Non-existence of some affinely equivariant location functionals in dimension $d \geq 2$

An *affine transformation* from $\mathbb{R}^d$ to itself is one of the form $Ax = Bx + v$ for all $x \in \mathbb{R}^d$ where B is a linear transformation ($d \times d$ matrix) and v is a fixed vector. Then A will be called *non-singular* if and only if B is. Here x and v are $d \times 1$ column vectors.

For any probability measure P and random variable X, which may be vector-valued, we have another probability measure $P \circ X^{-1}$, the distribution of X or image measure of P by X. For example, if P is defined on $\mathbb{R}^d$ and x_j is the jth coordinate function on $\mathbb{R}^d$, then $P \circ x_j^{-1}$ is the jth marginal of P, on $\mathbb{R}$.

Let $\mathcal{P}$ be a collection of probability measures on $\mathbb{R}^d$ and m a function from $\mathcal{P}$ into $\mathbb{R}^d$. Then m will be called an *affinely equivariant location functional* on $\mathcal{P}$ iff whenever $P \in \mathcal{P}$ and A is a non-singular affine transformation, we have $P \circ A^{-1} \in \mathcal{P}$ and $m(P \circ A^{-1}) = Am(P)$. Also, $m(\cdot)$ will be called *singularly affine(ly) equivariant* if the same holds when A may be singular.

When $d = 1$, the median is a singularly affine equivariant location functional defined on the class of *all* probability measures on $\mathbb{R}$. For $d = 1$, a singular linear transformation B is just multiplication by 0, and so for any P, $P \circ A^{-1}$ is concentrated in the point v. It turns out not to be restrictive to say that for such a distribution m should equal v. For $d \geq 2$, however, there are more singular matrices, and we will see that singular affine equivariance becomes very restrictive.

Recall that $\delta_x(A) := 1_A(x) := 1$ if $x \in A$ and 0 otherwise. For $n = 1, 2, ...,$ and $d = 1, 2, ...,$ let $\mathcal{P}_{n,d}$ be the class of all empirical measures $P_n = \frac{1}{n}\sum_{j=1}^n \delta_{x_j}$ on $\mathbb{R}^d$ where each $x_j = (x_{1j}, ..., x_{dj})' \in \mathbb{R}^d$. Clearly, for any transformation A from $\mathbb{R}^d$ into itself (affine or not) and P_n as given, $P_n \circ A^{-1} = \frac{1}{n}\sum_{j=1}^n \delta_{A(x_j)} \in \mathcal{P}_{n,d}$. Here is the main fact in this handout:

Theorem (Obenchain, 1971). Let $d \geq 2$ and suppose m is a singularly affine equivariant location functional defined on $\mathcal{P}_{n,d}$ for a given n. Then $m(P_n) = \int x\, dP_n = \overline{x} = \sum_{j=1}^n x_j/n$ for all $P_n \in \mathcal{P}_{n,d}$.

Remark. For $d = 1$ there are some robust singularly affine equivariant location functionals such as the median (and trimmed means, e.g. Randles and Wolfe, problem 7.4.2 pp. 246-247). But the sample mean $\overline{x}$ has breakdown point 0 for all n, so a singularly affine equivariant location functional on $\mathcal{P}_{n,d}$ for $d \geq 2$ can't have any robustness. Thus, researchers consider affinely (not singularly) equivariant functionals, not defined on all of $\mathcal{P}_{n,d}$, e.g. not defined on $P_n \circ A^{-1}$ for A singular.

Proof. For $X_j \in \mathbb{R}^d$, $j = 1, ..., n$, with $X_j = (X_{1j}, ..., X_{dj})'$, let X be the $d \times n$ *data matrix* X_{ij} for $i = 1, ..., d$ and $j = 1, ..., n$, so that X_j is the jth column of X. Let $P_n := \frac{1}{n}\sum_{j=1}^n \delta_{X_j} \in \mathcal{P}_{n,d}$. Then $m(P_n)$ is a function of X, say $m(P_n) \equiv M(X)$. Let B be any $d \times d$ matrix. Then the data matrix for $BX_1, ..., BX_n$ is BX, i.e. the jth column of BX is BX_j, so

$$M(BX) = m(P_n \circ B^{-1}) = Bm(P_n) = BM(X)$$

by singular affine equivariance.

Some special choices of B will be made. First, for each $u = 1, ..., d$, let $B^{(u)}_{ir} = 0$ if $i \geq 2$ or if $i = 1$ and $r \neq u$, with $B^{(u)}_{1u} := 1$. Let $X^{(u)}$ denote the uth row of X, so that $(X^{(u)})_j \equiv X_{uj}$ for $j = 1, ..., n$. For any $1 \times n$ vector V, let $\tilde{V}$ be the $d \times n$ matrix whose first row is V and whose other rows are all 0's. Then $B^{(u)}X = \tilde{X}^{(u)}$, so

$$M(\tilde{X}^{(u)}) = M(B^{(u)}X) = B^{(u)}M(X) = (M_u(X), 0, ..., 0)',$$

where $M(X) = (M_1(X), ..., M_d(X))'$. Thus

$$M_1(\tilde{X}^{(u)}) \equiv M_u(X). \tag{1}$$

Next, for any real numbers a and b, define a $d \times d$ matrix $B^{a,b}$ by $B^{a,b}_{11} := a$, $B^{a,b}_{12} := b$, and $B^{a,b}_{ij} := 0$ for all other i and j, i.e. for $i \geq 2$ or $j \geq 3$. Then $B^{a,b}X = (aX^{(1)} + bX^{(2)})^{\sim}$, so

$$M([aX^{(1)} + bX^{(2)}]^{\sim}) = M(B^{a,b}X) = B^{a,b}M(X) = (aM_1(X) + bM_2(X), 0, ..., 0)'. \tag{2}$$

By (1), $M_1(X) = M_1(\tilde{X}^{(1)})$ and $M_2(X) = M_1(\tilde{X}^{(2)})$. Equating first components in (2) gives

$$aM_1(\tilde{X}^{(1)}) + bM_1(\tilde{X}^{(2)}) = M_1(a\tilde{X}^{(1)} + b\tilde{X}^{(2)}).$$

For any (row vector($y \in \mathbb{R}^n$, we have a map $y \mapsto L(y) := M_1(\tilde{y})$ which is linear since $X^{(1)}$ and $X^{(2)}$ can be any two $1 \times n$ vectors and a, b any two real numbers. Thus $M_1(\tilde{y}) \equiv yz$ for some column vector $z \in \mathbb{R}^n$.

Now for any data matrix X, we have by (1)

$$M(X) = (M_1(X), ..., M_d(X))' = (M_1(\tilde{X}^{(1)}), ..., M_1(\tilde{X}^{(d)}))'$$

$$= (X^{(1)}z, ..., X^{(d)}z)' = Xz.$$

Next, any permutation of the columns X_j of X gives the same P_n and thus the same $M(X) = m(P_n)$, so the components of z are all equal, $z = (z_1, ..., z_1)'$. Thus $M(X) \equiv nz_1\overline{X}$.

Now suppose all X_j equal some $v \neq 0$ and let $Ax \equiv 2x - v$. Then $Av = v$, so $M(AX) = M(X) = nz_1 v = AM(X) = 2nz_1 v - v$. It follows that $z_1 = 1/n$ and $M(X) \equiv \overline{X}$, proving the theorem. □

Remarks. If an affinely invariant location functional m is defined on all of $\mathcal{P}_{n,d}$ and M is continuous as a function of $X_1, ..., X_n$, then m must be singularly affine equivariant and so is equal to $\overline{X}$.

Recall that a sequence Q_n of probability measures is said to converge weakly to Q_0 if $\int f dQ_n \to \int f dQ_0$ for every bounded continuous function f. (This form of convergence is used in the central limit theorem, for example.) Let $Q_n = (n-1)\delta_0/n + \delta_n/n$. Suppose m is an affinely equivariant location functional defined on $\mathcal{P}_{n,d}$ for all n for a given $d \geq 2$ and that m is continuous for weak convergence. Then it is continuous in $X_1, ..., X_n$ for fixed n, so it is singularly affine equivariant, and by Obenchain's theorem, $m(Q_n) = \int x dQ_n = 1$

for all n. But Q_n converge weakly to δ_0, for which $m(\delta_0) = 0$, contradicting the weak continuity, so no such m exists.

Note. The theorem is essentially contained in the statement and proof of Obenchain (1971, Lemma 1).

REFERENCE

Obenchain, R. L. (1971). Multivariate procedures invariant under linear transformations. *Ann. Math. Statist.* **42**, 1569-1578.