Journal of Multivariate Analysis 143 (2016) 12–31

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A definition of qualitative robustness for general point estimators, and examples

Henryk Zähle

Saarland University, Department of Mathematics, Postfach 151150, D–66041 Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 30 June 2014
Available online 8 September 2015

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

Keywords:
Qualitative robustness
Dominated statistical model
Nonparametric statistical model
Strong mixing
Weak topology
ψ-weak topology
Linear process

ABSTRACT

A definition of qualitative robustness for point estimators in general statistical models is proposed. Some criteria for robustness are established and applied to estimators in parametric, semiparametric, and nonparametric models. In specific nonparametric models, the proposed definition boils down to Hampel robustness. It is also explained how plug-in estimators in certain nonparametric models can be reasonably classified w.r.t. their degrees of robustness.

1. Introduction

Let (Θ, d_Θ) be a metric space, where Θ will be regarded as a parameter space. Let $(\Omega, \mathcal{F})$ be a measurable space, and $\mathbb{P}^\theta$ be any probability measure on $(\Omega, \mathcal{F})$ for every $\theta \in \Theta$. The set Ω can be seen as the sample space, where the sample is drawn from $\mathbb{P}^\theta$ with (unknown) $\theta \in \Theta$. As usual, the triplet $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta : \theta \in \Theta\})$ will be referred to as statistical model. Further, let $(\Sigma, \mathcal{S})$ be a measurable space and for every $n \in \mathbb{N}$ let $T_n : \Theta \to \Sigma$ be any map, where T_n and Σ can be regarded as an aspect function and the state space of the aspect function, respectively. For every $n \in \mathbb{N}$, let $\widehat{T}_n : \Omega \to \Sigma$ be any $(\mathcal{F}, \mathcal{S})$-measurable map, which can be seen as an estimator for the aspect $T_n(\theta)$ of θ. Often the sample space and the estimator can be written as

$$(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}}) \quad \text{and} \quad \widehat{T}_n(x) = \widehat{T}_n(x_1, \ldots, x_n) \quad \text{for all } x = (x_1, x_2, \ldots) \in \Omega \tag{1}$$

for some measurable space $(E, \mathcal{E})$, which is virtually the standard statistical setting, but this particular form will not be assumed here. Finally, let ρ be any metric on the set $\mathcal{M}_1(\Sigma)$ of all probability measures on $(\Sigma, \mathcal{S})$.

The following definition proposes a notion of (qualitative) robustness for the sequence of estimators $(\widehat{T}_n)$ which is in line with Hampel's notion of (qualitative) robustness. Note that the aspect functions T_n, $n \in \mathbb{N}$, do not play any role in the definition. They will only occur again in Section 2.

Definition 1.1. For any subset $\Theta_0 \subset \Theta$ we use the following terminology.

(i) The sequence $(\widehat{T}_n)$ is said to be (d_Θ, ρ)-robust on Θ_0 if for every $\theta_1 \in \Theta_0$ and $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\theta_2 \in \Theta_0, \quad d_\Theta(\theta_1, \theta_2) \le \delta \Longrightarrow \rho(\mathbb{P}^{\theta_1} \circ \widehat{T}_n^{-1}, \mathbb{P}^{\theta_2} \circ \widehat{T}_n^{-1}) \le \varepsilon \quad \text{for all } n \in \mathbb{N}. \tag{2}$$

E-mail address: zaehle@math.uni-sb.de.

http://dx.doi.org/10.1016/j.jmva.2015.08.004
0047-259X/

(ii) The sequence $(\widehat{T}_n)$ is said to be asymptotically (d_Θ, ρ)-robust on Θ_0 if for every $\theta_1 \in \Theta_0$ and $\varepsilon > 0$ there are some $\delta > 0$ and $n_0 \in \mathbb{N}$ such that

$$\theta_2 \in \Theta_0, \quad d_\Theta(\theta_1, \theta_2) \le \delta \Longrightarrow \rho(\mathbb{P}^{\theta_1} \circ \widehat{T}_n^{-1}, \mathbb{P}^{\theta_2} \circ \widehat{T}_n^{-1}) \le \varepsilon \quad \text{for all } n \ge n_0. \tag{3}$$

(iii) The sequence $(\widehat{T}_n)$ is said to be finite-sample (d_Θ, ρ)-robust on Θ_0 if for every $\theta_1 \in \Theta_0$, $\varepsilon > 0$, and $n_0 \in \mathbb{N}$ there is some $\delta > 0$ such that

$$\theta_2 \in \Theta_0, \quad d_\Theta(\theta_1, \theta_2) \le \delta \Longrightarrow \rho(\mathbb{P}^{\theta_1} \circ \widehat{T}_n^{-1}, \mathbb{P}^{\theta_2} \circ \widehat{T}_n^{-1}) \le \varepsilon \quad \text{for all } 1 \le n \le n_0. \tag{4}$$

On the one hand, Definition 1.1 is close to Hampel's definition of robustness in the context of nonparametric statistical models. Indeed, letting specifically $\Theta := \Theta_0 := \mathcal{M}_1(E)$ be the set of all probability measures on some complete and separable metric space E (equipped with any metric d_Θ generating the weak topology), $(\Omega, \mathcal{F})$ be as in (1), $\widehat{T}_n$ be as in (1) and invariant against permutations of the arguments, $\mathbb{P}^\mu := \mu^{\otimes\mathbb{N}}$ for all $\mu \in \Theta$, $\Sigma := \mathbb{R}^d$, and ρ be the Prohorov metric, then part (i) of Definition 1.1 coincides with the definition of robustness as given in Section 4 of [14]. Cuevas [9] put forward Hampel's nonparametric theory by replacing $\mathbb{R}^d$ by a general complete and separable metric space Σ. Krätschmer et al. [16,17] considered metrics that metrize finer topologies than the weak topology, and Zähle [26] allowed for laws $\mathbb{P}^\mu$ that are not necessarily infinite product measures (for a different approach for nonparametric estimators based on dependent observations, see [3,6,8,20,24]). The distinction between asymptotic and finite-sample robustness was implicitly also done in [9,14]. Huber [15] and other authors (e.g. [16,17,19]) regraded robustness simply as asymptotic robustness. Examples for robust estimators in nonparametric statistical models range from sample trimmed means [14] to L-estimators [15] to Z- and M-estimators [14,15] to R-estimators [15] to support vector machines [13].

On the other hand, Definition 1.1 allows for more statistical models $(\Omega, \mathcal{F}, \{\mathbb{P}^\theta : \theta \in \Theta\})$ than the one just discussed. In many classical examples of the theory of point estimation the parameter space Θ is a subset of $\mathbb{R}^k$ (and not the measure space $\mathcal{M}_1(E)$). The underlying statistical model has indeed often the shape $(E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}}, \{\mathbb{P}^\theta : \theta \in \Theta\})$ for some subset $\Theta \subset \mathbb{R}^k$. If $\mathbb{P}^\theta = \mu_\theta^{\otimes\mathbb{N}}$ for some $\mu_\theta \in \mathcal{M}_1(E)$, $\theta \in \Theta$, then this model corresponds to the standard situation where one can observe the realizations of i.i.d. E-valued random variables with distribution μ_θ but the true k-dimensional parameter vector θ is unknown; this setting is known as infinite product model. Robustness of the distribution of a given estimator w.r.t. small changes of the underlying model associated with θ is an obvious quality criterion, but it is not unique what the "right" notion of robustness is. In the mentioned infinite product model, for instance, a "change" of the underlying model can be measured in at least two ways. First, one may measure a change of θ w.r.t. the Euclidean distance on Θ. Second, one may measure a change of the probability measure μ_θ w.r.t. any metric on $\{\mu_\theta : \theta \in \Theta\}$ which metrizes the relative weak topology. The former approach is not covered by Hampel's theory, but it is covered by Definition 1.1 and seems to be more natural in the context of classical parametric models (as, for instance, the Gaussian model where $\mu_\theta := \mathrm{N}_{m,s^2}$ for $\theta = (m, s^2) \in \Theta := \mathbb{R} \times (0, \infty)$). The latter approach basically leads to a version of Hampel's definition when regarding $\widetilde{\Theta} := \{\mu_\theta : \theta \in \Theta\}$ as the parameter space. But strictly speaking this approach is neither covered by the existing literature due to the traditional assumption $\widetilde{\Theta} = \mathcal{M}_1(E)$. Definition 1.1, on the other hand, is more flexible and makes the second approach possible too.

Apart from the situation where $\Theta \subset \mathbb{R}^k$ ("parametric model"), the parameter space Θ is often the product of a subset of $\mathbb{R}^k$ and a subset of $\mathcal{M}_1(E)$ ("semiparametric model"). This is the case, for instance, in some parametric regression models, ARMA models, and so on. Then a change of the underlying model should be measured by any metric which metrizes the product topology on Θ. In this situation the classical definition of robustness does not apply again, but Definition 1.1 does.

The preceding discussion shows that Definition 1.1 is suitable not only for nonparametric statistical models but also for parametric and semiparametric statistical models. In this sense, this article treats a rather general setting and facilitates more examples than the existing literature on robustness in nonparametric statistical models.

The article is organized as follows. Section 2 provides some criteria for asymptotic and finite-sample robustness in the fashion of the celebrated Hampel theorem. Section 3 is devoted to examples, and Section 4 provides the proofs of the results of Section 2. In Section 3.1, we investigate plug-in estimators in *nonparametric* statistical models being more general compared to [9,14,16,26], and we classify plug-in estimators on Euclidean spaces w.r.t. their degrees of robustness. Section 3.2 provides results on robustness for estimators in dominated *parametric* statistical models, and Section 3.3 is devoted to robustness of a Yule–Walker-type estimator in the *semiparametric* statistical model of a linear process. The Introduction will be completed with some basic remarks on Definition 1.1.

Remark 1.2. When the metric ρ is fixed, then (d_Θ, ρ)-robustness of $(\widehat{T}_n)$ is clearly equivalent to (d'_Θ, ρ)-robustness of $(\widehat{T}_n)$ for any other metric d'_Θ which is equivalent to d_Θ. ◇

Remark 1.3. Of course, the sequence $(\widehat{T}_n)$ is robust on Θ_0 if and only if it is both asymptotically and finite-sample robust on Θ_0, and finite-sample robustness already holds when in (4) the phrase "for all $1 \le n \le n_0$" is replaced by "for $n = n_0$". Moreover, (d_Θ, ρ)-robustness of $(\widehat{T}_n)$ on Θ_0 means that the set of mappings

$$\left\{\Theta \longrightarrow \mathcal{M}_1(\Sigma),\ \theta \longmapsto \mathbb{P}^\theta \circ \widehat{T}_n^{-1}\ :\ n \in \mathbb{N}\right\}$$

is (d_Θ, ρ)-equicontinuous on Θ_0. ◇