The Annals of Statistics
2022, Vol. 50, No. 5, 2616–2640
https://doi.org/10.1214/22-AOS2199

AFFINE-EQUIVARIANT INFERENCE FOR MULTIVARIATE LOCATION UNDER L_p LOSS FUNCTIONS

BY ALEXANDER DÜRRE[a] AND DAVY PAINDAVEINE[b]

ECARES and Department of Mathematics, Université libre de Bruxelles, [a]*Alexander.Durre@ulb.be,* [b]*Davy.Paindaveine@ulb.be*

We consider the fundamental problem of estimating the location of a d-variate probability measure under an L_p loss function. The naive estimator, that minimizes the usual empirical L_p risk, has a known asymptotic behavior but suffers from several deficiencies for $p \neq 2$, the most important one being the lack of equivariance under general affine transformations. In this work, we introduce a collection of L_p location estimators $\hat{\mu}_n^{p,\ell}$ that minimize the size of suitable ℓ-dimensional data-based simplices. For $\ell = 1$, these estimators reduce to the naive ones, whereas, for $\ell = d$, they are equivariant under affine transformations. Irrespective of ℓ, these estimators reduce to the sample mean for $p = 2$, whereas for $p = 1$, the estimators provide the well-known spatial median and Oja median for $\ell = 1$ and $\ell = d$, respectively. Under very mild assumptions, we derive an explicit Bahadur representation result for $\hat{\mu}_n^{p,\ell}$ and establish asymptotic normality. We prove that, quite remarkably, the asymptotic behavior of the estimators does not depend on ℓ under spherical symmetry, so that the affine equivariance for $\ell = d$ is achieved at no cost in terms of efficiency. To allow for large sample size n and/or large dimension d, we introduce a version of our estimators relying on incomplete U-statistics. Under a centro-symmetry assumption, we also define companion tests $\phi_n^{p,\ell}$ for the problem of testing the null hypothesis that the location μ of the underlying probability measure coincides with a given location μ_0. For any p, affine invariance is achieved for $\ell = d$. For any ℓ and p, we derive explicit expressions for the asymptotic power of these tests under contiguous local alternatives, which reveals that asymptotic relative efficiencies with respect to traditional parametric Gaussian procedures for hypothesis testing coincide with those obtained for point estimation. We illustrate finite-sample relevance of our asymptotic results through Monte Carlo exercises and also treat a real data example.

1. Introduction. Both for univariate and multivariate probability measures, location functionals are of high interest and most often are the first ones that are considered to describe a distribution. For probability measures P over the real line, classical location functionals μ_P are the L_p ones defined as

$$\mu_P^p := \operatorname*{arg\,min}_{\mu \in \mathbb{R}} \mathrm{E}\big[|X - \mu|^p\big], \tag{1.1}$$

where X has distribution P (existence and uniqueness issues are left aside in this introduction but they will be carefully discussed in the sequel). These are thus minimizing the risk associated with the L_p loss function $L_p(\theta, \mu) = |\theta - \mu|^p$. The median and the mean are obtained for $p = 1$ and $p = 2$, respectively. When a random sample from P is available, μ_P^p can be estimated by $\hat{\mu}_n^p := \arg\min_{\mu \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^n |X_i - \mu|^p$. If P admits the density (in this work, all

Received September 2021; revised February 2022.
MSC2020 subject classifications. Primary 62G05, 62G20; secondary 62H15, 60D05.
Key words and phrases. Affine equivariance/affine invariance, contiguous alternatives, L_p loss functions, multivariate location functionals, Oja median, random simplices, spatial median.

densities will be with respect to the Lebesgue measure)

$$x \mapsto \frac{c_p}{\sigma} \exp\left(-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^p\right), \tag{1.2}$$

where $\mu \in \mathbb{R}$ is a location parameter, $\sigma > 0$ is a scale parameter and c_p is a normalizing constant, then $\hat{\mu}_n^p$ is the maximum likelihood estimator (MLE) for μ. Under the assumption that P is symmetric about μ (in the sense that any Borel set B has the same P-probability as its reflection about μ), then $\mu_P^p = \mu$ for any p, which provides a framework where the estimators $\hat{\mu}_n^p$ can be compared in a sensible way; this is the case in particular when P admits the density in (1.2).

For probability measures P over $\mathbb{R}^d$, it is natural to generalize (1.1) into

$$\mu_P^p := \underset{\mu \in \mathbb{R}^d}{\arg\min}\, \mathrm{E}\big[\|X-\mu\|^p\big], \tag{1.3}$$

where X has distribution P and $\|z\| = \sqrt{z'z}$ is the Euclidean norm of the d-vector z. For $p = 1$ and $p = 2$, this provides the *spatial* (or *geometric*) median of P and the mean vector $\mathrm{E}[X]$ of P, respectively; we refer to Brown (1983) or Möttönen, Nordhausen and Oja (2010) for the spatial median, that has been much used recently in other contexts, too (see, e.g., Cardot, Cénac and Zitt (2013), Minsker (2015), and Cardot, Cénac and Godichon-Baggioni (2017)). The asymptotic behavior of the corresponding estimator $\hat{\mu}_n^p$ (that is still obtained by replacing expectations with sample averages in (1.3)) has been investigated in Konen and Paindaveine (2022). As natural as it is, the approach suffers from several deficiencies. First and foremost, for $p \neq 2$, the functional μ_P^p and its estimator $\hat{\mu}_n^p$ are not *affine-equivariant*: indeed, denoting as $P_{A,b}$ the distribution of $AX + b$ when X has distribution P, it is not so that $\mu_{P_{A,b}}^p = A\mu_P^p + b$ for any $d \times d$ invertible matrix A and any d-vector b, and similarly, $\hat{\mu}_n^p$ does not satisfy $\hat{\mu}_n^p(AX_1 + b, \dots, AX_n + b) = A\hat{\mu}_n^p(X_1, \dots, X_n) + b$ for any such A and b. Affine equivariance is a key property, both at the population level (where it sometimes even defines what a multivariate location functional is; see, e.g., Definition 3.1 in Oja (2010)) and at the sample level (where affine equivariance is often regarded as a fundamental requirement for location estimators; see, among many others, Davies (1987), Donoho and Gasko (1992), Lopuhaä (1999), Zuo (2003)). Another issue is that if P admits the elliptical extension of the power-exponential density in (1.2), namely

$$x \mapsto \frac{c_{p,d}}{\sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}\|x-\mu\|_\Sigma^p\right), \tag{1.4}$$

where $\|x-\mu\|_\Sigma = \sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}$ is the Mahalanobis distance between x and μ in the metric associated with the symmetric positive definite matrix Σ, then $\hat{\mu}_n^p$ is the MLE for μ only in the spherical submodel obtained by imposing that Σ is isotropic, that is, proportional to the d-dimensional identity matrix I_d. In a nonspherical submodel associated with $\Sigma = \lambda V_0$ for some $\lambda > 0$ (with V_0 a fixed matrix with trace d that is different from I_d), the estimator $\hat{\mu}_n^p$ not only fails to be the MLE for μ but it is not even asymptotically equivalent to this MLE, which has very negative consequences in terms of asymptotic efficiency: the less isotropic V_0 is, the poorer the corresponding asymptotic performances of $\hat{\mu}_n^p$ are (for $p = 1$, this was shown in Niinimaa and Oja (1995) and Magyar and Tyler (2011)).

In this work, we will consider alternative multivariate extensions of the objective function in (1.1), that are based on random simplices. Denote as $\mathrm{Simpl}(x_1, \dots, x_\ell, x_{\ell+1})$ the simplex with vertices $x_1, \dots, x_\ell, x_{\ell+1} \in \mathbb{R}^d$ (that is, the convex hull of these $\ell + 1$ points in $\mathbb{R}^d$), and recall that its ℓ-measure (length for $\ell = 1$, area for $\ell = 2$, etc.) is given by

$$\begin{aligned} &m_\ell\big(\mathrm{Simpl}(x_1, \dots, x_\ell, x_{\ell+1})\big) \\ &\quad = \frac{1}{\ell!}\sqrt{\det\big((x_1 - x_{\ell+1} \dots x_\ell - x_{\ell+1})'(x_1 - x_{\ell+1} \dots x_\ell - x_{\ell+1})\big)}. \end{aligned}$$

With this notation, we generalize (1.1) into

$$\mu_P^{p,\ell} := \underset{\mu\in\mathbb{R}^d}{\arg\min}\, \mathrm{E}\big[m_\ell^p\big(\mathrm{Simpl}(X_1,\ldots,X_\ell,\mu)\big)\big], \tag{1.5}$$

where $X_1,\ldots,X_\ell$ form a random sample from P; here, $p\geq 1$, which will ensure that the objective function above is convex in μ, and $\ell\in\{1,\ldots,d\}$. Clearly, for $\ell=1$, these functionals reduce to those in (1.3), that fail to be affine-equivariant for $p\neq 2$. In contrast, for $\ell=d$, affine equivariance is achieved. In the median case $p=1$, the functional $\mu_P^{p,d}$ obtained for $\ell=d$ is the Oja median; see, for example, Oja (1983), Hettmansperger, Möttönen and Oja (1997), or Ollila, Oja and Hettmansperger (2002). As we will show, for $p=2$, we have that $\mu_P^{p,\ell}$ concides with the mean vector of P, irrespective of ℓ. These considerations extend to the sample case, in which a random sample $X_1,\ldots,X_n$ from P allows us to consider the estimator

$$\hat{\mu}_n^{p,\ell} := \underset{\mu\in\mathbb{R}^d}{\arg\min}\, \frac{1}{\binom{n}{\ell}} \sum_{1\leq i_1<\cdots<i_\ell\leq n} m_\ell^p\big(\mathrm{Simpl}(X_{i_1},\ldots,X_{i_\ell},\mu)\big) \tag{1.6}$$

(again, existence and uniqueness will be discussed later). For any fixed p and ℓ, we will study the asymptotic behavior of these estimators. More precisely, under very mild assumptions (that do not even require absolute continuity of P with respect to the Lebesgue measure), we will derive Bahadur representation results and prove asymptotic normality. When observations are generated from a distribution that is spherically symmetric, our results will reveal that, unexpectedly, the asymptotic behavior of $\hat{\mu}_n^{p,\ell}$ does not depend on ℓ. As a direct corollary, when observations are randomly sampled from (1.4) with $\Sigma=\lambda I_d$ for some $\lambda>0$, the estimators $\hat{\mu}_n^{p,\ell}$, $\ell=2,\ldots,d$, inherit the asymptotic optimality properties of the MLE $\hat{\mu}_n^{p,1}$. From affine equivariance, however, $\hat{\mu}_n^{p,d}$ will enjoy the same optimality properties in the general elliptical case where Σ is totally arbitrary, whereas, as mentioned above, non-isotropic values of Σ will negatively impact the asymptotic performances of competitors that are not affine-equivariant. Affine equivariance is achieved for $\ell=d$ only, but we still consider all possible $\ell\in\{1,\ldots,d\}$ in (1.5)–(1.6), not only because this offers a unified framework allowing us to investigate the properties of $\hat{\mu}_n^{p,1}$ and $\hat{\mu}_n^{p,d}$ simultaneously, but also because using a value of ℓ that is slightly smaller than d may be advantageous on a computational point of view while losing barely anything in terms of efficiency.

Besides point estimation, we will also consider the problem of testing the null hypothesis that the underlying location μ of the probability measure P at hand is equal to a given value μ_0; to identify unambiguously μ, we will do so under the assumption that the distribution is centro-symmetric about μ. In this framework, we define a family of companion tests $\phi_n^{p,\ell}$ to the location estimators in (1.6). Parallel to the estimation problem, these tests, for $p=2$ and irrespective of ℓ, reduce to the parametric Gaussian procedure for the considered problem, here the classical Hotelling T^2 test. For $p=1$, the proposed tests reduce to the spatial sign test from Möttönen and Oja (1995) and to the affine-invariant Oja test from Hettmansperger, Nyblom and Oja (1994) for $\ell=1$ and $\ell=d$, respectively. For $\ell=d$, our tests are actually affine-invariant irrespective of p. For any fixed ℓ and p, we derive the asymptotic distribution of these tests both under the null hypothesis and under sequences of contiguous alternatives. We obtain explicit expressions for the corresponding asymptotic local powers in the spherical case and, for $\ell=d$, also in the elliptical case. Our asymptotic results reveal that, under sphericity, the asymptotic relative efficiencies of the proposed tests with respect to the Hotelling test are the same as those the proposed estimators show relative to the sample mean (for $\ell=d$, this extends to the elliptical case from affine invariance). This supports the claim that these tests may be considered the companion tests to the estimators in (1.6).

The outline of the paper is as follows. In Section 2, we focus on point estimation. In Section 2.1, we study the structural properties (finiteness, convexity, coercivity) of the objective functions in (1.5)–(1.6), and then discuss existence and unicity of the resulting location functionals $\mu_P^{p,\ell}$ and estimators $\hat{\mu}_n^{p,\ell}$. In Section 2.2, we state the exact equivariance properties of these objects and we show that, for $p = 2$, $\mu_P^{p,\ell}$ and $\hat{\mu}_n^{p,\ell}$ reduce to the population mean and sample mean irrespective of ℓ. In Section 2.3, we study the asymptotic properties of the estimators $\hat{\mu}_n^{p,\ell}$. In particular, we show that these estimators satisfy a Bahadur representation result and we provide a consistent estimator for the covariance matrix in the resulting Gaussian asymptotic distribution. Since our estimators may be computationally demanding for large sample sizes n and/or large dimensions d, we define in Section 2.4 more practical versions of these estimators based on incomplete U-statistics. In Section 3, we turn to hypothesis testing. The proposed tests are based on "(p, ℓ)-scores", associated with the gradient of the simplex-based objective functions above. In Section 3.1, we introduce these scores and present their main properties, whereas, in Section 3.2, we define the proposed tests and investigate their asymptotic behavior. In Section 4, we illustrate our results through Monte Carlo exercises, both for point estimation (Section 4.1) and for hypothesis testing (Section 4.2), and we also treat a real data example (Section 4.3). In Section 5, we provide some final comments. All proofs are collected in the Supplementary Material Dürre and Paindaveine (2022).

For the sake of convenience, we introduce here some notation. For a symmetric positive definite $d \times d$ matrix Σ, we will denote as $\Sigma^{1/2}$ the unique symmetric positive definite matrix such that $\Sigma = (\Sigma^{1/2})^2$, and we will write $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$. With this notation, $\|x - y\|_\Sigma := \|\Sigma^{-1/2}(x - y)\|$ is the Mahalanobis distance between the d-vectors x, y in the metric associated with Σ. Throughout, expectations of the form $\mathrm{E}_P[g(X_1, \dots, X_r)]$ assume that the involved random vectors form a random sample from P. If the collection B of values of $(x_1, \dots, x_r)$ for which $g(x_1, \dots, x_r)$ is undefined is nonempty (typically, because $(x_1, \dots, x_r) \in B$ leads to dividing by zero in $g(x_1, \dots, x_r)$), then $\mathrm{E}_P[g(X_1, \dots, X_r)]$ tacitly stands for $\mathrm{E}_P[g(X_1, \dots, X_r)\mathbb{I}[(X_1, \dots, X_r) \notin B]]$, where $\mathbb{I}[C]$ is the indicator function associated with condition C.

2. Simplex-based L_p location functionals.

2.1. *Definition, existence and uniqueness.* Let P be a probability measure over $\mathbb{R}^d$ and consider the objective function

$$O_P^{p,\ell}(\mu) := \mathrm{E}_P\big[m_\ell^p\big(\mathrm{Simpl}(X_1, \dots, X_\ell, \mu)\big)\big]. \tag{2.1}$$

If observations $x_1, \dots, x_n$ in $\mathbb{R}^d$ are available, then the corresponding sample objective function is

$$O_n^{p,\ell}(\mu; x_1, \dots, x_n) := \frac{1}{\binom{n}{\ell}} \sum_{1 \le i_1 < \cdots < i_\ell \le n} m_\ell^p\big(\mathrm{Simpl}(x_{i_1}, \dots, x_{i_\ell}, \mu)\big) \tag{2.2}$$

(tacitly, we obviously assume throughout that $n \ge \ell$). While the sample objective function is always well defined for any $\mu \in \mathbb{R}^d$, this is not the case of the population version (2.1), that requires suitable moment conditions. We have the following result (where we say that P admits finite moments of order $q (> 0)$ if and only if $\int_{\mathbb{R}^d} \|x\|^q \, dP(x)$ exists and is finite).

THEOREM 2.1. *Fix a real number $p \ge 1$, an integer $d \ge 2$ and $\ell \in \{1, \dots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order p. Then, $O_P^{p,\ell}(\mu)$ is well defined for any $\mu \in \mathbb{R}^d$.*

Both the population and sample objective functions above satisfy convexity and coercivity properties that will play a key role in the sequel. Recall that a map $g : \mathbb{R}^d \to \mathbb{R}$ is said to be *coercive* if and only if $(g(\mu_k)) \to \infty$ for any sequence (μ_k) in $\mathbb{R}^d$ such that $(\|\mu_k\|) \to \infty$. We then have the following result.

THEOREM 2.2. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \dots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order p. Then*:

(i$_a$) *the mapping $\mu \mapsto O_P^{p,\ell}(\mu)$ is convex over $\mathbb{R}^d$*;
(ii$_a$) *if no $(\ell - 1)$-dimensional hyperplane of $\mathbb{R}^d$ has P-probability one, then $\mu \mapsto O_P^{p,\ell}(\mu)$ is coercive.*

Fix $x_1, \dots, x_n$ in $\mathbb{R}^d$. Then:

(i$_b$) *the mapping $\mu \mapsto O_n^{p,\ell}(\mu; x_1, \dots, x_n)$ is convex over $\mathbb{R}^d$*;
(ii$_b$) *if $x_1, \dots, x_n$ are not contained in an $(\ell - 1)$-dimensional hyperplane of $\mathbb{R}^d$, then $\mu \mapsto O_n^{p,\ell}(\mu; x_1, \dots, x_n)$ is coercive.*

As explained in the Introduction, suitable simplex-based L_p location functionals[1] are obtained by minimizing the objective functions above, which leads to considering

$$\mu_P^{p,\ell} := \arg\min_{\mu \in \mathbb{R}^d} O_P^{p,\ell}(\mu) \quad \text{and} \quad \mu_n^{p,\ell}(x_1, \dots, x_n) := \arg\min_{\mu \in \mathbb{R}^d} O_n^{p,\ell}(\mu; x_1, \dots, x_n).$$

In principle, neither existence nor uniqueness is guaranteed. This calls for the following result, where it is made precise when and how $\mu_P^{p,\ell}$ and $\mu_n^{p,\ell}(x_1, \dots, x_n)$ are defined.

THEOREM 2.3. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \dots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order p. Then*:

(i$_a$) *the set of minimizers of the map $\mu \mapsto O_P^{p,\ell}(\mu)$ is nonempty and convex*;
(ii$_a$) *if no $(\ell - 1)$-dimensional hyperplane of $\mathbb{R}^d$ has P-probability one, then this set of minimizer is also bounded, so that $\mu_P^{p,\ell}$, which we then define as the barycentre of this set, is a well-defined minimizer itself.*

Fix $x_1, \dots, x_n$ in $\mathbb{R}^d$. Then:

(i$_b$) *the set of minimizers of the map $\mu \mapsto O_n^{p,\ell}(\mu; x_1, \dots, x_n)$ is nonempty and convex*;
(ii$_b$) *if $x_1, \dots, x_n$ are not contained in an $(\ell - 1)$-dimensional hyperplane of $\mathbb{R}^d$, then this set of minimizer is also bounded, so that $\mu_n^{p,\ell}(x_1, \dots, x_n)$, which we then define as the barycentre of this set, is a well-defined minimizer itself.*

This theorem shows that, both in the population and sample cases, a minimizer always exists and that a unique representative of the set of minimizers, namely $\mu_P^{p,\ell}$ in the population case or $\mu_n^{p,\ell}(x_1, \dots, x_n)$ in the sample one, can be identified under the extremely mild assumption that the distribution P or the observations $x_1, \dots, x_n$ do not fully concentrate in an $(\ell - 1)$-dimensional hyperplane. If this assumption is not fulfilled, then it can be argued that the problem is at most an $(\ell - 1)$-dimensional one (rather than a d-dimensional one), hence should be solved by using simplices of dimension at most $\ell - 1$ in an appropriate reparametrization of this supporting hyperplane.

---

[1]In the sequel, we often use the term *functional* in the sample case, too, which amounts to considering $\mu_n^{p,\ell}(x_1, \dots, x_n)$ as a function of the empirical probability measure P_n associated with $x_1, \dots, x_n$. Incidentally, we stress that $\mu_n^{p,\ell}(x_1, \dots, x_n)$ is *not* obtained by evaluating the population functional $\mu_P^{p,\ell}$ at P_n, which, actually justifies that several theorems in this section explicitly discriminate between the population and sample cases.

2.2. *Equivariance properties.* We start by considering equivariance properties of the population functionals $\mu_P^{p,\ell}$. In this end, fix an integer $d \geq 2$, $\ell \in \{1, \ldots, d\}$ and a probability measure P over $\mathbb{R}^d$ that admits finite moments of order p. For the sake of simplicity, we further assume that no $(\ell - 1)$-dimensional hyperplane of $\mathbb{R}^d$ has P-probability one, so that the population functional $\mu_P^{p,\ell}$ is well defined (without this assumption, one could still state equivariance properties for the set of minimizers of the objective function $\mu \mapsto O_P^{p,\ell}(\mu)$). It can be trivially checked that the functional $\mu_P^{p,\ell}$ is then equivariant under translations, under homothetic transformations, and under orthogonal transformations in the following sense: for any d-vector b, any $\lambda > 0$ and any $d \times d$ orthogonal matrix O,

$$\mu_{P_{O,\lambda,b}}^{p,\ell} = \lambda O \mu_P^{p,\ell} + b, \tag{2.3}$$

where $P_{O,\lambda,b}$ denotes the distribution of $\lambda O X + b$ when X has distribution P. This extends to the sample case where the equivariance relation reads $\mu_n^{p,\ell}(\lambda O x_1 + b, \ldots, \lambda O x_n + b) = \lambda O \mu_n^{p,\ell}(x_1, \ldots, x_n) + b$ for any d-vectors $x_1, \ldots, x_n$ and for any O, λ, b as above.

As mentioned in the Introduction, however, it is desirable that multivariate location functionals are also equivariant under general affine transformations. Since the volume of a full-dimensional simplex transforms as

$$m_d\big(\mathrm{Simpl}(Ax_1 + b, \ldots, Ax_d + b, A\mu + b)\big) = |\det A| m_d\big(\mathrm{Simpl}(x_1, \ldots, x_d, \mu)\big)$$

for any d-vectors $x_1, \ldots, x_d, \mu, b$ and any $d \times d$ matrix A, the $(\ell = d)$-version of our location functionals are affine equivariant, in the sense that for any invertible $d \times d$ matrix A and any d-vector b,

$$\mu_{P_{A,b}}^{p,\ell} = A \mu_P^{p,\ell} + b, \tag{2.4}$$

where $P_{A,b}$ stands for the distribution of $AX + b$ when X has distribution P. In the sample case, this of course translates into $\mu_n^{p,\ell}(Ax_1 + b, \ldots, Ax_n + b) = A\mu_n^{p,\ell}(x_1, \ldots, x_n) + b$ for any d-vectors $x_1, \ldots, x_n$ and for any A, b as above, which is the sample affine-equivariance property defined in the Introduction.

Both in the population and sample cases, location functionals associated with lower-dimensional simplices (i.e., $\ell < d$) fail to be affine-equivariant in general. A notable exception is the case $p = 2$, for which the corresponding location functionals are affine-equivariant for any $\ell \in \{1, \ldots, d\}$, as we now explain. In the univariate case $d = 1$, only $\ell = 1$ may be considered and, as recalled in the Introduction, the objective functions

$$O_P^{2,\ell} = \int_{\mathbb{R}} (x - \mu)^2 \, dP(x) \quad \text{and} \quad O_n^{2,\ell}(\mu; x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

are then uniquely minimized at the expectation of P and at the sample mean of $x_1, \ldots, x_n$, respectively. Remarkably, this extends to the multivariate case $d > 1$, irrespective of ℓ. We have the following result.

THEOREM 2.4. *Fix an integer $d \geq 1$ and $\ell \in \{1, \ldots, d\}$.*

(a) *Let P be a probability measure over $\mathbb{R}^d$ that admits finite second-order moments. Then,*

$$\mu_P^{2,\ell} = \int_{\mathbb{R}^d} x \, dP(x)$$

is the unique minimizer of $\mu \mapsto O_P^{2,\ell}(\mu)$.

(b) *Fix* $x_1, \dots, x_n$ *in* $\mathbb{R}^d$. *Then, the sample mean*

$$\mu_n^{2,\ell}(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i$$

is the unique minimizer of $\mu \mapsto O_n^{2,\ell}(\mu; x_1, \dots, x_n)$.

Of course, this explains why, irrespective of ℓ, the population and sample location functionals associated with $p = 2$ are affine-equivariant. The proof of Theorem 2.4(b) is based on the classical Cauchy–Binet formula stating that, if A is an $\ell \times d$ real matrix and B is a $d \times \ell$ real matrix with $\ell \leq d$, then

$$\det(AB) = \sum_{1 \leq j_1 < \cdots < j_\ell \leq d} \det\big((A')_{j_1,\dots,j_\ell}\big) \det(B_{j_1,\dots,j_\ell}),$$

where $C_{j_1,\dots,j_\ell}$ is the matrix obtained by stacking the rows $j_1, \dots, j_\ell$ of C on top of each other. Interestingly, the proof of Theorem 2.4(a) relies on the following stochastic version of the Cauchy–Binet formula, which is of independent interest.

PROPOSITION 2.1. *Let P be a probability measure over $\mathbb{R}^d$ and $F, G : \mathbb{R}^d \to \mathbb{R}^\ell$ be functions such that* $\mathrm{E}[\|F(X)\|^2]$ *and* $\mathrm{E}[\|G(X)\|^2]$ *exist and are finite, where X is a random d-vector with distribution P. Then*

$$\det\big(\mathrm{E}\big[F(X)\big(G(X)\big)'\big]\big) = \frac{1}{\ell!}\mathrm{E}\big[\det\big(F(X_1)\dots F(X_\ell)\big)\det\big(G(X_1)\dots G(X_\ell)\big)\big], \tag{2.5}$$

where $X_1, \dots, X_\ell$ are independent copies of X.

2.3. *Asymptotic properties.* Assume now that a random sample $X_1, \dots, X_n$ from P is available. Obviously, a natural estimator of $\mu_P^{p,\ell}$ is then

$$\hat{\mu}_n^{p,\ell} := \mu_n^{p,\ell}(X_1, \dots, X_n).$$

To make sure that $\hat{\mu}_n^{p,\ell}$ is well defined with probability one, we will assume throughout that P is $(\ell - 1)$-smooth, in the sense that any $(\ell - 1)$-dimensional hyperplane has P-probability zero. Under this assumption, it can indeed be shown that the probability that $X_1, \dots, X_n$, with $n > \ell$, belong to an $(\ell - 1)$-dimensional hyperplane is zero (for the sake of completeness, we prove this in Lemma S.1.3), so that, from Theorem 2.3(ii$_b$), $\hat{\mu}_n^{p,\ell}$ is indeed well defined, with probability one, as the barycentre of the set of minimizers of the map

$$\mu \mapsto \widehat{O}_n^{p,\ell}(\mu) := O_n^{p,\ell}(\mu; X_1, \dots, X_n).$$

Our primary interest in this setup is to investigate the asymptotic behavior of $\hat{\mu}_n^{p,\ell}$. To describe this asymptotic behavior, we need to introduce the following notation. If there is a unique $(\ell - 1)$-dimensional hyperplane in $\mathbb{R}^d$ containing $x_1, \dots, x_\ell$, we will denote as $\Gamma_{x_1,\dots,x_\ell}$ the matrix of the orthogonal projection onto the $((d - \ell + 1)$-dimensional) orthogonal complement to the vector space spanned by $x_1 - x_\ell, \dots, x_{\ell-1} - x_\ell$; otherwise, we put $\Gamma_{x_1,\dots,x_\ell} = 0$. Since this definition requires that $\ell > 1$, we simply let $\Gamma_x := I_d$ for any $x \in \mathbb{R}^d$. We also put $m_0(\mathrm{Simpl}(x)) := 1$ for any $x \in \mathbb{R}^d$. We then have the following result.

THEOREM 2.5. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \dots, d\}$, with $(p, \ell) \neq (1, d)$. Let P be a probability measure over $\mathbb{R}^d$ that is $(\ell - 1)$-smooth and admits finite moments of order $2p$. If $p = 1$, then assume that no ℓ-dimensional hyperplane containing $\mu_P^{p,\ell}$ has P-probability one. Finally, if $1 \leq p < 2$, assume further that*

$$\mathrm{E}_P\bigg[\frac{m_{\ell-1}^p(\mathrm{Simpl}(X_1, \dots, X_\ell))}{\|\Gamma_{X_1,\dots,X_\ell}(X_\ell - \mu_P^{p,\ell})\|^{2-p}}\bigg] < \infty. \tag{2.6}$$

Then, with $\hat{\mu}_n^{p,\ell}$ based on a random sample $X_1, \dots, X_n$ from P,

$$\sqrt{n}\big(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell}\big) = H_P^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} T_P(X_i) + o_{\mathrm{P}}(1) \tag{2.7}$$

$$\to_{\mathcal{D}} \mathcal{N}_d\big(0, H_P^{-1} \mathrm{E}_P\big[T_P(X_1) T_P'(X_1)\big] H_P^{-1}\big) \tag{2.8}$$

as n diverges to infinity, where

$$T_P(x) := \frac{p}{\ell^{p-1}} \mathrm{E}_P\bigg[m_{\ell-1}^p\big(\mathrm{Simpl}(X_1, \dots, X_{\ell-1}, x)\big) \frac{\Gamma_{X_1,\dots,X_{\ell-1},x}(x - \mu_P^{p,\ell})}{\|\Gamma_{X_1,\dots,X_{\ell-1},x}(x - \mu_P^{p,\ell})\|^{2-p}}\bigg]$$

is such that $\mathrm{E}_P[T_P(X_1) T_P'(X_1)]$ exists and is finite, and where

$$\begin{aligned} H_P = {} & \frac{p(p-1)}{\ell^p} \mathrm{E}_P\bigg[\frac{m_{\ell-1}^p(\mathrm{Simpl}(X_1, \dots, X_\ell))}{\|\Gamma(\mu_P^{p,\ell} - X_\ell)\|^{2-p}} \frac{\Gamma(\mu_P^{p,\ell} - X_\ell)(\mu_P^{p,\ell} - X_\ell)'\Gamma}{\|\Gamma(\mu_P^{p,\ell} - X_\ell)\|^2}\bigg] \\ & + \frac{p}{\ell^p} \\ & \times \mathrm{E}_P\bigg[\frac{m_{\ell-1}^p(\mathrm{Simpl}(X_1, \dots, X_\ell))}{\|\Gamma(\mu_P^{p,\ell} - X_\ell)\|^{2-p}} \bigg(\Gamma - \frac{\Gamma(\mu_P^{p,\ell} - X_\ell)(\mu_P^{p,\ell} - X_\ell)'\Gamma}{\|\Gamma(\mu_P^{p,\ell} - X_\ell)\|^2}\bigg)\bigg] \end{aligned} \tag{2.9}$$

exists, is finite, and is invertible (in (2.9), we let $\Gamma := \Gamma_{X_1,\dots,X_\ell}$ to simplify the notation).

Before discussing this result and presenting some of its corollaries, we explain how the result can be used to construct confidence zones for $\mu_P^{p,\ell}$. Assuming that $\hat{C}_n$ is a weakly consistent estimator of the covariance matrix $C_P := H_P^{-1} \mathrm{E}_P[T_P(X_1) T_P'(X_1)] H_P^{-1}$ in the asymptotic normal distribution of $\hat{\mu}_n^{p,\ell}$ in (2.8), a confidence zone for $\hat{\mu}_n^{p,\ell}$ at asymptotic confidence level $1 - \alpha (\in (0, 1))$ is given by the hyper-ellipsoid

$$\big\{\mu \in \mathbb{R}^d : n\big(\mu - \hat{\mu}_n^{p,\ell}\big)' \hat{C}_n^{-1} \big(\mu - \hat{\mu}_n^{p,\ell}\big) \le \chi^2_{d,1-\alpha}\big\}, \tag{2.10}$$

where $\chi^2_{d,1-\alpha}$ denotes the upper α-quantile of the chi-square distribution with d degrees of freedom. To describe the required estimator $\hat{C}_n$, we introduce the following notation. For any positive integer k, we let $\mathcal{I}_n^k := \{I = (i_1, \dots, i_k) \in \mathbb{N}^k : 1 \le i_1 < \dots < i_k \le n\}$. For $I \in \mathcal{I}_n^\ell$, we will write $m_I := m_{\ell-1}(\mathrm{Simpl}(X_{i_1}, \dots, X_{i_\ell}))$ and $\Gamma_I := \Gamma_{X_{i_1},\dots,X_{i_\ell}}$. We then have the following result.

THEOREM 2.6. *Fix a real number $p \ge 1$, an integer $d \ge 2$ and $\ell \in \{1, \dots, d\}$, with $(p, \ell) \ne (1, d)$. Let $X_1, \dots, X_n$ be a random sample from a probability measure P over $\mathbb{R}^d$ satisfying the assumptions in Theorem 2.5. For $p = 1$, reinforce the assumption that P admits finite moments of order 2 into the assumption that it has finite moments of order $2 + \eta$ for some $\eta > 0$, and for $1 \le p < 2$, reinforce (2.6) into the assumption that*

$$\mathrm{E}_P\bigg[\frac{m_{\ell-1}^p(\mathrm{Simpl}(X_1, \dots, X_\ell))}{\|\Gamma_{X_1,\dots,X_\ell}(X_\ell - \mu_P^{p,\ell})\|^{2-p+\delta}}\bigg] < \infty \tag{2.11}$$

for some $\delta > 0$. Let $\gamma_{n,I} := \mathbb{I}[\|\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})\| > c_n]$ for $p \in [1, 2)$ and $\gamma_{n,I} := 1$ for $p \ge 2$, where c_n is a positive real sequence such that $c_n \to 0$ and $\sqrt{n} c_n \to \infty$, and define

$$\hat{T}_n(x) := \frac{p}{\ell^{p-1} \binom{n}{\ell-1}} \sum_{I \in \mathcal{I}_n^{\ell-1}} m_{\ell-1}^p\big(\mathrm{Simpl}(X_{i_1}, \dots, X_{i_{\ell-1}}, x)\big) \times \frac{\Gamma_{X_{i_1},\dots,X_{i_{\ell-1}},x}(x - \hat{\mu}_n^{p,\ell})}{\|\Gamma_{X_{i_1},\dots,X_{i_{\ell-1}},x}(x - \hat{\mu}_n^{p,\ell})\|^{2-p}}.$$

Then, as n diverges to infinity, $\hat{A}_n := \frac{1}{n}\sum_{i=1}^n \hat{T}_n(X_i)(\hat{T}_n(X_i))'$ *and*

$$\hat{H}_n = \frac{p(p-1)}{\ell^p\binom{n}{\ell}} \sum_{I \in \mathcal{I}_n^\ell} \gamma_{n,I} \frac{m_{\ell-1}^p(\mathrm{Simpl}(X_I))}{\|\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})\|^{2-p}} \frac{\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})(\hat{\mu}_n^{p,\ell} - X_{i_1})'\Gamma_I}{\|\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})\|^2}$$
$$+ \frac{p}{\ell^p\binom{n}{\ell}} \sum_{I \in \mathcal{I}_n^\ell} \gamma_{n,I} \frac{m_{\ell-1}^p(\mathrm{Simpl}(X_I))}{\|\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})\|^{2-p}}$$
$$\times \left(\Gamma_I - \frac{\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})(\hat{\mu}_n^{p,\ell} - X_{i_1})'\Gamma_I}{\|\Gamma_I(\hat{\mu}_n^{p,\ell} - X_{i_1})\|^2}\right)$$

converge in probability to $A_P := \mathrm{E}_P[T_P(X_1)T_P'(X_1)]$ *and* H_P, *respectively, so that* $\hat{C}_n := \hat{H}_n^{-1}\hat{A}_n\hat{H}_n^{-1}$ *is a weakly consistent estimator of* $C_P = H_P^{-1}\mathrm{E}_P[T_P(X_1)T_P'(X_1)]H_P^{-1}$.

This result shows that confidence zones for $\mu_P^{p,\ell}$ can be constructed under an extremely slight reinforcement of the assumptions from Theorem 2.5: we will show in Lemma S.2.9 that, under the moment assumptions considered in this theorem, Assumption (2.11) indeed only slightly reinforces (2.6). Also, the reinforcement in moment assumptions (which is in any case very mild, since it replaces finite second-order moments with finite moments of order $2+\eta$ for some $\eta > 0$) is imposed for $p = 1$ only, hence is not needed for the case $p > 1$ we primarily focus on in this work.

Under suitable symmetry assumptions on P, the population functionals $\mu_P^{p,\ell}$ coincide for any p and ℓ; for instance, if P is centro-symmetric about μ (in the sense that for any Borel set B of $\mathbb{R}^d$, B and its reflection about μ share the same P-probability), then the equivariance relation (2.3) entails that $\mu_P^{p,\ell} = \mu$, irrespective of p and ℓ. In such cases, the sample location functionals $\hat{\mu}_n^{p,\ell}$ all estimate the same quantity and it is desirable to compare them, for example, in terms of efficiency. While Theorem 2.5 is a very general result, it does not make such a comparison straightforward. This is an important motivation to derive the following result, that provides the key quantities needed to perform this comparison under sphericity, hence, for affine-equivariant estimators, also under ellipticity. We recall that the probability measure P is spherically symmetric about the origin of $\mathbb{R}^d$ if and only if for any Borel set B of $\mathbb{R}^d$, OB and B share the same P-probability for any $d \times d$ orthogonal matrix O (elliptically symmetric probability measures are then obtained by transforming spherically symmetric probability measures in an affine way).

THEOREM 2.7. *Fix a real number* $p \geq 1$, *an integer* $d \geq 2$ *and* $\ell \in \{1, \ldots, d\}$, *with* $(p, \ell) \neq (1, d)$. *Let* P *be a probability measure over* $\mathbb{R}^d$ *that admits finite moments of order* p, *is spherically symmetric about the origin of* $\mathbb{R}^d$, *and satisfies* $P[\{0\}] = 0$. *If* $1 \leq p < 2$, *then assume further that* (2.6) *holds. Then,* $\mu_P^{p,\ell} = 0$,

$$T_P(x) = \frac{p\Gamma(\frac{d-\ell+p+1}{2})\Gamma(\frac{d}{2})}{\ell^{p-1}\Gamma(\frac{d-\ell+1}{2})\Gamma(\frac{d+p}{2})}\mathrm{E}_P\big[m_{\ell-1}^p\big(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, 0)\big)\big]\frac{x}{\|x\|^{2-p}}$$

and

$$H_P = \frac{p(d+p-2)\Gamma(\frac{d-\ell+p+1}{2})\Gamma(\frac{d}{2})}{d\ell^{p-1}\Gamma(\frac{d-\ell+1}{2})\Gamma(\frac{d+p}{2})}\mathrm{E}_P\big[m_{\ell-1}^p\big(\mathrm{Simpl}(X_1, \ldots, X_{\ell-1}, 0)\big)\big]\mathrm{E}_P\big[\|X_1\|^{p-2}\big]I_d,$$

where Γ *is the Euler Gamma function.*

The proof of this result is extremely long and involved, and it requires original results from stochastic geometry. For spherical Gaussian distributions, an explicit expression of H_P (agreeing with the one in Theorem 2.7) can be obtained from the fact that H_P is actually the Hessian matrix of the map $\mu \mapsto O_P^{p,\ell} = \mathrm{E}_P[m_\ell^p(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))]$ at $\mu_P^{p,\ell} = 0$ (this is established in the proof of Theorem 2.5) and by using the distribution of $m_\ell(\mathrm{Simpl}(X_1, \ldots, X_\ell, \mu))$ in the standard normal case; see Theorem 1.1 in Paindaveine (2022). Since this distribution remains unknown away from the spherical Gaussian case, this strategy fails to provide an expression for H_P under other spherical distributions, which makes Theorem 2.7 important. In passing, we mention that the asymptotic distribution of $\hat{\mu}^{1,\ell}$, $\ell \in \{1, \ldots, d-1\}$ was investigated in Paindaveine (2022). But despite the fact that it focuses on the case $p = 1$, this earlier work excluded the key case $\ell = d$ yielding affine-equivariant estimators and, as far as sphericity is concerned, restricted to spherical Gaussian distributions. The investigation for an arbitrary value of p in the present work is thus much more extensive.

Now, using Theorem 2.7 in Theorem 2.5 and exploiting the equivariance properties from the previous subsection, we obtain the following corollary.

COROLLARY 2.1. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \ldots, d\}$, with $(p, \ell) \neq (1, d)$. Fix a d-vector μ and a symmetric positive definite $d \times d$ matrix Σ. Let P_0 be a probability measure over $\mathbb{R}^d$ that admits finite moments of order $2p$, is spherically symmetric about the origin of $\mathbb{R}^d$, and satisfies $P_0[\{0\}] = 0$. If $1 \leq p < 2$, then assume further that (2.6) holds for P_0. Let P be the distribution of $X = \Sigma^{1/2} Z + \mu$, where Z has distribution P_0. Then,* (i) *for $\ell \in \{1, \ldots, d-1\}$, $p \neq 2$ and $\Sigma = \sigma^2 I_d$ with $\sigma^2 > 0$,*

$$\sqrt{n}(\hat{\mu}_n^{p,\ell} - \mu) = \sqrt{n}(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell})$$

$$= \frac{d}{(d+p-2)\mathrm{E}_P[\|X_1 - \mu\|_\Sigma^{p-2}]\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\|X_i - \mu\|_\Sigma^{2-p}} + o_{\mathrm{P}}(1) \tag{2.12}$$

$$\to_{\mathcal{D}} \mathcal{N}_d\left(0, \frac{d\mathrm{E}_P[\|X_1 - \mu\|_\Sigma^{2(p-1)}]}{(d+p-2)^2(\mathrm{E}_P[\|X_1 - \mu\|_\Sigma^{p-2}])^2}\Sigma\right) \tag{2.13}$$

as n diverges to infinity; (ii) *for $\ell = d$ or $p = 2$, the same holds without any constraint on Σ.*

Interestingly, this shows that, in spherical cases where $\Sigma = \sigma^2 I_d$, the asymptotic behavior of $\hat{\mu}_n^{p,\ell}$ does not depend on ℓ. For the spatial location functionals associated with $\ell = 1$, the results in (2.12)–(2.13) reduce to the ones recently obtained in Section 8 of Konen and Paindaveine (2022). When the parent spherical distribution P_0 is t with $\nu(> 2p)$ degrees of freedom (i.e., has a density proportional to $x \mapsto (1 + \|x\|^2/\nu)^{-(d+\nu)/2}$), the asymptotic covariance matrix in (2.13) is

$$\frac{\Gamma(\frac{d+2}{2})\Gamma(\frac{d+2p-2}{2})\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu-2p+2}{2})}{\Gamma^2(\frac{d+p}{2})\Gamma^2(\frac{\nu-p+2}{2})}\Sigma,$$

whereas when it is power-exponential with parameter $\eta(> 0)$ degrees of freedom (i.e., has a density proportional to $x \mapsto \exp(-\|x\|^{2\eta}/2)$), this covariance matrix is

$$\frac{2^{(1-\eta)/\eta}\Gamma(\frac{d+2\eta}{2\eta})\Gamma(\frac{d+2p-2}{2\eta})}{\eta\Gamma^2(\frac{d+p+2\eta-2}{2\eta})}\Sigma$$

(the Gaussian case is of course obtained with $\eta = 1$, or as $\nu \to \infty$). Under strict ellipticity, that is when Σ is not proportional to I_d, the asymptotic behavior of $\hat{\mu}_n^{p,\ell}$ does depend on

ℓ, and in an adverse way for location functionals that are not affine-equivariant: as we will show in Section 4, the asymptotic accuracy of $\hat{\mu}_n^{p,\ell}$ with $\ell \in \{1, \dots, d-1\}$ is, under strict ellipticity and for any fixed p, dominated by that of the affine-equivariant estimator $\hat{\mu}_n^{p,d}$—unless, of course, $p = 2$ since $\hat{\mu}_n^{2,\ell} = \bar{X}_n$ irrespective of ℓ (see Theorem 2.4). In contrast, the affine-equivariant estimators $\hat{\mu}_n^{p,d}$ show performances that are not negatively affected by the possible nonisotropic value of Σ. This is also in line with the fact that it results from (2.12) that, when observations are randomly sampled from (1.4), $\hat{\mu}_n^{p,d}$ is asymptotically equivalent to the fixed-Σ MLE for μ (since $\hat{\mu}_n^{p,d}$ does not use the unknown fixed value of Σ, this estimator is thus adaptive with respect to Σ).

Finally, we comment on robustness aspects. It directly follows from (2.12) that, in the spherical case considered in Corollary 2.1, the influence function of $\mu_P^{p,\ell}$ at x is given by

$$\begin{aligned}\mathrm{IF}(x; \mu_P^{p,\ell}) &:= \lim_{\varepsilon \overset{>}{\to} 0} \frac{\mu_{P_{x,\varepsilon}}^{p,\ell} - \mu_P^{p,\ell}}{\varepsilon} \\ &= \frac{d\|x-\mu\|_\Sigma^{p-1}}{(d+p-2)\mathrm{E}_P[\|X_1-\mu\|_\Sigma^{p-2}]} \Sigma^{1/2} U_{\mu,\Sigma}(x),\end{aligned}$$

where $P_{x,\varepsilon} := (1-\varepsilon)P + \varepsilon\delta_x$ involves the Dirac probability measure δ_x at x and where $U_{\mu,\Sigma}(x) := \Sigma^{-1/2}(x-\mu)/\|x-\mu\|_\Sigma$ is a unit vector (from affine equivariance, this extends to an arbitrary matrix Σ for $\ell = d$). This shows that a bounded influence function is obtained for $p = 1$ only, and that the larger p is, the faster the norm of influence function increases to infinity as $\|x\|$ does. In this sense, the robustness of the proposed estimators decreases with p. In particular, these estimators are more robust (resp., less robust) than the sample mean for $p < 2$ (resp., for $p > 2$).

2.4. *A computationally efficient version.* Recalling the notation $\mathcal{I}_n^k$ introduced above Theorem 2.6, the location estimator considered in the previous sections is

$$\hat{\mu}_n^{p,\ell} := \arg\min_{\mu \in \mathbb{R}^d} \frac{1}{\binom{n}{\ell}} \sum_{(i_1,\dots,i_\ell)\in\mathcal{I}_n^\ell} m_\ell^p\big(\mathrm{Simpl}(X_{i_1},\dots,X_{i_\ell},\mu)\big), \tag{2.14}$$

where the argmin identifies the barycentre of the set of minimizers. As we showed, this estimator has many nice properties, particularly so for the case $\ell = d$ where affine equivariance is achieved. For large sample sizes n and/or large dimensions d, however, the resulting affine-equivariant estimator is hard to compute since it is defined as the minimizer of an objective function with $O(n^d)$ terms (more generally, for any fixed value of ℓ, the objective function has $O(n^\ell)$ terms, which is problematic unless ℓ is very small). A natural way[2] to improve on this, for a generic value of ℓ, relies on the concept of *incomplete U-statistics* (see, e.g., Blom (1976) and Enqvist (1978)) and consists in rather considering

$$\hat{\mu}_{n,N}^{p,\ell} := \arg\min_{\mu \in \mathbb{R}^d} \frac{1}{N} \sum_{(i_1,\dots,i_\ell)\in\mathcal{I}_{n,N}^\ell} m_\ell^p\big(\mathrm{Simpl}(X_{i_1},\dots,X_{i_\ell},\mu)\big), \tag{2.15}$$

where $\mathcal{I}_{n,N}^\ell$ is a subset of $\mathcal{I}_n^\ell$ that is obtained in one of the following ways:

- *sampling without replacement*: $\mathcal{I}_{n,N}^\ell$ results from sampling N multi-indices $(i_1,\dots,i_\ell)$ without replacement from $\mathcal{I}_n^\ell$ (N is then an integer between 1 and $\binom{n}{\ell}$);

[2]Yet original way, as it has not been proposed, for example, for the Oja median obtained for $p = 1$ and $\ell = d$.

- *sampling with replacement*: $\mathcal{I}_{n,N}^{\ell}$ results from sampling N multi-indices $(i_1, \dots, i_\ell)$ with replacement from $\mathcal{I}_n^{\ell}$ (N is then an integer between 1 and $\binom{n}{\ell}$);
- *Bernoulli sampling*: each multi-index $(i_1, \dots, i_\ell)$ from $\mathcal{I}_n^{\ell}$ is included in $\mathcal{I}_{n,N}^{\ell}$ (independently of the other multi-indices) with probability $p_n = N/\binom{n}{\ell}$ (here, N is a real number that is strictly between 0 and $\binom{n}{\ell}$, and the number of multi-indices in $\mathcal{I}_{n,N}^{\ell}$ is random with expectation N).

By exploiting asymptotic results for incomplete U-statistics (namely, the asymptotic equivalence results from Janson (1984) as well as a recent consistency result from Dürre and Paindaveine (2021)), one can establish the following result.

THEOREM 2.8. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \dots, d\}$, with $(p, \ell) \neq (1, d)$. Let $N = N_n$ be a sequence such that $N/n \to \infty$ as n diverges to infinity, and denote as $\hat{\mu}_{n,N}^{p,\ell}$ the estimator in* (2.15) *obtained from any of the three sampling schemes above*. *Then*, *under the assumptions of Theorem* 2.5,

$$\sqrt{n}\big(\hat{\mu}_{n,N}^{p,\ell} - \mu_P^{p,\ell}\big) = \sqrt{n}\big(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell}\big) + o_{\mathrm{P}}(1)$$

as n diverges to infinity.

This result shows that, as soon as $N/n \to \infty$, the incomplete estimators $\hat{\mu}_{n,N}^{p,\ell}$ show the exact same asymptotic behavior as their complete antecedents $\hat{\mu}_n^{p,\ell}$. In particular, $\hat{\mu}_{n,N}^{p,\ell}$ then admits the Bahadur representation result in (2.7) and asymptotic normality result in (2.8). Remarkably, these excellent asymptotic properties are achieved also by incomplete estimators minimizing objective functions that involve only $O(n^{1+\delta})$ terms for any $\delta > 0$ (or even $O(n \log n)$) rather than $O(n^d)$ terms, which has obvious computational advantages. Finite-sample performances of these estimators will be explored in Section 4.

3. Hypothesis testing. Let P be a probability measure over $\mathbb{R}^d$. To identify unambiguously a location parameter μ, we will throughout assume that P is centro-symmetric about μ (which is a common assumption in multivariate nonparametric statistics). Based on a random sample $X_1, \dots, X_n$ from P, we then consider the fundamental problem of testing the null hypothesis $\mathcal{H}_0 : \mu = \mu_0$ against the alternative hypothesis $\mathcal{H}_1 : \mu \neq \mu_0$, where μ_0 is a fixed d-vector. In this section, we propose and study tests based on scores associated with the simplex-based L_p objective functions we considered above. We first introduce those scores.

3.1. *Multivariate L_p scores.* For any $p \geq 1$, we will define a family of L_p score functions $S_P^{p,\ell}(x; \mu)$, $\ell \in \{1, \dots, d\}$. For $p = 1$, these provide *multivariate signs* (involving x only through its direction from μ), whereas, for $p = 2$, they are linear functions of $x - \mu$. Let us start with $p = 1$. In the multivariate case, the sign of x with respect to μ can be defined in several ways. Among the most classical solutions, one can find the *spatial signs*

$$S^{\mathrm{Spatial}}(x; \mu) = \nabla_x m_1\big(\mathrm{Simpl}(x, \mu)\big) = \nabla_x \|x - \mu\| = \frac{x - \mu}{\|x - \mu\|};$$

see, for example Möttönen, Oja and Tienari (1997) or Oja (2010) (derivation with respect to the parameter μ would be more natural to define a score function, yet is, in a location model, essentially equivalent to derivation with respect to x, on which we will focus in the sequel). Spatial signs behave well under orthogonal transformations but not under affine transformations, which was the motivation to introduce the affine-equivariant *Oja signs*

$$S_P^{\mathrm{Oja}}(x; \mu) = \nabla_x \mathrm{E}_P\big[m_d\big(\mathrm{Simpl}(X_1, \dots, X_{d-1}, x, \mu)\big)\big],$$

where $X_1, \dots, X_{d-1}$ form a random sample from P; see, for example, Hettmansperger, Nyblom and Oja (1994), Oja (1999) or Ollila, Oja and Hettmansperger (2002). In the univariate case $d = 1$, both concepts reduce to the usual sign of x with respect to μ, that is, to the sign of $x - \mu$. In view of what was done for point estimation in the previous section, it is natural to consider the class of ℓ-signs

$$S_P^{\ell}(x;\mu) = \nabla_x \mathrm{E}_P\big[m_\ell\big(\mathrm{Simpl}(X_1, \dots, X_{\ell-1}, x, \mu)\big)\big]$$

and, more generally, their L_p extension

$$\begin{aligned} S_P^{p,\ell}(x;\mu) &= \nabla_x \mathrm{E}_P\big[m_\ell^p\big(\mathrm{Simpl}(X_1, \dots, X_{\ell-1}, x, \mu)\big)\big] \\ &= \frac{1}{\ell^p} \nabla_x \mathrm{E}_P\big[m_{\ell-1}^p\big(\mathrm{Simpl}(X_1, \dots, X_{\ell-1}, \mu)\big) \|\Psi(x-\mu)\|^p\big] \qquad (3.1) \\ &= \frac{p}{\ell^p} \mathrm{E}_P\bigg[m_{\ell-1}^p\big(\mathrm{Simpl}(X_1, \dots, X_{\ell-1}, \mu)\big) \frac{\Psi(x-\mu)}{\|\Psi(x-\mu)\|^{2-p}}\bigg], \qquad (3.2) \end{aligned}$$

where we applied Lemma S.1.1 and where Ψ denotes the matrix of the orthogonal projection onto the orthogonal complement to the vector space spanned by $X_1 - \mu, \dots, X_{\ell-1} - \mu$. To avoid imposing the assumptions needed to differentiate under the expectation sign in (3.1), we adopt (3.2) as a definition of L_p scores.

DEFINITION 3.1. Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \dots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order p. Then, the population (p, ℓ)-score of x with respect to μ at P is defined as

$$S_P^{p,\ell}(x;\mu) = \frac{p}{\ell^p} \mathrm{E}_P\bigg[m_{\ell-1}^p\big(\mathrm{Simpl}(X_1, \dots, X_{\ell-1}, \mu)\big) \frac{\Psi(x-\mu)}{\|\Psi(x-\mu)\|^{2-p}}\bigg]$$

(throughout, we let $x/\|x\| := 0$ for $x = 0$). If a random sample $X_1, \dots, X_n$ from P is available, then we define the sample (p, ℓ)-score of x with respect to μ as

$$\begin{aligned} S_n^{p,\ell}(x;\mu) = \frac{p}{\ell^p \binom{n}{\ell-1}} &\sum_{1 \leq i_1 < \cdots < i_{\ell-1} \leq n} m_{\ell-1}^p\big(\mathrm{Simpl}(X_{i_1}, \dots, X_{i_{\ell-1}}, \mu)\big) \\ &\times \frac{\Psi_{i_1,\dots,i_{\ell-1}}(x-\mu)}{\|\Psi_{i_1,\dots,i_{\ell-1}}(x-\mu)\|^{2-p}}, \end{aligned}$$

where $\Psi_{i_1,\dots,i_{\ell-1}}$ denotes the matrix of the orthogonal projection onto the orthogonal complement to the vector space spanned by $X_{i_1} - \mu, \dots, X_{i_{\ell-1}} - \mu$.

Of course, we expect that sample (p, ℓ)-scores will allow one to reconstruct their population versions. The following result is what will be needed in the sequel.

PROPOSITION 3.1. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1, \dots, d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order $2p$ and is centro-symmetric about μ. Let $X_1, \dots, X_n$ be a random sample from P. Then*

$$\mathrm{E}_P\bigg[\bigg\|\frac{1}{n}\sum_{i=1}^n \big(S_n^{p,\ell}(X_i;\mu) - S_P^{p,\ell}(X_i;\mu)\big)\bigg\|^2\bigg] = O\big(n^{-2}\big)$$

as n diverges to infinity.

Note that, under the assumptions of Proposition 3.1,

$$\mathrm{E}_P\left[\left\{m^p_{\ell-1}(\mathrm{Simpl}(X_1,\ldots,X_{\ell-1},\mu))\frac{(\Psi_{1,\ldots,\ell-1}(X_\ell-\mu))_r}{\|\Psi_{1,\ldots,\ell-1}(X_\ell-\mu)\|^{2-p}}\right\}^2\right]<\infty \tag{3.3}$$

for any $r=1,\ldots,d$ (this follows by applying Lemma S.1.1, then Lemma S.1.2). Moreover, by conditioning with respect to $X_1,\ldots,X_{\ell-1}$, the symmetry assumption on P yields

$$\mathrm{E}_P\left[m^p_{\ell-1}(\mathrm{Simpl}(X_1,\ldots,X_{\ell-1},\mu))\frac{\Psi_{1,\ldots,\ell-1}(X_\ell-\mu)}{\|\Psi_{1,\ldots,\ell-1}(X_\ell-\mu)\|^{2-p}}\right]=0. \tag{3.4}$$

Thus, jointly with Proposition 3.1, the multivariate central limit theorem entails that

$$T^{p,\ell}_n(\mu):=\frac{1}{\sqrt{n}}\sum_{i=1}^n S^{p,\ell}_n(X_i;\mu)=\frac{1}{\sqrt{n}}\sum_{i=1}^n S^{p,\ell}_P(X_i;\mu)+o_\mathrm{P}(1)\xrightarrow{\mathcal{D}}\mathcal{N}_d(0,V^{p,\ell}_P), \tag{3.5}$$

with $V^{p,\ell}_P:=\mathrm{E}_P[(S^{p,\ell}_P(X_1;\mu))(S^{p,\ell}_P(X_1;\mu))']$ (note that this matrix does not depend on μ, which justifies the notation).

We end this section by showing that for $\ell=d$, the (p,ℓ)-scores introduced above behave well under affine transformations. More precisely, we have the following result.

PROPOSITION 3.2. *Fix a real number $p\geq 1$ and an integer $d\geq 2$. Let $X_1,\ldots,X_n$ be a random sample and x be a d-vector. Let A be an invertible $d\times d$ matrix. Then, $S^{p,d}_{n,A}(Ax;0)=|\det A|^p(A^{-1})'S^{p,d}_n(x;0)$, where $S^{p,d}_{n,A}(\cdot;0)$ denotes the (p,d)-score computed from the transformed sample $AX_1,\ldots,AX_n$.*

This result will entail that, for any p, the (p,d)-score tests we will define in the next section will be affine-invariant. For $\ell\leq d-1$, the equivariance relation $S^{p,d}_{n,A}(Ax;0)=|\det A|^p(A^{-1})'S^{p,d}_n(x;0)$ in Proposition 3.2 above will only hold for rigid-body transformations fixing the origin, that is, for transformations described by matrices A of the form $A=\lambda O$, where λ is a positive real number and O is a $d\times d$ orthogonal matrix, which will entail that the corresponding (p,ℓ)-score tests will be invariant under rigid-body transformations, but not under general affine transformations. Parallel to what we had for point estimation, a notable exception is the case $p=2$, for which the fact that

$$\begin{aligned}S^{2,\ell}_n(x;0)&=\left(\frac{2}{\ell^2\binom{n}{\ell-1}}\sum_{1\leq i_1<\cdots<i_{\ell-1}\leq n}m^2_{\ell-1}(\mathrm{Simpl}(X_{i_1},\ldots,X_{i_{\ell-1}},0))\Psi_{i_1,\ldots,i_{\ell-1}}\right)x\\&=:B^{2,\ell}_n x\end{aligned} \tag{3.6}$$

will be sufficient to ensure affine invariance of the $(2,\ell)$-score tests for any ℓ.

3.2. *(p,ℓ)-score tests.* Without any loss of generality, we focus on the problem of testing the null hypothesis $\mathcal{H}_0:\mu=0$ against the alternative hypothesis $\mathcal{H}_1:\mu\neq 0$ on the basis of a random sample $X_1,\ldots,X_n$ from a probability measure P over $\mathbb{R}^d$ that is centro-symmetric about μ (for $\mathcal{H}_0:\mu=\mu_0$ against $\mathcal{H}_1:\mu\neq\mu_0$, one may base the test on $X_i-\mu_0$, $i=1,\ldots,n$). The (p,ℓ)-score test we propose rejects the null hypothesis for large values of

$$Q^{p,\ell}_n=(T^{p,\ell}_n)'(V^{p,\ell}_n)^{-1}T^{p,\ell}_n,$$

where $T^{p,\ell}_n=T^{p,\ell}_n(0)$ (see (3.5) above) and

$$V^{p,\ell}_n:=\frac{1}{n}\sum_{i=1}^n\left\{S^{p,\ell}_n(X_i;0)-\frac{1}{n}\sum_{j=1}^n S^{p,\ell}_n(X_j;0)\right\}\left\{S^{p,\ell}_n(X_i;0)-\frac{1}{n}\sum_{j=1}^n S^{p,\ell}_n(X_j;0)\right\}' \tag{3.7}$$

is the sample covariance matrix of $S_n^{p,\ell}(X_i;0)$, $i=1,\dots,n$. One may actually replace $V_n^{p,\ell}$ with any other weakly consistent estimator of $V_P^{p,\ell}$ under the null hypothesis, such as $\tilde{V}_n^{p,\ell} := \frac{1}{n}\sum_{i=1}^n (S_n^{p,\ell}(X_i;0))(S_n^{p,\ell}(X_i;0))'$ (in the proof of Theorem 3.1 below, we establish that both $V_n^{p,\ell}$ and $\tilde{V}_n^{p,\ell}$ indeed converge in probability to $V_P^{p,\ell}$ under the null hypothesis).

It readily follows from Proposition 3.2 that, for any p, the test statistic $Q_n^{p,d}$ is affine-invariant, in the sense that, with obvious notation, $Q_n^{p,d}(AX_1,\dots,AX_n) = Q_n^{p,d}(X_1,\dots,X_n)$ for any invertible $d\times d$ matrix A. In contrast, for $\ell \le d-1$, we only have the rigid-body invariance property $Q_n^{p,\ell}(\lambda OX_1,\dots,\lambda OX_n) = Q_n^{p,\ell}(X_1,\dots,X_n)$ for any positive real number λ and orthogonal $d\times d$ matrix O. Again, the case $p=2$ is an exception, since it trivially follows from (3.6) that $Q_n^{2,\ell}$, irrespective of ℓ, is the Hotelling test statistic, which is affine-invariant. It is worth noting that the (p,ℓ)-score tests above will require finite moments of order $2p$. The case $p=2$ is again an exception: thanks to the cancellation of the matrix $B_n^{2,\ell}$ in (3.6) when computing $Q_n^{2,\ell}$, finite moments of order 2 rather than $2p=4$ will actually be sufficient, which is the standard assumption for Hotelling T^2's test. A similar phenomenon appears for point estimation: for $p=2$, the asymptotic behavior for $\hat{\mu}_n^{2,\ell} = \bar{X}_n$ holds under finite second-order moments only, whereas $\hat{\mu}_n^{p,\ell}$, with $p\neq 2$, requires finite moments of order $2p$ (in line with this, all asymptotic results for the Oja median $\hat{\mu}_n^{1,d}$ in the literature require finite moments of order $2p=2$).

The following result describes the asymptotic null distribution of $Q_n^{p,\ell}$.

THEOREM 3.1. *Fix a real number $p\ge 1$, an integer $d\ge 2$ and $\ell\in\{1,\dots,d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order $2p$ and is centro-symmetric about $\mu=0$. Denote as $P_{n,0}$ the hypothesis under which the observations $X_1,\dots,X_n$ form a random sample from P. Then, under $P_{n,0}$,*

$$Q_n^{p,\ell} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n S_P^{p,\ell}(X_i;0)\right)'(V_P^{p,\ell})^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n S_P^{p,\ell}(X_i;0)\right) + o_{\mathrm{P}}(1) \xrightarrow{\mathcal{D}} \chi_d^2$$

as n diverges to infinity.

Thus, the resulting test $\phi_n^{p,\ell}$ rejects the null hypothesis $\mathcal{H}_0:\mu=0$ at asymptotic level α whenever $Q_n^{p,\ell} > \chi^2_{d,1-\alpha}$ (recall that $\chi^2_{d,1-\alpha}$ is the upper α-quantile of the chi-square distribution with d degrees of freedom). When based on $\tilde{V}_n^{p,\ell}$ rather than on $V_n^{p,\ell}$, these tests reduce to the *spatial sign test* from Möttönen and Oja (1995) for $(p,\ell)=(1,1)$ and to the *Oja affine-invariant sign test* from Hettmansperger, Nyblom and Oja (1994) (see also Oja (1999)) for $(p,\ell)=(1,d)$. For $p=2$ and any $\ell\in\{1,\dots,d\}$, the proposed tests reduce to the classical Hotelling T^2 test (see, e.g., Chapter 5 in Anderson (2003)). The following result describes the asymptotic behavior of (p,ℓ)-score tests under contiguous local alternatives.

THEOREM 3.2. *Fix a real number $p\ge 1$, an integer $d\ge 2$ and $\ell\in\{1,\dots,d\}$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order $2p$, is centro-symmetric about $\mu=0$, and admits a density f whose square-root $f^{1/2}$ is weakly differentiable over $\mathbb{R}^d$ with a square-integrable weak gradient $\nabla f^{1/2}$. Fix $\tau\in\mathbb{R}^d$ and denote as $P_{n,\tau}$ the hypothesis under which the observations $X_1,\dots,X_n$ are mutually independent and have the common density $x\mapsto f(x-n^{-1/2}\tau)$. Let*

$$C_P^{p,\ell} := \int_{\mathbb{R}^d} S_P^{p,\ell}(x;0)\big(\varphi_f(x)\big)' f(x)\,dx, \quad \textit{with } \varphi_f(x) := -\frac{2\nabla f^{1/2}(x)}{f^{1/2}(x)}.$$

Then, under $P_{n,\tau}$, $Q_n^{p,\ell}$ is asymptotically noncentral chi-square with d degrees of freedom and noncentrality parameter $\delta = \tau'(C_P^{p,\ell})'(V_P^{p,\ell})^{-1}C_P^{p,\ell}\tau$.

As for point estimation, the result above is very general but it does not allow for a direct comparison between the various (p,ℓ)-score tests. The next result makes such a comparison possible in the spherical case and, for affine-invariant tests, also in the elliptical case.

COROLLARY 3.1. *Fix a real number $p \geq 1$, an integer $d \geq 2$ and $\ell \in \{1,\dots,d\}$. Fix a symmetric positive definite $d \times d$ matrix Σ, which, for $\ell \leq d-1$ and $p \neq 2$, is assumed to be of the form $\Sigma = \sigma^2 I_d$, with $\sigma^2 > 0$. Let P be a probability measure over $\mathbb{R}^d$ that admits finite moments of order $2p$ and admits the density $x \mapsto f(\|x\|_\Sigma)$. Assume that f is weakly differentiable over $\mathbb{R}_0^+$ with a weak derivative $(f^{1/2})'$ such that $\int_0^\infty \{(f^{1/2})'(r)\}^2 r^{d-1}\,dr < \infty$. Fix a d-vector τ and denote as $P_{n,\tau}$ the hypothesis under which the observations $X_1,\dots,X_n$ are mutually independent and have the common density $x \mapsto f(\|x - n^{-1/2}\tau\|_\Sigma)$. Finally, put*

$$\delta := \frac{(p+d-2)^2(\mathrm{E}_P[\|X_1\|_\Sigma^{p-2}])^2}{d\mathrm{E}_P[\|X_1\|_\Sigma^{2p-2}]}\tau'\Sigma^{-1}\tau \tag{3.8}$$

(for $p < 2$, this tacitly assumes that $\mathrm{E}_P[\|X_1\|_\Sigma^{p-2}] < \infty$, that is, $\int_0^\infty f(r) r^{p+d-3}\,dr < \infty$). Then, under $P_{n,\tau}$, $Q_n^{p,\ell}$ is asymptotically noncentral chi-square with d degrees of freedom and noncentrality parameter δ.

In the spherical case $\Sigma = \sigma^2 I_d$, the asymptotic power of the (p,ℓ)-score test under the sequence of local alternatives $P_{n,\tau}$ is therefore given by

$$\begin{aligned}&\lim_{n\to\infty} P_{n,\tau}[Q_n^{p,\ell} > \chi^2_{d,1-\alpha}]\\ &\quad= 1 - G_d\biggl(\chi^2_{d,1-\alpha}; \frac{(p+d-2)^2(\mathrm{E}_P[\|X_1\|_\Sigma^{p-2}])^2}{d\mathrm{E}_P[\|X_1\|_\Sigma^{2p-2}]}\tau'\Sigma^{-1}\tau\biggr),\end{aligned} \tag{3.9}$$

where $G_d(\cdot;\delta)$ is the cumulative distribution function of the noncentral chi-square distribution with d degrees of freedom and noncentrality parameter δ. Note that, under sphericity, these asymptotic local powers do not depend on ℓ (which is in line with what we had for point estimation in Section 2.3) and that, for $p=1$ and $p=2$, we have

$$\delta = \frac{(d-1)^2(\mathrm{E}_P[\|X_1\|^{-1}])^2}{d}\|\tau\|^2 \quad \text{and} \quad \delta = \frac{d}{\mathrm{E}_P[\|X_1\|^2]}\|\tau\|^2,$$

respectively, which is compatible with the "spatial" efficiencies obtained for $(p,\ell) = (1,1)$ and $(p,\ell) = (2,1)$ in, for example, Möttönen, Oja and Tienari (1997), page 547. Still in the spherical case, the asymptotic relative efficiency of the (p,ℓ)-score test with respect to the Hotelling test (that coincides with the $(2,\ell)$-score test for any ℓ) is

$$\frac{(p+d-2)^2\mathrm{E}_P[\|X_1\|_\Sigma^2](\mathrm{E}_P[\|X_1\|_\Sigma^{p-2}])^2}{d^2\mathrm{E}_P[\|X_1\|_\Sigma^{2p-2}]}$$

(this is obtained as the ratio of the corresponding noncentrality parameters τ) and, from affine invariance, this extends to the elliptical case for $\ell = d$; see Table 1 for some numerical values. It should be noted that these asymptotic relative efficiencies coincide, in the spherical case (or in the elliptical case for $\ell = d$), with those of $\hat{\mu}_n^{p,\ell}$ with respect to $\bar{X}_n$ (since the covariance matrices in (2.13) to consider to obtain asymptotic relative efficiencies for point estimation are proportional to each other, these asymptotic relative efficiencies are simply obtained as the ratio of the corresponding proportionality factors). The fact that asymptotic relative efficiencies match for point estimation and hypothesis testing supports the claim that the (p,ℓ)-score tests we introduced above are the companion tests to the estimators $\hat{\mu}_n^{p,\ell}$ introduced in Section 2. Obviously, explicit expressions of the asymptotic powers (or asymptotic relative efficiencies) under t distributions and power-exponential distributions can then be obtained exactly as for point estimation; see the discussion below Corollary 2.1.

TABLE 1

Asymptotic relative efficiencies of the (p, d)-score tests with respect to the Hotelling T^2 test under several d-dimensional elliptical distributions, namely under d-dimensional t distributions with $\nu = 3, 6, 9, 12$ degrees of freedom, d-dimensional Gaussian distributions, and d-dimensional power-exponential distributions with tail parameter $\eta = 2, 3, 4, 6$ (whenever these elliptical distributions are spherically symmetric, these AREs also are those of the (p, ℓ)-score tests for $\ell = 1, \ldots, d-1$, still with respect to the Hotelling test); "–" indicates that the required moment assumptions are not satisfied

		$\nu=3$	$\nu=6$	$\nu=9$	$\nu=12$	$\mathcal{N}$	$\eta=2$	$\eta=3$	$\eta=4$	$\eta=6$
$d=2$	$p=1$	2.	1.084	0.955	0.904	0.785	0.59	0.546	0.528	0.514
	$p=1.25$	2.007	1.162	1.043	0.996	0.887	0.712	0.67	0.652	0.638
	$p=1.5$	–	1.165	1.073	1.037	0.953	0.821	0.787	0.773	0.761
	$p=3$	–	–	0.588	0.672	0.884	1.214	1.334	1.391	1.443
	$p=4$	–	–	0.204	0.32	0.667	1.273	1.55	1.694	1.833
$d=3$	$p=1$	2.162	1.172	1.032	0.977	0.849	0.688	0.647	0.629	0.615
	$p=1.25$	2.079	1.203	1.08	1.032	0.919	0.78	0.743	0.727	0.714
	$p=1.5$	–	1.18	1.088	1.051	0.966	0.863	0.834	0.822	0.811
	$p=3$	–	–	0.602	0.689	0.905	1.168	1.263	1.309	1.351
	$p=4$	–	–	0.219	0.343	0.714	1.216	1.435	1.55	1.661
$d=4$	$p=1$	2.25	1.22	1.075	1.017	0.884	0.749	0.712	0.696	0.682
	$p=1.25$	2.12	1.227	1.102	1.052	0.937	0.822	0.79	0.776	0.764
	$p=1.5$	–	1.189	1.096	1.059	0.973	0.889	0.864	0.854	0.844
	$p=3$	–	–	0.612	0.7	0.92	1.138	1.216	1.254	1.290
	$p=4$	–	–	0.23	0.36	0.75	1.178	1.359	1.454	1.547
$d=6$	$p=1$	2.344	1.271	1.12	1.059	0.92	0.82	0.79	0.777	0.764
	$p=1.25$	2.164	1.253	1.125	1.074	0.956	0.872	0.847	0.836	0.825
	$p=1.5$	–	1.199	1.105	1.068	0.981	0.92	0.901	0.892	0.885
	$p=3$	–	–	0.625	0.715	0.94	1.101	1.159	1.187	1.215
	$p=4$	–	–	0.245	0.384	0.8	1.132	1.266	1.336	1.405
$d=10$	$p=1$	2.422	1.313	1.157	1.095	0.951	0.885	0.865	0.855	0.846
	$p=1.25$	2.202	1.274	1.144	1.093	0.973	0.919	0.901	0.893	0.885
	$p=1.5$	–	1.208	1.113	1.075	0.988	0.949	0.936	0.93	0.924
	$p=3$	–	–	0.638	0.73	0.959	1.066	1.103	1.122	1.141
	$p=4$	–	–	0.262	0.411	0.857	1.086	1.174	1.22	1.266

4. Monte Carlo exercises and a real data example. In this section, we explore by simulations the finite-sample relevance of our asymptotic results, both for point estimation (Section 4.1) and for hypothesis testing (Section 4.2), and we treat a real data example (Section 4.3). Simulated data will be obtained from elliptical t distributions and power-exponential distributions, that is, from distributions admitting density functions proportional to

$$x \mapsto \left(1 + \|x - \mu\|_\Sigma^2/\nu\right)^{-(d+\nu)/2} \quad \text{and} \quad x \mapsto \exp\left(-\frac{1}{2}\|x - \mu\|_\Sigma^{2\eta}\right),$$

respectively. Of course, this allows us to consider Gaussian distributions, that are power-exponential distributions with $\eta = 1$. We mostly focus on the trivariate case $d = 3$, but we will also consider the ten-variate case $d = 10$ when exploring the finite-sample behavior of the incomplete estimators introduced in Section 2.4.

4.1. *Point estimation.* Since all location estimators considered in this work are equivariant under translations, there is no loss of generality to restrict to $\mu = 0$. We then performed a simulation in the spherical case and one in the elliptical case, where we used $\Sigma = I_d$ and $\Sigma = \mathrm{diag}(100^2, 1, 1)$, respectively (going from the spherical setup to the elliptical one thus corresponds to multiplying the first marginal by 100, a factor that is not an extreme one for a change of measurement units). It should be noted that there is no loss of generality to restrict to the uncorrelated case where Σ is diagonal, as all location estimators considered in this work are equivariant under orthogonal transformations. For each value of Σ above, we generated $M = 5000$ independent random samples $X_1, \ldots, X_n$ of size $n = 100$ from the trivariate t_3 distribution, Gaussian distribution, and power-exponential distribution with $\eta = 2$. In each sample, we evaluated the location estimators $\hat{\mu}_n^{p,\ell}$, for any combination of $p \in \{1, 1.25, 2, 4\}$ and $\ell \in \{1, 2, 3\}$ (since $\hat{\mu}_n^{2,\ell} = \bar{X}_n$ for any ℓ, this provides 10 estimators). Figure 1 then presents boxplots of the standardized mean square errors (MSEs)

$$\|\hat{\mu}_n^{p,\ell}(m) - \mu_P^{p,\ell}\|_\Sigma^2 = \|\hat{\mu}_n^{p,\ell}(m) - 0\|_\Sigma^2, \quad m = 1, \ldots, M, \tag{4.1}$$

where $\hat{\mu}_n^{p,\ell}(m)$ denotes the value taken by $\hat{\mu}_n^{p,\ell}$ in the mth replication.

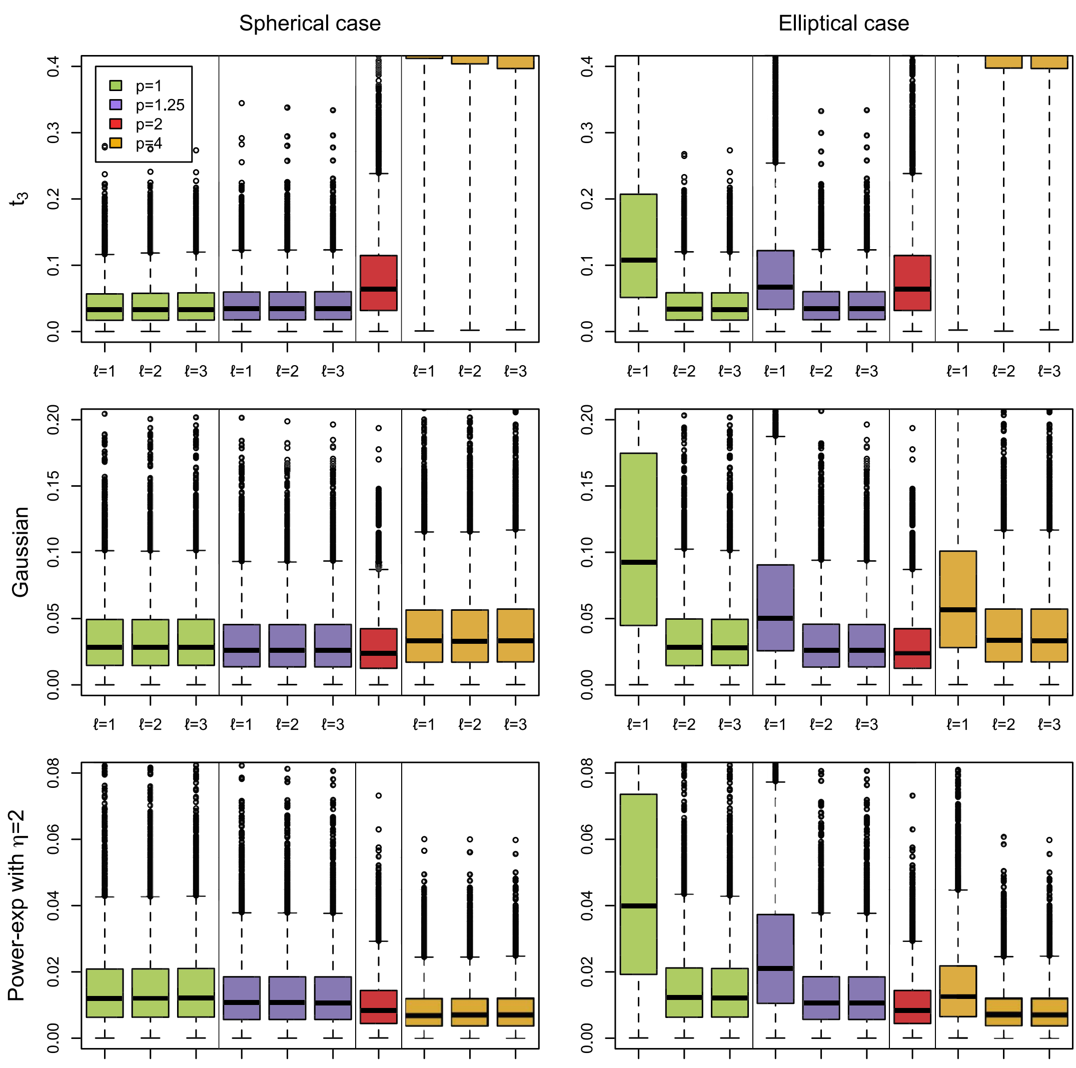

FIG. 1. *(Left) Boxplots of the standardized mean square errors of $\hat{\mu}_n^{p,\ell}$ in (4.1) obtained from $M = 5000$ independent random samples of size $n = 100$ from several trivariate distributions with location $\mu = 0$ and scatter matrix $\Sigma = I_d$ (since $\hat{\mu}_n^{2,\ell} = \bar{X}_n$ for any ℓ, each panel shows only one boxplot for $p = 2$); the distributions considered are the t_3 distribution (top), Gaussian distribution (middle), and power-exponential distribution with tail parameter $\eta = 2$ (bottom). (Right) The corresponding results for $\Sigma = \mathrm{diag}(100^2, 1, 1)$.*

Figure 1 fully supports our theoretical results and provides interesting further insight. In the spherical case, the independence on ℓ of the behavior of $\hat{\mu}_n^{p,\ell}$ clearly materializes already for the sample size considered. As expected, this independence does not survive departures from sphericity, where estimators associated with $\ell = 1$ exhibit much larger MSEs. Interestingly, estimators associated with $\ell = 2$ virtually show the same efficiency under ellipticity as under sphericity. Yet there is no guarantee that this will still be the case for more extreme departures from sphericity, whereas, from affine equivariance, estimators associated with $\ell = 3$ will behave in the exact same way for any value of Σ. It remains that, for the elliptical distribution considered in the present simulation, the estimators associated with $\ell = 2$, that dominate affine-equivariant estimators on a computational point of view (clearly, the larger ℓ, the heavier the computational burden to evaluate $\hat{\mu}_n^{p,\ell}$) compete equally with affine-equivariant estimators in terms of MSEs. Of course, results also significantly depend on p. As expected, the sample mean $\hat{\mu}_n^{2,\ell}$ and the estimators $\hat{\mu}_n^{4,\ell}$ dominate their competitors under spherical Gaussian distributions and spherical power-exponential distributions with tail parameter $\eta = 2$, respectively (from affine equivariance, this extends to elliptical distributions for $\ell = 3$); this is expected since our results entail that these estimators are asymptotically equivalent to the corresponding maximum likelihood estimators. As a rule, larger (resp., smaller) values of p provide more efficient estimators under light-tailed (resp., heavy-tailed) distributions. Finally, since they do not meet the required moments assumptions at t_3 distributions, the estimators associated with $p = 4$ exhibit there much higher MSEs than the other estimators (showing the boxplots for $p = 4$ would prevent comparing the other estimators).

It is natural to explore how well the asymptotic normality result from Theorem 2.5 materializes at the sample size considered. This result in particular entails that

$$\sqrt{n}(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell})_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, (\sigma_P^{p,\ell})^2),$$

where an explicit expression of $\sigma_P^{p,\ell}$ in the spherical case can be obtained from Corollary 2.1. To explore the quality of this asymptotic approximation, we restricted to the spherical setups considered in the left panels of Figure 1 and computed, for each combination of $p \in \{1, 1.25, 4\}$ and $\ell \in \{1, 2, 3\}$ a kernel density estimate[3] for the distribution of $\sqrt{n}(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell})_1/\sigma_P^{p,\ell}$ based on the $M = 5000$ values of the estimator at hand (we may safely ignore $p = 2$, as this would just explore finite-sample relevance of the central limit theorem). These kernel density estimates are plotted in the left panels of Figure 2, where histograms are also provided for $(p, \ell) = (1.25, d = 3)$. Clearly, the results are in an almost perfect agreement with the standard normal density (which is plotted as a solid black curve in each panel), except, obviously, for the case $p = 4$ under the t_3 distributions which does not satisfy the required moment assumptions. From affine equivariance, strictly the same agreement holds in the elliptical case for $\ell = d$.

Finally, we turn our attention to the incomplete estimators introduced in Section 2.4. We consider affine-equivariant estimation of location in dimension 10, that is, $\ell = d = 10$. For sample size $n = 100$ as above, our estimators are then defined as minimizers of an objective function with $\binom{n}{d} \approx 17.3 \times 10^{12}$ terms, making evaluations of our estimators too costly to perform a Monte Carlo exercise. To explore the asymptotic behavior of incomplete estimators, we generated $M = 5000$ independent random samples $X_1, \dots, X_n$ of size $n = 100$ from the ten-variate t_3 distribution, Gaussian distribution, and power-exponential distribution with $\eta = 2$, in each case with location $\mu = 0$ and scatter $\Sigma = I_d$. In each sample, we evaluated the incomplete estimators $\hat{\mu}_{n,N}^{p,d}$, for any combination of $p \in \{1, 1.25, 4\}$ and

[3] Kernel density estimates throughout are obtained from the `R` function `density` with default parameters.

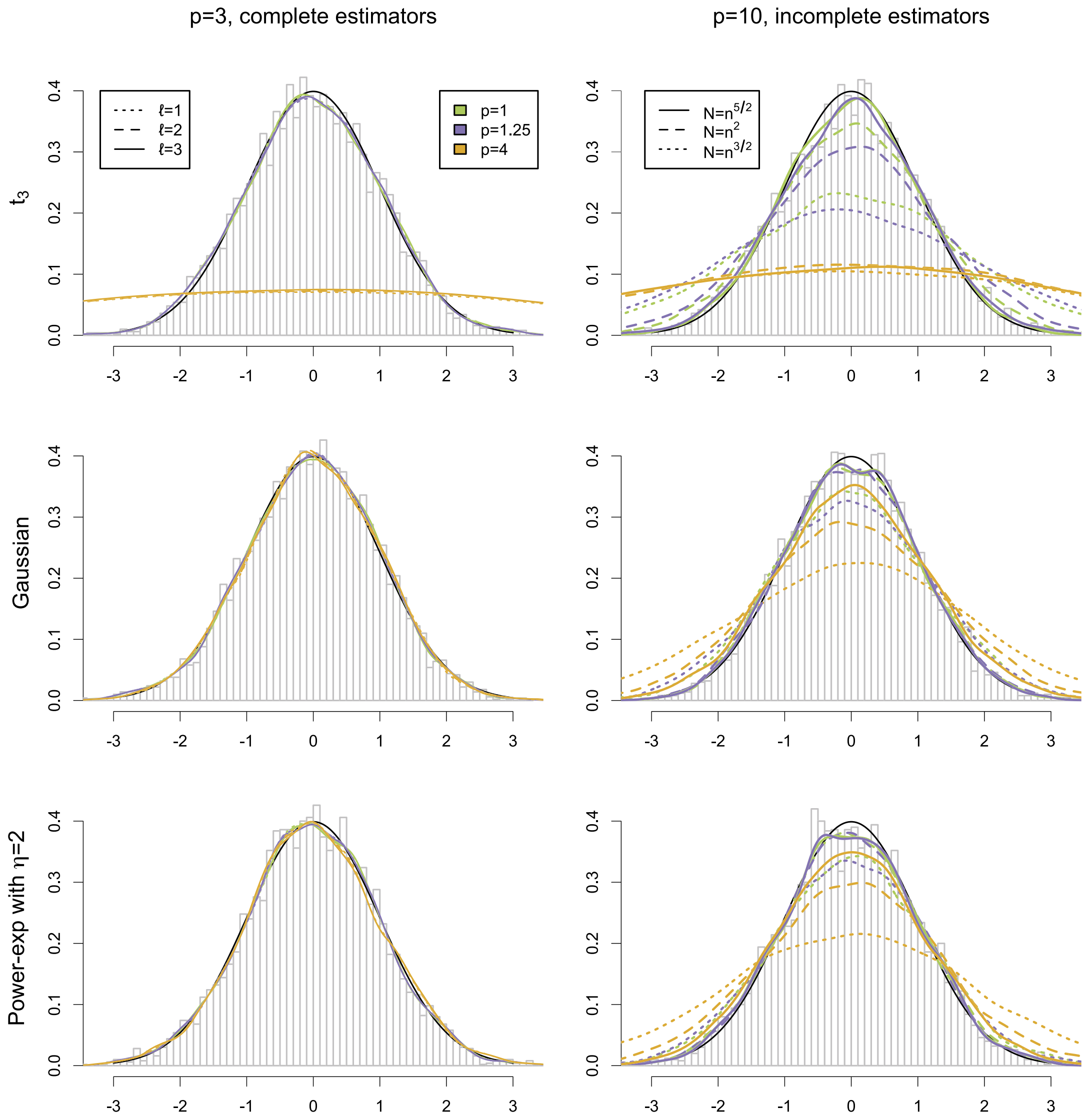

FIG. 2. *(Left) Kernel density estimates of the density of* $\sqrt{n}(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell})_1/\sigma_P^{p,\ell}$ *obtained from* $M = 5000$ *independent random samples of size* $n = 100$ *from several trivariate distributions with location* $\mu = 0$ *and scatter matrix* $\Sigma = I_d$; *the distributions considered are the* t_3 *distribution (top), Gaussian distribution (middle), and power-exponential distribution with tail parameter* $\eta = 2$ *(bottom). For each distribution, histograms are also provided for the* $M = 5000$ *values of* $\sqrt{n}(\hat{\mu}_n^{p,\ell} - \mu_P^{p,\ell})_1/\sigma_P^{p,\ell}$ *associated with* $(p, \ell) = (1.25, d = 3)$. *(Right) the corresponding results for* $\ell = d = 10$ *and for the incomplete estimators* $\hat{\mu}_{n,N}^{p,d}$ *in (2.15) based on sampling with replacement with* $N = n^{5/2}$, $N = n^2$ *and* $N = n^{3/2}$; *see Section 4.1 for details.*

$N \in \{n^{3/2}, n^2, n^{5/2}\}$ (using sampling with replacement as a sampling scheme). These incomplete estimators can be evaluated without any problem since they minimize objective functions with $N = n^{3/2} = 1000$ terms, $N = n^2 = 10{,}000$ terms, or $N = n^{5/2} = 100{,}000$ terms, which is a dramatic improvement over their "complete" antecedents above. The right panels of Figure 2 shows the resulting kernel density estimates for the distribution of $\sqrt{n}(\hat{\mu}_{n,N}^{p,\ell} - \mu_P^{p,\ell})_1/\sigma_P^{p,\ell}$ for each combination of p and N; histograms are also provided for $(p, N) = (1.25, n^{5/2})$. Forgetting again about the case $p = 4$ under t_3 distributions that does not meet the required moment assumptions, the results show that the agreement with the standard normal distribution is very good (it is only for $p = 4$ that the largest value of N is really needed), which supports the asymptotic result in Theorem 2.8. Incomplete estimators thus provide a practical way to evaluate the proposed estimators in higher dimensions (and/or for larger sample sizes).

4.2. *Hypothesis testing*. As in Section 3, we focus again, without any loss of generality, on the problem of testing the null hypothesis $\mathcal{H}_0 : \mu = 0$ against the alternative hypothesis $\mathcal{H}_1 : \mu \neq 0$. We performed one simulation in the spherical case and two in the elliptical case. In the spherical case, we generated $M = 5000$ independent random samples $X_1, \dots, X_n$ of size $n = 100$ from the trivariate t_3 distribution, Gaussian distribution, and power-exponential distribution with $\eta = 2$, each time with $\Sigma = I_d$ and $\mu_s = n^{-1/2}\Sigma^{1/2}(0, 0, s)'$, for any $s = 0, 0.5, 1, 1.5, \dots, 5$; the value $s = 0$ corresponds to the null hypothesis, whereas larger values of s provide increasingly severe alternatives. The elliptical cases use the same setup but with $\Sigma = \mathrm{diag}(100^2, 100^2, 1)$ and $\Sigma = \mathrm{diag}(100^2, 1, 1)$ (the fact that nonnull values of μ deviate from the null one in the third marginal makes it interesting to consider these two types of elliptical structures). In each sample, we performed the test $\phi_n^{p,\ell}$ at asymptotic level $\alpha = 5\%$ for any combination of $p \in \{1, 1.25, 2, 4\}$ and $\ell \in \{1, 2, 3\}$ (since $\phi_n^{2,\ell}$ is the classical Hotelling T^2 test for any ℓ, this provides 10 tests). Figure 3 plots the resulting rejection frequencies as functions of s, as well as the corresponding asymptotic powers obtained in the spherical case (and, for affine-invariant tests, in the elliptical cases) from Corollary 3.1; see (3.9). As for point estimation above, where we reported MSEs that were standardized in Σ, these rejection frequencies are also standardized through the role played by Σ in the nonnull values μ_s.

The results from Figure 3 are in a perfect agreement with our asymptotic results, which is remarkable in view of the relatively moderate sample size considered ($n = 100$):[4] in the spherical case, finite-sample power curves show a dependence on p but not on ℓ, and they virtually coincide with asymptotic power curves. The ranking of power curves is compatible with the AREs from Table 1 and with what we had for point estimation, favouring large (resp., small) values of p under light (resp., heavy) tails. The tests associated with $p = 4$ show barely any power under t_3 distributions, as they do not satisfy the required moment assumptions there. As for point estimation, independence on ℓ does not hold away from the spherical case, and only the affine-invariant tests associated with $\ell = 3$ maintain the same performance under ellipticity as under sphericity, hence should be favoured over competing tests. Yet it is interesting to see how both elliptical setups do differ: in particular, tests associated with $\ell = 1$ suffer more in the right panels of Figure 3 than in the middle panels, whereas the opposite holds true for tests associated with $\ell = 2$.

4.3. *A real-data example*. To illustrate the proposed methods on real data, we consider the data set `LASERI` (see the `R` package `ICSNP`), that contains the cardiovascular responses to a passive head-up tilt for 223 individuals. Here, we focus on responses in the following three haemodynamic variables: *wave velocity* (`PWVT`, in metres/second), *cardiac output* (`CO`, in litres/minute), and *systemic vascular resistance* (`SVRI`, in dynes×seconds/centimetre). For each of these variables, the data set offers four values per subject, namely (value 1:) the average value in the 10th minute of rest before the tilt, (value 2:) the average value in the 2nd minute during the tilt, (value 3:) the average value in the 5th minute during the tilt, and (value 4:) the average value in the 5th minute after the tilt. Several studies investigate whether or not the haemodynamic system 5 minutes after the tilt has come back already to pre-tilt

[4]It is only for the classical Hotelling test under t_3 distributions that some difference is observed, which is due to the fact that these distributions are close to distributions with infinite second-order moments under which the Hotelling test will collapse. This difference of course vanishes for larger sample sizes n.

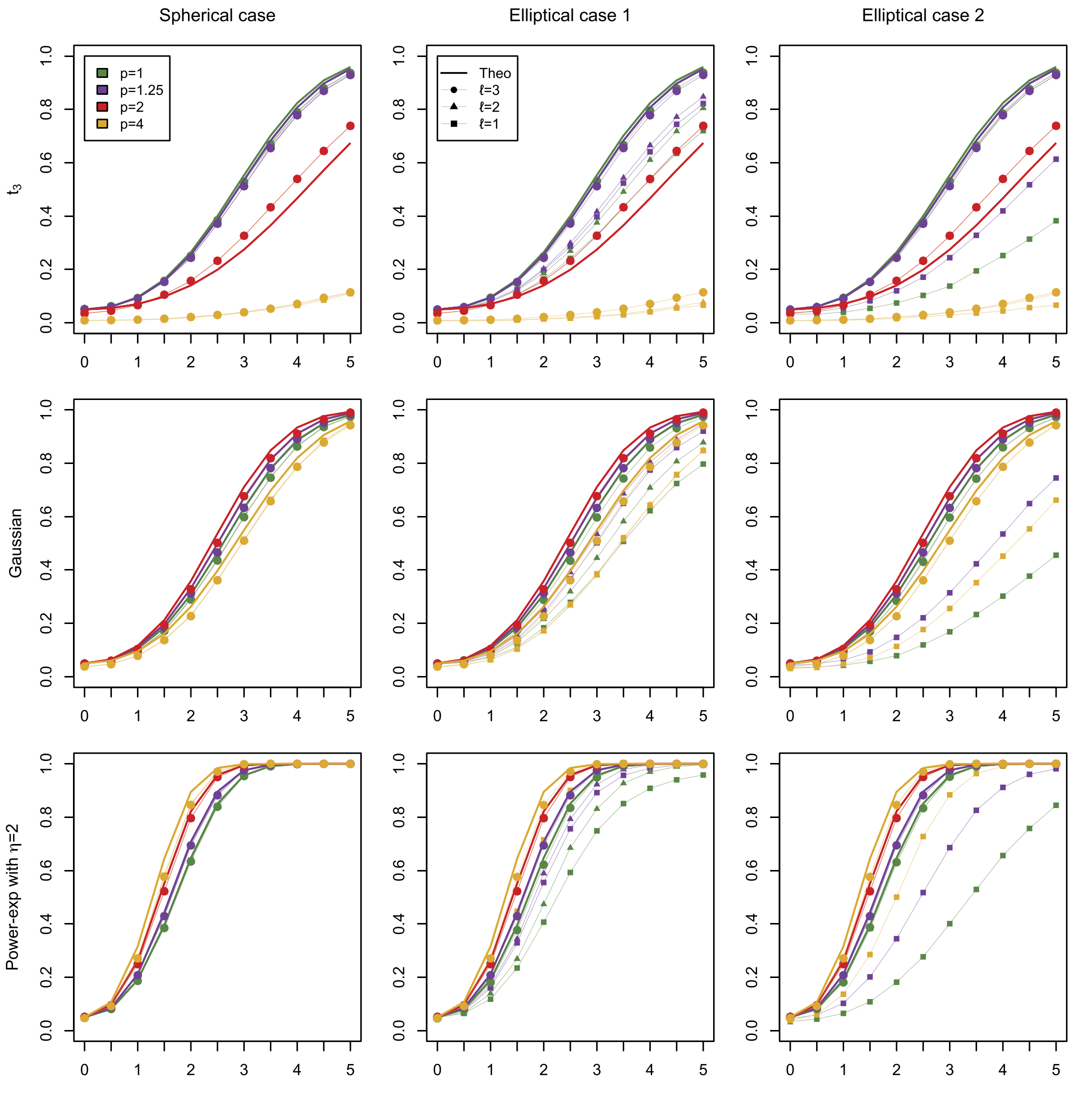

FIG. 3. *(Left) As functions of s, rejection frequencies of the tests $\phi_n^{p,\ell}$ obtained from $M = 5000$ independent random samples of size $n = 100$ from several trivariate distributions with scatter matrix $\Sigma = I_d$ and location $\mu = \mu_s = \sqrt{n}\Sigma^{1/2}(0, 0, s)'$ (since $\phi_n^{2,\ell}$ is the Hotelling T^2 test for any ℓ, each panel shows only one curve for $p = 2$); the distributions considered are the t_3 distribution (top), Gaussian distribution (middle), and power-exponential distribution with tail parameter $\eta = 2$ (bottom) (Center) The corresponding results for $\Sigma = \mathrm{diag}(100^2, 100^2, 1)$. (Right) The corresponding results for $\Sigma = \mathrm{diag}(100^2, 1, 1)$. In each panel, asymptotic power curves under sphericity (and, for affine-invariant tests, under ellipticity) are also plotted.*

levels, based on differences between values 1 and 4. Any reasonable test, however, rejects the null hypothesis that the haemodynamic system is back to normal (see, e.g., Fischer et al. (2020)), and we therefore rather focus on what happens during the tilt, based on differences between values of 2 and 3. The variables of interest here are thus, `PWVT2T3`, `COT2T3`, and `SVRIT2T3`, defined as the difference between the value at time 3 and at time 2 for each of the three variables above.

We thus test the null hypothesis that the vector made of `PWVT2T3`, `COT2T3`, and `SVRIT2T3` has location zero. Since the data does not reveal clear departures from central symmetry, we may apply our tests $\phi_n^{p,\ell}$ from Section 3, which we do for all combinations of

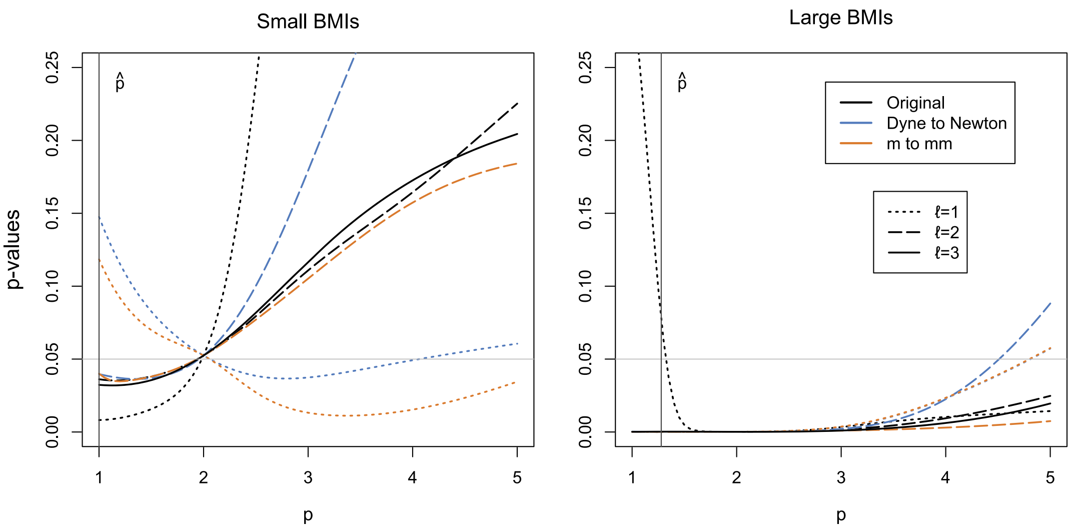

FIG. 4. *(Left) p-values of the tests* $\phi_n^{p,\ell}$, $\ell = 1, 2, 3$, *as functions of p for the real data set considered in Section* 4.3, *or more precisely for the subsample of subjects with a BMI smaller than* 25. *Blue curves (resp., orange curves) are those obtained when millimetres are used instead of metres in the variable* SVRIT2T3 *(resp., when Newtons are used instead of dynes in the variable* PWVT2T3). *(Right) The corresponding results for the subjects with a BMI larger than or equal to* 25.

$p \in \{1, 1.1, 1.2, \ldots, 5\}$ and $\ell = 1, 2, 3$. We actually do so separately on the 99 subjects showing a *body mass index (BMI)* below 25 and on the $223 - 99 = 124$ subjects with a BMI that is at least 25. The corresponding p-values are plotted as functions of p in black in Figure 4, with different line types to discriminate between the various values of ℓ. Clearly, the null hypothesis is almost always rejected for subjects with a large BMI, which indicates that these subjects suffer a change in the considered haemodynamic variables during the tilt (it is only for $\ell = 1$ and very small values of $p = 1$ that the null hypothesis is not rejected). The picture is much different for subjects with a small BMI: for those, our tests, at nominal level $\alpha = 5\%$, tend to reject the null hypothesis for most values of p below 2 and to not reject for values above 2. A moment estimate of p, in the range $p \in [1, \infty)$, in the model with densities (1.4) takes value 1 and 1.28 for subjects with small and large BMIs, respectively, which suggests in particular that the suitable decision is to reject the null hypothesis for small BMIs, too. For small BMIs, rejection for small values of p is clearer for $\ell = 1$ than for $\ell \in \{2, 3\}$, hence in particular, than for our affine-invariant tests obtained with $\ell = 3$. One should recall, however, that the outcome of such tests is may be much affected by the choice of measurement units, which is of course most undesirable. To illustrate this, we show the corresponding results (i) when metres are replace with millimetres in PWV and, alternatively, (ii) when dynes are replaced with Newtons in SVRI. As expected, such marginal scale changes have no impact on our affine-invariant tests associated with $\ell = 3$, but it is seen that they may dramatically change the p-values of the other tests, and especially of the tests associated with $\ell = 1$: in particular, for these tests, such scale changes among small BMIs basically always turn rejection into nonrejection and vice-versa!

Of course, it is also natural to consider point estimation in this framework. As it is difficult to provide informative static 3D plots, we encourage the reader to visit this link,[5] where a dynamic plot is available and where it is possible to show/hide our affine-equivariant estimators $\hat{\mu}_n^{p,3}$ for various values of p as well as the corresponding confidence ellipsoids from (2.10), in both cases separately for subjects with small or large BMIs.

[5] https://chart-studio.plotly.com/~alexanderduerre/14/#/

5. Final comments. Both for point estimation and hypothesis testing, we proposed a class of simplex-based inference procedures indexed by $p \geq 1$ and $\ell \in \{1, \ldots, d\}$, where p refers to the L_p loss considered and ℓ is the dimension of the simplices involved. As the real data example above suggested, it would be natural to tackle the problem of selecting suitable values of p and ℓ in a given situation. As the title of the present paper hints, we recommend choosing $\ell = d$ as affine equivariance/invariance will entail that the resulting procedures will show the same performance irrespective of the amount of ellipticity at hand (and will also behave as expected under possible marginal changes in measurement units). Whenever the sample size n and/or dimension d makes the computational burden too heavy, our affine-equivariant incomplete estimators from Section 2.4 can be used and will show the same asymptotic properties as their complete antecedents, which further strengthens our recommendation to use $\ell = d$. In a generic framework, the choice of p, which is somehow orthogonal to the choice of ℓ, may be based on cross-validation, whereas in the specific framework of power-exponential distributions, it is natural to substitute an estimator $\hat{p}$ for p in $\hat{\mu}_n^{p,\ell}$ and $\phi_n^{p,\ell}$. It is easy to check that μ and p are Fisher-orthogonal parameters in this model, which strongly suggests adaptivity (that is, which suggests that the resulting inference procedures enjoy the same asymptotic properties as the oracle ones based on the true value of p). For hypothesis testing, a natural approach, that would be an alternative to conducting an affine-invariant test based on a selected value of p, would be to consider the family of affine-invariant tests indexed by p and to use as a p-value the minimal p-value within this family. Turning this into a genuine testing procedure, however, would require a stochastic process asymptotic investigation that is beyond the scope of the present work.

While we did not comment on algorithmic aspects in Section 4, it is worth mentioning that, in the numerical experiments there, the location estimators $\hat{\mu}_n^{p,\ell}$ and $\hat{\mu}_{n,N}^{p,\ell}$ were evaluated by relying on standard optimization routines in `R` (more precisely, we used the built-in function `nlm`). While this successfully computed these estimators in a reasonable amount of time, it would of course be desirable to design algorithms that are tailor made for the objective functions considered in this work, and that would in particular exploit convexity of these objective functions. Possibly, the efficient algorithms recently developed for the Oja median ($p = 1$ and $\ell = d$) in Fischer et al. (2020) can be adapted to other values of p and ℓ.

Finally, while our work focused on inference for multivariate location, other problems of multivariate statistics may probably be tackled in an affine-equivariant way through the simplex-based approach we considered. As suggested by an anonymous referee, the minimal value of the location objective function is actually a natural measure of spread. In particular, for $p = 2$ and $\ell = 1$, this measure of spread is the total variance, that is, the trace of the covariance matrix. Beyond estimation of a scalar measure of spread, an interesting direction for future research is the more intricate problem of estimating a $d \times d$ scatter matrix through the simplex-based approach. This will be considered in future work.

Acknowledgments. We would like to thank the Editor, the Associate Editor and two anonymous referees for their insightful comments and suggestions that led to a substantial improvement of a previous version of this work.

Funding. This research was supported the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles.

SUPPLEMENTARY MATERIAL

Supplement to "Affine-equivariant inference for multivariate location under L_p loss functions" (DOI: 10.1214/22-AOS2199SUPP; .pdf). In this supplement, we prove all the results of the present paper.

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. Wiley, New York.

BLOM, G. (1976). Some properties of incomplete U-statistics. *Biometrika* **63** 573–580. MR0474582

BROWN, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30. MR0701072

CARDOT, H., CÉNAC, P. and GODICHON-BAGGIONI, A. (2017). Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *Ann. Statist.* **45** 591–614. MR3650394 https://doi.org/10.1214/16-AOS1460

CARDOT, H., CÉNAC, P. and ZITT, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli* **19** 18–43. MR3019484 https://doi.org/10.3150/11-BEJ390

DAVIES, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292. MR0902258 https://doi.org/10.1214/aos/1176350505

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. MR1193313 https://doi.org/10.1214/aos/1176348890

DÜRRE, A. and PAINDAVEINE, D. (2021). On the consistency of incomplete U-statistics under infinite second-order moments. arXiv:2112.14666.

DÜRRE, A. and PAINDAVEINE, D. (2022). Supplement to "Affine-equivariant inference for multivariate location under L_p loss functions." https://doi.org/10.1214/22-AOS2199SUPP

ENQVIST, E. (1978). On sampling from sets of random variables with application to incomplete U-statistics. Dissertation, Lund Univ.

FISCHER, D., MOSLER, K., MÖTTÖNEN, J., NORDHAUSEN, K., POKOTYLO, O. and VOGEL, D. (2020). Computing the Oja median in R: The package OjaNP. *J. Stat. Softw.* **92**.

HETTMANSPERGER, T. P., MÖTTÖNEN, J. and OJA, H. (1997). Affine-invariant multivariate one-sample signed-rank tests. *J. Amer. Statist. Assoc.* **92** 1591–1600. MR1615268 https://doi.org/10.2307/2965430

HETTMANSPERGER, T. P., NYBLOM, J. and OJA, H. (1994). Affine invariant multivariate one-sample sign tests. *J. Roy. Statist. Soc. Ser. B* **56** 221–234. MR1257809

JANSON, S. (1984). The asymptotic distributions of incomplete U-statistics. *Probab. Theory Related Fields* **66** 495–505. MR0753810

KONEN, D. and PAINDAVEINE, D. (2022). Multivariate ρ-quantiles: A spatial approach. *Bernoulli* **28** 1912–1934. MR4411516 https://doi.org/10.3150/21-bej1404

LOPUHAÄ, H. P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Ann. Statist.* **27** 1638–1665. MR1742503 https://doi.org/10.1214/aos/1017939145

MAGYAR, A. and TYLER, D. E. (2011). The asymptotic efficiency of the spatial median for elliptically symmetric distributions. *Sankhya B* **73** 165–192. MR2924190 https://doi.org/10.1007/s13571-011-0032-x

MINSKER, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli* **21** 2308–2335. MR3378468 https://doi.org/10.3150/14-BEJ645

MÖTTÖNEN, J., NORDHAUSEN, K. and OJA, H. (2010). Asymptotic theory of the spatial median. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis*: *A Festschrift in Honor of Professor Jana Jurečková*. *Inst. Math. Stat.* (*IMS*) *Collect.* **7** 182–193. IMS, Beachwood, OH. MR2808378

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.* **5** 201–213. MR1346895 https://doi.org/10.1080/10485259508832643

MÖTTÖNEN, J., OJA, H. and TIENARI, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.* **25** 542–552. MR1439313 https://doi.org/10.1214/aos/1031833663

NIINIMAA, A. and OJA, H. (1995). On the influence functions of certain bivariate medians. *J. Roy. Statist. Soc. Ser. B* **57** 565–574. MR1341324

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332. MR0721446 https://doi.org/10.1016/0167-7152(83)90054-8

OJA, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scand. J. Stat.* **26** 319–343. MR1712063 https://doi.org/10.1111/1467-9469.00152

OJA, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*. *Lecture Notes in Statistics* **199**. Springer, New York. MR2598854 https://doi.org/10.1007/978-1-4419-0468-3

OLLILA, E., OJA, H. and HETTMANSPERGER, T. P. (2002). Estimates of regression coefficients based on the sign covariance matrix. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 447–466. MR1924300 https://doi.org/10.1111/1467-9868.00344

PAINDAVEINE, D. (2022). On the measure of anchored Gaussian simplices, with applications to multivariate medians. *Bernoulli* **28** 965–996. MR4388926 https://doi.org/10.3150/21-bej1373

ZUO, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.* **31** 1460–1490. MR2012822 https://doi.org/10.1214/aos/1065705115