RECURRENCE RELATIONS BETWEEN THE PDF'S OF ORDER STATISTICS, AND SOME APPLICATIONS

By K. S. Srikantan

Central Statistical Organisation, New Delhi

1. Summary. In this paper we obtain a recurrence relation between the pdf's (probability density functions) of order statistics. This relation is then extended to their cdf's (cumulative distribution functions) and to the expected values of a given function of each order statistic and a function of pairs of them. The interrelation between these results and other known formulas is shown. The application of the results to the preparation of tables of cdf's, expected values, variances, etc., of order statistics is indicated, and the numerical errors propagated in such computations analysed. Finally, an illustration is given. It is hoped that the unified approach presented here will provide greater insight and flexibility to procedures of numerical evaluation of expected values, etc., of order statistics.

2. Recurrence relation. Let $x_{t,n}$ denote the tth order statistic in a random sample of n observations from a universe with cdf $F(x)$ and pdf $f(x)$ at x. Then the pdf of $x_{t,n}$ at x is given by

$$(1) \quad p_{t,n}(x) = \{n!/[(n-t)!(t-1)!]\}[F(x)]^{t-1}[1-F(x)]^{n-t}f(x), \quad t = 1(1)n.$$

From the identity

$$F^t(1-F)^{n-t-1} \equiv F^{t-1}(1-F)^{n-t-1} - F^{t-1}(1-F)^{n-t}$$

we deduce the recurrence

$$(2) \qquad p_{t+1,n}(x) = \{n/t\}p_{t,n-1}(x) - \{(n-t)/t\}p_{t,n}(x), \qquad t = 1(1)n-1.$$

This is equivalent to Cole's result (2) (and also (3)) in [1], as either result is deducible from the other. Integrating both sides of (2) over $(-\infty, x)$, we obtain the following relation connecting the cdf's of $x_{t+1,n}$, $x_{t,n}$ and $x_{t,n-1}$:

$$(3) \qquad F_{t+1,n}(x) = \{n/t\}F_{t,n-1}(x) - \{(n-t)/t\}F_{t,n}(x), \qquad t = 1(1)n-1,$$

where $F_{t,n}(x)$ denotes the cdf of $x_{t,n}$ at x. On the other hand, multiplying both sides of (2) by a given function of x, $\varphi(x)$, and integrating over all x, we arrive at the following relation between the expected values of $\varphi(x_{t+1,n})$, $\varphi(x_{t,n})$ and $\varphi(x_{t,n-1})$:

$$(4) \quad E[\varphi(x_{t+1,n})] = \{n/t\}E[\varphi(x_{t,n-1})] - \{(n-t)/t\}E[\varphi(x_{t,n})], \quad t = 1(1)n-1.$$

Relation (3) can also be obtained as a particular case of (4) by setting $\varphi(x) =$ the characteristic function of x in $(-\infty, x)$. Similarly we can deduce, from (5)

Received October 3, 1960; revised April 17, 1961.

through (15) infra, results connecting the cdf's (or the corresponding probabilities) of order statistics. When $\varphi(x) = x^r$, (4) is equivalent to (4) of Cole [1], as either result is deducible from the other. Godwin [4] also gives results similar to (4) when $\varphi(x) = x$ and x^2. Result (4) has been obtained by Govindarajulu for moments of order statistics [5].

3. Interpretation of the result. Recurrence (4) expresses the expected value of a given function of the $(t + 1)$st order statistic in a sample of size n in terms of the expected values of the same function of the tth order statistics in samples of sizes n and $n - 1$. By induction it follows that the expected value of a given function of any order statistic in a sample of size n can be expressed in terms of the expected values of the same function of the first order statistics in samples up to size n. A similar result in terms of the largest order statistics was obtained by Cole [1] for "normalized" moments of order statistics. We have the following explicit solution for recurrence (4).

$$(5)\quad E[\varphi(x_{t,v})] = (-1)^{v-t+1}\left\{\sum_{j=v-t+1}^{v}(-1)^j\binom{v}{j}\binom{j-1}{v-t}E[\varphi(x_{1,j})]\right\},\qquad 1 \leqq t \leqq v \leqq n.$$

This may be established by induction, using (4). For a symmetric population, setting $\bar{w}_j = E(x_{j,j}) - E(x_{1,j}) = -2E(x_{1,j})$, where expectations are taken about the population mean, we obtain

$$E(x_{t+1,v}) - E(x_{t,v}) = \binom{1}{2}\binom{v}{t}(-1)^{t+1}\Delta^t\bar{w}_{v-t},$$

a result surmised by "Student" and proved by E. S. Pearson [9].

Relation (4) or (5) enables us to compute the expected values of all order statistics, their squares, etc., from the expected values of the first order statistics, their squares, etc. From (3) it is evident that similar remarks apply to the tabulation of the cdf's of order statistics. Govindarajulu has also indicated that moments of order statistics can be obtained from those of the lowest order statistics [5].

4. Other formulas deducible from (4). We can express $E[\varphi(x_{t,v})]$, $1 \leqq t \leqq v \leqq n$, in terms of the expected values of the function φ of any n order statistics in samples up to size n, provided that there is no identity relation between them which is deducible from (4). For instance, we have the following solution for (4) in terms of the greatest order statistics in samples up to size n:

$$(6)\quad E[\varphi(x_{t,v})] = (-1)^t\sum_{j=t}^{v}(-1)^j\binom{v}{j}\binom{j-1}{t-1}E[\varphi(x_{j,j})],\qquad 1 \leqq t \leqq v \leqq n,$$

which is derived from (5) by considering the stochastic variable $y = -x$ and the function $\varphi(-y)$ instead of x and $\varphi(x)$. For $\varphi(x) = x^r$, (6) is equivalent to the result of Cole [1] that "normalized" moments of order statistics can be obtained by successive differencing of those of the largest order statistics.

Using as pivots the expected values of the function of the median statistics in even samples, i.e., $E[\varphi(x_{k,2k})]$ and $E[\varphi(x_{k+1,2k})]$, $k = 1(1)[[(n+1)/2]]$, we get

$$
\begin{aligned}
E[\varphi(x_{t,v})] = t\binom{v}{t} \sum_{j=0}^{[[(v-2t+1)/2]]} (-1)^j \Bigg\{ & \binom{v-2t-j+1}{j} \\
& \cdot E[\varphi(x_{t+j,2t+2j})] \Big/ \left[\binom{2t+2j}{t+j}(t+j)\right] \\
& + \binom{v-2t-j}{j-1} E[\varphi(x_{t+j+1,2t+2j})] \Big/ \left[\binom{2t+2j}{t+j+1}(t+j+1)\right]\Bigg\}, \\
\binom{v-2t}{-1} = 0, \qquad & 1 \leqq t \leqq [[v/2]] \quad \text{and} \quad v = 1(1)n,
\end{aligned} \tag{7}
$$

where $[[\lambda]]$ denotes the greatest integer not exceeding λ. This result can be proved by induction. We can deduce the corresponding expansion for $[[(v+1)/2]] + 1 \leqq t \leqq v \leqq n$ from (7) by considering the variable $y = -x$ and $\varphi(-y)$ instead of x and $\varphi(x)$. For a symmetric population, defining

$$
\psi(k) = \{E(x_{k+1,2k}) - E(x_{k,2k})\} \Big/ \binom{2k}{k},
$$

and setting in (7),

$$
E(x_{k+1,2k}) = -E(x_{k,2k}) = \left(\frac{1}{2}\right)\{E(x_{k+1,2k}) - E(x_{k,2k})\} = \left(\frac{1}{2}\right)\binom{2k}{k}\psi(k),
$$

where expectations are taken about the population mean, we obtain result (2) of Godwin [3].

Yet another solution, with the expected values of the function φ of order statistics in the largest sample as pivots, is given by

$$
\begin{aligned}
E[\varphi(x_{t,v})] = {} & \{1/(v+1)(v+2)\cdots n\} \\
& \cdot\Bigg\{\binom{n-v}{0}(v-t+1)(v-t+2)\cdots(n-t)E[\varphi(x_{t,n})] \\
& + \binom{n-v}{1} t(v-t+1)(v-t+2)\cdots(n-t-1)E[\varphi(x_{t+1,n})] \\
& + \binom{n-v}{2} t(t+1)(v-t+1)(v-t+2)\cdots(n-t-2)E[\varphi(x_{t+2,n})] \\
& + \cdots + \binom{n-v}{n-v} t(t+1)(t+2)\cdots(t+n-v-1)E[\varphi(x_{t+n-v,n})]\Bigg\}, \\
& \qquad 1 \leqq t \leqq v < n.
\end{aligned} \tag{8}
$$

This can be proved by induction and is made use of in Section 6.

5. Propagation of errors in tabulation. In order to secure a desired degree of accuracy in tabulation of expected values of any given function in each order

TABLE 1

Upper limit to errors propagated in $E[\varphi(x_{t,v})]$ through (5) (or modified form of (4)) for sample sizes 10 and 20 when $E[\varphi(x_{1,j})]$ is correct to k decimal places

Order Statistic	Upper Limit to Error Propagated in $E[\varphi(x)]$ in Units of $1/(2.10^k)$		Order Statistic	Upper Limit to Error Propagated in $E[\varphi(x)]$ in Units of $1/(2.10^k)$ Sample size $n = 20$
	Sample size $n = 10$	Sample size $n = 20$		
(1)	(2)	(3)	(4)	(5)
$x_{1,n}$	1	1	$x_{11,20}$	127,574,017
$x_{2,n}$	19	39	$x_{12,20}$	216,408,063
$x_{3,n}$	161	721	$x_{13,20}$	299,565,057
$x_{4,n}$	799	8,399	$x_{14,20}$	335,478,783
$x_{5,n}$	2,561	69,121	$x_{15,20}$	299,565,057
$x_{6,n}$	5,503	427,007	$x_{16,20}$	208,470,015
$x_{7,n}$	7,937	2,053,633	$x_{17,20}$	109,051,905
$x_{8,n}$	7,423	7,868,927	$x_{18,20}$	40,370,175
$x_{9,n}$	4,097	24,379,393	$x_{19,20}$	9,437,185
$x_{10,n}$	1,023	61,616,127	$x_{20,20}$	1,048,575

statistic through either (4) or one of (5) through (8), it is necessary to set upper limits to the errors propagated in such tabulation. These errors may arise from two sources: (i) approximations in the evaluation of pivotal expectations and (ii) approximations in the coefficients used. We shall first discuss the errors propagated by (5) or (4) using $E[\varphi(x_{1,j})]$, $j = 1(1)n$, as pivots.

The coefficients of $E[\varphi(x_{1,j})]$ in (5) are necessarily integral, whereas the coefficients of $E[\varphi(x_{t,v-1})]$ and $E[\varphi(x_{t,v})]$ in (4) can take non-integral values also. Hence it appears that the propagation of errors from the second source will be eliminated completely if we use (5). However, the same result can be obtained by using (4) provided that, in the numerical computations, we do not approximate $\{v/t\}$ and $\{(v - t)/t\}$ but use, instead, the form

$$\{vE[\varphi(x_{t,v-1})] - (v - t)E[\varphi(x_{t,v})]\}/t$$

to evaluate $E[\varphi(x_{t+1,v})]$. The expression within curly brackets will then be exactly divisible by t and no error will be introduced due to approximations in the coefficients.

We may next consider the errors propagated through (5) (or modified computational form of (4)) due to evaluation of $E[\varphi(x_{1,j})]$ correct to, say, k decimal places. The upper limit to the propagated error is the sum of the absolute values of the integral coefficients in formula (5) multiplied by the maximum error in the pivotal values and is, therefore,

$$\sum_{j=v-t+1}^{v} \binom{v}{j} \binom{j-1}{v-t} \{1/(2.10^k)\}.$$

To give some indication of this upper limit, Table 1 has been prepared for sample sizes 10 and 20. It shows how much accuracy is needed in the evaluation of $E[\varphi(x_{1,j})]$, $j = 1(1)n$, in order to secure at least a given degree of accuracy in tables of $E[\varphi(x_{t,v})]$, $1 \leqq t \leqq v \leqq n$. For instance, to ensure that tables of $E[\varphi(x_{t,v})]$ are correct to four places of decimals in samples up to size 10, we have to evaluate $E[\varphi(x_{1,j})]$ correct up to the eighth decimal place. This will, of course, ensure that $E[\varphi(x_{2,j})]$ is correct up to six decimal places and $E[\varphi(x_{3,j})]$ and $E[\varphi(x_{4,j})]$ up to five places in samples not exceeding ten. Harter [5] also comments on the propagation of error when using forward recurrence, which apparently was first suggested in this context by Federer [6].

6. Propagation of errors by (8) and its advantages. It will be evident from the foregoing discussion that the magnitude of the propagated error depends on the pivotal values chosen for using (4). In order to study the differential propagation of errors by different sets of pivotal values, let us consider, as an alternative, the expectations of a given function of the order statistics in the largest sample as pivots and apply the modified backward recurrence

$$\text{(9)} \quad \begin{aligned} E[\varphi(x_{t,v})] = \{&(v + 1 - t)E[\varphi(x_{t,v+1})] \\ &+ tE[\varphi(x_{t+1,v+1})]\}/(v + 1), \quad 1 \leqq t \leqq v \leqq n. \end{aligned}$$

The formal solution to this recurrence with $E[\varphi(x_{t,n})]$, $t = 1(1)n$, as pivots is given by (8). In using (9), the errors propagated by the approximation of the coefficients $\{(v + 1 - t)/(v + 1)\}$ and $\{t/(v + 1)\}$ can be eliminated if we first work out tables of $\mu_{t,v}$ given by the recurrence (with integral coefficients)

$$\text{(10)} \qquad \mu_{t,v} = (v + 1 - t)\mu_{t,v+1} + t\mu_{t+1,v+1}, \qquad 1 \leqq t \leqq v < n,$$

with boundary conditions, $\mu_{t,n} = E[\varphi(x_{t,n})]$, $t = 1(1)n$; and then obtain

$$\text{(11)} \qquad E[\varphi(x_{t,v})] = \mu_{t,v}/\{(v + 1)(v + 2) \cdots n\}, \quad 1 \leqq t \leqq v < n.$$

From (11), it will be clear that if the errors in $\mu_{t,v+1}$, $t = 1(1)v + 1$, are less than ϵ in absolute magnitude, then the propagated error in $\mu_{t,v}$, $t = 1(1)v$, cannot exceed $(v + 1)\epsilon$, as the absolute values of the integral coefficients in this relation add up to $v + 1$. By induction it follows that if $E[\varphi(x_{t,n})]$, $t = 1(1)n$, is correct to k decimal places, then the error propagated in $\mu_{t,v}$ cannot exceed $(v + 1)(v + 2) \cdots n/(2.10^k)$. Therefore, the error in $E[\varphi(x_{t,v})]$ as given by (11) cannot exceed 10^{-k} since the maximum error propagated by the pivotal values is $\{1/(2.10^k)\}$ and the rounding off error after division by $(v + 1)$ $(v + 2) \cdots n$ cannot exceed $\{1/(2.10^k)\}$. Hence, tabulation of expectations of any given function of each order statistic, using those of the order statistics in the largest sample (evaluated correct to k decimal places) as pivots, through (10) and (11) will be correct to within one unit in the last decimal place. For this reason, it is best to use this set of pivotal values unless there are other strong grounds for a different choice such as ease in computation of the pivotal expectations. However, attention may be drawn to the inherent difficulties in

TABLE 2

Reconstruction of expected values of order statistics from standard normal distribution in samples up to size 10 *through recurrence* (4) (*series F*) *and through* (10) *and* (11) (*series B*)

Sample Size & Formula	(Expected Value of Order Statistic).(−10¹⁰) 1	2	3	4	5
2 *F*	56418 95835*				
2 *T*	56418 95835				
2 *B*	56418 95835				
3 *F*	84628 43753*				
3 *T*	84628 43753				
3 *B*	84628 43753				
4 *F*	102937 53730*	29701 13822			
4 *T*	102937 53730	29701 13823			
4 *B*	102937 53730	29701 13823			
5 *F*	116296 44736*	49501 89706			
5 *T*	116296 44736	49501 89705			
5 *B*	116296 44736	49501 89705			
6 *F*	126720 63606*	64175 50386	20154 68346		
6 *T*	126720 63606	64175 50388	20154 68338		
6 *B*	126720 63606	64175 50388	20154 68338		
7 *F*	135217 83756*	75737 42706	35270 69586		
7 *T*	135217 83756	75737 42706	35270 69592		
7 *B*	135217 83756	75737 42707	35270 69592		
8 *F*	142360 03060*	85222 48628	47282 24940	15251 43996	
8 *T*	142360 03060	85222 48625	47282 24949	15251 43995	
8 *B*	142360 03060	85222 48626	47282 24950	15251 43995	
9 *F*	148501 31622*	93229 74564	57197 07852	27452 59116	
9 *T*	148501 31622	93229 74567	57197 07829	27452 59191	
9 *B*	148501 31622	93229 74568	57197 07829	27452 59191	
10 *F*	153875 27308*	100135 70448	65605 91028	37576 47108	12266 77128
10 *T*	153875 27308	100135 70446	65605 91054†	37576 46970	12266 77523
10 *B*	153875 27308*	100135 70446*	65605 91054*	37576 46970*	12266 77523*

F—Values obtained by using forward recurrence (4). *T*—Values as given by Teichroew. *B*—Values obtained by using backward recurrence (10) and (11).

* Pivotal values: *F*—Values for first order statistics as given by Teichroew. *B*—Values for order statistics in sample size 10 as given by Teichroew.

† The value given in [10] is 65605 91057 which has been corrected to 65605 91054 by Teichroew.

Expected values of order statistics higher than the median may be obtained from

$$E(x_{t,n}) + E(x_{n-t+1,n}) = 0.$$

numerical evaluation, with a high degree of precision, of the expected values, etc., of order statistics in large samples due to the large binomial coefficients involved in the pdf's of such statistics. If an alternative set of pivotal values is used, re-working the expected values through (10) and (11) will provied a sound and useful check on the accuracy of the computations. E. S. Pearson [9], p. 152, has also noted the advantages of backward recurrence.

7. An illustration. As an illustration we have reconstructed, to ten decimal places, in Table 2 the expected values of order statistics from the standard normal population in samples up to size ten through (a) recurrence (4), using the expectations of first order statistics as pivots and (b) relations (10) and (11), using as pivots the expected values of order statistics in a sample of size ten given by Teichroew [10].

In the table we have also shown the corresponding expected values given by Teichroew to provide some indication of the errors propagated by using different sets of pivotal values. It is seen that in using expectations of first order statistics as pivots, there are differences of about three units (compared to Teichroew's values), in the tenth, ninth and eighth places of the expected values of the second, third and fourth (as also fifth) order statistics respectively. On the other hand, the errors introduced, by using as pivots the expectations of the order statistics in a sample of size ten, do not exceed one unit in the tenth decimal place of the expected value of any order statistic. This demonstrates the superiority of the latter procedure for numerical evaluation of, and as a computational check on, the expected values, etc., of order statistics.

8. Extension to pairs of order statistics. The joint pdf of a pair of order statistics $(x_{u,n}, x_{v,n})$ in a sample of size n, at the point (x, y), is given by

$$\begin{aligned} p_{u,v,n}(x, y) &= \{n!/[(u-1)!(v-u-1)!(n-v)!]\}[F(x)]^{u-1} \\ &\quad \cdot [F(y) - F(x)]^{v-u-1}[1 - F(y)]^{n-v}f(x)f(y) \quad \text{for } x \leqq y, \\ &= 0 \qquad \text{otherwise}, \qquad 1 \leqq u < v \leqq n. \end{aligned} \tag{12}$$

We may deduce, as in the case of (2), the following recurrence between the pdf's of pairs of order statistics:

$$\begin{aligned} p_{u+1,v+1,n}(x, y) &= \{np_{u,v,n-1}(x, y) - (n-v)p_{u,v,n}(x, y) \\ &\quad - (u-v)p_{u,v+1,n}(x, y)\}/u, \qquad 1 \leqq u < v < n. \end{aligned} \tag{13}$$

Multiplying both sides of (13) by the function $\psi(x, y)$ and integrating over all (x, y), we get

$$\begin{aligned} E[\psi(x_{u+1,n}, x_{v+1,n})] &= \{nE[\psi(x_{u,n-1}, x_{v,n-1})] - (n-v)E[\psi(x_{u,n}, x_{v,n})] \\ &\quad - (v-u)E[\psi(x_{u,n}, x_{v+1,n})]\}/u, \quad 1 \leqq u < v < n. \end{aligned} \tag{14}$$

Teichroew's result (6) [10] is similar to (14). An explicit solution of (14), in terms of the expected values of the function of the first order statistics paired with all other order statistics in each sample size, is provided by

$$
\begin{aligned}
&E[\psi(x_{u,q}, x_{v,q})] \\
(15) \qquad &= (-1)^{u-1}\sum_{t=0}^{u-1} (-1)^t \binom{q}{t} \sum_{s=t}^{u-1} \binom{q-v+s-t}{s-t} \binom{v-2-s}{u-1-s} \\
&\qquad \cdot E[\psi(x_{1,q-t}, x_{v-s,q-t})], \qquad 1 \leqq u < v \leqq q \leqq n.
\end{aligned}
$$

This relation has advantages similar to those of (5) mentioned in Section 3 and enables us to compute the $\{n(n^2 - 1)/6\}$ expected values of functions of the type $\psi(x_{u,q}, x_{v,q})$, $1 \leqq u < v \leqq q \leqq n$, in terms of the $\{n(n-1)/2\}$ expected values of $\psi(x_{1,q}, x_{v,q})$, $1 < v \leqq q \leqq n$. Using as pivots the expected values of the function ψ of pairs of consecutive order statistics, i.e., $E[\psi(x_{v,q}, x_{v+1,q})]$, $1 \leqq v < q \leqq n$, and setting $E[\psi(x_{v,q}, x_{v+1,q})] = \{q!/[(v-1)!(q-v-1)!]\}$ $\gamma(q-v, v)$, we obtain Godwin's result (4) [4] when $\psi(x, y) = xy$. Govindarajulu has obtained result (14) for product moments of order statistics [5].

Errors propagated by recurrence (14) using any set of pivotal values can be examined as in Sections 5 and 6. It appears that if we use the expectations of the function ψ of pairs of order statistics in the largest sample, i.e.,

$$E[\psi(x_{u,n}, x_{v,n})], \qquad 1 \leqq u < v \leqq n,$$

as pivots and compute $E[\psi(x_{u,q}, x_{v,q})]$, $1 \leqq u < v \leqq q \leqq n$, through relations similar to (10) and (11), the propagated error cannot exceed one unit in the last decimal place.

9. Acknowledgment. The author is grateful to Professor William Kruskal and to the referees for suggesting improvements in this paper and for making the presentation more concise and lucid. He also wishes to thank Dr. K. R. Nair of the Central Statistical Organization for some useful suggestions.

REFERENCES

[1] R. H. Cole, "Relations between moments of order statistics," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 308–310.

[2] Walter T. Federer, *Evaluation of Variance Components from a Group of Experiments with Multiple Classifications*, Iowa Agricultural Experiment Station Research Bulletin No. 380, 1951.

[3] H. J. Godwin, "On the estimation of dispersion by linear systematic statistics," *Biometrika*, Vol. 36 (1949), pp. 92–100.

[4] H. J. Godwin, "Some low moments of order statistics," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 279–285.

[5] Zakkula Govindarajulu, "On moments of order statistics and quasi-ranges from normal populations," Report No. 13, University of Minnesota, Department of Statistics, May, 1961.

[6] H. Leon Harter, "Expected values of normal order statistics," *Biometrika*, Vol. 48 (1961), pp. 151–165.

[7] C. Hastings, Jr., F. Mosteller, J. W. Tukey, and C. P. Winsor, "Low moments

for small samples: a comparative study of order statistics," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 413–426.

[8] H. L. JONES, "Exact lower moments of order statistics in small samples from a normal distribution," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 270–273.

[9] E. S. PEARSON, "A further note on the distribution of range in samples taken from a normal population," *Biometrika*, Vol. 18 (1926), pp. 173–194.

[10] D. TEICHROEW, "Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 410–426.