WEIGHTED M-ESTIMATORS FOR MULTIVARIATE CLUSTERED DATA

MOHAMMED EL ASRI, DELPHINE BLANKE, AND EDITH GABRIEL

Abstract. In this work we study weighted M-estimators for \( \mathbb{R}^d \)-valued clustered data. We give sufficient conditions for their convergence as well as their asymptotic normality. Robustness of these estimators is addressed via the study of their breakdown point. Numerical studies compare them with their unweighted version and highlight that optimal weights maximizing the relative efficiency lead to a degradation of their breakdown point.

1. Introduction

M-estimators were first introduced by Huber (1964) as robust estimators of location and gave rise to a substantial literature. For results on their asymptotic behavior and robustness (using the study of the influence function and the breakdown point), we may refer in particular to the books of Huber (1981) and Hampel et al. (1986). For more recent references, we may cite the work of Ruiz-Gazen (2012) with a nice introductory presentation of robust statistics, and the book of Van der Vaart (2000) for results, in the independent and identically distributed setting, concerning convergence and asymptotic normality in the multivariate setting considered throughout this paper.

Most of references address the case where the data are independent and identically distributed. However, clustered, and hierarchical, data frequently arise in applications. Typically the facility location problem is an important research topic in spatial data analysis for the geographic location of some economic activity. In this field, recent studies perform spatial modelling with clustered data (see e.g. Liao and Guo, 2008; Javadi and Shahrabi, 2014, and references therein). Concerning robust estimation, Nevalainen et al. (2006) study the spatial median for the multivariate one-sample location problem with clustered data. They show that the intra-cluster correlation has an impact on the asymptotic covariance matrix. The weighted spatial median, introduced in their pioneer paper of 2007, has a superior efficiency with respect to its unweighted version, especially when clusters’ sizes are heterogenous or in the presence of strong intra-cluster correlation. The class of

2000 Mathematics Subject Classification. 62F35, 62G35.

Key words and phrases. M-estimation, Clustered data, Efficiency, Breakdown point.

ANR SYSCOMM, ROLSES ‘ROBUST AND OPTIMAL LOCATIONS FOR SUSTAINABLE ENVIRONMENT AND SYSTEMS’.
weighted M-estimators (introduced in El Asri, 2013) may be viewed as a generalization of this work to a broad class of estimators: weights are assigned to the objective function that defines M-estimators. The aim is, for example, to adapt M-estimators to the clustered structures, to the size of clusters, or to clusters including extremal values, in order to increase their efficiency or robustness.

In this paper, we study the almost sure convergence of weighted M-estimators and establish their asymptotic normality. Then, we provide consistent estimators of the asymptotic variance and derived, numerically, optimal weights that improve the relative efficiency to their unweighted versions. Finally, from a weight-based formulation of the breakdown point, we illustrate how these optimal weights lead to an altered breakdown point.

2. The framework

We consider \( n \) independent clusters, \( X_1, \ldots, X_n \), where \( X_i \) is defined by \( m_i \) \( \mathbb{R}^d \)-valued random variables \( X_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i, m_i \geq 1 \), derived from the same distribution \( P_\theta \). In the sequel, \( \theta \) belongs to a set \( \Theta \) of \( \mathbb{R}^d \), with non empty interior, supposed to be convex and bounded. For each cluster \( i, i = 1, \ldots, n \), we also make the assumption that \( (X_{i1}, X_{i2}) \overset{d}{=} (X_{ik}, X_{ik'}) \) for all \( k, k' = 1, \ldots, m_i \) with \( k \neq k' \). This condition implies that the correlation between the variables of a given cluster is the same for all pairs of variables of this cluster, but the correlation may vary from one cluster to another. The total number of variables is denoted by \( N_n := \sum_{i=1}^{n} m_i \) and we suppose that \( \lim_{n \to \infty} \frac{N_n}{n} = \ell \), with \( \ell \in ]0, \infty[ \). Finally, we define \( \theta \in \Theta \) by

\[
\theta = \arg\min_{a \in \Theta} M(a) := \arg\min_{a \in \Theta} \mathbb{E}_{\theta}(\rho(X_{11}, a)) = \arg\min_{a \in \Theta} \int \rho(x, a) dP_\theta(x),
\]

where, for all \( a \in \Theta, a \mapsto \rho(x, a) \) is a measurable function in \( x \) from \( \mathbb{R}^d \times \mathbb{R}^d \) to \( \mathbb{R} \).

**Definition 2.1.** The weighted M-estimator associated with the function \( \rho \) is defined by:

\[
\hat{\theta}_n^w = \arg\min_{a \in \Theta} M_n^w(a)
\]

with

\[
M_n^w(a) = \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij}\rho(X_{ij}, a)
\]

where the \( w_{ij}, i = 1, \ldots, n, j = 1, \ldots, m_i \) are (non random) positive weights defined by the statistician.
If \( a \mapsto \rho(\cdot, a) \) is differentiable for \( a = (a_1, \ldots, a_d)^T \) in a neighborhood of \( \theta \), then the vector of partial derivatives \( \psi = (\frac{\partial \rho}{\partial a_1}, \ldots, \frac{\partial \rho}{\partial a_d})^T \) is such that \( E_\theta(\psi(X_{11}, \theta)) = 0 \). We can thus derive the following definition:

**Definition 2.2.** The weighted M-estimator \( \hat{\theta}_n^w \) is the value of \( a \) satisfying the \( d \)-vectorial equality

\[
\sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij}, a) = 0.
\]

Several choices are possible for the weights and \( w_{ij} \equiv 1 \) leads to the unweighted case. Among particular choices of interest, we have the situation where the weights are the same inside each cluster, corresponding to \( w_{ij} \equiv w_i \), for an appropriate choice of \( w_i \) in the considered framework. Typically, we may choose \( w_i = c/m_i \) for some positive constant \( c \), if we want to penalize large clusters (i.e. the ones with a high number of variables).

### 3. Asymptotic results

In the following, we refer to the assumptions listed below.

**Assumption 3.1 (A3.1).** Let us assume that:

(a) For all \( \theta \in \Theta \), for all \( \epsilon > 0 \): \( \inf_{a \in \Theta: \|a-\theta\| > \epsilon} E_\theta \rho(X_{11}, a) > E_\theta \rho(X_{11}, \theta) \);

(b) For all \( a, \theta \in \Theta \), \( E_\theta(\rho^2(X_{11}, a)) < \infty \);

(c) \( \lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} = 1 \);

(d) \( \sum_{i \geq 1} \frac{w_i^2}{\ell_i^2} < \infty \) with \( w_i = \sum_{j=1}^{m_i} w_{ij} \).

The condition A3.1(a) ensures the uniqueness of the parameter \( \theta \) achieving the minimum. It is satisfied if \( \rho \) is a strictly convex function and the support of \( P_\theta \) is not concentrated on a line (see [Milasevic and Ducharme, 1987](#) for the spatial median). The condition A3.1(b) is required in the Kolmogorov criterion. Conditions A3.1(c-d) involve both the sizes \( m_i \) and the weights \( w_{ij} \). First, in the unweighted case, \( w_{ij} \equiv 1 \), the condition A3.1(c) is immediate and clearly satisfied if \( (m_i) \) is a bounded sequence. The unbounded case is possible, for example setting \( m_i = k \in \mathbb{N}^* \) for \( i = 2^k \) and \( m_i = \ell \) otherwise (as \( \frac{N_n}{n} \to \ell \)). For the case \( w_{ij} \equiv w_i \) with \( w_n \to 1 \), the Cesàro theorem gives that \( \lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} m_i w_i = \lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} m_i w_i = 1 \) yielding to condition A3.1(c), and consequently A3.1(d) is satisfied for e.g. bounded sequences \( (m_i) \) and \( (w_i) \). Another possibility is, for example, given with the choice \( w_i = \frac{\ell_i}{m_i} \) that fulfills both conditions.

#### 3.1. Almost sure convergence

The a.s. convergence is derived in the following theorem.
Theorem 3.1. If Assumptions A3.1 are satisfied and if, for all $x$, the function $a \mapsto \rho(x, a)$ is $k(x)$-Hölderian:

$$|\rho(x, a_1) - \rho(x, a_2)| \leq k(x) \|a_1 - a_2\|^\lambda, \quad 0 < \lambda \leq 1, \quad a_1, a_2 \in \Theta$$

with $E_q(k^2(X_{11})) < \infty$, then $\hat{\theta}_n \overset{\text{a.s.}}{\to} \theta$.

Proof. The proof is based on the equivalence:

$$\forall \epsilon > 0, \lim_{n \to \infty} P\left(\left\|\hat{\theta}_m - \theta\right\| < \epsilon, \forall m > n\right) = 1 \iff \hat{\theta}_n \overset{\text{a.s.}}{\to} \theta.$$

The condition A3.1(a) implies that $\theta$ satisfies

$$\forall \epsilon > 0, \forall a : \|\theta - a\| > \epsilon, \quad M(a) > M(\theta),$$

so, for $\eta > 0$ such that $\|a - \theta\| > \epsilon$, one gets $M(a) > M(\theta) + \eta$ and $\left\{\|\hat{\theta}_n - \theta\| > \epsilon\right\} \subset \left\{M(\hat{\theta}_n) > M(\theta) + \eta\right\}$. By this way,

$$\left\{M(\theta) \leq M(\hat{\theta}_n) \leq M(\theta) + \eta\right\} \subset \left\{\|\hat{\theta}_n - \theta\| \leq \epsilon\right\}.$$

Moreover, the estimator $\hat{\theta}_n$ is the minimizer of $M_n^w$, hence $M_n^w(\hat{\theta}_n) \leq M_n^w(\theta)$ and:

$$0 \leq M(\hat{\theta}_n) - M(\theta) = M(\hat{\theta}_n) - M_n^w(\hat{\theta}_n) + M_n^w(\hat{\theta}_n) - M(\theta) \leq 2 \sup_{a \in \Theta} |M_n^w(a) - M(a)|.$$

The uniform consistency of $M_n^w(a)$ is derived in the next lemma whose proof is postponed to the end.

Lemma 3.1. If conditions of Theorem 3.1 are fulfilled,

$$\sup_{a \in \Theta} |M_n^w(a) - M(a)| \overset{\text{a.s.}}{\to} 0.$$

We may deduce the a.s. convergence of $M(\hat{\theta}_n)$ to $M(\theta)$ and so the a.s. convergence of $\hat{\theta}_n$ thanks to (3.1). \qed

Remark 3.1. Theorem 3.1 includes estimators with lipschitzian objective function as the weighted spatial median and the weighted Huber estimator derived from $\rho(x, a) = \|x - a\|$ and $\rho(x, a) = \frac{1}{2} \|x - a\|^2$ if $\|x - a\| \leq k$ and $\rho(x, a) = k \|x - a\| - \frac{1}{2} k^2$ if $\|x - a\| > k$, respectively.

3.2. Asymptotic normality. Now, we derive the asymptotic normality of weighted M-estimators under additional conditions.

Assumption 3.2 (A3.2). Let us assume that:
(a) For some finite constant $c_w$,

$$\lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij}^2 = c_w;$$

(b) For $C_i = E_\theta \psi(X_{ij}, \theta)\psi^T(X_{ij'}, \theta)$,

$$\lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{j' \neq j} w_{ij} w_{ij'} C_i = C_{\theta}^w;$$

(c) There exists $\eta > 0$ such that:

$$\lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} (\sum_{j=1}^{m_i} w_{ij})^{2+\eta} < \infty \text{ and } E_\theta(\|\psi(X_{11}, \theta)\|^{2+\eta}) < \infty.$$

For the case $w_{ij} \equiv w_i$, with $\lim_{n \to \infty} w_n = 1$, the condition $A3.2(a)$ is fulfilled. Moreover if intra-cluster correlations are the same for all clusters, $C_i = C$, and if $\lim_{n \to \infty} m_n = m' > 1$, then $C_{\theta}^w = (m' - 1)C$ and the condition $A3.2(b)$ is also satisfied. The last condition $A3.2(c)$ is involved for the application of the Lindeberg theorem, its first part is fulfilled with the previous choices of weights.

**Theorem 3.2.** Under assumptions of Theorem 3.1 and $A3.2$, suppose moreover that the function $a \mapsto \psi(x, a)$ is twice differentiable in the neighborhood of $\theta$, with partial derivatives of order 2 dominated by a $P_\theta$ square-integrable function of $x$ independent from $a$, and such that, $E_\theta(\frac{\partial \psi(X_{11}, a)}{\partial a}|_{a=\theta})^2$ and $E_\theta(\frac{\partial \psi(X_{11}, a)}{\partial a}|_{a=\theta})^{-1}$ exist. Then

$$\sqrt{\frac{1}{N_n}(\hat{\theta}_n - \theta)} \xrightarrow{d} N(0, \Sigma_w)$$

with $\Sigma_w = V_{\theta}^{-1}(c_w B_\theta + C_{\theta}^w) V_{\theta}^{-1}$, where $B_\theta := E_\theta \psi(X_{11}, \theta)\psi^T(X_{11}, \theta)$, $V_{\theta} = E_\theta(\frac{\partial \psi(X_{11}, a)}{\partial a}|_{a=\theta})$.

**Proof.** The result is obtained with a Taylor expansion of order 2 for the function $\psi$. First, for $a = (a_1, \ldots, a_d)^T$ in a convex open neighborhood of $\theta$, let us denote

$$\psi(x, a) = \left(\frac{\partial \rho(x, a)}{\partial a_1}, \ldots, \frac{\partial \rho(x, a)}{\partial a_d}\right)^T =: (\psi_1(x, a), \ldots, \psi_d(x, a))^T.$$

There exists $\tilde{\theta}_i$ in this neighborhood such that

$$\psi_i(x, a) = \psi_i(x, \theta) + \dot{\psi}_i(x, \theta)^T (a - \theta) + \frac{1}{2} (a - \theta)^T \ddot{\psi}_i(x, \tilde{\theta}_i)(a - \theta),$$

with $\dot{\psi}_i(x, \theta) = (\frac{\partial \psi_i(x, \theta)}{\partial a_1}, \ldots, \frac{\partial \psi_i(x, \theta)}{\partial a_d})^T$ and

$$\ddot{\psi}_i(x, \tilde{\theta}_i) = \begin{pmatrix}
\frac{\partial^2 \psi_i(x, \tilde{\theta}_i)}{\partial a_1^2} & \cdots & \frac{\partial^2 \psi_i(x, \tilde{\theta}_i)}{\partial a_1 \partial a_d} \\
\vdots & \ddots & \vdots \\
\frac{\partial^2 \psi_i(x, \tilde{\theta}_i)}{\partial a_d \partial a_1} & \cdots & \frac{\partial^2 \psi_i(x, \tilde{\theta}_i)}{\partial a_d^2}
\end{pmatrix}.$$
Next for \( \tilde{\theta} := (\tilde{\theta}_1, \ldots, \tilde{\theta}_d) \), we obtain:

\[
\psi(x,a) = \psi(x,\theta) + \psi(x,\theta)^\tau (a - \theta) + \frac{1}{2} D(a - \theta) \ddot{\psi}(x,\tilde{\theta})(a - \theta),
\]

with

\[
D(a - \theta) = \begin{pmatrix}
(a - \theta)^\tau & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & (a - \theta)^\tau
\end{pmatrix} = (a - \theta)^\tau \otimes I,
\]

where \( I \) is the \( d \)-identity matrix, and \( \ddot{\psi}(x,\tilde{\theta}) = \begin{pmatrix}
\ddot{\psi}_1(x,\tilde{\theta}_1) \\
\vdots \\
\ddot{\psi}_d(x,\tilde{\theta}_d)
\end{pmatrix} \).

Next, if \( T_n^w(a) = \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij},a) \), we obtain

\[
T_n^w(a) = T_n^w(\theta) + \hat{T}_n^w(\theta)^\tau (a - \theta) + \frac{1}{2} D(a - \theta) \hat{T}_n^w(\tilde{\theta})(a - \theta)
\]

and, as \( T_n^w(\hat{\theta}_n^w) = 0 \) by definition of \( \hat{\theta}_n^w \), there exists \( \tilde{\theta}_n^w \) such that:

\[
(3.2) - \sqrt{N_n} T_n^w(\theta) = \left\{ \hat{T}_n^w(\theta)^\tau + \frac{1}{2} D(\hat{\theta}_n^w - \theta) \hat{T}_n^w(\tilde{\theta}_n^w) \right\} \sqrt{N_n}(\hat{\theta}_n^w - \theta).
\]

Then, from equation (3.2), we may derive the final result from the two following technical lemmas, whose proofs are postponed to the end.

**Lemma 3.2.** Under the assumptions A3.2,

\[
\sqrt{N_n} T_n^w(\theta) \xrightarrow{d \ n \to \infty} N(0, c_w V_\theta + C_w^w).
\]

**Lemma 3.3.** Under the assumptions of Theorem 3.2, \( \hat{T}_n^w(\theta) \xrightarrow{a.s. \ n \to \infty} V_\theta \) and \( \hat{T}_n^w(a) = O_p(1) \) uniformly in \( a \).

Finally, the almost sure convergence of \( \hat{\theta}_n^w \) to \( \theta \) yields the result. \( \square \)

In the independent case (obtained with \( m_i \equiv 1 \)), the asymptotic variance reduces to \( c_w V_{\theta}^{-1} B_{\theta} V_{\theta}^{-1} \), where \( c_w \) is equal to 1 for the unweighted case \( (w_{ij} \equiv 1) \). We may deduce that weights are of no use to reduce this variance. Actually, the condition A3.1(c) gives

\[
\frac{1}{n} \sum_{i=1}^n w_i \xrightarrow{n \to \infty} 1,
\]

so by Cauchy-Schwarz inequality, we get that \( c_w = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n w_i^2 \) is necessarily greater than one.

Clustering effects appear through the term \( C_w^w \). An optimal choice of weights is then possible to reduce this value compared to the unweighted case. It should be noticed that when \( C_i \equiv C \) for all clusters, this choice (minimizing the variance) does not depend neither on \( P_{\theta} \) nor on the objective function \( \rho \).

Regularity assumptions of Theorem 3.2 exclude estimators with not sufficiently smooth objective function. Nevalainen et al. (2006) dealt the special case of the weighted spatial
median. The following result gives the asymptotic distribution of weighted $M$-estimators for location parameters with a more stringent condition on $\psi$ but without requiring the existence of $\hat{\psi}$. Here, we suppose also that $P_\theta$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$.

**Theorem 3.3.** Assume that $\theta = \text{argmin}_{a \in \Theta} E_\theta(\rho(X_{11} - a))$, that $\theta$ is the unique zero of the function $a \mapsto E_\theta(\psi(X_{11} - a))$, and that the function $a \mapsto \hat{\psi}(x - a)$ is $k$- Hölderian (uniformly in $x$). Moreover, if $E_\theta(\frac{\partial \psi(X_{11} - a)}{\partial a}|_{a=\theta})$ exists and is invertible, and $E_\theta(\frac{\partial^2 \psi(X_{11} - a)}{\partial a^2}|_{a=\theta})^2$ exists, then Assumptions A3.1(c-d) and A3.2 yield to:

1) $\hat{\theta}_n^w \xrightarrow{a.s.} \theta$;
2) $\sqrt{n}(\hat{\theta}_n^w - \theta) \xrightarrow{d} N(0, \Sigma_\theta^w)$ with the covariance matrix $\Sigma_\theta^w$ defined in Theorem 3.1.

**Proof.** Since $\theta$ is the unique zero of $\psi$, we have for all $\epsilon > 0$

$$
(3.3) \quad \inf_{\|a - \theta\| > \epsilon} \|E_\theta \psi(X_{11} - a)\| > \|E_\theta \psi(X_{11} - \theta)\| = 0.
$$

Recall that $\theta$ is also defined by $\theta = \text{argmin}_{a \in \Theta} \|E_\theta \psi(X_{11} - a)\|$. We follow the steps of the proof of Theorem 3.1 replacing condition A3.1(a) with equation (3.3), $M(a)$ with $\|E_\theta \psi(X_{11} - a)\|$, and $M_n^w(a)$ with $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij}, a)$. Then, we may conclude that $\hat{\theta}_n^w \xrightarrow{n \to \infty} \theta$.

Next, using a Taylor’s formula with integral remainder, we obtain for all $a$ in a neighborhood of $\theta$:

$$
\psi(x - a) = \psi(x - \theta) + \hat{\psi}(x - \theta)^T(a - \theta) + R(x, \theta)(a - \theta),
$$

with $R(x, \theta) = o(1)$ uniformly in $x$ with the help of the Hölderian condition on $\hat{\psi}$. For $a = \hat{\theta}_n^w \xrightarrow{n \to \infty} \theta$ and $x = X_{ij}$, we get:

$$
\psi(X_{ij} - \hat{\theta}_n^w) = \psi(X_{ij} - \theta) + \hat{\psi}(X_{ij} - \theta)(\hat{\theta}_n - \theta) + (\hat{\theta}_n^w - \theta) o_p(1)
$$

and

$$
\frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij} - \hat{\theta}_n^w) = \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij} - \theta) + \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \hat{\psi}(X_{ij} - \theta)(\hat{\theta}_n - \theta) + (\hat{\theta}_n^w - \theta) o_p(1).
$$

Using notations of Theorem 3.2, we get

$$
(3.4) \quad - \sqrt{N_n T_n^w(\theta)} = \left(\hat{T}_n^w(\theta) + o_p(1)\right) \sqrt{N_n(\hat{\theta}_n^w - \theta)}
$$
and we apply Lemma 3.2 to get
\[ \sqrt{N_n T_n^w(\theta)} \xrightarrow{d} N(0, c_w B_{\theta} + C_{\theta}^w). \]
Next, the elements of the matrix \( \dot{\psi}(X_{ij} - \theta) \) are square integrable, thus \( T_n^w(\theta) \xrightarrow{a.s.} V_{\theta} \)
as in the first part of Lemma 3.3. The result follows from (3.4).

\[ \square \]

4. Relative efficiency

4.1. Estimation of the asymptotic variance. We study the efficiency of weighted M-
estimators relative to their unweighted versions. First, recall that the unweighted case
is obtained with \( w_{ij} \equiv 1 \) for which \( \Sigma_{\theta}^w := \Sigma_{\theta} = V_{\theta}^{-1} (B_{\theta} + C_{\theta}^m) V_{\theta}^{-1} \), since \( c_w = 1 \) and \( C_{\theta}^w := C_{\theta}^m = \lim_{n \to \infty} \frac{1}{N_n} \sum_{i=1}^{n} m_i (m_i - 1) C_i \). Variances are compared by using the relative
efficiency index defined by
\[
E_f = \left[ \frac{\det(V_{\theta}^{-1} (B_{\theta} + C_{\theta}^m) V_{\theta}^{-1})}{\det(V_{\theta}^{-1} (c_w B_{\theta} + C_{\theta}^w) V_{\theta}^{-1})} \right]^{\frac{1}{2}} = \left[ \frac{\det(\Sigma_{\theta})}{\det(\Sigma_{\theta}^w)} \right]^{\frac{1}{2}}.
\]
To this end, we first propose estimators for these variances and study their a.s. behavior
under assumptions of Theorem 3.2 (or Theorem 3.3). Also we assume that the same
importance is attached to each element of a cluster by giving them the same weight
(\( w_{ij} \equiv w_i \)). To estimate \( c_w B_{\theta} \), we denote \( \hat{B}_a^w \) the matrix defined by:
\[
\hat{B}_a^w = \frac{1}{N_n} \sum_{i=1}^{n} w_i^2 \sum_{j=1}^{m_i} \psi(X_{ij}, a) \psi^T(X_{ij}, a),
\]
and similarly, estimates of \( C_{\theta}^w \) and \( V_{\theta} \) will be derived from the functionals
\[
\hat{C}_a^w = \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \sum_{j' \neq j} w_i^2 \psi(X_{ij}, a) \psi^T(X_{ij}, a),
\]
and
\[
\hat{V}_a = \frac{1}{N_n} \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \left. \frac{\partial \psi^T(X_{ij}, y)}{\partial y} \right|_{y=a}.
\]

Proposition 4.1. If assumptions of Theorem 3.2 (or Theorem 3.3) are fulfilled, and if
moreover
1) \( \sum_{i \geq 1} \frac{w_i^4 m_i^2}{i^2} < \infty \) and \( \psi \) has a moment of order 4, for all \( a \in \Theta \):
\[ \hat{B}_a^w \xrightarrow{a.s.} c_w B_a; \]
2) \( \sum_{i \geq 1} \frac{w_i^4 m_i^4}{i^2} < \infty \) and \( \psi \) has a moment of order 4, for all \( a \in \Theta \):
\[ \hat{C}_a^w \xrightarrow{a.s.} C_a^w. \]
3) \( \sum_{i \geq 1} \frac{w_i^2 m_i^2}{i^2} < \infty \) and the matrix \( \frac{\partial \psi(X_{ij}, y)}{\partial y} \bigg|_{y=a} \) has a moment of order 2, for all \( a \in \Theta \):

\[
\hat{V}_a \xrightarrow{a.s.} V_a.
\]

**Proof.** Let us denote \( \hat{B}_a^w(k, l) \), the \((k, l)\) element of the matrix \( \hat{B}_a^w \),

\[
\hat{B}_a^w(k, l) = \frac{1}{N_n} \sum_{i=1}^{n} w_i^2 \sum_{j=1}^{m_i} \psi_k(X_{ij}, a) \psi_l(X_{ij}, a).
\]

The variables \( w_i^2 \sum_{j=1}^{m_i} \psi_k(X_{ij}, \theta) \psi_l(X_{ij}, \theta) \) are independent and such that

\[
\frac{1}{N_n} \sum_{i=1}^{n} w_i^2 \mathbb{E}_\theta \left( \sum_{j=1}^{m_i} \psi_k(X_{ij}, \theta) \psi_l(X_{ij}, \theta) \right) \xrightarrow{n \to \infty} c_w B_a(k, l)
\]

and with variances denoted by \( \tilde{V}_i \). Clearly,

\[
\tilde{V}_i \leq w_i^4 \sum_{j=1}^{m_i} \mathbb{E}_\theta \left( \psi_k^2(X_{ij}, \theta) \psi_l^2(X_{ij}, \theta) \right) + w_i^4 \sum_{j \neq j'} \mathbb{E}_\theta \left( \psi_k(X_{ij}, \theta) \psi_l(X_{ij}, \theta) \psi_k(X_{ij'}, \theta) \psi_l(X_{ij'}, \theta) \right)
\]

that can be bounded again by

\[
\tilde{V}_i \leq m_i^2 w_i^4 \sqrt{\mathbb{E}_\theta \left( \psi_k^4(X_{11}, \theta) \right)} \sqrt{\mathbb{E}_\theta \left( \psi_l^4(X_{11}, \theta) \right)}.
\]

We may conclude by using the Kolmogorov’s criterion:

\[
\sum_{i \geq 1} \frac{\tilde{V}_i}{i^2} \leq \sqrt{\mathbb{E}_\theta \left( \psi_k^4(X_{11}, \theta) \right)} \sqrt{\mathbb{E}_\theta \left( \psi_l^4(X_{11}, \theta) \right)} \sum_{i \geq 1} \frac{m_i^2 w_i^4}{i^2} < \infty.
\]

Proofs are similar for \( \hat{C}_a^w \) and \( \hat{V}_a \).

Note that, under some additional regularity assumptions on the function \( \psi \) and its derivatives, one may show (similarly to Lemma 3.1) that previous convergences can turned to be uniform over \( a \in \Theta \). By this way, \( \hat{B}_a^w, \hat{C}_a^w \) and \( \hat{V}_a \) are a.s. consistent estimators of \( c_w B_\theta, C_\theta^w \) and \( V_\theta \) respectively. Next, by strengthening the assumptions on the intra-cluster distribution, especially if \( \mathbb{P}_\theta \) is completely determined by its expectation and by the covariance matrix and, if the intra-cluster correlations are the same for each cluster, the existence of moments of order 2 for \( \psi \) is only required. This is the case in the following simulation study, where we estimate the relative efficiency defined in (4.1) for the Student distribution.
4.2. **Numerical results.** Simulations are performed in R-software \([\text{R Core Team}, 2014]\). We use an empirical version of the efficiency \(E_f\) defined in (4.1) with \(\Sigma_{\theta}\) and \(\Sigma_{\theta}^{w}\) respectively replaced by \(\hat{\Sigma}_{\theta}\) and \(\hat{\Sigma}_{\theta}^{w}\). Note that here, we use the true value \(\theta\) to compare the asymptotic efficiencies in an ideal setting. We consider four configurations of 100 v.a. subdivided in 10 clusters:

1. C1: 9 clusters of size 4 and 1 cluster of size 64;
2. C2: 5 clusters of size 4 and 5 clusters of size 16;
3. C3: 2 clusters of size 4, 1 cluster of size 8, and 7 clusters of size 12;
4. C4: 10 clusters of all sizes from 5 to 15.

We consider zero mean bivariate models \((\theta = (0,0)^T)\) with independent components and Gaussian or Student distribution, the latter with \(\nu\) degrees of freedom \((\nu \in \{1,3,9\})\). We suppose that all the clusters have the same correlation. Consequently, for all \(j \neq j':\)

\[
\text{Cov}(X_{ij}, X_{ij'}) = \rho \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]

and,

\[
\text{Cov}(X_{ij}, X_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]

with \(i = 1, \ldots, 10, \rho \in [0,1]\).

Here, we present the results obtained with \(\rho = 0.2\) and \(\rho = 0.8\).

Finally, the M-estimators considered in this study are: the spatial median, the empirical mean, the Huber estimator, the \(L_p\)-median with \(p = 3, 4, 5, 6\); Weights are taken to be the same within a given cluster \((w_{ij} \equiv w_i)\).

First we optimize the weights to minimize \(\det(\hat{\Sigma}_{\theta}^w)\), using the software *Matlab* \([2010]\). Table 1 shows that optimal weights have the same order of magnitude for clusters of the same size, whatever the estimator and the correlation \(\rho\). Concerning the simulation framework, these results are in agreement with our theoretical results. Moreover, we observe that weights decrease as the clusters’ sizes increase.

In tables 2 to 4, we report different measures of efficiency: \(\left(\frac{\det(\hat{\Sigma}_{\theta}^w)}{\det(\hat{\Sigma}_{\theta}^{opt})}\right)^{\frac{1}{2}}, \left(\frac{\det(\hat{\Sigma}_{\theta})}{\det(\hat{\Sigma}_{opt})}\right)^{\frac{1}{2}}\) and, \(E_f\) defined in (4.1) and estimated by \(\hat{E}_f = \left(\frac{\det(\hat{\Sigma}_{\theta})}{\det(\hat{\Sigma}_{opt})}\right)^{\frac{1}{2}}\). For the Gaussian distribution \(\hat{\Sigma}_{\theta}^{(w)}_{\text{opt}}\) refers to the estimated variance of the (weighted) mean while for Student distributions, the (weighted) median will be the chosen reference. Not surprisingly, Huber estimators appear more robust regarding the distribution. Analyzing the results obtained in the case of a bivariate Gaussian distribution with \(\rho = 0.2\) and \(\rho = 0.8\) (Table 2), we note that, whatever the configuration of clusters, the relative efficiency of all optimally weighted M-estimators is improved (compared to their unweighted version). However, the quality of improvement depends both on configurations of clusters and correlation \(\rho\). The efficiency of weighted estimators of the first configuration (C1) is better than the others, followed (in descending order) by C2, C4 and C3. Indeed the relative efficiency, for the weighted Huber estimator and for \(\rho = 0.2\), is equal to 2.461 for C1, 1.134 for C2, 1.049 for C4 and 1.038 for C3. We also note the impact of clusters’ sizes on the variance: the improvement
is even better when sizes are heterogeneous and with presence of large clusters but, the values of $\hat{E}_f$ are not too sensitive to the choice of estimators. Next in presence of a strong correlation, the relative efficiency is also improved, and thus smaller is the variance of the weighted M-estimators (relative to their unweighted version). This is particularly the case in presence of large clusters. Considering again the weighted Huber estimator, we get the relative efficiencies $4.019$ for C1, $1.33$ for C2, $1.103$ for C4 and $1.095$ for C3 when $\rho = 0.8$. Finally, we get similar results for the Cauchy and the Student distributions (see tables 3 and 4).
Table 2. Gaussian distribution: Relative efficiencies w.r.t. to the (weighted) mean and $\hat{E}_f$

| Configuration | $\left( \frac{\text{det}(\Sigma_w)}{\text{det}(\Sigma_{\text{opt}})} \right)^{\frac{1}{2}}$ | $\left( \frac{\text{det}(\hat{\Sigma}_w)}{\text{det}(\hat{\Sigma}_{\text{opt}})} \right)^{\frac{1}{2}}$ | $\hat{E}_f$ | $\left( \frac{\text{det}(\Sigma_\theta)}{\text{det}(\Sigma_{\text{opt}})} \right)^{\frac{1}{2}}$ | $\left( \frac{\text{det}(\hat{\Sigma}_\theta)}{\text{det}(\hat{\Sigma}_{\text{opt}})} \right)^{\frac{1}{2}}$ | $\hat{E}_f$ |
|---------------|---------------------------------|---------------------------------|---------|---------------------------------|---------------------------------|---------|
| **Configuration C1**  | ρ = 0.2, right: ρ = 0.8 | | | | | |
| **Median** | 1.167 | 1.068 | 2.293 | 1.243 | 1.208 | 3.923 |
| **Mean** | 1 | 1 | 2.507 | 1 | 1 | 4.036 |
| **Huber** | 1.036 | 1.017 | 2.461 | 1.069 | 1.064 | 4.019 |
| **Lp-Median; p = 3** | 1.06 | 1.008 | 2.385 | 1.063 | 1.044 | 3.964 |
| **Lp-Median; p = 4** | 1.257 | 1.052 | 2.099 | 1.275 | 1.216 | 3.849 |
| **Lp-Median; p = 5** | 1.614 | 1.133 | 1.76 | 1.629 | 1.488 | 3.686 |
| **Configuration C2**  | ρ = 0.2, right: ρ = 0.8 | | | | | |
| **Median** | 1.078 | 1.064 | 1.128 | 1.066 | 1.142 | 1.316 |
| **Mean** | 1 | 1 | 1.142 | 1 | 1 | 1.332 |
| **Huber** | 1.029 | 1.022 | 1.134 | 1.048 | 1.046 | 1.33 |
| **Lp-Median; p = 3** | 1.059 | 1.044 | 1.126 | 1.103 | 1.095 | 1.323 |
| **Lp-Median; p = 4** | 1.217 | 1.167 | 1.095 | 1.4 | 1.38 | 1.313 |
| **Lp-Median; p = 5** | 1.51 | 1.393 | 1.054 | 1.986 | 1.942 | 1.303 |
| **Configuration C3**  | ρ = 0.2, right: ρ = 0.8 | | | | | |
| **Median** | 1.112 | 1.103 | 1.032 | 1.129 | 1.125 | 1.091 |
| **Mean** | 1 | 1 | 1.04 | 1 | 1 | 1.095 |
| **Huber** | 1.028 | 1.026 | 1.038 | 1.048 | 1.048 | 1.095 |
| **Lp-Median; p = 3** | 1.059 | 1.052 | 1.034 | 1.1 | 1.095 | 1.09 |
| **Lp-Median; p = 4** | 1.211 | 1.193 | 1.025 | 1.385 | 1.375 | 1.087 |
| **Lp-Median; p = 5** | 1.486 | 1.447 | 1.013 | 1.931 | 1.911 | 1.084 |
| **Configuration C4**  | ρ = 0.2, right: ρ = 0.8 | | | | | |
| **Median** | 1.105 | 1.096 | 1.043 | 1.13 | 1.126 | 1.1 |
| **Mean** | 1 | 1 | 1.052 | 1 | 1 | 1.103 |
| **Huber** | 1.028 | 1.025 | 1.049 | 1.048 | 1.048 | 1.103 |
| **Lp-Median; p = 3** | 1.059 | 1.051 | 1.045 | 1.095 | 1.109 | 1.098 |
| **Lp-Median; p = 4** | 1.21 | 1.19 | 1.035 | 1.365 | 1.356 | 1.095 |
| **Lp-Median; p = 5** | 1.477 | 1.435 | 1.022 | 1.87 | 1.852 | 1.092 |

5. Breakdown point

5.1. Computation of the breakdown point. Results of the previous section showed that one may choose weights to minimize the variance of the M-estimators. The benefit over the unweighted version is especially important for large clusters with high intra-correlation where the best relative efficiency is achieved for small associated weights. By this way, if such clusters include outliers, one may expect a less impact due to their low weights. A possible measure of such robustness is the breakdown point. We recall here the definition given in [Donoho and Huber (1983)](https://www-stat.stanford.edu/~hastie/StatLearnText/chap5.pdf) (see also [Davies and Gather (2005)](https://www.stat.uni-muenchen.de/~gather/Papers/robust/5.1.pdf) for a discussion paper around this notion).
Table 3. Cauchy distribution: Relative efficiencies w.r.t. to the (weighted) median and $\hat{E}_f$

|               | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|^2 | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|\|E_f\|\| | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|^2 | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|\|E_f\|\| |
|---------------|----------------------------------|---------------------------------|----------------------------------|---------------------------------|
| Configuration C1 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 2.246 | 1 | 3.868 |
| Huber         | 1.161 | 1.203 | 2.327 | 1.181 | 1.2 | 3.928 |
| Configuration C2 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 1.107 | 1 | 1.309 |
| Huber         | 1.161 | 1.17 | 1.15 | 1.169 | 1.176 | 1.317 |
| Configuration C3 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 1.03 | 1 | 1.089 |
| Huber         | 1.153 | 1.156 | 1.033 | 1.152 | 1.154 | 1.092 |
| Configuration C4 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 1.04 | 1 | 1.098 |
| Huber         | 1.129 | 1.133 | 1.043 | 1.167 | 1.169 | 1.1 |

Table 4. Student distribution with 3 df: Relative efficiencies w.r.t. to the (weighted) median and $\hat{E}_f$

|               | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|^2 | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|\|E_f\|\| | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|^2 | \((\frac{\text{det}(\Sigma^w)}{\text{det}(\Sigma^w_{\text{opt}})})\)\|\|E_f\|\| |
|---------------|----------------------------------|---------------------------------|----------------------------------|---------------------------------|
| Configuration C1 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 2.273 | 1 | 3.894 |
| Huber         | 1.947 | 2.157 | 2.518 | 1.984 | 1.995 | 3.917 |
| Configuration C2 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 1.111 | 1 | 1.313 |
| Huber         | 2.078 | 2.106 | 1.127 | 2.059 | 2.051 | 1.307 |
| Configuration C3 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 1.031 | 1 | 1.09 |
| Huber         | 2.06 | 2.017 | 1.099 | 1.983 | 1.927 | 1.059 |
| Configuration C4 (left: $\rho = 0.2$, right: $\rho = 0.8$) | | | | |
| Median        | 1 | 1.042 | 1 | 1.099 |
| Huber         | 2.079 | 2.034 | 1.019 | 2.1 | 2.042 | 1.069 |

Definition 5.1. The finite sample replacement breakdown point of $\hat{\theta}(X)$ built with $n$ observations is defined by:

$$\epsilon_n = \min_{1 \leq k \leq n} \left\{ \frac{k}{n} : \sup_{Y_k} \| \hat{\theta}(Y_k) - \hat{\theta}(X) \| = \infty \right\}$$
where, \( Y_k \) denotes the corrupted sample from \( X \) by replacing \( k \) points of \( X \) with arbitrary values.

We denote by \( \hat{\theta}(X) \) and \( \hat{\theta}^w(X) \) the unweighted and weighted M-estimators based on \( X \). By reorganizing the indexation in \( X = \{X_1, \ldots, X_{N_n}\} \) with associated weights \( W = \{w_1, \ldots, w_{N_n}\} \), these estimators can be written as

\[
\hat{\theta}(X) = \arg\min_{a \in \Theta} \frac{1}{N_n} \sum_{i=1}^{N_n} \rho(X_i, a) \quad \text{and} \quad \hat{\theta}^w(X) = \arg\min_{a \in \Theta} \frac{1}{N_n} \sum_{i=1}^{N_n} w_i \rho(X_i, a).
\]

Their breakdown point is:

\[
\epsilon^*_N = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \sup_{Y_k} \left\| \hat{\theta}(Y_k) - \hat{\theta}(X) \right\| = \infty \right\}
\]

and

\[
\epsilon^{w*}_N = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \sup_{Y_k} \left\| \hat{\theta}^w(Y_k) - \hat{\theta}^w(X) \right\| = \infty \right\}
\]

where \( Y_k \) is again the corrupted sample from \( X \) with \( k \) arbitrary values. In this finite framework, we suppose that \( \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} = 1 \), that can be written as \( \sum_{i=1}^{N_n} w_i = N_n \) with the new indexation. The following result gives controlled bounds for \( \epsilon^{w*}_{N_n} \) that depend on the weights. For conciseness and clarity, we choose rational weights: \( w_i = \frac{L}{L} \) for some \( L \geq 1 \).

**Theorem 5.1.** Suppose that the breakdown point of \( \hat{\theta}(X) \) is such that \( \epsilon^*_N \leq \epsilon^{m*}_N \leq \epsilon^*_0 \) for all integers \( m \geq 1 \). The estimator \( \hat{\theta}^w(X) \) has a breakdown point \( \frac{k^*_w}{N_n} \) such that \( k^*_w \in \left[ [k^*_1, k^*_0] \right] \) where \( k^*_0 \) and \( k^*_1 \) are respectively defined by:

\[
k^*_0 = \min_{1 \leq k \leq N_n} \left\{ k : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid w_{i_1} + \cdots + w_{i_k} \geq \epsilon^*_0 N_n \right\},
\]

\[
k^*_1 = \min_{1 \leq k \leq N_n} \left\{ k : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid w_{i_1} + \cdots + w_{i_k} \geq \epsilon^*_N N_n \right\}.
\]

**Remark 5.1.** For the spatial median, one has \( \epsilon^*_N = \frac{N_n - 1}{N_n} \) \cite{Croux:1992}. In this case, \( \epsilon^*_0 = \frac{1}{2} \) and the condition \( \epsilon^*_N \leq \epsilon^{m*}_N \) is satisfied for all \( m \geq 1 \).

**Proof.** We follow the main steps of the proof given in Nevalainen et al. \cite{Nevalainen:2006} for the breakdown point of the spatial median. Clearly \( k \) is equal to \( \#\{Y_k \setminus (Y_k \cap X)\} \), where \( \#A \) denotes the cardinal of \( A \). From the definition of \( \hat{\theta}(X) \), we get that the equivalence

\[
\sup_{Y_k} \left\| \hat{\theta}(Y_k) - \hat{\theta}(X) \right\| = \infty \iff k \geq \epsilon^*_N N_n \text{ is itself equivalent to:}
\]

\[
(5.2) \quad \sup_{Y_k} \left\| \hat{\theta}(Y_k) - \hat{\theta}(X) \right\| = \infty \iff \#\{Y_k \setminus (Y_k \cap X)\} \geq \epsilon^*_N \#\{X\}.
\]
For \( w_i = \frac{\ell_i}{L} \), with \( i = 1, \ldots, N_n \) and \( \ell_i, L \in \mathbb{N}^* \), the weighted M-estimator can be written as \( \hat{\theta}^w(X) = \text{argmin}_{a \in \Theta} \frac{1}{N_n L} \sum_{i=1}^{N_n} \ell_i \rho(X_i, a) \). Therefore \( \hat{\theta}^w(X) \), associated with \( X \), is also the unweighted estimator \( \hat{\theta}(\tilde{X}) \) where \( \tilde{X} \) is defined by each \( X_i \) of \( X \) repeated \( \ell_i \) times (and similarly for the set \( \tilde{Y}_k \) deduced from \( Y_k \)). These transformations allow us to write the breakdown point given by (5.1) as:

\[
\epsilon_{N_n}^{w*} = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \sup_{\tilde{Y}_k} \left| \hat{\theta}(\tilde{Y}_k) - \hat{\theta}(\tilde{X}) \right| = \infty \right\}.
\]

So using (5.2) and the condition \( \sum_{i=1}^{N_n} \ell_i = LN_n \), we obtain

\[
\epsilon_{N_n}^{w*} = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \#\{\tilde{Y}_k - (\tilde{Y}_k \cap \tilde{X})\} \geq \epsilon_{LN_n}^{*} \#\{\tilde{X}\} \right\},
\]

where, if \( X_{i_1}, \ldots, X_{i_k} \) are the \( k \) points replaced in \( X \), one has to replace \( \ell_{i_1} + \cdots + \ell_{i_k} \) points in \( \tilde{X} \) by arbitrary values to obtain \( \tilde{Y}_k \). Moreover \( \#\{\tilde{X}\} = \sum_{i=1}^{N_n} \ell_i \), so the breakdown point is given by

\[
\epsilon_{N_n}^{w*} = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid \ell_{i_1} + \cdots + \ell_{i_k} \geq \epsilon_{LN_n}^{*} \sum_{i=1}^{N_n} \ell_i \right\}
\]

\[
(5.3) \quad = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid w_{i_1} + \cdots + w_{i_k} \geq \epsilon_{LN_n}^{*} N_n \right\}.
\]

Let \( k_{w}^{*} \) be the minimal value obtained in (5.3). As \( k_{w}^{*} \) is given by

\[
\min_{1 \leq k \leq N_n} \left\{ k : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid w_{i_1} + \cdots + w_{i_k} \geq \epsilon_{0}^{*} N_n \right\}
\]

the property \( \epsilon_{0}^{*} \geq \epsilon_{LN_n}^{*} \) implies \( w_{i_1} + \cdots + w_{i_{k_{w}^{*}}} \geq \epsilon_{0}^{*} N_n \) yielding in turn that \( w_{i_1} + \cdots + w_{i_{k_{w}^{*}}} \geq \epsilon_{LN_n}^{*} N_n \). Therefore, we may deduce that \( k_{w} \leq k_{0}^{*} \). In the same way, from the definition of \( k_{1}^{*} \) and the condition \( \epsilon_{N_n}^{*} \leq \epsilon_{LN_n}^{*} \), we also get \( k_{w}^{*} \geq k_{1}^{*} \).

\( \square \)

From Theorem 5.1, the breakdown point of a weighted M-estimator depends more on its weights than on potential outliers. Furthermore, the proof shows that for \( w_i = \frac{\ell_i}{L} \), its exact expression takes the form given in (5.3). It is worth noting that if \( \epsilon_{LN_n}^{*} \) and \( \epsilon_{0}^{*} \) are very closed, we get

\[
\epsilon_{N_n}^{w*} = \min_{1 \leq k \leq N_n} \left\{ \frac{k}{N_n} : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid w_{i_1} + \cdots + w_{i_k} \geq \epsilon_{0}^{*} N_n \right\}
\]

which generalizes the definition given by Nevalainen et al. (2006) for the weighted spatial median (since they use the asymptotical breakdown point of 0.5 to derive their results).

We conclude this section with two remarks enlightening the following facts. The asymptotical breakdown point of the unweighted estimator cannot be improved with weights,
and, there is a trade off to reach between optimal efficiency and maximal breakdown point.

**Remark 5.2.** Suppose that the weighted estimator achieves its maximal breakdown point, \( \epsilon_{N_n}^w = \frac{k_0^*}{N_n} \), with \( k_0^* \) such that

\[
k_0^* = \min_{1 \leq k \leq N_n} \left\{ k : \exists i_1, \ldots, i_k \in \{1, \ldots, N_n\} \mid w_{i_1} + \cdots + w_{i_k} \geq \epsilon_{0}^* N_n \right\}.
\]

Then, if the weights \( w_i \) are ranked in ascending order \( w^{(1)} < \cdots < w^{(N_n)} \), minimality of \( k \) is ensured by replacing the observations with the largest \( k_0^* \) weights where \( k_0^* \) satisfies

\[
(5.4) \quad \begin{cases} 
    w^{(N_n)} + \cdots + w^{(N_n - k_0^* + 1)} \geq \epsilon_{0}^* N_n \\
    w^{(N_n)} + \cdots + w^{(N_n - k_0^* + 2)} < \epsilon_{0}^* N_n.
\end{cases}
\]

First, we may remark that the minimal improvement of the unweighted breakdown point corresponds to \( k_0^* = \epsilon_{0}^* N_n + 1 \) where, for the sake of clarity, we choose \( N_n \) such that \( \epsilon_{0}^* N_n \) is again an integer. The second part of (5.4) becomes \( \sum_{i=0}^{\epsilon_0^* N_n - 1} w_i^{(N_n-i)} < \epsilon_{0}^* N_n \). As this sum includes \( \epsilon_{0}^* N_n \) terms, necessarily one gets that \( w^{(N_n - \epsilon_0^* N_n)} < 1 \), so \( w^{(1)} \leq \cdots \leq w^{(N_n - \epsilon_0^* N_n)} < w^{(N_n - \epsilon_0^* N_n + 1)} < 1 \). This leads us to a contradiction: on one hand

\[
\sum_{i=1}^{N_n} w^{(i)} = N_n,
\]

and on the other hand,

\[
\sum_{i=1}^{N_n - \epsilon_0^* N_n} w^{(i)} < N_n - \epsilon_0^* N_n
\]

and

\[
\sum_{i=N_n - \epsilon_0^* N_n + 1}^{N_n} w^{(i)} = \sum_{i=0}^{\epsilon_0^* N_n - 1} w^{(N_n - i)} < \epsilon_0^* N_n.
\]

**Remark 5.3.** If the maximal breakdown point of the unweighted version (obtained with \( w_i \equiv 1 \)) is a proportion \( \epsilon_0^* \) (with \( 0 < \epsilon_0^* \leq \frac{1}{2} \)) of the data, then from the first equation of (5.4), we have to overweight (with weights not smaller than one) \( \epsilon_0^* N_n \) observations in \( \hat{\theta}_w(X) \) to reach this value. Thus unfortunately, at least in the case where \( w_{ij} \equiv w_i \), it seems that there is no hope to simultaneously maximize the breakdown point and the relative efficiency (since the smallest weights are assigned to the largest clusters and, consequently to the largest number of values). This is illustrated in the following section.

5.2. **Numerical results.** In this part we evaluate the breakdown point of the weighted spatial median and the weighted Huber estimator, whose unweighted versions have a maximal breakdown point, \( \epsilon_0^* = 0.5 \). We consider the configurations (C1-C4) defined in section 4.2 and we consider a centered bivariate Gaussian distribution with \( \rho = 0.2 \) and
Table 5. Breakdown point $\epsilon_{n}^{w*}$ for the weighted spatial median and the weighted Huber estimator

\[
\begin{array}{|c|c|c|c|c|}
\hline
\text{Weighted spatial median} & \text{C1} & \text{C2} & \text{C3} & \text{C4} \\
\hline
\rho = 0.2 & w_{(N_n)+\ldots+w_{(N_n-k-1)}} & 51.921 & 50.173 & 50.817 & 50.081 \\
& \epsilon_{N_n}^{w*} & 23\% & 37\% & 46\% & 41\% \\
\hline
\rho = 0.8 & w_{(N_n)+\ldots+w_{(N_n-k-1)}} & 52.099 & 50.079 & 50.177 & 50.293 \\
& \epsilon_{N_n}^{w*} & 21\% & 23\% & 41\% & 36\% \\
\hline
\text{Weighted Huber estimator} & \text{C1} & \text{C2} & \text{C3} & \text{C4} \\
\hline
\rho = 0.2 & w_{(N_n)+\ldots+w_{(N_n-k-1)}} & 50.118 & 50.326 & 50.055 & 50.611 \\
& \epsilon_{N_n}^{w*} & 22\% & 36\% & 45\% & 41\% \\
\hline
\rho = 0.8 & w_{(N_n)+\ldots+w_{(N_n-k-1)}} & 51.985 & 50.443 & 50.117 & 50.219 \\
& \epsilon_{N_n}^{w*} & 21\% & 23\% & 41\% & 36\% \\
\hline
\end{array}
\]

$\rho = 0.8$. We select the optimal weights (with $w_{ij} \equiv w_i, i = 1, \ldots, 10$) maximizing the relative efficiency (see Table 1). The breakdown points computed for these two estimators are presented in Table 5: in each case, they are far less than 50%. Variations are observed according to the number of variables by clusters and according to the correlation. Both a strong correlation and the presence of large clusters worsen the breakdown point (see for example the configuration C1). We conclude that optimal weights improve significantly the efficiency but can drastically reduce the breakdown point.

**Appendix A. Proof of the auxiliary results**

A.1. **Proof of Lemma 3.1** We begin to establish the simple a.s. convergence of $M_n^w(a)$ to $M(a)$, for all $a \in \Theta$, with the help of the Kolmogorov criterion. First, the $X_{ij}, j = 1, \ldots, m_i, i = 1, \ldots, n$, have the same law $P_\theta$, so $E_\theta(\rho(X_{ij'}, a)\rho(X_{ij}, a)) \leq E_\theta(\rho^2(X_{11}, a))$, and:

$$
E_\theta \left( \sum_{j=1}^{m_i} w_{ij} \rho(X_{ij}, a) \right)^2 \leq \left( \sum_{j=1}^{m_i} w_{ij} \right)^2 E_\theta \left( \rho^2(X_{11}, a) \right).
$$

If $Y_i := \sum_{j=1}^{m_i} w_{ij} \rho(X_{ij}, a)$, these random variables are independent and such that $\mu_i = w_i E_\theta(\rho(X_{11}, a))$ with $w_i = \sum_{j=1}^{m_i} w_{ij}$. Thus, from the condition A3.1(d),

$$
\sum_{i=1}^{\infty} \frac{\text{Var} \left( Y_i \right)}{i^2} \leq \sum_{i=1}^{\infty} \frac{w_i^2}{i^2} E_\theta \left( \rho^2(X_{ij}, a) \right) < \infty.
$$
Then, for all \( a \in \Theta \),
\[
\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} \rho(X_{ij}, a) - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} M(a) \xrightarrow{a.s. \ n \to \infty} 0
\]
and the almost sure convergence of the estimator follows from \( \lim_{n \to \infty} \frac{N_n}{n} = \ell, \ \ell \in [0, \infty[ \)
and the condition A3.1(c).

The second part is devoted to the uniform convergence. From the compactness of \( \overline{\Theta} \) (where \( \overline{A} \) denotes the closure of the set \( A \)), we get that for all \( \epsilon > 0 \), there exist \( h_\epsilon > 0 \) and \( a_1, \ldots, a_r \) in \( \Theta \) such that \( \Theta \subset \bigcup_{k=1}^{r} B(a_k, h_\epsilon) \), with \( B(a_k, h_\epsilon) \) the open ball of center \( a_k \) and radius \( h_\epsilon \). Then, for all \( a \in \Theta \),
\[
|M_n^w(a) - M(a)| \\
\leq |M_n^w(a) - M_n^w(a_k)| + |M_n^w(a_k) - M(a_k)| + |M(a_k) - M(a)|.
\]
Since \( \rho(x, \cdot) \) is \( k(x) \)-Hölderian, we get the upper bound
\[
|M_n^w(a) - M_n^w(a_k)| \leq \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} k(X_{ij}) \|a - a_k\|^\lambda, \ \lambda \in [0, 1].
\]
Since \( E_\theta(k^2(X_{ij})) < \infty \), application of the Kolmogorov criterion gives
\[
k_n := \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} k(X_{ij}) \xrightarrow{a.s. \ n \to \infty} E_\theta(k(X_{11})).
\]
Therefore, there exists \( \Omega_1 \) with \( P(\Omega_1) = 1 \), such that for all \( \omega \in \Omega_1 \),
\[
\exists N_0(\omega), \ \forall n \geq N_0(\omega) : k_n \leq k',
\]
for some positive \( k' \). Moreover, the Hölderian condition on \( M(a) \) implies also:
\[
|M(a) - M(a_k)| \leq E_\theta(k(X_{11})) \|a - a_k\|^\lambda
\]
with \( a_k \) chosen such that \( \|a - a_k\| \leq h_\epsilon \). Finally, the a.s. convergence of \( M_n^w \) to \( M \) implies that for all \( 1 \leq k \leq r \), there exists \( \Omega_2 \) such that \( P(\Omega_2) = 1 \), and for all \( \omega \in \Omega_2 \):
\[
\forall \epsilon > 0, \ \exists N'(\omega), \ \forall n > N'(\omega) : \ |M_n^w(a) - M(a)| \leq k'h_\epsilon + \epsilon/3 + k'h_\epsilon.
\]
We conclude with the choice of \( h_\epsilon \) such that \( h_\epsilon < \frac{\epsilon}{3k'} \), and the uniform convergence of \( M_n^w(a) \) on \( \Omega_1 \cap \Omega_2 \) follows. \( \square \)

A.2. Proof of Lemma 3.2. We make use of the Lindeberg theorem. First, the vectors
\[
\xi_i^w := \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij}, \theta) \text{ with } i = 1, \ldots, n, \ \text{are independent, } E_\theta(\xi_i^w) = 0, \ \text{and with variances}
\]
\[ V_i, \ i = 1, \ldots, n, \ \text{defined by} \]
\[ V_i = \sum_{j=1}^{m_i} w_{ij}^2 E_{\theta} \psi(X_{ij}, \theta) \psi^T(X_{ij}, \theta) + \sum_{j \neq j'}^{m_i} w_{ij} w_{ij'} E_{\theta} \psi(X_{ij}, \theta) \psi^T(X_{ij'}, \theta). \]

Since all pairs \((X_{ij'}, X_{ij})\), \(j \neq j'\), have the same correlation in cluster \(i\), we set \(C_i = E_{\theta} \psi(X_{ij}, \theta) \psi^T(X_{ij'}, \theta)\) and \(B := E_{\theta} \psi(X_{ij}, \theta) \psi^T(X_{ij}, \theta)\). Then the conditions A3.2(a)-(b) give:

\[ \frac{1}{N_n} \sum_{i=1}^{n} V_i \xrightarrow{n \to \infty} c_w B + C^w. \]

Next as \(\lim_{n \to \infty} \frac{N_n}{n} = \ell\), one can write:

\[ \frac{1}{n} \sum_{i=1}^{n} V_i \xrightarrow{n \to \infty} \ell (c_w B + C^w) := \Sigma. \]

Then, we verify Lindeberg’s condition:

\[ E_n := \frac{1}{n} \sum_{i=1}^{n} E_{\theta} \left( \|\xi_i\|^2 \mathbf{I}(\|\xi_i\| > \ell \sqrt{n}) \right) \xrightarrow{n \to \infty} 0. \]

First, we get:

\[ \frac{1}{n} \sum_{i=1}^{n} E_{\theta} \left( \|\xi_i\|^{-\eta} \|\xi_i\|^{2+\eta} \mathbf{I}(\|\xi_i\| > \ell \sqrt{n}) \right) \leq \sqrt{n} \frac{1}{n} \sum_{i=1}^{n} (\epsilon)^{-\eta} E_{\theta} \left( \|\xi_i\|^{2+\eta} \right), \]

and Minkowski’s inequality gives:

\[ E_{\theta} \|\xi_i\|^{2+\eta} \leq \left( \sum_{j=1}^{m_i} w_{ij} \left( E_{\theta} \|\psi(X_{ij}, \theta)\|^{2+\eta} \right)^{1/(2+\eta)} \right)^{2+\eta} = E_{\theta} \|\psi(X_{11}, \theta)\|^{2+\eta} \left( \sum_{j=1}^{m_i} w_{ij} \right)^{2+\eta}. \]

So the condition A3.2(c) gives for all \(\epsilon > 0\):

\[ E_n \leq E_{\theta} \left( \|\psi(X_{11}, \theta)\|^{2+\eta} \right) (\sqrt{n} \epsilon)^{-\eta} \frac{1}{n} \sum_{i=1}^{n} w_i^{2+\eta} \xrightarrow{n \to \infty} 0. \]

Therefore, one obtains \(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \xrightarrow{d} N(0, l(c_w B + C^w))\), yielding to the asymptotic normality of \(\sqrt{N_n T_n^w(\theta)}\) with covariance matrix given by \(c_w B + C^w\).

\]

A.3. Proof of Lemma 3.3. We have

\[ (A.1) \quad \hat{T}_n^w(\theta) = \frac{1}{N_n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} w_{ij} \psi(X_{ij}, \theta) \]

\]
and

\[ \hat{T}_n^w(a) = \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \hat{\psi}(X_{ij}, a). \]

By assumption, the elements of the matrix \( \hat{\psi}(X_{ij}, \theta) \) are square integrable, as well as for the elements of \( \hat{\psi}(X_{ij}, a) \). We verify Kolmogorov’s condition with the same procedure as for establishing the convergence of \( M_n^w(a) = \frac{1}{N_n} \sum_{i=1}^n \sum_{j=1}^{m_i} w_{ij} \rho(X_{ij}, a) \). By this way, we get the a.s. convergence for the term defined in (A.1). Next for all \( a \), the elements of \( \hat{\psi}(X_{ij}, a) \) are dominated by some function \( F(X_{ij}) \), supposed to be square integrable and independent from \( a \). This implies the Kolmogorov’s condition for the sequence \( \sum_{j=1}^{m_i} w_{ij} F(X_{ij}) \) implying in turn that \( \hat{T}_n^w(a) = O_p(1) \) uniformly in \( a \).

REFERENCES

Croux, C. and P. J. Rousseeuw (1992). A class of high-breakdown scale estimators based on subranges. *Comm. Statist. Theory Methods* 21(7), 1935–1951.

Davies, P. L. and U. Gather (2005). Breakdown and groups. *Ann. Statist.* 33(3), 977–1035. With discussions and a rejoinder by the authors.

Donoho, D. and P. J. Huber (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann*, Wadsworth Statist./Probab. Ser., pp. 157–184. Wadsworth, Belmont, CA.

El Asri, M. (2013). Propriétés asymptotiques des M-estimateurs pondérés pour des données clusterisées. *C. R. Math. Acad. Sci. Paris* 351(11-12), 491–493.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986). *Robust statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73–101.

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Javadi, M. and J. Shahrabi (2014). New spatial clustering-based models for optimal urban facility location considering geographical obstacles. *J. Ind. Eng. Int.* 10(1), 1–12.

Liao, K. and D. Guo (2008). A clustering-based approach to the capacitated facility location problem. *T. GIS*, 323–339.

MATLAB (2010). *version 7.10.0 (R2010a)*. Natick, Massachusetts.

Milasevic, P. and G. R. Ducharme (1987). Uniqueness of the spatial median. *Ann. Statist.* 15(3), 1332–1333.
Nevalainen, J., D. Larocque, and H. Oja (2006). A weighted spatial median for clustered data. *Stat. Methods Appl.* 15(3), 355–379 (2007).

Nevalainen, J., D. Larocque, and H. Oja (2007). On the multivariate spatial median for clustered data. *Canad. J. Statist.* 35(2), 215–231.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ruiz-Gazen, A. (2012). Robust statistics: a functional approach. *Ann. I.S.U.P.* 56(2-3), 49–63.

Van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge (UK), New York (N.Y.): Cambridge University Press.

E-mail address, E. El Asri: mohammed.el-asri@alumni.univ-avignon.fr

E-mail address, D. Blanke: delphine.blanke@univ-avignon.fr

E-mail address, E. Gabriel: edith.gabriel@univ-avignon.fr