ASYMPTOTIC PROPERTIES AND ROBUSTNESS OF MINIMUM DISSIMILARITY ESTIMATORS OF LOCATION-SCALE PARAMETERS

F. BASSETTI† AND E. REGAZZINI†

Abstract. This paper deals with the asymptotic properties of an unexplored estimation method, for location and scale parameters, based on the minimization of the Monge–Gini–Kantorovich–Wasserstein distance. This method is rigorously defined and justified according to the general principle which directs the theory of regression. The resulting estimators—called minimum dissimilarity estimators—exist, and are measurable, consistent, and robust. Their asymptotic distribution is the same as the probability distribution of the absolute minimum point of an interesting functional of a standard Brownian bridge. This fact can be employed to obtain both explicit exact expressions and numerical approximations for the above asymptotic distribution.

Key words. argmax argument, asymptotic laws, influence function, minimum dissimilarity estimator, Monge–Gini–Kantorovich–Wasserstein metric, occupation time of a Brownian bridge, robustness, minimum dissimilarity estimators

DOI. 10.1137/S0040585X97981664

1. Introduction. Consider the location-scale model

\[ \tilde{x}_k = \mu + \sigma \tilde{\varepsilon}_k \quad (k = 1, 2, \ldots), \]

where \( \tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots \) are independent and identically distributed (i.i.d.) real-valued random variables with common nondegenerate probability distribution (p.d.) function \( F \), and \( (\mu, \sigma) = \theta \) is an unknown parameter in \( \Theta := \mathbb{R} \times [0, +\infty) \). The issue of estimating \( \theta \) represents one of the most elementary and classical problems in statistics, and there is a large body of literature about it. The reason why we go back to it is the wish to illustrate an estimation method based on the minimization of the Monge–Gini–Kantorovich–Wasserstein distance (G-distance for brevity) between the law of each \( \tilde{x}_i \) and the empirical distribution of any sample \( \tilde{x}^{(n)} = (\tilde{x}_1, \ldots, \tilde{x}_n) \). In fact, this distance has been substantially neglected by the innumerable studies which deal with minimum distance estimators, in spite of its usefulness in many fields of mathematics. See, for example, [21] for a comprehensive treatment, and [24] for applications in the theory of optimal transportation. As we will show, the method above, applied to the problem of estimating \( \theta \) in (1), appears to be robust and, consequently, can be recommended, for example, against the occurrence of anomalous values in \( \tilde{x}^{(n)} \). Moreover, it admits a theoretically interesting interpretation but, with respect to other minimum distance procedures, it turns out to be disadvantageous from a computational standpoint. In any case, we think it is worth analyzing, at least in the perspective of asymptotic statistics.

Section 2 contains the main definitions together with an interpretation to motivate the use of the method of estimation based on the minimization of the G-distance. Section 3 deals with existence, measurability, and consistency of estimators derived...
from such a process of minimization. A characterization of their limit distribution is obtained in section 4. Such a characterization is used, in section 5, to deduce the exact expression of the asymptotic distribution of location minimum $G$-distance estimators for some specific forms of $F$. Apropos of this, open problems about occupation times of a Brownian bridge are described. The same section includes the proof of a simple, interesting robustness property of minimum $G$-distance estimators of location parameters. Finally, in section 6 the characterization given in section 4 is used to derive confidence bounds for $\theta$ by Monte Carlo methods.

2. Main definitions and interpretive aspects. Let us formalize as follows the conditions mentioned at the beginning of the previous section:

$$(\text{H}_0) \ (\bar{\varepsilon}_1, \bar{x}_1), \ldots, (\bar{\varepsilon}_n, \bar{x}_n), \ldots$$ are real random vectors obeying model (1) for some $\theta := (\mu, \sigma)$ in $\Theta = \mathbb{R} \times [0, +\infty)$, defined on the probability space $(\Omega, \mathcal{F}, P_\theta)$; $P_\theta$ makes the $\bar{\varepsilon}_k$’s i.i.d. real-valued random variables with a common nondegenerate p.d. function $F$, and $F$ has finite expectation.

Thus, the $\bar{\varepsilon}_k$’s turn out to be i.i.d. random variables with a common p.d. function $F_\theta$ that, for every $x$ in $\mathbb{R}$, is defined by $F_\theta(x) := F(\sigma^{-1}(x - \mu))$ if $\sigma > 0$ and by $F_\theta(x) := H(x - \mu) = I_{[\mu, +\infty)}(x)$ if $\sigma = 0$. From now on, $I_A$ will indicate the indicator function of the set $A$. Define $F_n$ to be the empirical p.d. function of $\bar{x}^{(n)}$, i.e.,

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} H(x - \bar{x}_k) \quad (x \in \mathbb{R}).$$

Put

$$G(F_n, F_\theta) := \int_{\mathbb{R}} |F_n(x) - F_\theta(x)| \, dx,$$

that is, the $G$-distance between $F_n$ and $F_\theta$. Under these conditions, any $\hat{\theta}_n = \hat{\theta}_n(\bar{x}_1, \bar{x}_2, \ldots)$ in $\Theta$ such that

$$G(F_n, F_{\hat{\theta}_n}) = \inf_{\theta \in \Theta} G(F_n, F_\theta),$$

i.e., $\hat{\theta}_n \in \arg \min G(F_n, F_\theta)$, is said to be a minimum dissimilarity estimator (MDE) for $\theta$. It should be recalled that the term dissimilarity index was introduced by Gini to designate $G$; see [13]. More precisely, if $F_1$ and $F_2$ are p.d. functions with finite expectation, the dissimilarity index between them is defined to be

$$G(F_1, F_2) = \int_{\mathbb{R}} |F_1(x) - F_2(x)| \, dx = \int_{(0,1)} |F_1^{-1}(y) - F_2^{-1}(y)| \, dy,$$

where $F_1^{-1}$ and $F_2^{-1}$ stand for the quantile functions of $F_1$ and $F_2$, respectively, i.e.,

$$F_i^{-1}(y) := \sup \{ t \in \mathbb{R} : F_i(t) \geq y \} \quad (y \in (0,1), \ i = 1, 2).$$

To motivate the introduction of the above method of estimation, assume momentarily that $\varepsilon_k$ and $x_k$ $(k = 1, \ldots, n)$ are observed values of an explanatory variable and of a response variable, respectively. Then, like in the classical theory of regression, consider estimates of $\theta$ as minimizers of the loss function

$$\frac{1}{n} \sum_{k=1}^{n} |x_k - \mu - \sigma \varepsilon_k|.$$


with \( r \) being some fixed positive number. With \( r = 2 \), this criterion gives rise to the least squares estimates, whereas, if \( r = 1 \), it yields the Boscovich \( L^1 \)-approximation. For the case \( r = 1 \), by which our MDEs are inspired, see the relatively recent paper [19] and the book [5]. See also the papers contained in [11].

Now, if \( H^{(n)} \) is the empirical p.d. function of the sample \( ((\varepsilon_1, x_1), \ldots, (\varepsilon_n, x_n)) \), the previous loss function can be written as

\[
\int_{\mathbb{R}^2} |x - \mu - \varepsilon \sigma|^r dH^{(n)}(\varepsilon, x).
\]

Under the usual conditions of independence and identity in distribution (with parameter \( \theta = (\mu, \sigma) \)) for the \((\xi_k, \bar{x}_k)\)'s, \( H^{(n)} \) converges weakly (almost surely) as \( n \to +\infty \) to the bivariate p.d. supported by the straight line \( \{(x, y) \in \mathbb{R}^2: y = \mu + \sigma x \} \), with marginal p.d. functions \( F \) and \( F_\theta \). So, if \( \Gamma(F_1, F_2) \) denotes the Fréchet class of all the bivariate p.d. functions with fixed marginals \( F_1 \) and \( F_2 \), we can say that the limit of \( H^{(n)} \) is the maximal element of \( \Gamma(F, F_\theta) \), i.e.,

\[
\overline{H}(x, y) = \min \left\{ F(x, F_\theta(y)) \right\} \quad ((x, y) \in \mathbb{R}^2),
\]

with respect to the so-called concordance (or positive dependence) ordering defined by Gini as follows: If \( H_1 \) and \( H_2 \) are in \( \Gamma(F_1, F_2) \) and \( H_1(x, y) \geq H_2(x, y) \) for every \((x, y)\) in \( \mathbb{R}^2 \), with \( H_1(x_0, y_0) > H_2(x_0, y_0) \) for some \((x_0, y_0)\), then \( H_1 \) is said to be more concordant than \( H_2 \). Thus, in the statistical theory of regression the loss function (2) is determined by an empirical distribution which approximates \( \overline{H} \) for sufficiently large values of \( n \). To preserve this feature even when—as in our primary problem—only the response variable is observable, we can substitute \( H^{(n)} \) with

\[
H_{\ast}^{(n)}(x, y) := \min \left\{ F(x, F_n(y)) \right\} \quad ((x, y) \in \mathbb{R}^2),
\]

which, like \( H^{(n)} \), under the distributional conditions fixed at the beginning of the paper, converges weakly (almost surely) to \( \overline{H} \). Hence, if \( \int_{\mathbb{R}} |x| dF(x) < +\infty \), one will seek estimates of \( \theta \) which minimize

\[
\int_{\mathbb{R}^2} |x - \mu - \varepsilon \sigma|^r dH_{\ast}^{(n)}(\varepsilon, x) = \int_{\mathbb{R}} |F_n^{-1}(y) - F_\theta^{-1}(y)|^r dy,
\]

i.e., MDEs when \( r = 1 \). A proof of equality (3) can be found in [21].

As for literature, we know of a few papers devoted to the estimates described above: paper [4], which deals with estimates defined as minimizers of (3) with \( r = 2 \), and two papers [3] and [2], which discuss consistency of minimizers of the so-called Kantorovich functionals (see [21]) which appear as a natural generalization of (3). In fact, section 3 is a new version of some of the results contained in [2] adapted to our present particular situation.

3. Existence, measurability, and consistency of MDEs. Before we study properties like consistency, robustness, etc., it is worth tackling the problem of the existence of MDEs. Clearly, it is desirable that MDEs, when they exist, be measurable functions in order to state, for example, meaningful forms of asymptotic consistency. According to the next proposition, MDEs exist, are measurable, and are strongly consistent.

Proposition 3.1. Let \((H_0)\) be valid. Then, for each \( \theta \) in \( \Theta^\circ \) (the interior set of \( \Theta \)) and for every \( n \), there is an MDE \( \hat{\theta}_n \) which is \( \mathcal{F}/\mathcal{B}(\mathbb{R}) \)-measurable, and the
sequence \( (\hat{\theta}_n)_{n \geq 1} \) satisfies
\[
\lim_{n_0 \to +\infty} P_0 \left\{ \sup_{n \geq n_0} \| \hat{\theta}_n - \theta \|_2 \leq \varepsilon \right\} = 1 \quad (\varepsilon > 0),
\]
where \( \| \cdot \|_2 \) stands for the Euclidean norm in \( \mathbb{R}^2 \).

Proof. Apply the dominated convergence theorem to prove that \( \Theta \ni \theta \mapsto G(F_\theta, F_\theta') \) is continuous. Hence, for any \( c \), the set \( \{ \theta' \in \Theta : G(F_n, F_{\theta'}) \leq c \} \subset \mathbb{R}^2 \) is closed. Moreover, it is bounded since
\[
G(F_\theta, F_{\theta'}) = \rho \int_{(0,1)} \left| \sin \phi + F^{-1}(y) \cos \phi \right| dy
\]
with \( \rho = \| \theta' - \theta \|_2, \mu' - \mu = \rho \sin \phi, \sigma' - \sigma = \rho \cos \phi, \) and
\[
k := \inf_{\phi} \int_{(0,1)} \left| \sin \phi + F^{-1}(y) \cos \phi \right| dy > 0;
\]
thus, if \( G(F_\theta, F_{\theta'}) \leq c \), we can write \( \| \theta' - \theta \|_2 \leq c/k \), and from \( G(F_{\theta'}, F_\theta) \leq c + s_n \) (where \( s_n := G(F_n, F_\theta) \), which holds when \( G(F_n, F_{\theta'}) \leq c \), we get
\[
\{ \theta' \in \Theta : G(F_n, F_{\theta'}) \leq c \} \subset \{ \theta' \in \Theta : \| \theta' - \theta \|_2 \leq (c + s_n) k^{-1} \}.
\]
The fact that the sets \( \{ \theta' \in \Theta : G(F_n, F_{\theta'}) \leq c \} \) are compact entails
\[
\arg \min_{\theta'} G(F_n, F_{\theta'}) \neq \emptyset,
\]
which is tantamount to saying that MDEs \( \hat{\theta}_n \) exist for every \( n \). At this stage, the measurability of \( \hat{\theta}_n \) is a straightforward consequence of Corollary 1 in [7].

As for the consistency of \( (\hat{\theta}_n)_{n \geq 1} \), consider the triangle inequality again to obtain
\[
G(F_{\hat{\theta}_n}, F_\theta) \leq 2s_n \quad \text{and, hence,} \quad \| \hat{\theta}_n - \theta \|_2 \leq 2s_n/k.
\]
Then, apply the Glivenko–Cantelli theorem to get \( \sup |F_n - F_\theta| \to 0 \) (almost surely) and, by the Kolmogorov strong law of large numbers, \( \int x dF_n(x) \to \int x dF_\theta(x) \) (almost surely). These facts yield \( s_n \to 0 \) (almost surely); see, for example, [21, Corollary 7.5.3].

4. Characterization of the asymptotic distribution of MDEs. The main result of the present section characterizes the limiting law of the statistic
\[
\hat{\theta}_n := \sqrt{n}(\hat{\theta}_n - \theta),
\]
which is well-defined according to Proposition 3.1.

**Proposition 4.1.** Let \((H_0)\) be in force together with the following assumptions:
(i) \( F \) is absolutely continuous on \( \mathbb{R} \), with a bounded and continuous probability density \( f \), and the topological support of the probability measure determined by \( F \) is an interval, whose interior set will be denoted by \((a, b)\) with \(-\infty \leq a < b \leq +\infty\);
(ii) \( \int_{\mathbb{R}} \sqrt{f(x)(1 - f(x))} \) \(dx < +\infty \).

Then, for every \( \theta \) in \( \Theta^0 \), there is a suitable Brownian bridge \( B \) on \((\Omega, \mathcal{F}, P_\theta)\) such that
\[
M(h) := \int_{(0,1)} \left| \frac{\sigma B(y)}{f(F^{-1}(y))} - h_1 - h_2 F^{-1}(y) \right| dy
\]
\[
= \int_{(a,b)} \left| \frac{\sigma B(F(x))}{f(x)} - h_1 - h_2 x \right| f(x) \) \(dx \)
is, for every \( h = (h_1, h_2) \) in \( \mathbb{R}^2 \), a real-valued random variable and \( \{M(h) : h \in \mathbb{R}^2\} \) is a stochastic process on \((\Omega, \mathcal{F}, P_\theta)\) with continuous and coercive paths on \( \mathbb{R}^2 \), which
possess one and only one absolute minimum point \( \hat{h} \). It is just such a random minimum that represents the limit, in law, of \( (\hat{h}_n)_{n \geq 1} \).

Proof. The proof relies on an argmin argument which mimics Theorem 3.2.2 in [8]. Accordingly, consider

\[
M_n(\theta) := G(F_\theta, F_n), \quad M_n(h) := \sqrt{n}M_n(\theta + \frac{h}{\sqrt{n}})
\]

and note that, for any compact subset \( K \) of \( \mathbb{R}^2 \), there is \( \pi \) such that \( M_n(h) \) is well-defined for every \( h \) in \( K \) and for every \( n \geq \pi \). It remains to prove the following claims:

(a) Each path of \( M \) has one and only one absolute minimum point.
(b) For every compact \( K \subset \mathbb{R}^2 \), \( (M_n)_{n \geq 1} \) converges in distribution to \( M \) in the space \( L^\infty(K) \) of bounded functions from \( K \) into \( \mathbb{R} \).
(c) The sequence \( (\hat{h}_n)_{n \geq 1} \) is tight in \( (\mathbb{R}^2, \| \cdot \|_2) \).

Proof of claim (a). Consider the probability space \( ((a,b), \mathcal{B}(a,b), p_F) \), where \( p_F \) denotes the probability measure generated by \( F \). Such a space is nonatomic and the conjugate space of \( L^1((a,b), \mathcal{B}(a,b), p_F) \) is given by \( L^\infty((a,b), \mathcal{B}(a,b), p_F) \). Then, in view of Theorem 24 in [8], it suffices to prove that 0 is the only element of \( \mathcal{A} := \{ a: \mathbb{R} \to \mathbb{R} : a(x) = ax + \beta, (a,\beta) \in \mathbb{R}^2 \} \) which vanishes on an \( \varepsilon \)-set. Recall that \( B \subset (a,b) \) is termed an \( \varepsilon \)-set if \( B = \partial A \) and \( A \) is a measurable set satisfying \( \int_A a dF = \int_A a dF \) for every \( a \in \mathcal{A} \), i.e.,

\[
\int_A x dF(x) = \frac{1}{2} \quad \text{and} \quad \int_A x dF(x) = \int_{(a,b)} x dF(x).
\]

At this stage, it is enough to prove that \( |\partial A| \geq 2 \). In point of fact, \( |\partial A| \) is smaller than 2 if and only if \( A = (a,c) \) or \( A = (c,b) \) for some \( c \) in \( (a,b) \) and, therefore, either \( \int_A x dF(x) > \int_{(a,b)} x dF(x) \) or \( \int_A x dF(x) < \int_{(a,b)} x dF(x) \), in contradiction with (4).

Proof of claim (b). For every compact \( K \subset \mathbb{R}^2 \) there is \( \pi \) such that \( M_n(h) \) is well-defined for every \( h \) in \( K \) and for every \( n \geq \pi \). Thanks to (i), one can write

\[
F_{\theta + h/\sqrt{n}}(x) = F_\theta(x) + g(x) \frac{h}{\sqrt{n}} + \frac{\Delta_n(x,h)}{\sqrt{n}}
\]

with

\[
g(x) = -f\left(\frac{x - \mu}{\sigma}\right)\left(\frac{1}{\sigma}, \frac{x - \mu}{\sigma^2}\right)
\]

for every \( x \) in \( \mathbb{R} \), and

\[
\lim_{n \to +\infty} \sup_{\|h\|_2 \leq C} \int_{\mathbb{R}} |\Delta_n(x,h)| \, dx = 0
\]

for every \( C > 0 \). To prove (5), first observe that \( F_\theta \) is differentiable with respect to \( \theta \) and that the differential is given by

\[
dF\left(\frac{x - \mu}{\sigma}\right) = g(x) \, (d\mu, d\sigma).
\]
Thus,
\[
\int_{\mathbb{R}} |\Delta_n(x, h)| \, dx = \int_{\mathbb{R}} \sqrt{n} \left| F_{\theta + h/\sqrt{n}}(x) - F_\theta(x) - g(x) \left( \frac{h_1 + h_2}{\sqrt{n}} \right) \right| \, dx
\]

\[
= \int_{\mathbb{R}} \sqrt{n} \int_{(x-\mu)/\sigma}^{(x-\mu)/\sigma + \sqrt{n}} f(t) \, dt
\]

\[
+ \frac{1}{\sigma} f \left( \frac{x-\mu}{\sigma} \right) \left( h_1 + h_2 \frac{x-\mu}{\sigma} \right) \, dx
\]

\[
\leq \sqrt{n} \int_{\mathbb{R}} \int_{m_{n,t}}^{M_{n,t}} \left| \sigma f(t) - f(\xi) \left( \sigma + \frac{h_2}{\sqrt{n}} \right) \right| \, d\xi \, dt,
\]

with \(m_{n,t}\) and \(M_{n,t}\) representing \(t \wedge \sigma^{-1}\{t(\sigma + h_2/\sqrt{n}) + h_1/\sqrt{n}\}\) and \(t \vee \sigma^{-1}\{t(\sigma + h_2/\sqrt{n}) + h_1/\sqrt{n}\}\), respectively. Then, for any \(M > 0\),
\[
\int_{\mathbb{R}} |\Delta_n(x, h)| \, dx \leq \left( \int_{|t| > M} + \int_{|t| \leq M} \right) \frac{|h_1 + h_2t|}{\sigma} \]

\[
\times \int_{0}^{1} \left| \sigma f(t) - \left( \sigma + \frac{h_2}{\sqrt{n}} \right) f \left( t + x \left| \frac{h_1 + h_2t}{\sigma \sqrt{n}} \right| \right) \right| \, dx \, dt
\]

and, for every \(\varepsilon > 0\) there is \(\overline{M} = \overline{M}(\varepsilon)\) in \((a, b)\) and \(\overline{\pi} = \overline{\pi}(\varepsilon)\) such that, in view of (ii), the following inequalities hold for every \(M\) in \((a, b)\) greater than \(\overline{M}(\varepsilon)\) and \(n \geq \overline{\pi}:
\[
\sup_{\|h\| \leq C} \int_{|t| > M} \frac{|h_1 + h_2t|}{\sigma} \int_{0}^{1} \left| \sigma f(t) - \left( \sigma + \frac{h_2}{\sqrt{n}} \right) f \left( t + x \left| \frac{h_1 + h_2t}{\sigma \sqrt{n}} \right| \right) \right| \, dx \, dt
\]

\[
\leq \int_{|t| > M} C|1 + t| f(t) \, dt
\]

\[
+ \sup_{\|h\| \leq C} \int_{0}^{1} \int_{|t| > M} \left| \left( \sigma + \frac{h_2}{\sqrt{n}} \right) f \left( t + x \left| \frac{h_1 + h_2t}{\sigma \sqrt{n}} \right| \right) \right| \, dt \, dx
\]

\[
\leq \int_{|t| > M} C|1 + t| f(t) \, dt
\]

\[
+ \mathcal{C} \int_{0}^{1} \left( \int_{-M+C/\sqrt{n}}^{M} + \int_{-\infty}^{-M+C/\sqrt{n}} \right) [1 + |\xi|] f(\xi) \, d\xi \, dx \leq \varepsilon.
\]

Now, fix \(M\) as indicated above and verify—through dominated convergence—that
\[
\sup_{\|h\| \leq C} \int_{0}^{1} \left| \sigma f(t) - \left( \sigma + \frac{h_2}{\sqrt{n}} \right) f \left( t + x \left| \frac{h_1 + th_2}{\sigma \sqrt{n}} \right| \right) \right| \, dx
\]

converges to 0 as \(n \to +\infty\) for every \(t\), and that
\[
|h_1 + h_2t| \int_{0}^{1} \left| \sigma f(t) - \left( \sigma + \frac{h_2}{\sqrt{n}} \right) f \left( t + x \left| \frac{h_1 + th_2}{\sigma \sqrt{n}} \right| \right) \right| \, dx
\]

is bounded with respect to \((h, t)\) on \(\{h: \|h\| \leq C\} \times [-M, M]\). Hence, the Lebesgue-dominated convergence theorem yields \(\int_{\mathbb{R}} \sup_{\|h\| \leq C} |\Delta_n(x, h)| \, dx \to 0\) as \(n \to +\infty\).

Setting \(G_n(x) := \sqrt{n}(F_n - F_\theta)(x)\), one can rewrite \(M_n(h)\) as
\[
M_n(h) = \int_{\mathbb{R}} |G_n(x) - h g(x) - \Delta_n(x, h)| \, dx.
\]

Further, in view of Theorem 2.1 in [10], the sequence \((G_n)_{n \geq 1}\) converges in distribution, in \(L^1(\mathbb{R})\), to \(B(F_\theta)\), by virtue of (ii). Thus since \(f \mapsto \sum_{i=1}^{\nu} c_i \int_{\mathbb{R}} |f(x) -
$h^{(i)} g(x) \, dx$ is a continuous function from $L^1(\mathbb{R})$ into $\mathbb{R}$, for every $c_i$ in $\mathbb{R}$ and $h^{(i)}$ in $\mathbb{R}^2$ ($i = 1, \ldots, \nu$), we see that $\sum_{i=1}^{\nu} c_i \int_{\mathbb{R}} |G_n(x) - h^{(i)} g(x)| \, dx$ converges in distribution to $\sum_{i=1}^{\nu} c_i M(h^{(i)})$. In other words, the finite-dimensional distributions of
\[
\left\{ N_n(h) := \int_{\mathbb{R}} |G_n(x) - h \, g(x)| \, dx : h \in \mathbb{R}^2 \right\}
\]
correspond weakly to the finite-dimensional distributions of $M$. Now, since
\[
\left| \sum_{i=1}^{\nu} c_i \{ M_n(h^{(i)}) - N_n(h^{(i)}) \} \right| \leq \sum_{i=1}^{\nu} c_i \int_{\mathbb{R}} |\Delta_n(h^{(i)}(x))| \, dx,
\]
with the right-hand side converging to 0 as $n \to +\infty$, the Slutsky theorem can be used to show that the finite-dimensional distributions of $(M_n)_{n \geq 1}$ have the same weak limits as the finite-dimensional distributions of $(N_n)_{n \geq 1}$. So, for each $h$, $(M_n(h))_{n \geq 1}$ is tight in $\mathbb{R}$. Moreover,
\[
\left| M_n(h^{(1)}) - M_n(h^{(2)}) \right| \leq \int_{\mathbb{R}} \left| (h^{(1)} - h^{(2)}) g(x) \right| \, dx
\]
\[+ \int_{\mathbb{R}} \left\{ |\Delta_n(h^{(1)}(x))| + |\Delta_n(h^{(2)}(x))| \right\} \, dx.
\]
Then, since every compact set $K \subset \mathbb{R}^2$ has a finite measurable partition—say, $\{T_1, \ldots, T_k\}$—such that $\|h^{(1)} - h^{(2)}\|_2 \leq \varepsilon$ for every $h^{(1)}$ and $h^{(2)}$ in $T_i$, whichever $i$ may be, we can write
\[
\sup_{i \in \{h^{(1)}, h^{(2)}\} \in T_i} \sup \left| M_n(h^{(1)}) - M_n(h^{(2)}) \right| \leq \varepsilon \int_{\mathbb{R}} \|g\|_2 \, dx + 2 \sup_{i \in T_i} \int_{\mathbb{R}} |\Delta_n(x, h)| \, dx
\]
with $\sup_i \sup_{h \in T_i} \int_{\mathbb{R}} |\Delta_n(x, h)| \, dx \to 0$ as $n \to +\infty$ thanks to (5). This fact, through Theorem 1.5.6 in [23], tells us that $(M_n)_{n \geq 1}$ is tight in $l^\infty(K)$ for every compact set $K \subset \mathbb{R}^2$. Then, since weak convergence in $l^\infty(K)$ can be characterized as tightness plus convergence of marginals, we can conclude that $(M_n)_{n \geq 1}$ converges in law to $M$ in $l^\infty(K)$ for every compact subset $K$ of $\mathbb{R}^2$.

Proof of claim (c). For any $\theta'$ in $\Theta$, we can write
\[
G(F_{\theta'}, F_{\theta'}) - G(F_n, F_{\theta}) \leq M_n(\theta') \leq G(F_n, F_{\theta}) + G(F_{\theta'}, F_{\theta'})
\]
and, therefore,
\[
G(F_{\theta_n}, F_{\theta}) \leq G(F_n, F_{\theta}) + M_n(\hat{\theta}_n) \leq G(F_n, F_{\theta}) + M_n(\theta) \leq 2G(F_n, F_{\theta}).
\]
Since $G(F_{\theta}, F_{\theta'}) \leq c$ implies that $\|\theta' - \theta\|_2 \leq c/k$ holds true for some suitable $k$, as shown in the proof of Proposition 3.1, then
\[
\left\| \sqrt{n} (\hat{\theta}_n - \theta) \right\|_2 \leq 2 \sqrt{n} G(F_n, F_{\theta}) k^{-1},
\]
i.e.,
\[
P_{\theta} \left\{ \left\| \hat{\theta}_n \right\|_2 \leq \lambda \right\} \geq P_{\theta} \left\{ \sqrt{n} G(F_n, F_{\theta}) \leq \frac{\lambda k}{2} \right\} \quad (\lambda > 0).
\]
Thus, the tightness of $\hat{\theta}_n_{n \geq 1}$ is a consequence of the tightness of $\sqrt{n} G(F_n, F_{\theta})$ which, in turn, follows from Theorem 1.1 in [10] about the convergence in law of $G(F_n, F_{\theta})$. Proposition 4.1 is proved.
The above characterization of the limiting distribution of MDEs will be employed in section 6 to determine—via Monte Carlo methods—confidence bounds for location, scale, and location-scale parameters of some distinguished models. In the next section, the same characterization is discussed in connection with the exact form of the limiting distribution of an MDE for a location parameter.

5. MDEs for location parameters: Limiting distribution and robustness.

As to the problem of determining an adequate explicit representation for the probability distribution of \( \hat{h} \) when assumptions (i) and (ii) of Proposition 4.1 are valid, we confine ourselves to considering the case of a single location parameter, or of a single scale parameter. With respect to these particular cases, the present treatment is rather incomplete but, as explained later on, we intend to return to this subject in a forthcoming paper.

Assume that the scale parameter \( \sigma \) in (1) is known (e.g., \( \sigma = \sigma_0 > 0 \)). Then, \( \hat{h} \) is a real-valued random variable satisfying

\[
\hat{h} = \arg \min_{h \in \mathbb{R}} \int_{(0,1)} \left| \frac{\sigma_0 B(y)}{f(F^{-1}(y))} - h \right| dy.
\]

Hence, for each path \( B \) of \( B \), \( \hat{h} \) can be viewed as a median of the random variable

\[
y \mapsto \sigma_0 B(y)
\]

defined on the probability space \( (0,1, B(0,1), \lambda) \), with \( \lambda \) being the uniform probability measure on \( (0,1) \). Therefore, for every \( \theta = (\mu, \sigma_0) \) we have

\[
P_\theta \{ \hat{h} > \xi \} = P_\theta \left\{ \int_{(0,1)} I_{(-\infty,\gamma(s)]}(B(s)) \, ds < \frac{1}{2} \right\}
\]

with

\[
\gamma(\cdot) := \frac{\xi f(F^{-1}(\cdot))}{\sigma_0}.
\]

Thus, the problem of determining the limiting law of an MDE for a location parameter can be viewed as that of seeking the left limit at \( \frac{1}{2} \) of the probability distribution of the occupation time of \( B \) determined by (8).

Analogously, for MDEs of scale parameters—when \( \mu = \mu_0 \) is known—\( \hat{h} \) must satisfy

\[
\hat{h} = \arg \min_{h \in \mathbb{R}} \int_{(0,1)} \left| \frac{\sigma B(y)}{f(F^{-1}(y))} - h F^{-1}(y) \right| dy
\]

and, consequently, for each trajectory \( B \) of \( B \), it coincides with the median of the random variable defined on \( (0,1, B(0,1), \lambda^*) \) by

\[
y \mapsto \frac{\sigma B(y)}{f(F^{-1}(y)) F^{-1}(y)},
\]

where \( \lambda^* \) stands for the probability measure satisfying

\[
\lambda^*(dy) = \frac{1}{\overline{\mu}} |F^{-1}(y)| \, dy
\]

on \( (0,1) \), with \( \overline{\mu} := \int_0^1 |F^{-1}(y)| \, dy \). Hence,

\[
P_\theta \{ \hat{h} > \xi \} = P_\theta \left\{ \int_{(0,1)} I_{(-\infty,\gamma^*(s)]}(B(s)) |F^{-1}(s)| \, ds < \frac{\overline{\mu}}{2} \right\}
\]
with

\[ \gamma^*(t) := \frac{\xi f(F^{-1}(t))}{\sigma} \]  

Now, by recalling that \( B \) and \(-B\) have the same probability distribution, the previous statements can be used to easily prove the following proposition.

**Proposition 5.1.** Let assumption \((H_0)\) and assumptions (i) and (ii) of Proposition 4.1 be in force and let \( \hat{\mu}_n, \hat{\sigma}_n \) denote MDEs of location and scale parameters, respectively. Then the limiting probability distributions of \( \sqrt{n}(\hat{\mu}_n - \mu) \) and \( \sqrt{n}(\hat{\sigma}_n - \sigma) \) are symmetric.

Distributional results for Brownian occupation times have been discussed in recent papers and books; see [12], [22], [14], [6]. The latter reference contains useful formulae that can be used to determine (7) for the following models:

\[
\begin{cases}
F_\theta: \theta = (\mu, \sigma_0), \quad F_\theta(x) = F(x - \mu) = \int_{(-\infty, x]} \frac{1}{\sigma_0} e^{-(s-\mu)/\sigma_0} I_{(\mu, +\infty)}(s) \, ds \\
(x \in \mathbb{R}), \text{ for every } \mu \text{ in } \mathbb{R}
\end{cases}
\]  

\[ \int_{(0,1)} I_{(-\infty, \gamma(s))}(B(s)) \, ds \]

Indeed, in these cases, the Laplace–Stieltjes transform of the probability distribution of \( \int_{(0,1)} I_{(-\infty, \gamma(s))}(B(s)) \, ds \) can be found in the tables of [6]. For a correct use of these tables, it should be recalled that a Brownian bridge can be identified as a Brownian motion conditioned to take the value 0 at 1. Hence, setting

\[
L(z; \xi) := \lim_{e \to 0^+} \frac{1}{P_\theta(W_{(0,0)}^{(0,0)} \in (-e, e))} \mathcal{F}_z(0; 0; z, \xi).
\]

**Example 5.1.** Consider the problem of determining (15) when \( F_\theta \) is defined by (12). In this case, we have \( \gamma(s) = \xi(1 - s)/\sigma_0 \), and (14), through [6, Part II, section 2.1.5.8,] leads to

\[
L(z; \xi) = \frac{\xi}{\sqrt{2\pi}} e^{\xi^2/2} \int_0^1 \frac{e^{-z s} - e^{-z(1-s)}}{z s^{3/2}} \frac{e^{-z(1-s)}}{(1-s)^{3/2}} e^{-\xi^2/(2(1-s))} \, ds.
\]

\[
\hat{\xi} := \frac{\xi}{\sigma_0}.
\]
where \( L(t, z) \) is the value at \( z \) of the Laplace transform of \( x \mapsto I_{(t, 1)}(x) \). At this point, it is easy to see that the corresponding inverse Laplace transform—i.e., the density of the absolutely continuous component of the law of the occupation time—is

\[
x \mapsto \frac{\xi e^{\xi^2/2}}{\sqrt{2\pi}} \int_0^x \frac{e^{-\xi^2/2t}}{t^{3/2}(1-t)^{3/2}} \, dt \quad (0 < x < 1).
\]

Since the value of the integral on \((0, 1)\) of this function is 1, it then represents the probability density function of the occupation time, and we can consider (7) to determine the limiting distribution of the centered and normalized sequence \( \sqrt{n}(\hat{\theta}_n - \theta) \) of MDEs, i.e.,

\[
P_\theta \{ \hat{\theta} \leq \xi \} = 1 - \frac{\xi e^{\xi^2/2}}{\sqrt{2\pi}} \int_0^{1/2} \left( \int_0^x e^{-\xi^2/2t} t^{-3/2}(1-t)^{3/2} \, dt \right) \, dx \quad (\xi > 0)
\]

\[
= 1 - \frac{\xi e^{\xi^2/2}}{2\sqrt{2\pi}} \int_0^{+\infty} \frac{y e^{-y^2/2}}{(1+y)^{3/2}} \, dy
\]

\[
= 1 - \frac{\xi e^{\xi^2/2\sigma_0^2}}{2\sigma_0\sqrt{2\pi}} U \left( 2, \frac{3}{2}, \frac{\xi^2}{2\sigma_0^2} \right) = 1 - \frac{e^{-\xi^2/2\sigma_0^2}}{2\sqrt{\pi}} U \left( 3 \frac{1}{2}, \frac{1}{2}, \frac{\xi^2}{2\sigma_0^2} \right),
\]

where, as usual, \( U \) stands for the confluent hypergeometric function of the second kind (also known as the Tricomi function). See, for example, [18]. Combining this fact with Proposition 5.1, we can conclude with the following proposition.

Let \( \theta_n \) be the MDE of \( \mu \) in the exponential model (12). Then the limiting p.d. function (as \( n \to +\infty \)) of \( \sqrt{n}(\hat{\theta}_n - \theta) \) is given by the absolutely continuous function

\[
\xi \mapsto I_{(-\infty, 0)}(\xi) \frac{e^{-\xi^2/2\sigma_0^2}}{2\sqrt{\pi}} U \left( 3 \frac{1}{2}, \frac{1}{2}, \frac{\xi^2}{2\sigma_0^2} \right) + I_{(0, +\infty)}(\xi) \left\{ 1 - \frac{e^{-\xi^2/2\sigma_0^2}}{2\sqrt{\pi}} U \left( 3 \frac{1}{2}, \frac{1}{2}, \frac{\xi^2}{2\sigma_0^2} \right) \right\}.
\]

Furthermore, such a p.d. has probability density function

\[
\xi \mapsto \frac{2^{3/2}}{\pi \sigma_0^3} |\xi| \int_0^1 \frac{\sqrt{x}}{(1+x)\sqrt{(1-x)^3}} \exp \left\{ -\frac{1+x}{2\sigma_0^2} \xi^2 \right\} \, dx,
\]

and its moment of order \( 2k \) is

\[
\begin{cases} 
\frac{\sigma_0^{2k}}{k+1} & \text{for } k = 1, 2, \ldots, \\
1 & \text{for } k = 0.
\end{cases}
\]

**Example 5.2.** In order to determine (15) for the uniform model (13), first observe that in such a case one has \( \gamma(s) = \xi := \xi/\sigma_0 \). Thus, by starting from 1.1.5.8 in [6, Part III], the same procedure as in Example 5.1 can be applied to determine (7). One gets

\[
L(z, \xi) = e^{-z}(1 - e^{-2z^2}) + \sqrt{\frac{2}{\pi \xi z}} \int_0^1 t^{-3/2}(1 - e^{-zt}) e^{-z(1-t)-2z^2/(1-t)} \, dt,
\]

i.e., the Laplace–Stieltjes transform of the p.d.

\[
(1 - e^{-2z^2}) \delta_1(A) + \xi \sqrt{\frac{2}{\pi}} \int_A \left( \int_0^x \frac{e^{-2z^2/\xi}}{y(y-1)} \, dy \right)^{3/2} \, dx \quad (A \in \mathcal{B}(\mathbb{R}))
\]
of the occupation time of a Brownian bridge determined by $\gamma(s) \equiv \xi$. In other words, the following statement holds true.

Let $\hat{\theta}_n$ be the MDE of $\mu$ in the uniform model (13). Then the absolutely continuous function

$$
\xi \mapsto I_{(-\infty,0]}(\xi) \frac{1}{2\sqrt{\pi}} e^{-4\xi^2/\sigma_0^2} U\left(\frac{3}{2}, 1; \frac{1}{2}; \frac{2\xi^2}{\sigma_0^2}\right)
$$

$$
+ I_{(0,\infty)}(\xi) \left\{ 1 - \frac{1}{2\sqrt{\pi}} e^{-4\xi^2/\sigma_0^2} U\left(\frac{3}{2}, 1; \frac{1}{2}; \frac{2\xi^2}{\sigma_0^2}\right) \right\}
$$

is the limiting p.d. function (as $n$ goes to $+\infty$) of $\sqrt{n}(\hat{\theta}_n - \theta)$, under $P_0$.

Moreover, the moment of order $2k$ of this p.d. is

$$
\begin{cases}
\sigma_0^2 \frac{1 \cdot 3 \cdots (2k-1)}{8^k(k+1)} {}_2F_1\left(\frac{3}{2}, k + 2; \frac{1}{2}\right) & (k = 1, 2, \ldots), \\
1 & (k = 0),
\end{cases}
$$

where, as usual, ${}_2F_1$ denotes the Gauss hypergeometric function.

In [6], the transformation (14) with $\gamma(s) = \xi(1-s)$—see Example 5.1—is evaluated by means of the Cameron–Martin–Girsanov transformation of measure, whereas, if $\gamma(s) = \xi$—like in Example 5.2—it is determined through the Feynman–Kac formula. In fact, the direct application of such a formula to the evaluation of (14) for general forms of $\gamma$ can be problematic due to the violation of some regularity conditions required by the usual formulation of the Feynman–Kac theorem. See, for example, [17, subsection 5.6.B]. So, in order to get a systematic treatment of the p.d. of MDEs of location parameters in (1), we study a suitable extension of the Feynman–Kac theorem which yields (14) for general forms of $\gamma$, and we develop suitable analytical methods to produce explicit forms of the solutions of the corresponding differential problem. The results of this research will appear in a forthcoming paper.

The main objective of the rest of the present section is to study robustness of MDEs of location parameters. First, let us specify how we intend to measure robustness to the occurrence of anomalous observations. We represent the intrusion of “bad” observations through the p.d. function $x \mapsto H(x - \xi) =: H_\xi(x)$ which is degenerate at $\xi$. More precisely, after fixing a proportion $\varepsilon$ in $(0, 1)$ of anomalous observations, we suggest that $F_n$ be replaced with

$$
x \mapsto (1 - \varepsilon) F_n(x) + \varepsilon H_\xi(x).
$$

When the true value of the unknown location parameter is revealed, say $\mu$, since the $\tilde{x}_k$’s are supposed to be i.i.d. with common p.d. function $x \mapsto F(x - \mu)$, such a distribution can be viewed as an approximation of $F_n$, valid for large values of $n$. Thus,

$$
\mu_\varepsilon \in \arg \min_{\mu' \in \mathbb{R}} \int |F(x - \mu') - (1 - \varepsilon) F(x - \mu) - \varepsilon H(x - \xi)| \, dx
$$

is a minimum dissimilarity estimator of a location parameter, in the presence of gross errors, when the empirical distribution is assumed to be a copy of the true distribution. Under these circumstances, the quantity

$$
r = r(\xi, \varepsilon) := \sup \left\{ |\mu_\varepsilon - \mu| : \mu_\varepsilon \in \arg \min_{\mu' \in \mathbb{R}} G(F_{\mu'}, (1 - \varepsilon) F_\mu + \varepsilon H_\xi) \right\}
$$

(16)

can be interpreted as a measure of asymptotic robustness of an MDE of a location parameter. When $F$ satisfies the assumptions of Proposition 4.1, these estimators are
resistant to the intrusion of “bad” observations according to the following proposition, in which one sets,

\[ Z_\xi(y) := I_{(0, F(\xi - \mu))}(y) Z^+(y) + I_{[F(\xi - \mu), 1]}(y) Z^-(y) \quad (y \in (0, 1)), \]

\[ Z^+(y) := \frac{y}{f(F^{-1}(y))} \quad \text{and} \quad Z^-(y) := \frac{y - 1}{f(F^{-1}(y))} \quad (y \in (0, 1)), \]

and \( h_\xi \) for any element of \( \arg \min_h \int_{(0, 1)} |h - Z_\xi(y)| \, dy \) for some fixed \( \xi \) in \( \mathbb{R} \).

**Proposition 5.2.** Let the same assumptions as in Proposition 4.1 be in force with \( \sigma = 1 \). Moreover, assume there is \( \delta > 0 \) such that

\[
\sup_{y \in (0, \delta)} \frac{y}{f(F^{-1}(y))} + \sup_{y \in (1 - \delta, 1)} \frac{1 - y}{f(F^{-1}(y))}
\]

is finite. Then, for any \( \xi \neq \mu + F^{-1}(1/2) \), one has that

(a) \( \arg \min_h \int_{(0, 1)} |h - Z_\xi(y)| \, dy \) is a singleton;

(b) \( \xi \mapsto h_\xi \) is a bounded function;

(c) \( \mu_\varepsilon = \mu + h_\varepsilon + o(\varepsilon) \) (\( \varepsilon \to 0^+ \)).

It is worth noting that, in view of (c), \( \xi \mapsto h_\xi \) can be viewed as the influence function of the MDE of a location parameter.

**Proof of Proposition 5.2.** Define \( G_\varepsilon(\mu') := G(F_{\mu'}, (1 - \varepsilon) F_{\mu} + \varepsilon H_\xi) \) and set \( V_\varepsilon(h) := \frac{1}{\varepsilon} G_\varepsilon(\mu + h \varepsilon) \). Then \( \hat{h}_\varepsilon := (\mu_\varepsilon - \mu) / \varepsilon \) is an absolute minimum point of \( V_\varepsilon \). After noticing that (a) is plainly true if \( \xi \neq \mu + F^{-1}(1/2) \), we continue by proving (c), i.e., \( \mu_\varepsilon - \mu = h_\varepsilon + o(\varepsilon) \) as \( \varepsilon \to 0^+ \), which is tantamount to showing that \( \hat{h}_\varepsilon = h_\xi + o(1) \) as \( \varepsilon \to 0^+ \).

First, by the triangle inequality and the minimum property of \( \mu_\varepsilon \), one has

\[ G(F_{\mu}, F_{\mu_\varepsilon}) \leq 2G((1 - \varepsilon) F_{\mu} + \varepsilon H_\xi, F_{\mu}) = 2\varepsilon G(F_{\mu}, H_\xi) \]

and, since \( G(F_{\mu}, F_{\mu_\varepsilon}) = |\mu - \mu_\varepsilon| \), one obtains \( |\mu - \mu_\varepsilon| \varepsilon \leq 2G(F_{\mu}, H_\xi) \). Next, rewrite \( V_\varepsilon \) as

\[ V_\varepsilon(h) = \frac{1}{\varepsilon} \int_{(0, 1)} |\varepsilon h - Y_\varepsilon(y)| \, dy, \]

with \( Y_\varepsilon \) defined by

\[ Y_\varepsilon(y) = I_{(0, \alpha_\varepsilon)}(y) \left\{ F^{-1} \left( \frac{y}{1 - \varepsilon} \right) - F^{-1}(y) \right\} + I_{(\alpha_\varepsilon, \alpha_\varepsilon + \varepsilon)}(y) \{ \xi - \mu - F^{-1}(y) \} \]

\[ + I_{(\alpha_\varepsilon + \varepsilon, 1)}(y) \left\{ F^{-1} \left( \frac{y - \varepsilon}{1 - \varepsilon} \right) - F^{-1}(y) \right\} \]

with \( \alpha_\varepsilon := (1 - \varepsilon) F(\xi - \mu) \). Now, by combining the mean-value theorem and the dominated convergence theorem, it can be shown that \( V_\varepsilon(h) \) converges to \( \int_{(0, 1)} |h - Z_\xi(y)| \, dy := V_0(h) \), uniformly with respect to \( h \) on every compact interval, as \( \varepsilon \to 0^+ \). Then, in view of (a), \( \arg \min_h V_0(h) \) is a singleton and \( \hat{h}_\varepsilon \to h_\xi := \arg \min_h V_0(h) \) as \( \varepsilon \to 0^+ \).

Finally, note that the minimum point \( h_\xi \) must coincide with the median of \( Z_\xi(\cdot) \), when \( Z_\xi \) is thought of as a random variable on \((0, 1), \mathcal{B}(0, 1), \lambda)\). Hence, since

\[ Z^-(y) \leq Z_\xi(y) \leq Z^+(y) \quad (y \in (0, 1)), \]

we can write

\[ \text{median}(Z^-) \leq h_\xi \leq \text{median}(Z^+) \]

(17)
and this last inequality implies that $\xi \mapsto h_\xi$ is bounded, as asserted in (b). Proposition 5.2 is proved.

The last proposition shows that, from the point of view of robustness, an MDE of a location parameter is better than the empirical mean, at least for large values of $|\xi|$. As a matter of fact, if $\int_x x dF(x) = 0$, the influence function of the empirical mean, evaluated at $F_\mu$, has the form of the unbounded function $\xi \mapsto (\xi + \mu)$. Regarding the comparison with more robust estimators, such as the median, one can consider the special case in which $Z^-$ and $Z^+$ are strictly monotone functions, with the proviso that $Z^-$ must be increasing if $Z^+$ is decreasing. In such a case, $(Z^-,Z^+)$ is said to be an admissible pair. In fact, when $(Z^-,Z^+)$ is an admissible pair of strictly monotone functions, it is clear that

$$\text{median}(Z^-) = -\text{median}(Z^+) = -\left(2f\left(F^{-1}\left(\frac{1}{2}\right)\right)\right)^{-1}$$

holds true. Moreover, for any $\xi \neq \mu + F^{-1}\left(\frac{1}{2}\right)$,

$$-\left(2f\left(F^{-1}\left(\frac{1}{2}\right)\right)\right)^{-1} I_{(-\infty,F^{-1}(\frac{1}{2})+\mu)}(\xi) + \left(2f\left(F^{-1}\left(\frac{1}{2}\right)\right)\right)^{-1} I_{(F^{-1}(\frac{1}{2})+\mu,\infty)}(\xi)$$

is the well-known expression of the influence function of the empirical median seen as an estimator of a location parameter. See, for instance, both [15, Example 3.1] and [16, Chapter III, Example 1]. Combining these last facts with (17) gives the following proposition which states the superiority—from the point of view of robustness—of MDEs with respect to the median of the empirical distribution.

**Proposition 5.3.** If, in addition to the hypotheses assumed in Proposition 5.2, $(Z^-,Z^+)$ is an admissible pair of strictly monotone functions on $(0,1)$, then

$$|h_\xi| \leq \frac{1}{2f(F^{-1}(\frac{1}{2}))}.$$ 

Moreover, there is some $\xi$ for which the inequality is strict.

As for the hypothesis that both $Z^-$ and $Z^+$ are monotone functions, it should be noted that it is tantamount to assuming that the functions $(1 - F)$ and $F$ are log-concave or log-convex, and it is worth recalling that log-convexity is passed from densities to p.d. functions; see [20]. Since there is a large number of commonly used p.d.'s with log-concave densities (see [1]) the condition by which $Z^-$ and $Z^+$ must be monotone functions appears not to be seriously restrictive.

**6. Monte Carlo approximations.** Proposition 4.1 states that, under a few regularity conditions, MDEs $\hat{\theta}_n$ of unknown location-scale parameters, suitably centered and rescaled, converge in distribution to the random (unique) absolute minimum point $\hat{h}$ of

$$h \mapsto \int_{(0,1)} \left| \frac{\sigma B(x)}{f(F^{-1}(x))} - h_1 - h_2 F^{-1}(x) \right| dx,$$

where $B$ denotes a standard Brownian bridge.

Now, we use this characterization to get—by Monte Carlo methods—numerical approximations of the p.d. function of $\hat{h}$, i.e., the asymptotic distribution of $\hat{\theta}_n$. 
In particular, we employ such an approximation to determine asymptotic confidence bounds for the unknown parameters. Some examples of a limiting distribution function, for a few particular forms of $F_{\theta}$, are shown as solid curves in Figure 1, compared with the asymptotic distribution of the corresponding maximum likelihood estimator (MLE) (dashed curves). Apropos of these figures, it is worth noting that (6) implies that the limiting p.d. of $\sqrt{n}(\hat{\mu}_n - \mu)$ does not depend on $\mu$, when $\hat{\mu}_n$ is the MDE of the location parameter $\mu$, and that (9) entails the independence of $\sigma$ of the limiting p.d. of $\sqrt{n}(\hat{\sigma}_n - \sigma)/\sigma$ when $\hat{\sigma}_n$ is the MDE of the scale parameter $\sigma$. See the notation of Proposition 5.1.

The robustness of MDEs—proved in the previous section—is attained at some sacrifice of efficiency. As an example, we have computed both the asymptotic variance of MLEs and the asymptotic variance of MDEs for some specific models and have obtained the results displayed in Table 1.

In fact, these results, combined with the statements at the end of the previous section, seem to confirm the hypothesis that the trade-off between efficiency and robustness of MDEs is favorable.

In order to obtain confidence intervals or regions, with confidence level $(1 - \alpha)$, from the asymptotic distribution of some MDE, we must recall that the limiting laws
Asymptotic variance of MDE versus MLE.

<table>
<thead>
<tr>
<th></th>
<th>Location in the normal model with ( \sigma = 1 )</th>
<th>Scale in the normal model with ( \sigma = 1 )</th>
<th>Scale in the exponential model with ( \sigma = 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDE</td>
<td>1.08</td>
<td>0.64</td>
<td>1.07</td>
</tr>
<tr>
<td>MLE</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2
Confidence bounds for MDE and MLE.

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Type of estimator</th>
<th>Location in the normal model with ( \sigma = 1 )</th>
<th>Scale in the normal model</th>
<th>Scale in the exponential model</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.995</td>
<td>MDE</td>
<td>2.8863</td>
<td>2.2464</td>
<td>2.8944</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>2.8070</td>
<td>1.9849</td>
<td>2.8070</td>
</tr>
<tr>
<td>0.99</td>
<td>MDE</td>
<td>2.6606</td>
<td>2.0788</td>
<td>2.6835</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>2.5758</td>
<td>1.8214</td>
<td>2.5758</td>
</tr>
<tr>
<td>0.95</td>
<td>MDE</td>
<td>2.0303</td>
<td>1.5686</td>
<td>2.0452</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>1.9599</td>
<td>1.3859</td>
<td>1.9599</td>
</tr>
</tbody>
</table>

of \( \sqrt{n}(\hat{\mu}_n - \mu) \) and of \( \sqrt{n}(\hat{\sigma}_n - \sigma) / \sigma \) are independent of \( \mu \) and \( \sigma \), respectively. Therefore, (asymptotic) confidence intervals for the location parameter \( \mu \), with confidence level \((1 - \alpha)\), have the same form—for sufficiently large values of \( n \)—as the intervals based on MLEs, i.e., \((\hat{\mu}_n - c/\sqrt{n}, \hat{\mu}_n + c/\sqrt{n})\) with \( c \) a positive constant determined by \( P_{(\mu, \sigma)}(\hat{\mu} > c) = \alpha/2 \), where \( \hat{\mu} \) is the random minimizer of (6) with \( \sigma = \sigma_0 \). Analogously, (asymptotic) confidence intervals for the scale parameter \( \sigma \), with confidence level \((1 - \alpha)\), have—for sufficiently large values of \( n \)—the form

\[
\left( \frac{\hat{\sigma}_n}{1 + c/\sqrt{n}}, \frac{\hat{\sigma}_n}{1 - c/\sqrt{n}} \right)
\]

with \( 0 < c < \sqrt{n} \) determined by \( P_{(\mu, 1)}(\hat{\mu} > c) = \alpha/2 \), where \( \hat{\mu} \) is the random minimizer of (9) with \( \sigma = 1 \). Values of \( c \), corresponding to a few typical confidence levels and to a few different models, are presented in Table 2.

In connection with location-scale parameters, we have determined the ellipse of inertia, with confidence level \((1 - \alpha)\), for the (simulated) asymptotic distribution of the MDE of the parameter \( \theta = (\mu, \sigma) \) of the Gaussian model. Writing the ellipse of inertia as

\[
\mathcal{E}_\alpha = \{(x, y) \in \mathbb{R}^2: ax^2 + by^2 \leq c_\alpha, \quad a = 1.0419, \quad b = 0.6149 \}
\]

(see, for example, [9]) with \( c_\alpha \) satisfying

\[
P_{(0, 1)}(\hat{h} \in \mathcal{E}_\alpha) = 1 - \alpha,
\]

the region

\[
\left\{ (\mu', \sigma'): \left[ a(\mu' - \hat{\mu})^2 + \left( b - \frac{c_\alpha}{n} \right) \left( \sigma' - \frac{b}{b - c_\sigma/n} \hat{\sigma} \right)^2 \right] \leq \frac{b c_\sigma/n}{(b - c_\sigma/n)^2} \hat{\sigma}_n^2 \right\}
\]

includes the true value of the parameter \( \theta \) with probability \((1 - \alpha)\) with respect to the law of \( \hat{h} \) under \( P_\theta \). As an example, for \((1 - \alpha)\) equal to 0.95 (0.99, respectively) we have \( c_\alpha = 2.6231 \) \((c_\alpha = 3.99, \text{respectively})\).
REFERENCES


