Estimation using the disparity measure related to the robust indentification property

C. Agostinelli[1]

[1] Dipartimento di Statistica
Università Ca' Foscari
San Giobbe, Cannaregio, 873
I–30121 Venezia, Italy

Keywords: Identification property, Disparity measure, Robust estimation

Long Abstract

We introduce a Robust Estimator based on a disparity measure which is motivated by the *Robust identification property* (Agostinelli, 2004).

For sake of simplicity, hereafter we present some results under the assumption that each distribution function has density with respect to some common dominating measure. The results are extendible to distribution function without density.

Let $\mathcal{F}$ the set of all distribution function with density with respect to a measure ν. Let F, G and H three distribution functions belonging to $\mathcal{F}$ and f, g and h their corresponding densities then

$$\delta(F,G) = \inf\{0 \leq \varepsilon \leq 1/2 : \exists h(x) \text{ so that } (1-\varepsilon)g(x) \leq (1-\varepsilon)f(x) + \varepsilon h(x) \text{ a.s. } \nu\}$$

is a distance between F and G.

Two distribution F and G are said robust indetified at given level ε if $\delta(F,G) > \varepsilon$. This definition is particular interest when applied to a parametric family in the following way. Let $\mathcal{M} = \{M_\theta; \theta \in \Theta\}$ be a parametric family such that $\mathcal{M} \subseteq \mathcal{F}$ then the set

$$\Theta_o(\theta_a;\varepsilon) = \{\theta_b \in \Theta : \delta(M_{\theta_a}, M_{\theta_b}) \leq \varepsilon\}$$

is the values of $\theta \in \Theta$ which can not be distinguish from θ_a given a dataset of any size contaminated by at least a fraction of ε observations.

The distance is nicely related to the Variation Distance by the Scheffe's Theorem (see for instance, He and Simpson, 1993) in the sense they metrize the same topology.

The Robust estimation we are going to study is based on $\delta(F,G)$ but in a slight different way

$$\gamma(F, M_\theta) = \inf\{0 \leq \varepsilon \leq 1/2 : (1-\varepsilon)m(x;\theta) \leq f(x) \text{ a.s. } \nu\}$$

and the estimator $\hat{\theta}$ is

$$\hat{\theta} = \text{argmin}_{\theta\in\Theta}\gamma(\hat{F}, M_\theta)$$

where $\hat{F}$ is a smoothed version of the empirical cumulative distribution from the sample.

In the paper we will show that the Influence Function of this estimator is equal to zero for all the parametric family with density with respect to the Lebesque measure. Its asymptotic maxbias (maxbias$(\hat{\theta}, \theta_a, \varepsilon)$) would be related to the set $\Theta_o(\theta_a, \varepsilon)$ (and hence to the Robust Identification Property of the parametric family) by the formula

$$\text{maxbias}(\hat{\theta}, \theta_a, \varepsilon) = \max_{\theta\in\Theta_o(\theta_a;\varepsilon)} |\theta - \theta_a|$$

In the case of a normal location family we have that the maxbias of this estimator is twice the maxbias of the median for every level of contamination.

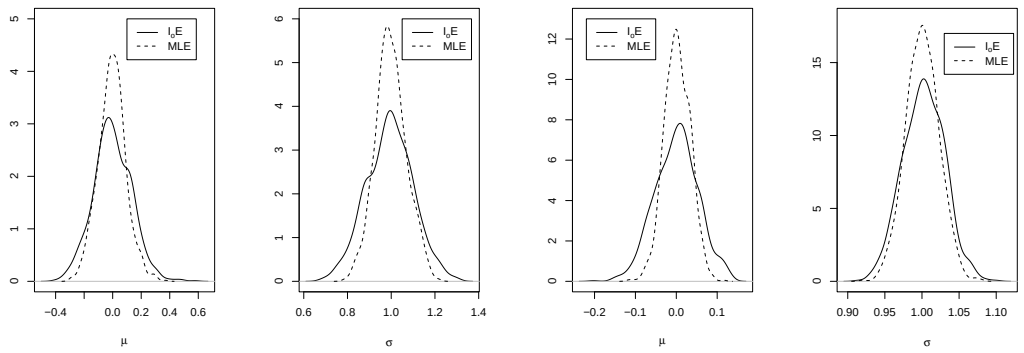

FIGURE 1. Non–parametric kernel density estimator for $\hat{\theta}$ and MLE for normal location and scale parametric family. Left, size=100, and right size=1000.

Despite of the behavior of its Influence Function this estimator seems to have efficiency comparable to Maximum Likelihood Estimator (MLE) as show by the limited Monte Carlo simulation presented in figure 1 run for the normal location and scale parametric family with samples from standard normal of size 100 and 1000.

In the following tables we report the results of a Monte Carlo simulation performed for two contaminated models: $(1-\varepsilon)N(0,1)+\varepsilon N(0,\sigma=5)$ (symmetric contamination) and $(1-\varepsilon)N(0,1)+\varepsilon N(8,1)$ (asymmetric contamination) with $\varepsilon = 0, 5\%, 10\%, 20\%, 30\%, 40\%$ and sample size 100. We run 1000 replication and we report, for each entry, its average and the estimated monte carlo standard error. The first three columns are the results for the introduced method, the third column is an estimate of the proportion of the "good" observations, the last two columns refers to the MLE. The simulation confirm the robust properties of the introduced estimator.

%	$\hat{\mu}$		$\hat{\sigma}$		$\hat{\xi}$		$\hat{\mu}_{MLE}$		$\hat{\sigma}_{MLE}$	
0	-0.001	(0.135)	0.993	(0.107)	0.875	(0.049)	-0.004	(0.097)	0.998	(0.071)
5	-0.005	(0.139)	1.002	(0.112)	0.846	(0.049)	-0.001	(0.147)	1.460	(0.263)
10	0.003	(0.144)	1.012	(0.120)	0.817	(0.049)	-0.006	(0.179)	1.825	(0.302)
20	-0.001	(0.157)	1.035	(0.132)	0.759	(0.051)	-0.006	(0.238)	2.387	(0.335)
30	0.000	(0.171)	1.066	(0.158)	0.703	(0.052)	-0.011	(0.278)	2.842	(0.345)
40	0.003	(0.196)	1.113	(0.186)	0.647	(0.053)	-0.017	(0.311)	3.235	(0.346)

Monte Carlo simulations results for the Normal distribution, size 100, symmetric contamination.

%	$\hat{\mu}$		$\hat{\sigma}$		$\hat{\xi}$		$\hat{\mu}_{MLE}$		$\hat{\sigma}_{MLE}$	
0	-0.001	(0.135)	0.993	(0.107)	0.875	(0.049)	-0.004	(0.097)	0.998	(0.071)
5	-0.005	(0.139)	0.985	(0.111)	0.827	(0.049)	0.395	(0.097)	2.020	(0.097)
10	-0.001	(0.142)	0.984	(0.116)	0.783	(0.047)	0.795	(0.097)	2.610	(0.096)
20	-0.001	(0.155)	0.981	(0.124)	0.692	(0.045)	1.596	(0.097)	3.368	(0.098)
30	0.002	(0.167)	0.975	(0.138)	0.602	(0.041)	2.396	(0.097)	3.816	(0.097)
40	0.001	(0.186)	0.974	(0.154)	0.513	(0.039)	3.196	(0.097)	4.060	(0.097)

Monte Carlo simulations results for the Normal distribution, size 100, asymmetric contamination.

References

C. Agostinelli (2004). Robust Identification Property. *Working Paper, 2004.1*, Dipartimento di Statistica, Università Ca' Foscari, Venezia.

X. He and D.G. Simpson (1993). Lower Bounds for Contamination Bias: Globally Minimax Versus Locally Linear Estimation. *Annals of Statistics*, **21**, 1, 314–337.