Ann. Inst. Statist. Math.
Vol. 45, No. 2, 243–247 (1993)

RELATIONSHIPS BETWEEN MOMENTS OF TWO RELATED SETS OF ORDER STATISTICS AND SOME EXTENSIONS

N. BALAKRISHNAN*[1], Z. GOVINDARAJULU[2] AND K. BALASUBRAMANIAN[3]

[1] *Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1*
[2] *Department of Statistics, University of Kentucky, Lexington, KY 40506-0027, U.S.A.*
[3] *Indian Statistical Institute, SJS Sansanwal Marg, New Delhi 110016, India*

(Received September 2, 1991; revised April 21, 1992)

Abstract. Govindarajulu expressed the moments of order statistics from a symmetric distribution in terms of those from its folded form. He derived these relations analytically by dividing the range of integration suitably into parts. In this paper, we establish these relations through probabilistic arguments which readily extend to the independent and non-identically distributed case. Results for random variables having arbitrary multivariate distributions are also derived.

Key words and phrases: Double exponential distribution, exponential distribution, order statistics, outliers, permanents, recurrence relations.

1. Introduction

If $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics obtained from a random sample of size n from a random variable X symmetric about 0 and $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ are the corresponding order statistics from $Y = |X|$, Govindarajulu (1963) expressed the moments of $X_{r:n}$ in terms of the moments of $Y_{r':n'}$, $1 \le r' \le n' \le n$. He derived these relations by dividing the range of integration and manipulating the integrals algebraically. Govindarajulu (1966) applied these results to compute the means, variances and covariances of order statistics from a double exponential distribution by making use of the known explicit expressions of these quantities from a standard exponential distribution.

Recently, Balakrishnan (1988a) extended the results of Govindarajulu to a single-outlier model which were used by Balakrishnan and Ambagaspitiya (1988) to study the robustness features of linear estimators for the parameters of a double exponential distribution. Balakrishnan's results for the single-outlier model has

* The first author would like to thank the Natural Sciences and Engineering Research Council of Canada for funding this research.

been extended by him (Balakrishnan (1989)) to the case when X_i's are independent, symmetric but non-identical by applying the methods used by Govindarajulu (1963) and Balakrishnan (1988*b*).

In this paper, we first of all derive Govindarajulu's result by a probabilistic argument. This method lends itself easily to extend the results to the case when X_i's are marginally symmetric but not necessarily independent. Further extension of the result when X_i's jointly have an arbitrary multivariate distribution is also presented.

2. Probabilistic proof for the i.i.d. case and an extension

Let us denote $E(X_{r:n})$ by $\mu_{r:n}$, $E(X_{r:n}^k)$ by $\mu_{r:n}^{(k)}$ for $k \geq 2$, $E(X_{r:n}X_{s:n})$ by $\mu_{r,s:n}$ for $1 \leq r < s \leq n$, and similarly $E(Y_{r:n})$ by $\nu_{r:n}$, $E(Y_{r:n}^k)$ by $\nu_{r:n}^{(k)}$ for $k \geq 2$, and $E(Y_{r:n}Y_{s:n})$ by $\nu_{r,s:n}$ for $1 \leq r < s \leq n$. Then, through division of the range of integration and direct algebraic manipulation of the resulting integrals, Govindarajulu (1963) established the following two relations between the two sets of moments of order statistics:

For $1 \leq r \leq n$ and $k = 1, 2, \ldots,$

$$\mu_{r:n}^{(k)} = 2^{-n}\left\{\sum_{i=0}^{r-1}\binom{n}{i}\nu_{r-i:n-i}^{(k)} + (-1)^k\sum_{i=r}^{n}\binom{n}{i}\nu_{i-r+1:i}^{(k)}\right\}; \tag{2.1}$$

and for $1 \leq r < s \leq n$,

$$\begin{aligned}\mu_{r,s:n} = 2^{-n}\Bigg\{\sum_{i=0}^{r-1}\binom{n}{i}\nu_{r-i,s-i:n-i} &+ \sum_{i=s}^{n}\binom{n}{i}\nu_{i-s+1,i-r+1:i} \\ &- \sum_{i=r}^{s-1}\binom{n}{i}\nu_{i-r+1:i}\nu_{s-i:n-i}\Bigg\}.\end{aligned} \tag{2.2}$$

We shall provide below a probabilistic proof for the relation in (2.1) while the relation in (2.2) may be proved easily using similar arguments.

Suppose $X_{r:n} > 0$. Then, the number of X's ≤ 0 can at most be $r-1$; let us suppose that number is i $(0 \leq i \leq r-1)$ with the remaining $n-i$ X's then constituting a random sample from Y. It is readily seen in this case that the conditional distribution of $X_{r:n}$ given that i $(< r)$ of the X's are negative is the same as the unconditional distribution of $Y_{r-i:n-i}$. Suppose $X_{r:n} \leq 0$. Then, the number of X's ≤ 0 will at least be r; let us suppose that number is i $(r \leq i \leq n)$. By noting then that these i X's constitute a random sample from $-Y$, it is readily seen in this case that the conditional distribution of $X_{r:n}$ given that i $(\geq r)$ of the X's are negative is the same as the unconditional distribution of $-Y_{i-r+1:i}$. The relation in (2.1) then follows immediately.

The aforementioned probabilistic argument enables us to extend Govindarajulu's relations in (2.1) and (2.2) to the case when the random variable X, with cumulative distribution function $F(x)$, is not necessarily symmetric. In this case, the conditional distribution of X given $X > 0$ and the conditional distribution of X given $X \leq 0$ will play the role of the distributions of Y and $-Y$, respectively. Now, by denoting the moments of order statistics from the former as before by $\nu_{s:m}$, $\nu_{s:m}^{(k)}$ and $\nu_{s,t:m}$ and the moments of order statistics from the latter by $\bar{\nu}_{s:m}$, $\bar{\nu}_{s:m}^{(k)}$ and $\bar{\nu}_{s,t:m}$, we have the following generalized relations:

For $1 \leq r \leq n$ and $k = 1, 2, \ldots,$

$$\mu_{r:n}^{(k)} = \sum_{i=0}^{r-1} \Pi_i \nu_{r-i:n-i}^{(k)} + \sum_{i=r}^{n} \Pi_i \bar{\nu}_{r:i}^{(k)}, \tag{2.3}$$

and for $1 \leq r < s \leq n$,

$$\mu_{r,s:n} = \sum_{i=0}^{r-1} \Pi_i \nu_{r-i,s-i:n-i} + \sum_{i=s}^{n} \Pi_i \bar{\nu}_{r,s:i} + \sum_{i=r}^{s-1} \Pi_i \bar{\nu}_{r:i} \nu_{s-i:n-i}, \tag{2.4}$$

where Π_i is the probability that exactly i X's ≤ 0 given by

$$\Pi_i = \binom{n}{i} (F(0))^i (1 - F(0))^{n-i}, \qquad i = 0, 1, \ldots, n.$$

3. Probabilistic proof for the i.ni.d. case and an extension

Suppose X_i, $i = 1, 2, \ldots, n$, are independent random variables with cdf $F_i(x)$ and pdf $f_i(x)$, $i = 1, 2, \ldots, n$, each symmetric about 0. Let $Y_i = |X_i|$, $i = 1, 2, \ldots, n$, and $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ and $Y_{1:n} \leq Y_{2:n} \leq \cdots \leq Y_{n:n}$ be the corresponding order statistics. Let us denote $E(X_{r:n})$ by $\mu_{r:n}$, $E(X_{r:n}^k)$ by $\mu_{r:n}^{(k)}$ for $k \geq 2$, and $E(X_{r:n} X_{s:n})$ by $\mu_{r,s:n}$ for $1 \leq r < s \leq n$. Further, let $Y_{r:n-i}^{[l_1, l_2, \ldots, l_i]}$ denote the r-th order statistic from $n - i$ Y variables obtained by deleting $Y_{l_1}, Y_{l_2}, \ldots, Y_{l_i}$ from $Y_1, Y_2, \ldots, Y_n$, with the corresponding k-th moment denoted by $\nu_{r:n-i}^{[l_1, l_2, \ldots, l_i](k)}$ and the product moment by $\nu_{r,s:n-i}^{[l_1, l_2, \ldots, l_i]}$.

With these notations, Balakrishnan (1989) has generalized the relations in (2.1) and (2.2) by using permanent representation of density of order statistics given by Vaughan and Venables (1972) and direct algebraic manipulation similar to those of Govindarajulu (1963). The generalized relations are as follows:

For $1 \leq r \leq n$ and $k = 1, 2, \ldots,$

$$\mu_{r:n}^{(k)} = 2^{-n} \left\{ \sum_{i=0}^{r-1} \sum_{1 \leq l_1 < \cdots < l_i \leq n} \nu_{r-i:n-i}^{[l_1, \ldots, l_i](k)} + (-1)^k \sum_{i=r}^{n} \sum_{1 \leq l_1 < \cdots < l_{n-i} \leq n} \nu_{i-r+1:i}^{[l_1, \ldots, l_{n-i}](k)} \right\}, \tag{3.1}$$

and for $1 \leq r < s \leq n$,

$$\mu_{r,s:n} = 2^{-n} \left\{ \sum_{i=0}^{r-1} \sum_{1 \leq l_1 < \cdots < l_i \leq n} \nu_{r-i,s-i:n-i}^{[l_1, \ldots, l_i]} \right. \tag{3.2}$$

$$+\sum_{i=s}^{n} \sum_{1\le l_1<\cdots<l_{n-i}\le n} \nu_{i-s+1,i-r+1:i}^{[l_1,\ldots,l_{n-i}]}$$

$$\left.-\sum_{i=r}^{s-1} \sum_{1\le l_1<\cdots<l_i\le n} \nu_{s-i:n-i}^{[l_1,\ldots,l_i]}\nu_{i-r+1:i}^{[l_{i+1},\ldots,l_n]}\right\},$$

where $\{l_1, l_2, \ldots, l_i\} \cap \{l_{i+1}, \ldots, l_n\} = \emptyset$.

We shall provide here a probabilistic proof for the relation in (3.1) while the relation in (3.2) may be similarly proved.

Suppose $X_{r:n} > 0$. Then, the number of X's ≤ 0 is at most $r-1$; let us suppose $X_{l_1}, X_{l_2}, \ldots, X_{l_i}$ are the only X's ≤ 0. It is then readily seen that the conditional distribution of $X_{r:n}$ given that $X_{l_1}, X_{l_2}, \ldots, X_{l_i}$ are negative is the same as the unconditional distribution of $Y_{r-i:n-i}^{[l_1,\ldots,l_i]}$. Suppose $X_{r:n} \le 0$. Then, the number of X's ≤ 0 is at least r. By using a similar argument, it is then seen that the conditional distribution of $X_{r:n}$ given that $X_{l_{n-i+1}}, \ldots, X_{l_n}$ are negative is the same as the unconditional distribution of $-Y_{i-r+1:i}^{[l_1,\ldots,l_{n-i}]}$. The relation in (3.1) then readily follows.

The probabilistic argument provided above makes it possible to extend Balakrishnan's relations in (3.1) and (3.2) to the case when the random variables X_i's are not necessarily symmetric. Let $\nu_{s:n-m}^{[l_1,\ldots,l_m](k)}$ and $\nu_{s,t:n-m}^{[l_1,\ldots,l_m]}$ denote the single and the product moments of order statistics from the conditional distribution of $n-m$ random variables obtained by deleting $X_{l_1}, \ldots, X_{l_m}$ from $X_1, X_2, \ldots, X_n$, given that all these $n-m$ variables are positive. Similarly, let $\bar{\nu}_{s:n-m}^{[l_1,\ldots,l_m](k)}$ and $\bar{\nu}_{s,t:n-m}^{[l_1,\ldots,l_m]}$ denote the corresponding moments of order statistics from the conditional distribution given that all the $n-m$ variables are negative. We may then prove the following generalized relations analogous to those in (2.3) and (2.4):

For $1 \le r \le n$ and $k = 1, 2, \ldots,$

$$\mu_{r:n}^{(k)} = \sum_{i=0}^{r-1} \sum_{1\le l_1<\cdots<l_i\le n} \Pi_{(l_1,l_2,\ldots,l_i)} \nu_{r-i:n-i}^{[l_1,\ldots,l_i](k)} \tag{3.3}$$

$$+\sum_{i=r}^{n} \sum_{1\le l_{i+1}<\cdots<l_n\le n} \Pi_{(l_1,\ldots,l_i)} \bar{\nu}_{r:i}^{[l_{i+1},\ldots,l_n](k)};$$

and for $1 \le r < s \le n$,

$$\mu_{r,s:n} = \sum_{i=0}^{r-1} \sum_{1\le l_1<\cdots<l_i\le n} \Pi_{(l_1,\ldots,l_i)} \nu_{r-i,s-i:n-i}^{[l_1,\ldots,l_i]} \tag{3.4}$$

$$+\sum_{i=s}^{n} \sum_{1\le l_{i+1}<\cdots<l_n\le n} \Pi_{(l_1,\ldots,l_i)} \bar{\nu}_{r,s:i}^{[l_{i+1},\ldots,l_n]}$$

$$+\sum_{i=r}^{s-1} \sum_{1\le l_1<\cdots<l_i\le n} \Pi_{(l_1,\ldots,l_i)} \nu_{s-i:n-i}^{[l_1,\ldots,l_i]} \bar{\nu}_{r:i}^{[l_{i+1},\ldots,l_n]},$$

where

$$\Pi_{(l_1,\ldots,l_i)} = p_{l_1} \cdots p_{l_i} q_{l_{i+1}} \cdots q_{l_n}, \tag{3.5}$$

with $p_i = P(X_i \leq 0) = 1 - q_i$.

Remark. It is easy to see that the relations in (3.3) and (3.4) simply reduce to those in (3.1) and (3.2) for the special case when all the X's are symmetric about 0 in which case $\Pi_{(l_1,\ldots,l_i)} = 2^{-n} \forall \{l_1, \ldots, l_i\} \subseteq \{1, 2, \ldots, n\} \forall i = 0, 1, \ldots, n$.

4. Results for the ni.ni.d. case

It is important to mention here that while proving the relations in (3.3) and (3.4) the only place where the independence of the random variables $X_1, X_2, \ldots, X_n$ is used is in the expression for $\Pi_{(l_1,\ldots,l_i)}$ in (3.5) in the form of a product of marginal probabilities. Therefore, if we redefine $\Pi_{(l_1,\ldots,l_i)}$ as

$$\Pi_{(l_1,\ldots,l_i)} = P\{X_{l_1} \leq 0, \ldots, X_{l_i} \leq 0, X_{l_{i+1}} > 0, \ldots, X_{l_n} > 0\}, \tag{4.1}$$

then the relations in (3.3) and (3.4) continue to hold even for the ni.ni.d. case, viz., when X_i's jointly have an arbitrary continuous multivariate distribution.

Acknowledgements

The authors would like to thank an Associate Editor and a referee for suggesting some changes which led to an improvement in the presentation of this paper.

References

Balakrishnan, N. (1988*a*). Recurrence relations among moments of order statistics from two related outlier models, *Biometrical J.*, **30**, 741–746.

Balakrishnan, N. (1988*b*). Recurrence relations for order statistics from n independent and non-identically distributed random variables, *Ann. Inst. Statist. Math.*, **40**, 273–277.

Balakrishnan, N. (1989). Recurrence relations among moments of order statistics from two related sets of independent and non-identically distributed random variables, *Ann. Inst. Statist. Math.*, **41**, 323–329.

Balakrishnan, N. and Ambagaspitiya, R. S. (1988). Relationships among moments of order statistics in samples from two related outlier models and some applications, *Comm. Statist. Theory Methods*, **17**, 2327–2341.

Govindarajulu, Z. (1963). Relationships among moments of order statistics in samples from two related populations, *Technometrics*, **5**, 514–518.

Govindarajulu, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population, *J. Amer. Statist. Assoc.*, **61**, 248–258.

Vaughan, R. J. and Venables, W. N. (1972). Permanent expressions for order statistics densities, *J. Roy. Statist. Soc. Ser. B*, **34**, 308–310.