



HAL
open science

Comparing new adaptive and robust estimators of location

Delphine Blanke, Edith Gabriel, Didier Josselin

► **To cite this version:**

Delphine Blanke, Edith Gabriel, Didier Josselin. Comparing new adaptive and robust estimators of location. Publications de l'Institut de Statistique de l'Université de Paris, 2012, Numéro spécial Robustesse, 56 (2-3), pp.65-86. <hal-01091375>

HAL Id: hal-01091375

<https://hal.science/hal-01091375v1>

Submitted on 21 Mar 2022

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.



HAL Authorization

COMPARING NEW ADAPTIVE AND ROBUST ESTIMATORS OF LOCATION

Pub. Inst. Stat. Univ. Paris
56, fasc. 2-3, 2012, 65-86

BY DELPHINE BLANKE^{*,§}, EDITH GABRIEL^{*,§} AND DIDIER JOSSELIN^{*,¶}

[§] *L.A.N.L.G., Université d'Avignon et des Pays de Vaucluse*
and [¶] *U.M.R. ESPACE, Université d'Avignon et des Pays de Vaucluse*

Abstract.

We compare 43 location estimators as regards their robustness through a Monte Carlo study. Their behaviour is examined for five sample sizes and 42 sampling distributions: 24 of which are centrally symmetric and 18 are asymmetric contaminated normal laws. We particularly focus on a new class of location estimators based on a combination of the empirical mean and median, derived from 1818 Laplace's work. The calculation of these estimators is given in the normal and uniform case, and they are approximate by bootstrap ("method 1", see Efron [6]) in the general case, based on an exact expression obtained for $\text{Cov}(\bar{X}_n, M_n)$.

Résumé.

Nous comparons, dans une étude Monte Carlo, la robustesse de 43 estimateurs d'un paramètre de localisation. Leur comportement est étudié pour cinq tailles d'échantillon et 42 lois, dont 24 sont symétriques et 18 sont des normales contaminées asymétriques. Nous nous intéressons en particulier aux résultats obtenus pour une nouvelle classe d'estimateurs, définis par une combinaison linéaire de la moyenne et de la médiane empirique et inspirés des travaux de Laplace (1818). La forme explicite de ces estimateurs est donnée pour le cas des lois normale et uniforme et une forme approchée est également obtenue par bootstrap ("méthode 1", cf Efron [6]) dans le cas général, en se basant sur une expression exacte de $\text{Cov}(\bar{X}_n, M_n)$.

1. Introduction. In many statistical analysis, the estimation of location for a given distribution is of importance, i.e. one wants to find a typical or central value that describes well the data. A large literature may be found around robust estimation of location. Roughly speaking, robustness can be defined through the stability of inference when the assumed model does not quite fit. Historical accounts of robustness appear in Huber [9], Hampel et al. [7]. We particularly refer to Ruiz-Gazen [23], in the present volume, for a detailed review. A comparative study, using Monte Carlo methods, was done by Andrews et al. [1]. This major study was next extended by Wegman and Carroll [25] to include estimators not previously examined, like adaptive ones developed in Hogg [8] (defined through preliminary sample-based calculation of some required statistics).

*ANR SYSCOMM, ROLSES

AMS 2000 subject classifications: Primary 62H12, 62F35; Secondary 62F40, 62G30

Keywords and phrases: Robustness, Location estimators, Adaptive estimators, Monte Carlo

In this paper, we focus on Laplace's estimator defined by a suitable combination of the empirical mean \bar{X}_n and median M_n and derive a new family of adaptive estimators. We perform a comprehensive simulation study to compare these adaptive estimators with classical location estimators, including the most popular, known for their efficiency and/or robustness. In previous Monte Carlo studies, principal measures of performance were relative efficiency or mean squared error criteria. Unfortunately, the obtained ranking often depends on the chosen criterion which implies some lack of robustness in the final decision. We suggest the use of a new criterion, the penalized risk ratio, designed to select estimators that are both efficient in the least favourable cases and whose overall behaviour is the most satisfactory.

The paper is organised as follows: Section 2 presents Laplace's idea and some derived adaptive estimators of location. These estimators are based on theoretical quantities that can be evaluated by bootstrap. To this aim, we propose to estimate $\text{Cov}(\bar{X}_n, M_n)$ by a direct theoretical calculation, avoiding any time-consuming Monte-Carlo approximation and referred by Efron [6] as bootstrap "method 1" (a method independently derived by Maritz and Jarrett [21] for the estimation of $\text{Var}(M_n)$). In Section 3, we present our numerical study by comparing the performance of 43 estimators for 5 sample sizes (21,51,101,501,1001) and 42 distributions (symmetric and asymmetric contaminated ones). Section 4 is devoted to some discussion as well as potential applications in Geography. All technical and theoretical results are postponed to the Appendix.

2. Laplace estimator and its variants.

2.1. *Framework.* Consider a sample X_1, \dots, X_n derived from a symmetric law $\mathbb{P}_\theta (X_1 - \theta \stackrel{L}{=} -(X_1 - \theta))$, $\theta \in \Theta \subset \mathbb{R}$, with absolutely continuous distribution function F and density denoted by f . We suppose also that X_i 's are integrable ($\mathbb{E} |X_1| < \infty$) and that f is positive in a neighbourhood of θ . Symmetry implies that for all real x , $f(\theta + x) = f(\theta - x)$. The theoretical center, θ , is well-defined: it is unique and corresponds to both the median of \mathbb{P}_θ ($\mathbb{F}(\theta) = \mathbb{P}_\theta(X_1 \leq \theta) = \frac{1}{2}$) and its expectation ($\mathbb{E}[X_1] = \int_{-\infty}^{\infty} xf(x) dx = \theta$). Then, two classical estimators of θ are the empirical mean \bar{X}_n and median M_n . More precisely, we set $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and $M_n = X_{(p+1)}$ if $n = 2p + 1$ while $M_n = \frac{1}{2}(X_{(p)} + X_{(p+1)})$ if $n = 2p$, and where $X_{(1)}, \dots, X_{(n)}$ is the order statistics associated with X_1, \dots, X_n (so that, one has almost surely $X_{(1)} < X_{(2)} < \dots < X_{(n)}$).

Since \mathbb{P} is symmetric, \bar{X}_n and M_n are unbiased estimators (whatever the parity of n) of θ , see Appendix A.1. Here, we focus on linear combination of \bar{X}_n and M_n :

$$(2.1) \quad T_n(\alpha) = (1 - \alpha)\bar{X}_n + \alpha M_n,$$

with $\alpha \in \mathbb{R}$. By this way, we expect with an "optimal" choice of α (to be defined) to combine robustness properties of M_n with efficiency of \bar{X}_n while keeping the unbiasedness property.

It is noteworthy that such estimators have been studied since the nineteenth century by Laplace [18] where a specific choice of α is discussed. Since then, these estimators have been several times “re-discovered” and studied independently under different forms. In particular, optimal linear combination of two unbiased estimators of θ is studied by Samuel-Cahn [24]; Chan and He [4] address nonparametric estimation of α while an adaptive binary (0-1) choice of α appears in Lai et al. [17]. The Laplace estimator is also cited and used in a simplified form by geographers in problems of optimal location Josselin and Ladiray [13, 15], Josselin [11, 12] or image filtering Josselin et al. [16], as well as in signal processing by Aysal and Barner [2].

2.2. On α choice. In Laplace [18]’s work, the choice of α results in the minimization of $\text{Var}(T_n(\alpha))$. Indeed, since \bar{X}_n and M_n are unbiased (in the symmetric case), we have $\text{Var}(T_n(\alpha)) = \mathbb{E}[(T_n(\alpha) - \theta)^2]$. The following expression is obtained:

$$(2.2) \quad \alpha_1 = \frac{\text{Var}(\bar{X}_n) - \text{Cov}(\bar{X}_n, M_n)}{\text{Var}(\bar{X}_n) + \text{Var}(M_n) - 2\text{Cov}(\bar{X}_n, M_n)}.$$

Josselin and Ladiray [13] propose to use the “MeAdian” (Médienne in French): namely a weighted mean of \bar{X}_n and M_n , with a weight inversely proportional to each variance:

$$(2.3) \quad \alpha_2 = \frac{\text{Var}(\bar{X}_n)}{\text{Var}(\bar{X}_n) + \text{Var}(M_n)}.$$

This choice presents several advantages: a simple geographical interpretation of the related estimator of θ , lying between \bar{X}_n et M_n and expected to be closer to the empirical mean (respectively the empirical median) as variance of the median (resp. the mean) is large.

In this paper, we propose a more “robust” approach, namely less sensitive to possible data asymmetry (see Blanke and Gabriel [3]). Our strategy relies on the minimization of the mean squared error $\mathbb{E}[(T_n(\alpha) - \theta)^2]$, also called L^2 -risk or Mean Squared Error (MSE), instead of variance, to get:

$$(2.4) \quad \alpha_3(\theta) = \frac{\mathbb{E}[\bar{X}_n(\bar{X}_n - M_n)] - \theta \mathbb{E}[\bar{X}_n - M_n]}{\mathbb{E}[(\bar{X}_n - M_n)^2]}$$

It is easy to see that $\alpha_1 = \alpha_3(\theta)$ (for all θ) as soon as $\mathbb{E}[M_n] = \mathbb{E}[\bar{X}_n] = \theta$. However, estimators of α derived from the corresponding empirical risks will be different. Using some prior estimate of the unknown θ in (2.4), we expect to improve efficiency of Laplace estimator at least toward potentially contaminated distributions (since the assumption of symmetry is not involved in the definition of α_3).

2.3. Use of bootstrap. Except for special cases (see Appendix A.3), the exact value of α_i ’s is somewhat hard to compute. Based on the asymptotic joint distribution of (\bar{X}_n, M_n) with limit covariance matrix depending on the unknown density

f , Chan and He [4] propose a nonparametric kernel density estimator of the asymptotic value of α_1 (assuming implicitly convergence of moments in their limit-in-law result). Moreover they consider only convex combinations, that is $\alpha \in [0, 1]$, while α is likely to be negative (particularly, $\alpha \approx -1/2$ in the uniform case, as shown in Appendix A.3.2).

Similarly to Josselin and Ladiray [14], we follow a bootstrap approach to evaluate the empirical counterparts of our α_i 's. By this way, the delicate choice of any smoothing parameter (as the bandwidth in the kernel case) is avoided and, moreover, better results for small sizes of sample are expected. Indeed, the methodology of Chan and He [4] can be rather inefficient since their proposed estimator concerns only the possibly asymptotic value of α_1 .

Let us recall that bootstrap's terminology was introduced in the paper of Efron [6]. The bootstrap "method 1" is based on the empirical measure of the sample, and was independently derived in Maritz and Jarrett [21] for variance estimation of the median. When theoretical calculation is not straightforward, Efron [6] propose a resampling of the data to obtain a Monte-Carlo approximation for the resulting random variable. In this paper, by following the same methodology as in Maritz and Jarrett [21], we give, in Appendix A.4, the bootstrap equivalent of the theoretical quantities involved in our α_i 's for odd n . Expressions are rather complicated and we denoted them as $\hat{\alpha}_i$:

$$\hat{\alpha}_1 = \frac{\widehat{\text{Var}}(\bar{X}_n) - \widehat{\text{Cov}}(\bar{X}_n, M_n)}{\widehat{\text{Var}}(\bar{X}_n) + \widehat{\text{Var}}(M_n) - 2\widehat{\text{Cov}}(\bar{X}_n, M_n)}$$

$$\hat{\alpha}_2 = \frac{\widehat{\text{Var}}(\bar{X}_n)}{\widehat{\text{Var}}(\bar{X}_n) + \widehat{\text{Var}}(M_n)}.$$

Here $\widehat{\text{Var}}(\bar{X}_n)$ represents the classical empirical estimator of the variance while $\widehat{\text{Var}}(M_n)$ and $\widehat{\text{Cov}}(\bar{X}_n, M_n)$ may be deduced from (A.8)-(A.10) if $n = 2p + 1$. By this way, using $\hat{\alpha}_1$ and $\hat{\alpha}_2$ we obtain two estimators:

$$(2.5) \quad \hat{\theta}_{\text{Lap}} = (1 - \hat{\alpha}_1)\bar{X}_n + \hat{\alpha}_1 M_n$$

$$(2.6) \quad \hat{\theta}_{\text{JL}} = (1 - \hat{\alpha}_2)\bar{X}_n + \hat{\alpha}_2 M_n$$

Concerning α_3 defined by (2.4), two strategies are considered to take into account the presence of the unknown θ in its formulation:

- *plug-in method*: θ is replaced by some prior estimate $\tilde{\theta}$ of θ :

$$(2.7) \quad \hat{\theta}_{\tilde{\theta}}^{(p)} = (1 - \hat{\alpha}_3(\tilde{\theta}))\bar{X}_n + \hat{\alpha}_3(\tilde{\theta})M_n$$

with

$$\hat{\alpha}_3(\tilde{\theta}) = \frac{\hat{\mathbb{E}}[\bar{X}_n^2] - \hat{\mathbb{E}}[\bar{X}_n M_n] - \tilde{\theta}(\hat{\mathbb{E}}[\bar{X}_n] - \hat{\mathbb{E}}[M_n])}{\hat{\mathbb{E}}[\bar{X}_n^2] + \hat{\mathbb{E}}[M_n^2] - 2\hat{\mathbb{E}}[\bar{X}_n M_n]}.$$

- *direct resolution*: we solve in θ the relation:

$$\theta = (1 - \alpha_3(\theta))\bar{X}_n + \alpha_3(\theta)M_n$$

resulting in the estimator

$$(2.8) \quad \hat{\theta}_{\text{dir}} = \frac{-\bar{X}_n \widehat{\mathbb{E}} [M_n(\bar{X}_n - M_n)] + M_n \widehat{\mathbb{E}} [\bar{X}_n(\bar{X}_n - M_n)]}{\widehat{\mathbb{E}} [(\bar{X}_n - M_n)^2] - (\bar{X}_n - M_n) \left(\widehat{\mathbb{E}} [\bar{X}_n] - \widehat{\mathbb{E}} [M_n] \right)}.$$

A large panel of classical, robust and/or adaptive, location estimators are also defined in Section 3.3 to allow an exhaustive comparison of the performance of each.

3. Methodology. In this section, we compare through a Monte-Carlo study the estimators defined in (2.5)-(2.8) to a wide variety of location estimators. The Monte-Carlo study involved $J = 42$ sampling distributions (see Section 3.2), $I = 43$ estimators (see Section 3.3) and $K = 5$ sample sizes: $n = 1001, 501, 101, 51, 21$. For each combination of distribution and sample size, $L = 3000$ replications have been simulated. In the following, $\hat{\theta}_{ijk\ell}$ defines the i -th estimator evaluated from the ℓ -th replicate with the j -th distribution and the k -th sample size.

3.1. *Discussion of risk criteria.* First, we compute the three classical and empirical L^2 -criteria (variance, bias, mean squared error, i.e. L^2 -risk) calculated over all replications:

$$V_{ijk} = \frac{1}{L} \sum_{\ell=1}^L \left(\hat{\theta}_{ijk\ell} - \overline{\hat{\theta}_{ijk}} \right)^2,$$

$$B_{ijk} = \frac{1}{L} \sum_{\ell=1}^L (\hat{\theta}_{ijk\ell} - \theta),$$

$$R_{ijk} = \frac{1}{L} \sum_{\ell=1}^L (\hat{\theta}_{ijk\ell} - \theta)^2 = V_{ijk} + B_{ijk}^2.$$

For fixed k , several criteria may be considered to select the best estimator. Among them, we first discuss the three criteria defined by:

- relative (variance) efficiency: $RE_k = \max_i \min_j \frac{\min_i V_{ijk}}{V_{ijk}}$,
- minimax risk: $MR_k = \min_i \max_j R_{ijk}$,
- minimax risk ratio: $MRR_k = \min_i \max_j \frac{R_{ijk}}{\min_i R_{ijk}}$.

The relative (variance) efficiency RE is a classical criterion for comparing unbiased estimators or asymptotically unbiased ones for large samples. To get the estimator less sensitive to the nature of data, the selected estimator maximizes the worst-case

ratio, defined by the minimum (among all distributions) value of ratio between the variance of the estimator and the smallest possible expected one for the relevant distribution. We do not retain this first criterion because we want to compare possibly biased estimators. In particular for small sample size, bias can be significant and should not be neglected. Therefore, we prefer to replace variance by an overall risk like the mean squared error. Despite of its widespread use, we will not consider either the minimax L^2 -risk MR. Its main drawback lies in its own definition, too pessimistic by nature: the selected estimator is the most efficient only in the worst case and may be inherently worse than all other estimators for all other distributions. In the same spirit as in the RE criterion, a possible alternative to the minimax criterion would be the minimax risk ratio MRR. This criterion selects the estimator minimizing the worst-case risk ratio. Such ratio can be interpreted as a weighted minimax risk, the overly pessimistic minimax viewpoint is then relaxed. Also this relative loss in MSE presents the advantage of being scale-invariant.

Finally, we suggest the use of the penalized risk ratio PRR:

$$(3.1) \quad \text{PRR}_k = \min_i \left(\max_j \frac{R_{ijk}}{\min_i R_{ijk}} + \text{sd}_{ik} \right)$$

where $\text{sd}_{ik} = \sqrt{\frac{1}{J} \sum_{j=1}^J \left(\frac{R_{ijk}}{\min_i R_{ijk}} - 1 \right)^2}$. With this last criterion, we hope to select a robust estimator toward the least favourable distributions, but also behaving well among all considered ones.

3.2. *Distributions.* In the Monte Carlo study, we set $\theta = 0$. About 24 real centrally symmetric distributions and 18 asymmetric contaminated distributions have been simulated.

3.2.1. *Centrally symmetric distributions.* Among the centrally symmetric distributions, 7 are non-mixtures:

1. the standard normal distribution $\mathcal{N}(0, 1)$: $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$,
2. the uniform distribution $\mathcal{U}[-\sqrt{3}, \sqrt{3}]$: $f(x) = \frac{1}{2\sqrt{3}}$ if $x \in [-\sqrt{3}, \sqrt{3}]$, 0 otherwise.
3. the standard Cauchy distribution $\mathcal{C}(0, 1)$: $f(x) = \frac{1}{\pi(1+x^2)}$
4. the logistic distribution: $f(x) = \frac{e^{-x}}{1+e^{-x}}$
5. the Laplace distribution: $f(x) = \frac{1}{2} e^{-|x|}$
6. the slash distribution: $f(x) = \frac{\varphi(0) - \varphi(x)}{x^2}$, where $\varphi(x)$ is the probability density function of the standard normal distribution
7. the Student distribution with 3 degrees of freedom: $f(x) = \frac{2}{\sqrt{3}\pi} \left(1 + \frac{x^2}{3}\right)^{-2}$

and 17 are two-components mixtures:

1. One-Out: $(n - 1)$ observations are simulated from a $\mathcal{N}(0, 1)$ and one from $\mathcal{N}(0, 9)$
2. One-Wild: $(n - 1)$ observations are simulated from a $\mathcal{N}(0, 1)$ and one from $\mathcal{N}(0, 100)$
3. Normal- α Normal contaminated: $(1 - \alpha)\mathcal{N}(0, 1) + \alpha\mathcal{N}(0, k^2)$, with $k = 3, 10$
4. Normal- α Cauchy contaminated: $(1 - \alpha)\mathcal{N}(0, 1) + \alpha\mathcal{C}(0, 1)$
5. Uniform- α Normal contaminated: $(1 - \alpha)\mathcal{U}[-\sqrt{3}, \sqrt{3}] + \alpha\mathcal{N}(0, 1)$
6. Normal- α Normal/Uniform contaminated: $(1 - \alpha)\mathcal{N}(0, 1) + \alpha\frac{\mathcal{N}(0, 1)}{\mathcal{U}[-\sqrt{3}, \sqrt{3}]}$

with $\alpha = 5\%, 10\%, 20\%$.

3.2.2. *Asymmetric contaminated distributions.* The asymmetric contaminated distributions are two-component mixtures with the main component being the standard normal distribution and the contamination component being a normal distribution centered at a point $\mu \neq \theta$. For $\alpha = 5\%, 10\%, 20\%$, these are:

$$(1 - \alpha)\mathcal{N}(0, 1) + \alpha\mathcal{N}(\mu, k^2),$$

with $\mu = 2, 4$ and $k = 1, 3, 10$.

3.3. *Estimators.* We have considered both standard and adaptive estimators, mainly defined from the empirical mean and median. Table 1 provides a code and a short description of each estimator.

3.3.1. *Standard estimators.* The empirical mean is well known to be a non-robust estimator, as very sensitive to outliers, and trimmed or winsorized means must be preferred. The $\alpha\%$ symmetrically trimmed mean is defined by

$$Tm(\alpha) = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} X_{(i)}$$

and the $\alpha\%$ symmetrically winsorized mean by

$$\frac{1}{n} \left\{ (k+1)X_{(k+1)} + \sum_{i=k+2}^{n-k-1} X_{(i)} + (k+1)X_{(n-k)} \right\},$$

where $k = [(n+1)\alpha]$ and $[\cdot]$ is the greatest integer function. The trimmed (resp. winsorized) estimators are defined with 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40% and 45% trimming (resp. winsorizing). We have also considered the trimean, which is defined by a weighted average of the quartiles: $\frac{1}{4}\{Q_1 + 2Q_2 + Q_3\}$.

M-estimators of location are solution of the equation $\arg \min_{\theta} \left(\sum_{i=1}^n \rho(x_i, \theta) \right)$, where ρ is called the objective function. Let us denote S the median of absolute deviation,

TABLE 1
Code and short description of the location estimators

Code	Description
\bar{X}_n	Mean
M_n	Median
tukey4	One-step Tukey's biweight, $k = 4$
tukey6	One-step Tukey's biweight, $k = 6$
tukey9	One-step Tukey's biweight, $k = 9$
mean05	5% symmetrically trimmed mean
mean10	10% symmetrically trimmed mean
mean15	15% symmetrically trimmed mean
mean20	20% symmetrically trimmed mean
mean25	25% symmetrically trimmed mean
mean30	30% symmetrically trimmed mean
mean35	35% symmetrically trimmed mean
mean40	40% symmetrically trimmed mean
mean45	45% symmetrically trimmed mean
winsor05	5% symmetrically winsorized mean
winsor10	10% symmetrically winsorized mean
winsor15	15% symmetrically winsorized mean
winsor20	20% symmetrically winsorized mean
winsor25	25% symmetrically winsorized mean
winsor30	30% symmetrically winsorized mean
winsor35	35% symmetrically winsorized mean
winsor40	40% symmetrically winsorized mean
winsor45	45% symmetrically winsorized mean
huber1	One-step Huber, $k = 1$
huber15	One-step Huber, $k = 1.5$
huber2	One-step Huber, $k = 2$
trimean	Weighted average of quartiles
HG1	Hogg-type adaptor using trimmed means 38%, 19%, mean and outer-mean
HG2	Hogg-type adaptor using trimmed means 38%, 25%, 10%
HG3	Hogg-type adaptor using trimmed means 38%, 10%, 5%
HG4	Hogg-type adaptor using trimmed means 38%, 25%, 19%
HG5	Hogg-type adaptor using trimmed means 38%, 25%
HG6	Hogg-type adaptor using trimmed means 38%, 19%
JLJ	Adaptive linear combination of trimmed means
	Adaptive linear combination of mean and median: $T_n(\alpha) = (1 - \alpha)\bar{X}_n + \alpha M_n$, where:
$\hat{\theta}_{Lap}$	α minimizes $\text{Var}(T_n(\alpha))$, see (2.5)
chan	α is estimated non-parametrically
$\hat{\theta}_{JL}$	α is a weight inversely proportional to each variance, see (2.6)
$\hat{\theta}_{tukey4}^{(p)}$	α minimizes $E[(T_n(\alpha) - \alpha)^2]$ and tukey4 is plugged, see (2.7)
$\hat{\theta}_{tukey6}^{(p)}$	idem with plug tukey6
$\hat{\theta}_{tukey9}^{(p)}$	idem with plug tukey9
$\hat{\theta}_{HG1}^{(p)}$	idem with plug HG1
$\hat{\theta}_{JLJ}^{(p)}$	idem with plug JLJ
$\hat{\theta}_{dir}$	direct resolution in α , see (2.8)

$S = \text{median}|x - M_n|$. With $\rho(x, \theta) := \rho(|x - \theta|)$, the objective function of the one-step Tukey's biweight estimator is given by

$$\rho(u) = \begin{cases} \frac{1}{6} [1 - (1 - u^2)]^2 & \text{if } |u| \leq 1 \\ \frac{1}{6} & \text{if } |u| > 1 \end{cases},$$

where $u = \frac{x - M_n}{kS}$ and we used $k = 4, 6, 9$. The objective function of the one-step Huber estimator is given by

$$\rho(u) = \begin{cases} \frac{1}{2}u^2 & \text{if } |u| \leq k \\ k|u| - \frac{1}{2}k^2 & \text{if } |u| > k \end{cases},$$

$u = \frac{x - M_n}{S}$ and we used $k = 1, 1.5, 2$.

3.3.2. Adaptive estimators. Adaptive estimators based on trimmed means include Hogg-type estimators and Jaeckel's adaptive trimmed mean (Andrews et al. [1]; Wegman and Carroll [25]). For $U(\alpha)$ the average of the $n\alpha$ largest order statistics and $L(\alpha)$ the average of the $n\alpha$ smallest, $OM(\alpha) = \{U(\alpha) + L(\alpha)\}/2$, and $Q = \frac{U(0.05) - L(0.05)}{U(0.5) - L(0.5)}$ and $Q' = \frac{U(0.20) - L(0.20)}{U(0.5) - L(0.5)}$, adaptive Hogg-type estimators (Hogg [8]; Wegman and Carroll [25]) are defined by:

$$HG_1 = \begin{cases} Tm(0.38) & Q > 3.2 \\ Tm(0.19) & 2.6 < Q \leq 3.2 \\ \bar{X}_n & 2 < Q \leq 2.6 \\ OM(0.25) & Q \leq 2 \end{cases} \quad HG_2 = \begin{cases} Tm(0.38) & Q' > 1.87 \\ Tm(0.25) & 1.81 < Q' \leq 1.87 \\ Tm(0.10) & Q' \leq 1.81 \end{cases}$$

$$HG_3 = \begin{cases} Tm(0.38) & Q' > 1.87 \\ Tm(0.10) & 1.81 < Q' \leq 1.87 \\ Tm(0.05) & Q' \leq 1.81 \end{cases} \quad HG_4 = \begin{cases} Tm(0.38) & Q' > 1.8 \\ Tm(0.25) & 1.55 < Q' \leq 1.8 \\ Tm(0.19) & Q' \leq 1.55 \end{cases}$$

$$HG_5 = \begin{cases} Tm(0.38) & Q' > 2.2 \\ Tm(0.25) & Q' \leq 2.2 \end{cases} \quad HG_6 = \begin{cases} Tm(0.38) & Q' > 2.2 \\ Tm(0.19) & Q' \leq 2.2 \end{cases}$$

Jaeckel [10] proposed a linear combination of two trimmed means:

$$JLJ = cTm(0.05) + (1 - c)Tm(0.25),$$

with

$$c = \frac{B\left(\frac{[0.05n]}{n}, \frac{[0.25n]}{n}\right) - A\left(\frac{[0.25n]}{n}\right)}{A\left(\frac{[0.05n]}{n}\right) - 2B\left(\frac{[0.05n]}{n}, \frac{[0.25n]}{n}\right) + A\left(\frac{[0.25n]}{n}\right)},$$

where for $n = 2p$ and $\alpha < \beta$,

$$A(\alpha) = \frac{1}{(1 - 2\alpha)^2} \sum_{j=p+1}^{n(\alpha-1)} (X_{(j)} - X_{(n-j+1)})^2$$

$$\begin{aligned}
 B(\alpha, \beta) = & \frac{1}{(1-2\alpha)(1-2\beta)} \left\{ \sum_{j=p+1}^{n(1-\beta)} (X_{(j)} - X_{(n-j+1)})^2 \right. \\
 & + \sum_{n(1-\beta)}^{n(1-\alpha)} (X_{(j)} - X_{(n-j+1)}) (X_{(n(1-\beta))} - X_{(n\beta+1)})^2 \\
 & \left. + \alpha (X_{(n(1-\beta))} - X_{(n\beta+1)}) (X_{(n(1-\alpha))} - X_{(n\alpha+1)}) \right\}
 \end{aligned}$$

Adaptive estimators defined in section 2 are linear combination of empirical mean and median $T_n(\alpha) = (1-\alpha)\bar{X}_n + \alpha M_n$. Laplace's estimator is such that α minimizes $\text{Var}(T_n(\alpha))$, see equation (2.5). Josselin and Ladiray defined α as a weight inversely proportional to the variance of \bar{X}_n and M_n , see (2.6). In Chan and He [4], α is defined by $1 - \min\{1, \max\{\pi_0, 0\}\}$, where

$$\pi_0 = \frac{\frac{1}{2f(\theta)} - \frac{\eta}{f(\theta)}}{2\left\{\sigma_X^2 + \frac{1}{4f^2(\theta)} - \frac{\eta}{f(\theta)}\right\}},$$

with $\eta = \mathbb{E}[|X - \theta|]$ and $\sigma^2 = \mathbb{E}[(X - \theta)^2]$. In practice, they consider a gaussian kernel to estimate $f(\theta)$: $\widehat{f}(\theta) = \frac{1}{nh} \sum_{i=1}^n \varphi\left(\frac{x_i - M_n}{h}\right)$, with $h = 0.79n^{-1/5} \min\{Q_3 - Q_1, \hat{\sigma}_X\}$.

In our proposal, α minimizes the mean squared error $\mathbb{E}[(T_n(\alpha) - \theta)^2]$ and thus depends on θ . For the obtained estimator (2.7), we propose to plug θ by JLJ, HG1 and Tukey's biweight estimators, as these estimators belong to the three ones having the minimum risk average when considering symmetric distributions, asymmetric contaminated distributions and all distributions for the different sample sizes. Finally, $\widehat{\theta}_{\text{dir}}$ avoids the plug with a direct solution, see (2.8).

3.4. Results. Table 2 provides ranks of the estimators according to the PRR criterion defined in (3.1). The findings are summarized as follows.

- For symmetric distributions, and thus when there is no bias, the best estimators are HG1, $\widehat{\theta}_{\text{HG1}}^{(p)}$, $\widehat{\theta}_{\text{tukey9}}^{(p)}$ and $\widehat{\theta}_{\text{Lap}}$ for every sample size.
- For asymmetric contaminated distributions as well as all distributions combined and for $n \geq 51$, $\widehat{\theta}_{\text{tukey4}}^{(p)}$, JLJ and $\widehat{\theta}_{\text{JLJ}}^{(p)}$ have the smallest ranks. When $n = 21$, JLJ and $\widehat{\theta}_{\text{JLJ}}^{(p)}$ are replaced by $\widehat{\theta}_{\text{tukey6}}^{(p)}$ and $\widehat{\theta}_{\text{tukey6}}^{(p)}$. Note that the estimators with good ranks when there is no contamination obtain high ranks in presence of contamination.
- The RE, MRR and PRR criteria provide similar results when considering symmetric distributions. In the case of asymmetric contamination, the MRR and PRR criteria here lead to very closed results.

In practice, data reliability is unknown. So, it is important to combine the results of the three different distribution types (symmetric, asymmetric contaminated and all). Table 3 gives the average ranks of Table 2 over all distribution types and over all distribution types and sample sizes. From this table, we can see that $\hat{\theta}_{\text{tukey4}}^{(p)}$, $\hat{\theta}_{\text{tukey6}}^{(p)}$, $\hat{\theta}_{\text{JLJ}}^{(p)}$, tukey4 and tukey6 always appear in the four smallest ranks. This is quite similar to the results observed from Table 2, except that the plugged estimators are better ranked because they are a far better ranked for symmetric distributions than tukey4 and tukey6. When the average is made over all sample sizes, we get the same results as the ones above described for Table 2. Finally, when ordering the estimators according to the average made over all distribution types and all sample sizes, we get $\hat{\theta}_{\text{tukey4}}^{(p)}$, tukey6, $\hat{\theta}_{\text{tukey6}}^{(p)}$ and $\hat{\theta}_{\text{JLJ}}^{(p)}$ at the four first positions.

Note that similar results have been obtained for $\theta = 4$. Finally, we also plugged θ in (2.7). The resulting "estimator", $\hat{\theta}_{\theta}^{(p)}$, was the first ranked one for every situations and whatever the criterion.

4. Discussion.

4.1. *Criteria.* The relative efficiency (RE) criterion is suitable for symmetric distributions when comparing unbiased estimators. Because of data uncertainty, in practice the minimax risk ratio (MRR) should be preferred. Indeed, asymmetric contamination leads to biased estimators; thus the variance is no longer a good criterion of robustness. In this case, RE is outperformed by MRR which is based on the L^2 -risk rather than on the variance. Even if in our simulation study the penalized risk ratio (PRR) and the MRR provide similar results, PRR should be preferred as it ensures the estimator to have a good global behaviour.

4.2. *Estimators.* Choosing an estimator is not simple task as its performance often depends on the nature of data and on the sample size. We have seen that HG1, $\hat{\theta}_{\text{HG1}}^{(p)}$, $\hat{\theta}_{\text{tukey9}}^{(p)}$ and $\hat{\theta}_{\text{Lap}}^{(p)}$ are good for symmetric distributions, whilst they are sensitive to asymmetric contamination, the case in which tukey4, $\hat{\theta}_{\text{tukey4}}^{(p)}$, JLJ and $\hat{\theta}_{\text{JLJ}}^{(p)}$ must be preferred. In all situations combined, $\hat{\theta}_{\text{tukey4}}^{(p)}$, tukey6, $\hat{\theta}_{\text{tukey6}}^{(p)}$ and $\hat{\theta}_{\text{JLJ}}^{(p)}$ can be retained. Thus, it appears that adaptive estimators are the most robust.

The simulation study showed that our adaptive estimators derived from Laplace's idea provide very good results. Their theoretical properties (efficiency, breakdown point, ...) should now be studied and this work is in progress (Blanke and Gabriel [3]). We could also extend this work to the spatial case and develop some new bivariate location estimators based on linear combination of mean and median. However this is not straightforward as there exist several definitions of the median in two-dimensional case.

TABLE 2

Rank of the estimators according to PRR, for each type of distribution (symmetric "S", asymmetric contaminated "AS", all combined) and each sample size.

	n=1001			n=501			n=101			n=51			n=21		
	S	AS	all	S	AS	all	S	AS	all	S	AS	all	S	AS	all
\bar{X}_n	43	42	42	43	42	42	43	42	43	43	42	43	43	42	43
M_n	38	7	7	38	7	7	38	6	6	38	10	11	35	11	28
tukey4	28	1	1	28	1	1	28	1	1	29	1	2	32	1	22
tukey6	10	5	5	12	5	5	11	11	12	11	12	12	11	2	1
tukey9	7	32	32	7	32	32	8	30	30	8	26	26	8	24	15
mean05	39	39	39	39	39	39	40	39	39	40	40	40	40	39	40
mean10	21	34	34	20	34	34	27	34	34	32	34	34	36	37	37
mean15	16	28	28	17	28	28	16	27	27	19	28	28	25	32	32
mean20	22	21	21	22	20	20	19	21	21	18	21	21	15	23	14
mean25	26	18	18	26	18	18	26	17	17	23	17	17	19	17	7
mean30	31	15	15	31	14	14	30	13	13	28	11	10	23	10	11
mean35	32	12	12	32	12	12	31	9	9	30	9	7	26	5	16
mean40	34	10	10	35	10	10	34	8	8	33	5	5	31	4	21
mean45	37	8	8	37	8	8	37	5	5	36	7	8	34	7	26
winsor05	41	41	41	41	41	41	41	41	41	41	41	41	41	41	41
winsor10	40	40	40	40	40	40	39	40	40	39	39	39	39	40	39
winsor15	30	38	38	29	38	38	33	38	38	37	38	38	37	38	38
winsor20	17	33	33	16	33	33	20	33	33	26	33	33	30	36	36
winsor25	18	25	25	19	25	25	17	24	24	16	24	24	21	26	20
winsor30	23	19	19	23	19	19	21	18	18	21	18	18	17	20	8
winsor35	29	16	16	30	16	16	29	15	14	27	13	14	24	12	13
winsor40	33	13	13	33	13	13	32	10	10	31	8	9	29	8	19
winsor45	36	9	9	36	9	9	36	7	7	35	6	6	33	6	24
huber1	15	23	23	15	21	21	14	20	20	14	19	19	12	21	9
huber15	9	29	30	9	30	30	15	31	31	17	30	30	16	28	25
huber2	19	35	35	18	35	35	22	36	36	25	37	37	27	35	34
trimean	24	22	22	24	22	22	23	22	22	20	22	22	20	22	12
HG1	1	37	37	1	37	37	1	37	37	1	36	36	1	34	33
HG2	13	14	14	13	15	15	10	19	19	9	20	20	10	13	4
HG3	8	20	20	8	24	24	7	25	25	4	25	25	7	19	5
HG4	27	11	11	27	11	11	24	12	11	24	14	13	22	9	10
HG5	25	17	17	25	17	17	25	16	16	22	16	16	18	16	6
HG6	20	24	24	21	23	23	18	23	23	15	23	23	9	27	23
JLJ	35	3	3	34	3	3	35	2	3	34	3	3	38	18	35
$\hat{\theta}_{Lap}$	4	26	26	4	26	26	3	26	26	3	29	29	4	30	29
chan	5	27	27	5	27	27	6	32	32	7	32	32	6	29	27
$\hat{\theta}_{JL}$	12	30	29	11	31	29	9	28	28	10	31	31	14	31	30
$\hat{\theta}_{tukey4}^{(p)}$	11	2	2	10	2	2	12	3	2	13	2	1	13	3	2
$\hat{\theta}_{tukey6}^{(p)}$	6	6	6	6	6	6	5	14	15	6	15	15	5	15	3
$\hat{\theta}_{tukey9}^{(p)}$	3	31	31	3	29	31	4	29	29	5	27	27	3	25	18
$\hat{\theta}_{HG1}^{(p)}$	2	36	36	2	36	36	2	35	35	2	35	35	2	33	31
$\hat{\theta}_{JLJ}^{(p)}$	14	4	4	14	4	4	13	4	4	12	4	4	28	14	17
$\hat{\theta}_{dir}$	42	43	43	42	43	43	42	43	42	42	43	42	42	43	42

TABLE 3
Average of ranks related to PRR.

	average over distribution types					average over all
	$n = 1001$	$n = 501$	$n = 101$	$n = 51$	$n = 21$	
\bar{X}_n	42.33	42.33	42.67	42.67	42.67	42.53
M_n	17.33	17.33	16.67	19.67	24.67	19.13
tukey4	10	10	10	10.67	18.33	11.8
tukey6	6.67	7.33	11.33	11.67	4.67	8.33
tukey9	23.67	23.67	22.67	20	15.67	21.13
mean05	39	39	39.33	40	39.67	39.4
mean10	29.67	29.33	31.67	33.33	36.67	32.13
mean15	24	24.33	23.33	25	29.67	25.27
mean20	21.33	20.67	20.33	20	17.33	19.93
mean25	20.67	20.67	20	19	14.33	18.93
mean30	20.33	19.67	18.67	16.33	14.67	17.93
mean35	18.67	18.67	16.33	15.33	15.67	16.93
mean40	18	18.33	16.67	14.33	18.67	17.2
mean45	17.67	17.67	15.67	17	22.33	18.07
winsor05	41	41	41	41	41	41
winsor10	40	40	39.67	39	39.33	39.6
winsor15	35.33	35	36.33	37.67	37.67	36.4
winsor20	27.67	27.33	28.67	30.67	34	29.67
winsor25	22.67	23	21.67	21.33	22.33	22.2
winsor30	20.33	20.33	19	19	15	18.73
winsor35	20.33	20.67	19.33	18	16.33	18.93
winsor40	19.67	19.67	17.33	16	18.67	18.27
winsor45	18	18	16.67	15.67	21	17.87
huber1	20.33	19	18	17.33	14	17.73
huber15	22.67	23	25.67	25.67	23	24
huber2	29.67	29.33	31.33	33	32	31.07
trimean	22.67	22.67	22.33	21.33	18	21.4
HG1	25	25	25	24.33	22.67	24.4
HG2	13.67	14.33	16	16.33	9	13.87
HG3	16	18.67	19	18	10.33	16.4
HG4	16.33	16.33	15.67	17	13.67	15.8
HG5	19.67	19.67	19	18	13.33	17.93
HG6	22.67	22.33	21.33	20.33	19.67	21.27
JLJ	13.67	13.33	13.33	13.33	30.33	16.8
$\hat{\theta}_{Lap}$	18.67	18.67	18.33	20.33	21	19.4
chan	19.67	19.67	23.33	23.67	20.67	21.4
$\hat{\theta}_{JL}$	23.67	23.67	21.67	24	25	23.6
$\hat{\theta}_{tukey4}^{(p)}$	5	4.67	5.67	5.33	6	5.33
$\hat{\theta}_{tukey6}^{(p)}$	6	6	11.33	12	7.67	8.6
$\hat{\theta}_{tukey9}^{(p)}$	21.67	21	20.67	19.67	15.33	19.67
$\hat{\theta}_{HG1}^{(p)}$	24.67	24.67	24	24	22	23.87
$\hat{\theta}_{JLJ}^{(p)}$	7.33	7.33	7	6.67	19.67	9.6
$\hat{\theta}_{dir}$	42.67	42.67	42.33	42.33	42.33	42.47

4.3. *Applications in Geography.* By the way, the L^2 -risk is often used in Geography, especially in the field of Remote Sensing and Image Analysis, under its close simpler expression: the RMSE (Root Mean Squared Error, Longley et al. [20]). For instance, this error assessment is used to match or assemble images which don't have the same geographical projection. In that particular case, we consider a theoretical reference image that encompasses the 'good' projection, as the θ . Then the second image, i.e. the estimated image, needs to be stretched into the reference projection, using a set of reference points. Since unbiased estimators are usually computed, only the sample variance is given instead of the MSE, separating the variance from the bias. It could be of interest to study the overall robustness of L^2 -risk, compared to the variance. Indeed, although L^2 -risk advantageously introduces a bias term, both still suffer from weak robustness due to important sensitivity to outliers. Instead of considering criterion (3.1), it also would be possible to replace the variance by the Least or the Median Absolute Deviations (resp. LAD and MAD) ; for instance, the MAD is used for several M -estimators we tested in this paper. In general, using robust estimates of centrality becomes crucial when users have to deal with empirical data coming from ground observation. On the one hand, these data can be handled within large sets of little distributions including only a few individuals. On another hand, it is really difficult for geographers to precisely determine which is the underlying statistical related law(s). Are they general ? Do they only have a local effect ? In these conditions, finding the most robust location estimator that is able to adapt to empirical batches of data or/and unknown statistical laws is of prior interest in spatial analysis. Notwithstanding this estimator useful adaptability, we also propose to search for the possible local or general spatial laws in order to choose, according to space peculiarities, the appropriate simulated theoretical laws during Monte-Carlo processing. This would allow to adapt the estimator efficiency assessment while keeping its global property of robustness.

APPENDIX A: DEVELOPMENTS

A.1. Bias of empirical median.

Case: n odd. If $n = 2p + 1$, one has $M_n = X_{(p+1)}$ which has the density (see e.g. David [5, p. 8]):

$$(A.1) \quad f_{(p+1)}(x) = \frac{(2p+1)!}{(p!)^2} \mathbb{F}^p(x) (1 - \mathbb{F}(x))^p f(x).$$

By symmetry of \mathbb{P}_θ , we have $\mathbb{F}(\theta - x) = 1 - \mathbb{F}(\theta + x)$ and $f(\theta + x) = f(\theta - x)$. From (A.1), we get that $f_{(p+1)}(\theta + x) = f_{(p+1)}(\theta - x)$ for all real x , therefore $f_{(p+1)}$ is also a symmetric density. ■

Case: n even. If $n = 2p$, we have $M_n = \frac{1}{2}(X_{(p)} + X_{(p+1)})$. Symmetry of \mathbb{P}_θ implies also that $f_{(r)}(x + \theta) = f_{X_{(n-r+1)}}(\theta - x)$ (see e.g. David [5, p. 19, Ex. 2.12]) giving in turn, for $r = p + 1$, that $f_{(p+1)}(\theta + x) = f_{(p)}(\theta - x)$ for all real x . We obtain

successively

$$\begin{aligned} \mathbb{E}(M_n) &= \frac{1}{2} \int_{-\infty}^{\infty} x f_{(p)}(x) dx + \frac{1}{2} \int_{-\infty}^{\infty} x f_{(p+1)}(x) dx \\ &= \frac{1}{2} \int_{-\infty}^{\infty} (\theta - y) f_{(p)}(\theta - y) dy + \frac{1}{2} \int_{-\infty}^{\infty} (\theta + y) f_{(p)}(\theta - y) dy = \theta. \end{aligned}$$

■

A.2. General expressions for $\mathbb{E}(\bar{X}_n M_n)$.

Case: $n = 2p + 1$.

PROPOSITION A.1. We have for i.i.d. X_i :

$$(A.2) \quad \mathbb{E}\left(\sum_{i=1}^{2p+1} X_i \mid M_n = t\right) = \frac{p}{\mathbb{F}(t)} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < t\}}) + t + \frac{p}{1 - \mathbb{F}(t)} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > t\}})$$

and

$$(A.3) \quad \begin{aligned} \mathbb{E}(\bar{X}_n M_n) &= \frac{(2p)!}{(p!)^2} \int_{-\infty}^{\infty} x \mathbb{F}^{p-1}(x) (1 - \mathbb{F}(x))^{p-1} f(x) \left\{ p(1 - \mathbb{F}(x)) \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < x\}}) \right. \\ &\quad \left. + p\mathbb{F}(x) \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > x\}}) + \mathbb{F}(x)(1 - \mathbb{F}(x))x \right\} dx. \end{aligned}$$

PROOF: Let $f_{(i,p+1)}$ be the density of $(X_{(i)}, X_{(p+1)})$. We use (A.1) and also, from David [5, page 9]:

$$(A.4) \quad f_{(i,p+1)}(x, t) = \frac{(2p+1)!}{(i-1)!(p-i)!p!} \mathbb{F}^{i-1}(x) f(x) (\mathbb{F}(t) - \mathbb{F}(x))^{p-i} (1 - \mathbb{F}(t))^p f(t) \mathbb{1}_{x < t}$$

for $i = 1, \dots, p$ while for $i = p + 2, \dots, 2p + 1$:

$$(A.5) \quad f_{(p+1,i)}(x, t) = \frac{(2p+1)!}{(i-p-2)!(2p+1-i)!p!} \mathbb{F}^p(x) f(x) (\mathbb{F}(t) - \mathbb{F}(x))^{i-p-2} (1 - \mathbb{F}(t))^{2p+1-i} f(t) \mathbb{1}_{x < t}.$$

First, one may write

$$\begin{aligned} \mathbb{E}\left(\sum_{i=1}^{2p+1} X_i \mid M_n = t\right) &= \sum_{i=1}^{2p+1} \mathbb{E}(X_{(i)} \mid M_n = t) \\ &= \sum_{i=1}^p \mathbb{E}(X_{(i)} \mid M_n = t) + t + \sum_{i=p+2}^{2p+1} \mathbb{E}(X_{(i)} \mid M_n = t). \end{aligned}$$

Then (A.2) follows from the study of $\mathbb{E}(X_{(i)} | M_n = t)$. First, for $i = 1, \dots, p$:

$$\begin{aligned} \mathbb{E}(X_{(i)} | M_n = t) &= \int_{-\infty}^{\infty} x \frac{f_{(i,p+1)}(x, t)}{f_{(p+1)}(t)} dx \\ &= \int_{-\infty}^{\infty} xp C_{p-1}^{i-1} \frac{\mathbb{F}^{i-1}(x)(\mathbb{F}(t) - \mathbb{F}(x))^{p-i} f(x) \mathbb{1}_{x < t}}{\mathbb{F}^p(t)} dx \end{aligned}$$

from (A.1), (A.4). Next, the Binomial formula implies:

$$\begin{aligned} \sum_{i=1}^p \mathbb{E}(X_{(i)} | M_n = t) &= p \int_{-\infty}^{\infty} x \frac{(\mathbb{F}(x) + \mathbb{F}(t) - \mathbb{F}(x))^{p-1}}{\mathbb{F}^p(t)} f(x) \mathbb{1}_{x < t} dx \\ &= \frac{p}{\mathbb{F}(t)} \int_{-\infty}^t x f(x) dx = \frac{p}{\mathbb{F}(t)} \mathbb{E}(X_1 \mathbb{1}_{X_1 < t}). \end{aligned}$$

For $i = p+2, \dots, 2p+1$ and using (A.5), we arrive at

$$\sum_{i=p+2}^{2p+1} \mathbb{E}(X_{(i)} | M_n = t) = \frac{p}{1 - \mathbb{F}(t)} \int_t^{\infty} x f(x) dx = \frac{p}{1 - \mathbb{F}(t)} \mathbb{E}(X_1 \mathbb{1}_{X_1 > t})$$

and (A.2) is proved. For $\mathbb{E}(\bar{X}_n M_n)$, we write with the help of (A.1):

$$\begin{aligned} \mathbb{E}(\bar{X}_n M_n) &= \frac{1}{n} \mathbb{E}(M_n \mathbb{E}(\sum_{i=1}^n | M_n)) \\ &= \frac{(2p)!}{(p!)^2} \int_{-\infty}^{\infty} \left(\frac{px}{\mathbb{F}(x)} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < x\}}) + x^2 + \frac{px}{1 - \mathbb{F}(x)} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > x\}}) \right) \\ &\quad \times \mathbb{F}^p(x)(1 - \mathbb{F}(x))^p f(x) dx \end{aligned}$$

and (A.3) follows easily. ■

Case: $n = 2p$.

PROPOSITION A.2. We have for i.i.d. X_i :

$$\begin{aligned} \text{(A.6)} \quad \mathbb{E}\left(\sum_{i=1}^{2p} X_i | X_{(p)} = x, X_{(p+1)} = y\right) &= \\ &= \frac{p-1}{\mathbb{F}(x)} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < x < y\}}) + x + y + \frac{p-1}{1 - \mathbb{F}(y)} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > y > x\}}) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}(\bar{X}_n M_n) &= \frac{(2p-1)!}{((p-1)!)^2} \iint \left(\frac{x+y}{2} \right) \mathbb{F}^{p-2}(x)(1 - \mathbb{F}(y))^{p-2} f(x)f(y) \\ &\quad \times \left\{ (p-1)(1 - \mathbb{F}(y)) \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < x\}}) + (p-1)\mathbb{F}(x) \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > y\}}) \right. \\ &\quad \left. + \mathbb{F}(x)(1 - \mathbb{F}(y))(x+y) \right\} \mathbb{1}_{x < y} dx dy. \end{aligned}$$

PROOF: Similar to the proof of Proposition A.1, on the basis of David [5, page 9]:

$$(A.7) \quad f_{(p,p+1)}(x, y) = \frac{(2p)! \mathbb{F}^{p-1}(x) f(x) (1 - \mathbb{F}(y))^{p-1} f(y)}{((p-1)!)^2} \mathbb{1}_{x < y},$$

and

$$f_{(i,p,p+1)}(t, x, y) = \frac{(2p)! \mathbb{F}^{i-1}(t) f(t) (\mathbb{F}(x) - \mathbb{F}(t))^{p-i-1} (1 - \mathbb{F}(y))^{p-1} f(x) f(y)}{(i-1)!(p-i-1)!(p-1)!} \mathbb{1}_{t < x < y},$$

for $i = 1, \dots, p-1$ while for $i = p+2, \dots, 2p$:

$$f_{(p,p+1,i)}(x, y, t) = \frac{(2p)! \mathbb{F}^{p-1}(x) f(t) (\mathbb{F}(t) - \mathbb{F}(y))^{i-p-2} (1 - \mathbb{F}(t))^{2p-i} f(x) f(y)}{(i-p-2)!(2p-i)!(p-1)!} \mathbb{1}_{x < y < t}.$$

■

A.3. Determination of α_i 's in special cases. In this part, we compute exact values of α_i 's, defined in (2.2)-(2.4), for data with gaussian and uniform distribution.

A.3.1. *The normal case.* For i.i.d. X_i with distribution $\mathcal{N}(\theta, \sigma^2)$, one gets that \bar{X}_n is independent from $X_{(r)} - \bar{X}_n$ for all $r = 1, \dots, n$, see e.g. David [5, page 31]. This implies that for odd n , $n = 2p + 1$:

$$\text{Cov}(\bar{X}_n, M_n) = \text{Cov}(\bar{X}_n, X_{(p+1)} - \bar{X}_n) + \text{Var}(\bar{X}_n) = \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

and similarly in the even case ($n = 2p$):

$$\text{Cov}(\bar{X}_n, M_n) = \frac{1}{2} \text{Cov}(\bar{X}_n, X_{(p)} - \bar{X}_n) + \frac{1}{2} \text{Cov}(\bar{X}_n, X_{(p+1)} - \bar{X}_n) + \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

Since $\alpha_1 = \frac{\text{Var}(\bar{X}_n) - \text{Cov}(\bar{X}_n, M_n)}{\text{Var}(\bar{X}_n) + \text{Var}(M_n) - 2\text{Cov}(\bar{X}_n, M_n)} = \alpha_3(\theta)$ (for all θ), we obtain in both cases that $\alpha_1 = \alpha_3 \equiv 0$ implying in turn that $\theta_{\text{Lap}} = \theta_{\text{Plug}, \theta} = \bar{X}_n$ as expected in the gaussian case where optimality of empirical mean is well known.

Concerning α_2 , no explicit expression is available for $\text{Var}(M_n)$. From Price and Bonett [22], we get the computed values of $\text{Var}(M_n)$ for small n while the asymptotic is equivalent to $(4nf^2(\theta))^{-1}$. For $n = 19$, $\theta = 0$ and $\sigma^2 = 1$, we deduce that $\alpha_2 \simeq 0.394$ and, asymptotically, $\alpha_2 = 0.389$. By this way, a non negligible mass is allowed for the sample median in the calculation of $\hat{\theta}_{\text{JL}}$.

A.3.2. *The uniform case.* For i.i.d. random variables from $\mathcal{U}_{[\theta-a, \theta+a]}$, one gets after some easy but tedious calculation (from expressions given in (A.1) and (A.7)) that for $n = 2p + 1$, $\mathbb{E}(M_n^2) = \theta^2 + \frac{a^2}{3+2p}$ while for $n = 2p$, $\mathbb{E}(M_n^2) = \theta^2 + \frac{a^2 p}{(p+1)(2p+1)}$. To compute the covariance, the easier way is to start from (A.2) with the evaluation of $\mathbb{E}(\bar{X}_n M_n | M_n)$. Since $F(t) = \frac{t-\theta+a}{2a}$ for $t \in [\theta - a, \theta + a]$ and $\mathbb{E}(X \mathbb{1}_{\{X < t\}}) = \frac{(t-\theta+a)(t+\theta-a)}{4a}$, $\mathbb{E}(X \mathbb{1}_{\{X > t\}}) = \frac{(\theta+a-t)(\theta+a+t)}{4a}$, we get for $n = 2p + 1$:

$$\begin{aligned} n\mathbb{E}(\bar{X}_n M_n) &= \mathbb{E}(M_n \mathbb{E}(\sum_{i=1}^n X_i | M_n)) \\ &= \frac{p}{2} \mathbb{E}(M_n^2 + (\theta - a)M_n) + \mathbb{E}(M_n^2) + \frac{p}{2} \mathbb{E}(M_n^2 + (\theta + a)M_n) \\ &= (p + 1)\mathbb{E}(M_n^2) + p\theta\mathbb{E}(M_n) \\ &= (2p + 1)\theta^2 + \frac{(p + 1)a^2}{3 + 2p} \end{aligned}$$

yielding to $\mathbb{E}(\bar{X}_n M_n) = \theta^2 + \frac{(n+1)a^2}{2n(n+2)}$ for odd n . If $n = 2p$, we proceed similarly on the basis of (A.6):

$$\begin{aligned} n\mathbb{E}(\bar{X}_n M_n) &= \mathbb{E}(M_n \mathbb{E}(\sum_{i=1}^n X_i | X_{(p)}, X_{(p+1)})) \\ &= \frac{p-1}{2} \mathbb{E}(M_n(X_{(p)} + \theta - a)) + 2M_n^2 + \frac{p-1}{2} \mathbb{E}(M_n(\theta + a + X_{(p+1)})) \\ &= (p + 1)\mathbb{E}(M_n^2) + (p - 1)\theta\mathbb{E}(M_n) \\ &= 2p\theta^2 + \frac{pa^2}{1 + 2p}, \end{aligned}$$

implying that $\mathbb{E}(\bar{X}_n M_n) = \theta^2 + \frac{a^2}{2(n+1)}$ for n even.

Again $\alpha_1 = \alpha_3(\theta)$ for all θ , and we deduce that $\alpha_1 = -\frac{1}{2}$ for odd n and $\alpha_1 = -\frac{n+2}{2(n-1)}$ in the even case. We may conclude that a negative mass is placed on M_n , the estimator privileging \bar{X}_n to M_n , a fact already noticed by Samuel-Cahn [24]. Indeed for unimodal symmetric densities, or more generally densities symmetric about zero and satisfying $f(x) \leq f(0)$ for all x , the asymptotic efficiency of M_n to \bar{X}_n is always greater than $\frac{1}{3}$, the lower bound being attained only for the uniform distribution, see Lehmann [19, page 359]. By this way, in the uniform case, M_n reaches his worst behaviour toward \bar{X}_n .

Concerning $\hat{\theta}_{JL}$, since this estimator is a convex combination of \bar{X}_n and M_n , we obtain a very different result: $\alpha_2 = \frac{n+2}{4n+2}$ for $n = 2p + 1$ while $\alpha_2 = \frac{(n+1)(n+2)}{(n+1)(n+2)+3n^2}$ for $n = 2p$ so that $\alpha_2 \sim \frac{1}{4}$ in both cases.

A.4. Bootstrap "method 1" for $n = 2p + 1$. Approximations of $\mathbb{E}(M_n)$ and $\text{Var}(M_n)$, $\widehat{\mathbb{E}}(M_n)$ and $\widehat{\text{Var}}(M_n)$, had been proposed by Maritz and Jarrett [21] and independently derived by Efron [6]. We recall their results, for $r \in \mathbb{N}^*$ and $n = 2p + 1$:

$$(A.8) \quad \widehat{\mathbb{E}}(M_n) = \frac{(2p+1)!}{(p!)^2} \sum_{i=1}^n X_{(i)} \left(\mathbb{B}\left(\frac{i}{n}, p+1, p+1\right) - \mathbb{B}\left(\frac{i-1}{n}, p+1, p+1\right) \right),$$

and

$$(A.9) \quad \widehat{\text{Var}}(M_n) = \frac{(2p+1)!}{(p!)^2} \sum_{i=1}^n X_{(i)}^2 \left(\mathbb{B}\left(\frac{i}{n}, p+1, p+1\right) - \mathbb{B}\left(\frac{i-1}{n}, p+1, p+1\right) \right) - (\widehat{\mathbb{E}}(M_n))^2,$$

where $\mathbb{B}(x, a, b)$ is the incomplete beta function defined by

$$\mathbb{B}(x, a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt.$$

Finally, note that the case $n = 2p$ is also addressed by Maritz and Jarrett [21].

Of course, the same methodology applies for $\mathbb{E}(\overline{X}_n)$ and $\text{Var}(\overline{X}_n)$ and gives, as expected, the classical empirical estimates:

$$\widehat{\mathbb{E}}(\overline{X}_n) = \widehat{\mathbb{E}}(X_1) = \overline{X}_n \quad \text{and} \quad n \widehat{\text{Var}}(\overline{X}_n) = \widehat{\text{Var}}(X_1) = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X}_n)^2.$$

Here, we propose to compute similar approximations of $\widehat{\mathbb{E}}(\overline{X}_n M_n)$ to deduce $\widehat{\text{Cov}}(\overline{X}_n, M_n)$, both involved in (2.5) and (2.7). For $n = 2p + 1$, we start from (A.3):

$$\begin{aligned} \mathbb{E}(\overline{X}_n M_n) &= \frac{(2p)!}{(p!)^2} \int_{-\infty}^{\infty} x \mathbb{F}^{p-1}(x) (1 - \mathbb{F}(x))^{p-1} f(x) \left\{ p(1 - \mathbb{F}(x)) \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < x\}}) \right. \\ &\quad \left. + p\mathbb{F}(x) \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > x\}}) + \mathbb{F}(x)(1 - \mathbb{F}(x))x \right\} dx \end{aligned}$$

We have:

$$\begin{aligned} \mathbb{E}(X_1 \mathbb{1}_{\{X_1 < x\}}) &= \int_{-\infty}^x t d\mathbb{F}(t) = \int_0^{\mathbb{F}(x)} \mathbb{F}^{-1}(z) dz \quad \text{and} \\ \mathbb{E}(X_1 \mathbb{1}_{\{X_1 > x\}}) &= \int_x^{\infty} t d\mathbb{F}(t) = \int_{\mathbb{F}(x)}^1 \mathbb{F}^{-1}(z) dz. \end{aligned}$$

This gives:

$$\begin{aligned} \mathbb{E}(\bar{X}_n M_n) &= \frac{(2p)!}{(p!)^2} \int_0^1 \mathbb{F}^{-1}(y) y^{p-1} (1-y)^{p-1} \times \\ &\quad \left\{ p(1-y) \int_0^y \mathbb{F}^{-1}(z) dz + py \int_y^1 \mathbb{F}^{-1}(z) dz + y(1-y) \mathbb{F}^{-1}(y) \right\} dy \end{aligned}$$

The empirical counterpart is:

$$\begin{aligned} \widehat{\mathbb{E}}(\bar{X}_n M_n) &= \frac{(2p)!}{(p!)^2} \left(p \sum_{i=1}^{n-1} \sum_{j=i+1}^n X_{(i)} \frac{X_{(j)}}{n} \int_{\frac{i-1}{n}}^{\frac{j}{n}} z^p (1-z)^{p-1} dz \right. \\ &+ p \sum_{i=1}^n X_{(i)}^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(\frac{i}{n} - z \right) z^p (1-z)^{p-1} dz + p \sum_{i=2}^n \sum_{j=1}^{i-1} X_{(i)} \frac{X_{(j)}}{n} \int_{\frac{i-1}{n}}^{\frac{j}{n}} z^{p-1} (1-z)^p dz \\ &\left. + p \sum_{i=1}^n X_{(i)}^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} \left(z - \frac{i-1}{n} \right) z^{p-1} (1-z)^p dz + \sum_{i=1}^n X_{(i)}^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} z^p (1-z)^p dz \right) \end{aligned}$$

and easy but tedious calculations give the following equivalent:

(A.10)

$$\begin{aligned} \widehat{\mathbb{E}}(\bar{X}_n M_n) &= \frac{(2p)!}{(p!)^2} \left(p \bar{X}_n \sum_{i=1}^n X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} z^p (1-z)^{p-1} dz + \frac{1}{n} \sum_{i=1}^n X_{(i)}^2 \left(\frac{i}{n} \right)^p \left(1 - \frac{i}{n} \right)^p \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} X_{(i)} X_{(j)} \left(\left(\frac{i}{n} \right)^p \left(1 - \frac{j}{n} \right)^p - \left(\frac{j-1}{n} \right)^p \left(1 - \frac{i-1}{n} \right)^p \right) \right). \end{aligned}$$

ACKNOWLEDGEMENTS

We thank Philippe Neveux for his reference [2] on the use of the Laplace estimator in signal processing.

REFERENCES

- [1] D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, and J. W. Tukey. *Robust estimates of location: Survey and advances*. Princeton University Press, Princeton, N.J., 1972.
- [2] T. C. Aysal and K. E. Barner. Generalized mean-median filtering for robust frequency-selective applications. *IEEE Trans. Signal Process.*, 55(3):937–948, 2007.
- [3] D. Blanke and E. Gabriel. On adaptive and robust estimators derived from 1818 laplace's work. En préparation.
- [4] Y. M. Chan and X. He. A simple and competitive estimator of location. *Statist. Probab. Lett.*, 19(2):137–142, 1994.
- [5] H. A. David. *Order statistics*. Probability and Mathematical Statistics. Wiley, New-York, 1970.

- [6] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.
- [7] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. The approach based on influence functions.
- [8] Robert V. Hogg. Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *J. Amer. Statist. Assoc.*, 69:909–927, 1974. With comments by H. Leon Harter, Joseph L. Gastwirth and Peter J. Huber, and with a rejoinder by Robert V. Hogg.
- [9] P. J. Huber. The 1972 Wald lecture. Robust statistics: A review. *Ann. Math. Statist.*, 43: 1041–1067, 1972.
- [10] L. A. Jaeckel. Some flexible estimates of location. *Ann. Math. Statist.*, 42:1540–1552, 1971.
- [11] D. Josselin. Une piste pour la recherche de la “valeur centrale optimale”. Discussion autour de la robustesse et du comportement de la “médienne”, Combinaison de normes L_p . *Cybergeo : Revue européenne de géographie*, 2004.
- [12] D. Josselin. Recherche du centre d’un ensemble de points De la P -médiane à la localisation médienne. *Cybergeo : Revue européenne de géographie*, 282, 2006. Theoquant 2003.
- [13] D. Josselin and D. Ladiray. Combining L_1 and L_2 norms, for a more robust spatial analysis : the “meadian attitude”. *Cybergeo : Revue européenne de géographie*, 222, 2002.
- [14] D. Josselin and D. Ladiray. The “meadian” : a robust norm designed using exploratory spatial data analysis. In *Actes du colloque “ L_1 -norm and related methods”*, Neuchâtel, Suisse, August 2002.
- [15] D. Josselin and D. Ladiray. La médienne : un compromis robuste entre la moyenne et la médiane. In *Actes du colloque SFC2002*, Toulouse, France, September 2002.
- [16] D. Josselin, J. Rojas-Mora, and A. Ulmer. MeAdian robust spatial filtering on satellite images. *Procedia - Social and Behavioral Sciences*, 21:222–229, 2011. International Conference: Spatial Thinking and Geographic Information Sciences 2011, Science direct, <http://www.sciencedirect.com/science/article/pii/S1877042811013498>.
- [17] T. L. Lai, H. Robbins, and K. F. Yu. Adaptive choice of mean or median in estimating the center of a symmetric distribution. *Proc. Nat. Acad. Sci. U.S.A.*, 80(18 i.):5803–5806, 1983.
- [18] P. S. de Laplace. *Deuxième supplément à la théorie analytique des Probabilités*. Courcier, Paris, 1818. (Reprinted in 1995, Editions Jacques Gabay).
- [19] E. L. Lehmann. *Theory of point estimation*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1983. A Wiley Publication in Mathematical Statistics.
- [20] Paul A. Longley, Michael F. Goodchild, David J. Maguire, and David W. Rhind. *Geographic Information Systems and Science*. GIS. John Wiley & Sons inc., 2011.
- [21] J. S. Maritz and R. G. Jarrett. A note on estimating the variance of the sample median. *J. Amer. Statist. Assoc.*, 73:194–196, 1978.
- [22] R. M. Price and D. G. Bonett. Estimating the variance of the sample median. *J. Statist. Comput. Simulation*, 68(3):295–305, 2001.
- [23] A. Ruiz-Gazen. Robust statistics: a functional approach. *Ann. ISUP*, 2012.
- [24] E. Samuel-Cahn. Combining unbiased estimators. *Amer. Statist.*, 48(1):34–36, 1994.
- [25] E. J. Wegman and R. J. Carroll. A monte carlo study of robust estimators of location. *Comm. Statist. Theory Methods*, 6:795–812, 1977.

§ LABORATOIRE D’ANALYSE NON LINÉAIRE ET GÉOMÉTRIE, EA2151,
UNIVERSITÉ D’AVIGNON ET DES PAYS DE VAUCLUSE, 33 RUE LOUIS PASTEUR, 84000 AVIGNON (FRANCE)
E-MAIL: delphine.blanke@univ-avignon.fr
edith.gabriel@univ-avignon.fr

¶ U.M.R. ESPACE 6012, UNIVERSITÉ D’AVIGNON ET DES PAYS DE VAUCLUSE
74 RUE LOUIS PASTEUR, 84029 AVIGNON (FRANCE)
E-MAIL: didier.josselin@univ-avignon.fr