Ann. Inst. Statist. Math.
33 (1981), Part A, 135-140

A NOTE ON THE MEDIAN OF A DISTRIBUTION

DALE UMBACH

(Received Dec. 18, 1978; revised Oct. 20, 1980)

Summary

Let F be a distribution function over the real line. Define $R_p(y) = \int |x-y|^p dF(x)$ for $p \geqq 1$. For $p>1$ there is a unique minimizer of $R_p(\cdot)$, say γ_p. Under mild conditions on F it is shown that $\lim_{p\to 1} \gamma_p$ exists and that the limit is a median. Thus, one could use this limit as a definition of a unique median. Also it is shown that $\lim_{p\to\infty} \gamma_p = (R+L)/2$ where R is the right extremity of F and L is the left extremity of F provided that $-\infty < L \leqq R < \infty$. A similar result is available if $L = -\infty$, $R = \infty$, yet F has symmetric tails.

1. Introduction

Suppose F is a distribution function, d.f., over the real line. The number m is called a median of F if $F(m)=F(m+0) \geqq \frac{1}{2}$ and $F(m-0) \leqq \frac{1}{2}$. Consider $R_p(y) = \int |x-y|^p dF(x)$ for $p \geqq 1$. It is well known that $R_1(y)$ is minimized at y_0 if and only if y_0 is a median, and that $R_2(y)$ is minimized when $y=\mu$, the mean of F. If $p>1$, then $R_p(y)$ is strictly convex and has a unique minimizer, say γ_p, characterized by

$$\int_{(\gamma_p, \infty)} (x-\gamma_p)^{p-1} dF(x) = \int_{(-\infty, \gamma_p)} (\gamma_p - x)^{p-1} dF(x) .$$

See DeGroot and Rao [1] for the details on these remarks.

Jackson [2] has shown that if F can be written as $F(x) = \frac{1}{n} \sum_{i=1}^{n} I_{[b_i, \infty)}(x)$ for some $-\infty < b_1 \leqq b_2 \leqq \cdots \leqq b_n < \infty$, then $\lim_{p\to 1} \gamma_p$ exists and the limit say γ, is characterized by; $\gamma = b_{(n+1)/2}$ (the unique median) if n is odd, and $(\gamma-b_1)(\gamma-b_2)\cdots(\gamma-b_{n/2}) = (b_{(n/2)+1}-\gamma)\cdots(b_n-\gamma)$ for n even. In addition

Key words: Median, minimum risk.

Jackson has shown that $\lim_{p\to\infty} \gamma_p = \frac{1}{2}(b_1+b_n)$.

In this note similar results are given for distribution functions in a large class.

2. The case when $p \to 1$

Suppose F is a d.f. and that $\int |x|^r dF(x) < \infty$ for some $r>1$. Then $R_p(y)<\infty$ for all real y and $1 \leqq p \leqq r$. Also, for $1<p\leqq r$, γ_p, the minimizer of $R_p(\cdot)$, exists and is unique since $R_p(\cdot)$ is strictly convex for $p>1$ and $\lim_{|y|\to\infty} R_p(y)=\infty$. In fact, DeGroot and Rao [1] have shown that

$$R_p'(y)/p = \int_{(-\infty,y)} (y-x)^{p-1} dF(x) - \int_{(y,\infty)} (x-y)^{p-1} dF(x)$$

exists for each y. Thus the minimizer is that value of y for which $R_p'(y)/p=0$. Note that if $R_p'(y)>0$ then $\gamma_p<y$ and if $R_p'(y)<0$ then $\gamma_p>y$.

To show that $\lim_{p\to 1} \gamma_p$ exists it will be shown that for each y there exists $q(y)>1$ such that $\operatorname{sgn} R_p'(y)$ is constant for $1<p<q(y)$. A limit exists, then, by the following Bolzano-Weierstrass type argument. Suppose there exists a finite interval, $I=(L, U)$, such that $\gamma_p \in I$ for $1<p<s$ for some $s>1$. Define $a_1 = \frac{1}{2}(U+L)$. If $\operatorname{sgn} R_p'(a_1)=0$ for all p sufficiently small then $\gamma_p \to a_1$. If $\operatorname{sgn} R_p'(a_1)=1$ for all p sufficiently small then $L \leqq \gamma_p < a_1$. If $\operatorname{sgn} R_p'(a_1)=-1$ for all p sufficiently small then $a_1<\gamma_p\leqq U$. If $L\leqq \gamma_p<a_1$, then with $a_2=\frac{1}{2}(L+a_1)$ consider $\operatorname{sgn} R_p'(a_2)$. If $a_1<\gamma_p\leqq U$, then with $a_2=\frac{1}{2}(U+a_1)$ consider $\operatorname{sgn} R_p'(a_2)$. This will tell us whether $\gamma_p>a_2$ or $\gamma_p<a_2$. Etc.

To show that $\operatorname{sgn} R_p'(y)$ is constant for p sufficiently close to 1, it will be helpful to expand it in a series and interchange the order of integration and summation. In doing this the following class of d.f.'s is considered.

DEFINITION. A d.f., F, is said to be of order $s \geqq 0$ if for each real x we have $\lim_{\varepsilon\to 0} (F(x+\varepsilon)-F(x))/\varepsilon^s = 0 = \lim_{\varepsilon\to 0} (F(x-\varepsilon)-F(x-0))/\varepsilon^s$ where $\varepsilon \to 0$ through positive values.

If F is discrete in the sense that there is a fixed positive minimum distance between any two jump points, then F is of order s for each $s<\infty$. By virtue of right continuity every d.f. is of order $s=0$. If F has a bounded density then F is of order s for each $s<1$. If F has a bounded density, f, except near x_0, and in a neighborhood of x_0,

$f(x) \leqq K|x-x_0|^{-r}$ for some $0<r<1$, then F is of order s for each $s<1-r$. The use to be made of such d.f.'s is that if F is of order $s>0$ then for each $0 \leqq q<s$ and each real y and each $\varepsilon>0$ we have $\int_{(y,y+\varepsilon)} (x-y)^{-q} \cdot dF(x)<\infty$. This is needed to establish sgn $R_p'(y)$ constant, as is seen in the following lemmas.

LEMMA 1. *If F is of order $s>0$ and $1<p<s+1$ and $R_p(y)<\infty$ then* $R_p'(y)/p=\sum_{n=0}^{\infty} \frac{(p-1)^n}{n!}\left[\int_{(-\infty,y)} (\ln (y-x))^n dF(x)-\int_{(y,\infty)} (\ln (x-y))^n dF(x)\right]$ $=\sum_{n=0}^{\infty} \frac{C_n(y)}{n!}(p-1)^n$.

PROOF. For $x>y$ we have $(x-y)^{p-1}=\sum_{n=0}^{\infty} \frac{[(p-1)\ln (x-y)]^n}{n!}$ and for $x<y$ we have $(y-x)^{p-1}=\sum_{n=0}^{\infty} \frac{[(p-1)\ln (y-x)]^n}{n!}$. Thus $R_p'(y)/p=$

$$\int_{(-\infty,y)} \sum_{n=0}^{\infty} \frac{[(p-1)\ln (y-x)]^n}{n!} dF(x) \tag{2.1}$$

$$-\int_{(y,\infty)} \sum_{n=0}^{\infty} \frac{[(p-1)\ln (x-y)]^n}{n!} dF(x)\,. \tag{2.2}$$

But since

$$\begin{aligned}\int_{(-\infty,y)} \sum_{n=0}^{\infty} &\left|\frac{[(p-1)\ln (y-x)]^n}{n!}\right| dF(x)\\ &=\int_{(-\infty,y)} \sum_{n=0}^{\infty} \frac{(p-1)^n}{n!}|\ln (y-x)|^n dF(x)\\ &=\int_{(-\infty,y)} e^{(p-1)|\ln (y-x)|} dF(x)\\ &=\int_{(-\infty,y-1]} (y-x)^{p-1} dF(x)+\int_{(y-1,y)} (y-x)^{1-p} dF(x)<\infty\,,\end{aligned}$$

the order of summation and integration can be interchanged to obtain (2.1)=

$$\sum_{n=0}^{\infty} \frac{(p-1)^n}{n!} \int_{(-\infty,y)} [\ln (y-x)]^n dF(x)\,.$$

One can similarly handle (2.2). Thus, $R_p'(y)/p=\sum_{n=0}^{\infty} \frac{C_n(y)}{n!}(p-1)^n$ where $C_n(y)=\int_{(-\infty,y)} [\ln (y-x)]^n dF(x)-\int_{(y,\infty)} [\ln (x-y)]^n dF(x)$.

LEMMA 2. *Under the conditions of Lemma 1 for each y there exists $q(y)>1$ such that* sgn $R_p'(y)$ *is constant for $1<p<q(y)$.*

PROOF. Let y be fixed. Let $n^*=\inf\{n : C_n(y)\neq 0\}$. If $n^*=\infty$, then $R_p'(y)/p=0$ for all $1<p<s+1$, and thus $\operatorname{sgn} R_p'(y)$ is constant. If $n^*<\infty$ then $\lim_{p\downarrow 1} \frac{R_p'(y)}{(p-1)^{n^*}p}=\lim_{p\downarrow 1}\sum_{n=n^*}^{\infty}\frac{(p-1)^{n-n^*}C_n(y)}{n!}=\frac{C_{n^*}(y)}{n^*!}\neq 0$. Thus there exists $q(y)>1$ such that $1<p<q(y)$ implies $\operatorname{sgn} R_p'(y)/p(p-1)^{n^*}$ is constant, which in turn implies that $\operatorname{sgn} R_p'(y)$ is constant.

Thus, if F is of some order $s>0$ and has a finite rth moment for some $r>1$, then we have that $\lim_{p\downarrow 1}\gamma_p$ exists provided we show that there exists a finite interval I such that $\gamma_p\in I$ for p sufficiently close to 1. This is established by showing that if M is any number larger than every median of F then $R_p'(M)>0$ for all p sufficiently close to 1 and thus $\gamma_p<M$; and that if N is any number smaller than every median then $R_p'(N)<0$ for p sufficiently close to 1, and thus $\gamma_p>N$. This will also establish that $\lim_{p\downarrow 1}\gamma_p$ is also a median of F. Thus $\lim_{p\downarrow 1}\gamma_p$ could be used as definition of a unique median. This is shown in Lemma 3.

LEMMA 3. *Suppose M is a number larger than every median. Then $\gamma_p<M$ for p sufficiently close to* 1.

PROOF. Note that for each y we have $\lim_{p\downarrow 1} R_p(y)=R_1(y)$. Let m be a median. Then we know $R_1(m)<R_1(M)$. But $R_p(M)\to R_1(M)$ and $R_p(m)\to R_1(m)$ as p decreases to 1. Thus for p sufficiently close to 1 we have $R_p(m)<R_p(M)$, which implies $\gamma_p<M$ since $m<M$ and $R_p(\cdot)$ is convex.

The case for N, a number smaller than any median, is handled similarly.

For example, suppose F has a density given by $f(x)=\frac{1}{2}I_{(-1,0)}^{(x)}+\frac{1}{8}I_{(1,5)}^{(x)}$. Each number between zero and one is a median. But since $R_p'(1)/p=\frac{1}{2p}[2^p-1-4^{p-1}]<0$ for $p>1$, we must have $\gamma_p>1$. Thus, $\lim_{p\downarrow 1}\gamma_p=1$ since the limit exists and must be a median.

Note that for absolute error loss, a Bayes estimator is given by the posterior median. If the above definition of the median is used the estimator is guaranteed measurable by being the a.e. limit of measurable functions.

3. The case when $p \to \infty$

The basic result of this section is that $\lim_{p\to\infty} \gamma_p = (R+L)/2$ where $R = \inf\{x \mid F(x)=1\}$ and $L = \sup\{x \mid F(x)=0\}$ when $-\infty < L \leqq R < \infty$. A similar result is available if $L=-\infty$ and $R=\infty$, yet F has symmetric tails. For γ_p to exist we need $R_p(y)<\infty$. So throughout this section we assume that F has all its moments finite. Under these conditions consider the following proposition.

PROPOSITION. Suppose there exist $R^*<\infty$ and $L^*>-\infty$ such that $L \leqq L^* \leqq R^* \leqq R$ and $F(R^*+x) = 1 - F(L^*-x-0)$ for each $x \geqq 0$. Then $\lim_{p\to\infty} \gamma_p = \frac{1}{2}(R^*+L^*)$.

PROOF. Let $Y = \frac{1}{2}(R^*+L^*)$. (Note that if $-\infty < L \leqq R < \infty$ then $\frac{1}{2}(R^*+L^*) = \frac{1}{2}(R+L)$.) Note that

$$R_p(Y) = \int_{(-\infty, L^*)} (Y-x)^p dF(x) + \int_{[L^*, Y)} (Y-x)^p dF(x) + \int_{(Y, R^*]} (x-Y)^p dF(x) + \int_{(R^*, \infty)} (x-Y)^p dF(x) .$$

The minimum of

$$\int_{(-\infty, L^*)} |y-x|^p dF(x) + \int_{(R^*, \infty)} |y-x|^p dF(x)$$

occurs at $y=Y$ by symmetry considerations. Thus for $p>1$ and $0<|\varepsilon| < \frac{1}{2}(R^*-L^*)$ we have

$$\int_{(-\infty, L^*)} (Y+\varepsilon-x)^p dF(x) + \int_{(R^*, \infty)} (x-Y-\varepsilon)^p dF(x) > \int_{(-\infty, L^*)} (Y-x)^p dF(x) + \int_{(R^*, \infty)} (x-Y)^p dF(x) .$$

But since $R_p(Y+\varepsilon) =$

$$\int_{(-\infty, L^*)} (Y+\varepsilon-x)^p dF(x) + \int_{[L^*, Y+\varepsilon)} (Y+\varepsilon-x)^p dF(x) + \int_{(Y+\varepsilon, R^*)} (x-Y-\varepsilon)^p dF(x) + \int_{[R^*, \infty)} (x-Y-\varepsilon)^p dF(x) ,$$

it is sufficient to show that for each such ε there exists $K<\infty$ such that $p \geqq K$ implies that

$$\int_{[L^*,Y+\varepsilon)}(Y+\varepsilon-x)^p dF(x)+\int_{(Y+\varepsilon,R^*]}(x-Y-\varepsilon)^p dF(x) \tag{3.1}$$

$$\geqq\int_{[L^*,Y)}(Y-x)^p dF(x)+\int_{(Y,R^*]}(x-Y)^p dF(x)\,. \tag{3.2}$$

But (3.2)$\leqq\left[\frac{1}{2}(R^*-L^*)\right]^p[F(R^*)-F(L^*-0)]$. Now note that without loss of generality we may assume that R^* and L^* are points of increase of F, for if not we could find a larger R^* and smaller L^* whose average remain the same and still have the properties above and are points of increase of F. Now suppose $\varepsilon<0$. Then (3.1)$=$

$$\begin{aligned}&\int_{[L^*,Y+\varepsilon)}(Y+\varepsilon-x)^p dF(x)+\int_{[Y+\varepsilon,R^*+\varepsilon/2]}(x-Y-\varepsilon)^p dF(x)\\&+\int_{(R^*+\varepsilon/2,R^*]}(x-Y-\varepsilon)^p dF(x)\,.\end{aligned} \tag{3.3}$$

Note that each term is non-negative and (3.3)$\geqq\left[\frac{1}{2}(R^*-L^*)-\varepsilon/2\right]^p$ $\cdot[F(R^*)-F(R^*+\varepsilon/2)]$. But since $F(R^*)-F(R^*+\varepsilon/2)>0$ and $-\varepsilon/2>0$ we have that for p sufficiently large

$$\begin{aligned}&\left[\frac{1}{2}(R^*-L^*)\right]^p[F(R^*)-F(L^*-0)]\\&\quad\leqq\left[\frac{1}{2}(R^*-L^*)-\varepsilon/2\right]^p[F(R^*)-F(R^*+\varepsilon/2)]\,.\end{aligned}$$

Thus for p sufficiently large (3.1)$\geqq$(3.2). One can handle the case $\varepsilon>0$ in a similar fashion. Thus the proof is completed.

In particular, if $-\infty<L\leqq R<\infty$, then F will have all its moments and $\gamma_p\to\frac{1}{2}(R+L)$ as $p\to\infty$.

Acknowledgement

Many thanks are due the referee for his careful reading of this paper.

BALL STATE UNIVERSITY

REFERENCES

[1] DeGroot, M. H. and Rao, M. M. (1963). Bayes estimation with convex loss, *Ann. Math. Statist.*, **34**, 839-846.

[2] Jackson, D. (1921). Note on the median of a set numbers, *Bull. Amer. Math. Soc.*, **27**, 160-164.