Ann. Inst. Statist. Math.
Vol. 44, No. 1, 177–183 (1992)

GENERAL RELATIONS AND IDENTITIES FOR ORDER STATISTICS FROM NON-INDEPENDENT NON-IDENTICAL VARIABLES

N. BALAKRISHNAN[1], S. M. BENDRE[1*] AND H. J. MALIK[2]

[1]*Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1*
[2]*Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1*

(Received May 18, 1990; revised November 26, 1990)

Abstract. Some recurrence relations and identities for order statistics are extended to the most general case where the random variables are assumed to be non-independent non-identically distributed. In addition, some new identities are given. The results can be used to reduce the computations considerably and also to establish some interesting combinatorial identities.

Key words and phrases: Order ststistics, recurrence relations and identities, non-independent non-identical variables.

1. Introduction

Let $X_1, X_2, \dots, X_n$ be a sequence of n random variables with corresponding order statistics $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$. Let $F_{r,n}(x)$ denote the marginal distribution function of $X_{r,n}$ and $F_{r,s,n}(x,y)$ denote the joint distribution function of $X_{r,n}$ and $X_{s,n}$, $1 \le r < s \le n$. The well-known recurrence relations when the random variables are i.i.d. are given by

$$rF_{r+1,n}(x) + (n-r)F_{r,n}(x) = nF_{r,n-1}(x), \qquad 1 \le r \le n-1$$

and

$$\begin{aligned} &rF_{r+1,s+1,n}(x,y) + (s-r)F_{r,s+1,n}(x,y) + (n-s)F_{r,s,n}(x,y) \\ &\qquad = nF_{r,s,n-1}(x,y), \qquad 1 \le r < s \le n. \end{aligned}$$

These are proved in terms of raw moments in David ((1981), pp. 46–49). The relations have been generalized to the cases when the random variables are exchangeable (Young (1967), David and Joshi (1968)) and when they are independent but non-identically distributed (Balakrishnan (1988)). Recently, Sathe and

* On leave from Department of Statistics, University of Bombay, Vidyanagari, Bombay 400098, India.

Dixit (1990) extended the relations to the most general case where no assumption is made on the underlying joint distribution function of the variables. The recurrence relations are given by

$$(1.1) \qquad rF_{r+1,n}(x) + (n-r)F_{r,n}(x) = \sum_{i=1}^{n} F_{r,n-1}^{(i)}(x), \qquad 1 \le r \le n-1,$$

and

$$(1.2) \qquad rF_{r+1,s+1,n}(x,y) + (s-r)F_{r,s+1,n}(x,y) + (n-s)F_{r,s,n}(x,y) = \sum_{i=1}^{n} F_{r,s,n-1}^{(i)}(x,y), \qquad 1 \le r < s \le n-1$$

where $F_{r,n-1}^{(i)}(x)$ and $F_{r,s,n-1}^{(i)}(x,y)$ denote the distribution function of $X_{r,n-1}$ and $(X_{r,n-1}, X_{s,n-1})$ respectively, in a sample of size $n-1$ obtained on dropping X_i from the original sample of size n.

The need for recurrence relations and identities is well-established in the literature and for details we refer to the recent monograph by Arnold and Balakrishnan (1989). In this paper we generalize some of the established results to the non-independent non-identically distributed (ni.ni.d.) variables and also establish some new identities. The results greatly reduce the amount of direct computations when the random variables are not necessarily i.i.d. The results can also be used to establish some interesting combinatorial identities following the methods illustrated by Joshi (1973) and Joshi and Balakrishnan (1981).

In order to facilitate the proofs of various results presented in the following sections, we present below a lemma which gives a combinatorial identity satisfied by a complete beta function.

LEMMA 1.1. *For real positive k and c and a positive integer b,*

$$\sum_{a=0}^{b} (-1)^a \binom{b}{a} B(a+k, c) = B(k, c+b)$$

where $B(\cdot,\cdot)$ is the beta function.

PROOF. Consider

$$\sum_{a=0}^{b} (-1)^a \binom{b}{a} B(a+k, c) = \sum_{a=0}^{b} (-1)^a \binom{b}{a} \int_0^1 u^{a+k-1}(1-u)^{c-1} du.$$

On changing the order of summation and integration we get

$$= \int_0^1 \left\{ \sum_{a=0}^{b} (-1)^a \binom{b}{a} u^a \right\} u^{k-1}(1-u)^{c-1} du,$$

hence the result. For an alternative proof based on generating functions, we refer to Riordan (1968).

Remark 1. The lemma holds true for incomplete beta integrals in general and we get

$$\sum_{a=0}^{b}(-1)^a\binom{b}{a}I_p(a+k,c)=I_p(k,c+b)$$

where $I_p(a,b)$ is defined as the incomplete beta integral given by

$$I_p(a,b)=\int_0^p u^{a-1}(1-u)^{b-1}du, \qquad p\in(0,1).$$

2. Relations for marginal distribution functions

Let $F_{r,m}^{[i_1,\dots,i_{n-m}]}(x)$, $1\le r\le m\le n$ denote the distribution function of r-th order statistic in a sample of size m obtained on dropping $X_{i_1},X_{i_2},\dots,X_{i_{n-m}}$ from the original sample of size n. Further, let

$$H_{r,m}(x)=\sum_{1\le i_1<i_2<\cdots<i_{n-m}\le n} F_{r,m}^{[i_1,\dots,i_{n-m}]}(x).$$

For $m=n$, $H_{r,n}(x)=F_{r,n}(x)$, $1\le r\le n$. Also, when the variables are identically distributed, $H_{r,m}(x)=\binom{n}{m}F_{r,m}(x)$.

Result 2.1. For $1\le r\le n$,

$$F_{r,n}(x)=\sum_{j=n-r+1}^{n}(-1)^{j+n-r+1}\binom{j-1}{n-r}H_{1,j}(x), \tag{2.1}$$

$$F_{r,n}(x)=\sum_{j=r}^{n}(-1)^{j+r}\binom{j-1}{r-1}H_{j,j}(x). \tag{2.2}$$

PROOF. We prove (2.1) and (2.2) follows on same lines. From (1.1), we have

$$F_{r,n}(x)=-\frac{n-r+1}{r-1}F_{r-1,n}(x)+\frac{1}{r-1}\sum_{i_1=1}^{n}F_{r-1,n-1}^{[i_1]}(x).$$

Upon using (1.1) to the r.h.s. of the above equation, we get

$$F_{r,n}(x)=\frac{(n-r+2)(n-r+1)}{(r-1)(r-2)}F_{r-2,n}(x)-2\frac{(n-r+1)}{(r-1)(r-2)}\sum_{i_1=1}^{n}F_{r-2,n-1}^{[i_1]}(x)$$
$$+\frac{1}{(r-1)(r-2)}\sum_{\substack{i_1=1\\ i_1\ne i_2}}^{n}\sum_{i_2=1}^{n}F_{r-2,n-2}^{[i_1,i_2]}(x).$$

By repeating this process of using (1.1) for the expression on the r.h.s. $r-1$ times and simplifying the resulting equation, we derive the relation in (2.1).

Remark 2. The relations in (2.1) and (2.2) have already been established by Balakrishnan (1988) for independent nonidentical variables and they reduce to the well-known identities of Srikantan (1962) for i.i.d. random variables.

Result 2.2.

$$\sum_{r=1}^{n} \frac{1}{r} F_{r,n}(x) = \sum_{r=1}^{n} B(r, n-r+1) H_{1,r}(x), \tag{2.3}$$

$$\sum_{r=1}^{n} \frac{1}{n-r+1} F_{r,n}(x) = \sum_{r=1}^{n} B(r, n-r+1) H_{r,r}(x). \tag{2.4}$$

PROOF. We prove (2.3) and (2.4) follows on same lines. From (2.1), we have

$$\sum_{r=1}^{n} \frac{1}{r} F_{r,n}(x) = \sum_{r=1}^{n} \frac{1}{r} \sum_{j=n-r+1}^{n} (-1)^{j+n-r+1} \binom{j-1}{n-r} H_{1,j}(x).$$

On interchanging the order of summation and making transformation, the r.h.s. reduces to

$$= \sum_{j=1}^{n} \left\{ \sum_{l=0}^{j-1} (-1)^{l} \binom{j-1}{l} \frac{1}{(n-j+l-1)} \right\} H_{1,j}(x).$$

From Lemma 1.1, the term inside the brackets {} is $B(j, n-j+1)$, thus establishing (2.3).

Remark 3. For identical variables, the result reduces to the identities given by Joshi (1973). The result also explains the presence of the factor $1/r$ in Joshi's identities.

For $i = 1, 2, \dots,$ define for a fixed n

$$C_{i+k-1} = \begin{cases} (n+i)(n+i+1)\cdots(n+i+k-2), & k = 2, 3, \dots \\ 1, & k = 1. \end{cases} \tag{2.5}$$

Then, by adopting a method similar to the one used in proving Result 2.2, we can prove the following two results.

Result 2.3. For $i, k = 1, 2, \dots,$

$$\sum_{r=1}^{n} F_{r,n}(x) / \{(r+i-1)(r+i)\cdots(r+i+k-2)\} \tag{2.6}$$

$$= \frac{1}{C_{i+k-1}} \sum_{r=1}^{n} \binom{k+r-2}{k-1} B(r, n-r+i) H_{1,r}(x),$$

$$(2.7)\qquad \sum_{r=1}^{n} F_{r,n}(x)/\{(n-r+i)(n-r+i+1)\cdots(n-r+i+k-1)\}$$
$$= \frac{1}{C_{i+k-1}} \sum_{r=1}^{n} \binom{k+r-2}{k-1} B(r, n-r+i) H_{r,r}(x).$$

Remark 4. For $i = 1$ and in case of identical random variables, (2.6) and (2.7) reduce to expressions (2) and (3) respectively, of Balakrishnan and Malik (1985).

Result 2.4. For $k = 1, 2, \ldots,$

$$(2.8)\qquad \sum_{r=1}^{n} F_{r,n}(x)/\{r(r+1)\cdots(r+k-1)(n-r+1)\cdots(n-r+k)\}$$
$$= \frac{1}{C_{2k}} \sum_{r=1}^{n} \binom{r+2k-2}{r-1} B(r, n-r+1)\{H_{1,r}(x) + H_{r,r}(x)\}$$

and for $k, l = 1, 2, \ldots,$

$$(2.9)\qquad \sum_{r=1}^{n} F_{r,n}(x)/\{r(r+1)\cdots(r+k-1)(n-r+1)\cdots(n-r+l)\}$$
$$= \frac{1}{C_{k+l}} \sum_{r=1}^{n} B(r, n-r+1)$$
$$\cdot \left\{ \binom{r+k+l-2}{k-1} H_{1,r}(x) + \binom{r+k+l-2}{l-1} H_{r,r}(x) \right\}$$

where C_{2k} and C_{k+l} are as defined in (2.5) with $i = 1$.

3. Relations for joint distribution functions

Let $F_{r,s,m}^{[i_1,\ldots,i_{n-m}]}(x, y)$, $1 \le r < s \le m \le n$, denote the joint distribution function of r-th and s-th order statistics from a sample of size m, obtained on dropping $X_{i_1}, X_{i_2}, \ldots, X_{i_{n-m}}$ from the original sample of size n. Further, let

$$H_{r,s,m}(x, y) = \sum_{1 \le i_1 < i_2 < \cdots < i_{n-m} \le n} F_{r,s,m}^{[i_1,\ldots,i_{n-m}]}(x, y)$$

where, for $m = n$, $H_{r,s,n}(x, y) = F_{r,s,n}(x, y)$, $1 \le r < s \le n$.

Result 3.1. For $1 \le r < s \le n$

$$F_{r,s,n}(x,y) = \sum_{j=r}^{s-1} \sum_{m=n-s+j+1}^{n} (-1)^{m+n-r-s+1} \cdot \binom{j-1}{r-1}\binom{m-j-1}{n-s} H_{j,j+1,m}(x,y), \tag{3.1}$$

$$F_{r,s,n}(x,y) = \sum_{j=s-r}^{s-1} \sum_{m=n-s+j+1}^{n} (-1)^{n-m-r+1} \cdot \binom{j-1}{s-r-1}\binom{m-j-1}{n-s} H_{1,j+1,m}(x,y), \tag{3.2}$$

$$F_{r,s,n}(x,y) = \sum_{j=s-r}^{n-r} \sum_{m=r+j}^{n} (-1)^{m+s} \cdot \binom{j-1}{s-r-1}\binom{m-j-1}{r-1} H_{m-j,m,m}(x,y). \tag{3.3}$$

The above given identities can be proved by starting with the recurrence relation in (1.2) and by repeatedly using it in a way similar to the one used in proving Result 2.1. These extend the identities given by Srikantan (1962) to those for joint distribution functions and also to the case of ni.ni.d. random variables. For i.i.d. variables, (3.2) has been established by Srikantan.

Now, by starting with the identities in Result 3.1 and following the lines as used in proving Result 2.2, we can prove the following result.

Result 3.2.

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \frac{1}{s-r} F_{r,s,n}(x,y) = \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} B(s-1, n-s+1) H_{r,r+1,s}(x,y), \tag{3.4}$$

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \frac{1}{r} F_{r,s,n}(x,y) = \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} B(s-1, n-s+1) H_{1,r+1,s}(x,y), \tag{3.5}$$

$$\sum_{r=1}^{n-1} \sum_{s=r+1}^{n} \frac{1}{n-s+1} F_{r,s,n}(x,y) = \sum_{r=1}^{n-1} \sum_{s=r+1}^{n} B(s-1, n-s+1) H_{r,s,s}(x,y). \tag{3.6}$$

Remark 5. In case of identical random variables, the result reduces to

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

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

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

which extends the identities given by Joshi (1973) for the joint distributions. As rightly pointed out by Joshi, these identities can be effectively used in checking the computations of the product moments of order statistics.

Remark 6. It should be pointed out, however, that if the relation in (1.1) (or equivalently (2.1) or (2.2)) and (1.2) (or equivalently (3.1), (3.2) or (3.3)) are employed in the computations of single moments and product moments respectively, then all the identities established in this paper will be automatically satisfied; see Arnold and Balakrishnan (1989) for details.

Acknowledgements

The authors would like to thank Mrs. Edna Pathmanathan for the excellent typing of the manuscript. NSERC is to be thanked for providing financial assistance for Dr. Bendre's visit to McMaster University through the operating research grants of Drs. Balakrishnan and Malik. The authors would also like to thank two referees for making some suggestions 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, *Lecture Notes in Statist.*, **53**, Springer, New York.

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

Balakrishnan, N. and Malik, H. J. (1985). Some general identities involving order statistics, *Comm. Statist. Theory Methods*, **14**, 333–339.

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

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

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

Joshi, P. C. and Balakrishnan, N. (1981). Applications of order statistics in combinatorial identities, *J. Combin. Inform. System Sci.*, **6**, 271–278.

Riordan, J. (1968). *Combinatorial Identities*, Wiley, New York.

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.

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.