Ann. Inst. Statist. Math.
38 (1986), Part A, 551-555

INEQUALITIES FOR ORDERED SUMS

H. A. DAVID*

(Received Nov. 5, 1984; revised June 10, 1985)

Summary

Let $x_i=y_i+z_i$, $i=1,\cdots,n$, and write $x_{(1)}\leqq\cdots\leqq x_{(n)}$, with corresponding notation for the ordered y_i and z_i. It is shown, for example, that $x_{(r)}\geqq\max_{i=1,\cdots,r}(y_{(i)}+z_{(r+1-i)})$, $r=1,\cdots,n$. Inequalities are also obtained for convex (or concave) functions of the $x_{(i)}$. The results lead immediately to bounds for the expected values of order statistics in nonstandard situations in terms of simpler expectations. A small numerical example illustrates the method.

1. Introduction

Let $X_{(1)}\leqq\cdots\leqq X_{(n)}$ be the order statistics formed from random variables $X_1,\cdots,X_n$. Smith and Tong [6] have developed inequalities for convex (or concave) functions of the $X_{(i)}$, $i=1,\cdots,n$, when X_i is expressible as a sum of two other random variables, $X_i=Y_i+Z_i$. The Y_i are not necessarily independent or identically distributed, nor are the Z_i. In applications the Z_i are often constants, $Z_i=\delta_i$, $i=1,\cdots,n$. The inequalities are in terms of the ordered Y_i, Z_i, or δ_i, and are useful whenever, for example, $\mathrm{E}\,Y_{(i)}$ and $\mathrm{E}\,Z_{(i)}$ can be handled more easily than $\mathrm{E}\,X_{(i)}$.

Of the order statistics only the maximum is convex (and the minimum concave). In this note we derive a simple inequality (Theorem 1) that holds for order statistics of any rank. We also strengthen one of the results of Smith and Tong [6] for convex functions of order statistics. Our results hold for any numbers x_i, y_i, z_i linked by $x_i=y_i+z_i$, $i=1,\cdots,n$. Applications to bounds for the $\mathrm{E}\,X_{(i)}$ are immediate. A small numerical example illustrates our methods on a normal sample with an unidentified outlier and permits some comparisons with the inequalities in Mallows and Richter [3], Arnold and Groeneveld [1], and Nagaraja [5].

* Research supported by U.S. Army Research Office.
Key words and phrases: Order statistics, convexity, majorization, outliers.

2. Inequalities for ordered sums

We begin with an inequality for an ordered x-value in terms of the ordered y and z values, where $x_i=y_i+z_i$, $i=1,\cdots,n$. It will be convenient to write the ordered x_i in either ascending or descending order:

(1a, 1b) $$x_{(1)}\leqq\cdots\leqq x_{(n)} \quad \text{or} \quad x_{[1]}\geqq\cdots\geqq x_{[n]} ,$$

with corresponding notation for the ordered y_i and z_i. Then clearly

(2a, 2b) $$x_{[1]}\leqq y_{[1]}+z_{[1]} , \qquad x_{(1)}\geqq y_{(1)}+z_{(1)} ,$$

and, with range $x_i=x_{[1]}-x_{(1)}$, etc., one has

(3) $$\text{range } x_i\leqq\text{range } y_i+\text{range } z_i .$$

The results (2) may be generalized as follows:

THEOREM 1. *If* $x_i=y_i+z_i$, $i=1,\cdots,n$, *then with the notation of* (1) *we have for* $r=1,\cdots,n$

(4a) $$x_{[r]}\leqq \min_{i=1,\cdots,r} (y_{[i]}+z_{[r+1-i]})$$

and

(4b) $$x_{(r)}\geqq \max_{i=1,\cdots,r} (y_{(i)}+z_{(r+1-i)}) .$$

PROOF. For some i in $\{1,\cdots,r\}$ and some j in $\{1,\cdots,n\}$ suppose that $x_j>y_{[i]}+z_{[r+1-i]}$. A necessary condition for this to hold is that either $y_j>y_{[i]}$ or $z_j>z_{[r+1-i]}$. Thus at most $(i-1)+(r-i)=r-1$ of the x_j can exceed $y_{[i]}+z_{[r+1-i]}$, i.e., $x_{[r]}\leqq y_{[i]}+z_{[r+1-i]}$, $i=1,\cdots,r$, which is (4a). The proof of (4b) is similar.

Comment 1. Equations (4a) and (4b) may equivalently be stated as

$$\max_{i=1,\cdots,r} (y_{(i)}+z_{(r+1-i)})\leqq x_{(r)}\leqq \min_{j=1,\cdots,n+1-r} (y_{(n+1-j)}+z_{(r-1+j)}) .$$

Comment 2. The bounds in (4) are attainable. In (4a) equality in the form $x_{[r]}=y_{[i]}+z_{[r+1-i]}$ is achieved if and only if there exists j such that $y_j=y_{[i]}$ and $z_j=z_{[r+1-i]}$. For example, $x_{[1]}=y_{[1]}+z_{[1]}$ if there exists j such that $y_j=y_{[1]}$ and $z_j=z_{[1]}$; in words, equality is achieved if y_j and z_j are the largest or, in case of ties, one of the largest y's and z's, respectively.

Frequently the sum $S_k=\sum_{i=1}^{k} x_{[i]}$ of the k largest x_i is of interest, $k=1,\cdots,n$. In obtaining inequalities for S_k one can do better than merely add the first k inequalities of (4). Thus, it is obvious that

(5a) $$S_k \leqq \sum_{i=1}^{k} (y_{[i]}+z_{[i]}) .$$

To deal with inequalities in the other direction note that from (4b)

$$S_1 \equiv x_{(n)} \geqq \max_{i=1,\cdots,n} (y_{(i)}+z_{(n+1-i)}) .$$

Let $x'_{(i)}$, $i=1,\cdots,n$ denote the n sums $y_{(i)}+z_{(n+1-i)}$ arranged in ascending order of magnitude. Then $x'_{(n)}$ is an attainable lower bound for $x_{(n)}$. If $x_{(n)}=x'_{(n)}$, with $x'_{(n)}=y_{(h)}+z_{(n+1-h)}$ (say), $h=1,\cdots,n$, we note that $x'_{(n-1)}$ is an attainable lower bound for $x_{(n-1)}$, where here

$$x'_{(n-1)} = \max_{\substack{j=1,\cdots,n \\ j\neq i}} (y_{(j)}+z_{(n+1-j)}) .$$

Repeating the process we have the sharp inequality

(5b) $$S_k = \sum_{i=n+1-k}^{n} x'_{(i)} .$$

The results (5) may be extended from S_k to a convex linear function l of ordered x_i, viz. $l=\sum_{i=1}^{n} c_i x_{(i)}$ with $c_1 \leqq \cdots \leqq c_n$. To see this, note that

$$l = c_1 S_n + (c_2-c_1) S_{n-1} + \cdots + (c_n - c_{n-1}) S_1$$

so that from (5) we have easily for $c_1 \geqq 0$

(6) $$\sum_{i=1}^{n} c_i x'_{(i)} \leqq l \leqq \sum_{i=1}^{n} c_i (y_{(i)}+z_{(i)}) .$$

If $c_m<0$, $c_{m+1}\geqq 0$, $m=1,\cdots,n$, equation (6) continues to hold since l may be split into

$$\sum_{i=1}^{m} c_i x_{(i)} + \sum_{i=m+1}^{n} c_i x_{(i)}$$

and the lower end counterparts of (5) applied to the first sum.

The right hand inequality of (6) is essentially the same as that in Theorem 2.1 of Smith and Tong [6], where majorization arguments (e.g., Marshall and Olkin [4]) are used. Our lower bound in (6) is, however, superior to theirs, viz. $\sum c_i (y_{(i)}+z_{[i]})$.

Clearly it is also possible to generalize (3) to

$$\sum_{i=1}^{m} d_i (x_{[i]}-x_{(i)}) \leqq \sum_{i=1}^{m} d_i (y_{[i]}-y_{(i)}) + \sum_{i=1}^{m} d_i (z_{[i]}-z_{(i)}) ,$$

where $d_1 \geqq \cdots \geqq d_m$ and $m=n/2$ or $(n+1)/2$ according as n is even or odd.

Finally, generalizations to the case $x_i=\sum_{j=1}^{p} x_{ij}$, $i=1,\cdots,n$ and p a positive integer ($p>1$), are straightforward. For example, in obvious notation, (4a) generalizes to

$$x_{[r]} \leqq \min \sum_{j=1}^{p} x_{[i_j]j} ,$$

where the minimum is taken over all positive integers i_j for which $\sum_{i=1}^{p} i_j = r+p-1$.

3. Bounds for the expectations of order statistics in nonstandard situations

Let $X_i = Y_i + Z_i$, $i=1,\cdots,n$, where the Y_i and Z_i are random variables. Then provided only that $\mathrm{E}\, Y_i$ and $\mathrm{E}\, Z_i$ exist, $i=1,\cdots,n$, previous results can immediately be converted into corresponding inequalities between expectations. For example, (4b) becomes

$$\mathrm{E}\, X_{(r)} \geqq \mathrm{E} \max_{i=1,\cdots,r} (Y_{(i)} + Z_{(r+1-i)}) \geqq \max_{i=1,\cdots,r} (\mathrm{E}\, Y_{(i)} + \mathrm{E}\, Z_{(r+1-i)}) ,$$

the last step following from Jensen's inequality.

It should be noted that no assumptions of independence and common distributions are needed in the above but some simplifying assumptions will usually be invoked in applications. A case of special interest occurs when the Y_i are iid and the Z_i are constants, say $Z_i = \delta_i$, $i=1,\cdots,n$. We consider an example of this kind. Other applications are given in Smith and Tong [6].

Sample with one outlier. Let Y_i, $i=1,\cdots,n$, be iid, $Z_i = \delta > 0$ for some unknown value of i and $Z_i = 0$ otherwise. Then the X_i are a sample with an unidentified outlier.

From (4) we have for $r=1,\cdots,n-1$

(7a) $$\mathrm{E}\, Y_{(r)} \leqq \mathrm{E}\, X_{(r)} \leqq \min (\mathrm{E}\, Y_{(r+1)}, \mathrm{E}\, Y_{(r)} + \delta) ,$$

and, for $r=n$,

(7b) $$\max (\mathrm{E}\, Y_{(n)}, \mathrm{E}\, Y_{(1)} + \delta) \leqq \mathrm{E}\, X_{(n)} \leqq \mathrm{E}\, Y_{(n)} + \delta .$$

As a small numerical illustration we take the Y_i to be independent normal $N(0, 1)$, $n=5$, and $\delta=2$. Results are presented in Table 1. In this case it is possible to obtain exact values for $\mathrm{E}\, X_{(r)}$ (David et al. [2]) but the bounds can easily be calculated for general patterns of the δ_i in the case of any distribution for which $\mathrm{E}\, Y_{(r)}$, the expected values of the order statistics, are available.

In our example the bounds for $\mathrm{E}\, X_{(5)}$ are rather wide and may be

Table 1. Expected value, $\mathrm{E}\,X_{(r)}$, of r-th order statistic in a standard normal sample of size 5 in the presence of an outlier with mean $\delta=2$

r	$\mathrm{E}\,Y_{(r)}$	$\mathrm{E}\,X_{(r)}$		
		lower bound(7)	exact value	upper bound(7)
1	−1.1630	−1.1630	−1.0316	−0.4950
2	−0.4950	−0.4950	−0.3054	0
3	0	0	0.2698	0.4950
4	0.4950	0.4950	0.9167	1.1630
5	1.1630	1.1630	2.1504	3.1630

improved as follows:

(a) $\mathrm{E}\,X_{(5)}\geqq 2$ since $\mathrm{E}\,X_{(n)}\geqq\delta$, a result that follows, for example, by a majorization argument (Marshall and Olkin [4], p. 348).

(b) $\mathrm{E}\,X_{(5)}\leqq 2.8$ since $\mathrm{E}\,X_{(n)}\leqq\mathrm{E}\,\bar{X}+(n-1)(\mathrm{E}\,S^2/n)^{1/2}$, where $(n-1)S^2=\sum(X_i-\bar{X})^2$ (Nagaraja [5]); $\mathrm{E}\,\bar{X}=\delta/n$ and $\mathrm{E}\,S^2=1+\delta^2/n$.

It will be clear that while methods (a) and (b) lead to better bounds for $\mathrm{E}\,X_{(5)}$ in the present instance, they will not necessarily do so. Both these approaches are confined to $X_{(n)}$ or other convex functions such as $\sum_{i=n+1-k}^{n} X_{(i)}$ (with corresponding results for lower extremes). For order statistics that are not extremes our methods tend to give much better bounds than other approaches. Note that, as a result of the outlier, the expected value of the sample median has increased from 0 to 0.2698 with bounds [0, 0.4950]; corresponding figures for the trimmed mean based on the three central order statistics are 0.2937 and [0, 0.5527].

IOWA STATE UNIVERSITY

REFERENCES

[1] Arnold, B. C. and Groeneveld, R. A. (1979). Bounds on expectations of linear systematic statistics based on dependent samples, *Ann. Statist.*, **7**, 220-223. Correction: **8**, 1402.

[2] David, H. A., Kennedy, W. J. and Knight, R. D. (1977). Means, variances, and covariances of normal order statistics in the presence of an outlier, *Selected Tables in Mathematical Statistics*, **5**, 75-204.

[3] Mallows, C. L. and Richter, D. (1969). Inequalities of Chebyshev type involving conditional expectations, *Ann. Math. Statist.*, **40**, 1922-1932.

[4] Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

[5] Nagaraja, H. N. (1981). Some finite sample results for the selection differential, *Ann. Inst. Statist. Math.*, **33**, 437-448.

[6] Smith, N. L. and Tong, Y. L. (1983). Inequalities for functions of order statistics under an additive model, *Ann. Inst. Statist. Math.*, **35** 255-265.