Jensen's inequality for multivariate medians

Milan Merkle
University of Belgrade, Serbia
emerkle@etf.rs

Given a probability measure μ on Borel sigma-field of $\mathbb{R}^d$, and a function $f:\ \mathbb{R}^d \mapsto \mathbb{R}$, the main issue of this work is to establish inequalities of the type $f(m) \leq M$, where m is a median (or a deepest point in the sense explained in the paper) of μ and M is a median (or an appropriate quantile) of the measure $\mu_f = \mu \circ f^{-1}$. For a most popular choice of halfspace depth, we prove that the Jensen's inequality holds for the class of quasi-convex and lower semi-continuous functions f.

To accomplish the task, we give a sequence of results regarding the "type D depth functions" according to classification in Y. Zuo and R. Serfling, Ann. Stat. **28** (2000), 461-482, and prove several structural properties of medians, deepest points and depth functions. We introduce a notion of a median with respect to a partial order in $\mathbb{R}^d$ and we present a version of Jensen's inequality for such medians. Replacing means in classical Jensen's inequality with medians gives rise to applications in the framework of Pitman's estimation.

Michigan State University, Department of Statistics and Probability
October 6, 2009

Jensen's inequality

Let μ be a probability measure on Borel sets of $\mathbb{R}^d$, $d \geq 1$, and let f be a real valued convex function defined on $\mathbb{R}^d$. The Jensen's inequality states that

$$f(m) \leq M \tag{1}$$

where

$$m = \int_{\mathbb{R}^d} \boldsymbol{x} \, \mathrm{d}\mu(\boldsymbol{x}) \quad \text{and} \quad M = \int_{\mathbb{R}} f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}).$$

Can we replace means m and M with corresponding medians?

Recall: $m \in \{\mathrm{Med}\, \mu\}$ if

$$\mu((-\infty, m]) \geq \frac{1}{2}, \qquad \mu([m, +\infty)) \geq \frac{1}{2}. \tag{2}$$

The set $\{\mathrm{Med}\, \mu\}$ of all medians m is a nonempty compact interval.

- Medians always exist
- Issues of robustness
- Inequalities can be sharper
- Build up a median based theory

Two results in $\mathbb{R}$ $(d = 1)$

Given a measure μ and a measurable real valued function f, let μ_f be a measure defined by $\mu_f(B) = \mu(\{x \mid f(x) \in B\})$, and let M be its median.

Theorem 1. *(R. J. Tomkins, Ann. Probab. 1975) Let μ be a probability measure on $\mathbb{R}$ and let f be a convex function defined on $\mathbb{R}$. Then for every median m of μ there exists a median M of μ_f such that (1) holds, i.e.,*

$$\max\{f(\{\mathrm{Med}\,\mu\})\} \leq \max\{\mathrm{Med}\,\mu_f\}. \tag{3}$$

Theorem 2. *(M. M, SPL 2005) Let μ be a probability measure on $\mathbb{R}$ and let f be a quasi-convex lower semi-continuous function defined on $\mathbb{R}$. Then for every median M of μ_f there exists a median m of μ such that (1) holds, i.e.,*

$$\min\{f(\{\mathrm{Med}\,\mu\})\} \leq \min\{\mathrm{Med}\,\mu_f\}. \tag{4}$$

Multivariate medians

To extend mentioned results to $d > 1$, we have first to choose among several possible notions of multivariate medians.

We may pick up a characteristics property of one-dimensional medians and extend it to a multivariate setup. However, by doing so, not all median properties can be preserved.

Let $\mathcal{U}$ be a specified collection of sets in $\mathbb{R}^d$, $d \geq 1$, and let μ be a probability measure on Borel sets of $\mathbb{R}^d$. For each $x \in \mathbb{R}^d$, define a **depth function**

$$D(x; \mu, \mathcal{U}) = \inf\{\mu(U) \mid x \in U \in \mathcal{U}\}. \tag{5}$$

(Type D of Zuo and Serfling, AS 2000)

In the case $d = 1$, with $\mathcal{U}$ being the set of intervals of the form $[a, +\infty)$ and $(-\infty, b]$ we have

$$D(x; \mu, \mathcal{U}) = \min\{\mu((-\infty, x]), \mu([x, +\infty))\},$$

and the set of deepest points has the following three properties:

- √ It is a compact interval.
- √ It is the set of all points x with the property that $D(x; \mu, \mathcal{U}) \geq \frac{1}{2}$.
- √ It is affine invariant set.

Which properties will be preserved in $d > 1$ depends on a choice of a family $\mathcal{U}$.

Assumptions

$$D(x; \mu, \mathcal{U}) = \inf\{\mu(U) \mid x \in U \in \mathcal{U}\}$$

(C_1) for every $x \in \mathbb{R}^d$ there is a $U \in \mathcal{U}$ so that $x \in U$.

(C_2') $D(x; P, \mathcal{U}) > 0$ for at least one $x \in \mathbb{R}^d$ and

(C_2'') $\lim\limits_{\|x\| \to +\infty} D(x; P, \mathcal{U}) = 0$

Condition (C_1) implies that $D \geq 0$, and (C_2) implies that D is not constant.

Tukey's depth: $\mathcal{U}$ is the set of all open (or all closed) halfspaces.

Let

$$\mathcal{V} = \{U^c \mid U \in \mathcal{U}\}$$

The depth function can be also specified in terms of $\mathcal{V}$.

Level sets, centers of a distribution and medians -1

Lemma 1. *Let $\mathcal{U}$ be any collection of non-empty sets in $\mathbb{R}^d$, such that the condition (C_1) holds:*

$$(C_1) \quad \text{for every } x \in \mathbb{R}^d \text{ there is a } U \in \mathcal{U} \text{ so that } x \in U$$

and let $\mathcal{V}$ be the collection of complements of sets in $\mathcal{U}$. Then, for any probability measure μ,

$$S_\alpha(\mu, \mathcal{U}) = \bigcap_{V \in \mathcal{V}, \mu(V) > 1-\alpha} V, \tag{6}$$

for any $\alpha \in (0, 1]$ such that there exists a set $U \in \mathcal{U}$ with $\mu(U) < \alpha$; otherwise $S_\alpha(\mu, \mathcal{U}) = \mathbb{R}^d$.

$S_\alpha(\mu, \mathcal{U})$ is called a **level set**.

If α_m is the maximum value of $D(x; \mu, \mathcal{U})$ for a given distribution μ, the set $S_{\alpha_m}(\mu, \mathcal{U})$ is called the **center** of μ and denoted by $C(\mu, \mathcal{U})$.

If $\alpha_m \geq 1/2$, we use the term **median**.

EXAMPLE: Let $\mathcal{V}$ be the family of all closed intervals in $\mathbb{R}$, and $\mathcal{U}$ the family of their complements. Then

$$S_\alpha = [q_\alpha, Q_{1-\alpha}],$$

where q_α is the smallest quantile of μ of order α, and $Q_{1-\alpha}$ is the largest quantile of μ of order $1 - \alpha$:

$$\begin{aligned} q_\alpha &= \min\{t \in \mathbb{R} \mid \mu\left((-\infty, t]\right) \geq \alpha\} \quad \text{and} \\ Q_{1-\alpha} &= \max\{t \in \mathbb{R} \mid \mu\left([t, +\infty)\right) \geq \alpha\}. \end{aligned} \tag{7}$$

For $\alpha = \frac{1}{2}$, $[q_{\frac{1}{2}}, Q_{\frac{1}{2}}]$ is the median interval.

Level sets, centers of a distribution and medians-2

Let $\mathcal{V}$ be a collection of **closed** subsets of $\mathbb{R}^d$ and let $\mathcal{U}$ be the collection of complements of sets in $\mathcal{V}$, and assume the conditions:

$$(C_1) \qquad \text{for every } x \in \mathbb{R}^d \text{ there is a } U \in \mathcal{U} \text{ so that } x \in U.$$

$$(C_2') \qquad D(x; P, \mathcal{U}) > 0 \text{ for at least one } x \in \mathbb{R}^d \qquad \text{and}$$
$$(C_2'') \qquad \lim_{\|x\| \to +\infty} D(x; P, \mathcal{U}) = 0$$

Theorem 3. *Under (C_1), the function $x \mapsto D(x; \mu, \mathcal{U})$ is upper semi-continuous. In addition, under conditions (C_2), the set $C(\mu, \mathcal{U})$ on which D reaches its maximum is equal to the minimal nonempty set S_α, that is,*

$$C(\mu, \mathcal{U}) = \bigcap_{\alpha : S_\alpha \neq \emptyset} S_\alpha(\mu, \mathcal{U}).$$

The set $C(\mu, \mathcal{U})$ is a non-empty compact set and it has the following representation:

$$C(\mu, \mathcal{U}) = \bigcap_{V \in \mathcal{V}, \mu(V) > 1 - \alpha_m} V, \quad \textit{where } \alpha_m = \max_{x \in \mathbb{R}^d} D(x; \mu, \mathcal{U}). \tag{8}$$

Some examples

Recall:

$$D(x;\mu,\mathcal{U}) = \inf\{\mu(U) \mid x \in U \in \mathcal{U}\}$$
$$S_\alpha(\mu,\mathcal{U}) = \bigcap_{V\in\mathcal{V},\mu(V)>1-\alpha} V$$

1° Consider the halfspace depth in $\mathbb{R}^2$, with the probability measure μ which assigns mass 1/3 to points $A(0,1)$, $B(-1,0)$ and $C(1,0)$ in the plane. Each point x in the closed triangle ABC has $D(x) = \frac{1}{3}$; points outside of the triangle have $D(x) = 0$. So, the function D reaches its maximum value $\frac{1}{3}$.

2° Let us now observe the same distribution, but with depth function defined with the family $\mathcal{V}$ of closed disks. The intersection of *all* closed disks V with $\mu(V) > 2/3$ is, in fact, the intersection of all disks that contain all three points A, B, C, and that is the closed triangle ABC. For any $\varepsilon > 0$, a disc V with $\mu(V) > 2/3 - \varepsilon$ may contain only two of points A, B, C, but then it is easy to see that the family of all such discs has the empty intersection. Therefore, S_α is non-empty for $\alpha \leq 1/3$, and again, the function D attains its maximum value 1/3 at the points of closed triangle ABC. In fact, depth functions in cases 1° and 2° are equivalent regardless of the dimension. The value of 1/3 is the maximal depth that can be generally expected in the two dimensional plane.

3° If $\mathcal{V}$ is the family of rectangles with sides parallel to coordinate axes, then the maximum depth is 2/3 and it is attained at $(0,0)$. Families $\mathcal{V}$ that are generalizations of intervals and rectangles will be considered next. We show that the maximal depth with alike families is always at least 1/2, regardless of dimension.

Partial order and intervals in $\mathbb{R}^d$

In $d = 1$, the median set can be represented as the intersection of all intervals with a probability mass $> 1/2$:

$$\{\mathrm{Med}\,\mu\} = \bigcap_{J=[a,b]:\ \mu(J)>1/2} J.$$

Let $\preceq$ be a partial order in $\overline{\mathbb{R}}^d$ and let $\boldsymbol{a}, \boldsymbol{b}$ be arbitrary points in $\overline{\mathbb{R}}^d$. We define a d-dimensional interval $[\boldsymbol{a}, \boldsymbol{b}]$ as the set of points in $\mathbb{R}^d$ that are between $\boldsymbol{a}$ and $\boldsymbol{b}$:

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{R}^d \mid \boldsymbol{a} \preceq \boldsymbol{x} \preceq \boldsymbol{b}\}$$

Assume the following three technical conditions:

(I1) Any interval $[\boldsymbol{a}, \boldsymbol{b}]$ is topologically closed, and for any $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ (i.e., with finite coordinates), the interval $[\boldsymbol{a}, \boldsymbol{b}]$ is a compact set.

(I2) For any ball $B \subset \mathbb{R}^d$, there exist $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ such that $B \subset [\boldsymbol{a}, \boldsymbol{b}]$.

(I3) For any set S which is bounded from above with a finite point, there exists a finite $\sup S$. For any set S which is bounded from below with a finite point, there exists a finite $\inf S$.

Example: Convex cone partial order. Let K be a closed convex cone in $\mathbb{R}^d$, with vertex at origin, and suppose that there exists a closed hyperplane π, such that $\pi \cap K = \{0\}$ (that is, $K \setminus \{0\}$ is a subset of one of open halfspaces determined by π). Define the relation $\preceq$ by $\boldsymbol{x} \preceq \boldsymbol{y} \iff \boldsymbol{y} - \boldsymbol{x} \in K$. The interval is then

$$[\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \mid \boldsymbol{x} - \boldsymbol{a} \in K \wedge \boldsymbol{b} - \boldsymbol{x} \in K\} = (\boldsymbol{a} + K) \cap (\boldsymbol{b} - K).$$

If the endpoints have some coordinates infinite, then the interval is either $\boldsymbol{a} + K$ (if $\boldsymbol{b} \notin \mathbb{R}^d$) or $\boldsymbol{b} - K$ (if $\boldsymbol{a} \notin \mathbb{R}^d$) or $\mathbb{R}^d$ (if neither endpoint is in $\mathbb{R}^d$).

The simplest, coordinate-wise ordering, can be obtained with K chosen to be the orthant with $x_i \geq 0, i = 1, \ldots, d$. Then

$$\boldsymbol{x} \preceq \boldsymbol{y} \iff x_i \leq y_i, \quad i = 1, \ldots, d. \tag{9}$$

Directional medians in $\mathbb{R}^d$

Theorem 4. *Let $\preceq$ be a partial order in $\overline{\mathbb{R}}^d$ such that conditions (I1)–(I3) hold. Let μ be a probability measure on $\mathbb{R}^d$ and let $\mathcal{J}$ be a family of intervals with respect to a partial order $\preceq$, with the property that*

$$\mu(J) > \frac{1}{2}, \quad \text{for each } J \in \mathcal{J}. \tag{10}$$

Then the intersection of all intervals from $\mathcal{J}$ is a non-empty compact interval.

The compact interval claimed in the Theorem 4 can be, in analogy to one dimensional case taken as a definition of the median induced by the partial order $\preceq$:

$$\{\mathrm{Med}\,\mu\}_{\preceq} \quad := \bigcap_{J=[\boldsymbol{a},\boldsymbol{b}]:\ \mu(J)>1/2} J. \tag{11}$$

Let $\mathcal{V}$ be the family of all closed intervals with respect to some partial order $\preceq$ that satisfies conditions (I1)–(I3) and let $\mathcal{U}$ be the family of their complements. Assuming that the condition (C_1) holds:

$$(C_1) \qquad \text{for every } x \in \mathbb{R}^d \text{ there is a } U \in \mathcal{U} \text{ so that } x \in U,$$

we find, via Lemma 1, that the level sets S_α with respect to the depth function $D(x;\mu,\mathcal{U})$ can be expressed as

$$S_\alpha(\mu,\mathcal{U}) = \bigcap_{V\in\mathcal{V},\mu(V)>1-\alpha} V.$$

Hence, $D(x;\mu,\mathcal{U}) \geq 1/2$ for all $x \in \{\mathrm{Med}\,\mu\}_{\preceq}$.

Directional median has the following properties:

- ✓ It is a compact interval.
- ✓ It is the set of all points x with the property that $D(x;\mu,\mathcal{U}) \geq \frac{1}{2}$.
- ✗ It is affine invariant set.

Convex sets and halfspaces

Recall: $D(x;\mu,\mathcal{U}) = \inf\{\mu(U) \mid x \in U \in \mathcal{U}\}$;

It is natural to have a convex center of a distribution, hence sets in $\mathcal{V}$ should be convex. Further, sets in $\mathcal{U}$ should not be bounded; otherwise the depth at x could be equal to $\mu(\{x\})$.

$$(C_1) \qquad \text{for every } x \in \mathbb{R}^d \text{ there is a } U \in \mathcal{U} \text{ so that } x \in U.$$

$$(C_2') \qquad D(x;P,\mathcal{U}) > 0 \text{ for at least one } x \in \mathbb{R}^d \qquad \text{and}$$
$$(C_2'') \qquad \lim_{\|x\|\to+\infty} D(x;P,\mathcal{U}) = 0$$

Theorem 5. *Let μ be any probability measure on Borel sets of $\mathbb{R}^d$. Let $\mathcal{V}$ be any family of closed convex sets in $\mathbb{R}^d$, and let $\mathcal{U}$ be the family of their complements. Assume that conditions (C_1) and (C_2'') hold. Then the condition (C_2') also holds, and there exists a point $x \in \mathbb{R}^d$ with $D(x;\mu,\mathcal{U}) \geq \frac{1}{d+1}$.*

(Extension of results in Donoho and Gasko (1992), Rousseeuw and Ruts (1999))

Example: For any $d > 1$ there is a probability distribution μ such that the maximal Tukey's depth is exactly $1/(d+1)$.

Let $A_1, \ldots, A_{d+1}$ be points in $\mathbb{R}^d$ such that they do not belong to the same hyperplane (i.e. to any affine subspace of dimension less than d), and suppose that $\mu(\{A_i\}) = \frac{1}{d+1}$ for each $i = 1, 2, \ldots, d+1$. Let S be a closed d-dimensional simplex with vertices at $A_1, \ldots, A_{d+1}$, and let $x \in S$. If x is a vertex of S, then there exists a closed halfspace H such that $x \in H$ and other vertices do not belong to H; then $D(x) = \mu(H) = 1/(d+1)$. Otherwise, let S_x be a d-dimensional simplex with vertices in x and d points among $A_1, \ldots, A_{d+1}$ that make together an affinely independent set. Then for S_x and the remaining vertex, say A_1, there exists a separating hyperplane π such that $\pi \cap S_x = \{x\}$ and $A_1 \notin \pi$. Let H be a halfspace with boundary π, that contains A_1. Then also $D(x) = \mu(H) = 1/(d+1)$. So, all points $x \in S$ have $D(x) = 1/(d+1)$. Points x outside of S have $D(x) = 0$, which is easy to see. So, the maximal depth in this example is exactly $1/(d+1)$.

Equivalence of depth functions

Theorem 6. *Let $\mathcal{V}$ be a collection of closed convex sets and $\mathcal{U}$ the collection of complements of all sets in $\mathcal{V}$. For each $V \in \mathcal{V}$, consider a representation*

$$V = \bigcap_{\alpha \in A_V} H_\alpha, \tag{12}$$

where H_α are closed subspaces and A_V is an index set. Let

$$\mathcal{H}^V = \{\overline{H_\alpha^c} + x \mid \alpha \in A_V,\ x \in \mathbb{R}^d\}$$

be the collection of closures of complements of halfspaces H_α and their translations. Further, let

$$\mathcal{H} = \bigcup_{V \in \mathcal{V}} \mathcal{H}^V.$$

If for any $H \in \mathcal{H}$ there exists at most countable collection of sets $V_i \in \mathcal{V}$, such that

$$V_1 \subseteq V_2 \subseteq \cdots \quad \text{and} \quad \mathring{H} = \bigcup V_i, \tag{13}$$

then

$$D(x;\mu,\mathcal{U}) = D(x;\mu,\mathcal{H}) = D(x;\mu,\mathring{\mathcal{H}}), \quad \text{for every } x \in \mathbb{R}^d,$$

where $\mathring{\mathcal{H}}$ is the family of open halfspaces from $\mathcal{H}$.

Two important particular cases:

a) Let $\mathcal{V}$ be the family of closed intervals with respect to the partial order defined with a convex cone K. Then

$$D(x;\mu,\mathcal{U}) = D(x;\mu,\mathcal{H}),$$

where $\mathcal{U}$ is the family of complements of sets in $\mathcal{V}$ and H is the family of all tangent halfspaces to K, and their translations.

In particular, if $\mathcal{V}$ is the family of intervals with respect to the coordinate-wise partial order, then the corresponding depth function is the same as the depth function generated by halfspaces with borders parallel to the coordinate hyperplanes.

b) Let $\mathcal{H}$ be the family of all closed halfspaces, and let $\mathcal{U}_c, \mathcal{U}_k$ and $\mathcal{U}_b$ be families of complements of all closed convex sets, compact convex sets and closed balls, respectively. Then

$$D(x;\mu,\mathcal{H}) = D(x;\mu,\mathcal{U}_c) = D(x;\mu,\mathcal{U}_k) = D(x;\mu,\mathcal{U}_b).$$

Tukey's median

The center of a distribution with respect to the family of all halfspaces in $\mathbb{R}^d$ has the following properties:

$\surd$ It is a compact convex set.

X It is the set of all points x with the property that $D(x; \mu, \mathcal{U}) \geq \frac{1}{2}$.

$\surd$ It is affine invariant set.

Class of suitable functions: C-functions

Definition 0.1. *A function* $f : \mathbb{R}^d \mapsto \mathbb{R}$ *will be called a* C-function *with respect to a given family* $\mathcal{V}$ *of closed subsets of* $\mathbb{R}^d$, *if for every* $t \in \mathbb{R}$, $f^{-1}((-\infty, t]) \in \mathcal{V}$ *or is empty set.*

EXAMPLES:

- If $\mathcal{V}$ is the family of all closed convex sets in $\mathbb{R}^d$, then the class of corresponding C-functions is precisely the class of lower semi-continuous quasi-convex functions, i.e., functions f that have the property that $f^{-1}((-\infty, t])$ is a closed set for any $t \in \mathbb{R}$ and

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}, \qquad \lambda \in [0,1], \quad x, y \in \mathbb{R}^d.$$

  In particular, every convex function on $\mathbb{R}^d$ is a C-function with respect to the class of all convex sets.
- A function f is a C-functions with respect to a family of closed intervals (with respect to some partial order in $\mathbb{R}^d$), if and only if

$$\{\boldsymbol{x} \in \mathbb{R}^d \mid f(\boldsymbol{x}) \leq t\} = [\boldsymbol{a}, \boldsymbol{b}], \qquad \text{for some } \boldsymbol{a}, \boldsymbol{b} \in \overline{\mathbb{R}}^d .$$

  It is not clear if this condition can be replaced with some other, easier to check, as it was done in the first case.

Jensen's inequality for level sets

Theorem 7. *Let $\mathcal{V}$ be a family of closed subsets of $\mathbb{R}^d$, and let $\mathcal{U}$ be the family of their complements. Assume that conditions (C_1) and (C_2) hold with a given probability measure μ. Let $\alpha > 0$ be such that the level set $S_\alpha = S_\alpha(\mu,\mathcal{U})$ is nonempty, and let f be a C-function with respect to $\mathcal{V}$.*

Then for every $m \in S_\alpha$ we have that

$$f(m) \leq Q_{1-\alpha}, \tag{14}$$

where $Q_{1-\alpha}$ is the largest quantile of order $1-\alpha$ for μ_f. □

Corollary 1. *(Jensen's inequality for "Tukey's median"). Let f be a lower semi-continuous and quasi-convex function on $\mathbb{R}^d$, and let μ be an arbitrary probability measure on Borel sets of $\mathbb{R}^d$. Suppose that the depth function with respect to halfspaces reaches its maximum α_m on the set $C(\mu)$ ("Tukey's median set"). Then for every $m \in C(\mu)$,*

$$f(m) \leq Q_{1-\alpha_m}, \tag{15}$$

where $Q_{1-\alpha_m}$ is the largest quantile of order $1-\alpha_m$ for μ_f.

EXAMPLE (the bound is sharp): Let A, B, C be non-collinear points in the two dimensional plane, and let $\mathcal{H}$ be the collection of open halfplanes. Let $l(AB)$ be the line determined by A and B. Let H_1 be the closed halfspace that does not contain the interior of the triangle ABC and has $l(AB)$ for its boundary, and let H_2 be its complement. Define a function f by

$$f(x) = e^{-d(x,l(AB))} \quad \text{if } x \in H_1\ , \qquad f(x) = e^{d(x,l(AB))} \quad \text{if } x \in H_2,$$

where $d(\cdot,\cdot)$ is euclidean distance. Then $f(A) = 1$, $f(B) = 1$ and $f(C) > 1$, and f is a convex function. Now suppose that μ assigns mass $1/3$ to each of the points A, B, C. The center $C(\mu, \mathcal{H})$ of this distribution is the set of points of the triangle ABC, with $\alpha_m = 1/3$. Hence, for $m \in C(\mu, \mathcal{H})$, $f(m)$ takes all values in $[1, f(C)]$. On the other hand, quantiles for μ_f of the order $2/3$ are points in the closed interval $[1, f(C)]$; hence the most we can state is that $f(m) \leq f(C)$, with $f(C)$ being the largest quantile of order $2/3$. □

Jensen's inequality for directional medians

Theorem 8. *Let $\mathcal{V}$ be a family of closed intervals with respect to a partial order in $\mathbb{R}^d$, such that conditions (I1)–(I3) are satisfied. Let $\{\mathrm{Med}\,\mu\}$ be the median set of a probability measure μ with respect to the chosen partial order, and let f be a C-function with respect to the family $\mathcal{V}$. Then for every $M \in \mathrm{Med}\,\{\mu_f\}$, there exists an $m \in \{\mathrm{Med}\,\mu\}$, such that*

$$f(m) \leq M, \tag{16}$$

or equivalently, $\min f(\{\mathrm{Med}\,\mu\}) \leq \min\{\mathrm{Med}\,\mu_f\}$. *Further, for every $m \in \{\mathrm{Med}\,\mu\}$,*

$$f(m) \leq \max\{\mathrm{Med}\,\mu_f\}, \tag{17}$$

or, equivalently, $\sup f(\{\mathrm{Med}\,\mu\}) \leq \max\{\mathrm{Med}\,\mu_f\}$.

TWO EXAMPLES

For a d-dimensional random variable X with expectation $\mathrm{E}\,X$ and $\mathrm{Med}\,X = \mathrm{E}\,X$, we may use both classical Jensen's inequality $f(\mathrm{E}\,X) \le \mathrm{E}\,f(X)$ or one of inequalities derived above, provided that f is a convex C-function and that $\mathrm{E}\,f(X)$ exists. It can happen that the upper bound in terms of medians or quantiles is lower than $\mathrm{E}\,f(X)$. To illustrate the point, consider univariate case, with $X \sim \mathcal{N}(0,1)$ and $f(x) = (x-2)^2$. Then the classical Jensen's inequality with means gives $4 \le 5$. Since here $\mathrm{Med}\,(X-2)^2 = 4.00032$ (numerically evaluated), the inequality $f(\mathrm{E}\,X) \le \mathrm{Med}\,f(X)$ is sharper. Of course, if $\mathrm{E}\,f(X)$ does not exist, the median alternative is the only choice.

Let a and b are points in $\mathbb{R}^d$, and let $\|\cdot\|$ be usual euclidean norm. Since the function

$$x \mapsto \|x-a\|^2 - \|x-b\|^2$$

is affine, it is a C-function for the halfspace depth. Let m be a point in the center of a distribution μ, and let α_m be the value of the depth function in the center. Let X be a d-dimensional random variable on some probability space $(\Omega, \mathcal{F}, P)$ with the distribution μ. Consider the function $f(x) = \|x-a\|^2 - \|x-m\|^2$. Then we have that $0 \le \|m-a\|^2 \le Q_{1-\alpha_m}$, which implies that $P(f(X) \ge 0) \ge \alpha_m$, or, equivalently,

$$P(\|X-m\| \le \|X-a\|) \ge \alpha_m \qquad \text{for any } a \in \mathbb{R}^d. \tag{18}$$

The expression on the left hand side of (18) is known as Pitman's measure of nearness; in this case it measures the probability that X is closer to m than to any other chosen point a. For distributions with $\alpha_m = \frac{1}{2}$, (18) means that each point in "Tukey's median set" is a best non-random estimate of X (or, a most representative value) in the sense of Pitman's criterion, with the euclidean distance as a loss function. The analogous result in one dimensional case is well known.