Journal of Statistical Planning and Inference 140 (2010) 3076–3087

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Distribution of order statistics from selected subsets of concomitants

Ke Wang [a,*], H.N. Nagaraja [b]

[a] Department of Statistics, Fudan University, Shanghai 200433, China
[b] Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 5 November 2009
Received in revised form
6 April 2010
Accepted 12 April 2010
Available online 20 April 2010

Keywords:
Concomitants of order statistics
Limiting distribution
Sample quantiles
Censored bivariate samples
Bivariate normal distribution
Farlie–Gumbel bivariate exponential distribution

ABSTRACT

For a random sample of size n from an absolutely continuous bivariate population (X, Y), let $X_{i:n}$ be the i th X-order statistic and $Y_{[i:n]}$ be its concomitant. We study the joint distribution of $(V_{s:m}, W_{t:n-m})$, where $V_{s:m}$ is the s th order statistic of the *upper* subset $\{Y_{[i:n]}, i=n-m+1,\ldots,n\}$, and $W_{t:n-m}$ is the t th order statistic of the *lower* subset $\{Y_{[j:n]}, j=1,\ldots,n-m\}$ of concomitants. When $m=\lceil np_0\rceil$, $s=\lceil mp_1\rceil$, and $t=\lceil (n-m)p_2\rceil$, $0<p_i<1, i=0,1,2$, and $n\to\infty$, we show that the joint distribution is asymptotically bivariate normal and establish the rate of convergence. We propose second order approximations to the joint and marginal distributions with significantly better performance for the bivariate normal and Farlie–Gumbel bivariate exponential parents, even for moderate sample sizes. We discuss implications of our findings to data-snooping and selection problems.

1. Introduction

Suppose (X_i, Y_i), $i=1,\ldots,n$, is a random sample from an absolutely continuous bivariate population (X,Y). If we order the sample by the X-variate, and obtain the order statistics, $X_{1:n}<X_{2:n}<\cdots<X_{n:n}$, for the X sample, then the Y-variate associated with the r th order statistic $X_{r:n}$ is called the *concomitant of the rth order statistic*, and is denoted by $Y_{[r:n]}$. The term *concomitant of order statistic* was introduced by David (1973) and independently Bhattacharya (1974) used the term *induced order statistic*. Concomitants of order statistics arise on several occasions. For a comprehensive review of the area see David and Nagaraja (1998); Sections 9.8 and 11.7 of David and Nagaraja (2003) contain a brief review. Recently He and Nagaraja (2009a) have proposed an efficient estimator of the correlation coefficient of a bivariate normal (BVN) distribution based on linear functions of the complete set of n concomitants and their order statistics. He and Nagaraja (2009b) have obtained the joint distribution of a single pair $(Y_{i:n}, Y_{[j:n]})$, $i,j=1,\ldots,n$, extracted from all of the n concomitants and used it to determine the expected cost of mismatch in broken bivariate samples.

In this paper we study the joint distribution of $(V_{s:m}, W_{t:n-m})$, where $V_{s:m}$ is the s th order statistic of the *upper* subset $\{Y_{[i:n]}, i=n-m+1,\ldots,n\}$, and $W_{t:n-m}$ is the t th order statistic of the *lower* subset $\{Y_{[j:n]}, j=1,\ldots,n-m\}$ of concomitants. Nagaraja and David (1994) discussed the finite-sample and asymptotic distributions of the upper extreme order statistic $V_{m:m}$ and Joshi and Nagaraja (1995) studied the joint distribution of extremes, $(V_{m:m}, W_{n-m:n-m})$. Chu et al. (1999) investigated the asymptotic distributions of extremes from central subsets of concomitants. Here we obtain results for central order statistics from upper and lower subsets, and discuss some interesting findings.

* Corresponding author.
E-mail address: kewang@fudan.edu.cn (K. Wang).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.04.008

We obtain in Section 2 the exact joint distribution of $(V_{s:m},W_{t:n-m})$ for the finite sample case using a conditioning argument. In Section 3 the asymptotic distribution of $(V_{s:m},W_{t:n-m})$ is obtained under appropriate regularity conditions assuming $m=\lceil np_0\rceil$, $s=\lceil mp_1\rceil$, and $t=\lceil (n-m)p_2\rceil$, $0<p_i<1, i=0,1,2$, as $n\to\infty$. A result on the rate of convergence is also provided. In Section 4 we propose a higher order approximation to these distributions. In Section 5, we illustrate our results with BVN and Farlie–Gumbel bivariate exponential (FGBVE) examples. For both distributions, we show that the second order approximation given in Section 4 performs markedly superior to the normal approximation even for moderate sample sizes. Section 6 discusses some applications of our results to data snooping and selection problems, and provides some generalizations of our results.

Let $F(x,y)$ be the joint cdf of (X,Y), and $f(x,y)$ be its joint pdf; $f_X(x)$ be the marginal pdf of X, and $f_{Y|X}(y|x)$ be the conditional pdf of Y given X. Let $F_1(y|x)(f_1(y|x))$ and $F_2(y|x)(f_2(y|x))$ be the conditional cdf (pdf) of Y given $X>x$ and $X<x$, respectively. For $0<p<1$, and a univariate cdf G, let $G^{-1}(p)$ be the p th quantile. A bivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by BVN$(\boldsymbol{\mu},\boldsymbol{\Sigma})$. A standard normal pdf is denoted by ϕ, cdf by Φ, and a normal distribution with mean μ and variance σ^2 is denoted by N(μ,σ^2).

2. Finite-sample distribution of $(V_{s:m},W_{t:n-m})$

We can derive the joint cdf of $(V_{s:m},W_{t:n-m})$ by conditioning on the value of $X_{n-m:n}$ and using the following result established by Kaufmann and Reiss (1992) (Theorem 2):

Lemma 1. *Given $X_{n-m:n}=x$, $V_{s:m}$ has the same distribution as the sth order statistic of a random sample of size m from the cdf $F_1(\cdot|x)$, and $W_{t:n-m}$ has the same distribution as the tth order statistic of the sample consisting of $n-m$ independent observations, of which $n-m-1$ are from the cdf $F_2(\cdot|x)$ and the remaining one is from the cdf $F_{Y|X}(\cdot|x)$. Moreover, $V_{s:m}$ and $W_{t:n-m}$ are conditionally independent given $X_{n-m:n}=x$.*

From Lemma 1, we have

$$F_{V_{s:m}W_{t:n-m}}(v,w)=\mathbf{P}(V_{s:m}\le v,W_{t:n-m}\le w)$$
$$=\int \mathbf{P}(V_{s:m}\le v|X_{n-m:n}=x)\mathbf{P}(W_{t:n-m}\le w|X_{n-m:n}=x)\,dF_{X_{n-m:n}}(x),$$

where

$$\mathbf{P}(V_{s:m}\le v|X_{n-m:n}=x)=\sum_{i=s}^{m}\binom{m}{i}[F_1(v|x)]^i[1-F_1(v|x)]^{m-i}$$

and

$$\mathbf{P}(W_{t:n-m}\le w|X_{n-m:n}=x)$$
$$=\sum_{j=t}^{n-m-1}\binom{n-m-1}{j}[F_2(w|x)]^j[1-F_2(w|x)]^{n-m-1-j}+\binom{n-m-1}{t-1}[F_2(w|x)]^{t-1}[1-F_2(w|x)]^{n-m-t}F_{Y|X}(w|x). \quad (1)$$

Remarks. 1. The conditional pdfs f_1 and f_2 are given by

$$f_1(y|x)=\frac{\int_x^\infty f(u,y)\,du}{1-F_X(x)};\quad f_2(y|x)=\frac{\int_{-\infty}^x f(u,y)\,du}{F_X(x)}. \quad (2)$$

2. Note that the conditional cdf in (1) is just the cdf of an order statistic when there is a single outlier (Arnold and Balakrishnan, 1989, p. 109).

3. Asymptotic joint normality

3.1. Limit distribution of $(V_{s:m},W_{t:n-m})$

Theorem 1. *Suppose $m=\lceil np_0\rceil$, $s=\lceil mp_1\rceil$, and $t=\lceil (n-m)p_2\rceil$, $0<p_i<1, i=0,1,2$, as $n\to\infty$. Let $x_0=F_X^{-1}(q_0)$ with $q_0=1-p_0$, and $h_i(x)=F_i^{-1}(p_i|x), i=1,2$, denote the p_ith quantile of the conditional cdf $F_i(\cdot|x)$ when viewed as a function of x. We assume that $h_i(x)$ has first order derivative at x_0 and that $f(x,y)>0$ in a neighborhood of $(x_0,h_i(x_0))$. Define $\sigma_0=\sqrt{p_0q_0}/f_X(x_0)$, and let $a_i=h_i(x_0)$, $c_i=\sigma_0h_i'(x_0)$, $i=1,2$, and*

$$b_1=\frac{\sqrt{p_1q_1}}{\sqrt{p_0}}f_1(a_1|x_0),\quad b_2=\frac{\sqrt{p_2q_2}}{\sqrt{q_0}}f_2(a_2|x_0).$$