Ulam Quarterly — Volume 2, Number 1, 1993

On the Extrema of the Expected Values of Functions of Independent Identically Distributed Random Variables

Daoud Bshouty
Technion-Israel Institute of Technology
Haifa (3200), Israel

Walter Hengartner
Université Laval
Quebéc, Canada G1K 7P4

Vijay K. Rohatgi
Bowling Green State University
Bowling Green, OH 43403

and

Gábor J. Székely
Mathematical Institute
Eötvös University
1088 Budapest, Hungary

Abstract

Let $X_1, X_2, \ldots, X_n$ be an arbitrary sequence of iid random variables such that $E|X_j| = 1$. The minimum m_n of $E|X_1 + X_2 + \ldots + X_n|$ is taken for random variables X_j whose support contain exactly two points. In particular we get that $m_2 = 1$ and $m_3 = 1.3316$. Our approach is variational and can be generalized to other symmetric convex kernels. Related results were obtained by Hoeffding (1955), Loève (1977) and Hildebrand (1984). An alternative probabilistic approach to the problem using a convexity argument is included in the appendix.

Introduction

Let $X_1, \ldots, X_n$ be a sequence of not necessarily identically distributed independent random variables and let K be a real valued function defined on $\Re^n$. Hoeffding (1955) considered the following extremal problem:

$$Max(Min)EK(X_1, \ldots, X_n)$$

under the constraints

$$Eg_{ij}(X_j) = c_{ij}; \quad 1 \le i \le k; \quad 1 \le j \le n,$$

where g_{ij} are given real valued functions on $\Re$ and where c_{ij} are given real numbers. He proved that, under rather general conditions, one needs to consider only discrete random variables X_j which are supported on at most $k+1$ points (see also Karlin and Shapley (1953), and Dubins (1962)). Hoeffding emphasized that the same optimization problem is much more difficult if $X_1, X_2, \dots, X_n$ are assumed to be *identically distributed* (the iid case). In fact, in the iid case, a result of Hartley and David (1954) shows that the least upper bound for $Emax(X_1, X_2, \dots, X_n)$ given EX_j and EX_j^2 cannot be approached arbitrarily close with discrete distributions having a bounded number of steps. For some special cases, where $n = 2$, the iid case was considered in Hoeffding and Shrikhande (1955). The main difficulty in the identically distributed case is the non-linear nature of Hoeffding's variation technique.

§1. A Mad Theorem

Several authors have studied the problem of minimization of the mean absolute derivation (MAD):

$$Min\, E|X_1 + \dots + X_n| \qquad (\star)$$

under the constraints

$$E|X_j| = 1; \quad 1 \le j \le n,$$

where $X_1, X_2, \dots, X_n$ are iid. Hoeffding and Shrikhande (1955) considered the case $n = 2$ using direct techniques of variations which lead to the study of quadratic functions. Loève (1977, prob. 15, p. 278) studied the case where X is centered at the median which enabled him the use of symmetrization techniques. More recently, Hildebrand (1984), using Jensen's inequality, studied the case $n = 2$ where X_j is symmetric and centered at its expectation. In the above cases the extremal distribution function is $T = \frac{1}{2}(H_{-1} + H_1)$ where H_a denotes the Heavyside distribution function with jump at the point a. In what follows, we extend the variational techniques introduced by Hoeffding to prove that the extremal problem $(\star)$ has an extremal distribution that is supported at two points. We also introduce an example where the extremal distribution is not symmetric.

Theorem 1. *Let $X_1, X_2, \dots, X_n$ be a sequence of iid random variables such that $E|X_j| = 1$ Then the minimum of*

$$|X_1 + X_2 + \dots + X_n|^r; \quad r \ge 1,$$

is attained for a random variable that is supported on exactly two points, one of which is nonnegative, and the other nonpositive.

Lemma 1 (Loève (1977), prob. 2, p. 275.). *Let $r \geq 1$ and $X_1, X_2, \ldots, X_n$ be iid random variables. Then $E|X_1 + X_2 + \ldots + X_n|^r < \infty$ iff $E|X_i|^r < \infty$.*

Let D_r be the set of all distribution functions F such that

$$\int_{\Re} |x|^r dF < \infty \qquad and \qquad \int_{\Re} |x| dF = 1,$$

and let C_k be the subset of D_r which consists of all distribution functions supported on at most k points. We endow D_r with the norm:

$$||F|| = \int_{\Re} max(1, |x|^r) dF.$$

With respect to this topology we have the following result.

Lemma 2. *$\bigcup_{k=1}^{\infty} C_k$ is dense in D_r.*

Proof. Fix F in D_r and define

$$F_1(x) = \begin{cases} 0; & x < 0 \\ F(x); & x \geq 0 \end{cases}$$

and $F_2(x) = 0$ if $F(0) = 0$ or else

$$F_2(x) = \begin{cases} F(x); & x < 0 \\ 0; & x \geq 0. \end{cases}$$

Then $F(x) = F_1(x) + F_2(x)$. Put

$$\int_{\Re} dF_1 = \alpha_1, \quad \int_{\Re} dF_2 = \alpha_2, \quad \int_{\Re} x dF_1 = \beta_1 \quad \text{and} \quad \int_{\Re} x dF_2 = -\beta_2.$$

Since F is in D_r we have

$$\alpha_1 + \alpha_2 = 1 \quad and \quad \beta_1 + \beta_2 = 1.$$

Now let X denote the normed linear space of all right continuous non-negative and monotone non-decreasing functions G on $\Re^+$ such that $G(x) = 0$ for $x < 0$ and $\int_{\Re} |x|^r dF < \infty$. Then the subspace Y of all G in X which are finitely supported is dense in X. We now apply the following result of Deutsch (1966):

> *Let Y be a dense subspace of a topological vector space X. Then for every $x \in X$, neighborhood U of x and $T_1, \ldots, T_n \in X'$ (the topological dual of X), there is a $y \in Y$ such that $y \in U$ and $T_j(x) = T_j(y); \quad j = 1, ..., n$.*

We conclude that in a small neighborhood of F_1 there exists a $G_1 \in Y$ such that

$$\int_{\Re} dG_1 = \alpha_1 \quad and \quad \int_{\Re} x dG_1 = \beta_1.$$

Similarly, if X is the normed linear space of right continuous nonnegative and monotone nondecreasing functions G on $\Re^-$ such that $G(x) = 0$ for $x \geq 0$ and $\int |x|^r dF < \infty$ and Y is the subspace of X of finitely supported functions, then in a small neighborhood of F_2, there exists a $G_2 \in Y$ such that

$$\int_{\Re} dG_2 = \alpha_2 \qquad \int_{\Re} x dG_2 = -\beta_2.$$

Hence $G = G_1 + G_2$ is in a small neighborhood of F and belongs to $\bigcup_{k=1}^{\infty} C_k$.

Lemma 3. $K^{(r)}(F) \equiv E_F|X_1 + \ldots + X_n|^r$ *is a continuous functional.*

Proof. If F and G are two distribution functions then

$$\begin{aligned} &dF(x_1)dF(x_2)\ldots dF(x_n) - dG(x_1)dG(x_2)\ldots dG(x_n) \\ &\quad = \sum_{i=1}^{n} dG(x_1)\ldots dG(x_{i-1})[d(F-G)(x_i)]dF(x_{i+1})\ldots dF(x_n), \end{aligned}$$

with an appropriate interpretation for the extreme indices 0 and $n+1$. By the convexity of $|x|^r (r \geq 1)$ we have

$$\left|\sum_{k=1}^{n} x_k\right|^r \leq \left(\sum_{k=1}^{n} |x_k|\right)^r = n^r \left(\sum_{k=1}^{n} \frac{|x_k|}{n}\right)^r \leq n^{r-1}\left(\sum_{k=1}^{n} |x_k|^r\right).$$

Thus, if F and G are in D_r then

$$\begin{aligned} &|K^{(r)}(F) - K^{(r)}(G)| \\ &\leq n^{r-1} \sum_{k=1}^{n}\sum_{i=1}^{n} \int_{\Re^n} |x_k|^r dG(x_1)\ldots[d(F-G)(x_i)]\ldots dF(x_n) \end{aligned}$$

and

$$\begin{aligned} &\int_{\Re^n} |x_k|^r dG(x_1)\ldots[d(F-G)(x_i)]\ldots dF(x_n) \\ &\qquad = \begin{cases} \int_{\Re} |d(F-G)(x)| \int_{\Re} |x|^r dG(x) & \text{if } i \neq k \\ \int_{\Re} |x|^r |d(F-G)(x)| & \text{if } i = k \end{cases} \end{aligned}$$

from which the continuity follows.

Lemmas 2 and 3 imply that

$$\inf_{F \in D_r} K^{(r)}(F) = \inf_{F \in \cup C_k} K^{(r)}(F).$$

Therefore, without loss of generality, we may only consider discrete distributions. In the following lemma we exclude the case $r = 1$ because $|x|$ is not strictly convex and not differentiable everywhere.

Lemma 4. *Suppose that the minimum of $K^{(r)}(F)(r > 1)$ in C_k is attained for F_k. Then F_k is supported on at most two points, one of them is nonnegative and the other one is nonpositive.*

Proof. Suppose indirectly, that $supp\ \ F_k = \{b_1, b_2, \dots, b_m\}$ where $m > 2$ and $\beta_i = F_k(b_i + 0) - F_k(b_i - 0)$. We introduce a variation F_k^* of F_k such that for some fixed $1 \le i \ne j \le m$, all $b's$ and $\beta's$ are kept fixed except for the values of b_i, β_i, b_j and β_j that are changed as follows. Choose i and j such that $b_i \ne 0$ and $|b_i| \ne |b_j|$ and put

$$\begin{aligned} \beta_i^* &= \beta_i + \epsilon c \qquad & b_i^* &= b_i + \epsilon \\ \beta_j^* &= \beta_j - \epsilon c \qquad & b_j^* &= b_j \end{aligned}$$

where $c = -sgn(b_i)\beta_i/(|b_i| - |b_j| + \epsilon sgn(b_i))$. For small ϵ's F_k^* is in C_k and thus

$$K^{(r)}(F_k^*) \ge K^{(r)}(F_k).$$

Recall the Heavyside distribution H_a and put $J(x_n) = E_{F_k}|X_1 + X_2 + \dots + X_{n-1} + x_n|^r$. Using the symmetry of the kernel function $|x_1 + x_2 + \dots + x_n|^r$ we have

$$\begin{aligned} K^{(r)}(F_k^*) &= K^{(r)}(F_k - \epsilon c H_{b_j} + \epsilon c H_{b_i+\epsilon} + \beta_i(H_{b_i+\epsilon} - H_{b_i}) + O(\epsilon^2)) \\ &= K^{(r)}(F_k) + n\epsilon \int_{\Re} J(x) d(c(H_{b_i+\epsilon}(x) \\ &\qquad - H_{b_j}(x)) + \beta_i(H_{b_i+\epsilon}(x) - H_{b_i}(x))/\epsilon) + O(\epsilon^2) \\ &= K^{(r)}(F_k) + n\epsilon \left[c\left(J(b_i + \epsilon) - J(b_j)\right) + \beta_i \frac{J(b_i + \epsilon) - J(b_i)}{\epsilon} \right] \\ &\qquad + O(\epsilon^2) \\ &\ge K^{(r)}(F_k). \end{aligned}$$

Thus, we get

$$n\epsilon \left[c\left(J(b_i + \epsilon) - J(b_j)\right) + \beta_i \frac{J(b_i + \epsilon) - J(b_i)}{\epsilon} \right] + O(\epsilon^2) \ge 0.$$

Therefore, using the differentiability of $|x_1 + \dots + x_n|^r (r > 1)$, we get

$$n\epsilon\beta_i \left[-sgn(b_i) \frac{J(b_i) - J(b_j)}{|b_i| - |b_j|} + J'(b_i) \right] + o(\epsilon) \ge 0.$$

for any positive or negative (small) ϵ. Hence

$$J'(b_i) = \frac{J(b_j) - J(b_i)}{|b_j| - |b_i|} sgn(b_i). \tag{1}$$

We shall show that there is only one nonnegative number in *supp* F_k. Suppose $b_i > 0$. If $b_j \geq 0$ then we see from (1) that the tangent line of J at b_i has the same slope as the secant connecting $(b_i, J(b_i))$ and $(b_j, J(b_j))$. Since J is a finite convex combination of strictly convex curves $(r > 1)$ we get a contradiction. Similarly, *supp* F_k contains at most one nonpositive point.

Remark 1. One can easily see that under the constraint $E|X| = 1$ the minimum of $E|X_1 + ... + X_n|^r$ is not attained for one point distributions $(H_1 \quad or \quad H_{-1})$. Thus the minimum is taken for distributions supported on exactly two points.

If *supp* F_k contains two points, say $b_1 \geq 0$ and $b_2 < 0$ then $(\star)$ shows that $J'(b_2) = -J'(b_1)$. We shall need this relation later.

Proof of Theorem 1. Let $r > 1$. Then by the previous lemmas we have

$$\begin{aligned} m = \inf_{D_r} K^{(r)}(F) &= \inf_{\cup C_k} K^{(r)}(F) \\ &= \inf_k (\inf_{C_k} K^{(r)}(F)) \\ &= \inf_k (K^{(r)}(F_k)) \end{aligned}$$

where F_k is in C_2 and hence $F_k = F_2$. Therefore

$$m = K^{(r)}(F_2).$$

Only the case $r = 1$ has not been settled. Denote by $F_2^{(r)}$ the extremal distribution function in C_2 which minimizes $E|X_1 + \ldots + X_n|^r \quad (r > 1)$. We claim that there exists a sequence r_k such that $F_2^{(r_k)}$ converges to some F_2 in C_2. To see this, let

$$supp \quad F_2^{(r_k)} = \{a_k, b_k\}, \quad \text{and} \quad \alpha_k = F_2^{(r_k)}(a_k + 0) - F_2^{(r_k)}(a_k - 0)$$

such that $F_2^{(r_k)}$ is a convergent sequence of distribution functions where $|a_k| \leq 1$ and $|b_k| \geq 1$. Then α_k converge to some α and a_k to some a where $0 \leq \alpha \leq 1$ and $|a| \leq 1$. If b_k has a finite limit b then $F_2^{(r_k)}$ converges to some F in C_2. Assume indirectly that b_k diverges to infinity (a similar argument works for b_k diverging to minus infinity). By Lemma 4 we have $b_k > 0$, $a_k < 0$ for large k and $J'(b_k) = -J'(a_k)$ where

$$\begin{aligned} J(x_n) &= E|X_1 + \ldots + X_{n-1} + x_n|^{r_k} \\ &= \sum_{i=0}^{n-1} \binom{n-1}{i} (1-\alpha_k)^{n-1-i} \alpha_k^i |(n-1-i)a_k + ib_k + x_n|. \end{aligned}$$

Since a_k is a bounded sequence and b_k diverges to infinity, we conclude that the functions

$$G_i(x_n) = (n-1-i)a_k + ib_k + x_n, \quad i = 1, 2, \ldots, n$$

satisfy $G_i(a_k) > 0$ and $G_i(b_k) > 0$ for sufficiently large k. This implies that both $J'(a_k)$ and $J'(b_k)$ are positive which leads to a contradiction. Hence F_2 is in C_2.

Let F minimize $E_F|X_1 + \ldots + X_n|$ on D_1. Since $|x_1 + \ldots + x_n|^r$ is monotone as r decreases to one (decreasing on the subset $|x_1 + \ldots + x_n| \leq 1$ and increasing on its complement), we conclude by the Lebesgue monotone convergence theorem that

$$K^{(r)}(F) \to K^{(1)}(F_2).$$

We also have

$$K^{(r_k)}(F_2^{(r_k)}) \to K^{(1)}(F_2).$$

because this sequence is a finite combination of continuous functions in r_k, a_k, and α_k. But

$$K^{(r_k)}(F_2^{(r_k)}) \leq K^{(r_k)}(F)$$

and therefore in the limit we get that

$$K^{(1)}(F_2) \leq K^{(1)}(F)$$

and hence equal. This completes the proof of the theorem.

§2. Special Cases, Generalizations

We consider the case $n = 3$ and $r = 1$. By Theorem 1,

$$\begin{aligned}\min_{D_r} E|X_1 + X_2 + X_3| &= \min_{C_2} E|X_1 + X_2 + X_3| \\ &= \min(\alpha^3 \cdot 3a + 3\alpha^2(1-\alpha)|2a+b| + 3\alpha(1-\alpha)^2|a + (1-\alpha)^3|3b|)\end{aligned}$$

under the constraint

$$\alpha a - (1-\alpha)b = 1, \quad a \geq 0, \quad b < 0 \quad \textit{and} \quad 0 < \alpha < 1.$$

Fixing α, and noting that b is linear in a, the last expression is a convex combination of piecewise linear functions of a. Hence, it attains its minimum at a corner point, specifically when

$$a + 2b = 0 \quad \textit{or} \quad b + 2a = 0.$$

These are a symmetric pair and therefore we need only consider the first case which yields

$$a = 1/(2-\alpha) \quad \textit{and} \quad b = -2/(2-\alpha).$$

The minimum of

$$\begin{aligned}3\alpha^3/(1-\alpha) + 3\alpha^2(1-\alpha)|2/(2-\alpha) + 2/(2-\alpha)\\ + (\alpha/(2-\alpha)-1)/(1-\alpha)|\\ + 3\alpha(1-\alpha)^2|1/(2-\alpha) + (2\alpha/(2-\alpha)-1)/(1-\alpha)|\\ + (1-\alpha)^3 3|(\alpha/(2-\alpha)-1)/(1-\alpha)|\end{aligned}$$

is attained for $\alpha^* = .6527$ and thus $a = .7422$, $b = -1.4844$. Finally $E|X_1 + X_2 + X_3| \geq 1.3316$.

For large n we have the following asymptotic result.

Proposition 1. *Let T_n be the distribution function that minimizes $E|X_1 + \ldots + X_n|$ where $X_1, \ldots, X_n$ are iid such that $E|X_j| = 1$ and $T = \frac{1}{2}(H_{-1} + H_1)$. Then T_n converges in law to T as n tends to infinity.*

Proof. It is readily verified that

$$E|X_1 + \ldots + X_n| = \sum_{j=0}^{n} \binom{n}{j} |n-2j|/2^j$$

has magnitude $\sqrt{n}$. On the other hand if $X_1, \ldots, X_n$ are iid then $X_1 + \ldots + X_n$ is asymptotically normally distributed with expected value $\sim nEX_1$. Hence

$$\begin{aligned}E_T|X_1 + \ldots + X_n| \geq E_{T_n}|X_1 + \ldots + X_n| \geq |E_{T_n}(X_1 + \ldots + X_n)|\\ \sim nE_{T_n}X\end{aligned}$$

which is a contradiction unless $E_{T_n}X$ tends to zero. Since T_n is supported at two points we conclude that T_n converges in law to T.

The method of the proof of Lemma 4 shows that the following general result also holds.

Theorem 2. *Let $Q(x_1, x_2, \ldots, x_n) : \Re^n \to \Re$ be a symmetric strictly convex differentiable kernel function. Let $X1_, X_2, \ldots, X_n$ be iid random variables with distribution function F. Put $K^Q(F) = E_F Q(X_1, X_2, \ldots, X_n)$, $D_Q = F : K^Q(F) < \infty$ and $E_F|X| = 1$ and let C_k be the subset of D_Q whose elements are supported on at most k points. If*

$$\inf_{F \in D_Q} K^Q(F) = \inf_{F \in \cup_k C_k} K^Q(F)$$

then $K^Q(F)$ is minimized by an F supported on two points.

It would be interesting to replace the condition $E_F|X| = 1$ by some more general condition of the form $E_F g(X) = c$ where $g : \Re^m \to \Re^k$, $c \in \Re^k$.

Appendix

The following probabilistic approach uses a convexity argument and was suggested by Burgess Davis.

Theorem 3. *Let $X, X_1, X_2, \dots, X_n$ be an arbitrary sequence of iid random variables, and $f : \Re^n \to \Re$, $\quad g : \Re \to \Re$ convex functions (in each variable) such that $E(f(X_1, X_2, \dots, X_n))$ and $E(g(X))$ exist. If g is a broken line with k linear segments then the minimum of $E(f(X_1, X_2, \dots, X_n))$ is taken for random variables X with at most k values.*

Corollary. *Theorem 1.*

Proof of Theorem 3. Let $\mathcal{F}$ be an arbitrary σ-algebra, and put $\hat{X}_i = E(X_i|\mathcal{F})$. Then by Jensen's inequality $E(f(X_1, X_2, \dots, X_n)|\mathcal{F}) \geq f(\hat{X}_1, X_2, \dots, X_n) \geq \dots \geq f(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n)$ a.s., and taking expectations in both sides $E(f(X_1, X_2, \dots, X_n)) \geq Ef(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n)$. On the other hand, again by Jensen's inequality $E(g(X)) \geq E(g(\hat{X}))$ with equality iff $\mathcal{F}$ is such that

$$E(g(X)|\mathcal{F}) = g(E(X|\mathcal{F})) \quad a.s. \tag{$\star\star$}$$

In this case it is clear that given $E(g(X))$, the minimum of $E(f(X_1, X_2, \dots, X_n))$ is taken for random variables of the form $\hat{X} = E(X|\mathcal{F})$.

In case g is a convex broken line with k linear segments, then one can easily define a function ϕ with k values such that $(\star\star)$ holds if $\mathcal{F}$ is generated by this function ϕ. E.G. if $g(x) = |x|$ then

$$\phi(x) = \begin{cases} +1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

is a good choice.

References

[1] Deutsch, F., 1966, *Simultaneous interpolation and approximation in topological linear spaces,* SIAM J. Appl. Math. 14, 1180-1190.

[2] Dubins, L. E., 1962, *On extreme points of convex sets,* J. Math. Anal. and Appl. 5, 237-244.

[3] Hoeffding, W., 1955, *The extrema of the expected value of a function of independent random variables,* Ann. Math. Statist. 26, 268-275.

[4] Hoeffding, W. and Shrikhande, S. S., 1955, *Bounds for the distribution function of a sum of independent, identically distributed random variables,* Ann. Math. Statist. **26,** 439-449.

[5] Hartley, H. O. and David, H. A., 1954, *Universal bounds for mean range and extrema observation,* Ann. Math. Statist. 25, 55-99.

[6] Hildebrand, D. K., 1984, *Lord's MAD paradox and Jensen's inequality,* Amer. Statist. 38, 296-297.

[7] Karlin, S. and Shapley, L. S., *Geometry of Moment Spaces*, Mem. Amer. Math. Soc., No. 12., 1953.

[8] Loéve, M., 1977, *Probability Theory I* (4th ed.), New York, Springer-Verlag.