Real Analysis Exchange
Vol. 29(1), 2003/2004, pp. 131–138

Jan Mycielski, Department of Mathematics, University of Colorado, Boulder, CO 80309. email: jmyciel@euclid.colorado.edu

Pawel Szeptycki, Department of Mathematics, University of Kansas, Lawrence, KS 66045. email: szeptyck@math.ukans.edu

MINIMIZING MOMENTS

Abstract

We will prove a certain characterization of the function x^2 and of some similar functions, in the style of Cauchy's characterization of the function ax by its additivity and boundedness over any interval of positive length.

The following proposition is fundamental in statistics. If X is a random variable such that the expected value $E(X^2) < \infty$, then the function $E((X-t)^2)$ of the real variable t attains its minimum at the point $t = E(X)$. Matatyahu Rubin conjectured that if f is a function such that for every X, $E(f(X-t))$ attains its minimum at $E(X)$, then f is of the form $f(x) = \alpha x^2 + \beta$, where α, β are constants and $\alpha \geq 0$. The purpose of this note is to generalize and prove this conjecture. The generalization is two-fold. First, the assumption will be restricted to two-valued random variables. Second, the function x^2 will be replaced by any even function $F(x)$ which is either strictly convex or equals $|x|$. This generalization will be made precise in Theorem 2 below.

Our first observation is the following.

Theorem 1. *If F is an even and strictly convex function, then for every random variable X such that $E(F(mX)) < \infty$ for some $m > 1$, the function $E(F(X-t))$ is defined and continuous for all real t and attains its minimum at a unique point t_0 which will be denoted by $\xi_F(X)$.*

Proof. The hypotheses on F and on X imply that

$$\begin{aligned} E(F(X-t)) &= E(F(\frac{1}{m}mX + \frac{m-1}{m}(-\frac{m}{m-1}t))) \\ &\leq \frac{1}{m}E(F(mX)) + \frac{m-1}{m}F(-\frac{m}{m-1}t) < \infty. \end{aligned}$$

Key Words: Continuity, convexity, moments, percentiles, variance
Mathematical Reviews subject classification: 26A51, 60A99
Received by the editors September 14 2002
Communicated by: Krzysztof Chris Ciesielski

To prove continuity, suppose that $t_n \to t$. Then

$$F(X - t_n) \leq \frac{1}{m} F(mX) + \frac{m-1}{m} \sup_n F(\frac{m}{m-1} t_n)$$

and by the dominated convergence theorem $E(F(X - t_n)) \to E(F(X - t))$. Without loss of generality we may assume that $F(0) = 0$. Then for $2^l \leq |t| < 2^{l+1}$ we can write

$$F(X - t) = F(t(\frac{X}{t} - 1)) \geq 2F(\frac{t}{2}|\frac{X}{t} - 1|) \geq 2^l F(2^{-l}t|\frac{X}{t} - 1|) \geq 2^l F(\frac{X}{t} - 1),$$

and by the dominated convergence theorem, $E(F(\frac{X}{t} - 1)) \to F(-1)$ and $E(F(X - t)) \to \infty$ as $|t| \to \infty$. It follows that the minimum of $E(F(X - t))$ is attained at some point t. This t is unique by the strict convexity of F. $\square$

From now on let X be a two-valued random variable such that

$$P(X = a) = p,\ P(X = b) = q = 1 - p.$$

We cannot expect an explicit general formula for the functional $\xi_F(X)$ in terms of F and the parameters a, b, and p. But for some special F this is possible. We will show (see Step 1 in the proof of Theorem 2) that if $F(x) = |x|^c$ with $c > 1$, then the functional $\xi_F(X) = \xi_c(X)$ is given by the formula

$$\xi_c(X) = \frac{p^r a + q^r b}{p^r + q^r} = \lambda a + (1 - \lambda) b,$$

where $r = \frac{1}{c-1}$ and $\lambda = \frac{p^r}{p^r + q^r}$. Then, of course, $\xi_2(X) = E(X)$.

Also, if $F(x) = e^{|x|}$ and $a < b$,

$$\xi_F(X) = \frac{a + b}{2} + \frac{1}{2} \log\Big(\frac{q}{p}\Big),$$

if $a - b \leq \log\big(\frac{q}{p}\big) \leq b - a$ and $\xi_F(X) = a$ or b if $\log\big(\frac{q}{p}\big)$ lies outside of this interval.

Beside all functions F satisfying the hypotheses of Theorem 1 we also consider the function $F(x) = |x|$ which is convex but not strictly convex. The latter was suggested to us by Fred S. Van Vleck. In this case, a fact of importance in statistics, $E(|X - t|)$ is minimized by any median of X, i.e., by every number m such that

$$P(X \leq m) \geq \frac{1}{2} \text{ and } P(X \geq m) \geq \frac{1}{2}.$$

Let MX denote the set of all medians of X.

For a two-valued X, as above, $MX = \{a\}$, if $p > \frac{1}{2}$, $MX = \{b\}$, if $q > \frac{1}{2}$, and $MX = [a, b]$, if $p = q = \frac{1}{2}$.

Our generalization of Rubin's conjecture stated at the beginning is the following.

Theorem 2. *Let F be even and strictly convex. Let f be such that for every two-valued random variable X, the expected value $E(f(X-t))$ attains its minimum at $t = \xi_F(X)$. Then $f(x) = \alpha F(x) + \beta$ where α and β are constants, $\alpha \geq 0$. The conclusion also holds if $F(x) = |x|$ and $E(f(X-t))$ attains its minimum at every $t \in MX$.*

Remark 1. As it will be apparent from the proof, in the case when f and F are known a priori to be differentiable, the phrase $E(f(X-t))$ *attains its minimum at* $t = \xi_k(X)$ could be replaced by $E(f(X-t))$ *has a critical point at* $t = \xi_k(X)$, in which case we do not claim that $\alpha \geq 0$.

Remark 2. As already mentioned, the functional $\xi_k(X) = \xi_F(X)$, where $F(x) = |x|^k$ with $k \neq 1$, appears to be a natural generalization of the median and of the mean of X. In particular it seems to be natural to define the k-th *central moment* of X as $E(|X - \xi_k(X)|^k)$ and not as $E(|X - \xi_2(X)|^k)$ which appears sometimes in the literature.

First let us point out some simple properties of the functional $\xi_F(X) = \xi(a, b, p)$. By definition this is the unique real t minimizing $pF(a-t)+qF(b-t)$. In other words $\xi = \xi_F(X)$ iff

$$pF(a-t) + qF(b-t) \geq pF(X-\xi) + qF(b-\xi) \tag{1}$$

for all real t. Since F is even and convex, for every real t,

$$F(\frac{a-b}{2}) + F(\frac{b-a}{2}) = 2F(\frac{b-a}{2}) \leq F(a-t) + F(b-t).$$

Also, since $F(x)$ is strictly decreasing for $x < 0$ and strictly increasing for $x > 0$, we get the following assertion.

Lemma 1. *$a < \xi_F(X) < b$, $\xi_F(X) = \frac{a+b}{2}$ if $p = q = \frac{1}{2}$, $\xi_F(X) = a$ if $p = 1$ and $\xi_F(X) = b$ for $q = 1$.*

For fixed a and b consider the function $\xi(p) = \xi_F(X) = \xi(a, b, p)$. If $p_n \to p_0$ and if $\xi(p_n) \to \eta$, then passing to the limit in both sides of the inequality (1) we conclude that $\eta = \xi(p_0)$ and by the compact graph theorem we have the next lemma.

Lemma 2. *For fixed $a < b$ the function $p \to \xi(p)$ is a continuous function of $p \in [0,1]$.*

We will need this lemma in the form of the intermediate value property.

Corollary 1. *For every pair $x_0, x > 0$ there exist $b > 0$ and $p \in (0,1)$ such that the two-valued random variable X with $P(X = 0) = p$, $P(X = b) = q$ satisfies $\xi_F(X) = x_0$ and $b - \xi_F(X) = x$.*

Indeed, it suffices to take $b = x_0 + x$. Then $0 < x < b$ and by continuity of ξ there is a $p \in (0,1)$ such that $\xi_F(X) = x$.

Corollary 2. *For $0 < p < 1$, $\xi(a,b,p) - a$ assumes all values in the interval $(0, b-a)$.*

We will also use the following.

Lemma 3. *With the notations as above we have $\xi(-b,-a,q) = -\xi(a,b,p)$.*

This is an immediate consequence of the assumption that F is even. Our next lemma is of more general nature.

Lemma 4. *Suppose that f is a real function defined and bounded on an interval $[a,b]$ and satisfying the condition $f(\frac{x+y}{2}) \leq \frac{f(x)+f(y)}{2}$ for all $x, y \in [a,b]$. Then f is continuous on $[a,b]$.*

The conclusion is well known if f is additive and bounded on some interval. For the sake of completeness we give the following simple argument. Let x be arbitrary and $x_n, y_n \to x$ as $n \to \infty$ be such that $f(y_n) \to \liminf_{y \to x} f(y) = l$ and $f(\frac{x_n+y_n}{2}) \to \limsup_{y \to x} f(y) = L$. We use the assumption on f to conclude that

$$L = \lim_{n\to\infty} f(\frac{x_n + y_n}{2}) \leq \frac{1}{2}(\lim_{n\to\infty} f(y_n) + \limsup_{n\to\infty} f(x_n)) \leq \frac{l+L}{2},$$

so that $l = L$. Writing $f(x) = f(\frac{x+\epsilon+x-\epsilon}{2}) \leq \frac{f(x+\epsilon)+f(x-\epsilon)}{2}$ and taking the $\liminf_{\epsilon\to 0}$ we get $f(x) \leq L$. Also, with $x_n \to x$ such that $f(x_n) \to l$ we conclude, taking the $\liminf_{n\to\infty}$ in the inequality $f(\frac{x+x_n}{2}) \leq \frac{f(x)+f(x_n)}{2}$, that $f(x) \geq l$. It follows that $l = f(x) = L$ and f is continuous at x.

PROOF OF THEOREM 2. We may assume that f is not a constant, otherwise the conclusion is trivial.

We begin with the proof in the case when $F(x) = |x|$. In this case

$$E(f(X-t)) = pf(a-t) + qf(b-t)$$

is minimized by $t = a$ if $p > \frac{1}{2}$, $t = b$ when $p < \frac{1}{2}$ and by any t of the form $sa + (1-s)b$, $0 \le s \le 1$ for $p = \frac{1}{2}$. In particular, for $p = \frac{1}{2}$,

$$f(a-t) + f(b-t) \ge f(a-a) + f(b-a),$$

which for $t = b$ implies $f(b-a) \ge f(a-b)$. Similarly we get the reverse inequality to conclude that f is even. Also, using $p = \frac{1}{2}$, $t = 0$ and $s = \frac{1}{2}$ we get $\frac{f(a)+f(b)}{2} \ge f(\frac{a+b}{2})$. In particular $f(x) \ge f(0)$ for all x. Replacing $f(x)$ by $f(x) - f(0)$ we may assume that $f(0) = 0$. Again with $p = \frac{1}{2}$ the hypothesis on f implies that for $0 \le s \le 1$,

$$f(a - sa - (1-s)b)) + f(b - sa - (1-s)b) = f((1-s)(a-b)) + f(s(b-a)$$

is constant and equals $f(b-a) = f(a-b)$. We can rewrite this again as $f(y) = f((1-s)y) + f(sy)$, where $y = b-a$, because f is even. If $A, B > 0$, then letting $y = A+B$ and $s = A/(A+B)$ we conclude that $f(A+B) = f(A)+f(B)$ provided A and B are of the same sign. In particular, for every M and every $x \in [0, M]$ we have $0 \le f(x) = f(M) - f(M-x) \le f(M)$ so that f is bounded on any finite interval in $[0, \infty)$. By Lemma 4 f is continuous on $[0, \infty)$ and being additive it must be of the form $f(x) = \alpha x$ for $x > 0$. Since it is even, the conclusion that $f(x) = \alpha|x|$ follows readily.

We are left with the case when F is strictly convex. We write f as the sum of its even and odd parts: $f = f_e + f_o$ where $f_e(x) = \frac{f(x)+f(-x)}{2}$ and $f_o(x) = \frac{f(x)-f(-x)}{2}$. Let us show the following.

Lemma 5. *If f satisfies the assumption of Theorem 2, then so does f_e .*

Proof. We rewrite the hypothesis on f in the form of the inequality

$$pf(a-t) + qf(b-t) \ge pf(a - \xi(a,b,p)) + qf(b - \xi(a,b,p)), \qquad (2)$$

for all t, a, b and $p \in [0, 1]$. By Lemma 2 the same inequality is true if $f(x)$ is replaced by $f(-x)$. Adding those inequalities side by side, we conclude that (2) holds also if f is replaced by f_e. Thus Lemma 5 is proved. □

Now let us show that f_e is continuous. Letting $a = b$ in (2) we see that f_e is bounded from below by $f(0) = f_e(0)$. We may assume that $f(0) = 0$. Again, letting $t = b$ in (2) and observing that, by Corollary 2, for $0 < p < 1$, $\xi(a, b, p) - a$ assumes all values in the interval $(0, b-a)$ we conclude that in this interval f_e is bounded from above by $f_e(b-a)$. It follows that f_e is bounded on any finite interval and by Lemma 3, f_e is continuous.

The remainder of the argument is now divided into into 3 steps:

1) f is everywhere continuous and continuously differentiable except possibly on a countable set.

2) f is even.

3)f is an arbitrary function.

Step 1). For a strictly convex F the one-sided derivatives $F'_+(x)$ and $F'_-(x)$ exist for every x, and for $x < y$

$$F'_-(x) \leq F'_+(x) < F'_-(y) \leq F'_+(y).$$

In particular the derivative F' exists except possibly on an at most countable set where it has positive jumps. Denote this set by S. For a two-valued random variable X with parameters a, b, p, $\xi = \xi_F(X)$ is the solution t of the equation $\frac{d}{dt}E(F(X-t)) = 0$; i.e.,

$$pF'(a-\xi) + qF'(b-\xi) = 0, \tag{3}$$

provided $a-\xi$ and $b-\xi$ are not in E. The hypothesis on f implies that for all a, b, p,

$$pf'(a-\xi) + qf'(b-\xi)) = 0, \tag{4}$$

if ξ is the unique solution of (3) and both $a-\xi$ and $b-\xi$ are outside a set S' where the derivative f' fails to exist. Fix now $x_0 > 0$ outside $S \cup S'$ and for an arbitrary x outside this set apply Corollary 1. Equation (3) implies that $\frac{p}{q} = -\frac{F'(x)}{F'(-x_0)}$. Substituting into (4) we get

$$f'(x) = -\frac{p}{q}f'(-x_0) = \frac{f'(-x_0)}{F'(-x_0)}F'(x).$$

It follows that f differs by a constant from a constant multiple of F which is the conclusion of the Theorem.

Observe also that in the cases when $F(x) = |x|^c$, $c > 1$, and $F(x) = e^{|x|}$ equation (3) implies the formulas for $\xi_F(X)$ announced earlier in this paper.

Step 2). In this case $f = f_e$ and as noticed at the beginning of the proof, f is continuous. The inequality $f(x) + f(y) \geq 2f(\frac{x+y}{2})$, together with the continuity of f implies that f is convex. But then f is differentiable outside of a countable set and the one-sided derivatives of f exist at every point of that exceptional set. Thus by Step 1) we get the desired result.

Step 3). The main idea is to write $f = f_e + f_o$, use Step 2) to get $f_e = \alpha F + \beta$ and then show that $f_o = 0$. By Lemma 5 f_e satisfies the hypotheses of the theorem and by Step 2), $f_e = \alpha F + \beta$. We can assume without loss of generality that $\alpha = 1$ and $\beta = 0$. Hence $f = F + f_o$. Now we consider condition (2) with

$p = q = \frac{1}{2}$. Then, by Lemma 1, $\xi = \frac{a+b}{2}$ and with $a - t = x$ and $b - t = -y$ we get

$$F(x) + F(y) - 2F(\frac{x+y}{2}) + f_o(x) - f_o(y) \geq f_o(\frac{x+y}{2}) + f_o(-\frac{x+y}{2}) = 0.$$

By symmetry in x, y this implies that $|f_o(x) - f_o(y)| \leq F(x) - 2F(\frac{x+y}{2}) + F(y)$. Let $x = y + 2h$ and denote by Δ_h the operator of difference with increment h. Then the last inequality can be written as

$$|\Delta_{2h} f_o(y)| \leq \Delta_h^2 F(y). \tag{5}$$

(5) implies that f_o is continuous.

Assume for a moment that F is continuously differentiable. Then dividing both sides of (5) by h and letting $h \to 0$ we conclude (applying l'Hôpital's theorem to the right hand side of the inequality) that the derivative f_o' exists and vanishes identically. Thus f_o is a constant and since f_o is odd, it is 0. This concludes Step 3) for a differentiable F.

Now we will reduce the general case to the case of a continuously differentiable F. This is done by regularization. Let $\varphi \geq 0$ be an arbitrary, continuously differentiable function on $\mathbb{R}$ vanishing outside of the interval $[-1, 1]$ and satisfying $\int_{\mathbb{R}} \varphi(y)dy = 1$. For a continuous g define

$$(\varphi \star g)(x) = \int_{\mathbb{R}} \varphi(x - y)g(y)dy.$$

Then $\varphi \star g$ is continuously differentiable and satisfies $\varphi \star (\Delta_h g) = (\Delta_h \varphi) \star g = \Delta_h(\varphi \star g)$. Also the functions of the form $\varphi \star g$ approximate g uniformly on any finite interval. In particular, if they are all constant, then so is g. With this in mind we apply the operator $\varphi\star$ to both sides of (5) - this is legitimate since f_o is continuous. We get

$$|\Delta_{2h}(\varphi \star f_o)(x)| = |\varphi \star \Delta_{2h} f_o(x)| \leq \varphi \star |\Delta_{2h} f_o(x)| \leq \varphi \star \Delta_h^2 F(x) = \Delta_h^2(\varphi \star F)(x).$$

Since now $\varphi \star F$ is continuously differentiable, the previous argument allows us to conclude that $\varphi \star f_o$ is constant and so is f_o. This concludes Step 3) of the proof. □