Journal of the Korean
Statistical Society
Vol. 25, No. 2, 1996

On the Dependence Structure of Concomitants of Order Statistics

Song-Ho Kim [1] and Tae-Sung Kim [2]

Abstract

Let $(X_{1j}, X_{2j}, \ldots, X_{mj}, Y_j)$ $j = 1, 2, \ldots, n$, be a sample of size n on an $(m+1)$-dimensional vector $(X_1, X_2, \ldots, X_m, Y)$, $m \geq 1$. If $Y_{(r)}$ denote the rth order statistic from Y, then the $\underline{X}_{[r:n]}$ paired with $Y_{(r)}$ is termed the concomitant vector of the order statistics. The general distributions of concomitant of order statistics will be found. The mean, variance and covariance of $X_{[r:n]}$ will be studied. Then we will apply the results to the multivariate normal variate case.

Key Words: Concomitant; Order statistics; Covariance; Moment; Multivariate normal distribution; Gamma distribution.

[1] Department of Statistics, Anyang University, Anyang-Shi, 430-714, Korea.
[2] Department of Statistics, Won-Kwang University, Iri-Shi, 570-749, Korea.

1. INTRODUCTION

Let $(X_{1i}, X_{2i}, \ldots, X_{mi}, Y_i)$ $(i = 1, 2, ..., n)$ be a random sample from $(X_1, X_2, \ldots, X_m, Y)$. If we arrange the Y-variates in ascending order as

$$Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)},$$

then the $\underline{X}$-vectors paired with these order statistics are denoted by

$$\underline{X}_{[1:n]}, \underline{X}_{[2:n]}, \ldots, \underline{X}_{[n:n]},$$

where $\underline{X}_{[r:n]} = (X_{1[r:n]}, X_{2[r:n]}, \ldots, X_{m[r:n]})$ and termed the concomitants vectors of the order statistics.

The research of concomitants of order statistics were initiated by David (1973) and Bhattacharya(1974) for $m = 1$ case simultaneously. The extended works were obtained by David and Galambos(1974), and David et al.(1977). Yang (1977) has investigated the general distribution theory, exact moment and applications. And Bhattacharya(1984) summarized the previous results.

When Kim and David(1990) treated the behavior of dependence structure of order statistics, they considered the dependence structure of concomitant of order statistics for $m = 1$. Under the assumption that $X_1, X_2, \ldots, X_m$ are independent normal variates and $Y = \sum_{i=1}^{m} X_i$, Song et al.(1992) obtained the mean, variance and covariance of concomitant variables. Song(1993) has investigated the gamma case of $m > 1$ for the above model, resulting in some positive correlations between concomitant variables.

However, the results of Song et al(1992) and Song(1993) have been devoted to the special cases, normal and gamma. In this paper, we will deal with the general case for any distributions and models. In section 2, we will give the dependence result on $m = 1$ case. In section 3, for $m > 1$, the general formulas of mean, variance, and covariance of concomitants of order statistics will be investigated. The multivariate normal case will be treated. In secton 3, it is shown that $Cov(X_{1i[r:n]}, X_{2j[s:n]})$ can be positive for all $r, s = 1, \ldots, n$, if (X_1, X_2) are bivariate normal variates with $\rho = 0.96$. We study some aspects of the dependence structure of the concomitants of order statistics under normal assumption. Three Tables showing these dependence structure are provided.

2. Bivariate Cases

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a radom sample of n pairs from a common bivariate cdf, $F(x,y)$ and pdf, $f(x,y)$.

Theorem 1. [Kim and David(1990)] Let Y_i and Z_i $(i = 1, \ldots, n)$ be mutually independent rv's. If $X_i = g(Y_i) + Z_i$, a regression model of Y on X. When g is monotone function of Y_i, then $X_{[1:n]}, X_{[2:n]}, \ldots, X_{[n:n]}$ are associated.

Since any subet of associated random variables are associated, $(X_{[r:n]}, X_{[s:n]})$ are associated. It implies $Cov(X_{[r:n]}, X_{[s:n]}) >= 0$ for any $1 \leq r \leq s \leq n$. Yang (1977) showed

$$Cov(X_{[r:n]}, X_{[s:n]}) = Cov(E(X_1 \mid Y_1 = Y_{(r)}), E(X_1 \mid Y_1 = Y_{(s)})).$$

Example 1. Let (X,Y) be distributed by $N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$. Then $X_i = \rho\frac{\sigma_x}{\sigma_y}(Y - \mu_y) + Z_i$, where Z_i is distributed by $N(\mu_x, (1-\rho^2)\sigma_x^2)$. Hence $X_{[1:n]}, X_{[2:n]}, \ldots, X_{[n:n]}$ are associated. We have,

$$Cov(X_{[r:n]}, X_{[s:n]}) = \rho^2\sigma_x^2 Cov(Z_{(r)}, Z_{(s)}) \tag{2.1}$$

where $Z_{(r)}$ and $Z_{(s)}$ are order statistics from standard normal distribution.

Kim and David(1990) showed that if $f(x)$ is a Polya frequence funtion of order 2 (PF_2, i.e. if $f(x)$ is nonnegative and $\log f(x)$ is concave on $(-\infty, +\infty)$.), then the covariance of any two order statistics is less than the variance of either, and the covariance of $Y_{(r)}$ and $Y_{(s)}$ is monotone in r and s separately, decreasing as r and s separate from one another. Since normal is a PF_2, the variances and covariances of concomitants of order statistics from bivariate normal variate have the same property by (2.1).

Example 2. If (X,Y) is distributed by $BVE(\lambda_1, \lambda_2, \lambda_{12})$ (i.e. see Barlow and Proshan(1975)), it can be shown that $X_{[1:n]}, X_{[2:n]}, \ldots, X_{[n:n]}$ are associated.

When $X_i = g(Y_i, Z_i)$, representing a general regression model of X on Y, define the concomitant of the rth order statistics with

$$X_{[r:n]} = g(Y_{(r)}, Z_{(r)}), \quad r = 1, .., n.$$

For the above model, Kim and David(1990) also treated the dependence structure of concomitants of order statistics.

3. Multivariate Cases

More generally, we can consider the multivariate concomitant vectors of order statistics. Let $(X_{1i}, X_{2i}, \ldots, X_{mi}, Y_i)$ $(i = 1, 2, \ldots, n)$ be a random sample from common $(m+1)$ variate pdf $f(\underline{x}, y)$. Then the jpdf of $\underline{X}_{[r:n]}, Y_{(r)}$ is

$$f_{\underline{X}[r:n], Y_{(r)}}(\underline{x}, y) = f(\underline{x} \mid y) f_{(r)}(y),$$

where $f_{(r)}(y)$ is the pdf of $Y_{(r)}$. Hence,

$$f_{\underline{X}_{[r:n]}}(\underline{x}) = \int_{-\infty}^{\infty} f(\underline{x} \mid y) f_{(r)}(y) dy \tag{3.1}$$

More generally,

$$\begin{aligned} & f_{\underline{X}[r_1:n], \underline{X}[r_2:n], \ldots, \underline{X}[r_k:n]}(\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k) \\ = & \int_{-\infty}^{\infty} \int_{-\infty}^{y_{r_k}} \cdots \int_{-\infty}^{y_{r_2}} \prod_{i=1}^{k} f(\underline{x}_i \mid y_i) f_{r_1, r_2, \ldots, r_k}(y_{r_1}, y_{r_2}, \ldots, y_{r_k}) dy_{r_1} dy_{r_2} dy_{r_k} \end{aligned} \tag{3.2}$$

where $f_{r_1, r_2, \ldots, r_k}(y_{r_1}, y_{r_2}, \ldots, y_{r_k})$ is the joint pdf of the k ordered Y-variates $Y_{(r_1)}$, $Y_{(r_2)}$, …, $Y_{(r_k)}$ with $1 \leq r_1 < r_2 < \ldots < r_k \leq n$. By (3.1) and (3.2) , we can easily prove the following:

$$\begin{aligned} E(X_{i[r:n]}) &= E(E(X_{i1} \mid Y_1 = Y_{(r)})), \\ & i = 1, 2, \ldots, m \text{ and } r = 1, 2, \ldots, n. \end{aligned} \tag{3.3}$$

$$\begin{aligned} Var(X_{i[r:n]}) &= E(Var(X_{i1} \mid Y_1 = Y_{(r)})) + Var(E(X_{i1} \mid Y_1 = Y_{(r)})), \\ & i = 1, 2, \ldots, m \text{ and } r = 1, 2, \ldots, n. \end{aligned} \tag{3.4}$$

$$\begin{aligned} Cov(X_{i[r:n]}, X_{i[s:n]}) &= Cov(E(X_{i1} \mid Y_1 = Y_{(r)}), E(X_{i1} \mid Y_1 = Y_{(s)})), \\ & i = 1, 2, \ldots, m \text{ and } r \neq s,\ r, s = 1, 2, \ldots, n. \end{aligned} \tag{3.5}$$

$$\begin{aligned} Cov(X_{i[r:n]}, X_{j[s:n]}) &= E(Cov(X_{i1} \mid Y_1 = Y_{(r)}, X_{j1} \mid Y_1 = Y_{(s)})) \\ &+ Cov(E(X_{i1} \mid Y_1 = Y_{(r)}), E(X_{j1} \mid Y_1 = Y_{(s)})) \\ & i \neq j, i, j = 1, 2, \ldots, m \text{ and } r, s = 1, 2, \ldots, n. \end{aligned} \tag{3.6}$$

(3.3), (3.4) and (3.5) are obtained by Yang(1977). By above facts, we can extend the results of Song et al.(1992) to multivariate normal variate easily. And it can be shown that Song(1993)'s treatment on the sum of Gamma variates is only a special case.

Song et al.(1992) mentioned that (3.6)'s result applied to estimate the exceedence probability of extreme lake levels in hydrology.

Theorem 2. Suppose $\underline{X}$ be distributed by $N_m(\underline{\mu}, \Sigma)$ (i.e. multivariate normal distribution with mean vector, $\underline{\mu}$ and variance-covariance matrix, Σ). Let $Y = \sum_{i=1}^{m} a_i X_i$, a linear combination of $X_1, X_2, \ldots, X_m$ with $| a_i |> 0$, for at least one of $i = 1, 2, \ldots, m$. Then,

$$E(X_{i[r:n]}) = \mu_i + \frac{(a_i\sigma_i^2 + \sum_{l\neq i} a_l\sigma_{li})}{\sqrt{\sum_{l=1}^{m} a_l^2\sigma_l^2 + 2\sum_{l=1}^{m}\sum_{l<k}^{m} a_l a_k \sigma_{lk}}} E(Z_{(r)}) \tag{3.7}$$

$$i = 1, 2, \ldots, m \text{ and } r = 1, 2, \ldots, n;$$

$$V(X_{i[r:n]}) = \sigma_i^2 + \frac{(a_i\sigma_i^2 + \sum_{l\neq i} a_l\sigma_{li})^2}{\sum_{l=1}^{m} a_l^2\sigma_l^2 + 2\sum_{l=1}^{m}\sum_{l<k}^{m} a_l a_k \sigma_{lk}} (Var(Z_{(r)} - 1)) \tag{3.8}$$

$$i = 1, 2, \ldots, m \text{ and } r = 1, 2, \ldots, n;$$

$$Cov(X_{i[r:n]}, X_{i[s:n]}) = \frac{(a_i\sigma_i^2 + \sum_{l\neq i} a_l\sigma_{li})^2}{\sum_{l=1}^{m} a_l^2\sigma_l^2 + 2\sum_{l=1}^{m}\sum_{l<k}^{m} a_l a_k \sigma_{lk}} (Cov(Z_{(r)}, Z_{(s)}) \tag{3.9}$$

$$i = 1, 2, \ldots, m \text{ and } r \neq s,\ r, s = 1, 2, \ldots, n;$$

$$Cov(X_{i[r:n]}, X_{j[r:n]}) = \sigma_{ij} + \frac{(a_i\sigma_i^2 + \sum_{l\neq i} a_l\sigma_{li})(a_j\sigma_j^2 + \sum_{l\neq j} a_l\sigma_{lj})}{\sum_{l=1}^{m} a_l^2\sigma_l^2 + 2\sum_{l=1}^{m}\sum_{l<k}^{m} a_l a_k \sigma_{lk}} (Var(Z_{(r)} - 1) \tag{3.10}$$

$$i = 1, 2, \ldots, m \text{ and } r = 1, 2, \ldots, n;$$

$$Cov(X_{i[r:n]}, X_{j[s:n]}) = \sigma_{ij} + \frac{(a_i\sigma_i^2 + \sum_{l\neq i} a_l\sigma_{li})(a_j\sigma_j^2 + \sum_{l\neq j} a_l\sigma_{lj})}{\sum_{l=1}^{m} a_l^2\sigma_l^2 + 2\sum_{l=1}^{m}\sum_{l<k}^{m} a_l a_k \sigma_{lk}} (Cov(Z_{(r)}, Z_{(s)}) - 1) \tag{3.11}$$

$$i \neq j, \quad i, j = 1, 2, \ldots, m \text{ and } r \neq s,\ r, s = 1, 2, \ldots, n,$$

where $Z_{(r)}$ and $Z_{(s)}$ are the rth and sth order statistics of a sample of size n from stanadard normal variate.

Proof. Since $(\underline{X}', Y)$ is a $(m+1)$-dimensional multivariate normal variate, it can be shown easily by above (3.3),(3.4),(3.5) and (3.6).

When $X_1, X_2, \ldots, X_m$ are independent, Song et al.(1992) obtained (3.7), (3.8) and (3.10), without discussing about the distributions of concomitants. They presented table of the values of correlations of $X_{i[r:n]}$ and $X_{j[r:n]}$ for sample size up to 12. From the table, the absolute value of the correlation of $X_{i[r:n]}$ and $X_{j[r:n]}$ is increasing for $r < \frac{n}{2}$. It may be from the ordering of $Var(Z_{(r)})$.

Corollay 1. Let $X_1, X_2, \ldots, X_m$ are independent normal variates with each μ_i and σ_i^2. Let $Y = \sum_{i=1}^{m} a_i X_i$, where $| a_i |> 0$. If a_i and a_j has same sign, then the covariances of $X_{i[r:n]}$ and $X_{j[s:n]}$ are negative. And the covariance of $X_{i[r:n]}$ and $X_{j[s:n]}$ is monotone in r and s separately, decreasing r and s separate one another.

The proof uses Remark 2 and Theorem 2.

Table 1. $Cov(X_{1[r:n]}, X_{2[s:n]})$ for $m=2$ independent $N(0,1)$ samples of size $n=12$

r \ s	1	2	3	4	5	6
1	-0.33818					
2	-0.41988	-0.40137				
3	-0.44553	-0.43255	-0.42101			
4	-0.45847	-0.44840	-0.43940	-0.43009		
5	-0.46646	-0.45825	-0.45087	-0.44322	-0.43469	
6	-0.47200	-0.46511	-0.45889	-0.45242	-0.44519	-0.43688
7	-0.47617	-0.47207	-0.46494	-0.45938	-0.45315	-0.44580
8	-0.47949	-0.47439	-0.46978	-0.46496	-0.45955	
9	-0.48228	-0.47786	-0.47386	-0.46967		
10	-0.48475	-0.48094	-0.47748			
11	-0.48710	-0.48387				
12	-0.48969					

Numerical illustration is provided by the simple model, $Y = X_1 + X_2$ when X_1 and X_2 are independent standard normal variates(Table 1). Variances and covariances that are not included in the table can be obtained using the identities,

$$Var(Z_{(r)}) = Var(Z_{(n-r+1)})$$

and

$$Cov(Z_{(r)}, Z_{(s)}) = Cov(Z_{(n-r+1)}, Z_{(n-s+1)}).$$

Song et al.(1992) showed that any covariance of $X_{i[r:n]}$ and $X_{j[r:n]}$ from the model $Y = \sum_{i=1}^{m} X_i$ is negative. And Song(1993) obtained positive dependence case from some gamma random variables. However, if we consider Theorem 2, then we can have a simple example for the positive covariance between two concomitant variables of order statistics. Let X_1 and X_2 be distributed by standard bivariate normal variate with correlation . Then for the model, $Y = X_1 + X_2$,

$$Cov(X_{1[r:n]}, X_{2[r:n]}) = \rho + \frac{(1+\rho)}{2}(Var(Z_{(r)} - 1).$$

Since $Var(Z_{(r)}) < 1$, ρ should be positive and satisfied $Var(Z_{(r)}) > \frac{(1-\rho)}{\rho}$ for the positive covariance of two concomitant random variables of order statistics. Numerical illustrations of this fact are given in Table 2.

Table 2. $Cov(X_{1[r:n]}, X_{2[r:n]})$ for $m = 2$ of size upto 12 from $N_2(0, 0, 1, 1, \rho = 0.96)$.

n \ r	1	2	3	4	5	6
1	0					
2	0.64806					
3	0.52828	0.41970				
4	0.46188	0.33325				
5	0.41858	0.28529	0.26110			
6	0.38761	0.25399	0.22129			
7	0.36408	0.23160	0.19533	0.18624		
8	0.34544	0.21462	0.17675	0.16344		
9	0.33021	0.20118	0.16265	0.14715	0.14278	
10	0.31746	0.19023	0.15150	0.13478	0.12803	
11	0.30658	0.18109	0.14241	0.12500	0.11685	0.11442
12	0.29716	0.17332	0.13482	0.11701	0.10800	0.10411

Futhermore, as indicated by Table 1, the variances and covariances of $X_{1[r:n]}$ and $X_{2[s:n]}$ are negative. The question therefore arises whether for $r \neq s$, $Cov(X_{1[r:n]}, X_{2[s:n]})$ can ever be positive. The answer is easily seen to be yes when X_1 and X_2 are distributed by standard bivariate normal distribution with $\rho = 0.96$. Table 3 presents this fact, which all variances and covariances of $X_{1[r:n]}$ and $X_{2[s:n]}$ are positive value.

Table3. $Cov(X_{1[r:n]}, X_{2[s:n]})$ for $m = 2$ samples of size $n = 12$ from $N_2(0, 0, 1, 1, \rho = 0.96)$

r \ s	1	2	3	4	5	6
1	0.29716					
2	0.13703	0.17332				
3	0.08675	0.11220	0.13482			
4	0.06141	0.08113	0.09878	0.11701		
5	0.04575	0.06183	0.07630	0.09130	0.10800	
6	0.03487	0.04839	0.06058	0.07326	0.08743	0.10411
7	0.02671	0.03827	0.04872	0.05962	0.07182	0.08623
8	0.02020	0.03019	0.03923	0.04868	0.05928	
9	0.01474	0.02339	0.03124	0.03945		
10	0.00528	0.01736	0.02414			
11	0.00528	0.01161				
12	0.00021					

REFERENCES

(1) Barlow, R. E. and Proshan F. (1981). *Statistical Theory of Reliability and Life Testing : Probability Models.* To Begin With Silver Spring, MD.

(2) Bhattacharya, P. K.(1974), Convergence of sample paths of normalized sums of induced order statistics, *Annals of Statistics*, Vol. **2**, pp 1034-1039

(3) Bhattacharya, P. K.(1984), Induced Order Statistics: Theory and Applications, *Handbook of Statistics(P. R. Krishnaiah and P. K. Sen edited)*, Vol. **4**, pp 383-403.

(4) David, H. A.(1973), Concomitant of order statistics, *Bulletin of International Statistics Institute*, Vol. **4**, pp 295-300.

(5) David, H. A.(1981) , *Order Statistics*, 2dn, John Wiley and Sons, New York.

(6) David, H. A. and Glambos J.(1974), The asymptotic theory of concomitants of order statistics, *Journal of Applied Statistics*, Vol. **11**, pp 762-770

(7) David, H. A., O'Connell M.J. and Yang S. S.(1977), Distribution and expected value of the ranks of a concomitant of an order statistics, *Annals of Mathematical Statistics*, Vol. **5**, pp 216-223.

(8) Kim, S. H. and David, H. A.(1990), On the dependence structure of order statistics and concomitants of order statistics , *Journal of Statistical Planning and Inference*, Vol. **24**, pp. 363-368.

(9) Song, R.(1993), Moments of variables summing to gamma order statistics , *Communications in Statistics B*, Vol. **22**, pp. 797-803.

(10) Song, R. , Buchberger, S. G. and Deddens, J. A.(1992) , Moment of variables summing to normal order statistics , *Statistics and Probability Letters*, Vol. **15**, pp. 203-208.

(11) Yang, S. S.(1977) , General distribution theory of the concomitants of order statistics , *The Annals of Statistics*, Vol. **5**, pp 996-1002.