Ordering of Concomitants of Order Statistics, with Applications

Donna Lucas*

1. Introduction and summary.

Let (X_i, Y_i) $(i = 1,2,\ldots,n)$ be n independent random variables from some bivariate distribution. When the X's are arranged in ascending order as

$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n} ,$$

we denote the corresponding Y's by

$$Y_{[1:n]}, Y_{[2:n]}, \ldots, Y_{[n:n]} ,$$

and call these the concomitants of the order statistics.

David, O'Connell, and Yang (1977) investigate the probability distribution of $R_{r,n}$, the rank of $Y_{[r:n]}$ among the n Y's. We apply these results to a problem in the reconstruction of a broken random sample, first presented by DeGroot, Feder, and Goel (1971).

When X and Y are distributed according to Gumbel's bivariate exponential distribution, we show that

$$\pi_{1n} > \pi_{1,n-1} > \cdots > \pi_{11} ,$$

where $\pi_{rs} = P\{R_{r,n} = s\}$. Additionally, for n = 2 and Y stochastically increasing (decreasing) in X, we note that $\pi_{11} > \pi_{12}$ and $\pi_{22} > \pi_{21}$ ($\pi_{12} > \pi_{11}$ and $\pi_{21} > \pi_{22}$).

*This research was supported by the U.S. Army Research Office under Contract DAAG-29-77-C-0035 and the National Science Foundation under Contract MSC78-01434.

2. Application of the distribution of the rank of the concomitants to a matching problem.

Suppose a sample of size n is drawn from some bivariate distribution. However, before the sample values are observed, each pair in the sample is broken into its two components. We observe the X's in some random order and the Y's in some independent random order, thus not knowing the original correspondence of X's and Y's. We consider the problem of matching one particular X, rather than reconstructing the entire sample.

DeGroot, Feder, and Goel (1971) assume that the joint distribution of X and Y can be represented by a probability density function of the form

$$f(x,y) = \alpha(x)\beta(y)e^{xy} \quad \text{for } (x,y) \in R^2, \tag{2.1}$$

where α and β are arbitrary real-valued functions. Denote the ordered observations of the un-paired sample by $x_{1:n} \leq x_{2:n} \leq \ldots \leq x_{n:n}$ and $y_{1:n} \leq y_{2:n} \leq \ldots \leq y_{n:n}$. Suppose one wishes to match $x_{1:n}$. The posterior probability of obtaining a correct match is maximized by pairing $y_{1:n}$ with $x_{1:n}$. Similarly, this criterion leads to pairing $y_{n:n}$ with $x_{n:n}$. A general solution is not pursued.

We suggest the following procedure for matching one observation, not being restricted to bivariate distributions with probability density functions of the form (2.1). Suppose one wishes to match the r^{th} largest X. Then pair with it the k^{th} largest Y, where

$$P\{R_{r,n} = k\} = \max_{1 \leq s \leq n} P\{R_{r,n} = s\}.$$

David, O'Connell, and Yang (1977) derive the following expression for $P\{R_{r,n} = s\}$.

$$P\{R_{r,n} = s\} = n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C_k \Theta_1^k \, \Theta_2^{r-1-k} \, \Theta_3^{s-1-k} \, \Theta_4^{n-r-s+1+k} \, f(x,y)\,dx\,dy, \tag{2.2}$$

where

$$\Theta_1(x,y) = P\{X < x,\ Y < y\},\ \Theta_2(x,y) = P\{X < x,\ Y > y\},$$

$$\Theta_3(x,y) = P\{X > x,\ Y < y\},\ \Theta_4(x,y) = P\{X > x,\ Y > y\},$$

$$t = \min(r-1,\ s-1)\ ,$$

and

$$C_k(r,s,n) = \frac{(n-1)!}{k!(r-1-k)!(s-1-k)!(n-r-s+1+k)!}\ .$$

Numerical results are given for the case in which the joint distribution of X and Y is bivariate normal, $n = 9$, $\rho = 0.1(0.1)0.9,\ 0.95$. It is noted that for small and intermediate values of ρ, holding r constant, π_{rs} is not necessarily maximized by s=r. However, we observe that for each set of calculations,

$$\pi_{11} > \pi_{12} > \cdots > \pi_{19}\ ,$$

and

$$\pi_{99} > \pi_{98} > \cdots > \pi_{91}\ .$$

Since $\pi_{rs}(\rho) = \pi_{r,n+1-s}(-\rho)$ $(r,s = 1,\ldots,n)$, for negative ρ,

$$\pi_{19} > \pi_{18} > \cdots > \pi_{11}\ ,$$

and

$$\pi_{91} > \pi_{92} > \cdots > \pi_{99}\ .$$

3. The distribution of the rank of the first concomitant when sampling from Gumbel's bivariate exponential distribution.

We consider the distribution of the rank of $Y_{[1:n]}$ when the joint distribution of X and Y is Gumbel's bivariate exponential distribution. The marginal distributions of both X and Y are standard exponential, and the joint probability density function is

(3.1) $f(x,y) = e^{-(x+y+\Theta xy)} \{(1+\Theta x)(1+\Theta y) - \Theta\}$ $(x>0,\ y>0,\ 0\le\Theta\le 1)$,

as stated by Johnson and Kotz (1972). When $\Theta = 0$, X and Y are independent, and the correlation decreases as Θ increases. Also, $P\{Y > y | X = x\}$ decreases as x increases, so we say that Y is stochastically decreasing in X (see Barlow and Proschan (1975)).

From (2.2), we have

$$P\{R_{1,n} = s\} = n\binom{n-1}{s-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Theta_3^{s-1}\ \Theta_4^{n-s}\ f(x,y)\,dxdy\ . \tag{3.2}$$

When the joint probability density function of X and Y is (3.1),

$$\Theta_3(x,y) = e^{-x} - e^{-(x+y+\Theta xy)}\ , \tag{3.3}$$

and

$$\Theta_4(x,y) = e^{-(x+y+\Theta xy)}\ .$$

Making the appropriate substitutions,

$$P\{R_{1,n} = s\} = n\binom{n-1}{s-1} \int_0^{\infty} \int_0^{\infty} (e^{-x} - e^{-(x+y+\Theta xy)})^{s-1}\ (e^{-(x+y+\Theta xy)})^{n-s}$$

$$\cdot\ (e^{-(x+y+\Theta xy)}\{(1 + \Theta x)(1 + \Theta y) - \Theta\})dxdy$$

$$= n\binom{n-1}{s-1} \int_0^{\infty} \int_0^{\infty} e^{-nx}(1-e^{-(y+\Theta xy)})^{s-1}\ (e^{-(y+\Theta xy)})^{n-s+1}\ \{(1+\Theta x)(1+\Theta y)-\Theta\}dxdy$$

$$= n\binom{n-1}{s-1} \sum_{k=0}^{s-1} (-1)^k\binom{s-1}{k} \int_0^{\infty} e^{-nx}\ \left(\int_0^{\infty} e^{-(n-s+1+k)(y+\Theta xy)}\ \{(1+\Theta x)(1+\Theta y)-\Theta\}dy\right)dx.$$

Integrating by parts and simplifying, we obtain

$$P\{R_{1,n} = s\} = \frac{1}{n} + e^{\frac{n}{\theta}}\, E_1(\tfrac{n}{\theta})\left(n\binom{n-1}{s-1} \sum_{k=0}^{s-1} (-1)^k\ \binom{s-1}{k} \frac{1}{(n-s+1+k)^2} - 1\right)\ , \tag{3.4}$$

where $E_1(a) = \int_1^{\infty} \frac{e^{-ax}}{x}\, dx$.

We now compare $P\{R_{1,n} = s\}$ and $P\{R_{1,n} = s-1\}$.

$$P\{R_{1,n} = s\} - P\{R_{1,n} = s-1\} = \{\frac{1}{n} + e^{\frac{n}{\theta}} E_1(\tfrac{n}{\theta})(n\binom{n-1}{s-1} \sum_{k=0}^{s-1} (-1)^k \binom{s-1}{k} \frac{1}{(n-s+1+k)^2} - 1)\}$$

$$- \{\frac{1}{n} + e^{\frac{n}{\theta}} E_1(\tfrac{n}{\theta})(n\binom{n-1}{s-2} \sum_{k=0}^{s-2} (-1)^k \binom{s-2}{k} \frac{1}{(n-s+2+k)^2} - 1)\}$$

$$= ne^{\frac{n}{\theta}} E_1(\tfrac{n}{\theta})[\binom{n-1}{s-1} \sum_{k=0}^{s-1} (-1)^k \binom{s-1}{k} \frac{1}{(n-s+1+k)^2}$$

$$- \binom{n-1}{s-2} \sum_{k=0}^{s-2} (-1)^k \binom{s-2}{k} \frac{1}{(n-s+1+k)^2}]$$

$$= ne^{\frac{n}{\theta}} E_1(\tfrac{n}{\theta})[\binom{n-1}{s-1} \sum_{k=0}^{s-1} (-1)^k \binom{s-1}{k} \frac{1}{(n-s+1+k)^2}$$

$$+ \binom{n-1}{s-1} \sum_{k=1}^{s-1} (-1)^k \binom{s-1}{k} \frac{k}{n-s+1} \frac{1}{(n-s+1+k)^2}]$$

$$= \frac{n}{n-s+1} e^{\frac{n}{\theta}} E_1(\tfrac{n}{\theta})[\binom{n-1}{s-1} \sum_{k=0}^{s-1} (-1)^k \binom{s-1}{k} \frac{1}{(n-s+1+k)}].$$

(3.5) $$P\{R_{1,n} = s\} - P\{R_{1,n} = s-1\} = \frac{1}{n-s+1} e^{\frac{n}{\theta}} E_1(\tfrac{n}{\theta}) .$$

Thus,

(3.6) $$\pi_{1n} > \pi_{1,n-1} > \pi_{1,n-2} > \dots > \pi_{12} > \pi_{11} .$$

We note that any monotonic increasing transformations applied separately to X and Y do not change the values of π_{rs}. Due to this fact, (3.6) holds not only for X and Y having a joint probability density function of the form (3.1), but for all other variates having distributions which can be derived by such transformations. We conjecture that (3.6) holds for an even wider class of distributions.

4. General results for n=2.

Consider the case in which n=2, and Y is stochastically increasing in X. Suppose $X_{1:2} = x_{1:2}$ and $X_{2:2} = x_{2:2}$, where $x_{1:2} < x_{2:2}$. Then,

(4.1) $$P\{R_{2,2} = 2|X_{1:2} = x_{1:2}, X_{2:2} = x_{2:2}\}$$
$$= P\{Y_{[2:2]} > Y_{[1:2]}|X_{1:2} = x_{1:2}, X_{2:2} = x_{2:2}\} .$$

Since (X_1,Y_1) and (X_2,Y_2) are independent and identically distributed, the conditional probability density function of $Y_{[r:2]}$ given $X_{r:2} = x_{r:2}$ is $f_{Y_{[r:2]}}(y|X_{r:2} = x_{r:2}) = f_Y(y|X = x_{r:2})$, r = 1,2. It then follows, from the assumption that Y is stochastically increasing in X, that

(4.2) $$P\{Y_{[2:2]} > y|X_{2:2} = x_{2:2}\} = P\{Y > y|X = x_{2:2}\}$$
$$> P\{Y > y|X = x_{1:2}\} = P\{Y_{[1:2]} > y|X_{1:2} = x_{1:2}\} \text{ for all } y.$$

Hence,

$$P\{Y_{[2:2]} > Y_{[1:2]}|X_{1:2} = x_{1:2}, X_{2:2} = x_{2:2}\}$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{u} f_Y(t|X = x_{1:2})f_Y(u|X = x_{2:2})dt\,du$$
$$= \int_{-\infty}^{\infty} F_Y(u|X = x_{1:2})f_Y(u|X = x_{2:2})du$$
$$> \int_{-\infty}^{\infty} F_Y(u|X = x_{2:2})f_Y(u|X = x_{2:2})du = \tfrac{1}{2} ,$$

for any $x_{1:2} < x_{2:2}$. Thus we have shown

(4.3) $$P\{R_{2,2} = 2\} > \tfrac{1}{2} .$$

Since $P\{R_{2,2} = 1\} + P\{R_{2,2} = 2\} = 1$,

(4.4) $$\pi_{22} > \pi_{21} .$$

Due to the relationships $P\{R_{2,2} = 1\} + P\{R_{1,2} = 1\} = 1$ and $P\{R_{1,2} = 1\} + P\{R_{1,2} = 2\} = 1$, we also have

(4.5) $$\pi_{11} > \pi_{12} .$$

It can be shown similarly that for Y stochastically decreasing in X,

(4.6) $$\pi_{21} > \pi_{22} \text{ and } \pi_{12} > \pi_{11} .$$

References

Barlow, R.E., and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart, and Winston, Inc.

David, H.A., O'Connell, M.J., and Yang, S.S. (1977). Distribution and expected value of the rank of a concomitant of an order statistic. *Ann. Statist.* 5, 216-223.

DeGroot, M.H., Feder, P.I., and Goel, P.K. (1971). Matchmaking. *Ann. Math. Statist.* 42, 578-593.

Johnson, N.L., and Kotz, S. (1972). Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Ordering of Concomitants of Order Statistics, with Applications	5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
	6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1204
7. AUTHOR(s) Donna Lucas	8. CONTRACT OR GRANT NUMBER(s) DAAG-29-77-C-0035
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office, Research Triangle Park, NC	12. REPORT DATE January 1979
	13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release: Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Concomitants of Order Statistics, Broken Random Sample

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Let (X_i,Y_i) be n independent rv's from some bivariate distribution. Let $X_{r:n}$ denote the rth ordered X-variate, and $Y_{[r:n]}$ the Y-variate paired with $X_{r:n}$. The distribution of $R_{r,n}$, the rank of $Y_{[r:n]}$, is applied to a matching problem. Also, it is shown that $\pi_{1n} > \pi_{1,n-1} > \cdots > \pi_{11}$ when sampling from Gumbel's bivariate exponential distribution, where $\pi_{rs} = P\{R_{r,n} = s\}$.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)