



HAL
open science

Optimal robust M-estimators using divergences

Aida Toma

► **To cite this version:**

Aida Toma. Optimal robust M-estimators using divergences. *Statistics and Probability Letters*, 2010, 79 (1), pp.1. 10.1016/j.spl.2008.04.011 . hal-00613922

HAL Id: hal-00613922

<https://hal.science/hal-00613922>

Submitted on 8 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Optimal robust M-estimators using divergences

Aida Toma

PII: S0167-7152(08)00231-9
DOI: 10.1016/j.spl.2008.04.011
Reference: STAPRO 5058

To appear in: *Statistics and Probability Letters*

Received date: 1 April 2008
Revised date: 11 April 2008
Accepted date: 21 April 2008

Please cite this article as: Toma, A., Optimal robust M-estimators using divergences. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.04.011

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Optimal robust M-estimators using divergences

Aida Toma^{a,b,*}

^a*Mathematics Department, Academy of Economic Studies, Piața Romană No. 6,
Bucharest, Romania, e-mail: aida_toma@yahoo.com*

^b*"Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied
Mathematics, Calea 13 Septembrie No. 13, Bucharest, Romania*

Abstract

We introduce new robustness and efficiency measures based on divergences and use it to construct equivariant optimal robust M-estimators.

Key words: M-estimator, optimal robustness, divergence

1 Introduction

Let $(F_\theta)_\theta$ be a parametric model, where $\theta \in \Theta \subset \mathbb{R}^d$, and X_1, \dots, X_n be a random sample on F_θ .

A map T which sends an arbitrary distribution function on \mathbb{R}^p to Θ is a statistical functional corresponding to an estimator $\hat{\theta}_n$ of the parameter θ , whenever $\hat{\theta}_n = T(F_n)$, F_n being the empirical distribution function. The influence func-

* Corresponding author. E-mail: aida_toma@yahoo.com

tion of T is defined as $\text{IF}(x; T, F_\theta) := \left. \frac{\partial T(\tilde{F}_{\varepsilon x})}{\partial \varepsilon} \right|_{\varepsilon=0}$, where $\tilde{F}_{\varepsilon x} := (1 - \varepsilon)F_\theta + \varepsilon\delta_x$, δ_x being the point mass distribution at x . The unstandardized gross error sensitivity (GES) is defined by $\gamma_u^*(T, \theta) := \sup_x \|\text{IF}(x; T, F_\theta)\|$. Let $\hat{\theta}_n = T(F_n)$ be an estimator of θ . Under regularity conditions it holds

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_D \mathcal{N}(0, V_T(\theta)) \quad (1)$$

where \rightarrow_D denotes the convergence in distribution and

$$V_T(\theta) = \int \text{IF}(x; T, F_\theta) \text{IF}(x; T, F_\theta)^t dF_\theta(x) \quad (2)$$

t denoting the transpose.

The optimal robust M-estimators for multidimensional parametric models were introduced by Stahel (1981). A part of this estimators are defined by minimizing the trace of $V_T(\theta)$ among the M-estimators with $\gamma_u^*(T, \theta)$ bounded by a given constant. These estimators have the drawback that are not equivariant¹, since $\gamma_u^*(T, \theta)$ is not invariant under model reparametrizations. In the same paper were proposed some alternative robustness measures, whose invariance with respect to reparametrizations have led to the construction of equivariant optimal robust M-estimators. These measures are the self standardized GES and the information standardized GES. The efficiency assuring the optimality of the corresponding estimators is attained by choosing ap-

¹ Equivariance means that, considering the reparametrization $\bar{\theta} := \beta(\theta)$ and T_1^* and T_2^* optimal estimators of θ and $\bar{\theta}$ respectively, it holds $T_2^* = \beta(T_1^*)$.

appropriate minimization criterions. The resulting equivariant estimators are called optimal self standardized robust M-estimators and optimal information standardized robust M-estimators, respectively. Yohai (2008) proved that the optimal information standardized robust M-estimators could be also obtained by solving the variational problem presented in Stahel (1981), by replacing the robustness and efficiency measures with some new ones based on the use of the Kullback-Leibler divergence. We mention that, the information standardized GES has also been considered by He and Simpson (1992) in the context of direction estimates on sphere.

In this paper we generalize the results from Yohai (2008) by considering large classes of divergences, namely the Cressie-Read divergences (Cressie and Read (1994)) and the density power divergences (Basu et al. (1998)). On the basis of these divergences, we define new robustness and efficiency measures and use it to construct optimal robust M-estimators. As the Kullback-Leibler divergence, used by Yohai (2008), the Cressie-Read divergences play in the favor of the optimal information standardized robust M-estimators. On the other hand, the robustness and efficiency measures based on the density power divergences lead to new classes of equivariant optimal robust M-estimators.

2 GES and efficiency measures based on divergences

Let $(F_\theta)_\theta$ be a parametric model and $(f_\theta)_\theta$ be the corresponding densities. A divergence from the Cressie-Read class between F_{θ^*} and the true distribution

F_θ is defined by

$$\text{CR}_\gamma(\theta^*, \theta) = \int_{-\infty}^{\infty} \varphi_\gamma \left(\frac{f_{\theta^*}(z)}{f_\theta(z)} \right) f_\theta(z) dz \quad (3)$$

where $\varphi_\gamma(z) := \frac{z^\gamma - \gamma z + \gamma - 1}{\gamma(\gamma - 1)}$ for $\gamma \in \mathbb{R} \setminus \{0, 1\}$, $\varphi_0(z) := -\log z + z - 1$ and $\varphi_1(z) := z \log z - z + 1$. The Kullback-Leibler divergence (KL) is associated to φ_1 , the modified Kullback-Leibler (KL_m) to φ_0 , the χ^2 divergence to φ_2 , the modified χ^2 divergence (χ_m^2) to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$.

A density power divergence between F_{θ^*} and F_θ is defined by

$$\text{DP}_\alpha(\theta^*, \theta) = \int_{-\infty}^{\infty} \left\{ f_\theta^{1+\alpha}(z) - \left(1 + \frac{1}{\alpha} \right) f_{\theta^*}(z) f_\theta^\alpha(z) + \frac{1}{\alpha} f_{\theta^*}^{1+\alpha}(z) \right\} dz \quad (\alpha > 0). \quad (4)$$

It is known that the CR_γ divergences, as well as the DP_α divergences are invariant with respect to one to one parameter transformations.

Let T be a statistical functional corresponding to an estimator of θ . The CR_γ divergence between $F_{T(F)}$ and F_θ is

$$\text{CR}_\gamma(T(F), \theta) = \int_{-\infty}^{\infty} \varphi_\gamma \left(\frac{f_{T(F)}(z)}{f_\theta(z)} \right) f_\theta(z) dz \quad (5)$$

while the DP_α divergence between $F_{T(F)}$ and F_θ is

$$\text{DP}_\alpha(T(F), \theta) = \int_{-\infty}^{\infty} \left\{ f_\theta^{1+\alpha}(z) - \left(1 + \frac{1}{\alpha} \right) f_{T(F)}(z) f_\theta^\alpha(z) + \frac{1}{\alpha} f_{T(F)}^{1+\alpha}(z) \right\} dz. \quad (6)$$

If T is Fisher consistent, then $\text{CR}_\gamma(T(F_\theta), \theta) = 0$ and $\text{DP}_\alpha(T(F_\theta), \theta) = 0$. Therefore, in order to assess the robustness of $T(F)$ under infinitesimal outlier con-

taminations, in the next lemmas we compute the first and second Gateaux derivatives of the functionals $\text{CR}_\gamma(T(F), \theta)$ and $\text{DP}_\alpha(T(F), \theta)$. Derivation with respect to ε yields:

Lemma 1 *It hold*

- (i) $\left. \frac{\partial \text{CR}_\gamma(T(\tilde{F}_{\varepsilon x}), \theta)}{\partial \varepsilon} \right|_{\varepsilon=0} = 0.$
- (ii) $\left. \frac{\partial^2 \text{CR}_\gamma(T(\tilde{F}_{\varepsilon x}), \theta)}{\partial \varepsilon^2} \right|_{\varepsilon=0} = \text{IF}(x; T, F_\theta)^t I(\theta) \text{IF}(x; T, F_\theta),$ where $I(\theta) = \int \frac{\dot{f}_\theta(z) \dot{f}_\theta(z)^t}{f_\theta(z)} dz$

is the information matrix and \dot{f}_θ denotes the derivative with respect to θ .

Lemma 2 *It hold*

- (i) $\left. \frac{\partial \text{DP}_\alpha(T(\tilde{F}_{\varepsilon x}), \theta)}{\partial \varepsilon} \right|_{\varepsilon=0} = 0.$
- (ii) $\left. \frac{\partial^2 \text{DP}_\alpha(T(\tilde{F}_{\varepsilon x}), \theta)}{\partial \varepsilon^2} \right|_{\varepsilon=0} = \text{IF}(x; T, F_\theta)^t M_\alpha(\theta) \text{IF}(x; T, F_\theta),$

where $M_\alpha(\theta) = (1 + \alpha) \int f_\theta^{\alpha-1}(z) \dot{f}_\theta(z) \dot{f}_\theta(z)^t dz.$

These two lemmas allow to consider, for small ε , the following approximations

$$\begin{aligned} \text{CR}_\gamma(T(\tilde{F}_{\varepsilon x}), \theta) &\simeq \text{IF}(x; T, F_\theta)^t I(\theta) \text{IF}(x; T, F_\theta) \varepsilon^2 \\ \text{DP}_\alpha(T(\tilde{F}_{\varepsilon x}), \theta) &\simeq \text{IF}(x; T, F_\theta)^t M_\alpha(\theta) \text{IF}(x; T, F_\theta) \varepsilon^2. \end{aligned}$$

Then we define the GES based on the CR_γ divergence of the functional T as

$$\gamma_{\text{CR}_\gamma}^*(T, \theta) = \sup_x \text{IF}(x; T, F_\theta)^t I(\theta) \text{IF}(x; T, F_\theta) \quad (7)$$

respectively the GES based on the DP_α divergence of the functional T as

$$\gamma_{\text{DP}_\alpha}^*(T, \theta) = \sup_x \text{IF}(x; T, F_\theta)^t M_\alpha(\theta) \text{IF}(x; T, F_\theta). \quad (8)$$

Observe that $\gamma_{\text{CR}\gamma}^*$ is independent upon the γ choice and coincide with the well known information standardized GES, while $\gamma_{\text{DP}\alpha}^*$, $\alpha > 0$, define a new class of standardized GES. It is known that the information standardized GES is invariant with respect to differentiable one to one parameter transformations. In the following we show that every $\gamma_{\text{DP}\alpha}^*$ enjoy the same property.

Indeed, let $\bar{\theta} := \beta(\theta)$ be a one to one differentiable transformation and $J(\theta) := \frac{\partial \beta(\theta)}{\partial \theta}$ be its Jacobian. The transformed model writes as $(\bar{F}_{\bar{\theta}})_{\bar{\theta}}$ with $\bar{F}_{\bar{\theta}} := F_{\theta}$. Then the densities, scores and the matrices $M_{\alpha}(\theta)$ transform to

$$\bar{f}_{\bar{\theta}}(z) = f_{\theta}(z), \quad \frac{\dot{\bar{f}}_{\bar{\theta}}(z)}{\bar{f}_{\bar{\theta}}(z)} = J(\theta)^{-t} \frac{\dot{f}_{\theta}(z)}{f_{\theta}(z)}, \quad M_{\alpha}(\bar{\theta}) = J(\theta)^{-t} M_{\alpha}(\theta) J(\theta)^{-1} \quad (9)$$

(A^{-t} means $(A^{-1})^t$). If T is an estimating functional of θ , then is natural to examine $\bar{T} := \beta(T)$ as an estimating functional of $\bar{\theta}$. Then $\text{IF}(x; \bar{T}, \bar{F}_{\bar{\theta}}) = J(\theta) \text{IF}(x; T, F_{\theta})$. Consequently, replacing all these quantities in $\gamma_{\text{DP}\alpha}^*$, it holds $\gamma_{\text{DP}\alpha}^*(\bar{T}, \bar{\theta}) = \gamma_{\text{DP}\alpha}^*(T, \theta)$.

Let $\hat{\theta}_n$ be an estimator of θ satisfying (1) with $V_T(\theta)$ given by (2).

Lemma 3 *It holds $n\text{E}_{\theta}(\|\hat{\theta}_n - \theta\|_{I(\theta)}^2) \rightarrow \text{tr}\{I(\theta)V_T(\theta)\}$, where E_{θ} denotes the expectation with respect to F_{θ} and tr denotes the trace.*

Proof. Let $B(\theta) = V_T^{1/2}(\theta)$ and $z_n = n^{1/2}B(\theta)^{-1}(\hat{\theta}_n - \theta)$. Then $z_n \rightarrow_D \mathcal{N}(0, I_d)$ where I_d is the identity matrix. Then

$$n\|\hat{\theta}_n - \theta\|_{I(\theta)}^2 = n(\hat{\theta}_n - \theta)^t I(\theta) (\hat{\theta}_n - \theta) = z_n^t B(\theta)^t I(\theta) B(\theta) z_n. \quad (10)$$

Let $V(\theta)$ and $\Lambda(\theta)$ be the eigenvectors and the eigenvalues of $B(\theta)^t I(\theta) B(\theta)$

and put $w_n := V(\theta)^t z_n$. Then $w_n \rightarrow_D \mathcal{N}(0, I_d)$ and

$$n\|\hat{\theta}_n - \theta\|_{I(\theta)}^2 = w_n^t \Lambda(\theta) w_n \rightarrow_D \sum_{i=1}^d \lambda_i(\theta) v_i^2 \quad (11)$$

where v_1, \dots, v_d are i.i.d. variables with distribution $\mathcal{N}(0, 1)$ and $\lambda_i(\theta)$, $1 \leq i \leq d$, are the eigenvalues of $I(\theta) V_T(\theta)$. Then, in virtue of Helly-Bray lemma

$$nE_\theta(\|\hat{\theta}_n - \theta\|_{I(\theta)}^2) \rightarrow \sum_{i=1}^d \lambda_i(\theta) = \text{tr}\{I(\theta) V_T(\theta)\}. \quad (12)$$

Remark 1 *The mean square error of the estimator $\hat{\theta}_n$, evaluated in the metric given by the inverse of the information matrix, which in its asymptotic form is used to make the criterion to be minimized, is asymptotically equivalent to $2E_\theta(\text{CR}_\gamma(\hat{\theta}_n, \theta))$, independently upon the γ choice.*

Indeed, some simple calculation shows that

$$\left. \frac{\partial \text{CR}_\gamma(\theta^*, \theta)}{\partial \theta^*} \right|_{\theta^*=\theta} = 0 \quad \text{and} \quad \left. \frac{\partial^2 \text{CR}_\gamma(\theta^*, \theta)}{\partial \theta^{*2}} \right|_{\theta^*=\theta} = I(\theta). \quad (13)$$

Then a Taylor expansion of order two yields

$$n\text{CR}_\gamma(\hat{\theta}_n, \theta) = \frac{n}{2} (\hat{\theta}_n - \theta)^t I(\theta) (\hat{\theta}_n - \theta) + no(\|\hat{\theta}_n - \theta\|^2). \quad (14)$$

Since $no(\|\hat{\theta}_n - \theta\|^2) \rightarrow 0$ in probability, from (11) and (14) it follows that

$2n\text{CR}_\gamma(\hat{\theta}_n, \theta) \rightarrow_D \sum_{i=1}^d \lambda_i(\theta) v_i^2$. Consequently $2nE_\theta(\text{CR}_\gamma(\hat{\theta}_n, \theta)) \rightarrow \text{tr}\{I(\theta) V_T(\theta)\}$

by the Helly-Bray lemma.

Lemma 4 *It holds $2nE_\theta(\|\hat{\theta}_n - \theta\|_{M_\alpha(\theta)}^2) \rightarrow \text{tr}\{M_\alpha(\theta) V_T(\theta)\}$.*

The proof is similar to that of Lemma 3. Remark 1 adapts accordingly, saying that the mean square error of the estimator $\hat{\theta}_n$, evaluated in the metric given by the inverse of $M_\alpha(\theta)$, is asymptotically equivalent to $2E_\theta(DP_\alpha(\hat{\theta}_n, \theta))$.

Hence, the criterion to be minimized in order to obtain optimal robust M-estimators is $\text{tr}\{I(\theta)V_T(\theta)\}$ when using $\gamma_{CR_\gamma}^*$ as sensitivity measure, respectively $\text{tr}\{M_\alpha(\theta)V_T(\theta)\}$ when using the sensitivity $\gamma_{DP_\alpha}^*$ (see also Remark 1 p. 242 and the discussion at the end of p. 243 from Hampel et al. (1986) for comments on the minimization criterion used in the construction of optimal robust M-estimators).

3 Optimal robust M-estimators

Let θ be a fixed parameter. Making use of the robustness and efficiency measures previously defined, equivariant optimal robust M-estimators of the parameter θ arise as solutions to the following problems:

1. Find an M-estimator such that its ψ -function verifies

$$\begin{aligned} \int \psi(x, \theta) s(x, \theta)^t dF_\theta(x) &= I^{1/2}(\theta) \\ \int \psi(x, \theta) dF_\theta(x) &= 0 \end{aligned}$$

where $s(x, \theta)$ is the score function, and such that its corresponding functional T minimizes $\text{tr}\{I(\theta)V_T(\theta)\}$ subject to $\gamma_{CR_\gamma}^*(T, \theta) \leq c$, where c is a given constant.

2. Let α be fixed. Find an M-estimator such that its ψ -function verifies

$$\begin{aligned}\int \psi(x, \theta) s(x, \theta)^t dF_\theta(x) &= M_\alpha^{1/2}(\theta) \\ \int \psi(x, \theta) dF_\theta(x) &= 0\end{aligned}$$

and such that its corresponding functional T minimizes $\text{tr}\{M_\alpha(\theta)V_T(\theta)\}$ subject to $\gamma_{\text{DP}_\alpha}^*(T, \theta) \leq c$, where c is a given constant.

The solution to the first problem is unique and is precisely the solution obtained by Stahel (1981) as optimal information standardized robust M-estimator and by Yohai (2008) as optimal robust M-estimator using the Kullback-Leibler divergence. In the following, we determine the solution to the second problem. To this purpose we use the context of the optimal unstandardized robust M-estimators.

Let \mathcal{T}_θ^r be the set of functionals (corresponding to M-estimators) defined by

$$\mathcal{T}_\theta^r := \left\{ T : \text{IF}(\cdot; T, F_\theta) \text{ exists, } \int \text{IF}(x; T, F_\theta) dF_\theta(x) = 0, \right. \\ \left. \int \text{IF}(x; T, F_\theta) s(x, \theta)^t dF_\theta(x) = I_d \text{ and } V_T(\theta) \text{ exists} \right\}.$$

The problem is to find a functional T from $\mathcal{T}_{c,\theta}^u := \{T \in \mathcal{T}_\theta^r : \gamma_a^*(T, \theta) \leq c\}$ that minimizes $\text{tr}\{V_T(\theta)\}$. Then the corresponding estimator will be the optimal unstandardized robust M-estimator. The solution to this problem is presented in chapter 4 from Hampel et al. (1986), the conclusion being the following. The optimal unstandardized robust M-estimator is defined by the ψ -function

$$\psi_c^{A(\theta), a(\theta)}(x, \theta) = h_c\{A(\theta)[s(x, \theta) - a(\theta)]\} \quad (15)$$

where $h_c(z) := z \min(1, c/\|z\|)$ is the multidimensional function of Huber and the pair $(A(\theta), a(\theta))$, where $A(\theta)$ is $d \times d$ nonsingular matrix and $a(\theta) \in \mathbb{R}^d$, is the unique solution of the system

$$\begin{aligned} \int \psi_c^{A(\theta), a(\theta)}(x, \theta) s(x, \theta)^t dF_\theta(x) &= I_d \\ \int \psi_c^{A(\theta), a(\theta)}(x, \theta) dF_\theta(x) &= 0. \end{aligned}$$

The associated functional T_c^u takes part from $\mathcal{T}_{c, \theta}^u$ and minimizes $\text{tr}\{V_T(\theta)\}$ on this set.

Definition 1 *Let $\alpha > 0$ be fixed. Suppose that there exists a differentiable solution $(A_\alpha(\theta), a_\alpha(\theta))$ of the system*

$$\begin{aligned} \int \psi_c^{A_\alpha(\theta), a_\alpha(\theta)}(x, \theta) s(x, \theta)^t dF_\theta(x) &= M_\alpha^{1/2}(\theta) \\ \int \psi_c^{A_\alpha(\theta), a_\alpha(\theta)}(x, \theta) dF_\theta(x) &= 0 \end{aligned}$$

where $\psi_c^{A_\alpha(\theta), a_\alpha(\theta)}$ is of the form (15). Then $\psi_c^{A_\alpha(\theta), a_\alpha(\theta)}$ is called the ψ -function of the optimal $M_\alpha(\theta)$ standardized robust M-estimator and is denoted by $\psi_c^{\text{DP}\alpha}$. The corresponding functional is denoted $T_c^{\text{DP}\alpha}$.

We prove that $\psi_c^{\text{DP}\alpha}$ as defined above exists and is unique and that the functional $T_c^{\text{DP}\alpha}$ minimizes $\text{tr}\{M_\alpha(\theta)V_T(\theta)\}$ in the class $\mathcal{T}_{c, \theta}^{\text{DP}\alpha} \subset \mathcal{T}_\theta^r$ defined by the regular functional with the sensitivity $\gamma_{\text{DP}\alpha}^*$ not exceeding a given bound c . This entails the fact that the optimal $M_\alpha(\theta)$ standardized robust M-estimator is the solution to the second problem enounced at the beginning of this section.

Let $\bar{\theta} := \beta(\theta) = M_\alpha^{1/2}(\theta)\theta$. The optimal unstandardized robust M-estimator

of the parameter $\bar{\theta}$ is characterized by the ψ -function

$$\psi_c^{A(\bar{\theta}), a(\bar{\theta})}(x, \bar{\theta}) = h_c\{A(\bar{\theta})[s(x, \bar{\theta}) - a(\bar{\theta})]\} \quad (16)$$

that verifies the restrictions

$$\int \psi_c^{A(\bar{\theta}), a(\bar{\theta})}(x, \bar{\theta}) s(x, \bar{\theta})^t dF_{\bar{\theta}}(x) = I_d \quad (17)$$

$$\int \psi_c^{A(\bar{\theta}), a(\bar{\theta})}(x, \bar{\theta}) dF_{\bar{\theta}}(x) = 0 \quad (18)$$

and its functional T^u minimizes $\text{tr}\{V_T(\bar{\theta})\}$ in the class $\mathcal{T}_{c, \bar{\theta}}^u$.

Coming back to the parameter θ , for which we construct the optimal $M_\alpha(\theta)$ standardized robust M-estimator, the ψ -function (16) rewrites as

$$h_c\{A(\bar{\theta})[M_\alpha^{-1/2}(\theta)s(x, \theta) - a(\bar{\theta})]\} = h_c\{A_\alpha(\theta)[s(x, \theta) - a_\alpha(\theta)]\} \quad (19)$$

where $A_\alpha(\theta) := A(M_\alpha^{1/2}(\theta)\theta)M_\alpha^{-1/2}(\theta)$ and $a_\alpha(\theta) := M_\alpha^{1/2}(\theta)a(M_\alpha^{1/2}(\theta)\theta)$, and is precisely the ψ -function $\psi_c^{\text{DP}\alpha}$ since it verifies the system from Definition 1, as results from (17) and (18). The existence and the unicity of this ψ -function is assured by the existence and unicity of the pair $(A(\bar{\theta}), a(\bar{\theta}))$ solution of the system given by (17) and (18).

On the other hand, the functional $T_c^{\text{DP}\alpha}$ has the form $T_c^{\text{DP}\alpha} = \beta^{-1}(T^u)$.

Then using the fact that $\text{IF}(x; T^u, F_{\bar{\theta}}) = M_\alpha^{1/2}(\theta)\text{IF}(x; T_c^{\text{DP}\alpha}, F_\theta)$, we obtain $\gamma_u^*(T^u, \bar{\theta}) = \gamma_{\text{DP}\alpha}^*(T_c^{\text{DP}\alpha}, \theta)$ and $\text{tr}\{V_{T^u}(\bar{\theta})\} = \text{tr}\{M_\alpha(\theta)V_{T_c^{\text{DP}\alpha}}(\theta)\}$. These equalities entail that the functional $T_c^{\text{DP}\alpha}$ is from the set $\mathcal{T}_{c, \theta}^{\text{DP}\alpha}$ and minimizes $\text{tr}\{M_\alpha(\theta)V_T(\theta)\}$ on this set.

References

- [1] Basu, A., Harris, I., Hjort, N., Jones, M., 1998. Robust and efficient estimation by minimising a density power divergence. *Biometrika* 85, 549-559.
- [2] Cressie, T., Read, T., 1994. Multinomial goodness of fit tests. *J. R. Statist. Soc. Ser. B* 46, 440-464.
- [3] Hampel, F., Ronchetti, E., Rousseeuw, P.J., Stahel, W., 1986. *Robust Statistics: the Approach Based on Influence Functions*. Wiley, New York.
- [4] He, X., Simpson, D.G., 1992. Robust direction estimation. *Annals of Statistics* 20, 351-369.
- [5] Yohai, V., 2008. Optimal robust estimates using the Kullback-Leibler divergence. *Statistics & Probability Letters*, In Press.
- [6] Stahel, W.A., 1981. Robust estimation, infinitesimal optimality and covariance matrix estimators. Ph.D. Thesis, ETH, Zurich.