

Minimum dissimilarity estimators

Stimatori di minima dissomiglianza

Eugenio Regazzini¹

Dipartimento di Matematica “Felice Casorati”

Università degli Studi di Pavia

eugenio@dimat.unipv.it

Riassunto: Questa nota è dedicata allo studio di stimatori puntuali di parametri, definiti come minimi assoluti della dissomiglianza di una distribuzione empirica dai vari elementi di un dato modello statistico. Dopo una breve discussione sul significato di questa proposta, si procede alla presentazione di alcune soluzioni dei problemi di esistenza e misurabilità inerenti agli stimatori in questione. Quindi, si danno due gruppi di condizioni sufficienti per la consistenza. Lo studio delle proprietà asintotiche prosegue con una caratterizzazione delle leggi limite, in base alla quale tali leggi coincidono con quelle dei minimi di certi funzionali integrali del ponte browniano standard. Nel caso di modelli con parametri di scala o posizione lo studio viene ulteriormente approfondito nella direzione della determinazione esplicita delle distribuzioni dei suddetti minimi. Infine, viene effettuato un confronto tra gli stimatori proposti e taluni stimatori classici, dal punto di vista dell'efficienza e della robustezza.

Keywords: Argmax theorem, asymptotic laws of minimum dissimilarity estimates, consistency of minimum dissimilarity estimates, Feynman–Kac formula, minimum dissimilarity estimates.

1. Introduction

Let $(\xi_n)_{n \geq 1}$ be a sequence of random variables associated with a sequence of observations whose range is a subset of $X = \mathbb{R}^N$. Let $\mathcal{X}$ denote the class of Borel sets of X and d the Euclidean metric on $\mathbb{R}^N$. For each value θ of an unknown parameter taking values in Θ , consider the probability distribution p_θ for $(\xi_n)_{n \geq 1}$ which makes the ξ_n s independent and identically distributed according to the probability measure α_θ . Hence, $p_\theta = \alpha_\theta^\infty$ for each θ in Θ . By $\mathbb{P}$ denote the set of all probability measures on $(X, \mathcal{X})$, and consider $\mathcal{M} := \{\alpha_\theta : \theta \in \Theta\} \subset \mathbb{P}$ as the statistical model of interest.

Given the n -sample $\xi(n) := (\xi_1, \dots, \xi_n)$, most frequentistic estimation methods are founded on the minimization of some measure of the discrepancy between $\mathcal{M}$ and the empirical distribution ν_n of $\xi(n)$. The present paper deals with estimates defined as minimizers of Gini's indices of dissimilarity. The connections of these distances with optimal transportation problems can be invoked to give significant reasons for their adoption in order to introduce new estimation methods. In fact, the problem of minimizing the discrepancy between ν_n and $\mathcal{M}$ can be described in the following terms. To each α_θ in $\mathcal{M}$ associate a hole whose volume function is described by $\alpha_\theta(dy)$, and think of ν_n as a model

¹Address for correspondence: Eugenio Regazzini, Dipartimento di Matematica, Università degli Studi di Pavia, Via Ferrata 1, 27100 Pavia
Research supported in part by MIUR, PRIN 2002 “*Bayesian nonparametric methods and their applications*”

of a pile of sand. See Villani (2003). Now, the problem of estimating θ could be seen as the problem of determining the hole whose conformation better corresponds to the conformation of ν_n . In order to quantify the dissimilarity between conformations, suppose you must transfer the sand to one of the above holes, and the cost of moving one unit of mass from location x to location y (in X) is $c(x, y) = d(x, y)^r$ for some $r \geq 1$. Then, ν_n can be considered as less dissimilar from α_{θ_1} than from α_{θ_2} if there exists a transference plan of ν_n to α_{θ_1} less expensive than any transference plan of ν_n to α_{θ_2} . Hence, if $G^{(r)}(\nu_n, \alpha_\theta)$ denotes the “risk” – with respect to the cost function d^r – of the less expensive transference plan of ν_n to $\mathcal{M}$, the hole characterized by $\hat{\theta}_n$, yielding the minimum risk $G^{(r)}(\nu_n, \alpha_{\hat{\theta}_n})$, is the one which exhibits, within $\mathcal{M}$, the least dissimilarity from ν_n , with respect to the cost function d^r .

It is well-known that $G^{(r)}(\nu_n, \alpha_\theta)$ is a dissimilarity index (in the sense of Gini), i.e. a particular form of Kantorovich (semi-)distance. See, in addition to Villani (2003), Section 1.1.8 of Dudley (1989) and the books of Rachev (1991), Rachev and Rüschendorf (1998). Throughout the present paper, $\hat{\theta}_n$ will be called *minimum dissimilarity estimator* (abbreviated as MDE).

A deep study on the asymptotic distribution of the Gini dissimilarity index of order 1 is contained in del Barrio *et al.* (1999b). Moreover, del Barrio *et al.* (1999a), and del Barrio *et al.* (2000) use dissimilarity indices of order 2 to define suitable tests for location-scale families of distributions. A study in the very same field is developed in Csörgő (2002) and de Wet (2002) for tests based on weighted dissimilarity indices. The sole application to the theory of estimation is, to our knowledge, Bertino (1971), which gives an explicit form of MDE of order 2 for scale-position parameters.

The present paper summarizes some of the material contained in recent papers jointly written with Antonella Bodini and Federico Bassetti. See Bassetti *et al.* (2004), Bassetti and Regazzini (2003a) and Bassetti and Regazzini (2003b). Section 2 contains some preliminary information about the Kantorovich theory. MDEs are defined, and their existence and measurability are studied, in Section 3. Sufficient conditions for consistency of these estimates are given in Section 4. Sections 5 and 6 deal with the limiting distributions of rescaled forms of MDEs. Finally, a partial comparative analysis – with respect to efficiency and robustness – is developed in Section 7.

2. Preliminaries

In this paper, as said in the previous section, the costs are expressed as powers of the Euclidean distance d on $X = \mathbb{R}^N$, i.e. $c(x, y) = d(x, y)^r$ for some $r \geq 1$. Define

$$\mathbb{P}^{(r)} = \left\{ p \in \mathbb{P} : \int_X c(x, y) p(dx) < +\infty \right\}$$

for some (or, equivalently, for every) y in X and, for every p and q in $\mathbb{P}$, consider the class $\mathbb{P}(p, q)$ of all probability measures on $(X^2, \mathcal{X}^2)$ with marginals p and q . Given any p and q in $\mathbb{P}^{(r)}$, every element π in $\mathbb{P}(p, q)$ can be viewed as an “admissible” transference plan, with respect to the cost c . It is well-known that there is γ in $\mathbb{P}(p, q)$ such that $G^{(r)}(p, q) = \inf \left\{ \int_{X^2} c d\pi : \pi \in \mathbb{P}(p, q) \right\}$ turns out to be represented as

$$G^{(c)}(p, q) = \int_{X^2} c d\gamma. \quad (2.1)$$

See Corollary 5.2.2 in Rachev (1991); for $X = \mathbb{R}$, this result had been proved by Dall'Aglia (1956) and, for distributions with finite support, by Salvemini (1939). Whence $G^{(r)}(p, q)$ gives the *risk of the optimal transference plan* of p to q . It is a semi-distance on $\mathbb{P}^{(r)}$, whereas $(G^{(r)})^{1/r}$ is a distance on $\mathbb{P}^{(r)}$. See Corollary 5.2.1 in Rachev (1991). This distance represents a natural extension of the concept of *Gini's dissimilarity index*. See Gini (1915). For $X = \mathbb{R}$, the fact that $(G^{(r)})^{1/r}$ is a distance has been established by Landenna (1956). $G^{(1)}$ is also known as simple dissimilarity index, or *Wasserstein metric*.

The descriptive power of $G^{(r)}$ is evidenced by the following property:

(2.2) *If p and p_n belong to $\mathbb{P}^{(r)}$ for every n , then $G^{(r)}(p, p_n) \rightarrow 0$ as $n \rightarrow +\infty$ if and only if $p_n \Rightarrow p$ and $\int_X c(x, y) p_n(dx) - \int_X c(x, y) p(dx) \rightarrow 0$ as $n \rightarrow +\infty$ for some (or, equivalently, for every) y in X . (As usual, the symbol $\Rightarrow$ designates weak convergence.)*

A fact of great significance, with respect to the study of the asymptotic properties of MDE's, is given by

(2.3) *There is a set $\mathcal{N}$ in $\mathcal{X}^\infty$ of p_{θ_0} -probability 0 such that*

$$(2.3)_1 \quad \nu_{n,x} \Rightarrow \alpha_{\theta_0} \quad (x \in \mathcal{N}^c)$$

($\nu_{n,x}$ denotes the path of ν_n corresponding to the sequence of observations x in X^∞); moreover, if α_{θ_0} belongs to $\mathbb{P}^{(r)}$, then

$$(2.3)_2 \quad \int_X c(t, y) \nu_{n,x}(dt) \rightarrow \int_X c(t, y) \alpha_{\theta_0}(dt) \quad (x \in \mathcal{N}^c, y \in X).$$

In view of (2.2), conditions (2.3)₁–(2.3)₂ are equivalent to

$$G^{(r)}(\nu_{n,x}, \alpha_{\theta_0}) \rightarrow 0 \quad (x \in \mathcal{N}^c).$$

To some extent, existence and measurability of MDE's depend on semicontinuity of $G^{(r)}$:

(2.4) $G^{(r)} : \mathbb{P}^{(r)} \times \mathbb{P}^{(r)} \rightarrow \bar{\mathbb{R}}^+$ is a lower semicontinuous function whenever $\mathbb{P}^{(r)}$ is endowed with the topology of weak convergence.

3. Minimum dissimilarity estimators: definition, existence and measurability

In the sequel, for the sake of notational simplicity, $G_n^{(r)}(\theta)$ will stand for $G_n^{(r)}(\nu_{n,x}, \alpha_\theta)$; moreover, the whole of the following conditions from (a) to (d) will be referred to as *General Conditions*, (GC) for short: (a) $X = \mathbb{R}^N$, $\mathcal{X}$ = Borel class on $\mathbb{R}^N$; (b) Θ is some subspace of a metric space S with distance function d_Θ ; (c) $\mathcal{M}$ is identifiable (i.e., $\alpha_{\theta_1} \neq \alpha_{\theta_2}$ if $\theta_1 \neq \theta_2$) and is a subset of $\mathbb{P}^{(r)}$. Now, the main definition can be given.

Let (GC) be valid. If, for some x in X^∞ and n in $\mathbb{N}$, the element $\hat{\theta}_n = \hat{\theta}_n(\xi^{(n)}(x))$ of Θ satisfies

$$G_n^{(r)}(\hat{\theta}_n) = \inf \{G_n^{(r)}(\theta) : \theta \in \Theta\} \quad (3.1)$$

then it is called *minimum dissimilarity estimate* (MDE) of θ . In the sequel, $\hat{\theta}_n$ will be thought of as a function of x , written, for notational simplicity, as $\hat{\theta}_n = \hat{\theta}_n(x)$.

Proposition (2.4) can be used to show that $C(t) := \{p \in \mathbb{P}^{(r)} : G(\alpha_{\theta_0}, p) \leq t\}$ is closed, for any θ_0 in Θ , in the topology of weak convergence; moreover, it is not difficult to prove that, in the same topology, $C(t)$ is relatively compact. In other words, $C(t)$ is

compact for every t . This fact, combined with the lower semicontinuity of $\alpha_\theta \mapsto G_n^{(r)}(\theta)$, leads to conclude that $C_n(t) := \{\alpha_\theta : G_n^{(r)}(\theta) \leq t\}$ is compact. Whence, by lower semicontinuity, one has

Theorem 3.1. *Under (GC), a MDE for θ , say $\hat{\theta}_n(x)$, exists for every x in X and n in $\mathbb{N}$, whenever $\mathcal{M}$ is closed in the topology of weak convergence.*

In general, $x \mapsto \hat{\theta}_n(x)$ is not measurable, i.e. it is not a random variable and, therefore, the problem of the measurability of a MDE deserves careful consideration. The following measurability criterion for MDEs follows from Corollary 1 in Brown and Purves (1973).

Theorem 3.2 *Let (GC) be in force, Θ being a σ -compact subset of $\mathbb{R}^M$ (or, more generally, of some Polish space). Moreover, assume $\alpha_{\theta_m} \Rightarrow \alpha_\theta$ for any sequence $(\theta_m)_{m \geq 1}$ and θ in Θ such that $d_\Theta(\theta_m, \theta) \rightarrow 0$. Then, $x \mapsto \hat{\theta}_n(x)$ is a random element on $(X^\infty, \mathcal{X}^\infty)$ with values in Θ equipped with the σ -algebra coinciding with the trace of the Borel class of $\mathbb{R}^M$ on Θ .*

4. Consistency of MDEs

As for the consistency of MDEs, as defined by (3.1), it should be noted that (2.3)₁–(2.3)₂ entail (a.s.– p_{θ_0})

$$\begin{aligned} \limsup_n G^{(r)}(\alpha_{\hat{\theta}_n}, \alpha_{\theta_0}) &\leq \liminf_n 2^r \{G^{(r)}(\alpha_{\hat{\theta}_n}, \nu_{n,x}) + G^{(r)}(\nu_{n,x}, \alpha_{\theta_0})\} \\ &\leq \liminf_n 2 \cdot 2^r G^{(r)}(\nu_{n,x}, \alpha_{\theta_0}) = 0. \end{aligned}$$

Hence, one can say that $(\alpha_{\hat{\theta}_n})_{n \geq 1}$ is a consistent sequence of estimators for α_{θ_0} , with respect to the metric generated by the dissimilarity index $G^{(r)}$. More precisely,

Theorem 4.1. *If (GC) are valid, then for any θ_0 in Θ there exists a p_{θ_0} -null set $\mathcal{N}$ such that*

$$\lim_{n \rightarrow +\infty} G^{(r)}(\alpha_{\hat{\theta}_n(x)}, \alpha_{\theta_0}) = 0$$

holds true for any x in $\mathcal{N}^c$, $\hat{\theta}_n(x)$ being defined as in (3.1).

As for consistency of $(\hat{\theta}_n)_{n \geq 1}$, it is easy to deduce, from Theorem 4.1, the following criterion.

Theorem 4.2. *Let (GC) be in force together with*

(i) $G^{(r)}(\alpha_{\theta_n}, \alpha_\theta) \not\rightarrow 0$ whenever $d_\Theta(\theta, \theta_n) \not\rightarrow 0$.

Then, there exists a p_{θ_0} -null set $\mathcal{N}$ such that

$$\hat{\theta}_n(x) \rightarrow \theta_0 \quad \text{as } n \rightarrow +\infty$$

obtains for every x in $\mathcal{N}^c$, $\hat{\theta}_n$ being defined according to (3.1).

This proposition states that $\hat{\theta}_n \rightarrow \theta_0$ almost surely with respect to the inner probability associated to p_{θ_0} , i.e. $\hat{\theta}_n(x) \rightarrow \theta_0$, as $n \rightarrow +\infty$, for every x in the complement of a p_{θ_0} -null subset of X^∞ . See, for example, Sections 1.2, 1.9 of van der Vaart and Wellner (1996).

It is interesting to observe that condition (i) in Theorem 4.2 is met by a large subclass of exponential families; see Theorem 8.3 in Barndorff-Nielsen (1978). So, MDEs turn out

to be consistent for the parameters of these exponential families. It is worth noting that, in this case, MDEs are measurable if, for example, Θ is open; in fact, under this assumption, Theorem 3.2 can be used directly.

If it is difficult to verify condition (i) in Theorem 4.2, then an alternative criterion for checking consistency is provided by

Theorem 4.3. *Let (GC) be valid together with*

(i) $\alpha_{\theta_n} \Rightarrow \alpha_\theta$ for every θ and $(\theta_n)_{n \geq 1} \subset \Theta$, such that $d_\Theta(\theta_n, \theta) \rightarrow 0$.

Moreover, let I be a subspace of Θ such that θ_0 is an interior point of I and the restriction of $\theta \mapsto G^{(r)}(\alpha_\theta, \alpha_{\theta_0})$ to this subspace is a coercive function. Then there exists a p_{θ_0} -null set $\mathcal{N}$ such that

$$\hat{\theta}_n(x) \rightarrow \theta_0 \quad \text{as } n \rightarrow +\infty$$

obtains for every x in $\mathcal{N}^c$, $\hat{\theta}_n$ being defined according to (3.1).

5. Asymptotic approximation for laws of MDEs: generalities

The following sections are devoted to the study of the limiting laws of sequences of MDEs derived from $G_n^{(r)}$, when (GC) are in force combined with the following additional conditions: $X = \mathbb{R}$; Θ is a σ -compact subset of $S = \mathbb{R}^M$ and d_Θ stands for the Euclidean metric on $\mathbb{R}^M$. For notational simplicity, the whole of the resulting modified version of (GC) will be indicated by $(GC)_s$. Moreover, the symbols A_n and A_θ are adopted for the (empirical) distribution function associated with ν_n and for the distribution function associated with α_θ , respectively. In preparation for the application of what is called the *argmax* theorem – see, for example, van der Vaart and Wellner (1996) – it should be noted that the following hypotheses from (H_1) to (H_5) are common to the arguments developed in the rest of the paper.

For every θ_0 in Θ :

(H₀) $A_{\theta_0}(x) = \int_{-\infty}^x a_{\theta_0}(t) dt$ for every x in $\mathbb{R}$.

(H₁)

$$\int_{(0,1)} \frac{[\sqrt{y(1-y)}]^r}{a_{\theta_0}(A_{\theta_0}^{-1}(y))^r} dy < +\infty.$$

(H₂) The generalized inverse $A_{\theta_0}^{-1}$ (i.e. $A_{\theta_0}^{-1}(t) = \sup\{x : A_{\theta_0}(x) \leq t\}$ for t in $[0, 1)$) is such that

$$A_{\theta_0+h/\sqrt{n}}^{-1}(y) = A_{\theta_0}^{-1}(y) + \frac{1}{\sqrt{n}}(hg(y) + \Delta_n(y, h))$$

holds true for every h in $\mathbb{R}^M$ and y in $(0, 1)$, with $g = (g_1, \dots, g_M)$ satisfying $\int_{(0,1)} |g_k(y)|^r dy < +\infty$ ($k = 1, \dots, M$) and, for every $c \geq 0$,

$$\sup_{|h| \leq c} \int_0^1 |\Delta_n(y, h)|^r dy = o(1) \quad \text{as } n \rightarrow +\infty.$$

(H₃)

$$\int_{(0,1) \setminus (\frac{1}{n+1}, \frac{n}{n+1})} |\mathbb{W}_n(y)|^r dy = o_{p_{\theta_0}}(1) \quad \text{as } n \rightarrow +\infty$$

with $\mathbb{W}_n := \sqrt{n}(A_n^{-1} - A_{\theta_0}^{-1})$.

(H₄) $A_{\theta_0} \in C^2(b_0, c_0)$, $a_{\theta_0} > 0$ on (b_0, c_0) with $b_0 := \sup\{x : A_{\theta_0}(x) = 0\} < c_0 =: \sup\{x : A_{\theta_0}(x) < 1\}$, and

$$\sup_{0 < y < 1} \frac{y(1-y)|\dot{a}_{\theta_0}(A_{\theta_0}^{-1}(y))|}{a_{\theta_0}(A_{\theta_0}^{-1}(y))^2} < +\infty.$$

(H₅) For every $c > 0$, there are $\delta = \delta(c) > 0$ and $k = k(c)$ implying

$$\sup_{(\star)} d_{\Theta}(\theta, \theta_0) \leq \frac{k}{\sqrt{n}} \quad \text{with} \quad (\star) := \left\{ \theta \in \mathbb{R}^M : d_{\Theta}(\theta, \theta_0) \leq \delta, G_r(A_{\theta}, A_{\theta_0}) \leq \frac{c}{\sqrt{n}} \right\}.$$

It should be recalled that, for $N = 1$, $G_n^{(r)}$ admits the representation

$$G_n^{(r)}(\theta) = \int_{(0,1)} |A_{\theta}^{-1}(y) - A_n^{-1}(y)|^r dy$$

Furthermore, in view of (GC)_s and (H₂), it can be shown, through Theorem 3.2, that the estimators $\hat{\theta}_n$, defined by (3.1), are random elements.

Now, for every n , define $\hat{h}_n$ as

$$\hat{h}_n := \sqrt{n}(\hat{\theta}_n - \theta_0),$$

set $\mathbb{B}$ for a standard *Brownian bridge* on $[0, 1]$ and put

$$\mathbb{Q}(\cdot) := \mathbb{B}(\cdot)/a_{\theta_0}(A_{\theta_0}^{-1}(\cdot)), \quad \mathbb{M}_n(h) := \sqrt{n}G_n^{(r)}(\theta_0 + h/\sqrt{n})^{1/r}.$$

For any compact subset K of $\mathbb{R}^M$, consider $\mathbb{M}_n$ as a stochastic process indexed by h in K , with sample paths in the space $l^\infty(K)$ of all bounded real functions on K , i.e. of all functions $z : K \rightarrow \mathbb{R}$ such that

$$\|z\|_K := \sup_{h \in K} |z(h)| < +\infty.$$

The symbol $\xrightarrow{\mathcal{L}_0}$ denotes the convergence in law when $(\tilde{x}_n)_{n \geq 1}$ has the probability distribution p_{θ_0} .

In this framework, Theorem 3.2.2 in van der Vaart and Wellner (1996) becomes

Theorem 5.1. *Let (GC)_s, (H₀), (H₁) be valid. Assume that*

$$\mathbb{M}_n \xrightarrow{\mathcal{L}_0} \mathbb{M} \quad (\text{as } n \rightarrow +\infty) \tag{5.1}$$

in $l^\infty(K)$ for every compact subset K of $\mathbb{R}^M$, with

$$\mathbb{M}(h) := \left\{ \int_{(0,1)} |\mathbb{Q}(y) - hg(y)|^r dy \right\}^{1/r}$$

and g in $(L^r(0, 1))^M$. Let $(\hat{h}_n)_{n \geq 1}$ be tight under p_{θ_0} , and $\mathbb{M}$ possess a unique minimum in a (random) point $\hat{h}$. Then, $\hat{h}_n \xrightarrow{\mathcal{L}_0} \hat{h}$ as $n \rightarrow +\infty$.

Note that when $\int_{(0,1)} |hg|^r < +\infty$, (H₁) can be exploited, according to the results in Subsection 5.3 in Csörgő and Horváth (1993), to establish that $\mathbb{M}(h) < +\infty$ is valid, with p_{θ_0} -probability 1, for each h in $\mathbb{R}^M$.

As for the validity of (5.1) and of the tightness for $(\hat{h}_n)_{n \geq 1}$, the following statements can be useful.

Theorem 5.2. *Suppose $(GC)_s$, (H_0) – (H_4) hold. Then,*

$$\mathbb{M}_n \xrightarrow{\mathcal{L}_0} \mathbb{M} \quad (\text{as } n \rightarrow +\infty)$$

in $l^\infty(K)$ for every compact subset K of $\mathbb{R}^M$, g being the same as in (H_2) .

Theorem 5.3. *Let $(GC)_s$ and (H_5) be in force. Moreover, assume $(\hat{\theta}_n)_{n \geq 1}$ is weakly consistent and $\sqrt{n}G_n^{(r)}(\theta_0)^{1/r}$ converges in law. Then $(\hat{h}_n)_{n \geq 1}$ is tight under p_{θ_0} .*

In Theorem 5.3 the assumption of consistency for $(\hat{\theta}_n)_{n \geq 1}$ can be omitted when (H_5) is substituted with the stronger condition

(H'_5) *For every $c > 0$, there is $k = k(c)$ implying*

$$\sup_{\{\theta: G^{(r)}(\alpha_\theta, \alpha_{\theta_0})^{1/r} \leq c/\sqrt{n}\}} d_\Theta(\theta, \theta_0) \leq \frac{k}{\sqrt{n}}.$$

Moreover, (H_0) , (H_1) , (H_3) and (H_4) imply that $\sqrt{n}G_n^{(r)}(\theta_0)^{1/r}$ converges in law and, therefore, Theorem 5.3 can be restated as in the following

Theorem 5.4. *If $(GC)_s$, (H_0) , (H_1) , (H_3) , (H_4) and (H'_5) hold, then the sequence $(\hat{h}_n)_{n \geq 1}$ is tight under p_{θ_0} .*

6. Asymptotic approximation for laws of MDEs: location–scale families

A few of the conditions under which the argmin argument works can be easily verified when $\mathcal{M}$ is a *location–scale family*, that is

$$\begin{aligned} \Theta &= \{ \theta = (\mu, \sigma) : \mu \in \mathbb{R}, \quad \sigma > 0 \} \\ A_\theta(x) &= A(\sigma^{-1}(x - \mu)) \end{aligned} \tag{6.1}$$

A being any fixed nondegenerate distribution function such that $\int_{\mathbb{R}} |x|^r dA(x) < +\infty$. Indeed, since

$$A_\theta^{-1}(y) = \sigma A^{-1}(y) + \mu$$

holds for every y in $(0, 1)$ and θ in $\mathbb{R} \times (0, +\infty)$, then (H_2) is trivially satisfied with $\Delta_n = 0$ and $g(y) = (1, A^{-1}(y))$. Furthermore, as far as (H_5) is concerned, the following can be proved

Lemma 6.1. *If $\mathcal{M}$ is defined by (6.1) and satisfies (H_0) , then (H'_5) holds true.*

As an example, it is worth recalling that the argmin theorem can be applied, using the results established up to this point, to: (a) the family of uniform laws or, more generally, of distributions with bounded support; (b) the Gaussian family; (c) the Weibull scale family.

The problem of uniqueness of $\hat{h}$ is discussed in Subsection 4.1 of Bassetti and Regazzini (2003a) in the case of estimates from the dissimilarity index of order 1 (i.e. $r = 1$), when $\mathcal{M}$ is a location–scale family, under some extra–condition.

In the same paper, the reader can find a method for obtaining the asymptotic distribution of minimum dissimilarity estimates (with $r = 1$) of a *position* parameter. Apropos of this, the results presented in Section 5 can be used to prove

Theorem 6.1. *Let (H_1) be valid, and (6.1) holds with $\sigma = \sigma_0$. Moreover, assume that*

$$\mathbb{M}(h) = \int_{(0,1)} \left| \frac{\sigma_0 \mathbb{B}(y)}{a(A^{-1}(y))} - h \right| dy$$

has a unique minimum $\hat{h}$. Then, the limiting law of $\sqrt{n}(\hat{\theta}_n - \mu_0)$, under p_{θ_0} with $\theta_0 = (\mu_0, \sigma_0)$, is symmetric, coincides with the distribution of $\hat{h}$, and

$$p_{\theta_0} \left\{ \hat{h} > \xi \right\} = \frac{1}{2} - \frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+, c \rightarrow +\infty} \int_{\epsilon}^c \operatorname{Im} \left[\frac{e^{-it/2}}{t} \phi(t; \xi) \right] dt$$

where $\phi(\cdot; \xi)$ stands for the characteristic function of $\int_{(0,1)} \mathbb{I}_{(-\infty, \gamma(s)]}(\mathbb{B}(s)) ds$, with $\gamma(\cdot) = \xi a(A^{-1}(\cdot))/\sigma_0$.

Hence, in order to obtain explicit forms for the distribution function of $\hat{h}$, one ought to determine $\phi(\cdot; \xi)$. As shown in Bassetti and Regazzini (2003b), this can be done by resorting to a suitable extension of the celebrated Feynman–Kac theorem. See the contributed talk of Federico Bassetti at this Meeting.

Theorem 6.2. *Let $\gamma(\cdot) = \xi a(A^{-1}(\cdot))/\sigma_0$ belong to $C^1(0, 1)$ and $\dot{\gamma}$ be in $L^p(0, 1)$ for some $p > 2$. Then, for any $\xi > 0$,*

$$\phi(t; \xi) = e^{it} \left(1 + \frac{1}{2} \int_0^1 \frac{1}{\sqrt{1-u}} \exp \left\{ -\frac{\gamma(1-u)^2}{2(1-u)} - iut \right\} \nu_0^-(u/2, -it) du \right)$$

where (ν_0^-, ν_0^+) is the unique continuous solution of the system

$$\left\{ \begin{array}{l} \frac{(1-e^{-2zy})}{2\sqrt{\pi y}} \exp\left\{-\frac{(r(y))^2}{4y}\right\} + \int_0^y \frac{1}{2\sqrt{\pi(y-\tau)}} \nu_0^+(\tau) \exp\left\{-\frac{(r(y)-r(\tau))^2}{4(y-\tau)}\right\} d\tau \\ = \int_0^y \frac{1}{2\sqrt{\pi(y-\tau)}} \nu_0^-(\tau) \exp\left\{-2z(y-\tau) - \frac{(r(y)-r(\tau))^2}{4(y-\tau)}\right\} d\tau \\ - \frac{(1-e^{-2zy})}{2\sqrt{\pi y^{3/2}}} r(y) \exp\left\{-\frac{(r(y))^2}{4y}\right\} = \nu_0^+(y) + \nu_0^-(y) \\ + \int_0^y \frac{r(y)-r(\tau)}{2\sqrt{\pi(y-\tau)^{3/2}}} \nu_0^+(\tau) \exp\left\{-\frac{(r(y)-r(\tau))^2}{4(y-\tau)}\right\} d\tau \\ - \int_0^y \frac{r(y)-r(\tau)}{2\sqrt{\pi(y-\tau)^{3/2}}} \nu_0^-(\tau) \exp\left\{-2z(y-\tau) - \frac{(r(y)-r(\tau))^2}{4(y-\tau)}\right\} d\tau \end{array} \right.$$

and $r(y) = \gamma(1-2y)$ for y in $(0, 1/2]$.

As an example, if

$$a_{\theta}(x) = \frac{1}{\sigma_0} e^{-(x-\mu)/\sigma_0} \mathbb{I}_{(\mu, +\infty)}(x),$$

Theorem 6.2 yields

$$p_{\theta_0} \left\{ \sqrt{n}(\hat{\theta}_n - \mu_0) \leq \xi \right\} \rightarrow p_{\theta_0} \left\{ \hat{h} \leq \xi \right\} = 1 - \frac{1}{2\sqrt{\pi}} \Psi \left(\frac{3}{2}, \frac{1}{2}, 2\tilde{\xi}^2 \right) e^{-4\tilde{\xi}^2}$$

$\tilde{\xi} = \xi/\sigma_0 > 0$, and Ψ stands for the confluent hypergeometric function of the second type (the Tricomi function).

7. Comparing minimum dissimilarity with maximum likelihood estimates in some particular models

In this final section, minimum dissimilarity estimates are compared to maximum likelihood estimates of location parameters in some statistical models. On the one hand, one finds that maximum likelihood estimates behave more efficiently than their competitors with respect to the construction of confidence intervals, but on the other hand, minimum dissimilarity estimates are preferable from the point of view of robustness. It should be noted that both statements are based on considerations of an asymptotic nature.

Consider the problem of constructing an asymptotic confidence interval for the position parameter μ of the exponential model given at the end of the previous section. In view of the expression of the limiting distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$, for any α in $(0, 1)$, set $d_\alpha > 0$ in such a way that

$$\alpha = \frac{1}{\pi} \Psi \left(\frac{3}{2}, \frac{1}{2}, 2d_\alpha^2 \right) e^{-4d_\alpha^2}$$

and consider the random interval $\tilde{\mathcal{I}}_{\alpha,n} = (\hat{\mu}_n - \sigma_0 d_\alpha / \sqrt{n}, \hat{\mu}_n + \sigma_0 d_\alpha / \sqrt{n})$. Then,

$$p_{\theta_0} \left\{ \mu_0 \in \tilde{\mathcal{I}}_{\alpha,n} \right\} \sim 1 - \alpha \quad (n \rightarrow +\infty)$$

holds true for every μ_0 in $\mathbb{R}$. For the sake of comparison, recall that for the maximum likelihood estimator $\mu_n^* = \min\{\tilde{x}_1, \dots, \tilde{x}_n\}$ of μ

$$p_{\theta_0} \left\{ \frac{\mu_n^* - \mu_0}{-\sigma_0 \log(1 - 1/n)} \leq y \right\} \rightarrow \mathbb{I}_{(0,+\infty)}(y)(1 - e^{-y}).$$

Hence, letting $d_\alpha^* = -\log \alpha$ and $\tilde{\mathcal{I}}_{\alpha,n}^* = (\mu_n^* - \sigma_0 \log(\alpha) \log(1 - 1/n), \mu_n^*)$, and

$$p_{\theta_0} \left\{ \mu_0 \in \tilde{\mathcal{I}}_{\alpha,n}^* \right\} \sim 1 - \alpha$$

holds as $n \rightarrow +\infty$ for any μ in $\mathbb{R}$. At this stage, $\tilde{\mathcal{I}}_{\alpha,n}^*$ is immediately seen to be shorter than $\tilde{\mathcal{I}}_{\alpha,n}$ from an asymptotic point of view.

As for the robustness of the methods presented here, some numerical experiments suggest that $G^{(1)}$ -minimum estimates are the estimates most resistant to the intrusion of “bad” observations. Because the “ L^1 –estimates” lack differentiability, their robustness cannot be measured by the usual influence functions. So, to compare minimum dissimilarity estimates from that perspective, one can resort to quantitative measures, or indices, derived from the Donoho and Liu concept of *bias distortion curve* (see Donoho and Liu (1988a) and Donoho and Liu (1988b)), defined by

$$r = r(\pi_\xi; \epsilon) := \sup \left\{ |\bar{\theta} - \theta_0| : \bar{\theta} \in \operatorname{argmin}_\theta L(\alpha(\theta), \rho_\epsilon(\theta, \pi_\xi)) \right\}$$

where L is some measure of discrepancy between probabilities (dissimilarity indices, in this paper),

$$\rho_\epsilon(\theta, \pi_\xi)(A) = (1 - \epsilon)\alpha(\theta_0; A) + \epsilon\pi_\xi(A) \quad (A \in \mathcal{X})$$

and π_ξ is any probability measure, affected by the position parameter ξ , which contaminates the limit α_{θ_0} of ν_n .

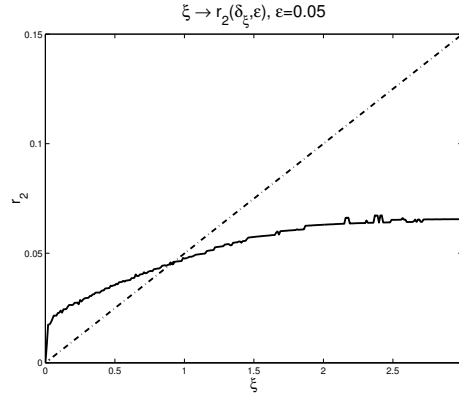


Figure 1: Solid lines refer to $G^{(1)}$ -estimates; dashdotted to $G^{(2)}$ -estimates

Using r to check the robustness of minimum dissimilarity estimates of the mean of a Gaussian law with variance 1, one obtains the results shown in Figure 1 as graphs of $\xi \mapsto r(\pi_\xi; \epsilon)$ for $\epsilon = 0.05$, $\pi_\xi = \delta_\xi$ and $L = G^{(i)}$ with $i = 1, 2$. Since, in this case, r is an even function, then its curve is plotted only for positive ξ 's.

As for the maximum likelihood estimates, it should be noted that, in the particular case of a Gaussian model with known variance, the $G^{(2)}$ -minimum estimate is the same as the maximum likelihood estimate. Thus, the curves in Figure 1 can be used to compare the $G^{(1)}$ -minimum estimate and the maximum likelihood estimate from the standpoint of robustness.

Finally Figure 2 exhibits the case of an exponential model of unknown position parameter μ and unitary scale, with $\epsilon = 0.01$, $\pi_\xi = \delta_\xi$ and $L = G^{(i)}$ with $i = 1, 2$.

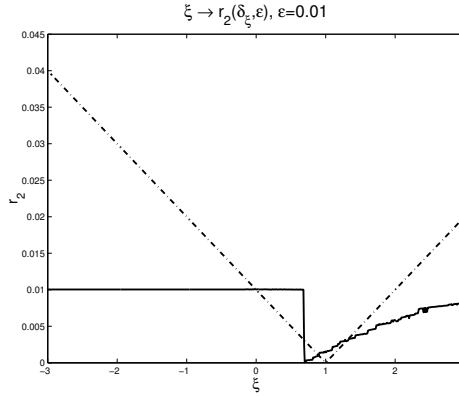


Figure 2: Solid lines refer to $G^{(1)}$ -estimates; dashdotted to $G^{(2)}$ -estimates

Unlike the case of Figure 1, in comparing the minimum dissimilarity estimates and maximum likelihood estimates for the position parameter of the exponential model, one cannot find any estimate of the former type which coincides with $\mu_n^* := \min\{\tilde{x}_1, \dots, \tilde{x}_n\}$. Thus, one can decide to think of maximum likelihood estimates as minima of

$$L(A_\mu, A_n) := \begin{cases} -\int_{\mathbb{R}^+} \log(\exp\{-(x - \mu)\}) dA_n(x) & \text{if } \mu \leq \min\{\tilde{x}_1, \dots, \tilde{x}_n\} \\ +\infty & \text{if } \mu > \min\{\tilde{x}_1, \dots, \tilde{x}_n\}, \end{cases}$$

where $A_\mu(x) := (1 - e^{-(x-\mu)}) \mathbb{I}_{[0,+\infty)}(x - \mu)$, and then proceed to evaluate r to obtain $r(\delta_\xi; \epsilon) = |\xi| \mathbb{I}_{(-\infty, 0]}(\xi)$. Comparing this function with the one represented in Figure 2, one can conclude that, even if the $G^{(i)}$ -dissimilarity estimates ($i = 1, 2$) also behave very well in this case, the robustness of the maximum likelihood estimate is invincible if $\xi > 0$. On the other hand, if $\xi \leq 0$, the maximum likelihood estimate is extremely poor from the point of view of robustness, while the behavior of the $G^{(i)}$ -minimum estimates ($i = 1, 2$) is rather good in this case as well.

References

- Barndorff-Nielsen O. (1978) *Information and Exponential Families in Statistical Theory*, Wiley, Chichester.
- Bassetti F., Bodini A. and Regazzini E. (2004) Consistency of minimum Kantorovich distance estimators, Technical Report 2004/4–MI, IMATI–CNR, Milano.
- Bassetti F. and Regazzini E. (2003a) Asymptotic approximations for laws of minimum dissimilarity estimates, Technical Report 2003/3–MI, IMATI–CNR, Milano.
- Bassetti F. and Regazzini E. (2003b) Probability law for the occupation time of a pinned Brownian motion with arbitrary drift, Technical Report 2003/5–MI, IMATI–CNR, Milano.
- Bertino S. (1971) Gli indici di dissomiglianza e la stima dei parametri, *Studi di Probabilità, Statistica e Ricerca Operativa in Onore di Giuseppe Pompilj*, 187–202.
- Brown L.D. and Purves R. (1973) Measurable selections of extrema, *The Annals of Statistics*, 1, 902–912.
- Csörgő M. and Horváth L. (1993) *Weighted Approximations in Probability and Statistics*, Wiley, Chichester.
- Csörgő S. (2002) Weighted correlation tests for scale families, *Test*, 11, 219–248.
- Dall’Aglio G. (1956) Sugli estremi dei momenti delle funzioni di ripartizione doppie, *Annali Scuola Normale Superiore di Pisa*, 10, 35–74.
- de Wet T. (2002) Goodness-of-tests for location and scale families based on weighted l_2 -Wasserstein distance measure, *Test*, 11, 89–107.
- del Barrio E., Cuesta-Albertos J.A. and Matrán C. (2000) Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests, *Test*, 9, 1–96.
- del Barrio E., Cuesta-Albertos J.A., Matrán C. and Rodríguez-Rodríguez J.M. (1999a) Tests of goodness of fit based on the l_2 -Wasserstein distance, *The Annals of Statistics*, 27, 1230–1239.
- del Barrio E., Giné E. and Matrán C. (1999b) Central limit theorems for the Wasserstein distance between the empirical and the true distributions., *The Annals of Probability*, 27, 1009–1071.
- Donoho D. and Liu R. (1988a) The “automatic” robustness of minimum distance functionals, *The Annals of Statistics*, 16, 552–586.
- Donoho D. and Liu R. (1988b) Pathologies of some minimum distance estimators, *The Annals of Statistics*, 16, 587–608.
- Dudley R.M. (1989) *Real Analysis and Probability*, Wadsworth, Pacific Grove.
- Gini C. (1915) Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche, *Atti del R. Ist. Veneto di Sc. Let. ed Arti*, 74 (P II), 185–213.
- Landenna G. (1956) La dissomiglianza, *Statistica*, 16, 21–57.

- Rachev S.T. (1991) *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester.
- Rachev S.T. and Rüschendorf L. (1998) *Mass Transportation Problems. Voll. I and II*, Springer-Verlag, New York.
- Salvemini T. (1939) Sugli indici di omofilia, *Atti I Riunione Scientifica S.I.S.*.
- van der Vaart A.W. and Wellner J.A. (1996) *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.
- Villani C. (2003) *Topics in Optimal Transportation*, American Mathematical Society, Providence.