The Annals of Statistics
1994, Vol. 22, No. 1, 478–494

DISTRIBUTION OF THE MAXIMUM OF CONCOMITANTS OF SELECTED ORDER STATISTICS[1]

By H. N. Nagaraja and H. A. David

Ohio State University and Iowa State University

For a random sample of size n from an absolutely continuous bivariate population (X, Y), let $X_{i:n}$ denote the ith order statistic of the X sample values. The Y-value associated with $X_{i:n}$ is denoted by $Y_{[i:n]}$ and is called the concomitant of the ith order statistic. For $1 \leq k \leq n$, let $V_{k,n} = \max(Y_{[n-k+1:n]}, \dots, Y_{[n:n]})$. In this paper, we discuss the finite-sample and the asymptotic distributions of $V_{k,n}$. We investigate the limit distribution of $V_{k,n}$ as $n \to \infty$, when k is held fixed and when $k = [np], 0 < p < 1$. In both cases we obtain simple sufficient conditions and determine the associated norming constants. We apply our results to some interesting situations, including the bivariate normal population and the simple linear regression model.

1. Introduction. Let (X_i, Y_i), $i = 1, \dots, n$, be a random sample from an absolutely continuous bivariate population (X, Y) with c.d.f. $F(x, y)$. Let $X_{i:n}$ and $Y_{i:n}$ denote the rth order statistics of the X and Y sample values, respectively. Suppose the pairs are ordered by the X_i, and let the Y-value associated with $X_{r:n}$ be denoted by $Y_{[r:n]}$. We call $Y_{[r:n]}$ the concomitant of the rth order statistic. A substantial body of literature exists on concomitants of order statistics, which are also called induced order statistics by some authors. A convenient window into the literature is provided by the review articles of Bhattacharya (1984) and David (1993).

The most important use of concomitants arises in selection procedures when $k (< n)$ individuals are chosen on the basis of their X-values. Then the corresponding Y-values represent performance on an associated characteristic. For example, if the top k out of n bulls, as judged by their genetic makeup, are selected for breeding, then $Y_{[n-k+1:n]}, \dots, Y_{[n:n]}$ might represent the average milk yields of their female offspring; alternatively, X might be the score on a screening test and Y the score on a later test. Suppose only the top k performers in the screening test are selected for further training and a second test. Then $V_{k,n} = \max(Y_{[n-k+1:n]}, \dots, Y_{[n:n]})$ represents the score of the best performer in the second test. The ratio $E(V_{k,n})/E(Y_{n:n})$, which clearly increases to 1 with k, is a measure of effectiveness of the screening procedure. One may wish to choose k to make this ratio sufficiently close to 1. See also Yeo and David (1984) for a different approach to selection.

It is evident that $V_{k,n}$ is a useful statistic. Recently, Feinberg (1991) and

Received May 1992; revised March 1993.

[1]This research was partially supported by U.S. Army Research Office Contract DAAL03-89K-0010.

AMS 1991 *subject classifications*. Primary 62E20, 62G30; secondary 62E15.

Key words and phrases. Maximum, concomitants of order statistics, weak convergence, tail equivalence, bivariate normal distribution, simple linear regression model.

Feinberg and Huber (1993) have investigated some properties of $V_{k,n}$ in a study of cutoff rules under imperfect information. Feinberg (1991) used simulation to examine the behavior of $E(V_{k,n})$ for selected values of n assuming the sample is drawn from a bivariate normal distribution. Motivated by his work, we investigate the finite-sample as well as the asymptotic properties of $V_{k,n}$ for an arbitrary absolutely continuous bivariate c.d.f. F.

In Section 2, using symmetry arguments, we obtain a very useful expression for the c.d.f. of $V_{k,n}$. Next, we investigate the limit distribution of $V_{k,n}$ as $n \to \infty$, when k is held fixed (extreme case) and when $k = [np], 0 < p < 1$ (quantile case). The former situation is considered in Section 3 and the latter is handled in Section 4. In both cases we obtain simple sufficient conditions and determine norming constants which ensure the convergence of $V_{k,n}$ to a nondegenerate random variable V. We look into a simple linear regression model in Section 5 and discuss possible distributions for V in the extreme and the quantile cases. Our results are illustrated with the help of three examples involving uniform, normal and exponential distributions. In view of its practical importance, the situation where (X, Y) is a bivariate normal population is discussed thoroughly. For this population, we compare, in the last section, the values of $E(V_{k,n})$ simulated by Feinberg (1992) with those obtained using the distribution of V.

1.1. *Some notation and conventions.* Let F_1 and F_2 be the c.d.f.'s, and, for $0 \le p \le 1$, let $\xi_1(p)$ and $\xi_2(p)$ be the p-th quantiles of the marginal distributions of X and Y, respectively. In the quantile case, x' stands for $\xi_1(q)$, where $q = 1 - p$. The conditional c.d.f.'s $F_2(y\,|\,x) = P(Y \le y\,|\,X = x)$ and $F_2^*(y\,|\,x) = P(Y \le y\,|\,X > x)$ play a major role in our discussion. The c.d.f. of $V_{k,n}$ is denoted by $F_{k,n}$. In general, the c.d.f. and p.d.f. of a random variable R are denoted by F_R and f_R, unless special notation is used for simplicity. Further, $R_{k:k}$ represents the maximum of a random sample of size k from the distribution of R. The symbols $=_d$, $\to_d$ and $\to_P$ represent equality in distribution, convergence in distribution and convergence in probability, respectively. A normal distribution with mean μ and variance σ^2 is denoted by $N(\mu, \sigma^2)$. A $N(0,1)$ random variable and its p.d.f. and c.d.f., are represented by Z, ϕ and Φ, respectively. An exponential random variable with mean θ is identified as $\text{Exp}(\theta)$, and T represents an Exp(1) random variable.

2. Finite-sample c.d.f. of $V_{k,n}$.

Consider

$$\begin{aligned}
&P(V_{k,n} \le y)\\
&\quad = P\big(Y_{[n-k+1:n]} \le y, \dots, Y_{[n:n]} \le y\big)\\
&\quad = \int_{x_0 < x_1 < \cdots < x_k} P\big(Y_{[n-k+1:n]} \le y, \dots, Y_{[n:n]} \le y | X_{n-k:n} = x_0, \dots, X_{n:n} = x_k\big)\\
&\qquad \times f_{X_{n-k:n}}(x_0) k! \left\{ \prod_{i=1}^{k} \frac{f_1(x_i)}{1 - F_1(x_0)}\, dx_i \right\} dx_0.
\end{aligned}$$

Note that, conditioned on the values of the order statistics, the concomitants are independent. Hence, we can write

$$F_{k,n}(y) = \int_{x_0} \left\{ k! \int_{x_0 < x_1 < \cdots < x_k} \prod_{i=1}^{k} P(Y_{[n-i+1:n]} \le y \,|\, X_{n-i+1:n} = x_i) \times \frac{f_1(x_i)}{1 - F_1(x_0)}\, dx_i \right\} f_{X_{n-k:n}}(x_0)\, dx_0.$$

The integrand of the multiple integral in brackets above is a symmetric function of the variables $x_1, \ldots, x_k$. Thus, it can be expressed as

$$\left\{ \int_{x_0 < x} F_2(y \,|\, x) \frac{f_1(x)}{1 - F_1(x_1)}\, dx \right\}^k = \left\{ P(Y \le y \,|\, X > x_0) \right\}^k = F_2^*(y \,|\, x_0)^k.$$

This simplification leads to the following compact representation for the c.d.f. of $V_{k,n}$:

$$F_{k,n}(y) = \int_{x=-\infty}^{\infty} \{F_2^*(y \,|\, x)\}^k f_{X_{n-k:n}}(x)\, dx. \tag{2.1}$$

3. Asymptotic distribution of $V_{k,n}$ in the extreme case. Now let us suppose that k is held fixed while $n \to \infty$. Let us assume that there exist constants $a_n, b_n > 0$, such that $(X_{n:n} - a_n)/b_n \to_d W_1$, where W_1 is a nondegenerate random variable with c.d.f. G. If this holds, we say that F_1 is in the domain of attraction of G and we write $F_1 \in D(G)$. It is well known that G can be one of the three extreme value c.d.f.'s and following, for example, Resnick (1987), we will denote them by Φ_α, Ψ_α and Λ. Von Mises (1936) gave sufficient conditions for $F_1 \in D(G)$ for each of the three cases. A convenient reference for these conditions is Resnick [(1987), Propositions 1.15–1.17]. Further, the von Mises conditions are necessary and sufficient for the p.d.f. of $(X_{n:n} - a_n)/b_n$ to converge to g, the p.d.f. of G. It is also known that if $(X_{n:n} - a_n)/b_n \to_d W_1$, then $(X_{n-k+1:n} - a_n)/b_n \to_d W_k$. The random variable W_k behaves like the kth lower record value from the c.d.f. G, if we count the first value in the sequence from G as the first (lower) record value. Combining all these results, we can conclude the following.

LEMMA 1. *If the von Mises conditions are satisfied, there exist constants $a_n, b_n > 0$, such that the p.d.f. of $(X_{n-i+1:n} - a_n)/b_n$ converges to g_i, the p.d.f. of W_i, the ith lower record value from the c.d.f. G, for any fixed i. Further, the joint p.d.f. of $W_1, \ldots, W_{k+1}$ and the marginal p.d.f. of W_{k+1} are given, respectively, by*

$$g_{(k+1)}(w_1, \ldots, w_{k+1}) = g(w_{k+1}) \prod_{r=1}^{k} \left[\frac{g(w_r)}{G(w_r)} \right], \qquad w_1 > w_2 > \cdots > w_{k+1},$$

and

$$g_{k+1}(w) = \frac{(-\log G(w))^k}{k!} g(w),$$

where $G = \Phi_\alpha, \Psi_\alpha$ *or* Λ.

Now we are ready to obtain an expression for the limiting c.d.f. of appropriately normalized $V_{k,n}$.

RESULT 1. *Suppose the conditions of Lemma* 1 *hold and assume that there exist constants* A_n *and* $B_n > 0$ *such that*

$$F_2^*(A_n + B_n y | a_n + b_n x) \to H(y \mid x) \tag{3.1}$$

as $n \to \infty$, *for all* x *and* y. *Then, as* $n \to \infty$,

$$F_{k,n}(A_n + B_n y) \to \int_{-\infty}^{\infty} \{H(y \mid x)\}^k g_{k+1}(x)\,dx. \tag{3.2}$$

PROOF. From (2.1) we have

$$\begin{aligned} F_{k,n}(A_n + B_n y) &= \int_{-\infty}^{\infty} \left\{F_2^*(A_n + B_n y \mid x)\right\}^k f_{X_{n-k:n}}(x)\,dx \\ &= \int_{-\infty}^{\infty} \left\{F_2^*(A_n + B_n y | a_n + b_n x)\right\}^k \left\{b_n f_{X_{n-k:n}}(a_n + b_n x)\right\} dx. \end{aligned}$$

The first factor in the above integrand is a nonnegative bounded function converging to $\{H(y \mid x)\}^k$ for all x, by the assumption made in (3.1). Further, the assumed von Mises condition implies that the second factor, a p.d.f., converges to the p.d.f. $g_{k+1}(x)$ as $n \to \infty$. Hence we can appeal to the extended dominated convergence theorem [Rao (1973), page 136)] to conclude that (3.2) holds. □

Let us examine condition (3.1). As $n \to \infty, a_n + b_n x \to \xi_1(1)$, for all x. Now suppose the joint distribution of (X, Y) is such that as $x \to \xi_1(1), F_2(y \mid x) \to H(y)$. Then (3.1) holds with $H(y \mid x) = H(y)$ and $A_n = 0, B_n = 1$. This fact is formalized in the following lemma.

LEMMA 2. *If, for some* y,

$$\lim_{x \to \xi_1(1)} F_2(y \mid x) = H(y), \tag{3.3}$$

then

$$\lim_{x \to \xi_1(1)} F_2^*(y \mid x) = H(y). \tag{3.4}$$

PROOF. If (3.3) holds, then, for every $\varepsilon > 0$, we can find an x_1 such that, for all $u > x_1$,

$$H(y) - \varepsilon < F_2(y|u) < H(y) + \varepsilon$$

and consequently, for all $x > x_1$,

$$(H(y) - \varepsilon) \int_x^\infty f_1(u)\,du < \int_x^\infty F_2(y|u) f_1(u)\,du < (H(y) + \varepsilon) \int_x^\infty f_1(u)\,du.$$

This means

$$H(y) - \varepsilon < F_2^*(y \,|\, x) < H(y) + \varepsilon,$$

since

$$F_2^*(y \,|\, x) = \frac{\int_x^\infty F_2(y|u) f_1(u)\,du}{\int_x^\infty f_1(u)\,du}. \tag{3.5}$$

Therefore (3.4) holds. □

So we can conclude that if (3.3) holds and the appropriate von Mises condition holds for F_1, then, as $n \to \infty, P(V_{k,n} \le y) \to \{H(y)\}^k$. Thus, if H is a c.d.f., $V_{k,n} \to_d V$, where V behaves like the maximum of a random sample of size k from the c.d.f. H.

EXAMPLE 1. Let (X, Y) have the joint p.d.f. given by

$$f(x,y) = 2, \qquad 0 < y < x < 1. \tag{3.6}$$

This is the joint p.d.f. of the order statistics from a random sample of size 2 from the standard uniform distribution. Clearly, given $X = x, Y$ is uniformly distributed over $(0, x)$ and hence (3.3) holds with $H(y)$ as the standard uniform c.d.f. Further, $F_1(x) = x^2, 0 < x < 1$, and hence

$$\frac{(1-x) f_1(x)}{1 - F_1(x)} \to 1,$$

as $x \to 1$. This is the von Mises sufficient condition for $F_1 \in D(\Psi_1)$. Thus, as $n \to \infty, P(V_{k,n} \le y) \to y^k, 0 < y < 1$.

Condition (3.3) does not hold for the bivariate normal distribution. Let us see what happens in that situation.

EXAMPLE 2. Let (X, Y) be bivariate normal with zero means, unit variances and positive correlation coefficient ρ. It is well known that $F_1 = \Phi \in D(\Lambda)$ and satisfies the associated von Mises condition. Further, convenient choices for the norming constants are [see, e.g., David (1981), page 264],

$$a_n = \sqrt{2 \log n} - \frac{1}{2} \frac{\log(4\pi \log n)}{\sqrt{2 \log n}} \quad \text{and} \quad b_n = \frac{1}{\sqrt{2 \log n}}. \tag{3.7}$$

Clearly, $a_n \to \xi_1(1) = \infty$ while $b_n \to 0$ as $n \to \infty$. Now choose $A_n = \rho a_n$ and $B_n = B = \sqrt{1-\rho^2}$. Hence, from (3.5),

$$F_2^*(A_n + By|a_n + b_n x) = \frac{\int_{a_n+b_n x}^{\infty} F_2(A_n + By|u) f_1(u)\,du}{1 - F_1(a_n + b_n x)} \tag{3.8}$$
$$= \frac{\int_x^{\infty} F_2(A_n + By|a_n + b_n w) n b_n f_1(a_n + b_n w)\,dw}{n\{1 - F_1(a_n + b_n x)\}},$$

on setting $u = a_n + b_n w$. Since the conditional distribution of Y given $X = x$ is $N(\rho x, 1-\rho^2)$, $F_2(A_n + By|a_n + b_n w)$ is simply $\Phi(y - \rho B^{-1} b_n w)$. It is nonnegative and bounded by 1 and approaches $\Phi(y)$ as $n \to \infty$. Further, the convergences of the c.d.f. and the p.d.f. of $(X_{n:n} - a_n)/b_n$ together imply that $n b_n f_1(a_n + b_n w) \to e^{-w}$, an integrable function. Hence, on using the extended dominated convergence theorem once again, we conclude that, as $n \to \infty$, the numerator in (3.8) converges to $\int_x^{\infty} \Phi(y) e^{-w}\,dw$. Since $F_1 \in D(\Lambda)$, the denominator in (3.8) approaches $-\log \Lambda(x)$ or e^{-x}. Hence the ratio converges to $\Phi(y)$. Therefore we can conclude from Result 1 that, as $n \to \infty$,

$$\left\{\frac{(V_{k,n} - \rho a_n)}{\sqrt{(1-\rho^2)}}\right\} \to_d Z_{k:k}, \tag{3.9}$$

where the norming constant a_n is given in (3.7). Since the second term there tends to 0 as $n \to \infty$, we conclude that

$$\left(V_{k,n} - \rho\sqrt{2 \log n}\right) \to_d \sqrt{1-\rho^2} Z_{k:k}. \tag{3.10}$$

It may be noted that (3.10) holds even when $\rho \leq 0$.

4. Asymptotic distribution of $V_{k,n}$ in the quantile case.

4.1. *Basic limit theorem.* Let us assume $k = [np]$, $0 < p < 1$, so that the top $100p\%$ of the X-values are selected. We will now examine the limiting distribution of $V_{k,n}$ as $n \to \infty$.

RESULT 2. *Assume that $f_1(x') > 0$, where $x' \equiv \xi_1(q)$, with $q = 1 - p$. For constants A_k and $B_k > 0$ free of x, and for fixed y, define*

$$H_n(y\,|\,x) = \{F_2^*(A_k + B_k y\,|\,x)\}^k. \tag{4.1}$$

Assume that, as $n \to \infty$, for all real u,

$$H_n\left(y\,\middle|\,x' + \frac{u}{\sqrt{n}}\right) \to H(y\,|\,x'). \tag{4.2}$$

Then, $F_{k,n}(A_k + B_k y) \to H(y\,|\,x')$.

PROOF. Under our assumptions, it is well known that

$$Z_n = \left\{ \frac{\sqrt{n} f_1(x')(X_{n-k:n} - x')}{\sqrt{pq}} \right\} \to_d Z.$$

In fact, from Reiss [(1989), page 111] it follows that, for every Borel set E,

$$\Phi_n(E) \to \Phi(E), \tag{4.3}$$

where Φ_n is the measure induced by the distribution of Z_n. Now, from (2.1) and (4.1), we have

$$\begin{aligned} F_{k,n}(A_k + B_k y) &= E\{H_n(y \mid X_{n-k:n})\} \\ &= E\{H_n(y \mid x' + c_0 Z_n/\sqrt{n})\}, \end{aligned}$$

where $c_0 = \sqrt{pq}/f_1(x')$. Note that $h_n(z) = H_n(y \mid x' + c_0 z/\sqrt{n})$ is nonnegative and bounded (by 1). Hence if $h_n(z) \to h(z)$, from (4.3) and Royden[(1968), Proposition 18, page, 232] it follows that $Eh_n(Z_n) \to Eh(Z)$. Now in (4.2) we assumed that $h(z)$ is $H(y \mid x')$, and thus we can conclude that $F_{k,n}(A_k + B_k y) \to H(y \mid x')$. □

Let us now examine the condition given in (4.2). Define $H_n^*(y \mid x) = nP(Y > A_k + B_k y \mid X > x)$ and $H^*(y \mid x) = \{-\log H(y \mid x)\}/p$. Since $H_n(y \mid x) = (1 - n^{-1} H_n^*(y \mid x))^{[np]}$, (4.2) holds iff $H_n^*(y \mid x' + u/\sqrt{n}) \to H^*(y \mid x')$. Since $P(X > x') = p > 0$, this holds iff

$$nP(X > x' + u/\sqrt{n}, Y > A_k + B_k y) \to \{-\log H(y \mid x')\}/p. \tag{4.4}$$

Thus, (4.2) holds iff (4.4) holds for all u, and the latter condition appears to be easier to check.

Now, if $H(y \mid x') > 0$, (4.4) is equivalent to the condition

$$nP(X \in I_n, Y > A_k + B_k y) \to 0, \tag{4.5}$$

where I_n is the open interval with the endpoints x' and $x' + u/\sqrt{n}$. Recently, Joshi and Nagaraja (1992), in their study of joint distribution of maxima of concomitants, have used (4.5) to reach a conclusion similar to the one reached in Result 2.

4.2. *Sufficient conditions.* The assumptions regarding the existence of the norming constants and the validity of either (4.2) or (4.4) are hard to verify directly. Thus we will search for easily verifiable sufficient conditions. A simple one ensuring that (4.2) holds is given by the following lemma, whose proof is omitted.

LEMMA 3. *Suppose that $H_n(y \mid x) \to H(y \mid x)$ for x in I, a neighborhood of x', and that $H(y \mid x)$ is continuous at x', where H_n is given by* (4.1). *Then* (4.2) *holds.*

We now introduce the concept of tail equivalence from extreme value theory. This will help us in our search for the constants A_k and B_k appearing in Result 2, which have to be free of x. Fix x in I, and assume

$$\lim_{y \to \xi_2(1)} \frac{P(Y > y)}{P(Y > y \mid X > x)} = \lim_{y \to \xi_2(1)} \frac{1 - F_2(y)}{1 - F_2^*(y \mid x)} = c(x), \tag{4.6}$$

where $c(x)$ is positive and finite. If (4.6) holds, then $F_2^*(y \mid x)$ and $F_2(y)$ are said to be tail equivalent. From Resnick [(1987), Proposition 1.19, page 67] it follows that if $F_2 \in D(G)$ as $k \to \infty$ such that $\{F_2(A_k + B_k y)\}^k \to G(y)$ and if (4.6) holds, then $H_n(y \mid x) \to H(y \mid x)$, with $H(y \mid x) = G(A(x) + B(x)y)$. Further,

$$\begin{aligned} &A(x) = 0 \quad \text{and} \quad \{B(x)\}^{\alpha} = c(x) \quad \text{if } G = \Phi_\alpha, \\ &A(x) = 0 \quad \text{and} \quad \{B(x)\}^{-\alpha} = c(x) \quad \text{if } G = \Psi_\alpha, \\ &A(x) = \log c(x) \quad \text{and} \quad B(x) = 1 \quad \text{if } G = \Lambda. \end{aligned} \tag{4.7}$$

Since G is a continuous c.d.f., if $c(x)$ is continuous in I, then from (4.7) it is clear that $H(y \mid x)$ is also continuous in I. From Lemma 3 we can then conclude that (4.2) holds. Thus, if $F_2 \in D(G)$ and if, for x in I, (4.6) holds, where $c(x)$ is continuous at x', then all the conditions for the application of Result 2 are satisfied. We formalize this discussion with the following assertion.

RESULT 3. *Suppose $F_2 \in D(G)$; that is, there exist constants A_k and $B_k > 0$ such that $P(Y_{k:k} \le A_k + B_k y) \to G(y)$, for all real y. With $x' = \xi_1(q)$, assume $f_1(x') > 0$ and there exists a neighborhood I of x' in which* (4.6) *holds with $c(x)$ being continuous at x'. Then $P(V_{k,n} \le A_k + B_k y) \to G(A(x') + B(x')y)$ as $k \to \infty$. The functions $A(\cdot)$ and $B(\cdot)$ are dependent on $c(x')$ and the form of G, and are given in* (4.7).

Convenient choices for A_k and B_k are also well known and depend only on the c.d.f. F_2. See, for example, Galambos [(1987), pages 53–54]. Result 3 means essentially that, under certain conditions, $V_{k,n}$ (appropriately normalized) has the same limit distribution as that of the sample maximum of a random sample of size k from the c.d.f. $F_2^*(y \mid x')$. The key conditions are $F_2 \in D(G)$, and $F_2^*(y \mid x)$ and $F_2(y)$ are tail equivalent for all x in I. These conditions can be replaced by $F_2^*(y \mid x') \in D(G)$ and the tail equivalence of $F_2^*(y \mid x')$ and $F_2^*(y \mid x)$. These would be more natural conditions. However, the form of the conditional c.d.f. F_2^* in general is more involved than that of the marginal c.d.f. F_2. One would thus expect the determination of the norming constants for F_2 to be easier than those for $F_2^*(y \mid x')$.

4.3. *Examples.*

EXAMPLE 1 (Continued). When the joint p.d.f. is given by (3.6), $1 - F_2(y) =$

$(1-y)^2$, $0 \le y \le 1$, and hence

$$\frac{(1-y)f_2(y)}{1-F_2(y)} = 2.$$

This implies the von Mises sufficient condition for $F_2 \in D(\Psi_2)$. The norming constants are $A_k = \xi_2(1) = 1$ and $B_k = \xi_2(1) - \xi_2(1-(1/k)) = 1/\sqrt{k}$. The c.d.f. of $\sqrt{k}(Y_{k:k} - 1)$ converges to Ψ_2 as $k \to \infty$. Now, for $0 < x < 1$,

$$F_2^*(y \mid x) = \begin{cases} \dfrac{2y}{1+x}, & 0 \le y \le x < 1, \\ 1 - \dfrac{(1-y)^2}{1-x^2}, & 0 < x < y \le 1. \end{cases}$$

Therefore,

$$\lim_{y \to \xi_2(1)} \frac{1-F_2(y)}{1-F_2^*(y \mid x)} = c(x) = 1 - x^2,$$

which means that $F_2(y)$ and $F_2^*(y \mid x)$ are tail equivalent for all x, and $c(x)$ is continuous in (0,1). Further, clearly $f_1(x')$ is positive. Thus, Result 3 is applicable. Note that $c(x') = p$ and, since $G = \Psi_2$, from (4.7) we have $A(x') = 0$ and $\{B(x')^{-2}\} = p$, or $B(x') = 1/\sqrt{p}$. Thus, we can conclude that

$$P\left(\sqrt{k}(V_{k,n} - 1) \le y\right) \to \Psi_2(y/\sqrt{p}) = \exp(-y^2/p), \qquad y \le 0,$$

as $k \to \infty$. In other words, $\sqrt{k}(1 - V_{k,n})$ converges to a Rayleigh distribution.

EXAMPLE 2 (Continued). When (X, Y) is standard bivariate normal, $F_1(y) = F_2(y) = \Phi(y) \in D(\Lambda)$ and $\xi_2(1) = \infty$. On noting that X, given $Y = y$, is $N(\rho y, 1 - \rho^2)$ and using the form $f(x,y) = f_2(y) f_1(x|y)$, we see that

$$\begin{aligned} P(X > x, Y > y) &= \int_y^\infty f_2(v) \int_x^\infty f_1(u|v)\, du\, dv \\ &= \int_y^\infty \phi(v)\left\{1 - \Phi\left((x - \rho v)/\sqrt{1-\rho^2}\right)\right\} dv. \end{aligned}$$

Hence

$$\frac{1-F_2(y)}{1-F_2^*(y \mid x)} = \frac{[1-\Phi(x)][1-\Phi(y)]}{\int_y^\infty \phi(v)\{1 - \Phi((x-\rho v)/\sqrt{1-\rho^2})\}\, dv}.$$

On using l'Hôpital's rule, it follows that

$$\begin{aligned} \lim_{y \to \infty} \frac{1-F_2(y)}{1-F_2^*(y \mid x)} &= [1-\Phi(x)] \lim_{y \to \infty} \frac{1}{\{1 - \Phi((x-\rho y)/\sqrt{1-\rho^2})\}} \\ &= 1 - \Phi(x), \end{aligned}$$

since $\rho > 0$. Thus (4.6) holds with $c(x) = 1 - \Phi(x)$.

Now it is easy to check that all the conditions assumed in Result 3 are satisfied. Also note that $c(x') = 1 - \Phi(x') = p$. Since $G = \Lambda$, from (4.7) it follows that $A(x') = \log p$ and $B(x') = 1$. Hence, our conclusion is that, as $k \to \infty$,

$$P(V_{k,n} \le A_k + B_k y) \to \Lambda(y + \log p) = \exp(-(e^{-y}/p)), \tag{4.8}$$

for all real y. The norming constants A_k and B_k can be chosen as a_k and b_k, respectively, where a_n and b_n are given by (3.7).

REMARKS. In contrast to (4.8), when $\rho = 0$, or when X and Y are independent, $P(V_{k,n} \le A_k + B_k y) \to \Lambda(y)$ for the same choices of A_k and B_k. Thus, while the limit distribution is free of ρ and depends only on p as long as $\rho > 0$, it is free of p when $\rho = 0$. When $\rho < 0$, F_2 and F_2^* are not tail equivalent and hence Result 3 is not applicable. Recent work by Joshi and Nagaraja (1992) indicates that in that case, while the limiting c.d.f. of $V_{k,n}$ is still Λ, the norming constants are different and depend on both p and ρ.

5. A simple linear regression model. Our example on the bivariate normal population is a special case of the following simple linear regression model discussed extensively in the literature on concomitants [see, e.g., David (1981), pages 110 and 282]:

$$Y_{[i:n]} = \beta X_{i:n} + U_{[i]}, \qquad 1 \le i \le n. \tag{5.1}$$

We assume $U_{[i]}$, not necessarily normal, has c.d.f. F_3, $U_{[i]}$ and $X_{i:n}$ are independent, and take $\beta > 0$. Under these assumptions, we investigate the possible limit distributions for $V_{k,n}$.

An expression for $F_{k,n}$ is given by (2.1) where, on recalling (3.5), we note that

$$F_2^*(y \mid x) = \frac{\int_x^\infty F_3(y - \beta u) f_1(u)\, du}{1 - F_1(x)}.$$

5.1. *Asymptotic distribution in the extreme case.* If $\xi_1(1)$ is finite, $X_{n-i+1:n} \to_P \xi_1(1)$, for $i = 1, \dots, k$, and hence, from the definition of $V_{k,n}$,

$$V_{k,n} \to_d \beta\xi_1(1) + \max(U_{[n]}, \dots, U_{[n-k+1]}) \approx \beta\xi_1(1) + U_{k:k}, \tag{5.2}$$

with $U_{k:k}$ having c.d.f. F_3^k.

Now suppose $\xi_1(1)$ is infinite. If $X_{n:n}$ has an additive weak law [Galambos (1987), page 244] so that $X_{n:n} - c_n \to_P 0$, then $X_{n-i+1:n} - c_n \to_P 0$, $i = 2, \dots, k$. Consequently,

$$V_{k,n} - \beta c_n \to_d U_{k:k}. \tag{5.3}$$

Suppose $F_1 \in D(G)$. If $G = \Psi_\alpha$, then $\xi_1(1)$ is necessarily finite and hence (5.2) holds. If $G = \Lambda$ and $\xi_1(1)$ is finite, then again (5.2) holds. To see what happens in other cases, let us suppose $(X_{n:n} - a_n)/b_n \to_d W_1$ and rewrite (5.1) as

$$\frac{Y_{[i:n]} - \beta a_n}{\beta b_n} = \frac{X_{i:n} - a_n}{b_n} + \frac{U_{[i]}}{\beta b_n}, \qquad n-k+1 \le i \le n. \tag{5.4}$$

When $G = \Phi_\alpha, a_n$ can be taken to be 0 and b_n to be $\xi_1(1-n^{-1})$, which approaches infinity. Therefore in (5.4), $\{U_{[i]}/\beta b_n\} \to_P 0$. Hence we can conclude that

$$\left(\left\{\frac{(Y_{[n:n]} - \beta a_n)}{\beta b_n}\right\}, \dots, \left\{\frac{(Y_{[n-k+1:n]} - \beta a_n)}{\beta b_n}\right\}\right) \to_d (W_1, \dots, W_k),$$

where, as noted in Lemma 1, the W's are the successive lower record values from the c.d.f. G. Thus we have, with $a_n = 0$,

$$\left\{\frac{(V_{k,n} - \beta a_n)}{\beta b_n}\right\} \to_d W_1. \tag{5.5}$$

When $G = \Lambda$, we can choose $a_n = \xi_1(1-n^{-1})$ and $b_n = E(X - a_n | X > a_n)$. Suppose $b_n \to b$. If b is zero, then $X_{n-i+1:n} - a_n \to_P 0$ and hence (5.3) holds with $c_n = a_n$. This is what happens when (X, Y) has a bivariate normal distribution, a conclusion drawn at the end of Section 3. If b is infinite, (5.5) holds. If b is finite and positive, using (5.4) and the independence of the X's and U's, we conclude that

$$\left(\left\{\frac{(Y_{[n:n]} - \beta a_n)}{\beta b_n}\right\}, \dots, \left\{\frac{(Y_{[n-k+1:n]} - \beta a_n)}{\beta b_n}\right\}\right)$$
$$\to_d \left(W_1 + (\beta b)^{-1}U_1, \dots, W_k + (\beta b)^{-1}U_k\right).$$

This means that $\{(V_{k,n} - \beta a_n)/\beta b\} \to_d V$, and

$$V =_d \max\left(W_1 + (\beta b)^{-1}U_1, \dots, W_k + (\beta b)^{-1}U_k\right), \tag{5.6}$$

where the U's are i.i.d. random variables, independent of the W's. Since the W's are dependent, it is not evident how one can find the c.d.f. of V in (5.6) in a closed form. Of course, we can condition on the values of the W's and express F_V in terms of a k-dimensional integral. However, a more convenient form of V is possible in view of the following result.

LEMMA 4. *Let $W_i, i \ge 1$, be the lower record values from the c.d.f. Λ, and let $U_i, i \ge 1$, be i.i.d. random variables independent of the W's. Then, if $c > 0$,*

$$\max(W_1 + cU_1, \dots, W_k + cU_k) =_d W_{k+1} + \max(T_1 + cU_1, \dots, T_k + cU_k), \tag{5.7}$$

where the T's are i.i.d. Exp(1) *random variables independent of all the other random variables.*

PROOF. We prove the lemma by showing that the random variables on both sides of (5.7) have the same c.d.f. When $G(x) = \Lambda(x) = \exp(-\exp(-x))$, the joint p.d.f. of $w_1, \dots, W_{k+1}$ can be expressed by Lemma 1 as

$$g_{(k+1)}(w_1, \dots, w_{k+1}) = g(w_{k+1}) \prod_{r=1}^{k} \exp(-w_r), \quad w_1 > w_2 > \cdots > w_{k+1}.$$

Consider the c.d.f. of V by using the form given on the left-hand side of (5.7). By conditioning on the values of $W_1, \dots, W_{k+1}$, we can write

$$F_V(v) = \int_{w_1 > \cdots > w_{k+1}} \left\{ \prod_{i=1}^{k} F_3(c^{-1}(v - w_i)) \exp(-w_i)\, dw_i \right\} g(w_{k+1})\, dw_{k+1}.$$

As in Section 2, we use the symmetry of the function in brackets and the symmetry of the region of integration. This yields

$$F_V(v) = \int_{w_{k+1}} \left\{ \int_{w > w_{k+1}} F_3(c^{-1}(v - w)) \exp(-w)\, dw \right\}^k \frac{1}{k!} g(w_{k+1})\, dw_{k+1}.$$

With $t = w - w_{k+1}$, we obtain

$$\begin{aligned} F_V(v) &= \int_{w_{k+1}} \left\{ \int_{t=0}^{\infty} F_3(c^{-1}(v - t - w_{k+1})) \exp(-t)\, dt \right\}^k \\ &\qquad \times \frac{\exp(-k w_{k+1})}{k!} g(w_{k+1})\, dw_{k+1} \\ &= \int_{w_{k+1}} \{P(T + cU \le v - w_{k+1})\}^k g_{k+1}(w_{k+1})\, dw_{k+1} \qquad (5.8) \\ &= \int_{w_{k+1}} P\{\max(T_1 + cU_1, \dots, T_k + cU_k) \le v - w_{k+1}\} g_{k+1}(w_{k+1})\, dw_{k+1} \\ &= P(\{W_{k+1} + \max(T_1 + cU_1, \dots, T_k + cU_k)\} \le v). \end{aligned}$$

Thus the lemma is established. □

COROLLARY. *In the model given by* (5.1), *suppose the c.d.f. of* $(X_{n:n} - a_n)/b_n$ *converges to* Λ *and* $b_n \to b$, *finite and positive. Let* $T^* =_d T + (U_1/\beta b)$. *Then* $(V_{k,n} - \beta a_n)/\beta b_n \to_d V$, *where*

$$V =_d W_{k+1} + T^*_{k:k}, \qquad (5.9)$$

and W_{k+1} *and* $T^*_{k:k}$ *are independent.*

REMARKS.

(i) The appearance of the i.i.d. Exp(1) random variables in Lemma 4 is not a surprise. It is known [Weissman (1978)] that when $G = \Lambda$ the random variables $W_1 - W_2, \dots, W_k - W_{k+1}$ are i.i.d. Exp(1) random variables and are independent of W_{k+1}. However, it is not clear how this fact enters into the equality of the distributions appearing in (5.7).

(ii) We can obtain the representation for V as given by (5.9) if we apply Result 1 and the approach used in Example 2 to determine an explicit form for $H(y \mid x)$ [(5.8) is nothing but (3.2) of Result 1]. However, to use that approach one has to assume that the related von Mises condition holds, which is avoided here.

(iii) The representation for V as given by (5.6) involves the maximum of k dependent, stochastically decreasing, random variables. In contrast, the form given in (5.9) is the sum of two independent random variables, where one of them behaves like the sample maximum of a random sample. The latter version is much easier to handle, especially if we are interested in the moments of V. The following example illustrates that fact.

EXAMPLE 3. Let X, Y behave, respectively, like sample minimum and maximum in a random sample of size 2 from an Exp(1) distribution. These can be viewed as the failure times of the first component to fail and of the series system made up of two components with i.i.d. exponential life-length distributions. It is well known that $2X$ and $U = Y - X$ are i.i.d. Exp(1) random variables. Suppose the selection is made on the basis of X-value and interest is about Y, the system failure time. We can then use the model given in (5.1) with $\beta = 1$ to describe their relationship.

We know that $2X_{n:n} - \log n \to_d W_1$, where W_1 has c.d.f. Λ. Thus, from the corollary it follows that $2V_{k,n} - \log n \to_d V =_d W_{k+1} + T^*_{k:k}$. Since $T^* =_d T_1 + 2U_1$, it has the distribution of the sample maximum of a random sample of size 2 from an Exp(2) population. Consequently, $T^*_{k:k} =_d 2T_{2k:2k}$ and $V =_d W_{k+1} + 2T_{2k:2k}$.

The first two moments of W_{k+1} are given by Weissman [(1978), page 813] and those of $T_{2k:2k}$ can be obtained, for example, from David [(1981), page 49]. On using these and the fact that the two random variables are independent, we conclude that $E(V) = \gamma - \sum_{i=1}^{k} i^{-1} + 2\sum_{i=1}^{2k} i^{-1}$, where γ is Euler's constant $(0.5772\dots)$ and $\mathrm{Var}(V) = (\pi^2/6) - \sum_{i=1}^{k} i^{-2} + 4\sum_{i=1}^{2k} i^{-2}$.

5.2. *Quantile case.* When $k = [np]$, where $0 < p < 1$, we do not have any special technique for handling the asymptotic theory for $V_{k,n}$ as was done in the extreme case. Thus, we may use Result 3, which involves the tail equivalence condition given by (4.6). In the bivariate normal example we could use the knowledge of the conditional distribution of X given $Y = y$ to verify that (4.6) holds. Now let us see what happens for the bivariate distribution introduced in Example 3.

EXAMPLE 3 (*Continued*). Since $F_2(y) = (1 - e^{-y})^2, y \geq 0$, it is easily verified that the von Mises condition for $F_2 \in D(\Lambda)$ is satisfied. Since F_2 is the c.d.f. of the maximum of a random sample of size 2 from Exp(1), $Y_{k:k} =_d T_{2k:2k}$. This means the norming constants can be chosen as $A_k = \log(2k)$ and $B_k = 1$.

We now evaluate the $c(x)$ in (4.6). For this consider, for $0 \leq x < y$,

$$P(X > x, Y > y) = \int_y^\infty \int_x^\infty_{u<v} 2e^{-(u+v)}\,du\,dv$$
$$= e^{-y}(2e^{-x} - e^{-y}).$$

Thus, for $x < y$,

$$\frac{1 - F_2(y)}{1 - F_2^*(y \mid x)} = \frac{[1 - F_1(x)][1 - F_2(y)]}{P(X > x, Y > y)}$$
$$= \frac{e^{-2x}e^{-y}(2 - e^{-y})}{e^{-y}(2e^{-x} - e^{-y})},$$

which converges to e^{-x} as $y \to \infty$. This is our $c(x)$.

Now $x' = (-\log p)/2$ and $c(x') = \sqrt{p}$. Since $G = \Lambda$, from (4.7) we have $A(x') = (\log p)/2$ and $B(x') = 1$. Hence, it follows from Result 3 that, as $k \to \infty$, $P(V_{k,n} - \log 2k \leq y) \to \exp(-(e^{-y}/\sqrt{p}))$, for all real y.

REMARKS. In the first two examples, $c(x)$ coincided with $1 - F_1(x)$. This meant $P(X > x \mid Y > y) \to 1$ as $y \to \xi_2(1)$. This is not the case with Example 3. This example also illustrates the possibility that even when $F_2(y \mid x)$ and $F_2^*(y \mid x)$ are tail equivalent to $F_2(y)$ the limiting value of $\{(1 - F_2(y))/(1 - F_2^*(y \mid x))\}$ and that of $\{(1 - F_2(y))/(1 - F_2(y \mid x))\}$ can be different.

6. Numerical comparisons. Feinberg (1989) considered the standard bivariate normal population discussed in Example 2 and simulated $E(V_{k,n})$ for all k, with $n = 25, 50$ and 100, and selected values of ρ^2 ($\rho > 0$). For the sake of illustration, in Table 1 we have chosen $n = 50$ and $\rho^2 = 0.01, 0.05, 0.10, 0.25$, 0.50 and 0.90 and have taken $k = 1, 2, 3, 4, 5, 50$. In that table, E_1 stands for the values of $E(V_{k,n})$ simulated by Feinberg (1992), each entry being based on one million replications. He also used the exact expression for the c.d.f. of $V_{k,n}$ given in (2.1) to evaluate $E(V_{k,n})$ through numerical integration, and he reports strong agreement between those values and the simulated values given by E_1.

In Table 1 we also display two approximations to $E(V_{k,n})$ as suggested by the asymptotic distribution of $V_{k,n}$ in the extreme and quantile cases. They are $E_2 = \rho a_n + \sqrt{1 - \rho^2}\,E(Z_{k:k})$, as suggested by (3.9), and E_3, given by (4.8) as $A_k + B_k E(V) = a_k + b_k\{\gamma - \log(k/n)\}$. The approximation E_3 did not perform well for moderate k values and hence its value is given only for $k = 50$ $(=n)$. Even there, it does not do well. This is not surprising in view of the following facts: (i) The norming constants as well as the limit distribution are free of ρ;

TABLE 1
Simulated values of and asymptotic approximations to $E(V_{k,50})$ for samples from the standard bivariate normal population

k	ρ^2	0.01	0.05	0.10	0.25	0.50	0.90
1	E_1	0.2262	0.5034	0.7119	1.1240	1.5902	2.1333
	E_2	0.2101	0.4698	0.6644	1.0504	1.4856	1.9931
	Exact	0.2249	0.5029	0.7112	1.1245	1.5903	2.1337
2	E_1	0.7672	1.0118	1.1892	1.5247	1.8767	2.2105
	E_2	0.7715	1.0197	1.1996	1.5390	1.8845	2.1715
3	E_1	1.0342	1.2551	1.4145	1.7066	1.9957	2.2323
	E_2	1.0521	1.2946	1.4672	1.7833	2.0840	2.2607
4	E_1	1.2048	1.4097	1.5548	1.8161	2.0626	2.2407
	E_2	1.2343	1.4731	1.6409	1.9419	2.2134	2.3186
5	E_1	1.3283	1.5199	1.6538	1.8909	2.1056	2.2445
	E_2	1.3672	1.6033	1.7676	2.0576	2.3079	2.3608
50	E_1	2.2491	2.2489	2.2493	2.2489	2.2489	2.2488
	E_3	2.3072	2.3072	2.3072	2.3072	2.3072	2.3072
	Exact	2.2491	2.2491	2.2491	2.2491	2.2491	2.2491

(ii) the rate of convergence of the sample maximum to the c.d.f. Λ is known to be rather slow. The rows corresponding to $k = 1$ and 50 also contain the exact values of $E(V_{k,n})$, $\rho E(Z_{50:50})$ and $E(Z_{50:50})$, respectively.

From Table 1 it appears the approximation E_2 performs extremely well for $k = 2$ over the range of values of ρ considered. As k increases the gap between E_1 and E_2 appears to widen. This is expected as we move away from the extreme case.

In Table 2 we carry out a similar comparison for a sample size of 100, where we have displayed E_2 for $k \leq 5$ and E_3 for larger selected values of k. The conclusion is essentially the same. The performance of E_3 is uniformly poor and it fails to be monotonic in k as well. For both sample sizes, the approximation E_2 comes closest to E_1 when $k = 2$, more so for $n = 100$.

Acknowledgments. We are indebted to an Associate Editor for suggesting the use of (4.2) in place of a more stringent condition that we had earlier for the validity of Result 2. We take this opportunity to thank also Professor Fred Feinberg for his queries, which provided motivation for this work, and for sharing his simulation results.

TABLE 2
Simulated values of and asymptotic approximations to $E(V_{k,100})$ for samples from the standard bivariate normal population

k	ρ^2	0.01	0.05	0.10	0.25	0.50	0.90
1	E_1	0.2506	0.5604	0.7905	1.2529	1.7735	2.3793
	E_2	0.2366	0.5291	0.7483	1.1831	1.6732	2.2448
	Exact	0.2508	0.5607	0.7930	1.2538	1.7731	2.3789
2	E_1	0.7940	1.0720	1.2735	1.6611	2.0676	2.4609
	E_2	0.7980	1.0790	1.2835	1.6717	2.0721	2.4232
3	E_1	1.0626	1.3201	1.5037	1.8479	2.1932	2.4854
	E_2	1.0787	1.3540	1.5511	1.9160	2.2716	2.5124
4	E_1	1.2354	1.4774	1.6472	1.9622	2.2648	2.4957
	E_2	1.2608	1.5324	1.7248	2.0746	2.4011	2.5703
5	E_1	1.3601	1.5895	1.7496	2.0412	2.3121	2.5008
	E_2	1.3938	1.6626	1.8516	2.1903	2.4955	2.6126
10	E_1	1.7055	1.8936	2.0217	2.2401	2.4189	2.5070
	E_3	2.7039	2.7039	2.7039	2.7039	2.7039	2.7039
50	E_1	2.3212	2.3922	2.4320	2.4842	2.5056	2.5078
	E_3	2.5550	2.5550	2.5550	2.5550	2.5550	2.5550
100	E_1	2.5075	2.5080	2.5073	2.5074	2.5076	2.5078
	E_3	2.5565	2.5565	2.5565	2.5565	2.5565	2.5565
	Exact	2.5076	2.5076	2.5076	2.5076	2.5076	2.5076

REFERENCES

BHATTACHARYA, P. K. (1984). Induced order statistics: Theory and applications. In *Handbook of Statistics* (P. R. Krishnaiah and P. K. Sen, eds.) 4 383–403. North-Holland, Amsterdam.

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

DAVID, H. A. (1993). Concomitants of order statistics: Review and recent developments. In *Multiple Comparisons, Selection, and Applications in Biometry* (F. M. Hoppe, ed.) 507–518. Dekker, New York.

FEINBERG, F. M. (1989). Personal communication.

FEINBERG, F. M. (1991). Modelling optimal cutoff severity: A comparison of discrete and continuous cutoff rules under imperfect information. Working paper, Fuqua School of Business, Duke Univ.

FEINBERG, F. M. (1992). Personal Communication.

FEINBERG, F. M. and HUBER, J. (1993). A theory of cutoff severity under imperfect information. Working paper, Fuqua School of Business, Duke Univ.

GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Krieger, Malabar, FL.

JOSHI, S. N. and NAGARAJA, H. N. (1992). Joint distribution of maxima of concomitants of selected order statistics. Technical Report 490, Dept. Statistics, Ohio State Univ.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.

REISS, R.-D. (1989). *Approximate Distributions of Order Statistics*. Springer, New York.

RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

ROYDEN, H. L. (1968). *Real Analysis*, 2nd ed. Macmillan, New York.

VON MISES, R. (1936). La distribution de la plus grande de n valuers. *Rev. Math. Union Interbalcanique* **1** 141–160. Reproduced in *Selected Papers of Richard von Mises* **2** 271–294 (1964). Amer. Math. Soc., Providence, RI.

WEISSMAN, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* **73** 812–815.

YEO, W. B. and DAVID, H. A. (1984). Selection through an associated characteristic with applications to the random effects model. *J. Amer. Statist. Assoc.* **79** 399–405.

DEPARTMENT OF STATISTICS
OHIO STATE UNIVERSITY
141 COCKINS HALL
1958 NEIL AVENUE
COLUMBUS, OHIO 43210-1247

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
SNEDECOR HALL
AMES, IOWA 50011-1210