Journal of Multivariate Analysis 158 (2017) 1–19

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Domains of weak continuity of statistical functionals with a view toward robust statistics

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

[a] Faculty of Mathematics, University of Duisburg–Essen, Germany

[b] Department of Statistics and Actuarial Science, University of Waterloo, Canada

[c] Department of Mathematics, Saarland University, Germany

ARTICLE INFO

Article history:
Received 20 June 2016
Available online 6 March 2017

AMS subject classifications:
62F10
62F35
62G05
62G35
62P05

Keywords:
(ψ_k)-weak topology
w-set
Qualitative robustness
Hampel's theorem
Maximum likelihood estimator
Law-invariant convex risk measure
Aggregation robustness
Orlicz space

ABSTRACT

Many standard estimators such as several maximum likelihood estimators or the empirical estimator for any law-invariant convex risk measure are not (qualitatively) robust in the classical sense. However, these estimators may nevertheless satisfy a weak robustness property (Krätschmer et al. (2012, 2014)) or a local robustness property (Zähle (2016)) on relevant sets of distributions. One aim of our paper is to identify sets of local robustness, and to explain the benefit of the knowledge of such sets. For instance, we will be able to demonstrate that many maximum likelihood estimators are robust on their natural parametric domains. A second aim consists in extending the general theory of robust estimation to our local framework. In particular we provide a corresponding Hampel-type theorem linking local robustness of a plug-in estimator with a certain continuity condition.

1. Introduction and problem statement

Recently, in [22] qualitative robustness of plug-in estimators was considered as a local property, i.e., on strict subsets of the natural domain of the corresponding statistical functional, and a respective Hampel-type criterion was proven. The latter says that if the statistical functional is continuous for a certain topology finer than the weak topology, then qualitative robustness holds on every set of distributions on which the relative weak topology coincides with the finer topology. Such sets of distributions were characterized in [22], but the provided characterization is rather technical and not at all useful for checking the concurrence of the topologies for any given set. The aim of the present paper is to provide more useful characterizations of such sets, and to illustrate their use in the context of qualitative robustness. Compared to [22] we will also allow for more general topologies on sets of distributions which will turn out to increase the flexibility to check qualitative robustness for statistical functionals. As applications, robustness of maximum likelihood estimators and of empirical estimators of law-invariant convex risk measures are studied in detail. In particular we will demonstrate that

* Corresponding author.
E-mail addresses: volker.kraetschmer@uni-due.de (V. Krätschmer), alex.schied@gmail.com (A. Schied), zaehle@math.uni-sb.de (H. Zähle).

http://dx.doi.org/10.1016/j.jmva.2017.02.005
0047-259X/

many maximum likelihood estimators are robust on their natural parametric domains and even on broader sets. A further field of application is quantitative risk management. In recent contributions in this field the property of robustness has been pointed out as an important requirement for risk assessment; see, for instance, [3,6,14]. Again the empirical estimators of well-founded statistical functionals like those associated with law-invariant convex risk measures fail to be robust but might satisfy this property on domains of interest.

To explain our intension more precisely, let E be a Polish space and $\mathcal{M}_1$ be the set of all Borel probability measures on E. Consider the statistical model

$$(\Omega, \mathcal{F}, \{\mathrm{P}^\theta : \theta \in \Theta\}) = (E^{\mathbb{N}}, \mathcal{B}(E)^{\otimes\mathbb{N}}, \{\mathrm{P}^\mu : \mu \in \mathcal{M}\}), \tag{1}$$

where $\mathcal{M} \subseteq \mathcal{M}_1$ is any set of Borel probability measures on E and

$$\mathrm{P}^\mu \equiv \mu^{\otimes\mathbb{N}}$$

is the infinite product measure of μ. Note that the coordinate projections on $E^{\mathbb{N}}$ are i.i.d. with law μ under P^μ. For every $\boldsymbol{x} = (x_1, x_2, \ldots) \in E^{\mathbb{N}}$ and $n \in \mathbb{N}$, we define the empirical probability measure $\widehat{m}_n(\boldsymbol{x})$ by

$$\widehat{m}_n(\boldsymbol{x}) = \widehat{m}_n(x_1, \ldots, x_n) \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}.$$

Assume that $\mathcal{M}$ contains the set

$$\mathfrak{E} = \{\widehat{m}_n(x_1, \ldots, x_n) : x_1, \ldots, x_n \in E,\ n \in \mathbb{N}\}$$

of all empirical probability measures. Let (Σ, d_Σ) be a complete and separable metric space and $T : \mathcal{M} \to \Sigma$ be any map (statistical functional). The empirical probability measure $\widehat{m}_n$ induces a nonparametric estimator $\widehat{T}_n : \Omega \to \Sigma$ for $T(\mu)$ in the statistical model (1) through

$$\widehat{T}_n(\boldsymbol{x}) = T(\widehat{m}_n(x)), \quad \boldsymbol{x} = (x_1, x_2, \ldots) \in \Omega, \tag{2}$$

provided $\widehat{T}_n$ is $(\mathcal{F}, \mathcal{B}(\Sigma))$-measurable.

The following Definition 1.1 generalizes Hampel's classical notion of (qualitative) robustness for the sequence $(\widehat{T}_n)$ as introduced in [9]. Recall from Theorem 2.14 in [10] that the set of all Borel probability measures on Σ equipped with the weak topology is Polish and can be metrized by the Prohorov metric π. Moreover denote by $\mathcal{O}_{\mathrm{w}}$ the weak topology on $\mathcal{M}_1$.

Definition 1.1. For a given set $M \subseteq \mathcal{M}$ and $\mu \in M$, the sequence of estimators $(\widehat{T}_n)$ is said to be M-*robust at* μ if for every $\varepsilon > 0$ there exists an open neighborhood $U = U(\mu, \varepsilon; M)$ of μ for the relative weak topology $\mathcal{O}_{\mathrm{w}} \cap M$ such that

$$\nu \in U \quad \Longrightarrow \quad \pi(\mathrm{P}^\mu \circ \widehat{T}_n^{-1}, \mathrm{P}^\nu \circ \widehat{T}_n^{-1}) \le \varepsilon \quad \text{for all } n \in \mathbb{N}.$$

The sequence $(\widehat{T}_n)$ is said to be *robust on* M if it is M-robust at every $\mu \in M$.

In their pioneer work, Hampel [9] and Cuevas [4] used (mainly the first part of) Definition 1.1 with specifically $M = \mathcal{M} = \mathcal{M}_1$ and established several criteria for robustness; see Theorems 1–2 in [9] and Theorems 1–2 in [4]. In the present paper, our focus will be on the second part of Definition 1.1, i.e., on robustness of $(\widehat{T}_n)$ on subsets M of $\mathcal{M}$. In this context the following two criteria are already known for $M = \mathcal{M}$.

(I) If $T : \mathcal{M} \to \Sigma$ is continuous for the relative weak topology $\mathcal{O}_{\mathrm{w}} \cap \mathcal{M}$, then $(\widehat{T}_n)$ is robust on $\mathcal{M}$.
(II) If $(\widehat{T}_n)$ is weakly consistent and robust on $\mathcal{M}$, then $T : \mathcal{M} \to \Sigma$ is continuous for the relative weak topology $\mathcal{O}_{\mathrm{w}} \cap \mathcal{M}$.

Assertion (I) is a straightforward generalization of Theorem 2 in [4] (where the author assumed $\mathcal{M} = \mathcal{M}_1$) and assertion (II) is a special case of Theorem 1 in [4].

Recall that we assumed the set $\mathfrak{E}$ of all empirical probability measures to be contained in $\mathcal{M}$. As $\mathfrak{E}$ is dense in $\mathcal{M}_1$ with respect to the weak topology $\mathcal{O}_{\mathrm{w}}$ (see Theorem A.38 in [8] reformulated for probability measures), this implies that weak continuity of the map $T : \mathcal{M} \to \Sigma$ is a relatively strict requirement. For instance, in the case $E = \mathbb{R}$ the mean functional $T(\mu) = \int x\,\mu(dx)$ is not weakly continuous on $\mathfrak{E}$ (indeed, letting $x_{n,1} = n$ and $x_{n,i} = 0$ for $i = 2, \ldots, n$ and $n \in \mathbb{N}$, the sequence $(\widehat{m}_n(x_{n1}, \ldots, x_{nn}))_{n\in\mathbb{N}}$ converges to δ_0 with respect to $\mathcal{O}_{\mathrm{w}}$, but $\int x\,\widehat{m}(x_{n1}, \ldots, x_{nn})(dx) = 1 \not\to 0 \int x\,\delta_0(dx)$). In view of (I)–(II), this simple example indicates that there are only a few relevant statistical functionals $T : \mathcal{M} \to \Sigma$ for which the corresponding sequence of estimators $(\widehat{T}_n)$ is robust on the whole domain $\mathcal{M}$. Nevertheless, for general statistical functionals one might ask for those subsets M of $\mathcal{M}$ on which robustness of $(\widehat{T}_n)$ holds. The following simple example shows that this question can be reasonable.

Example 1.2. Let $E = (0, \infty)$ and $\mathcal{E}$ be the class of all exponential distributions with mean θ (see Example 3.6), $\theta \in (0, \infty)$. The unique maximum likelihood estimator for the parameter θ is known to be $\widehat{T}_n(\boldsymbol{x}) = \overline{\boldsymbol{x}}_n$, where $\overline{\boldsymbol{x}}_n = \frac{1}{n}\sum_{i=1}^n x_i$ for $\boldsymbol{x} = (x_1, x_2, \ldots)$. It can be represented by $\widehat{T}_n(\boldsymbol{x}) = T(\widehat{m}_n(\boldsymbol{x}))$ for the functional $T(\mu) = \int x\,\mu(dx)$ with domain $\mathcal{M} = \{\mu \in \mathcal{M}_1 : \int |x|\,\mu(dx) < \infty\}$ and state space $\Sigma = (0, \infty)$. Since T is weakly consistent by the law of large numbers but *not* weakly continuous on $\mathcal{M}$, assertions (I)–(II) imply that the sequence $(\widehat{T}_n)$ is not robust on $\mathcal{M}$. However,