J. R. Statist. Soc. B (1993)
55, *No.* 3, *pp.* 687-691

Duality Principle in Order Statistics

By K. BALASUBRAMANIAN and N. BALAKRISHNAN†

Indian Statistical Institute, New Delhi, India *McMaster University, Hamilton, Canada*

[Received May 1991. Final revision May 1992]

SUMMARY

We establish a duality principle for order statistics in the arbitrary case, using which many known dual results on order statistics can be deduced without a formal proof. The literature on order statistics is rich with such dual pairs derived under different assumptions such as continuity, absolute continuity, exchangeability, etc.

Keywords: EXCHANGEABLE RANDOM VARIABLES; IDENTITIES; ORDER STATISTICS; RECURRENCE RELATIONS

1. INTRODUCTION

Recurrence relations and identities for order statistics have been derived under many different assumptions for the underlying set of variables; see, for example, David (1981) and Arnold and Balakrishnan (1989). Two well-known relations for the independent and identically distributed (IID) case due to Srikantan (1962) and Govindarajulu (1963) are given by

$$F_{r:n}(x) = \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1} \binom{n}{i} F_{i:i}(x) \tag{1}$$

and

$$F_{r:n}(x) = \sum_{i=n-r+1}^{n} (-1)^{i-n+r-1} \binom{i-1}{n-r} \binom{n}{i} F_{1:i}(x), \tag{2}$$

where $F_{r:n}(x)$ is the cumulative distribution function of the rth-order statistic in a sample of size n.

It can be seen that there is a duality in these two relations, i.e. $F_{a:b}(x)$ in one, when replaced by $F_{b-a+1:b}(x)$, produces the other. Such a duality is also seen in many other results on order statistics for the cases of

(a) IID variables,
(b) exchangeable variables,
(c) independent and non-identical variables and
(d) arbitrary variables,

established by researchers including Joshi (1973), David and Joshi (1968), Young (1967), Balakrishnan (1988), Bapat and Beg (1988), Balasubramanian and Beg (1990)

†*Address for correspondence*: Department of Mathematics and Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1, Canada.

 0035-9246/93/55687 $2.00

and Sathe and Dixit (1990). In most of these papers, independent proofs have been given for these dual results.

In this paper, we formally establish a duality principle for order statistics by considering the arbitrary case which makes the proof of the dual result redundant once the primal result has been established.

2. NOTATION

Let $S = (X_1, X_2, \ldots, X_n)$ be a random vector, $T \subset \{1, 2, \ldots, n\}$ and ${}_S F_{\mathbf{r}:T}(\mathbf{x})$, with $\mathbf{r} = (r_1, r_2, \ldots, r_k)$ and $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, be the joint cumulative distribution function of the k order statistics $X_{r_1:|T|}, X_{r_2:|T|}, \ldots, X_{r_k:|T|}$ corresponding to the X_i, $i \in T$, with $1 \leqslant r_1 < r_2 < \ldots < r_k \leqslant |T|$. Similarly, let ${}_S \bar{F}_{\mathbf{r}:T}(\mathbf{x})$ be the joint survival function of the k order statistics $X_{r_1:|T|}, X_{r_2:|T|}, \ldots, X_{r_k:|T|}$ corresponding to the X_i, $i \in T$, with $1 \leqslant r_1 < r_2 < \ldots < r_k \leqslant |T|$. Let $\mathscr{S}$ be a family of random vectors of dimension n such that, if $S = (X_1, X_2, \ldots, X_n)$ is in $\mathscr{S}$, then $\bar{S} = (-X_1, -X_2, \ldots, -X_n)$ is also in $\mathscr{S}$. Such a family $\mathscr{S}$ may be called a reflective family. For example, the family consisting of all n-dimensional random vectors each of whose components are

(a) discrete

is clearly a reflective family. Similarly, the collection of all random vectors whose components are

(b) continuous,
(c) absolutely continuous,
(d) IID,
(e) independent and non-identically distributed,
(f) symmetric and
(g) exchangeable,

and any meaningful intersection of these collections, is a reflective family. Hence, it is easy to see that there are many interesting examples of reflective families.

3. DUALITY PRINCIPLE

With the previous notation, we prove the duality principle for order statistics for reflective families in the following theorem.

Theorem. Suppose that a relation of the form

$$\sum c_{\mathbf{r}:T}\, {}_S F_{\mathbf{r}:T}(\mathbf{x}) \equiv 0 \tag{3}$$

for all S in a reflective family $\mathscr{S}$, for every real $\mathbf{x}$, and where the summation is over all subsets T of $\{1, 2, \ldots, n\}$ and over $\mathbf{r} = (r_1, \ldots, r_k)$ with $1 \leqslant r_1 < r_2 < \ldots < r_k \leqslant |T|$, is satisfied. Then, the following dual relation is also satisfied by every $S \in \mathscr{S}$:

$$\sum c_{\mathbf{r}:T}\, {}_S F_{\mathbf{R}:T}(\mathbf{x}) \equiv 0, \tag{4}$$

where $\mathbf{R} = (R_1, R_2, \ldots, R_k) = (|T| - r_k + 1, \ldots, |T| - r_1 + 1)$.

Proof. By changing S to $\bar{S}$ in equation (3), we simply obtain

$$\sum c_{\mathbf{r}:T}\, {}_{\bar{S}}F_{\mathbf{r}:T}(\mathbf{x}) = \sum c_{\mathbf{r}:T}\, {}_{S}\bar{F}_{\mathbf{R}:T}(-\mathbf{x}) \equiv 0.$$

Since this equality holds for every real $\mathbf{x}$, we immediately have

$$\sum c_{\mathbf{r}:T}\, {}_{S}\bar{F}_{\mathbf{R}:T}(\mathbf{x}) \equiv 0. \tag{5}$$

Now by writing

$$F_{X_1, X_2, \ldots, X_k}(\mathbf{x}) = 1 + \sum_{l=1}^{k} (-1)^l \sum_{1 \leqslant i_1 < \ldots < i_l \leqslant k} \bar{F}_{\mathbf{X}_{(i)}}(\mathbf{x}_{(i)}), \tag{6}$$

where $\mathbf{X}_{(i)} = (X_{i_1}, \ldots, X_{i_l})$, $\mathbf{x}_{(i)} = (x_{i_1}, \ldots, x_{i_l})$ and $\mathbf{R}_{(i)} = (R_{i_1}, \ldots, R_{i_l}) = (|T| - r_{i_l} + 1, \ldots, |T| - r_{i_1} + 1)$, and observing that equation (5) implies that

$$\sum c_{\mathbf{r}:T} = 0 \tag{7}$$

and

$$\sum c_{\mathbf{r}:T}\, {}_{S}\bar{F}_{\mathbf{R}_{(i)}:T}(\mathbf{x}_{(i)}) \equiv 0 \tag{8}$$

(by setting all or other x_is to 0), the dual relation in equation (4) simply follows from equation (6) on using equations (7) and (8).

4. ILLUSTRATION

We illustrate a few dual pairs of relations that have already been published under varying assumptions, and we also establish some new relations by making use of the duality principle.

4.1. *Independent and Identically Distributed Case*

Downton (1966) established that

$$\sum_{r=1}^{n} (r-1)^{(k)} F_{r:n}(x) = \frac{n^{(k+1)}}{k+1} F_{k+1:k+1}(x),$$

where $n^{(k)} = n(n-1) \ldots (n-k+1)$ for $k = 1, 2, \ldots$ with $n^{(0)} = 1$, from which, on using the duality principle, we obtain his other result

$$\sum_{r=1}^{n} (n-r)^{(k)} F_{r:n}(x) = \frac{n^{(k+1)}}{k+1} F_{1:k+1}(x).$$

Srikantan (1962) established that

$$F_{r,s:n}(x, y) = \sum_{j=s-r}^{s-1} \sum_{m=n-s+j+1}^{n} (-1)^{n-m-r+1} \binom{j-1}{s-r-1} \binom{m-j-1}{n-s} \binom{n}{m} F_{1,j+1:m}(x, y)$$

from which, on using the duality principle, we obtain the new relation

$$F_{r,s:n}(x, y) = \sum_{j=s-r}^{n-r} \sum_{m=r+j}^{n} (-1)^{s-m} \binom{j-1}{s-r-1} \binom{m-j-1}{r-1} \binom{n}{m} F_{m-j,\, m:m}(x, y).$$

Similarly, Joshi and Balakrishnan (1982) proved that

$$\sum_{s=2}^{n-r+1} \binom{n-s}{r-1} F_{1,s:n}(x,y) + \sum_{s=2}^{r+1} \binom{n-s}{n-r-1} F_{1,s:n}(x,y) = \binom{n}{r} F_{1:n-r}(x) F_{1:r}(y)$$

from which, on using the duality principle, we obtain the new relation

$$\sum_{s=r}^{n-1} \binom{s-1}{r-1} F_{s,n:n}(x,y) + \sum_{s=n-r}^{n-1} \binom{s-1}{n-r-1} F_{s,n:n}(x,y) = \binom{n}{r} F_{n-r:n-r}(x) F_{r:r}(y).$$

4.2. *Exchangeable Case*

Balakrishnan (1987) showed that

$$\sum_{r=1}^{n} \frac{1}{(r+i-1)^{(i)}} F_{r:n}(x) = \frac{1}{(n+i-1)^{(i)}} \sum_{r=1}^{n} \binom{r+i-2}{i-1} \frac{F_{1:r}(x)}{r}$$

from which, by using the duality principle, we obtain his other result

$$\sum_{r=1}^{n} \frac{1}{(n-r+i)^{(i)}} F_{r:n}(x) = \frac{1}{(n+i-1)^{(i)}} \sum_{r=1}^{n} \binom{r+i-2}{i-1} \frac{F_{r:r}(x)}{r}.$$

4.3. *Independent and Non-identical Case*

By making use of the permanent expressions of order statistics given by Vaughan and Venables (1972), Balakrishnan (1988) and Bapat and Beg (1988) generalized relation (1) to

$$F_{r:n}(x) = \sum_{i=r}^{n} (-1)^{i-r} \binom{i-1}{r-1} \sum_{1 \leqslant l_1 < l_2 < \ldots < l_{n-i} \leqslant n} F_{i:i}^{[l_1,\ldots,l_{n-i}]}(x), \tag{9}$$

where $F_{i:i}^{[l_1,\ldots,l_{n-i}]}(x)$ denotes the distribution function of the largest order statistic in a sample of size i obtained by dropping the variables $X_{l_1}, \ldots, X_{l_{n-i}}$ from $X_1, \ldots, X_n$. Now, by simply using the duality principle, we observe their other relation

$$F_{r:n}(x) = \sum_{i=n-r+1}^{n} (-1)^{i-n+r-1} \binom{i-1}{n-r} \sum_{1 \leqslant l_1 < l_2 < \ldots < l_{n-i} \leqslant n} F_{1:i}^{[l_1,\ldots,l_{n-i}]}(x). \tag{10}$$

Applying duality to equations (3.2) and (3.5) in Balakrishnan *et al.* (1992) gives

$$F_{r,s:n}(x,y) = \sum_{j=s-r}^{n-r} \sum_{m=r+j}^{n} (-1)^{m+s} \binom{j-1}{s-r-1} \binom{m-j-1}{r-1} \sum_{1 \leqslant l_1 < l_2 < \ldots < l_{n-m} \leqslant n} F_{m-j,m:m}^{[l_1,\ldots,l_{n-m}]}(x,y)$$

and

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \frac{1}{n-s+1} F_{r,s:n}(x,y) = \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \left\{(s-1)\binom{n-1}{s-1}\right\}^{-1} \times \sum_{1 \leqslant l_1 < \ldots < l_{n-s} \leqslant n} F_{s-r,s:s}^{[l_1,\ldots,l_{n-s}]}(x,y).$$

In these equations, $F_{i,j:m}^{[l_1,\ldots,l_{n-m}]}(x, y)$ denotes the joint distribution function of ith- and jth-order statistics in a sample of size m obtained by dropping the variables $X_{l_1}, \ldots, X_{l_{n-m}}$ from $X_1, \ldots, X_n$.

4.4. *Arbitrary Case*

Relations (9) and (10) have been proved for the arbitrary case recently by Sathe and Dixit (1990) from one of which the other follows by using the duality principle established in this paper.

ACKNOWLEDGEMENTS

The second author would like to thank the Natural Sciences and Engineering Research Council of Canada for funding this research. The authors would like to thank Mrs Edna Pathmanathan for the excellent typing of the manuscript. Thanks are also due to a referee and the Editor for suggesting some changes which led to a considerable improvement in the presentation of this paper.

REFERENCES

Arnold, B. C. and Balakrishnan, N. (1989) Relations, bounds and approximations for order statistics. *Lect. Notes Statist.*, **53**.

Balakrishnan, N. (1987) A note on moments of order statistics from exchangeable variates. *Communs Statist. Theory Meth.*, **16**, 855–861.

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

Balakrishnan, N., Bendre, S. M. and Malik, H. J. (1992) General relations and identities for order statistics from non-independent non-identical variables. *Ann. Inst. Statist. Math.*, **44**, 177–183.

Balasubramanian, K. and Beg, M. I. (1990) On expectations of functions of order statistics. *Sankhya* B, **52**, 103–114.

Bapat, R. B. and Beg, M. I. (1988) Order statistics for nonidentically distributed variables and permanents. *Sankhya* B, **51**, 79–93.

David, H. A. (1981) *Order Statistics*, 2nd edn. New York: Wiley.

David, H. A. and Joshi, P. C. (1968) Recurrence relations between moments of order statistics for exchangeable variates. *Ann. Math. Statist.*, **39**, 272–274.

Downton, F. (1966) Linear estimates with polynomial coefficients. *Biometrika*, **53**, 129–141.

Govindarajulu, Z. (1963) On moments of order statistics and quasi-ranges from normal populations. *Ann. Math. Statist.*, **34**, 633–651.

Joshi, P. C. (1973) Two identities involving order statistics. *Biometrika*, **60**, 428–429.

Joshi, P. C. and Balakrishnan, N. (1982) Recurrence relations and identities for the product moments of order statistics. *Sankhya* B, **44**, 39–49.

Sathe, Y. S. and Dixit, U. J. (1990) On a recurrence relation for order statistics. *Statist. Probab. Lett.*, **9**, 1–4.

Srikantan, K. S. (1962) Recurrence relations between the PDF's of order statistics, and some applications. *Ann. Math. Statist.*, **33**, 169–177.

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

Young, D. H. (1967) Recurrence relations between the P.D.F.'s of order statistics of dependent variables, and some applications. *Biometrika*, **54**, 283–292.