Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 345 (2007) 353–358
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics

Minimum Hellinger distance estimators for some multivariate models: influence functions and breakdown point results

Aida Toma

Mathematics Department, Academy of Economic Studies, Piaţa Romană 6, Bucharest, Romania

Received 12 January 2007; accepted after revision 23 July 2007

Available online 30 August 2007

Presented by Paul Deheuvels

Abstract

In this Note we study robustness properties of the minimum Hellinger distance estimators (MHDE) for location and covariance in the case of some multivariate distributions. We determine the general form of the influence function of the MHDE and establish that, at the model, this is the same with the influence function of the maximum likelihood estimator. We also prove that, in the hypothesis of a worst possible choice of contamination, the asymptotic breakdown point of the MHDE is larger than $1/2$ at the model. ***To cite this article: A. Toma, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Résumé

Estimateurs de distance Hellinger minimum pour quelques modèles multivariées : fonctions d'influence et le point de rupture. Dans cette Note, nous étudions des propriétés de robustesse des estimateurs du minimum de la distance de Hellinger (MHDE) pour location et covariance dans le cas de quelques modèles multivariées. Nous déterminons la forme générale de la fonction d'influence du MHDE et établissons que, au modèle, elle est la même que la fonction d'influence de l'estimateur du maximum de vraisemblance. Nous démontrons également que, dans l'hypothèse du choix de la contamination la plus mauvaise, le point de rupture asymptotique du MHDE est superiéur à $1/2$ au modèle. ***Pour citer cet article : A. Toma, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Version française abrégée

Dans cette Note, nous étudions des propriétés de robustesse des estimateurs du minimum de la distance de Hellinger (MHDE) pour location et covariance dans le cas de quelques modèles multivariées. Les distributions auxquelles nous nous référons particulièrement sont des distributions elliptiquement symétriques et les distributions du système multivariées de Johnson. En utilisant la fonction résiduelle d'ajustement (RAF) nous déterminons la forme générale de la fonction d'influence du MHDE et montrons que, au modèle, elle est la même avec la fonction d'influence de l'estimateur du maximum de vraisemblance. Nous démontrons également que, dans l'hypothèse du choix de la contamination la plus mauvaise, le point de rupture asymptotique du MHDE est superiéur à $1/2$ au modèle.

E-mail address: aida_toma@yahoo.com.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.07.024

1. Introduction and notations

There are two fundamental properties in parametric estimation: the efficiency when the model is correctly chosen and the robustness when the true distribution of the observations is in a neighborhood of the model, but not necessarily in it. The popularity of the minimum Hellinger distance and related methods is due to the ability to combine these two properties. Among others, Beran [3], Tamura and Boos [10], Simpson [8,9] have proved that the simultaneous goals of efficiency and robustness can be achieved by using parametric procedures based on the Hellinger distance. Lindsay [6] generalized this type of procedures, by using disparities instead of the Hellinger distance, generating estimators that are both first order efficient and robust (see also Basu and Lindsay [1], Basu et al. [2], Park and Basu [7], Broniatowski [4], Broniatowski and Keziou [5] for related methods).

In this Note we bring into discussion robustness properties of the minimum Hellinger distance estimators (MHDE) for location and covariance in the case of some multivariate distributions. The distributions we specially refer are elliptically symmetric distributions and the distributions from the multivariate Johnson system. For these distributions, theoretical results concerning the existence, consistency, asymptotic normality, the empirical breakdown point of the corresponding MHDE, as well as the practical importance of these estimators are demonstrated in Tamura and Boos [10] and Toma [11]. Here we complete the robustness study of these MHDE by using the so called residual adjustment function (RAF). We determine the general form of the influence function of the MHDE and establish that, at the model, this equals to the influence function of the maximum likelihood estimator. We also prove that, in the hypothesis of a worst possible choice of contamination, the asymptotic breakdown point of the MHDE is larger than $1/2$ at the model.

Consider the parametric family $\{f_\theta\colon \theta \in \Theta = \mathbb{R}^p \times \mathrm{SPD}(p)\}$, where $\mathrm{SPD}(p)$ is the set of all $p \times p$ symmetric and positive definite matrices. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random sample of p-vectors and let $\hat{g}_n$ be a nonparametric kernel density estimator computed from $\mathbf{X}$. The MHDE based on $\mathbf{X}$ and $\hat{g}_n$ is any value $\theta = (\mu, V) \in \Theta$ defined by

$$\hat{\theta}_n = (\hat{\mu}_n, \widehat{V}_n) = \arg\inf_{\theta\in\Theta} d_H(\hat{g}_n, f_\theta) \tag{1}$$

where $d_H(g, f) = [\int \{g^{\frac{1}{2}}(x) - f^{\frac{1}{2}}(x)\}^2\,\mathrm{d}x]^{\frac{1}{2}}$ is the Hellinger distance. The MHDE could be also obtained as

$$\hat{\theta}_n = \arg\inf_{\theta\in\Theta} \phi_H(\hat{g}_n, f_\theta) \tag{2}$$

where ϕ_H is the twice squared Hellinger distance, namely $\phi_H(g, f) = 2d_H^2(g, f)$. The function ϕ_H is known to be a divergence and its use in this context enables us to obtain some results concerning the robustness properties of the MHDE. Define the Pearson residual at a point x as $\delta(x) = \frac{\hat{g}_n(x) - f_\theta(x)}{f_\theta(x)}$. Then the Hellinger divergence between $\hat{g}_n$ and f_θ could be written as $\phi_H(\hat{g}_n, f_\theta) = \int C(\delta) f_\theta$ where $C(\delta) = 2[(\delta+1)^{1/2} - 1]^2$ (we will sometime use the notation $\int f$ for $\int f(x)\,\mathrm{d}x$ whenever it is defined).

Next we specify some notations. For any $p \times p$ symmetric matrix A, let $\mathrm{uvec}A$ denote the $p(p+1)/2$ dimensional column vector formed from the elements in the upper triangular half of A, including the diagonal elements. In order to use the gradient and to derive the influence function of the MHDE, we will identify the parameter θ with $(\mu^t, (\mathrm{uvec}V)^t)^t$, where the superscript t denotes the transpose. For any real valued function $f(x)$, let $f'(x)$ denote its first derivative with respect to x. Let ∇ represent the gradient with respect to θ and let $\nabla^2 f_\theta(x)$ denotes the matrix of second partial derivatives of f_θ. Let $s(x, \theta) = \nabla \log f_\theta(x)$ be the $p + p(p+1)/2$ dimensional vector of likelihood score function. Then $I_\theta = \int s(x,\theta)s(x,\theta)^t f_\theta(x)\,\mathrm{d}x$ is the Fisher information matrix.

Under some regularity conditions, the minimum Hellinger divergence estimating equation becomes

$$-\nabla\phi_H = \int A(\delta)\nabla f_\theta = 0 \tag{3}$$

where $A(\delta) = (\delta+1)C'(\delta) - C(\delta)$. The function $A(\delta)$ is the so called residual adjustment function (RAF) of the divergence introduced by Lindsay [6] and determine some specific properties of the estimator, for example how strongly the effect of large probabilistic outliers (which manifest themselves as large positive values of δ) is downweighted. In our case, the expression of the RAF is

$$A(\delta) = 2\big[(\delta+1)^{1/2} - 1\big]. \tag{4}$$

Let $\mathcal{G}$ be the space of densities. Define the functional $T : \mathcal{G} \to \Theta$ corresponding to the MHDE estimator as $T(g) = (\mu(g), V(g)) = \arg\inf_{\theta\in\Theta} \phi_H(g, f_\theta)$. Since $T(f_\theta) = \theta$ for all $\theta \in \Theta$, the functional T is Fisher consistent. From the definition (2) we observe that the MHDE has the functional form $T(\hat{g}_n)$, where $\hat{g}_n$ is the kernel density estimator based on the data. For a given density g, $T(g)$ is a solution of the equation having the structural form (3), with δ obtained from the Pearson residual by replacing $\hat{g}_n$ with g. More precisely, the estimating equation for $T(g)$ is

$$\int A\left(\frac{g(y) - f_{T(g)}(y)}{f_{T(g)}(y)}\right)\nabla f_{T(g)}(y)\,\mathrm{d}y = 0. \tag{5}$$

This implicit equation defining $T(g)$ will be proved to be useful in order to determine the influence function corresponding to the MHDE.

2. Influence functions

In this section we determine the influence function of the MHDE. We establish that the maximum likelihood estimator and MHDE have the same influence functions at the model f_θ.

If x is from the sample space, let $\delta_x(y)$ be the indicator function for x. Let $g_{\varepsilon x}(y) = (1 - \varepsilon)g(y) + \varepsilon\delta_x(y)$ be an ε-contaminated version of the density $g(y)$. If T is a functional on the space of densities, then its influence function is defined to be $IF(x; T, g) = \lim_{\varepsilon\to 0} \frac{T(g_{\varepsilon x})-T(g)}{\varepsilon}$.

In the following, for a given density g, we suppose that $T(g)$ there exists and we identify $T(g)$ with $(\mu(g)^t, (\mathrm{uvec}V(g))^t)^t$. We also suppose that for the density g, the corresponding influence function $IF(x; T, g)$ there exists.

Theorem 2.1. *The influence function of the functional T corresponding to the MHDE is $IF(x; T, g) = \mathrm{DEN}^{-1}\mathrm{NUM}$, where*

$$\mathrm{NUM} = \left(\frac{g(x)}{f_{T(g)}(x)}\right)^{-\frac{1}{2}} s\big(x, T(g)\big) - E_G\left[\left(\frac{g(X)}{f_{T(g)}(X)}\right)^{-\frac{1}{2}} s\big(X, T(g)\big)\right]$$

$$\mathrm{DEN} = E_G\left[\left(\frac{g(X)}{f_{T(g)}(X)}\right)^{-\frac{1}{2}} s\big(X, T(g)\big)s\big(X, T(g)\big)^t\right] - 2\int\left\{\left(\frac{g(y)}{f_{T(g)}(y)}\right)^{\frac{1}{2}} - 1\right\}\nabla^2 f_{T(g)}(y)\,\mathrm{d}y.$$

Proof. For the ε-contaminated density, we have $\int A(\frac{g_{\varepsilon x}(y)}{f_{T(g_{\varepsilon x})}(y)} - 1)\nabla f_{T(g_{\varepsilon x})}(y)\,\mathrm{d}y = 0$. Now derivation yields

$$\int A'\left(\frac{g(y)}{f_{T(g)}(y)} - 1\right)\frac{(\delta_x(y) - g(y))f_{T(g)}(y) - g(y)(\nabla f_{T(g)}(y))^t IF(x; T, g)}{f^2_{T(g)}(y)}\nabla f_{T(g)}(y)\,\mathrm{d}y$$
$$+ \int A\left(\frac{g(y)}{f_{T(g)}(y)} - 1\right)\nabla^2 f_{T(g)}(y) IF(x; T, g)\,\mathrm{d}y = 0.$$

Then we obtain the equality

$$\left[\int A'\left(\frac{g(y)}{f_{T(g)}(y)} - 1\right)\frac{\nabla f_{T(g)}(y)\nabla f_{T(g)}(y)^t}{f^2_{T(g)}(y)}g(y)\,\mathrm{d}y - \int A\left(\frac{g(y)}{f_{T(g)}(y)} - 1\right)\nabla^2 f_{T(g)}(y)\,\mathrm{d}y\right] IF(x; T, g)$$
$$= A'\left(\frac{g(x)}{f_{T(g)}(x)} - 1\right)\frac{\nabla f_{T(g)}(x)}{f_{T(g)}(x)} - \int A'\left(\frac{g(y)}{f_{T(g)}(y)} - 1\right)\frac{\nabla f_{T(g)}(y)}{f_{T(g)}(y)}g(y)\,\mathrm{d}y.$$

By replacing $A(\delta)$ from (4), we obtain the influence function $IF(x; T, g) = \mathrm{DEN}^{-1}\mathrm{NUM}$ where

$$\mathrm{NUM} = \left(\frac{g(x)}{f_{T(g)}(x)}\right)^{-\frac{1}{2}} s\big(x, T(g)\big) - E_G\left[\left(\frac{g(X)}{f_{T(g)}(X)}\right)^{-\frac{1}{2}} s\big(X, T(g)\big)\right]$$

$$\mathrm{DEN} = E_G\left[\left(\frac{g(X)}{f_{T(g)}(X)}\right)^{-\frac{1}{2}} s\big(X, T(g)\big)s\big(X, T(g)\big)^t\right] - 2\int\left\{\left(\frac{g(y)}{f_{T(g)}(y)}\right)^{\frac{1}{2}} - 1\right\}\nabla^2 f_{T(g)}(y)\,\mathrm{d}y. \qquad \square$$

Corollary 2.2. *If the true density g is the model f_θ, then $T(g)=\theta$ and the influence function corresponding to the functional T is $IF(x;T,f_\theta)=I_\theta^{-1}s(x,\theta)$.*

At the model, we recognize the influence function of the maximum likelihood estimator. This is unbounded and hence, the robustness of the corresponding MHDE cannot be explained through their influence function. This aspect has been already noted in literature for other minimum Hellinger procedures. Among others, Lindsay [6] underlined the limitation of the influence function approach for such procedures and indicated some other ways to explore their robustness properties. In our context, besides of the RAF, whose form indicate the quality of the MHDE to downweight some possible outliers that manifest as large values of δ, the asymptotic breakdown point, that we treat in the next section, offers information on the robustness of the MHDE with respect to some contamination models.

3. Breakdown analysis

The breakdown point of a statistical functional represents the smallest fraction of contamination in data that may cause an arbitrarily extreme value in the estimate. Consider the contamination model $H_{\varepsilon,n}=(1-\varepsilon)G+\varepsilon K_n$ where $\{K_n\}$ is a sequence of contaminating distributions. Denote by $h_{\varepsilon,n}$, g and k_n the corresponding densities.

Extending the definition of Simpson [8], we say that there is breakdown in T for ε level contamination if there exists a sequence $\{K_n\}$ such that $d(T(h_{\varepsilon,n}),T(g))\to\infty$ as $n\to\infty$, where

$$d\big(T(h_{\varepsilon,n}),T(g)\big)=\big\|\mu(h_{\varepsilon,n})-\mu(g)\big\|+\max\left\{\big|\lambda_1\big(V(h_{\varepsilon,n})\big)-\lambda_1\big(V(g)\big)\big|,\left|\frac{1}{\lambda_p(V(h_{\varepsilon,n}))}-\frac{1}{\lambda_p(V(g))}\right|\right\}$$

$\lambda_1(A)\geqslant\lambda_2(A)\geqslant\cdots\geqslant\lambda_p(A)$ being the eigenvalues of the matrix A. Then the breakdown point is defined as the smallest level of contamination ε for which T breakdown.

When we estimate a covariance matrix we have to be sure not only that there are no 'explosions' of the respective estimator, but also that there are no 'implosions' of this (that is the estimator is not arbitrarily close to zero). This remark justifies the use of the distance d in the definition of the breakdown point, such that the breakdown is produced when the location 'explodes' or the covariance 'explodes' or 'implodes'.

The following assumptions reflect the intuitively worst possible choice of the contamination. We write below $\theta_n=T(h_{\varepsilon,n})$.

- A1. $\int\min\{g(x),k_n(x)\}\to 0$ as $n\to\infty$. That is the contaminating distribution becomes asymptotically singular to the true distribution.
- A2. For any fixed $\theta\in\Theta$, $\int\min\{f_\theta(x),k_n(x)\}\to 0$ as $n\to\infty$. That is the contaminating distribution is asymptotically singular to the specified models.
- A3. $\int\min\{g(x),f_{\theta_n}(x)\}\to 0$ as $n\to\infty$ if $d(\theta_n,T(g))\to\infty$ as $n\to\infty$. This means that large values of the parameter θ give distributions which become singular to the true distribution.

Theorem 3.1. *Assume that the true distribution, the model and the contaminating sequence satisfy conditions* A1–A3. *Then the asymptotic breakdown point of the MHDE is at least* $1/2$ *at the model.*

Proof. Let $\theta_n=T(h_{\varepsilon,n})$ be the minimizer of $\phi_H(h_{\varepsilon,n},f_\theta)$. Given the level of contamination ε, let, if possible, breakdown occurs, that is there exists a sequence $\{K_n\}$ such that $d(\theta_n,T(g))\to\infty$ as $n\to\infty$.

Then

$$\phi_H(h_{\varepsilon,n},f_\theta)=\int_{A_n}D\big(h_{\varepsilon,n}(x),f_{\theta_n}(x)\big)+\int_{A_n^c}D\big(h_{\varepsilon,n}(x),f_{\theta_n}(x)\big) \tag{6}$$

where $A_n=\{x\colon\ g(x)>\max\{k_n(x),f_{\theta_n}(x)\}\}$ and $D(g,f)=C(\frac{g}{f}-1)f=2[g^{1/2}-f^{1/2}]^2$.

From A1 we have $\int_{A_n}k_n(x)\to 0$ and from A3 we have $\int_{A_n}f_{\theta_n}(x)\to 0$ as $n\to\infty$. Hence, using Cauchy–Schwarz inequality we can see that

$$\left|\int_{A_n}D\big(h_{\varepsilon,n}(x),f_{\theta_n}(x)\big)-\int_{A_n}D\big((1-\varepsilon)g(x),0\big)\right|\to 0 \tag{7}$$

as $n \to \infty$. From A1 and A3, we can see that $\int_{A_n^c} g(x) \to 0$ and

$$\left|\int_{A_n} D\big((1-\varepsilon)g(x),0\big) - \int D\big((1-\varepsilon)g(x),0\big)\right| = \left|\int_{A_n} 2(1-\varepsilon)g - \int 2(1-\varepsilon)g\right| \to 0 \tag{8}$$

as $n \to \infty$. Using (7) and (8) we have

$$\int_{A_n} D\big(h_{\varepsilon,n}(x), f_{\theta_n}(x)\big) \to 2(1-\varepsilon). \tag{9}$$

From A1 and A3, we have $\int_{A_n^c} g(x) \to 0$. Using an analogous reasoning as above,

$$\left|\int_{A_n^c} D\big(h_{\varepsilon,n}(x), f_{\theta_n}(x)\big) - \int D\big(\varepsilon k_n(x), f_{\theta_n}(x)\big)\right| \to 0.$$

Taking into account that C is a convex function, by Jensen inequality we have

$$\int D\big(\varepsilon k_n(x), f_{\theta_n}(x)\big) = \int C\left(\frac{\varepsilon k_n(x)}{f_{\theta_n}(x)} - 1\right) f_{\theta_n}(x) \geqslant C\left(\int \varepsilon k_n(x) - f_{\theta_n}(x)\right) = C(\varepsilon - 1) = 2[\sqrt{\varepsilon} - 1]^2.$$

Then with (9) it follows that

$$\liminf_{n\to\infty} \phi_H(h_{\varepsilon,n}, f_{\theta_n}) \geqslant 4(1-\sqrt{\varepsilon}). \tag{10}$$

For any fixed θ, let $B_n = B_n(\theta) = \{x : k_n(x) > \max\{g(x), f_\theta(x)\}\}$.

From A1, $\int_{B_n} g(x) \to 0$ and from A2, $\int_{B_n} f_\theta(x) \to 0$ as $n \to \infty$. From A1 and A2, we have $\int_{B_n^c} k_n(x) \to 0$ as $n \to \infty$ and using Cauchy–Schwarz inequality, we obtain

$$\left|\int_{B_n} D\big(h_{\varepsilon,n}(x), f_\theta(x)\big) - \int_{B_n} D\big(\varepsilon k_n(x), 0\big)\right| \to 0$$

as $n \to \infty$. Notice that $\int_{B_n} D(\varepsilon k_n(x), 0) = 2\varepsilon \int_{B_n} k_n \to 2\varepsilon$ as $n \to \infty$. So

$$\int_{B_n} D\big(h_{\varepsilon,n}(x), f_\theta(x)\big) \to 2\varepsilon \tag{11}$$

as $n \to \infty$.

Similarly we have

$$\left|\int_{B_n^c} D\big(h_{\varepsilon,n}(x), f_\theta(x)\big) - \int D\big((1-\varepsilon)g(x), f_\theta(x)\big)\right| \to 0$$

as $n \to \infty$. Then from (11) we have

$$\lim_{n\to\infty} \phi_H(h_{\varepsilon,n}, f_\theta) = 2\varepsilon + \int D\big((1-\varepsilon)g(x), f_\theta(x)\big) \geqslant 2\varepsilon + \inf_\theta \int D\big((1-\varepsilon)g(x), f_\theta(x)\big). \tag{12}$$

Let θ^* be the minimizer, over Θ, of $\int D((1-\varepsilon)g(x), f_\theta(x))$. The inequality in (12) becomes equality for $\theta = \theta^*$. Observe that among all fixed θ, the divergence ϕ_H is minimized in the limit by θ^*.

When the true distribution belongs to the model, then $g = f_{\theta_t}$ and

$$\int D\big((1-\varepsilon) f_{\theta_t}, f_{\theta_t}\big) = 2[\sqrt{1-\varepsilon} - 1]^2$$

which is also the lower bound for $D((1-\varepsilon) f_{\theta_t}, f_\theta)$, over $\theta \in \Theta$. So in this case $\theta^* = \theta_t$ and $\lim_{n\to\infty} \phi_H(h_{\varepsilon,n}, f_{\theta^*}) = 4[1 - \sqrt{1-\varepsilon}]$. Thus at the model, for $\varepsilon < \frac{1}{2}$, $\lim_{n\to\infty} \phi_H(h_{\varepsilon,n}, f_{\theta^*}) < 4[1 - \sqrt{\varepsilon}]$. Then using (10) we obtain

$$\lim_{n\to\infty} \phi_H\big(h_{\varepsilon,n}, f_{\theta^*}\big) < \liminf_{n\to\infty} \phi_H(h_{\varepsilon,n}, f_{\theta_n})$$

and deduce that θ_n is not a minimizer for every n. The consequence is that every level of contamination ε for which breakdown occurs must satisfy $\varepsilon \geqslant \frac{1}{2}$. $\square$

Acknowledgement

The author thanks to the referees for the suggestions leading to improvement of the paper.

References

[1] A. Basu, B. Lindsay, Minimum disparity estimation in the continuous case: efficiency, distributions, robustness, Annals of the Institute of Statistical Mathematics 46 (1994) 683–705.
[2] A. Basu, S. Sarkar, A.N. Vidyashankar, Minimum negative exponential disparity estimation in parametric models, Journal of Statistical Planning and Inference 58 (1997) 349–370.
[3] R. Beran, Minimum Hellinger distance estimates for parametric models, Annals of Statistics 5 (1977) 445–463.
[4] M. Broniatowski, Estimation through Kullback–Leibler divergence, Mathematical Methods of Statistics 12 (2003) 391–409.
[5] M. Broniatowski, A. Keziou, Parametric estimation and testing through divergences, Preprint 2004-1, L.S.T. A–Université Paris 6.
[6] B. Lindsay, Efficiency versus robustness: the case of minimum Hellinger distance and related methods, Annals of Statistics 22 (1994) 1081–1114.
[7] C. Park, A. Basu, Minimum disparity estimation: asymptotic normality and breakdown point results, Bulletin of Informatics and Cybernetics 99 (1999) 1–15.
[8] D.G. Simpson, Minimum Hellinger distance estimation for the analysis of count data, Journal of the American Statistical Association 82 (1987) 802–807.
[9] D.G. Simpson, Hellinger deviance tests: efficiency, breakdown points, and examples, Journal of the American Statistical Association 84 (1989) 104–113.
[10] R. Tamura, D. Boos, Minimum Hellinger distance estimation for multivariate location and covariance, Journal of the American Statistical Association 81 (1986) 223–229.
[11] A. Toma, Minimum Hellinger distance estimators for multivariate distributions from Johnson system, Journal of Statistical Planning and Inference (2007), in press.