Statistics & Probability Letters 14 (1992) 103–105
North-Holland
27 May 1992

Bounds on distribution functions of order statistics for dependent variates

G. Caraux and O. Gascuel
Département d'Informatique Fondamentale, Montpellier, France

Received July 1991

Abstract: Upper and lower bounds are given for $F_{X_{r:n}}$, the distribution function of the rth order statistic from n possibly dependent random variables. We show that these bounds may be reached when the random variables have a common distribution function. For any distribution function F we may construct a set of n exchangeable variates (with c.d.f. F), whose dependency structure is such that the bounds are attained.

Keywords: Distribution function of order statistics, upper and lower bounds, dependent variates.

1. Introduction

Let $X_1, \dots, X_n$ be identically distributed random variables with order statistics denoted as $X_{1:n} \leqslant \cdots \leqslant X_{j:n} \leqslant \cdots \leqslant X_{n:n}$. Let $F_{X_{r:n}}$ be the distribution function of the rth smallest order statistics.

It is well known that if X_j are independent, with common distribution function F, then

$$F_{X_{r:n}}(x) = \sum_{j=r}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}. \tag{1}$$

When identical and independence conditions for the distributions are dropped but the variates are exchangeable, the formula

$$F_{X_{r:n}}(x) = \sum_{j=r}^{n} (-1)^{j-r} \binom{j-1}{r-1} \binom{n}{j} \times \Pr\left(\bigcap_{k=1}^{j} \{X_k \leqslant x\} \right) \tag{2}$$

is given by David (1970, p.82).

Correspondence to: Prof. G. Caraux, Dept. d'Informatique Fondamentale, LIRMM, 860 rue de Saint Priest, 34100 Montpellier, France.

Maurer and Margolin (1976) extend the expression (2) to random variables that are not necessarily exchangeable. The expression (3.2) that they found, like (2), may be used in practice only when the dependency structure of the variates is known.

An extensive literature has developed on inequalities involving linear functions of order statistics and their expectations (see David (1988)). Bounds on expectations of order statistics for dependent random variables may also be found in Arnold (1988) and Hoover (1989).

In this paper we propose upper and lower bounds for $F_{X_{r:n}}(x)$ that hold whatever the dependency structure. These bounds may be used in statistical tests, for instance, when the dependency structure is unknown.

These bounds may be reached when the variates have the same distribution function. This generalizes some results obtained by Lai and Robbins (1976) for extremal statistics.

2. Bounds on $F_{X_{r:n}}(x)$

Minimum $X_{1:n}$ and maximum $X_{n:n}$ are easy to treat. Application of Bonferroni's inequality gives

$$F(x) \leqslant F_{X_{1:n}}(x) \leqslant nF(x) \tag{3}$$

and similarly,

$$F(x) \geqslant F_{X_{n:n}}(x) \geqslant 1 - n(1 - F(x)).$$

In the general case, for any r, we have:

Proposition 1. *Let $X_1, \ldots, X_n$ be a set of n random variables which are identically distributed (with c.d.f. F). Then*

$$F_{X_{r:n}}(x) \leqslant \inf\left(\frac{n}{r}F(x), 1\right). \tag{4}$$

Proof. Let $1_{\{X_j \leqslant x\}}$ be the indicator of the event $\{X_j \leqslant x\}$ and let

$$\nu_n(x) = \sum_{j=1}^{n} 1_{\{X_j \leqslant x\}}$$

be the number of X_j which are $\leqslant x$. Then $\nu_n(x)$ is a non-negative random variable and, according to Markov's inequality, we have, for $\lambda > 0$,

$$\Pr(\nu_n(x) \geqslant \lambda) \leqslant \frac{E(\nu_n(x))}{\lambda}. \tag{5}$$

If $\lambda = r$ ($r \in \{1, \ldots, n\}$), then the left-hand side in (5) is $F_{X_{r:n}}(x)$. Moreover, since $E(1_{\{X_j \leqslant x\}}) = F(x)$, then

$$E(\nu_n(x)) = nF(x), \tag{6}$$

and thus

$$F_{X_{r:n}}(x) \leqslant \frac{n}{r}F(x).$$

Since $F_{X_{r:n}}(x) \leqslant 1$, the inequality (4) is proven. □

Proposition 2. *Let $X_1, \ldots, X_n$ be a set of n random variables which are identically distributed (with c.d.f. F). Then*

$$F_{X_{r:n}}(x) \geqslant \sup\left(0, 1 - \frac{n}{n-r+1}(1 - F(x))\right). \tag{7}$$

Proof. If x is a continuity point of F then it is clear that the inequality (7) is contained in (4) since

$$F_{X_{r:n}}(x) = 1 - F_{-X_{n-r+1:n}}(-x).$$

In the general case we define $p_j = \Pr(\nu_n(x) = j)$. Developing the let-hand side of (6) we see that

$$\begin{aligned} nF(x) &= (r-1)\sum_{j=0}^{r-1} p_j - \sum_{j=0}^{r-1}(r-1-j)p_j \\ &\quad + n\sum_{j=r}^{n} p_j - \sum_{j=r}^{n}(n-j)p_j \\ &= (r-1)\left(1 - F_{X_{r:n}}(x)\right) + nF_{X_{r:n}}(x) - \mathscr{K}. \end{aligned}$$

Since $\mathscr{K} \geqslant 0$ it follows that

$$F_{X_{r:n}}(x) \geqslant 1 - \frac{n}{n-r+1}(1 - F(x)). \qquad \square$$

Propositions 1 and 2 can be generalized to cases in which the X_j are not identically distributed.

Proposition 3. *Let $X_1, \ldots, X_n$ be a set of n random variables with distribution functions $F_{X_1}, \ldots, F_{X_n}$. Then*

$$\sup\left(0, 1 - \frac{\sum_{j=1}^{n}\left(1 - F_{X_j}(x)\right)}{n-r+1}\right) \leqslant F_{X_{r:n}}(x) \leqslant \inf\left(\frac{\sum_{j=1}^{n} F_{X_j}(x)}{r}, 1\right).$$

The proof of this proposition follows the same lines as that of Propositions 1 and 2. We just have to change the equality (6) to

$$E(\nu_n(x)) = \sum_{j=1}^{n} F_{X_j}(x).$$

3. Optimality of the bounds

In this part, we construct a set of exchangeable variates, the dependency structure of which is such that the bound (4) is attained. Using the same principle, we may also prove that the bound (7) may be reached.

Proposition 4. *Let F be any distribution function. Then, for any integer r and n ($1 \leqslant r \leqslant n$), a set of n dependent random variables $X_1, \ldots, X_n$ can be found, with distribution function F, for which the bound (4) is reached.*

Proof. Let $\mathscr{U}$ be a random variable uniformly distributed in the range [0, 1] and let

$$U_j^* = \frac{r}{n}\mathscr{U} \qquad \forall j \in \{1,\dots,r\}, \tag{8a}$$

$$U_j^* = \frac{j-1}{n} + \frac{1}{n}\mathscr{U} \quad \forall j \in \{r+1,\dots,n\}. \tag{8b}$$

Then U_j^* are uniformly distributed in the range

$[0, r/n]$ when $j \in \{1,\dots,r\}$,

or in the range

$[(j-1)/n, j/n]$ when $j \in \{r+1,\dots,n\}$.

Now let $\sigma = \{\sigma_1, \sigma_2,\dots,\sigma_n\}$ be a random permutation of the set $\{1, 2,\dots, n\}$ defined as a trial consisting of drawing, without replacement, each element of the set $\{1, 2,\dots, n\}$. In this scheme, σ_j is the number associated with the outcome of the jth draw.

Let

$$U_j = U_{\sigma_j}^*. \tag{9}$$

Since $U_{r:n} = U_r^*$, we see from (8) that

$$F_{U_{r:n}}(u) = \inf\left(\frac{n}{r}u, 1\right) \quad \forall u \in [0, 1]. \tag{10}$$

Using

$$\Pr(U_j \leqslant u) = \Pr\left(\bigcap_{k=1}^{n} \left(\{\sigma_j = k\} \cap \{U_k^* \leqslant u\}\right)\right)$$
$$= \frac{1}{n}\sum_{k=1}^{n} \Pr(U_k^* \leqslant u), \tag{11}$$

we easily see that the random variables U_j are distributed identically. Moreover, it follows from the definition (8) and from the right-hand side of (11) that for any U_j,

$$F_{U_j}(u) = u \quad \forall u \in [0, 1]. \tag{12}$$

Then the random variables U_j are identically distributed with a uniform distribution function in the range [0, 1]. By anamorphosis let

$$X_j = F^{-1}(U_j), \tag{13}$$

where

$$F^{-1}(u) = \inf\{x \mid F(x) \geqslant u\}.$$

Each random variable X_j has distribution function F.

Also,

$$X_{r:n} = F^{-1}(U_{r:n}) \quad \text{and} \quad F_{X_{r:n}}(x) = F_{U_{r:n}}(F(x)).$$

Finally, using (10),

$$F_{X_{r:n}}(x) = \inf\left(\frac{n}{r}F(x), 1\right). \qquad \square$$

Similarly, we may prove

Proposition 5. *Let F be any distribution function. Then, for any integer r and n* ($1 \leqslant r \leqslant n$), *a set of n dependent random variables* $X_1,\dots,X_n$ *can be found, with distribution function F, for which the bound* (7) *is reached.*

Note that Propositions 4 and 5 do not apply when the random variables are not identically distributed. In this case $X_{r:n} = F^{-1}(U_{r:n})$ cannot be written. Moreover, we may provide examples in which the bounds may not be fulfilled, whatever the dependency structure. For instance, let X_j be constant and equal to j. The dependency structure may not vary. Then $X_{r:n} = r$ and $F_{X_{r:n}}(x) = 0$ when $x < r$ (1 otherwise). It may be easily verified that this distribution function is between the bounds of Proposition 3 but reaches neither one nor the other.

References

Arnold, B.C. (1988), Bounds on the expected maximum, *Comm. Statist. Theory and Methods* **17**(7), 2135–2150.

David, H.A. (1970), *Order Statistics* (Wiley, New York).

David, H.A. (1988), General bounds and inequalities in order statistics, *Comm. Statist. Theory Methods* **17**(7), 2119–2134.

Hoover, D.R. (1989), Bounds on expectations of order statistics for dependent samples, *Statist. Probab. Lett.* **8**, 261–265.

Lai, T.L. and H. Robbins (1976), Maximally dependent random variables, *Proc. Nat. Acad. Sci. U.S.A.* **73**(2), 286–288.

Maurer, W. and B.H. Margolin (1976), The multivariate inclusion–exclusion formula and order statistics from dependent variates, *Ann. Statist.* **4**(6), 1190–1199.