AN INEQUALITY FOR DEVIATIONS FROM MEDIANS

BY JOHN W. TUKEY

Princeton University

In a recent note in these Annals, Birnbaum and Zuckerman [1] proved that if:

(1) $X_1, X_2, \cdots, X_n$ are independent random variables with the same
distribution (i.e., form a sample),
(2) their common distribution is symmetric about zero,

then

$$E(|X_1 + X_2 + \cdots + X_n|) \geq \varphi(n) \cdot E(|X_1|),$$

where

$$\varphi(2k + 1) = \varphi(2k + 2) = \frac{1 \cdot 3 \cdot 5 \cdot 7 \cdots (2k + 1)}{1 \cdot 2 \cdot 4 \cdot 6 \cdots (2k)}.$$ \[1\]

It is the purpose of the present note to extend this to the following, more
general, result:

**Theorem.** If

(i) $X_1, X_2, \cdots, X_n$ are independent random variables,
(ii) the median of each $X_i$ is zero,

then

$$E(|X_1 + X_2 + \cdots + X_n|) \geq \frac{\varphi(n)}{n} E(|X_1| + |X_2| + \cdots + |X_n|).$$

It will be convenient to let $d_i = E(|X_i|)$ and

$$d = \frac{1}{n} \sum d_i = \frac{1}{n} E(|X_1| + |X_2| + \cdots + |X_n|),$$

so that the desired inequality becomes

$$E(|X_1 + X_2 + \cdots + X_n|) \geq \varphi(n) \cdot d.$$

Define $e_i$ by

$$e_i = \int_0^\infty x dF_i(x),$$

where $F_i(x)$ is the cumulative distribution function of $X_i$. Since

$$d_i = E(|X_i|) = -\int_0^\infty x dF_i(x) + \int_{\infty}^0 x dF_i(x),$$

with $d_i = E(|X_i|) = \frac{1}{n} \sum d_i = \frac{1}{n} E(|X_1| + |X_2| + \cdots + |X_n|),$

...
it follows that
\[ \int_0^\infty x dF_t(x) = e_t - d_t. \]

The basic idea of the proof, which is common to both notes, is to divide the \( n \)-dimensional space of \( x_1, x_2, \ldots, x_n \) into its \( 2^n \) "octants," break up the expectation of \( |X_1 + X_2 + \cdots + X_n| \) into the corresponding parts, and apply elementary inequalities. Let \( O_s \) be the octant in which a set \( S \) of variables are \( \leq 0 \). From (4), (5) and hypothesis (ii) it follows that
\[ 2^{n-1} \int_{O_s} \cdots \int x_i \prod dF_i(x) = \begin{cases} e_i, & \text{if } x_i \geq 0 \text{ in } O_s, \\ e_i - d_i, & \text{if } x_i \leq 0 \text{ in } O_s. \end{cases} \]

Hence
\[ 2^{n-1} \int_{O_s} \cdots \int \sum x_i \prod dF_i(x) = \sum_{i=1}^n e_i - \sum d_i = e - \sum d_i. \]

where \( e = \sum e_i \), and the second and third sums are over all \( d_i \) for which \( x_i \leq 0 \) in the chosen octant \( O_s \). The contribution of the octant \( O_s \) to \( E(|X_1 + X_2 + \cdots + X_n|) \) is
\[ \int_{O_s} \cdots \int |\sum x_i| \prod dF_i(x) \geq \left| \int_{O_s} \cdots \int (\sum x_i) \prod dF_i(x) \right| \]
\[ = 2^{-(n-1)} |e - \sum d_i|. \]

For each value of \( s \), there will be \( \binom{n}{s} \) octants with \( s \) variables \( \leq 0 \). The sum of their contribution to \( E(|X_1 + X_2 + \cdots + X_n|) \) is
\[ I_s = \frac{1}{2^{n-1}} \sum |e - \sum d_i| \geq \frac{1}{2^{n-1}} \binom{n}{s} e - \left( \binom{n}{s} - 1 \right) \sum d_i, \]
where the inequality follows from \( \sum |a_s| \geq |\Sigma a_s| \), and it is noticed that each \( d_i \) occurs in \( \binom{n-1}{s-1} \) different inner sums. Recalling that \( \Sigma d_i = nd \), this may be written
\[ I_s \geq \frac{1}{2^{n-1}} \binom{n}{s} |e - sd|. \]
Finally,
\[ E(|X_1 - X_2 + \cdots + X_n|) = \sum_{i=0}^{n} I_i \geq 2^{-(n-1)} \sum_{i=0}^{n} \binom{n}{s} |e - sd| \]
\[ \geq 2^{-(n-1)} \sum_{2s < n} \binom{n}{s} (|e - sd| + |e - (n - s)d|) \]
\[ \geq 2^{-(n-1)} \sum_{2s < n} \binom{n}{s} (n - 2s)d, \]
where the last inequality follows from \(|a| + |b| \geq b - a\). To complete the proof, it is only necessary to evaluate the last sum. One method of evaluation may be found in Birnbaum and Zuckerman’s note.

If each \( X_i = \pm 1 \), each with probability one-half, then all of the inequalities of the proof become equalities. So that, in this case,
\[ E(|X_1 + X_2 + \cdots + X_n|) = \varphi(n) \cdot d. \]

Since the limiting distribution in this case is a normal distribution with standard deviation \( n^2 \) and \( E(|X_1 + X_2 + \cdots + X_n|) = (2n/\pi)^{1/2} \), it follows that this is the asymptotic value of \( \varphi(n) \).

The inequality of the theorem is only efficient when the \( E(|X_i|) \) are of nearly the same size. In other cases it can often be usefully supplemented by the

**Lemma.** If

(i) \( X_1, X_2, \ldots, X_n \) are independent

(ii) for each \( i \), either \( X_i \) has median zero, or the sum of the means of the other \( X_j \) is zero (this is implied by either (a) the median of each \( X_i \) is zero, or (b) the mean of each \( X_i \) is zero), then

\[ E(|X_1 + X_2 + \cdots + X_n|) \geq \text{Max } E(|X_i|). \]

The lemma follows from the case where \( n = 2 \), by applying that case to

\[ Y_i = X_{i_0}, \quad Y_2 = \sum_{i \neq i_0} X_{i_0}, \]

where the maximum of \( E(|X_i|) \) is attained for \( i = i_0 \).

The special case follows from the inequality

\[ |x_1 + x_2| \geq |x_1| + x_2 \cdot \text{sgn } x_1, \]
since this implies

\[ E(|X_1 + X_2|) \geq E(|X_1|) + E(X_2) \cdot E(\text{sgn } X_1) = E(X_1) \]

using first (i) and then (ii).

In conclusion, it is interesting to note that the mean cannot replace the median in the hypothesis of the theorem. For let \( X_1, X_2, X_3 \) be independent,
and take the values 1 (with probability 2/3) and -2 (with probability 1/3). 
$X_1 + X_2 + X_3$ takes the values 3 (with probability 8/27), 0 (with probability 
12/27), -3 (with probability 6/27) and -6 (with probability 1/27). Hence 
$E(|X_1|) = 4/5$, and $E(|X_1 + X_2 + X_3|) = 48/27 = 16/9 = 4/3E(|X_i|)$, 
which is not $\geq 3/2E(|X_i|)$.

**REFERENCE**


---

**ON THE INDEPENDENCE OF THE EXTREMES IN A SAMPLE**

**BY E. J. GUMBEL**

New School for Social Research

In a previous article [1] the assumption was used that the $m$th observation in 
ascending order (from the bottom) and the $m$th observation in descending order 
(from the top) are independent variates, provided that the rank $m$ is small com-
pared to the sample size $n$. In the following it will be shown that this assump-
tion holds for the usual distributions.

Let $x$ be a continuous, unlimited variate, let $\Phi(x)$ be the probability of a value 
equal to, or less than, $x$; let $\varphi(x)$ be the density of probability, henceforth called 
the initial distribution. The $m$th observation from the bottom is written $m x$ 
and the $k$th observation from the top is written $x_k$. Thus, the bivariate dis-
tribution $w_n(m x, x_k)$ of $m x$ and $x_k$, is such that there are $m - 1$ observations less 
than $m x$; $k - 1$ observations greater than $x_k$ and $n - m - k$ observations between 
$m x$ and $x_k$.

For simplicity's sake write

$$
\Phi(m x) = m \Phi; \quad \Phi(x_k) = \Phi_k.
$$

$$
\varphi(m x) = m \varphi; \quad \varphi(x_k) = \varphi_k.
$$

Then

$$
w_n(m x, x_k) = C_m \frac{m - 1 - m \varphi(\Phi_k - m \Phi)}{n - m - k} \varphi_k(1 - \Phi_k)^{k-1},
$$

where

$$
C = \frac{n!}{(m - 1)!(k-1)!(n - m - k)!}.
$$

In the expression (1) no assumption about dependence or independence of $m x$ 
and $x_k$ is implied except that these values are taken from the same population.

The distribution (1) is now modified by introducing three conditions. First,

---

1 Research done with the support of a grant from the American Philosophical Society.