SOME DISTRIBUTION PROBLEMS OF ORDER STATISTICS FROM DISCRETE POPULATIONS

D. G. KABE

(Received March 20, 1969)

1. Introduction and summary

Khatri [3] gives a systematic account of some theory of ordered statistics for the discrete case. However, his derivations and his expressions, for the probability function of the ith ordered statistic and the joint probability function of ith and jth ordered statistics, are complicated. We give a simpler expression for the joint probability function. We apply some of our results to ordered statistics theory for the geometric distribution and to the discrete uniform distribution. The geometric distribution has been considered by Margolin and Winokur [4] who obtain the results by probabilistic arguments rather than analytical methods. Johnson and Leone [2] find the distribution of the range for the discrete uniform distribution by an involved method. We derive this distribution in a straightforward manner. Finally we consider the large sample behaviour of the ith ordered statistic. Since discrete ordered statistic theory is of importance in inverse sampling and Markov chains (see Margolin and Winokur [4]), we hope our paper may be at least of pedagogical interest.

2. Some discrete ordered statistics theory

Let $x_1 \leqq x_2 \leqq x_3 \leqq \cdots \leqq x_i \leqq \cdots \leqq x_j \leqq \cdots \leqq x_N$ denote an ordered sample of size N on a variate X whose distribution function is $P(x)=\sum_{r=0}^{x} P\{X=r\}$. Then the probability that $X_i < x$ is the same as the probability that the number of observations less than x is at least equal to i, which equals the cumulative binomial probability

$$(1) \qquad \sum_{r=i}^{N} \binom{N}{r} (P(x))^r (1-P(x))^{N-r} = \frac{N!}{(i-1)!(N-i)!} \int_0^{P(x)} u^{i-1}(1-u)^{N-i} du \, ,$$

a well-known result (see e.g., Wilks ([6], p. 152, example 6.22)). It follows that the probability function of the ith ordered variate is

$$(2)\qquad P\{X_i=x\}=\frac{N!}{(i-1)!(N-i)!}\int_{P(x-1)}^{P(x)} u^{i-1}(1-u)^{N-i}du\,.$$

We note the analogy between (1) and the corresponding result for the continuous case which is obtained by the same argument. Similarly we may note that

$$(3)\qquad \begin{aligned} &P\{X_i\leqq x_i,\, X_j\leqq x_j\}\\ &=\sum_{r=1}^{j}\sum_{s=j}^{N}\frac{N!\{P(x_i)\}^r\{P(x_j)-P(x_i)\}^{s-r}\{1-P(x_j)\}^{N-s}}{r!(s-r)!(N-s)!}\\ &=C\int_0^{P(x_i)}du\int_u^{P(x_j)}u^{i-1}(v-u)^{j-i-1}(1-v)^{N-j}dv\,, \end{aligned}$$

where

$$(4)\qquad C=N!/(i-1)!(j-i-1)!(N-j)!\,.$$

It follows that joint probability function of X_i and X_j is

$$(5)\qquad P\{X_i=x_i,\, X_j=x_j\}=C\int\int_{\Delta}u^{i-1}(v-u)^{j-i-1}(1-v)^{N-j}dvdu\,,$$

where the region Δ of integration is determined by the conditions

$$(6)\qquad x_i<x_j,\; P(x_i-1)\leqq u\leqq P(x_i)\leqq P(x_j-1)\leqq v\leqq P(x_j)\,.$$

An integral of type (5) has been evaluated by Wilks ([6], p. 330) by splitting the region Δ of integration into two parts. However, we shall use the procedure of integration by parts. Integrating (5) with respect to u first by parts and then with respect to v by parts, we find that

$$(7)\qquad \begin{aligned} &P\{X_i=x_i,\, X_j=x_j\}\\ &=\sum_{K=1}^{i}\sum_{t=1}^{N-j+1}\frac{N!}{(i-K)!(N-j+1-t)!(j-i+K-1+t)!}\\ &\quad\cdot[-\{P(x_i)\}^{i-K}\{P(x_j)-P(x_i)\}^{j-i+K-1+t}\{1-P(x_j)\}^{N-j-t+1}\\ &\quad+\{P(x_i)\}^{i-K}\{P(x_j-1)-P(x_i)\}^{j-i+K-1+t}\{1-P(x_j-1)\}^{N-j-t+1}\\ &\quad+\{P(x_i-1)\}^{i-K}\{P(x_j)-P(x_i-1)\}^{j-i+K-1+t}\{1-P(x_j)\}^{N-j-t+1}\\ &\quad-\{P(x_i-1)\}^{i-K}\{P(x_j-1)-P(x_i-1)\}^{j-i+K-1+t}\\ &\qquad\cdot\{1-P(x_j-1)\}^{N-j-t+1}]\\ &=-R(x_i,x_j)+R(x_i-1,x_j)+R(x_i,x_j-1)-R(x_i-1,x_j-1)\,, \end{aligned}$$

where

$$(8)\qquad R(x_i,x_j)$$

$$=P\{X_i>x_i, X_j\leqq x_j\}$$
$$=\sum_{K=1}^{i}\sum_{t=1}^{N-j+1}\frac{N!\{P(x_i)\}^{i-K}\{P(x_j)-P(x_i)\}^{j-i+K+t-1}\{1-P(x_j)\}^{N-j-t+1}}{(i-K)!(N-j-t+1)!} .$$

The expression (7) for the joint probability of X_i and X_j is simpler and readily usable than the alternative one given by Khatri ([3], p. 168, equation 6). Moreover, the interpretation of each term in equation (7) is obvious. Similar expressions like (7) may be derived for the joint probability distributions of three or more ordered values.

In case $X_i=X_j$, then (7) reduces to

$$(9)\qquad P\{X_i=X_j=\alpha\}$$
$$=\sum_{K=1}^{i}\sum_{t=1}^{N-j+1}\frac{N!}{(i-K)!(N-j+1-t)!(j-i+K-1+t)!}$$
$$\cdot\{P(\alpha-1)\}^{i-K}\{P(\alpha)-P(\alpha-1)\}^{j-i+K-1+t}\{1-P(\alpha)\}^{N-j+1-t} .$$

Setting $r=i-K$ and $s=N-j-t+1$ and changing the variables of summation from K and t to r and s we see that the result (9) agrees with the one given by Abdel-Aty ([1], p. 66, equation (14)). In (9) we set $i=1$, and $j=N$, and find that

$$(10)\qquad P\{X_1=X_N=\alpha\}=\{P(\alpha)-P(\alpha-1)\}^N , \qquad \alpha=0, 1, \cdots .$$

Again setting $i=1$, $j=N$ in (7) we find that

$$(11)\qquad P\{X_1=x_1, X_N=x_N\}$$
$$=\{P(x_N)-P(x_1-1)\}^N-\{P(x_N)-P(x_1)\}^N-\{P(x_N-1)P(x_1-1)\}^N$$
$$+\{P(x_N-1)-P(x_1)\}^N .$$

The formulae (10) and (11) are due to Siotani [5]. Khatri gives formulae for the mean and variances of X_i and X_j-X_i, however, in general these formulae are difficult to apply. Depending on the particular distributions under study, like the geometric and the discrete uniform, it is easier to obtain moments by using characteristic functions than by using the formulae directly.

3. Geometric and uniform order statistic theory

We consider the usual geometric distribution

$$(12)\qquad P\{X=x\}=pq^x, p+q=1 , \qquad x=0, 1, 2, \cdots .$$

It is easily seen that

$$(13)\qquad P(x)=p\sum_{r=0}^{x}q^r=1-q^{x+1} .$$

Note that the geometric distribution considered by Margolin and Winokur is a translated geometric distribution of the variate $X+1$. Thus if $X_1 \leqq X_2 \leqq \cdots \leqq X_N$ are N ordered variates from (12), then using (2) with $K = N!/(i-1)!(N-i)!$, we find that

$$P\{X_i = x\} = K \int_{1-q^x}^{1-q^{x+1}} u^{i-1}(1-u)^{N-i} du . \tag{14}$$

Now writing $u = 1-(1-u)$ and expanding u^{i-1} by the binomial theorem we write (14) as

$$\begin{aligned} P\{X_i = x\} &= K \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^j \int_{1-q^x}^{1-q^{x+1}} (1-u)^{N+j-i} du \\ &= K \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^j q^{x(N+j+1-i)} (1-q^{N-i+j+1})/(N-i+j+1) . \end{aligned} \tag{15}$$

The moment generating function $\Phi(\theta)$ of X_1 is

$$\Phi(\theta) = K \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^j (1-q^{N+j-i+1})/(1-e^{\theta} q^{N+j+1-i})(N-i+j+1) . \tag{16}$$

By using (16) we may obtain the first two moments of X_i and hence find $E(X_i+1)$ and $E(X_i+1)^2$, the results agree with the ones given by Margolin and Winokur ([4], p. 921, equations (9) and (10)). Now by using (10) we find that

$$P\{\text{range}=0\} = \sum_{\alpha=0}^{\infty} P\{X_N = X_1 = \alpha\} = P^N \sum_{\alpha=0}^{\infty} q^{\alpha N} = P^N/(1-q^N) , \tag{17}$$

a result which agrees with the one given by Margolin and Winokur ([4], p. 918). Similarly by using (11) and setting $X_N - X_1 = R$ and summing over the values of X_1, we find the probability function of the range R to be

$$P\{R = x\} = \frac{(1-q^{x+1})^N - (1-q^x)^N - q^N(1-q^x)^N + q^N(1-q^{x-1})^N}{(1-q^N)} . \tag{18}$$

The result (18) has been obtained by Margolin and Winokur ([4], p. 919, equation (4)) by probabilistic arguments.

Now consider the discrete uniform distribution

$$P\{X = x\} = \frac{1}{K} , \qquad x = 1, 2, \cdots, K . \tag{19}$$

Here $P(x) = x/K$, hence using (11) and setting $X_N - X_1 = R$, we find the joint probability of R and X_1 to be

$$P\{R = x, X_1 = x_1\} = [(x+1)^N - 2x^N + (x-1)^N]/K^N . \tag{20}$$

Obviously $1 \leqq X_1 < X_1 + R \leqq K$, hence X_1 is to be summed in (20) from 1 to $K-R$. Thus from (20) we find that

$$P\{R=x\}=(K-x)[(x+1)^N-2x^N+(x-1)^N]/K^N\,, \tag{21}$$

a result which agrees with the one given by Johnson and Leone ([2], p. 169, equation (6.29)).

Now we shall consider the large sample behaviour of the ith ordered statistic as i is fixed but N tends to ∞. The behaviour of the distribution of the ith ordered statistic as $N \to \infty$ need not be identical with that of the continuous case, see e.g., Wilks ([6], p. 269, Sections 9.6.1 and 9.6.2), although Khatri ([3], p. 168) due to similarity of equation (1) with the corresponding case hopes so, and Johnson and Leone ([2], p. 168, Section 6.8) remark that the theory of the distribution of the ordered discrete variable follows similar lines as the theory of ordered continuous variables. To prove this consider the left-hand side expression of the equation (1). In case we could keep $NP(x)=W$ fixed, then taking limit as $N \to \infty$ we see that this limit is the tail of the Poisson distribution

$$\sum_{x=i}^{\infty} e^{-W}W^x/x! = \int_0^W e^{-W}W^i/\Gamma(i+1)\,, \tag{22}$$

and in this case the large sample behaviour of the ith ordered variable would be identical with that of the continuous analogue. However, $NP(x)$, on account of the discreteness of $P(x)$, cannot be always kept constant and hence may tend to ∞. Thus we consider the smallest ordered statistic from a Poisson distribution with parameter λ. Here

$$P(x)=\begin{cases} 0 & \text{if } x<0\,, \\ e^{-\lambda} & \text{if } 0 \leqq x < 1\,, \end{cases} \tag{23}$$

and thus as x decreases $P(x)$ abruptly jumps from $e^{-\lambda}$ to 0, and for any $x \geqq 0$ $NP(x) \geqq Ne^{-\lambda} \to \infty$ as $N \to \infty$. It follows that $NP(x)$ cannot be fixed at least in this case.

ST. MARY'S UNIVERSITY, HALIFAX, N.S., CANADA

REFERENCES

[1] S. H. Abdel-Aty, "Ordered variables in discontinuous distributions," *Stat. Neerlandica*, 8 (1954), 61-82.

[2] N. L. Johnson and F. C. Leone, *Statistics and Experimental Designs*, I, John Wiley, New York, 1964.

[3] C. G. Khatri, "Distributions of order statistics for the discrete case," *Ann. Inst. Statist. Math.*, 14 (1963), 167-171.

[4] B. H. Margolin and H. S. Winokur, "Exact moments of the order statistics of the geometric distribution and their relation to inverse sampling and reliability of redundant systems," *J. Amer. Statist. Ass.*, 62 (1967), 915-925.
[5] Minoru Siotani, "Order statistics for discrete case with a numerical application to the binomial distribution," *Ann. Inst. Statist. Math.*, 8 (1956), 95-104.
[6] S. S. Wilks, *Mathematical Statistics*, John Wiley, New York, 1962.