Verlag, New York.

[27] Shorack, G.R. and Wellner, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

[28] Slud, E. (1987). Generalization of an inequality of Birnbaum and Marshall, with applications to growth rates for submartingales. *Stochastic Processes and their Applications*, 24, 51–60.

[29] Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36, 423–439.

[30] van der Vaart, A.W. (1998). *Asymptotic statistics.* Cambridge University Press, Cambridge.

[31] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

[32] Zähle, H. (2010). Rates of almost sure convergence of plug-in estimates for distortion risk measures. *Metrika*, forthcoming, doi: 10.1007/s00184-010-0302-z.

[33] Zähle, H. (2011). Marcinkiewicz-Zygmund strong laws for statistical functionals and empirical distribution functions. *Submitted.*