Available online at www.sciencedirect.com

SCIENCE DIRECT®

Statistics & Probability Letters 71 (2005) 277–281

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Jensen's inequality for medians

Milan Merkle[1]

Department of Applied Mathematics, University of Belgrade, PO Box 35-54, 1120 Belgrade, Serbia and Montenegro

Received 5 August 2003; received in revised form 28 October 2004
Available online 24 December 2004

Abstract

We prove an analogue of Jensen's inequality, with medians instead of means. A novel definition of a median is given, which allows a natural extension to higher dimensions.

MSC: primary 60E15; 62E99; 62H99

Keywords: Median; Probability distribution; Jensen's inequality; Multivariate median

---

1. An analogue of Jensen's inequality for medians

For any random variable X with a finite expectation EX and for any convex function f, it holds that $f(EX) \leqslant Ef(X)$. This is one version of famous Jensen's inequality, which plays a significant role in probability and statistics. It is natural to ask if the expectation can be replaced with some other kind of mean value. In this work, we present an analogue of Jensen's inequality, where expectation is replaced by median, and we show that this inequality holds for a class of functions that contains convex functions as a proper subset. The main tool in derivation of the inequality is a new characterization of a median, which turns out to be suitable for extension to multivariate distributions. The number of known inequalities regarding medians, unlike expectations, is very small, probably the best known is mean-median-mode inequality (see Basu and DasGupta, 1997; Dharmadhikari and Joag-Dev, 1988), and this paper is a contribution to the topic.

---

[1] A major part of this research was done during the author's visit to the Department of Statistics at Federal University of Rio de Janeiro (UFRJ), Brasil.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2004.11.010

For convenience, we start with the definition of a median.

Definition 1.1. Let X be a real-valued random variable with distribution function F. A median of X (or a median of F) is any real number m which satisfies the inequalities

$$P(X \leqslant m) \geqslant \tfrac{1}{2} \quad \text{and} \quad P(X \geqslant m) \geqslant \tfrac{1}{2}, \tag{1}$$

or, equivalently,

$$F(m) \geqslant \tfrac{1}{2} \quad \text{and} \quad F(m_-) \leqslant \tfrac{1}{2}. \tag{2}$$

For any random variable, the median is either unique, or there exists a closed bounded interval $[a, b]$, such that all points of $[a, b]$ are medians.

We will use the notation m or $\mathrm{Med}(X)$ for any median of X.

Lemma 1.2. *Let X be a random variable and let I be a closed interval.*

(i) *If $P(X \in I) = \frac{1}{2}$, then at least one median of X is in I.*
(ii) *Suppose that I has the following property: If J is any closed interval such that I is a proper subset of J, then $P(J) > \frac{1}{2}$. Then I contains all medians of X.*

Proof. (i) Let F be the distribution function of X. If $I = (-\infty, b]$, then $F(b) = \frac{1}{2}$ by assumption, and so, b is a median of X. If $I = [a, +\infty)$, then $F(a_-) = \frac{1}{2}$ and again it follows that a is a median of X. If $I = [a, b]$, then $F(a_-) \leqslant \frac{1}{2}$ and $F(b) \geqslant \frac{1}{2}$, hence, at least one median of X is between a and b.

(ii) Suppose that m is a median of X and $m \notin I$. The assumptions imply that there exists a closed interval $J \supset I$, such that $P(X \in J) > \frac{1}{2}$ and $m \notin J$. Then J is disjoint with one of sets $A = [m, +\infty)$ or $B = (-\infty, m]$, and so either $P(X \in A) < \frac{1}{2}$ or $P(X \in B) < \frac{1}{2}$, which implies that m is not a median. Therefore, all medians of X are in I. □

Definition 1.3. A real-valued function f, defined on $\mathbb{R}$, will be called a C-function if for any $t \in \mathbb{R}$, the set

$$f^{-1}((-\infty, t]) = \{x \in \mathbb{R} \mid f(x) \leqslant t\}$$

is a closed interval, a singleton, or an empty set.

Remark 1.4. A function f is lower semicontinuous if $f^{-1}((-\infty, t])$ is a closed set, for any $t \in \mathbb{R}$. Thus, any C-function is lower semicontinuous on $\mathbb{R}$, and so (Schechter, 1997), $f(x) \leqslant \liminf_{y \to x} f(y)$, for any $x \in \mathbb{R}$. In this special case, however, we have a stronger property described in the following lemma.

Lemma 1.5. *If f is a C-function, then finite left and right limits exist at any point $x \in \mathbb{R}$ and*

$$f(x) \leqslant \min(f(x_-), f(x_+)).$$

Proof. Suppose that f does not have a left limit at some point $x \in \mathbb{R}$. Let l and L be left inferior and superior limit, respectively, $l < L$. Then for any $t \in (l, L)$, the set $f^{-1}((-\infty, t])$ is not an interval, because in any left neighborhood of x, there are points x' with $f(x') > t$, as well as points x'' with $f(x'') < t$. Therefore, f is not a C-function. The same argument holds for the right limit. Hence, any C-function has left and right limits at any point. Let us show that they are finite.

If f is a C-function, then due to Remark 1.4 above, the left and right limits at any point must be greater than $-\infty$. Suppose that $f(x_-) = +\infty$ for some x. Let $x' < x$ and $f(x') = K$. Let $I_K = f^{-1}((-\infty, K])$. Then $x' \in I_K$, and there is an interval $(x - \delta, x)$ which is not in I_K. Since I_K is an interval, it follows that $f(x) > K$. However, x' can be chosen in such a way that K is arbitrary large, so it follows that $f(x) = +\infty$, which is a contradiction. A similar argument shows that $f(x_+) < +\infty$. □

Lemma 1.6. *If f is convex on $\mathbb{R}$, then it is a C-function.*

Proof. For any $t \in \mathbb{R}$, let $I_t = f^{-1}((-\infty, t])$. If $a, b \in I_t$, $a < b$, then for any $\lambda \in (0, 1)$, we have that

$$f(\lambda a + (1-\lambda)b) \leqslant \lambda f(a) + (1-\lambda)f(b) \leqslant \lambda t + (1-\lambda)t = t,$$

hence $\lambda a + (1-\lambda)b \in I_t$ and therefore, I_t is an interval. Since any function convex on $\mathbb{R}$ must be continuous (see Mitrinović, 1970), I_t has to be a closed interval. □

Other examples of C-functions 1.7. It is easy to see that any monotone and continuous function is a C-function. Further examples include functions $x \mapsto |x-a| \pm |x-b|$, for any real a, b. Finally, any continuous function which is nonincreasing on $(-\infty, \theta)$ and nondecreasing on $(\theta, +\infty)$ is a C-function, for a fixed θ. This property is satisfied with any loss function $x \mapsto L(x, \theta)$ which is of use in statistics.

Theorem 1.8 (*Jensen's inequality for medians*). *Let f be a C-function and let X be any real random variable. If* $\mathrm{Med}(X) = m$ *is unique, then*

$$f(m) \leqslant \mathrm{Med} f(X), \tag{3}$$

where $\mathrm{Med} f(X)$ *is any median of $f(X)$. If* $\mathrm{Med} f(X)$ *is unique, then* (3) *holds for any median m of X. In general, for any median of $f(X)$ there exists a median m of X such that* (3) *holds.*

Proof. Let M be a median of $f(X)$, and let $I_M = f^{-1}((-\infty, M])$. Then

$$P(X \in I_M) = P(f(X) \leqslant M) \geqslant \tfrac{1}{2},$$

and, by Lemma 1.2, there exists a median m of X in I_M, which yields $f(m) \leqslant M$. So, we proved that for any median of $f(X)$ there is a median of X such that (3) holds, which, in particular, implies that, if $\mathrm{Med}(X)$ is unique, then (3) holds for any median of $f(X)$. Now suppose that $M = \mathrm{Med} f(X)$ is unique. Then I_M has the property as in (ii) of Lemma 1.2, and, consequently, it contains all medians of X, which means that (3) in this case holds for all medians m of X. □

2. A characterization of median

Theorem 2.1. *For any random variable X, the set of its medians* (*a point or a closed interval*) *coincides with the intersection of all closed intervals I that have the following property*: *If J is any closed interval such that I is a proper subset of J, then $P(J) > \frac{1}{2}$.*

Proof. Let I_M denote the set of all medians of X. It is proved in (ii) of Lemma 1.2 that $I_M \subset I$, for every closed interval I that has a property as stated in the theorem. On the other hand, if $I_M = [a, b]$, $a \leqslant b$, then intervals $(-\infty, b]$ and $[a, +\infty)$ both have the stated property and their intersection is I_M. This shows that the intersection of all such intervals is a subset of I_M, and the proof is finished. □

Remark 2.2. Theorem 2.1 gives a novel possibility of extending the definition of medians in higher dimensions. There have been many attempts to extend the concept of median to multivariate distributions or data (see Chakraborty and Chaudhuri, 1996; Small, 1990 and references therein). We need a definition and a technical lemma before proceeding further. In d-dimensional Euclidean space $\mathbb{R}^d$ we will denote the points by $\mathbf{x}$ and their coordinates by x_i, $i = 1, \dots, d$. The space $\mathbb{R}^d$ together with all points $\mathbf{x}$ with $x_i = \pm\infty$ for some i, will be denoted $\overline{\mathbb{R}}^d$.

Definition 2.3. Let $\mathbf{a}$, $\mathbf{b}$ be two points in $\overline{\mathbb{R}}^d$. A subset I of $\mathbb{R}^d$ will be called a closed interval, to be denoted by $[\mathbf{a}, \mathbf{b}]$, if

$$I = \{(x_1, \dots, x_d) \in \mathbb{R}^d \mid a_i \leqslant x_i \leqslant b_i,\ i = 1, \dots, d\}.$$

If a_i, b_i are finite for all $i = 1, 2, \dots, d$, the interval is compact.

For example, in $\mathbb{R}^2$, closed intervals are all closed parallelograms with sides parallel to axes, as well as all sets that can be obtained from such parallelograms by letting one or more side lengths to converge to infinity or to zero. In particular, any point, or a line parallel to one of axes, or an empty set, are closed intervals.

In $\overline{\mathbb{R}}^d$, we can introduce a partial order by

$$\mathbf{a} \leqslant \mathbf{b} \iff a_i \leqslant b_i \text{ for all } i = 1, \dots, d.$$

Then the closed interval $[\mathbf{a}, \mathbf{b}]$ can be defined as the set of points $\mathbf{x}$ such that $\mathbf{a} \leqslant \mathbf{x} \leqslant \mathbf{b}$. See (Schechter, 1997, pp. 78–80) for more properties of intervals.

Lemma 2.4. *Let* $\mathscr{I} = \{I_\alpha = [\mathbf{a}^\alpha, \mathbf{b}^\alpha], \alpha \in A\}$, *be a collection of closed intervals in* $\mathbb{R}^d$, *where* A *is an index set. Assume that there is at least one* α *such that* $[\mathbf{a}^\alpha, \mathbf{b}^\alpha]$ *is a compact interval. If* $I_\alpha \cap I_\beta \neq \emptyset$ *for each* $\alpha, \beta \in \mathscr{I}$, *then the intersection of all sets in* $\mathscr{I}$ *is a non-empty compact interval* $[\mathbf{a}, \mathbf{b}]$, *where* $\mathbf{a}$ *and* $\mathbf{b}$ *are finite points.*

Proof. Suppose that each two intervals in $\mathscr{I}$ have a non-empty intersection. This implies that each interval is non-empty, that is, $\mathbf{a}^\alpha \leqslant \mathbf{b}^\alpha$, for any $\alpha \in A$. Further, it is easy to see that intervals I_α and I_β, $\alpha \neq \beta$ have a non-empty intersection if and only if $\mathbf{a}^\alpha \leqslant \mathbf{b}^\beta$ and $\mathbf{a}^\beta \leqslant \mathbf{b}^\alpha$. Hence, it follows that for each $\alpha, \beta \in A$ and each $k = 1, \dots, d$, we have that

$$a_k^\alpha \leqslant b_k^\beta.$$

Now, let

$$a_k = \sup_{\alpha \in A} a_k^\alpha, \quad b_k = \inf_{\alpha \in A} b_k^\alpha, \quad k = 1, \dots, d.$$

From the assumptions it follows that both a_k and b_k are finite for all k. Let $\mathbf{a} = (a_1, \dots, a_d)$ and $\mathbf{b} = (b_1, \dots, b_d)$. Then $[\mathbf{a}, \mathbf{b}]$ belongs to all intervals in $\mathscr{I}$. □

Lemma 2.5. *Let* X *be a random variable with values in* $\mathbb{R}^d$. *Let* $\mathscr{I}$ *be the family of all closed intervals with the following property: If* J *is any closed interval such that* I *is a proper subset of* J, *then* $P(X \in J) > \frac{1}{2}$. *Then the intersection of all intervals in* $\mathscr{I}$ *is a non-empty compact interval.*

Proof. Clearly, for any random variable X, there exists a compact interval I such that $P(X \in I) > \frac{1}{2}$; then $I \in \mathscr{I}$. Due to Lemma 2.4, we only need to show that each two intervals in $\mathscr{I}$ have a non-empty intersection. Indeed, suppose that $I_1, I_2 \in \mathscr{I}$ and that $I_1 \cap I_2 = \emptyset$. Then there exist two closed disjoint intervals J_1 and J_2 with the property that $P(X \in J_i) > \frac{1}{2}$, which is impossible.

Definition 2.6. Let X be a random variable with values in $\mathbb{R}^d$. The intersection of all sets of the family $\mathscr{I}$ as in Lemma 2.5, is called median set of X. The median set is a non-empty compact interval (in the sense of Definition 2.3), which can degenerate into a line or a point.

Properties of median sets and their determination in multivariate distributions and data sets will be investigated in a separate work. Here we will use the above definition to discuss an extension of Jensen's inequality in a multivariate setup.

Jensen's inequality for medians—multivariate case 2.7. Jensen's inequality for medians—multivariate case In view of previous work, Definition 1.3 can be adopted to describe a real-valued C-function defined on $\mathbb{R}^d$, $d > 1$. However, here we lose a connection with convexity, because a convex function need not be a C-function. For example, the function $(x, y) \mapsto x^2 + y^2$ is convex, but not a C-function. By inspection of the proof of Theorem 1.8, one can see that the theorem holds for any multivariate random variable X and any multivariate C-function, where $\mathrm{Med}(X)$ is defined as in Definition 2.6.

References

Basu, S., DasGupta, A., 1997. The mean, median and mode of unimodal distributions: a characterization. Theory Probab. Appl. 41 (2).

Chakraborty, B., Chaudhuri, P., 1996. On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. Proc. Am. Math. Soc. 124, 2539–2547.

Dharmadhikari, S.W., Joag-Dev, K., 1988. Unimodality, Convexity and Applications. Academic Press, New York.

Mitrinović, D.S., 1970. Analytic Inequalities. Berlin-Heidelberg-New York.

Schechter, E., 1997. Handbook of Analysis and Its Foundations. Academic Press, New York.

Small, C.G., 1990. A survey of multidimensional medians. Internat. Statist. Rev. 58, 263–277.