ON CONCOMITANTS OF ORDER STATISTICS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree Doctor of Philosophy in the

Graduate School of The Ohio State University

By

Ke Wang, M.S.

* * * * *

The Ohio State University

2008

Dissertation Committee:

H. N. Nagaraja, Adviser

Douglas Wolfe

Shi Tao

Approved by

____

Adviser

Graduate Program in Statistics

ABSTRACT

Let $(X_i, Y_i), 1 \leq i \leq n$, be a sample of size n from an absolutely continuous random vector (X, Y). Let $X_{i:n}$ be the ith order statistic of the X-sample and $Y_{[i:n]}$ be its concomitant. We study three problems related to the $Y_{[i:n]}$'s in this dissertation. The first problem is about the distribution of concomitants of order statistics (COS) in dependent samples. We derive the finite-sample and asymptotic distribution of COS under a specific setting of dependent samples where the X's form an equally correlated multivariate normal sample. This work extends the available results on the distribution theory of COS in the literature, which usually assumes independent and identically distributed (i.i.d) or independent samples. The second problem we examine is about the distribution of order statistics of subsets of concomitants from i.i.d samples. Specifically, we study the finite-sample and asymptotic distributions of $V_{s:m}$ and $W_{t:n-m}$, where $V_{s:m}$ is the sth order statistic of the concomitants subset $\{Y_{[i:n]}, i = n - m + 1, \dots, n\}$, and $W_{t:n-m}$ is the tth order statistic of the concomitants subset $\{Y_{[j:n]}, j = 1, \dots, n - m\}$. We show that with appropriate normalization, both $V_{s:m}$ and $W_{t:n-m}$ converge in law to normal distributions with a rate of convergence of order $n^{-1/2}$. We propose a higher order expansion to the marginal distributions of these order statistics that is substantially more accurate than the normal approximation even for moderate sample sizes. Then we derive the finite-sample and asymptotic joint distribution of $(V_{s:m}, W_{t:n-m})$. We apply these results and determine the probability of an event of interest in commonly used selection procedures. We also apply the results to study the power of identifying the disease-susceptible gene in two-stage designs for gene-disease association studies. The third problem we consider is about estimating the conditional mean of the response variable (Y) given that the explanatory variable (X) is at a specific quantile of its distribution. We propose two estimators based on concomitants of order statistics. The first one is a kernel smoothing estimator, and the second one can be thought of as a bootstrap estimator. We study the asymptotic properties of these estimators and compare their finite sample behavior using simulation.

To my wife Haoying and my parents.

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. H. N. Nagaraja, for his priceless guidence, his insightful feedback, and his warm encouragement through the past few years. I would also like to thank the other members of my committee, Dr. Douglas A. Wolfe and Dr. Tao Shi, for their valuable comments and suggestions. I really thank the Department of Statistics at the Ohio State University for providing me a wonderful environment for my PhD study.

My special thanks go to my wife, Haoying Ou, for her love and support while I finished the degree. I would also like to express my sincere gratitude to my parents. Their enduring love and support helped me overcome all the difficulties I faced in the past.

VITA

July 31, 1975	Born - Changde, China
1997	B.E. Industrial Engineering, Nanjing University of Science and Technology, Nanjing, China
2000	M.A. Finance, Sun Yat-Sen University, Guangzhou, China
2003	M.S. Statistics, The Ohio State University, Columbus, OH, U.S.A.

PUBLICATIONS

Research Publications

N. Cressie, B. E. Buxton, C. A. Calder, P. F. Craigmile, C. Dong, N. J. McMillan, M. Morara, T. J. Santner, K. Wang, G. Young, and J. Zhang. (2007), "From sources to biomarkers: A hierarchical Bayesian approach for human exposure modeling", *Journal of Statistical Planning and Inference*, 137 (11), 3361-3379.

FIELDS OF STUDY

Major Field: Statistics

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF SYMBOLS AND ABBREVIATIONS

Notation	Meaning
BVN	bivariate normal distribution
COS	Concomitants of Order Statistics
cdf	cumulative distribution function
pdf	probability density function
$X_{1:n} \leq \ldots \leq X_{n:n}$	order statistics of the X sample values
$(X_{r:n}, Y_{[r:n]})$	rth X-order statistic and its Y-concomitant
$F_X(\cdot)$, $F_Y(\cdot)$	cumulative distribution functions (cdf's) of X, Y
$f_X(\cdot)$, $f_Y(\cdot)$	probability density functions (pdf's) of X, Y
$F_X^{-1}(p) := \inf\{x \vert F_X(x) \geq p\}$	inverse cdf or quantile function of X distribution, for $0 < p < 1$
$\xi_X(p) := F_X^{-1}(p)$	pth percentile of X distribution, for $0 < p < 1$
$F(x, y)$	joint cdf of (X, Y)
$f(x, y)$	joint pdf of (X, Y)
$F_{Y\vert X}(\cdot\vert x)$	conditional cdf of Y given that $X = x$
$f_{Y\vert X}(\cdot\vert x)$	conditional pdf of Y given that $X = x$
$F_1(\cdot\vert x)$	conditional cdf of Y given that $X > x$
$f_1(\cdot\vert x)$	conditional pdf of Y given that $X > x$
$F_2(\cdot\vert x)$	conditional cdf of Y given that $X < x$
$f_2(\cdot\vert x)$	conditional pdf of Y given that $X < x$
$f_{X_{r:n}}(\cdot)$	pdf of order statistic $X_{r:n}$
$f_{X_{r:n}, Y_{[r:n]}}(x, y)$	joint pdf of $(X_{r:n}, Y_{[r:n]})$
$f_{Y_{[r_1:n]}, \ldots, Y_{[r_k:n]}}(y_1, \ldots, y_k)$	joint pdf of $Y_{[r_1:n]}, \ldots, Y_{[r_k:n]}$
$\Phi(\cdot)$, $\phi(\cdot)$	cdf and pdf of standard normal distribution
$\mu_X := E(X)$, $\mu_Y := E(Y)$	means of X, Y
$\sigma_X^2 := \text{Var}(X)$, $\sigma_Y^2 := \text{Var}(Y)$	variances of X, Y
ρ	correlation coefficient (as in BVN)
$\Phi(x, y; \rho)$, $\phi(x, y; \rho)$	cdf and pdf of standard bivariate normal distribution with correlation ρ
$\mathbf{\Sigma}_{\boldsymbol{Y}} := \{\text{Cov}(Y_j, Y_k)\}_{j,k=1}^{l}$	variance-covariance matrix of the random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_l)$
$\mathbf{\Sigma}_{XY} = \mathbf{\Sigma}_{YX}^{\text{T}} := \{\text{Cov}(X, Y_j)\}_{j=1}^{l}$	covariance vector between the random variable X and the random vector $\boldsymbol{Y}$

Notation	**Meaning**
$per[A]$	permanent of the square matrix A
$[x]$	ceiling function, the largest integer not exceeding x
$X \stackrel{\mathrm{d}}{=} Y$	random variates X and Y have the same distribution
$\boldsymbol{J}_k$	a $k \times k$ matrix of 1's
$\boldsymbol{I}_k$	a $k \times k$ identity matrix
$m^{(l)}(\cdot)$	lth derivative of function m
$\hat{F}_n(\cdot) := \sum_{i=1}^{n} I(X_i \leq \cdot)$	empirical cdf based on random sample $(X_1, \ldots, X_n)$
$C_F(x, y)$	copula function associated with cdf F
$m(x) := E(Y\|X = x)$	conditional expectation of Y given that $X = x$ (or regression function of Y on X)
$\sigma^2(x) := \mathrm{Var}(Y\|X = x)$	conditional variance of Y given that $X = x$
VaR	Value-at-Risk
$K(\cdot)$	kernel function used in kernel smoothing methods
h_n, h	bandwidth used in kernel smoothing methods
$K_h(\cdot)$	$\frac{1}{h} K(\cdot/h)$

CHAPTER 1

OVERVIEW OF CONCOMITANTS OF ORDER STATISTICS

1.1 Introduction

Suppose (X_i, Y_i), $i = 1, \ldots, n$, is a random sample from a bivariate population (X, Y) with cdf $F(x, y)$. If we order the sample by the X-variate, and obtain the order statistics, $X_{1:n}, \ldots, X_{n:n}$, for the X sample, then the Y-variate associated with the rth 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 statistics* was first introduced by David (1973)[1].

Some generalizations to the above definition of concomitants of order statistics have been proposed. Barnett et al. (1976) considered a situation in which there are ℓ variates associated with each X. They proposed to order the $\ell + 1$ variate measurements $(X_i, Y_{1i}, \ldots, Y_{\ell i})$ based on the X-values and associated with $X_{r:n}$ will be the vector of concomitants $(Y_{1[r:n]}, \ldots, Y_{\ell[r:n]})$.

To allow for the selection based on more than one characteristic, Egorov and Nevzorov (1984), Reiss (1989), and Kaufmann and Reiss (1992) made a further generalization by considering the ordering of more than a single X. These authors proposed to order vectors

[1]Independently, Bhattacharya (1974) use the term *rth induced order statistic* for what is defined here, but the term *concomitant of order statistic* is more commonly used.

$\boldsymbol{x}_i$, $i = 1, \ldots, n$, by the size of some real-valued function $g(\boldsymbol{x}_i)$, resulting in the so-called g-ordering:

$$\boldsymbol{x}_k \leq_g \boldsymbol{x}_j \quad \text{if} \quad g(\boldsymbol{x}_i) \leq g(\boldsymbol{x}_j).$$

In particular, if $g(\boldsymbol{x}_i) = x_{1i}$, then the vectors are ordered by the first component x_{1i}, and the other components become the concomitants.

So far the random vectors are assumed to be independent and identically distributed. Eryilmaz (2005) generalized this by considering the concomitants of order statistics whenever $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent but otherwise arbitrarily distributed.

Concomitants of order statistics can arise in several applications. In selection procedures, items or subjects may be chosen on the basis of their X characteristic, and an associated characteristic Y that is hard to measure or can be observed only later may be of interest. For example, X may be the score of a candidate on a screening test, and Y is the measure of the final performance of the candidate; or X could be the score based on a particular search engine (like Google, Yahoo etc.), and Y is the score based on more exhaustive search of the internet. After the selection based on X values, the resulting measurements on the characteristic Y for the chosen subjects are actually the concomitants associated with the top X order statistics. (Without loss of generality we assume large Y values are desirable, and X and Y are positively correlated.) Under this setting, Yeo and David (1984) considered the problem of choosing the best k objects out of n candidates on the basis of auxiliary measurements X, while the measurements of primary interest Y are not available. The authors are interested in the probability that the m subjects with the largest X-values consists of the k objects with the largest Y-values.

Another application of concomitants of order statistics is in ranked-set sampling, first introduced by McIntyre (1952). It is a sampling scheme for situations where measurement of the variable of primary interest for sampled items is expensive or time-consuming while ranking of a set of items related to the variable of interest can be easily done. It can be shown that for such situations ranked-set sampling can achieve efficiency and reduce cost when compared to the simple random sampling. The original ranked-set sampling works as follows. A set of k items is drawn from the population, and we rank the items either by judgment or by actual measurement of some auxiliary variable X which is easy to measure. The item ranked the smallest is measured for the variable of our interest Y. Then another set of k items is drawn and ranked, and only the item ranked the second smallest is quantified. The procedure replicates until the item ranked the largest in the kth set is quantified. This completes a cycle of the sampling. The cycle is then repeated m times. The ranking involved in the above scheme can always be regarded as based on an auxiliary X-variate, representing a hypothetical (in the case of judgement ranking) or actual measurement. So the Y-value obtained in the ith set is actually the concomitant associated with ith X order statistic (David and Levine, 1972; Stokes, 1977). A comprehensive review of ranked set sampling can be found in Wolfe (2004), and in Chen et al. (2004).

Concomitants of order statistics have also been used in estimation and hypotheses testing problems. Spruill and Gastwirth (1982) have used concomitants to estimate the correlation coefficient between two sensitive variables, data on which are kept separately, and merge of the data is not possible due to confidentiality considerations. Another natural application of concomitants of order statistics is in dealing with the estimation of parameters for multivariate data sets that are subject to some form of type II censoring; examples include Harrell and Sen (1979), Gomes (1981, 1984), and Gill et al. (1990). For a recent comprehensive review of these applications see David and Nagaraja (1998) and Sections 9.8 and 11.7 of David and Nagaraja (2003).

For the remaining part of this chapter, we will review the basic finite-sample and asymptotic distribution theory of concomitants of order statistics and use bivariate normal distribution to provide illustrative examples. Then the outline and organization of the dissertation will be described.

1.2 Finite-Sample Distribution Theory

The finite-sample distribution theory for concomitants of order statistics has been investigated by several authors, for example by David (1973), David et al. (1977), Yang (1977), Bhattacharya (1984), and recently by Balasubramanian and Beg (1998), Eryilmaz (2005). Here we review some of the important results for the finite-sample distribution of concomitants of order statistics.

1.2.1 Concomitants with Regression Models

If the population distribution, (X, Y), is such that the following regression model for Y and X holds

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \ldots, n, \tag{1.1}$$

where $m(\cdot)$ is the regression function, and ε_i is the error term which is assumed to be independent of X_i, then we have

$$Y_{[r:n]} = m(X_{r:n}) + \varepsilon_{[r]}$$

where $\varepsilon_{[r]}$ is the ε_i associated with $X_{r:n}$, which can be shown to be independent of $X_{r:n}$ and have the same distribution as ε_i.

As discussed in David and Nagaraja (1998), one example of such a model is the linear regression model. In that case X_i and Y_i have means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and are linked by the linear relationship

$$Y_i = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X_i - \mu_X) + \varepsilon_i, \tag{1.2}$$

where ρ is the correlation between X and Y. A special case is where the distribution of (X, Y) is bivariate normal with mean $(\mu_X, \mu_Y)^{\mathrm{T}}$, and variance-covariance matrix

$$\begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

Then (1.2) will hold with ε_i being distributed as $N(0, \sigma_Y^2(1-\rho^2))$.

From (1.2) we have:

$$Y_{[i:n]} = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X_{i:n} - \mu_X) + \varepsilon_{[i]}. \tag{1.3}$$

So it follows that for any $r, s = 1, \ldots, n$,

$$\begin{aligned}
&\mathbf{E}(Y_{[r:n]}) = \mu_Y + \rho\sigma_Y\alpha_{r:n}, \\
&\mathrm{Var}(Y_{[r:n]}) = \sigma_Y^2(\rho^2\beta_{rr:n} + 1 - \rho^2), \\
&\mathrm{Cov}(X_{r:n}, Y_{[s:n]}) = \rho\sigma_X\sigma_Y\beta_{rs:n}, \\
&\mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]}) = \rho^2\sigma_Y^2\beta_{rs:n}, \qquad r \neq s,
\end{aligned} \tag{1.4}$$

where

$$\alpha_{r:n} = \mathbf{E}\left(\frac{X_{r:n} - \mu_X}{\sigma_X}\right) \quad \text{and} \quad \beta_{rs:n} = \mathrm{Cov}\left(\frac{X_{r:n} - \mu_X}{\sigma_X}, \frac{X_{s:n} - \mu_X}{\sigma_X}\right). \tag{1.5}$$

If we assume bivariate normality[2] for (X, Y), we will have the relations between the moments of $Y_{[r:n]}$ and $Y_{r:n}$ (Sondhauss, 1994):

[2]More generally, we only need to assume that the marginal distributions for X and Y are identical.

$$\mathbf{E}(Y_{[r:n]}) - \mu_Y = \rho(\mathbf{E}(Y_{r:n}) - \mu_Y),$$

$$\text{Var}(Y_{[r:n]}) - \sigma_Y^2 = \rho^2(\text{Var}(Y_{r:n}) - \sigma_Y^2),$$

$$\text{Cov}(Y_{[r:n]}, Y_{[s:n]}) = \rho^2\text{Cov}(Y_{r:n}, Y_{s:n}), \qquad r \neq s.$$

1.2.2 General Results

Without assuming the structural relation between X and Y as indicated by (1.1), Yang (1977) studied the exact distribution of $Y_{[r:n]}$. It was shown that if the (X_i, Y_i)'s are assumed to be i.i.d observations from some arbitrary absolutely continuous bivariate distribution with cdf $F(x, y)$, then for $1 \leq r_1 < \ldots < r_k \leq n$, the joint density for $(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})$ is given by

$$\begin{aligned} &f_{Y_{[r_1:n]},\ldots,Y_{[r_k:n]}}(y_1, \ldots, y_k) \\ &\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_2} \prod_{h=1}^{k} f_{Y|X}(y_h|x_h) f_{x_{r_1:n},\ldots,x_{r_k:n}}(x_1, \ldots, x_k)dx_1 \ldots dx_k. \end{aligned} \tag{1.6}$$

This follows directly from an important result regarding the conditional independence of $(Y_{[1:n]}, \ldots, Y_{[n:n]})$ given the values of $(X_{1:n}, \ldots, X_{n:n})$, which is due to Bhattacharya (1974). The result is given by the following proposition:

Proposition 1.2.1. *The concomitants of order statistics, $(Y_{[1:n]}, \ldots, Y_{[n:n]})$, are conditionally independent given $X_1 = x_1, \ldots, X_n = x_n$ with conditional cdf's $F_{Y|X}(\cdot|X = x_{1:n}), \ldots,$ $F_{Y|X}(\cdot|X = x_{n:n})$, respectively, where the $x_{i:n}$'s are ordered x_i's such that $x_{1:n} \leq \ldots \leq$ $x_{n:n}$.*

It follows from Proposition 1.2.1 that for any $k \leq n$ with $1 \leq r_1 < \ldots < r_k \leq n$, the $Y_{[r_h:n]}, h = 1, \ldots, k$, are conditionally independent given $X_{r_h:n} = x_h, h = 1, \ldots, k$, with joint conditional pdf

$$f_{Y_{[r_1:n]},\ldots,Y_{[r_k:n]}|X_{r_1:n}=x_1,\ldots,X_{r_k:n}=x_k}(y_1,\ldots,y_k)=\prod_{h=1}^{k} f_{Y|X}(y_h|x_h).$$

By Proposition 1.2.1, we can obtain the following results for the moments of concomitants of order statistics

$$\begin{aligned}
&\mathbf{E}(Y_{[r:n]}) = \mathbf{E}[m(X_{r:n})],\\
&\mathrm{Var}(Y_{[r:n]}) = \mathrm{Var}[m(X_{r:n})] + \mathbf{E}[\sigma^2(X_{r:n})],\\
&\mathrm{Cov}(X_{r:n}, Y_{[s:n]}) = \mathrm{Cov}[X_{r:n}, m(X_{s:n})],\\
&\mathrm{Cov}(Y_{[r:n]}, Y_{[s:n]}) = \mathrm{Cov}[m(X_{r:n}), m(X_{s:n})], \qquad r \neq s
\end{aligned} \tag{1.7}$$

where $m(x) = \mathbf{E}(Y|X = x)$ and $\sigma^2(x) = \mathrm{Var}(Y|X = x)$.

Example 1.2.1. Suppose the distribution we are sampling from is bivariate normal given by:

$$(X, Y)^{\mathrm{T}} \sim N\left(\begin{pmatrix}\mu_X\\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right). \tag{1.8}$$

By (1.6), the pdf of the concomitant of rth order statistic $Y_{[r:n]}$ is given by:

$$f_{Y_{[r:n]}}(y) = \int_{-\infty}^{\infty} \phi\left[\frac{(y-\mu_Y)/\sigma_Y - \rho(x-\mu_X)/\sigma_X}{\sqrt{1-\rho^2}}\right]\frac{f_{X_{r:n}}(x)}{\sigma_Y\sqrt{1-\rho^2}}dx$$

where

$$f_{X_{r:n}}(x) = \frac{n!}{(r-1)!(n-r)!\sigma_X}\Phi\left(\frac{x-\mu_X}{\sigma_X}\right)^{r-1}\left[1-\Phi\left(\frac{x-\mu_X}{\sigma_X}\right)\right]^{n-r}\phi\left(\frac{x-\mu_X}{\sigma_X}\right),$$

and $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf of the standard normal distribution, respectively. Alternatively we can derive the same result as above for the pdf of $Y_{[r:n]}$ using (1.3) by noticing that the distribution of $(Y_{[r:n]} - \mu_Y)/\sigma_Y$ is a convolution of $\rho(X_{r:n} - \mu_X)/\sigma_X$ and $N(0, 1 - \rho^2)$.

Using similar arguments as in Bhattacharya (1984), we can generalize Proposition 1.2.1 by considering the case in which the (X_i, Y_i)'s are still independent but not necessarily identically distributed, as stated in the Proposition below:

Proposition 1.2.2. *Suppose the random vectors $(X_i, Y_i), i = 1, \ldots, n$, are independent, and (X_i, Y_i) has cdf $F^{(i)}(x, y)$. Then the concomitants of order statistics, $(Y_{[1:n]}, \ldots, Y_{[n:n]})$, are conditionally independent given $X_1 = x_1, \ldots, X_n = x_n$ with conditional cdf's $F_{Y|X}^{(\lambda(i,\boldsymbol{x}))}(\cdot|x_{i:n}), i = 1, \ldots, n$, respectively, where $\lambda(i, \boldsymbol{x})$ is defined to be the index in $\{1, \ldots, n\}$ such that for the given vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, the ith smallest value $(x_{i:n})$ is $x_{\lambda(i,\boldsymbol{x})}$, and $F_{Y|X}^{(i)}(\cdot|x)$ is the conditional cdf of Y given $X = x$ corresponding to distribution $F^{(i)}(x, y)$.*

Proof. Let $\boldsymbol{X_n} = (X_1, \ldots, X_n)$. As discussed in Bhattacharya (1984), the random permutation, $(\lambda(1, \boldsymbol{X_n}), \ldots, \lambda(n, \boldsymbol{X_n}))$ of $(1, \ldots, n)$ will be determined only by $\boldsymbol{X_n}$, and we have $Y_{\lambda(i,\boldsymbol{X_n})} = Y_{[i:n]}$. By the independence of (X_i, Y_i)'s, Y_j will be independent of $\{(X_i, Y_i), i \neq j\}$ for each j. So we have:

$$
\begin{aligned}
&\mathbf{P}(Y_{[j:n]} \leq y_j, j = 1, \ldots, n | X_i = x_i, i = 1, \ldots, n) \\
&=\mathbf{P}(Y_{\lambda(j,\boldsymbol{x_n})} \leq y_{\lambda(j,\boldsymbol{x_n})}, j = 1, \ldots, n | X_{\lambda(i,\boldsymbol{x_n})} = x_{\lambda(i,\boldsymbol{x_n})}, i = 1, \ldots, n) \\
&=\prod_{i=1}^{n} \mathbf{P}(Y_{\lambda(i,\boldsymbol{x_n})} \leq y_{\lambda(i,\boldsymbol{x_n})} | X_{\lambda(i,\boldsymbol{x_n})} = x_{\lambda(i,\boldsymbol{x_n})}) \\
&=\prod_{i=1}^{n} F^{\lambda(i,\boldsymbol{x_n})}(y_{\lambda(i,\boldsymbol{x_n})} | x_{i:n})
\end{aligned}
$$

which establishes the desired result. □

Remark 1.2.1. Notice that unlike Lemma 1.2.1, we cannot derive from Proposition 1.2.2 the result that for any $k \leq n$ with $1 \leq r_1 < \ldots < r_k \leq n$, the $Y_{[r_h:n]}$'s are conditionally independent given $X_{r_h:n} = x_h, h = 1, \ldots, k$. One exception is that if the distributions

$F^{(i)}(x,y), i = 1,\ldots,n$ are such that the conditional distributions $F^{(i)}_{Y|X}(\cdot|x)$ are all the same, then the conditional independence of the Y-concomitants given the values of X order statistics still holds, and the conditional distribution of $(Y_{[r_1:n]},\ldots,Y_{[r_k:n]})$ given the values of order statistics $(X_{r_1:n},\ldots,X_{r_k:n})$ is given by:

$$\mathbf{P}(Y_{[r_j:n]} \leq y_j, j=1,\ldots,k|X_{r_i:n}=x_i, i=1,\ldots,k) = \prod_{i=1}^{k} F(y_i|x_i), \tag{1.9}$$

with $F(y|x)$ being the common conditional cdf of Y given X. The expression on the right hand side of (1.9) is exactly the same as it would be in the i.i.d case. One example for such situation is as following. Suppose we have two applicant pools. Let $X_i, i = 1,\ldots,m$, be the scores of the screening test for the applicant pool 1, and $X_i, i = m+1,\ldots,m+n$, be the scores for the applicant pool 2. We assume that $X_i \sim N(\mu_1,\sigma_1^2), i = 1,\ldots,m$, and $X_i \sim N(\mu_2,\sigma_2^2), i = m+1,\ldots,m+n$. Let Y be the final score for the measure of the applicant's performance, and we assume for both pool 1 and pool 2, Y depends on the initial score X through the following simple linear regression

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1,\ldots,m+n, \tag{1.10}$$

with $\varepsilon \sim N(0,\sigma_e^2)$. Even though the (X_i,Y_i)'s are not identically distributed, they share the common conditional distribution of Y given the value of X, which is $N(\beta_0+\beta_1 x,\sigma_e^2)$. As a result, the conditional distribution of the Y-concomitants given the values of X order statistics is the same as in the i.i.d case.

1.2.3 Multivariate Generalization

Now consider the multivariate case in which there are ℓ variates associated with each X, and we have n independent sets of variates $(X_i, Y_{1i},\ldots,Y_{\ell i})$. As in the bivariate case we order the sample based on X sample values, and associated with $X_{r:n}$ is the vector of concomitants $(Y_{1[r:n]}, \ldots, Y_{\ell[n:n]})$. This situation has applications in hydrology and has been intensively studied by Song et al. (1992), Song and Deddens (1993), and Balakrishnan (1993).

Let $m_j(x_i) = \mathbf{E}(Y_{ji}|X_i = x_i)$ and $\sigma_{jk}(x_i) = \text{Cov}(Y_{ji}, Y_{ki}|X_i = x_i)$, we have the following results for the moments of concomitants similar to (1.7):

$$\begin{aligned}
&\mathbf{E}(Y_{j[r:n]}) = \mathbf{E}[m_j(X_{r:n})], \\
&\text{Cov}(Y_{j[r:n]}, Y_{j[s:n]}) = \text{Cov}[m_j(X_{r:n}), m_j(X_{s:n})] + \mathbf{E}[\sigma_{jk}(X_{r:n})]
\end{aligned} \tag{1.11}$$

Suppose that $(X_i, Y_{1i}, \ldots, Y_{\ell i}), i = 1, \ldots, n$, is a random sample from a multivariate normal distribution with mean vector

$$\boldsymbol{\mu} = (\mu_X, \mu_1, \ldots, \mu_\ell)^{\text{T}}$$

and variance-covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_Y \end{pmatrix}$$

where

$$\begin{aligned}
&\mu_X = E(X_i) \quad \text{and} \quad \mu_j = E(Y_{ij}), j = 1, \ldots, \ell \\
&\boldsymbol{\Sigma}_{XY} = \boldsymbol{\Sigma}_{YX}^{\text{T}} = \{\text{Cov}(X_i, Y_{ij})\}_{j=1}^{\ell} = (\sigma_{xj})_{j=1}^{\ell}, \\
&\boldsymbol{\Sigma}_Y = \{\text{Cov}(Y_{ij}, Y_{ik})\}_{j,k=1}^{\ell} = (\sigma_{jk})_{j,k=1}^{\ell}.
\end{aligned}$$

Notice that for $i = 1, \ldots, n$

$$Y_{ij} = \mu_j + \rho_j \sigma_{jj} \frac{X_i - \mu_X}{\sigma_X} + \varepsilon_{ij}$$

where $\rho_j = \text{Corr}(X_i, Y_{ij})$, $\boldsymbol{\varepsilon_i} = (\varepsilon_{i1}, \ldots, \varepsilon_{i\ell})^{\mathrm{T}}$ are independent of $X_i, i = 1, \ldots, n$, and are i.i.d multivariate normal with mean vector $\mathbf{0}$ and variance-covariance matrix:

$$\mathbf{\Sigma}_{Y|X} = \mathbf{\Sigma}_Y - \sigma_X^{-2}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XY}.$$

So we have

$$Y_{j[r:n]} = \mu_j + \rho_j\sigma_{jj}\frac{X_{r:n} - \mu_X}{\sigma_X} + \varepsilon_{j[n]}$$

where $\boldsymbol{\varepsilon}_{[r]} = (\varepsilon_{1[r]}, \ldots, \varepsilon_{\ell[r]})^{\mathrm{T}}$ are independent of $X_{i:n}, i = 1, \ldots, n$, and have the same distribution as $\boldsymbol{\varepsilon}_i$.

Thus it follows that

$$\begin{aligned}
&\mathbf{E}(Y_{j[r:n]}) = \mu_j + \rho_j\sigma_{jj}\alpha_{r:n},\\
&\text{Cov}(Y_{j[r:n]}, Y_{k[r:n]}) = \rho_j\rho_k\sigma_{jj}\sigma_{kk}\beta_{rr:n} + \sigma_{jk}(x) = \sigma_{jk} - \rho_j\rho_k\sigma_{jj}\sigma_{kk}(1 - \beta_{rr:n}),\\
&\text{Cov}(Y_{j[r:n]}, Y_{k[s:n]}) = \rho_j\rho_k\sigma_{jj}\sigma_{kk}\beta_{rs:n}, \qquad r \neq s,
\end{aligned}$$

where $\sigma_{jk}(x) = \sigma_{jk} - \rho_j\rho_k\sigma_{jj}\sigma_{kk}$ is the jkth element of the variance-covariance matrix $\mathbf{\Sigma}_{Y|X}$, and $\alpha_{r:n}$ and $\beta_{rs:n}$ are defined by (1.5).

1.2.4 Non-identical Distribution Case

Concomitants of order statistics for the situation in which the random vectors are independent but otherwise arbitrarily distributed are considered by Eryilmaz (2005). Let $(X_i, Y_i), i = 1, \ldots, n$, be independent random vectors with cdf's $F_i(x, y), i = 1, \ldots, n$, respectively, the distribution of concomitant $Y_{[r:n]}$ can be expressed in terms of permanents as in the following result due to Eryilmaz (2005).

Theorem 1.2.1. *Let $(X_i, Y_i), i = 1, \ldots, n$, be independent random vectors with cdf's $F_i(x, y), i = 1, \ldots, n$, and marginals F_{X_i}, F_{Y_i} respectively. Then the cdf of concomitant*

$Y_{[r:n]}$ is given by

$$\mathbf{P}(Y_{[r:n]} \leq y) = \frac{1}{(r-1)!(n-r)!} \sum_{i=1}^{n} \int \mathbf{P}(Y_i \leq y | X_i = x) \, per \begin{array}{c} \begin{bmatrix} F_{X_1}(x) & 1 - F_{X_1}(x) \\ \vdots & \vdots \\ F_{X_{i-1}}(x) & 1 - F_{X_{i-1}}(x) \\ F_{X_{i+1}}(x) & 1 - F_{X_{i+1}}(x) \\ \vdots & \vdots \\ F_{X_n}(x) & 1 - F_{X_n}(x) \end{bmatrix} \\ \begin{array}{cc} r-1 & \quad n-r \end{array} \end{array} dF_{X_i}(x) \tag{1.12}$$

where the permanent

$$per \begin{array}{c} \begin{bmatrix} F_{X_1}(x) & 1 - F_{X_1}(x) \\ \vdots & \vdots \\ F_{X_{i-1}}(x) & 1 - F_{X_{i-1}}(x) \\ F_{X_{i+1}}(x) & 1 - F_{X_{i+1}}(x) \\ \vdots & \vdots \\ F_{X_n}(x) & 1 - F_{X_n}(x) \end{bmatrix} \\ \begin{array}{cc} r-1 & \quad n-r \end{array} \end{array}$$

is defined as in Vaughan and Venables (1972), i.e., for a square matrix $A = \{a_{ij}\}_{i,j=1}^{n}$, the permanent is defined to be

$$per[A] = \sum_{\sigma} \prod_{i=1}^{n} a_{i\sigma(i)} \tag{1.13}$$

where $\sigma = (\sigma(1), \ldots, \sigma(n))$ is a permutation of $(1, \ldots, n)$, and the summation in (1.13) is over all possible permutations of $(1, \ldots, n)$.

Eryilmaz (2005) also derived the joint distribution of $(Y_{[1:n]}, Y_{[n:n]})$ for the non-identical case. He used FGM type bivariate distribution and the Marshall and Olkin's bivariate exponential distribution as illustrations.

1.3 Asymptotic Distribution Theory

1.3.1 Marginal Distributions

We first consider the marginal limiting distribution for concomitant $Y_{[r:n]}$. Intuitively the asymptotic distribution of $Y_{[r:n]}$ depends on the growth pattern of r as $n \to \infty$, and more importantly by the dependence structure of the bivariate cdf $F(x, y)$. Usually three situations are considered for the growth pattern of r as $n \to \infty$: (i) the quantile case in which $r = [np]$, with $0 < p < 1$; (ii) the extremal case in which either r or $n - r$ is fixed; (iii) the intermediate case in which $r \to \infty$, $n - r \to \infty$ in such a way that r/n approaches either 0 or 1.

A convenient way to model the dependence structure of (X, Y) is to assume that the simple linear regression model (1.2) holds for the random vector (X, Y). In this case analytical results regarding the asymptotic distribution of the concomitant of order statistic $Y_{[r:n]}$ can be obtained as in David and Galambos (1974) and David (1994).

Recall that under the simple linear regression model (1.2), we can express $Y_{[r:n]}$ as:

$$Y_{[r:n]} = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X_{r:n} - \mu_X) + \epsilon_{[r]}. \tag{1.14}$$

Without loss of generality, we might assume that $\mu_X = \mu_Y = 0$, and $\sigma_X = \sigma_Y = 1$ in (1.14).

From (1.14) and the fact that $X_{r:n}$ is independent of $\varepsilon_{[r]}$, we see that the asymptotic distribution of $Y_{[r:n]}$ depends on that of $X_{r:n}$, as well as the distribution of ϵ in a linear fashion. It is useful to differentiate between the following two cases regarding the limiting behavior of $X_{r:n}$:

(i) $X_{r:n}$ converges in probability, i.e., there exists a sequence $\{a_n\}$ such that

$$X_{r:n} - a_n \xrightarrow{P} 0, \qquad \text{as } n \to \infty. \tag{1.15}$$

In this case it follows directly from the Slutsky theorem that:

$$Y_{[r:n]} - \rho a_n \xrightarrow{\mathcal{L}} \epsilon, \qquad \text{as } n \to \infty. \tag{1.16}$$

The condition (1.15) holds in the following cases:

(a) $r = [np]$, $0 < p < 1$, as $n \to \infty$. In this case we have $X_{r:n} \xrightarrow{P} F_X^{-1}(p)$ with the assumption that the cdf $F_X(x)$ is absolutely continuous and has positive density at $F_X^{-1}(p)$;

(b) $r = n - k + 1$ for some fixed k. In this case, a necessary and sufficient condition for (1.15) to hold has been established by Hall (1979). In particular if the marginal distribution of X is standard normal, then (1.15) holds with $a_n = \sqrt{2 \log n}$; see David (1994) for a detailed discussion.

(ii) If $X_{r:n}$ fails to converge in probability, and instead there exist constants a_n and $b_n > 0$ such that

$$\frac{X_{r:n} - a_n}{b_n} \xrightarrow{\mathcal{L}} W, \qquad \text{as } n \to \infty \tag{1.17}$$

for some nondegenerate random variable W. Here, if $b_n \to b$ as $n \to \infty$, then we have

$$\frac{Y_{[r:n]} - \rho a_n}{b_n} \xrightarrow{\mathcal{L}} \rho W + \frac{\epsilon}{b}, \qquad \text{as } n \to \infty. \tag{1.18}$$

So in this case with appropriate normalization, $Y_{[r:n]}$ converges in distribution to the convolution of ρW and ε.

In upper (lower) extreme case, i.e., $r = n - k + 1$ ($r = k$) for some fixed k, (1.17) will hold with W being distributed as the kth lower (upper) record value from one of the three extreme value distributions (Nagaraja and David, 1994). In the intermediate case with r/n approaching 1 as $n \to \infty$, if the distribution of X is bounded above, (1.17) holds with W being normally distributed (Reiss, 1989, p109).

The limiting distribution of $Y_{[r:n]}$ will generally depend on the conditional distribution of Y given X, and the marginal distribution of X, as suggested by the following theorem for the extreme case ($r = n - k + 1$). The theorem is due to Galambos (1978), and is extended in David (1994) and Sondhauss (1994).

Theorem 1.3.1. *Let $F_X(x)$ satisfy one of the von Mises conditions* [3] *and assume that the sequences of constants a_n, $b_n > 0$, are such that as $n \to \infty$,*

$$[F_X(a_n + b_n x)]^n \to G(x)$$

for all x. Further, suppose there exist constants A_n and B_n such that

$$F_{Y|X}(A_n + B_n y | a_n + b_n x) \to H(y|x) \tag{1.19}$$

uniformly for all x and y. Then

$$\mathbf{P}(Y_{[n-k+1:n]} \leq A_n + B_n y) \to \int_{-\infty}^{\infty} H(y|x) dG_{(k)}(x)$$

where $G_{(k)}$ is the cdf of the kth lower record value from the extreme value cdf G.

Notice that if the joint distribution of (X, Y) is such that as $x \to F_X^{-1}(1)$, $F_{Y|X}(y|x) \to H(y)$ for all y, then the condition (1.19) of Theorem 1.3.1 will hold with $A_n = 0$, $B_n = 1$, and it follows that $\mathbf{P}(Y_{[n-k+1:n]} \leq y) \to H(y)$. Arnold et al. (1992, p221) provided an example of bivariate exponential distribution for such a situation.

[3] see Resnick (1987, p62-64) for details of these conditions.

Example 1.3.1. For the bivariate normal population given by (1.8), without loss of generality, we assume $\mu_X = \mu_Y = 0$, and $\sigma_X = \sigma_Y = 1$. We want to derive the asymptotic behavior of $Y_{[r:n]}$ for both the quantile and extremal cases.

For the quantile case, i.e., $r = [np]$ for some $0 < p < 1$, by the fact that

$$X_{r:n} \xrightarrow{P} \Phi^{-1}(p) \qquad \text{as } n \to \infty,$$

we have

$$Y_{[r:n]} \xrightarrow{\mathcal{L}} N(\rho\Phi^{-1}(p), 1-\rho^2)$$

as discussed above.

For the extremal case, i.e., $r = n - k + 1$ for some fixed integer k and $n \to \infty$, as discussed in David (1994) we have

$$X_{r:n} - a_n \xrightarrow{P} 0,$$

with $a_n = \sqrt{2\log n}$. So it follows that

$$Y_{[r:n]} - \rho\sqrt{2\log n} \xrightarrow{\mathcal{L}} N(0, 1-\rho^2) \qquad \text{as } n \to \infty \tag{1.20}$$

We can also obtain (1.20) by using Theorem 1.3.1. First notice that the standard normal cdf $\Phi \in D(\Lambda)$ and satisfies the associated von Mises condition (Resnick, 1987; David and Nagaraja, 2003). The corresponding norming constants can be chosen as

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

where Λ is the Gumbel extreme value distribution. Let $A_n = \rho a_n$ and $B_n = \sqrt{1-\rho^2}$. Since the conditional distribution of Y given $X = x$ is $N(\rho x, 1-\rho^2)$, it follows that

$$F_{Y|X}(A_n + B_n y | a_n + b_n x) = \Phi\left(y - \frac{\rho}{\sqrt{1-\rho^2}} b_n x\right) \to \Phi(y)$$

uniformly for all x and y as $n \to \infty$. So we have

$$\frac{Y_{[r:n]} - \rho a_n}{\sqrt{1-\rho^2}} \xrightarrow{\mathcal{L}} N(0,1), \qquad \text{as } n \to \infty$$

which is equivalent to (1.20).

1.3.2 Joint Distributions

The asymptotic distribution of a finite set of concomitants has also been studied by several authors. Under the assumption that the random vector (X, Y) is distributed such that $Y - \mathbf{E}(Y|X)$ and X are independent (which is equivalent the assumption that the general regression model given by (1.1) holds), David and Galambos (1974) showed that for any fixed k and any choice $1 \le r_1 < \cdots < r_k \le n$, $Y_{[r_1:n]}, \ldots, Y_{[r_k:n]}$ are asymptotically independent provided that Var$[\mathbf{E}(Y_{[r_i:n]}|X_{r_i:n})]$ approaches 0 as $n \to \infty$ for all $i = 1, \ldots, k$.

In the quantile case where $r_i/n \to p_i$, $0 < p_i < 1$, for $i = 1, \ldots, k$, Yang (1977) proved the following theorem.

Theorem 1.3.2. *Let (X_i, Y_i), $i = 1, \ldots, n$, be n i.i.d observations from an absolutely continuous distribution with cdf $F(x, y)$ and pdf $f(x, y)$. Let $1 \le r_1 < \cdots < r_k \le n$ be sequences of integers such that as $n \to \infty$, $r_i/n \to p_i$, $0 < p_i < 1$, for $i = 1, \ldots, k$. Then we have*

$$\mathbf{P}(Y_{[r_1:n]} \le y_1, \ldots, Y_{[r_k:n]} \le y_k) \to \prod_{i=1}^{k} F_{Y|X}(y_i|F_X^{-1}(p_i)). \tag{1.21}$$

In the extremal case, assuming that the conditions of Theorem 1.3.1 hold, it follows from the conditional independence of concomitants and the Bounded Convergence Theorem that:

$$\begin{aligned}&\mathbf{P}(Y_{[n:n]} \leq A_n + B_n y_1, \ldots, Y_{[n-k+1:n]} \leq A_n + B_n y_k)\\ \to &\int_{x_1 > \cdots > x_k} \prod_{i=1}^{k} H(y_i|x_i) dG_k(x_1, \ldots, x_k)\end{aligned}$$

where G_k is the joint cdf of the first k lower record values from the extreme value distribution G (Nagaraja, 1982; David, 1994).

1.4 Outline and Organization of the Dissertation

In Chapters 2 to 5 of this dissertation, we will study three separate topics related to the concomitants of order statistics. Here we give a brief overview of these topics, and describe the organization of the rest of the dissertation.

From the review of the distribution theory about concomitants of order statistics in this chapter, we see that concomitants of order statistics were only studied under assumption of i.i.d or independent samples in the literature. In Chapter 2, we derive the finite-sample and asymptotic distribution of concomitants of order statistics for a special case of dependent samples. The relevant results are illustrated with a simple example.

In Chapter 3 and 4, we study order statistics of concomitants of subsamples. Let $V_{s:m}$ be the sth order statistic of the concomitants subset $\{Y_{[i:n]}, i = n-m+1, \ldots, n\}$, and $W_{t:n-m}$ be the tth order statistic of the concomitants subset $\{Y_{[j:n]}, j = 1, \ldots, n-m\}$. In Chapter 3, we study the distributions of $V_{s:m}$ and $W_{t:n-m}$ separately. Both the finite-sample and asymptotic distribution (in the quantile case) are derived for $V_{s:m}$ and $W_{t:n-m}$. The rates of convergence in these distributions are also studied, and we propose a higher order approximation to these distributions with better performance even for moderate sample sizes.

The joint distribution of $V_{s:m}$ and $W_{t:n-m}$ is then studied in Chapter 4. We first derive the finite-sample distribution of $(V_{s:m}, W_{t:n-m})$ using a conditioning argument. Then the asymptotic distribution of $(V_{s:m}, W_{t:n-m})$ for both the quantile and extremal cases are obtained. Finally the results are applied to study the probability of an event of interest in a selection procedure. We also apply the results to study the power of identifying the disease-susceptible gene in two-stage designs for gene-disease association studies as discussed in Satagopan et al. (2002) and Satagopan et al. (2004).

In Chapter 5, we consider the problem of estimating the conditional mean of response variable given that the explanatory variable is at specific quantiles of its distribution. This is closely related to the usual bivariate regression problem, but differs from it in that the evaluation point is itself an unknown parameter. We propose two estimators based on concomitants of order statistics. The first class of estimators is a kernel smoothing estimator generalized from Yang (1981), and the second estimator, motivated by Mausser (2001), can be thought of as a "bootstrap" estimator. The asymptotic properties of these estimators are studied, and we compare the finite sample behavior of these estimators based on Monte Carlo studies.

In Chapter 6 we conclude with a summary of the results derived in previous chapters. A general discussion on future work related to this dissertation is also provided.

CHAPTER 2

CONCOMITANTS OF ORDER STATISTICS FOR DEPENDENT SAMPLES

In this Chapter, we will study the concomitants of order statistics for dependent samples. In particular we consider the case in which the bivariate random vectors (X_i, Y_i), $i = 1, \ldots, n$, are observations from some common distribution with cdf $F(x, y)$, but no longer mutually independent. Section 2.1 describes the basic settings under which the concomitants of order statistics are studied, and the finite-sample and asymptotic distributions of concomitants are examined in Section 2.2 and 2.3.

2.1 Basic Setting

Suppose (X_i, Y_i), $i = 1, \ldots, n$, are a sequence of random vectors such that

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \ldots, n \tag{2.1}$$

where m is some smooth regression function; ε_i's are i.i.d error terms with standard normal distribution. Unlike the usual i.i.d assumption on the X_i's, we assume that they come from some common distribution with cdf $F_X(x)$, but have some form of dependence structure among them. Specifically we assume that the X_i's are multivariate normal with mean vector $\mathbf{0}$, and variance-covariance matrix $\mathbf{\Sigma}$ where

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}. \tag{2.2}$$

The condition $\rho > -1/(n-1)$ is needed to ensure the positive definiteness of $\mathbf{\Sigma}$, but we will assume that $\rho > 0$ in our discussion. Notice that the X_i's are identically distributed as standard normal, and are equally correlated. As usual, we assume the error terms ε_i's are independent of the X_i's. Under this setting, the random vectors are no longer i.i.d, but they follow the common distribution $F(x, y)$, and the random variates Y_i's are conditionally independent given the values of X_i's. Using similar arguments as in the proof of Proposition 1.2.2 we can establish the conditional independence of the Y concomitants given the values of X_i's. The conditional independence of the Y concomitants given the values of X order statistics will also hold since the (X_i, Y_i)'s are identically distributed (see Remark 1.2.1).

2.2 Finite-Sample Distribution of $Y_{[r:n]}$

In this section we will derive the distribution of concomitants $Y_{[i:n]}$ under the conditions specified in Section 2.1.

From (2.1), we have

$$Y_{[r:n]} = m(X_{r:n}) + \varepsilon_{[r]} \tag{2.3}$$

where $\varepsilon_{[r]}$ is the error term associated with $X_{r:n}$. Since the ε_i's are independent of the X_i's, $\varepsilon_{[r]}$ has the same distribution as ε_i's, i.e., $\varepsilon_{[r]} \sim N(0, 1)$. So we have

$$[Y_{[r:n]} | X_{r:n} = x] \sim N(m(x), 1). \tag{2.4}$$

The density for $Y_{[r:n]}$ can then be derived as

$$f_{Y_{[r:n]}}(y) = \int \phi(y - m(x)) f_{X_{r:n}}(x) dx, \tag{2.5}$$

where $f_{r:n}(\cdot)$ is the density for the rth order statistic of the X sample values.

For equally correlated multivariate normal variates, as demonstrated in David and Nagaraja (2003, p100-101), the X_i's can be represented as

$$X_i = \sqrt{\rho} Z_0 + \sqrt{1-\rho} Z_i, \quad \text{for } \rho > 0, \tag{2.6}$$

where $Z_i, i = 0, \ldots, n$, are i.i.d standard normal variates. So the rth order statistic of X sample can be expressed as

$$X_{r:n} = \sqrt{\rho} Z_0 + \sqrt{1-\rho} Z_{r:n}, \tag{2.7}$$

where $Z_{r:n}$ is the rth order statistic of $(Z_1, \ldots, Z_n)$. Then the cdf for $X_{r:n}$ can be derived as

$$\begin{aligned} F_{X_{r:n}}(x) &= \mathbf{P}(X_{r:n} \leq x) \\ &= \int \mathbf{P}\left(Z_{r:n} \leq \frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \phi(z) dz \\ &= \sum_{k=r}^{n} \binom{n}{k} \int \left[\Phi\left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \right]^k \left[1 - \Phi\left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \right]^{n-k} \phi(z) dz, \end{aligned} \tag{2.8}$$

and the pdf for $X_{r:n}$ is given by:

$$f_{X_{r:n}}(x) = \frac{1}{\sqrt{1-\rho}} \int f_{Z_{r:n}} \left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \phi(z) dz, \tag{2.9}$$

with

$$\begin{aligned} & f_{Z_{r:n}} \left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \\ &= \frac{n!}{(r-1)!(n-r)!} \left[\Phi\left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \right]^{r-1} \left[1 - \Phi\left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right) \right]^{n-r} \phi\left(\frac{x - \sqrt{\rho} z}{\sqrt{1-\rho}} \right). \end{aligned} \tag{2.10}$$

Example 2.2.1. (Simple Linear Regression) Suppose the X_i and Y_i, $i = 1, \ldots, n$, are linked through the following simple linear regression model:

$$Y_i = \beta X_i + \varepsilon_i. \tag{2.11}$$

Then we have

$$[Y_{[r:n]} | X_{r:n} = x] \sim N(\beta x, 1), \tag{2.12}$$

and the density for $Y_{[r:n]}$ is given by

$$f_{Y_{[r:n]}}(y) = \int \phi(y - \beta x) f_{X_{r:n}}(x) dx \tag{2.13}$$

with $f_{X_{r:n}}(x)$ given by (2.9).

2.3 Asymptotic Distribution of $Y_{[r:n]}$

From (2.3), we see that the asymptotic distribution of $Y_{[r:n]}$ will depend on the limiting behavior of $X_{r:n}$ as well as the regression function m. We consider two cases, namely the quantile case and the extremal case, while deriving the asymptotic distribution of $Y_{[r:n]}$.

2.3.1 The Quantile Case

In this case $r/n \to p$ as $n \to \infty$. By (2.7), and the fact that

$$Z_{r:n} \xrightarrow{P} \Phi^{-1}(p) \tag{2.14}$$

we have

$$X_{r:n} \xrightarrow{\mathcal{L}} N(\sqrt{1-\rho}\Phi^{-1}(p), \rho). \tag{2.15}$$

With the assumption of continuity on the mean regression function m, the limiting distribution of $Y_{[r:n]}$ is just the convolution of $m(X^*)$ and the standard normal distribution with $X^* \sim N(\sqrt{1-\rho}\Phi^{-1}(p), \rho)$, i.e.,

$$Y_{[r:n]} \xrightarrow{\mathcal{L}} m(X^*) + Z \tag{2.16}$$

where

$$X^* \sim N(\sqrt{1-\rho}\Phi^{-1}(p), \rho)$$

and

$$Z \sim N(0, 1).$$

The result given by (2.16) can be readily extended to derive the asymptotic distribution of $(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})$, where $r_i, i = 1, \ldots, k$, are constants such that as $n \to \infty$, $r_i/n \to p_i$ with $p_1 < \cdots < p_k$.

Theorem 2.3.1. *Suppose (X_i, Y_i), $i = 1, \ldots, n$, are a sequence of random vectors such that*

$$Y_i = m(X_i) + \varepsilon_i, i = 1, \ldots, n, \tag{2.17}$$

where m is some continuous regression function; ε_i's are i.i.d error terms with standard normal distribution, and are assumed to be independent of X_i's. Assume that the X_i's are distributed as equally correlated multivariate normal with zero mean vector and correlation ρ. Then for $r_i, i = 1, \ldots, k$, such that as $n \to \infty$, $r_i/n \to p_i$ with $0 < p_1 < \cdots < p_k < 1$, we have

$$(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})^T \xrightarrow{\mathcal{L}} (m(X_1^*) + Z_1, \ldots, m(X_k^*) + Z_k)^T \tag{2.18}$$

where $(X_1^*, \ldots, X_k^*)^T$ is a random vector defined to be:

$$(X_1^*, \ldots, X_k^*)^T := (\sqrt{1-\rho}\Phi^{-1}(p_1) + \sqrt{\rho}Z_0, \ldots, \sqrt{1-\rho}\Phi^{-1}(p_k) + \sqrt{\rho}Z_0)^T$$

with Z_0 being a standard normal variate; and Z_i's are i.i.d standard normal variates which are assumed to be independent of the X_i^*'s.

Proof. Note that by (2.1) we have

$$(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})^{\mathrm{T}} \stackrel{\mathrm{d}}{=} (m(X_{r_1:n}) + Z_1, \ldots, m(X_{r_k:n}) + Z_k)^{\mathrm{T}}. \tag{2.19}$$

While by (2.7), and the fact that

$$Z_{r_i:n} \xrightarrow{P} \Phi^{-1}(p_i), \qquad i = 1, \ldots, k, \tag{2.20}$$

we have

$$(X_{r_1:n}, \ldots, X_{r_k:n}) \xrightarrow{\mathcal{L}} (\sqrt{1-\rho}\Phi^{-1}(p_1) + \sqrt{\rho}Z_0, \ldots, \sqrt{1-\rho}\Phi^{-1}(p_k) + \sqrt{\rho}Z_0)^{\mathrm{T}}.$$

Since m is assumed to be continuous, by Ferguson (1996, Theorem 6, p39), the desired result follows readily. □

Example 2.3.1. (Simple Linear Regression, continued) With the simple linear regression model as given by (2.11), for any $r_i, i = 1, \ldots, k$, such that as $n \to \infty$, $r_i/n \to p_i$ with $0 < p_1 < \cdots < p_k < 1$, by Theorem 2.3.1 the limiting distribution for $(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})$ can be obtained. It follows that

$$(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})^{\mathrm{T}} \xrightarrow{\mathcal{L}} N_k(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*), \tag{2.21}$$

where

$$\boldsymbol{\mu}^* = \left(\beta\sqrt{1-\rho}\Phi^{-1}(p_1), \ldots, \beta\sqrt{1-\rho}\Phi^{-1}(p_k)\right)^{\mathrm{T}}$$

and

$$\boldsymbol{\Sigma}^* = \beta^2\rho\boldsymbol{J}_k + \boldsymbol{I}_k$$

with $\boldsymbol{J}_k$ being the $k \times k$ matrix of 1's, and $\boldsymbol{I}_k$ being the $k \times k$ identity matrix. We see that the limiting distribution of $(Y_{[r_1:n]}, \ldots, Y_{[r_k:n]})$ is also an equally correlated multivariate normal with common correlation $\beta^2\rho/(1 + \beta^2\rho)$. This dependence structure is the same as in a random effects model.

2.3.2 The Extremal Case

Suppose $r = n - k + 1$ for some fixed k as $n \to \infty$. By the well-known fact that

$$Z_{r:n} - \sqrt{2 \log n} \xrightarrow{P} 0$$

we have from (2.7) that

$$X_{r:n} - a_n \xrightarrow{\mathcal{L}} N(0, \rho) \tag{2.22}$$

with $a_n = \sqrt{2(1-\rho)\log n}$.

To derive the asymptotic distribution of $Y_{[r:n]}$, we first need to examine the limiting behavior of $m(X_{r:n})$, which in turn depends on the mean regression function m. But as demonstrated by the following Lemma, under appropriate assumptions on m, $m(X_{r:n})$ is asymptotically normal with appropriate normalization.

Lemma 2.3.1. *Suppose $X_{r:n}$ is the rth order statistic of a random vector, $(X_1, \ldots, X_n)$, from the equally correlated multivariate normal distribution as given by (2.2). Now assume that the mean regression function in (2.1) is differentiable up to lth order for some fixed $l > 1$, and $m^{(l)}(x)$ is continuous and bounded away from zero in the following sense*

$$0 < \liminf_{x \to \infty} m^{(l)}(x) \leq \limsup_{x \to \infty} m^{(l)}(x) < M$$

or

$$-M < \liminf_{x \to \infty} m^{(l)}(x) \leq \limsup_{x \to \infty} m^{(l)}(x) < 0$$

for some $M > 0$. Then as $n \to \infty$,

$$\frac{m(X_{r:n}) - m(a_n)}{m'(a_n)} \xrightarrow{\mathcal{L}} N(0, \rho) \tag{2.23}$$

where $r = n - k + 1$ for some fixed k as $n \to \infty$, and $a_n = \sqrt{2(1-\rho)\log n}$.

Proof. Note by Taylor series expansion

$$\begin{aligned} m(X_{r:n}) &= m(a_n) + m'(a_n)(X_{r:n} - a_n) + \cdots \\ &\quad + m^{(l-1)}(a_n)\frac{(X_{r:n} - a_n)^{l-1}}{(l-1)!} + m^{(l)}(X^*)\frac{(X_{r:n} - a_n)^l}{l!}, \end{aligned}$$

where X^* is somewhere between $X_{r:n}$ and a_n, and hence is a random quantity depending on $X_{r:n}$. It is obvious that X^* tends to infinity as $n \to \infty$ in probability.

So assuming $m'(a_n) \neq 0$, we have

$$\frac{m(X_{r:n}) - m(a_n)}{m'(a_n)} = (X_{r:n} - a_n) + \sum_{j=2}^{l-1} \frac{m^{(j)}(a_n)}{m'(a_n)} \frac{(X_{r:n} - a_n)^j}{j!} + \frac{m^{(l)}(X^*)}{m'(a_n)} \frac{(X_{r:n} - a_n)^l}{l!}. \tag{2.24}$$

By the fact that $m^{(l)}(x)$ is continuous and bounded, we have $m^{(l)}(X^*)$ is bounded in probability. Also from (2.22) we know that

$$X_{r:n} - a_n \xrightarrow{\mathcal{L}} N(0, \rho).$$

So it suffices to show that

$$\frac{m^{(j)}(x)}{m'(x)} \to 0, \quad \text{as } x \to \infty \tag{2.25}$$

for $j = 2, \ldots, l-1$, and

$$m'(x) \to \infty, \quad \text{as } x \to \infty. \tag{2.26}$$

If $0 < \liminf_{x\to\infty} m^{(l)}(x) \leq \limsup_{x\to\infty} m^{(l)}(x) < M$, then there exists some x_0 such that for any $x > x_0$, we have

$$C_1 < m^{(l)}(x) < C_2$$

where $0 < C_1 < C_2$ are some constants. So we have, for all $x > x_0$,

$$C_1(x - x_0) + m^{(l-1)}(x_0) < m^{(l-1)}(x) < C_2(x - x_0) + m^{(l-1)}(x_0)$$

which implies

$$m^{(l-1)}(x) = O(x),$$

for $x > x_0$. And for $k = 2, \ldots, l-1$, note that

$$m^{(l-k)}(x) = \sum_{i=0}^{k-1} \frac{(x-x_0)^i}{i!} m^{(l-k+i)}(x_0) + \frac{(x-x_0)^k}{k!} m^{(l)}(x^*) \tag{2.27}$$

with x^* be some value between x and x_0. Thus it follows that $m^{(l-k)}(x) = O(x^k)$ for $x > x_0$, $k = 2, \ldots, l-1$. So (2.25) and (2.26) easily follow. Similar arguments can be used for the case $-M < \liminf_{x\to\infty} m^{(l)}(x) \le \limsup_{x\to\infty} m^{(l)}(x) < 0$. So the Lemma is proved. □

Remark 2.3.1. The assumptions we made about the mean regression function $m(\cdot)$ will hold for the polynomials. In particular, if $m(\cdot)$ is a polynomial of degree s, then the assumptions will hold with $l = s$.

Then the limiting distribution of $Y_{[r:n]}$ for the extremal case can be derived as the following Theorem.

Theorem 2.3.2. *With the assumptions on $m(\cdot)$ as specified in Lemma 2.3.1, we have*

$$\frac{Y_{[r:n]} - m(a_n)}{m'(a_n)} \xrightarrow{\mathcal{L}} N(0, \rho) \tag{2.28}$$

as $n \to \infty$ with $r = n - k + 1$ for some fixed k, and $a_n = \sqrt{2(1-\rho)\log n}$.

Proof. By Lemma 2.3.1 we have

$$\frac{m(X_{r:n}) - m(a_n)}{m'(a_n)} \xrightarrow{\mathcal{L}} N(0, \rho) \tag{2.29}$$

and $m'(x) \to \infty$ as $x \to \infty$ under the assumptions on $m(\cdot)$. Then the desired results easily follow from (2.3). □

Remark 2.3.2. Similar result will hold for the concomitant of upper intermediate order statistic, i.e., $Y_{[r:n]}$, with $r = n - k + 1$, where $k \to \infty$, and $k/n \to 0$ as $n \to \infty$. By the following fact about upper intermediate order statistics (see, for example, David and Nagaraja, 2003, p312)

$$Z_{n-k+1:n} - \Phi^{-1}\left(\frac{n-k}{n}\right) \xrightarrow{P} 0, \tag{2.30}$$

we see that (2.28) holds with $a_n = \Phi^{-1}\left(\frac{n-k}{n}\right)$.

Example 2.3.2. (Simple Linear Regression, continued) With the simple linear regression model as given by (2.11), the limiting distribution of $Y_{[n-k+1:n]}$, with k being some fixed integer, is normal and

$$Y_{[r:n]} - \beta a_n \xrightarrow{\mathcal{L}} N(0, \beta^2\rho) \tag{2.31}$$

with $a_n = \sqrt{2(1-\rho)\log n}$.

CHAPTER 3

DISTRIBUTION OF A SINGLE ORDER STATISTIC OF SUBSETS OF CONCOMITANTS OF ORDER STATISTICS

3.1 Introduction

In this chapter we will study the distributions of $V_{s:m}$ and $W_{t:n-m}$ respectively, where $V_{s:m}$ is the sth order statistic of the subset $\{Y_{[i:n]}, i = n-m+1, \ldots, n\}$, and $W_{t:n-m}$ is the tth order statistic of the subset $\{Y_{[j:n]}, j = 1, \ldots, n-m\}$ of concomitants.

Nagaraja and David (1994) discussed the finite-sample and asymptotic distributions of $V_{s:m}$ with $s = m$, i.e., the maxima of the corresponding subset of concomitants. They showed that the limiting distribution of $V_{m:m}$ does not depend on the correlation between X and Y. Here we obtain the corresponding results for general s and t.

Related to this work is the study by Chu et al. (1999), in which the authors investigated the asymptotic distributions of $W_{r_n,s_n,n} = \min(Y_{[r_n:n]}, Y_{[r_n+1:n]}, \ldots, Y_{[s_n:n]})$, $V_{r_n,s_n,n} = \max(Y_{[r_n:n]}, \ldots, Y_{[s_n:n]})$, as well as some other functionals of $W_{r_n,s_n,n}$ and $V_{r_n,s_n,n}$, where $\{r_n\}$ and $\{s_n\}$ are two integer sequences such that $0 \leq r_n \leq s_n \leq n$, and $r_n/n \to p$ for $0 < p < 1$ as $n \to \infty$. The results were applied to study the problem of locating the maximum of a nonparametric regression function as discussed in Chen et al. (1996).

In Section 3.2, the finite sample distributions of $V_{s:m}$ and $W_{t:n-m}$ are derived using conditioning argument. In Section 3.3 the asymptotic distribution of $V_{s:m}$ for the quantile case, i.e. the situation in which $m = [np_0]$, $s = [mp_1]$, for $0 < p < 1$, and $n \to \infty$, is obtained under appropriate regularity conditions. We also establish the rate of convergence in the distribution of $V_{s:m}$ there, and propose a second order approximation to the distribution of $V_{s:m}$. In Section 3.4 similar results for the asymptotic distribution of $W_{t:n-m}$ are derived for the quantile case. For the extremal case where s and t are kept fixed while n and m approach infinity, we defer the discussion until Chapter 4, and we will derive the asymptotic joint distribution of $(V_{s:m}, W_{t:n-m})$ there.

3.2 Finite-Sample Distributions of $V_{s:m}$ and $W_{t:n-m}$

To derive the finite-sample distributions of $V_{s:m}$ and $W_{t:n-m}$ we need the following lemma:

Lemma 3.2.1. *Given $X_{n-m:n} = x$, $V_{s:m}$ behaves like the sth order statistic of a random sample of size m from the cdf $F_1(\cdot|x)$; given $X_{n-m+1:n} = x$; $W_{t:n-m}$ behaves like the tth order statistic of a random sample of size $n-m$ from the cdf $F_2(\cdot|x)$.*

Lemma 3.2.1 can be proved using Theorem 2 in Kaufmann and Reiss (1992). Here we give another proof using the Markovian property of order statistics.

Proof. First notice that the conditional joint density for $(Y_{[n-m+1:n]}, \ldots, Y_{[n:n]})$ given $X_{n-m:n} = x$ is given by:

$$\begin{aligned} & f_{Y_{[n-m+1:n]},\ldots,Y_{[n:n]}|X_{n-m:n}=x}(y_1, \ldots, y_m) \\ = & \int \ldots \int f_{Y_{[n-m+1:n]},\ldots,Y_{[n:n]}|X_{n-m:n}=x,X_{n-m+1:n}=v_1,\ldots,X_{n:n}=v_m}(y_1, \ldots, y_m) \\ & \qquad \times f_{X_{n-m+1:n},\ldots,X_{n:n}|X_{n-m:n}=x}(v_1, \ldots, v_m) dv_1 \ldots dv_m \\ = & \int \ldots \int \prod_{i=1}^{m} f_{Y|X=v_i}(y_i) f_{X_{n-m+1:n},\ldots,X_{n:n}|X_{n-m:n}=x}(v_1, \ldots, v_m) dv_1 \ldots dv_m. \end{aligned}$$

Next recall that given $X_{n-m:n} = x$, $(X_{n-m+1:n}, \ldots, X_{n:n})$ behave like the order statistics of a random sample of size m from the cdf

$$G(t) = \frac{F_X(t) - F_X(x)}{1 - F_X(x)} I_{\{t>x\}}$$

(David and Nagaraja, 2003, p17). Hence given $X_{n-m:n} = x$, $Y_{[n-m+i:n]}$ $(i = 1, \ldots, m)$ has the same distribution as the concomitant of ith order statistic of a random sample of size m from the bivariate distribution (X^*, Y^*) with joint cdf $G(x, y)$, such that

$$G_{X^*}(t) = \frac{F_X(t) - F_X(x)}{1 - F_X(x)} I_{\{t>x\}}; \qquad G_{Y^*|X^*}(y) = F_{Y|X}(y) \tag{3.1}$$

Also notice that the mapping associated with taking order statistics is invariant to the mapping associated with taking concomitants, i.e., if we define

$$\phi((x_1, y_1), \ldots, (x_m, y_m)) \equiv y_{i:m}$$

and

$$\psi((x_1, y_1), \ldots, (x_m, y_m)) \equiv ((x_{1:m}, y_{[1:m]}), \ldots, (x_{m:m}, y_{[m:m]})),$$

then

$$\phi \circ \psi((x_1, y_1), \ldots, (x_m, y_m)) = \phi((x_1, y_1), \ldots, (x_m, y_m)).$$

So it follows that given $X_{n-m:n} = x$, $V_{s:m}$, which is the sth order statistic of $(Y_{[n-m+1:n]}, \ldots, Y_{[n:n]})$, behaves like the sth order statistic of a random sample of size m from the marginal distribution of Y^* derived from $G(x, y)$ of (3.1), which is just $F_1(\cdot|x)$. Similar argument can be used to derive the result on $W_{t:n-m}$. □

By Lemma 3.2.1, we can derive the cdf, $F_{V_{s:m}}(v)$, of $V_{s:m}$ as follows:

$$F_{V_{s:m}}(v) = \int \mathbf{P}(V_{s:m} \leq v | X_{n-m:n} = x) dF_{X_{n-m:n}}(x) \tag{3.2}$$

where

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

and $F_{X_{n-m:n}}(\cdot)$ is the cdf of the order statistic $X_{n-m:n}$.

Similarly the cdf of $W_{t:n-m}$ can be derived as:

$$F_{W_{t:n-m}}(w) = \int \mathbf{P}(W_{t:n-m} \leq w | X_{n-m+1:n} = x) dF_{X_{n-m+1:n}}(x) \tag{3.3}$$

where

$$\mathbf{P}(W_{t:n-m} \leq w | X_{n-m+1:n} = x) = \sum_{i=t}^{n-m} \binom{n-m}{i} [F_2(w|x)]^i [1 - F_2(w|x)]^{n-m-i}$$

and $F_{X_{n-m+1:n}}(\cdot)$ is the cdf of the order statistic $X_{n-m+1:n}$.

3.3 Asymptotic Distribution of $V_{s:m}$ in the Quantile Case

In this section we derive the asymptotic distribution of $V_{s:m}$ for the *quantile case* in which $s = [mp_1]$, $m = [np_0]$ for some p_0 and p_1 such that $0 < p_i < 1, i = 0, 1$, as $n \to \infty$. The rate of convergence in the distribution of $V_{s:m}$ is also established, and we propose an improved approximation to the cdf of $V_{s:m}$ that works well even for moderate sample sizes.

3.3.1 The Main Result

The main result regarding the asymptotic distribution of $V_{s:m}$ in the *quantile case* is given by the following Theorem.

Theorem 3.3.1. *Let* $(X_i, Y_i), i = 1, 2, \ldots, n,$ *be a random sample from the absolutely continuous bivariate distribution* $F(x, y)$*, and we assume the joint density* $f(x, y)$ *is continuous in both arguments. Suppose* $m = [np_0]$ *and* $s = [mp_1]$, $0 < p_i < 1, i = 0, 1,$ *as* $n \to \infty$*. Let*

$$x_0 = F_X^{-1}(q_0), \tag{3.4}$$

with $q_0 = 1 - p_0$*, and*

$$a = F_1^{-1}(p_1|x_0), \tag{3.5}$$

and we assume that $f(x_0, a) > 0$*. Then we have*

$$\frac{V_{s:m} - a}{b_n} \xrightarrow{\mathcal{L}} Z_1 + g_1(x_0)Z_2 \tag{3.6}$$

where

$$b_n = \left[\frac{\sqrt{np_0} f_1(a|x_0)}{\sqrt{p_1 q_1}}\right]^{-1}, \tag{3.7}$$

Z_1 *and* Z_2 *are independent standard normal variables, and*

$$g_1(x_0) = \frac{\sqrt{p_0 q_0}}{f_X(x_0)} \left[\frac{\sqrt{p_0} f_1(a|x_0)}{\sqrt{p_1 q_1}}\right] \frac{\partial F_1^{-1}(p_1|x)}{\partial x}|_{x=x_0} \tag{3.8}$$

To prove Theorem 3.3.1 we need the following lemma.

Lemma 3.3.1. *Let* $\{F_n\}_{n=1}^{\infty}$ *be a family of continuous cdfs. For each* n*, let* $X_{s:n}$ *be the* s*th order statistic of a random sample of size* n *from the distribution* F_n*. Suppose* $s = [np_0]$

and $n \to \infty$, and $f_n(\xi_n(p_0)) > 0$ for each n, where $\xi_n(p_0)$ is the p_0-th quantile of the distribution F_n. Then we have

$$\frac{\sqrt{n}(X_{s:n} - \xi_n(p_0))f_n(\xi_n(p_0))}{\sqrt{p_0 q_0}} \xrightarrow{\mathcal{L}} N(0,1).$$

Proof. By the representation of quantile as in Ghosh (1971, Theorem 1), we have:

$$X_{s:n} = \xi_n(p_0) - \frac{\hat{F}_n(\xi_n(p_0)) - p_0}{f_n(\xi_n(p_0))} + R_n$$

where $\hat{F}_n(\xi_n(p_0)) = \frac{1}{n}\sum_{k=1}^{n} I_{\{X_k \le \xi_n(p_0)\}}$ is the empirical cdf associated with the random sample of size n from the distribution F_n, and R_n is the remainder such that $R_n = o_p(1/\sqrt{n})$.

So we have

$$\sqrt{n}(X_{s:n} - \xi_n(p_0))f_n(\xi_n(p_0)) = -\sqrt{n}(\hat{F}_n(\xi_n(p_0)) - p_0) + o_p(1).$$

And by Lindeberg's double array central limit theorem we have

$$\sqrt{n}(\hat{F}_n(\xi_n(p_0)) - p_0) \xrightarrow{\mathcal{L}} N(0, p_0 q_0).$$

Thus the desired result follows. □

Proof of Theorem 3.3.1

.

- By the continuity of $f(x,y)$, it can be proved that $F_1^{-1}(p_1|x)$ is differentiable with respect to x at x_0 as follows:

Let $h(x) = F_1^{-1}(p_1|x)$. This means

$$\frac{\int_{-\infty}^{h(x)} \int_x^{\infty} f(u,v)dudv}{1 - F_X(x)} = p_1,$$

or

$$\int_{-\infty}^{h(x)} \int_x^{\infty} f(u,v)dudv = p_1(1 - F_X(x)).$$

So we have

$$\frac{d}{dx} \int_{-\infty}^{h(x)} \int_x^{\infty} f(u,v)dudv = -p_1 f_X(x). \tag{3.9}$$

Notice that

$$\begin{aligned}
&\frac{d}{dx} \int_{-\infty}^{h(x)} \int_x^{\infty} f(u,v)dudv \\
=\;& \lim_{\Delta \to 0} \frac{\int_{-\infty}^{h(x+\Delta)} \int_{x+\Delta}^{\infty} f(u,v)dudv - \int_{-\infty}^{h(x)} \int_x^{\infty} f(u,v)dudv}{\Delta} \\
=\;& \lim_{\Delta \to 0} \frac{\int_{h(x)}^{h(x+\Delta)} \int_{x+\Delta}^{\infty} f(u,v)dudv - \int_{-\infty}^{h(x)} \int_x^{x+\Delta} f(u,v)dudv}{\Delta} \\
=\;& \lim_{\Delta \to 0} \frac{\int_{h(x)}^{h(x+\Delta)} \int_x^{\infty} f(u,v)dudv - \int_{h(x)}^{h(x+\Delta)} \int_x^{x+\Delta} f(u,v)dudv}{\Delta} \\
&- \lim_{\Delta \to 0} \int_{-\infty}^{h(x)} f(u^*,v)dv, \quad \text{where } u^* \in (x, x+\Delta) \\
=\;& \int_x^{\infty} \lim_{\Delta \to 0} \frac{\int_{h(x)}^{h(x+\Delta)} f(u,v)dv}{\Delta} du - \int_{-\infty}^{h(x)} f(x,v)dv \\
=\;& \int_x^{\infty} \lim_{\Delta \to 0} \frac{f(u,v^*) \times [h(x+\Delta) - h(x)]}{\Delta} du - \int_{-\infty}^{h(x)} f(x,v)dv \\
&\quad \text{where } v^* \in (h(x), h(x+\Delta)) \\
=\;& \left(\lim_{\Delta \to 0} \frac{h(x+\Delta) - h(x)}{\Delta}\right) \times \int_x^{\infty} f(u,h(x))du - \int_{-\infty}^{h(x)} f(x,v)dv.
\end{aligned}$$

So from (3.9)

$$h'(x) = \frac{\int_{-\infty}^{h(x)} f(x,v)dv - p_1 f_X(x)}{\int_x^{\infty} f(u,h(x))du}. \tag{3.10}$$

- Let $V_n = (V_{s:m} - a)/b_n$,

$$c_0 = \sqrt{p_0 q_0}/f_X(x_0); \qquad Z_n = \sqrt{n} f_X(x_0)(X_{n-m:n} - x_0)/\sqrt{p_0 q_0}; \tag{3.11}$$

and

$$a_n(z) = F_1^{-1}(p_1 | x_0 + c_0 z/\sqrt{n}). \tag{3.12}$$

Notice

$$\begin{aligned}
&P(V_n \leq v) \\
&=\mathbf{E}\left[P(V_n \leq v | Z_n)\right] \\
&=\int P(V_n \leq v | Z_n = z) dF_{Z_n}(z) \\
&=\int P(V_n \leq v | X_{n-m:n} = x_0 + c_0 z/\sqrt{n}) dF_{Z_n}(z) \\
&=\int P(\frac{V_{s:m} - a_n(z)}{b_n} + \frac{a_n(z) - a}{b_n} \leq v | X_{n-m:n} = x_0 + c_0 z/\sqrt{n}) dF_{Z_n}(z).
\end{aligned} \tag{3.13}$$

- Note the following facts:

  (i) Z_n defined in (3.11) converges in distribution to $N(0,1)$. This follows from the well-known fact about the limiting distribution of a central order statistic.

  (ii) $\forall \ z \in \mathbb{R}$, given $X_{n-m:n} = x_0 + c_0 z/\sqrt{n}$

$$\frac{V_{s:m} - a_n(z)}{b_n} \xrightarrow{\mathcal{L}} N(0,1).$$

  This follows from Lemma 1 and the fact that conditional on $X_{n-m:n} = x_0 + c_0 z/\sqrt{n}$, $V_{s:m}$ behaves the same as the sth order statistic of the random sample of size m from the distribution $F_1(\cdot | x_0 + c_0 z/\sqrt{n})$.

(iii)

$$\frac{a_n(z)-a}{b_n} = \frac{F_1^{-1}(p_1|x_0+c_0z/\sqrt{n}) - F_1^{-1}(p_1|x_0)}{\left[\frac{\sqrt{np_0}f_1(a|x_0)}{\sqrt{p_1q_1}}\right]^{-1}}$$
$$= h'(u^*)c_0z/\sqrt{n}\left[\frac{\sqrt{np_0}f_1(a|x_0)}{\sqrt{p_1q_1}}\right]$$
$$\to g_1(x_0)z$$

as $n \to \infty$, where h' is the partial derivative of $F_1^{-1}(p_1|x)$ w.r.t. x, $u^* \in (x_0, x_0+c_0z/\sqrt{n})$, and g_1 is given by (3.8).

Upon applying the Bounded Convergence Theorem to the expression (3.13), we conclude that

$$\mathbf{P}(V_n \le v) \to \int \Phi(v - g_1(x_0)z)d\Phi(z), \qquad \text{as } n \to \infty.$$

So (3.6) easily follows from the convolution formula.

□

Remark 3.3.1. Notice that (3.6) is equivalent to

$$\sqrt{n}(V_{s:m}-a) \xrightarrow{\mathcal{L}} bZ_1 + h_1(x_0)Z_2 \tag{3.14}$$

where

$$a = F_1^{-1}(p_1|x_0); \quad b = \frac{\sqrt{p_1q_1}}{\sqrt{p_0}f_1(a|x_0)}, \tag{3.15}$$

Z_1 and Z_2 are independent standard normal variables, and

$$\begin{aligned} h_1(x_0) &= \frac{\sqrt{p_0q_0}}{f_X(x_0)} \cdot \frac{\partial F_1^{-1}(p_1|x)}{\partial x}\Big|_{x=x_0} && (3.16)\\ &= \frac{\sqrt{p_0q_0}}{f_X(x_0)} \cdot \frac{\int_{-\infty}^{a} f(x_0,v)dv - p_1f_X(x_0)}{\int_{x_0}^{\infty} f(u,a)du} && (3.17)\\ &= \frac{\sqrt{q_0}(F_3(a|x_0)-F_1(a|x_0))}{\sqrt{p_0}f_1(a|x_0)}. && (3.18)\end{aligned}$$

Example 3.3.1. Suppose we are sampling from a standard bivariate normal distribution given by

$$(X,Y) \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right). \tag{3.19}$$

Note that

$$f_1(y|x_0) = \left[1 - \Phi\left(\frac{x_0 - \rho y}{\sqrt{1-\rho^2}}\right)\right]\phi(y)/p_0,$$

and $a = F_1^{-1}(p_1|x_0)$ satisfies the following equation:

$$\int_{-\infty}^{a} f_1(y|x_0)dy - p_1 p_0 = 0 \tag{3.20}$$

or is the unique solution to

$$\Phi(a) - \int_{-\infty}^{a} \Phi\left(\frac{x_0 - \rho y}{\sqrt{1-\rho^2}}\right)\phi(y)dy = p_0 p_1$$

which can be solved numerically using Newton-Raphson method.

Further b in (3.15) is given by

$$b = \frac{\sqrt{p_1 q_1}}{\sqrt{p_0} f_1(a|x_0)} = \frac{\sqrt{p_1 q_1 p_0}}{\left[1 - \Phi\left(\frac{x_0 - \rho a}{\sqrt{1-\rho^2}}\right)\right]\phi(a)}. \tag{3.21}$$

From (3.10) we know that:

$$\begin{aligned}
\frac{\partial F_1^{-1}(p_1|x)}{\partial x}\Big|_{x=x_0} &= \frac{\int_{-\infty}^{a} f(x_0, v)dv - p_1 f_X(x_0)}{\int_{x_0}^{\infty} f(u,a)du} \\
&= \frac{\Phi\left(\frac{a-\rho x_0}{\sqrt{1-\rho^2}}\right)\phi(x_0) - p_1\phi(x_0)}{\left[1 - \Phi\left(\frac{x_0-\rho a}{\sqrt{1-\rho^2}}\right)\right]\phi(a)}.
\end{aligned}$$

So we have

$$h_1(x_0) = \frac{\sqrt{p_0q_0}}{f_X(x_0)} \cdot \frac{\partial F_1^{-1}(p_1|x)}{\partial x}|_{x=x_0} \tag{3.22}$$

$$= \frac{\sqrt{p_0q_0}}{\phi(x_0)} \cdot \frac{\Phi\left(\frac{a-\rho x_0}{\sqrt{1-\rho^2}}\right)\phi(x_0) - p_1\phi(x_0)}{\left[1-\Phi\left(\frac{x_0-\rho a}{\sqrt{1-\rho^2}}\right)\right]\phi(a)} \tag{3.23}$$

$$= \frac{\sqrt{p_0q_0}\left(\Phi\left(\frac{a-\rho x_0}{\sqrt{1-\rho^2}}\right) - p_1\right)}{\left[1-\Phi\left(\frac{x_0-\rho a}{\sqrt{1-\rho^2}}\right)\right]\phi(a)}. \tag{3.24}$$

3.3.2 Rate of Convergence

We just showed that after appropriate normalization, $V_{s:m}$ will converge in law to the normal distribution for the quantile case. Next we will explore its rate of convergence.

First we observe the following Lemma regarding the rate of convergence in the expectation of a bounded continuous function of an order statistic in the quantile case.

Lemma 3.3.2. *Suppose $X_{r:n}$ is the rth order statistic of a random sample of size n from a distribution with cdf and pdf, $F(x)$ and $f(x)$ respectively, and assume $r/n - p = O(n^{-1})$ as $n \to \infty$. Let F have $m+1$ bounded derivatives in a neighborhood of $\xi_p = F^{-1}(p)$ for $m > 1$ and $f(\xi_p) > 0$. Define*

$$Z_n = \frac{\sqrt{n}f(\xi_p)}{\sigma_p}(X_{r:n} - \xi_p) \tag{3.25}$$

with $\sigma_p = \sqrt{p(1-p)}$. Then for any bounded and continuous function H defined on the real line, we have

$$\mathbf{E}(H(Z_n)) \to \mathbf{E}(H(Z)) \tag{3.26}$$

and

$$|\mathbf{E}(H(Z_n)) - \mathbf{E}(H(Z))| = O(n^{-1/2}), \tag{3.27}$$

where Z is a standard normal random variable.

Proof. The convergence of the expectation given by (3.26) will follow from the well-known fact that $Z_n \xrightarrow{\mathcal{L}} N(0,1)$, and the assumption that H is a bounded and continuous function.

To prove (3.27), let f_n be the pdf of Z_n. With the assumption that F have $m+1$ bounded derivatives in a neighborhood of $\xi_p = F^{-1}(p)$ for $m > 1$ and $f(\xi_p) > 0$, according to a result in Reiss (1989, p147-148), we have

$$\sup_{B \in \mathcal{B}} \left| \mathbf{P}(Z_n \in B) - \int_B dG_n(z) \right| \le C_m n^{-m/2} \tag{3.28}$$

for some constant C_m (not depending on n), where $\mathcal{B}$ is the Borel field on $\mathbb{R}$, and

$$G_n(z) = \Phi(z) + \phi(z) \sum_{i=1}^{m-1} n^{-i/2} S_{i,n}(z) \tag{3.29}$$

with $S_{i,n}(z)$ being a polynomial of degree less than or equal to $3i-1$, and having coefficients uniformly bounded over n. Moreover,

$$|f_n(z) - g_n(z)| \le D_m n^{-m/2} \phi(z)(1 + |z|^{3m}) \tag{3.30}$$

for all $z \in [-\log n, \log n]$, where D_m is some constant not depending on n, and $g_n(z) = G_n'(z)$.

Notice

$$\mathbf{E}(H(Z_n)) = \int_{-\log n}^{\log n} H(z) f_n(z) dz + \mathbf{E}(H(Z_n) I(|Z_n| > \log n)), \tag{3.31}$$

and by (3.30) we have

$$\int_{-\log n}^{\log n} H(z) g_n(z) dz - O(n^{-m/2}) \le \int_{-\log n}^{\log n} H(z) f_n(z) dz \le \int_{-\log n}^{\log n} H(z) g_n(z) dz + O(n^{-m/2}). \tag{3.32}$$

It is easy to verify that

$$\int_{-\log n}^{\log n} H(z)g_n(z)dz = \int_{-\log n}^{\log n} H(z)\phi(z)dz + O(n^{-1/2}). \tag{3.33}$$

Plugging (3.32) and (3.33) into (3.31) we have:

$$\begin{aligned} & \mathbf{E}(H(Z)) - \mathbf{E}(H(Z)I(|Z| > \log n)) + \mathbf{E}(H(Z_n)I(|Z_n| > \log n)) + O(n^{-1/2}) \\ \leq\ & \mathbf{E}(H(Z_n)) \\ \leq\ & \mathbf{E}(H(Z)) - \mathbf{E}(H(Z)I(|Z| > \log n)) + \mathbf{E}(H(Z_n)I(|Z_n| > \log n)) + O(n^{-1/2}). \end{aligned} \tag{3.34}$$

And by (3.28) and the fact that H is bounded we have

$$\mathbf{E}(H(Z_n)I(|Z_n| > \log n)) - \mathbf{E}(H(Z)I(|Z| > \log n)) = O(n^{-1/2}) \tag{3.35}$$

which finishes the proof. □

Remark 3.3.2. The assumption that F have $m+1$ bounded derivatives in a neighborhood of $\xi_p = F^{-1}(p)$ for $m > 1$ is needed for establishing (3.28) and (3.30). For our purpose this assumption only needs to hold for $m = 2$.

Now we have the following result regarding the rate of convergence for the asymptotic distribution of $V_{s:m}$.

Theorem 3.3.2. *Along with the assumptions of Theorem 3.3.1 and Lemma 3.3.2, we assume that $F_1^{-1}(p_1|x)$, as a function of x, has second order derivative and the derivative is bounded in a neighborhood of $x_0 = F_X^{-1}(p_0)$. Then we have*

$$\sup_v \left| \mathbf{P}\left(\frac{V_{s:m} - a}{b_n} \leq v \right) - \Phi\left(\frac{v}{\sqrt{1 + g_1^2(x_0)}} \right) \right| \leq O(n^{-1/2}). \tag{3.36}$$

Proof. As in the proof of Theorem 3.3.1, let V_n and Z_n be the normalized $V_{s:m}$ and $X_{n-m:n}$, respectively; that is,

$$V_n = \frac{V_{s:m} - a}{b_n}; \quad \text{and} \quad Z_n = \frac{\sqrt{n} f_X(x_0)}{\sigma_{p_0}}(X_{n-m:n} - x_0) \tag{3.37}$$

with $x_0 = F_X^{-1}(1-p_0)$ and $\sigma_{p_0} = \sqrt{p_0(1-p_0)}$.

From the proof of Theorem 3.3.1, we know that

$$\mathbf{P}(V_n \le v) = \mathbf{E}\left[\mathbf{P}\left(\frac{V_{s:m} - a_n(Z_n)}{bn} + g_n(Z_n) \le v | Z_n\right)\right] \tag{3.38}$$

where $a_n(z)$ is given in (3.12), and $g_n(z) = (a_n(z) - a)/b_n$ approaches $g_1(x_0)z$ as $n \to \infty$. Also notice that

$$\begin{aligned} g_n(z) &= \frac{F_1^{-1}(p_1|x_0 + c_0 z/\sqrt{n}) - F_1^{-1}(p_1|x_0)}{b_n} \\ &= \frac{\sqrt{np_0} f_1(a|x_0)}{\sqrt{p_1 q_1}}\left[\frac{c_0 z}{\sqrt{n}}\frac{\partial}{\partial x} F_1^{-1}(p_1|x)|_{x=x_0} + \frac{1}{2}\left(\frac{c_0 z}{\sqrt{n}}\right)^2 \frac{\partial^2}{\partial x^2} F_1^{-1}(p_1|x)|_{x=u^*}\right] \\ &= g_1(x_0)z + \frac{z^2}{\sqrt{n}} M(z), \end{aligned} \tag{3.39}$$

where u^* is some value within $(x_0, x_0 + c_0 z/\sqrt{n})$, $M(z)$ is some constant free of n. This is because of the assumption that $F_1^{-1}(p_1|x)$, as a function of x, has second order derivative and the derivative is bounded in a neighborhood of $x_0 = F_X^{-1}(p_0)$. So for any $z \in \mathbb{R}$, $g_n(z)$ converges to $g_1(x_0)z$ at a rate of order $n^{-1/2}$.

By (3.28) we have for any $z \in \mathbb{R}$

$$\sup_v \left|\mathbf{P}\left(\frac{V_{s:m} - a_n(z)}{bn} + g_n(z) \le v | Z_n = z\right) - \Phi(v - g_n(z))\right| \le O(n^{-1/2}). \tag{3.40}$$

So

$$\sup_v |\mathbf{P}(V_n \le v) - E[\Phi(v - g_n(Z_n))]| \le O(n^{-1/2}). \tag{3.41}$$

While

$$\begin{aligned} &|\mathbf{E}\{\Phi[v - g_n(Z_n)]\} - \mathbf{E}\{\Phi[v - g_1(x_0)Z]\}| \\ &\le |\mathbf{E}\{\Phi[v - g_n(Z_n)]\} - \mathbf{E}\{\Phi[v - g_1(x_0)Z_n]\}| \\ &\quad + |\mathbf{E}\{\Phi[v - g_1(x_0)Z_n]\} - \mathbf{E}\{\Phi[v - g_1(x_0)Z]\}|, \end{aligned} \tag{3.42}$$

by (3.39) we have

$$|E[\Phi(v - g_n(Z_n))] - E[\Phi(v - g_1(x_0)Z_n)]| \leq O(n^{-1/2}), \tag{3.43}$$

and (3.27) of Lemma 3.3.2 implies that

$$|E[\Phi(v - g_1(x_0)Z_n)] - E[\Phi(v - g_1(x_0)Z)]| \leq O(n^{-1/2}). \tag{3.44}$$

So we have

$$|E[\Phi(v - g_n(Z_n))] - E[\Phi(v - g_1(x_0)Z)]| \leq O(n^{-1/2}) \tag{3.45}$$

and the desired result follows. □

Example 3.3.2. Suppose the population distribution is bivariate standard normal with correlation ρ. From Theorem 3.3.1 we know that for sufficiently large n, $V_{s:m}$ is approximately

$$N\left(a, (b^2 + h_1^2(x_0))/n\right)$$

with a, b and $h_1(x_0)$ given by (3.20), (3.21) and (3.24), respectively.

Figure 3.1 shows the histograms for the 1000 simulated sample values of $V_{s:m}$ and the density curves of the corresponding normal distributions for different values of ρ, with $n = 400$, $p_0 = 0.2$, and $p_1 = 0.3$, that is, $m = 100$, $s = 30$. From the plots we observe that the histograms of $V_{s:m}$ samples come very close to the corresponding normal density curves.

Next we study numerically the accuracy of the normal approximation to the distribution of $V_{s:m}$. We first estimate the cdf of the normalized $V_{s:m}$, namely $\sqrt{n}(V_{s:m} - a)$, by Monte Carlo simulations with 160000 trials. Then we compute the difference between the estimated cdf and the cdf provided by the normal approximation $N\left(0, (b^2 + h_1^2(x_0))/n\right)$

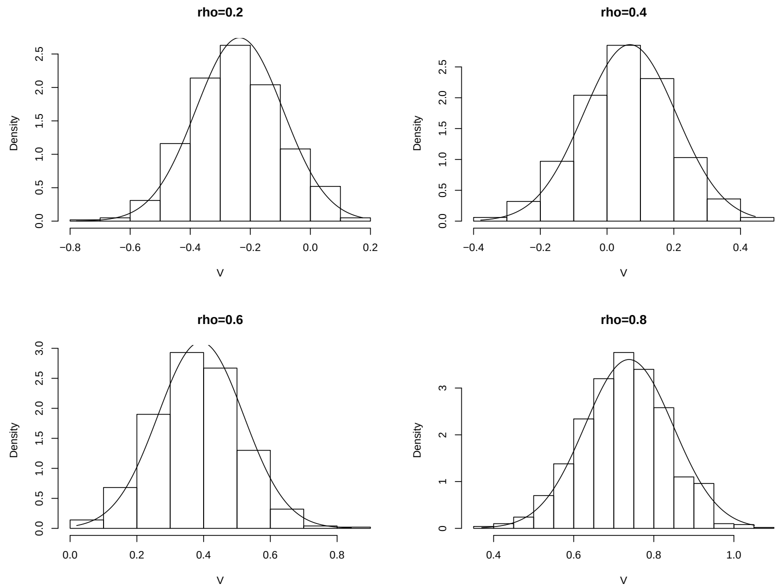

Figure 3.1: Histograms of Simulated $V_{s:m}$ vs. the Corresponding Normal Density Curves for $\rho = 0.2, 0.4, 0.6, 0.8$.

for range of values within three standard deviations around the mean. We carry out these calculations for sample sizes $n = 100, 400, 900, 1600, 2500$ and 10000, and we set $\rho = 0.4$, $p_0 = 0.5$, and $p_1 = 0.3$. Figure 3.2 gives the estimated true cdf for $V_{s:m}$ along with the cdf by normal approximation for different sample sizes. From Figure 3.2 we observe that the normal approximation always underestimates the true cdf of $V_{s:m}$ in our sample cases. But this is not true in general. Limited simulation studies show that in some sample cases, the normal approximation overestimate the true cdf of $V_{s:m}$. Figure 3.3 gives the plots of the difference between the simulated cdf and the approximation provided by the associated normal cdf vs. different values of $V_{s:m}$ for different sample sizes. We observe that the normal approximation achieves better performance at the two tails than in the center of the distribution, and overall the approximation is fairly good for a sample of size 400.

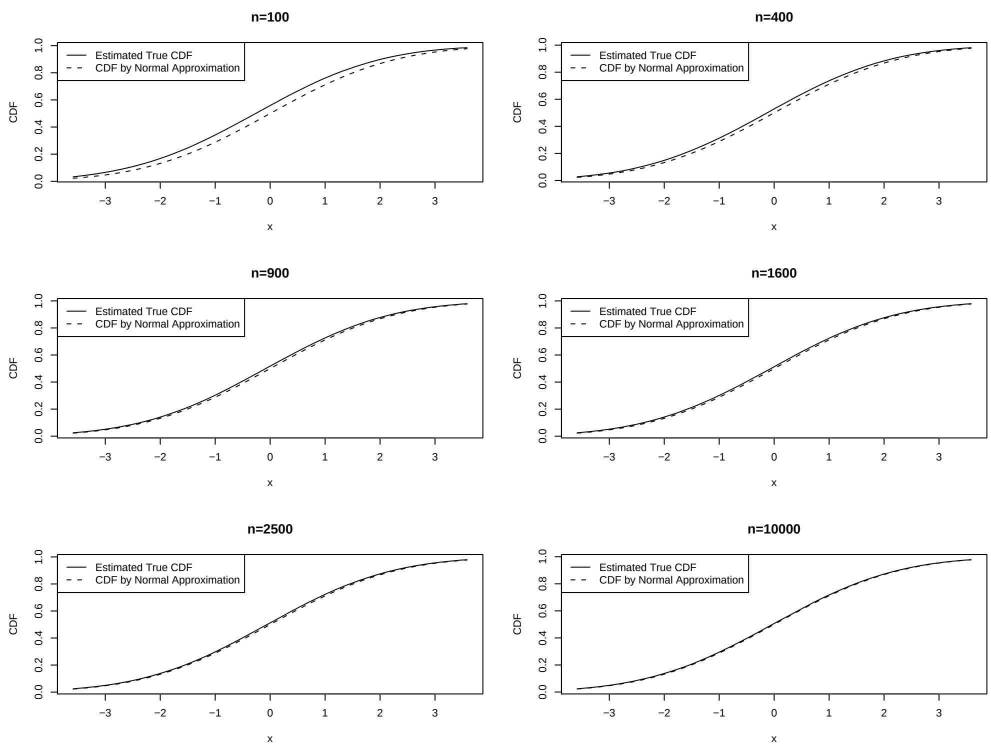

Figure 3.2: Estimated True cdf vs. the cdf by the Normal Approximation of $V_{s:m}$ for the Standard Bivariate Normal Population when $\rho = 0.4$, $p_0 = 0.5$ and $p_1 = 0.3$ ($m = 0.5n$, $s = 0.15n$).

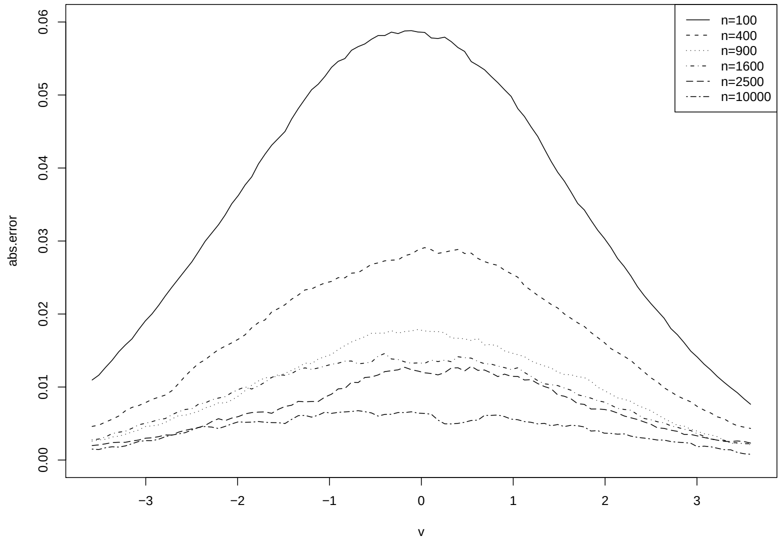

Figure 3.3: Absolute Errors of the Normal Approximation to the Distribution of $V_{s:m}$ for the Standard Bivariate Normal Population when $\rho = 0.4$, $p_0 = 0.5$ and $p_1 = 0.3$ ($m = 0.5n$, $s = 0.15n$).

To examine the rate of convergence for the distribution of $V_{s:m}$, we plot the negative logarithm of the maximum absolute errors for the normal approximation over the range of evaluated values vs. the logarithm of the corresponding sample sizes. Note that the slope of the fitted regression line in this plot will give us an idea about the rate of convergence in the distribution of $V_{s:m}$. Figure 3.4 gives the resulting plot based on the simulated data with $\rho = 0.4$, $p_0 = 0.5$, and $p_1 = 0.3$. We observe that the points in the plot fall compactly along a straight line with a slope of $\frac{1}{2}$, suggesting that the rate of convergence of the distribution of $V_{s:m}$ is $n^{-1/2}$ which is consistent with the result of Theorem 3.3.2.

To explore the effects of parameters ρ, p_0, and p_1 on the accuracy of the normal approximation, we look at the maximum absolute differences between the estimated cdf of

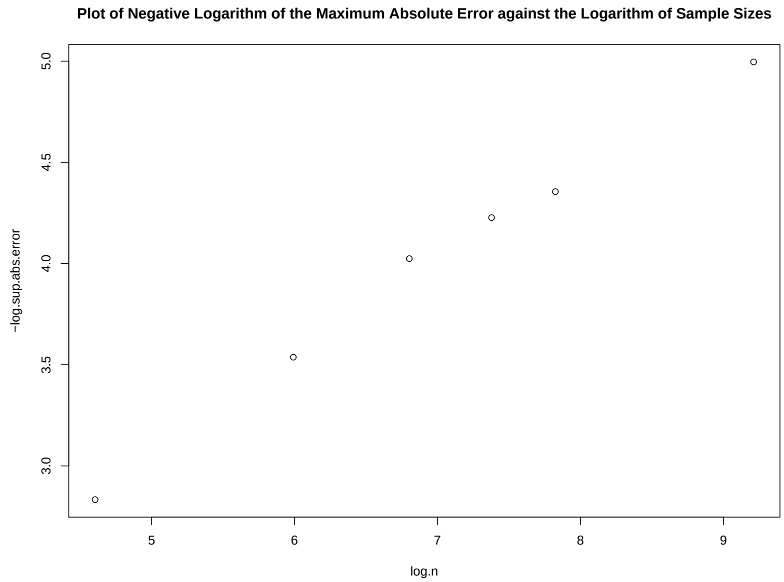

Figure 3.4: Maximum Absolute Errors of the Normal Approximation to the Distribution of $V_{s:m}$ vs. the Sample Sizes when $\rho = 0.4$, $p_0 = 0.5$ and $p_1 = 0.3$ ($m = 0.5n$, $s = 0.15n$).

normalized $V_{s:m}$ and the value provided by the normal approximation for different parameter values. Figure 3.5 gives the relevant graphs, where larger discrepancies are represented by lighter shades. From Figure 3.5 we observe that the patterns of the level plots for different values of ρ are roughly the same, indicating that the effect of correlation on the normal approximation is minimal. We also see that the approximation is much improved for values of p_0 and p_1 close to 0.5 when compared to those for extreme p_0 and p_1 values.

3.3.3 Improving the Normal Approximation

If we have additional information on the joint distribution $F(x, y)$, we can improve upon the normal approximation to the distribution of $V_{s:m}$ given in Section 3.3.2. The idea

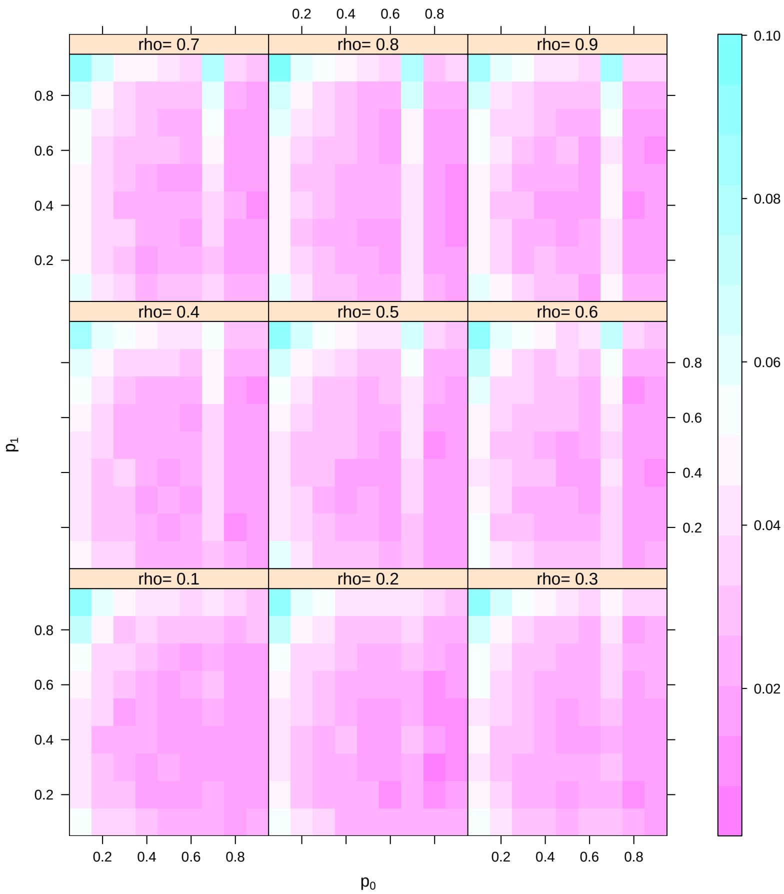

Figure 3.5: Maximum Absolute Errors of the Normal Approximation for the cdf of $V_{s:m}$ for Different p_0, p_1, and ρ

is similar to the *Edgeworth expansions* for the asymptotic distribution of sample mean; see, for example, Ferguson (1996, p31-32).

We will first give a motivational numerical example of a higher order expansion for the cdf of a sample quantile. Then the specific expression for a higher order expansion to the distribution of $V_{s:m}$ will be derived.

For the sample quantile, we have the following well-known property resulting in the normal approximation to the distribution of sample quantile

$$Z_n = \frac{\sqrt{n} f_X(\xi_p)}{\sigma_p}(X_{[np]:n} - \xi_p) \xrightarrow{\mathcal{L}} N(0,1), \tag{3.46}$$

where $\sigma_p = \sqrt{p(1-p)}$, $\xi_p = F_X^{-1}(p)$, and $[np]$ denotes the largest integer not exceeding np. Here we assume the distribution of X is continuous, and the density f_X is positive over a neighborhood of ξ_p. Let $Z_n = \sqrt{n} f_X(\xi_p)(X_{[np]:n} - \xi_p)/\sigma_p$. Then by (3.46), we have

$$\mathbf{P}(Z_n \le z) \approx \Phi(z), \tag{3.47}$$

and the effect of the above approximation is governed by the rate of convergence in (3.46), which is of order $n^{-1/2}$. Using a result in Reiss (1989, p147-148), we can improve upon the above normal approximation by adding an extra term to achieve higher order rate of convergence. In particular with the assumption that the density f_X has bounded third derivative in a neighborhood of ξ_p, we have

$$\sup_z \left| \mathbf{P}(Z_n \le z) - (\Phi(z) + \frac{\phi(z)}{\sqrt{n}} S_{1,n}(z)) \right| = O(n^{-1}) \tag{3.48}$$

where

$$S_{1,n}(z) = \left(\frac{2p-1}{3\sigma_p} + \frac{\sigma_p f'_X(\xi_p)}{2 f_X(\xi_p)^2} \right) z^2 + \frac{np - [np]}{\sigma_p} + \frac{p+1}{3\sigma_p}. \tag{3.49}$$

So one can consider a better approximation to the distribution given by:

$$\mathbf{P}(Z_n \le z) \approx \Phi(z) + \frac{\phi(z)}{\sqrt{n}} S_{1,n}(z) := G_n(z). \tag{3.50}$$

Example 3.3.3. We will use a simple example to demonstrate the improvement in the approximation achieved by (3.50) over the normal approximation. Suppose $X_1, \ldots, X_{10}$ are i.i.d observations from the standard exponential distribution with density $\exp(-x), x \ge 0$. We consider approximating the distribution of the 40% sample percentile, namely

$X_{4:10}$. Figure 3.6 gives the plot of the cdf for the normalized $X_{4:10}$ given by $(X_{4:10} - 0.3297)/0.2154$, along with the normal approximation $\Phi(x)$, and the approximation $G_n(x)$ given by (3.50), and Figure 3.7 gives the scatter plot of the absolute errors of normal approximation and the approximation by (3.50). We observe that even for a small sample size like 10, there is a significant improvement in the accuracy of the second order approximation given by (3.50) over the normal approximation in most cases, except for the neighborhoods where the exact cdf intersects with the cdf given by the normal approximation.

Example 3.3.4. We repeat the above calculations for the standard normal population. Figure 3.8 and Figure 3.9 give the corresponding results. We observe that for the standard normal population the higher order approximation G_n is much better than the normal approximation for almost all the regions except for the upper tail where both approximations are very close to the true cdf. Further, the normal approximation appears to underestimate the cdf of $X_{4:10}$.

In both cases, we note that approximation provided by G_n is amazingly close to the actual cdf in the center part of the distribution of $X_{r:n}$.

Note that the G_n approximation given by (3.50) is not necessarily a valid cdf as it can be easily verified that the derivative of G_n is not necessarily non-negative over $\mathbb{R}$. The derivative of G_n can be expressed as:

$$G_n'(z) = \phi(z)\left(1 + \frac{1}{\sqrt{n}}(-Az^3 + (2A - B)z)\right), \tag{3.51}$$

where

$$A = \frac{2p-1}{3\sigma_p} + \frac{\sigma_p f_X'(\xi_p)}{2f_X(\xi_p)^2},$$

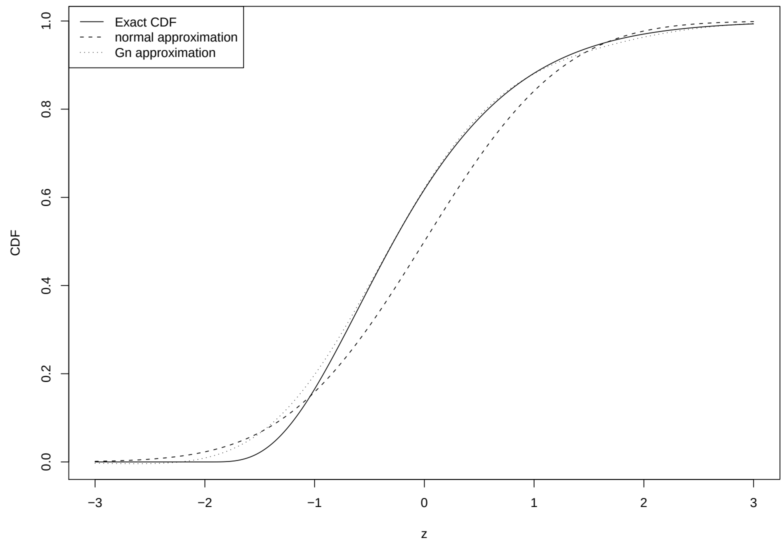

Figure 3.6: Normal Approximation and G_n Approximation to the cdf of $X_{r:n}$ for the Standard Exponential Parent when $n = 10$ and $r = 4$.

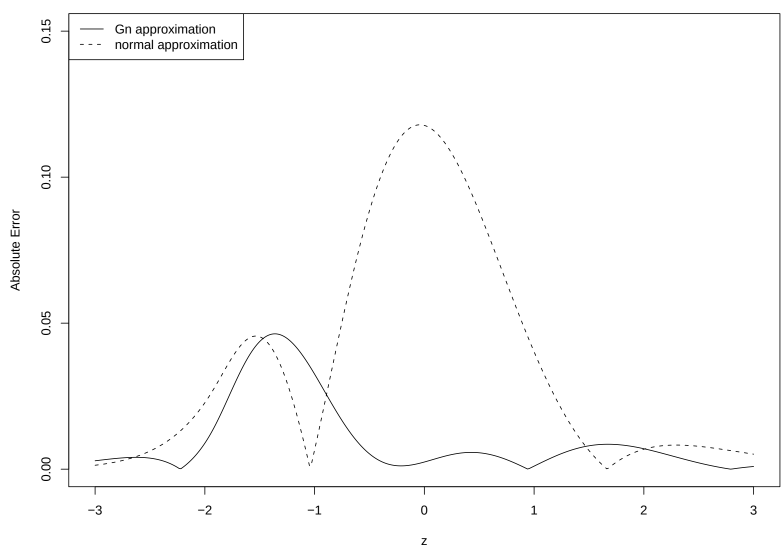

Figure 3.7: Absolute Errors of the Normal Approximation and G_n Approximation to the cdf of $X_{r:n}$ for the Standard Exponential Parent when $n = 10$ and $r = 4$.

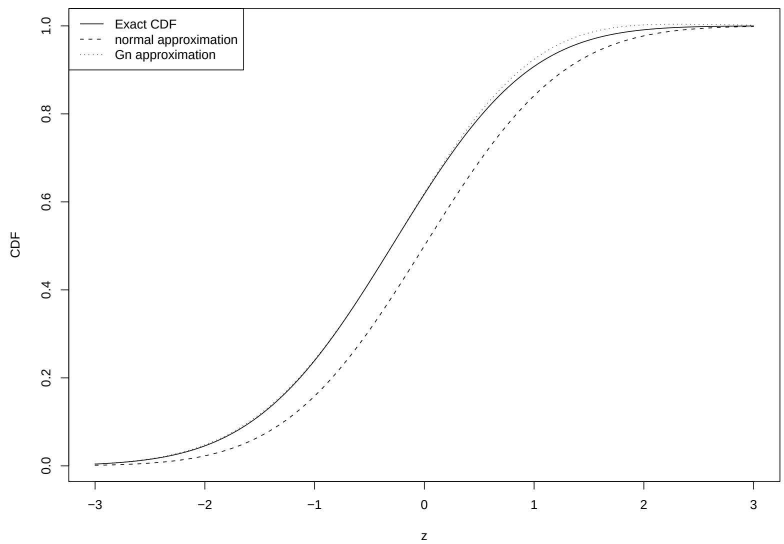

Figure 3.8: Normal Approximation and G_n Approximation to the cdf of $X_{r:n}$ for the Standard Normal Parent when $n = 10$ and $r = 4$.

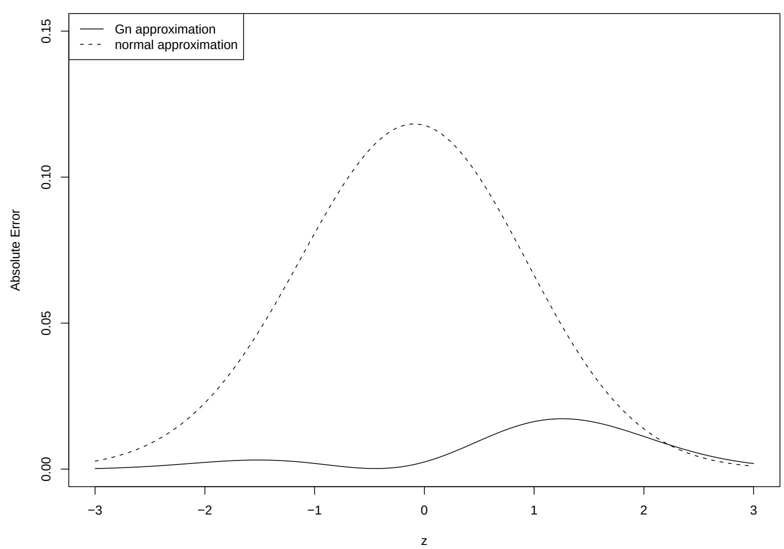

Figure 3.9: Absolute Errors of the Normal Approximation and G_n Approximation to the cdf of $X_{r:n}$ for the Standard Normal Parent when $n = 10$ and $r = 4$.

and

$$B = \frac{np - [np]}{\sigma_p} + \frac{p+1}{3\sigma_p} > 0.$$

So when $A > 0$, for example, when the population distribution is standard normal, and $p < 0.5$, the derivative G_n' can be negative for large enough z. But for p not very close to 0, with a moderate sample size n, the term $1/\sqrt{n}(-Az^3 + (2A - B)z)$ will be dominated by 1 for $z \in [-3, 3]$. As a result, G_n will be monotonically increasing in the range of $[-3, 3]$, which covers the range of values we are usually interested in, as demonstrated by Figure 3.6 and Figure 3.8.

Next we will derive similar result for approximating the distribution of $V_{s:m}$. With a minor adjustment of the arguments used in the proof of Lemma 3.3.2 we have the following result for approximating the expectation of functions of sample quantiles:

Lemma 3.3.3. *Suppose the assumptions in Lemma 3.3.2 hold. Let H be a bounded and continuous function defined on $\mathbb{R}$, and define Z_n as*

$$Z_n = \frac{\sqrt{n} f(\xi_p)}{\sigma_p}(X_{r:n} - \xi_p) \tag{3.52}$$

with $\sigma_p = \sqrt{p(1-p)}$. Then we have

$$\left| \mathbf{E}(H(Z_n)) - \int_{\mathbb{R}} H(z) dG_n(z) \right| = O(n^{-1}) \tag{3.53}$$

where G_n is defined in (3.50).

Then using similar arguments as in the proof of Theorem 3.3.2, we can establish the following result:

Theorem 3.3.3. *With the assumptions of Theorem 3.3.1 and Lemma 3.3.3, and we assume that $F_1^{-1}(p_1|x)$, as a function of x, has third order derivative and the derivative is bounded* *in a neighborhood of* $x_0 = F_X^{-1}(p_0)$. *Suppose* $m = [np_0]$ *and* $s = [mp_1]$. *Then,*

$$\sup_v \left| \mathbf{P}\left(\frac{V_{s:m} - a}{b_n} \le v \right) - \int_{\mathbb{R}} G_{1,n}(v - g_1(x_0)z) dG_{0,n}(z) \right| \le O(n^{-1}) \tag{3.54}$$

where

$$G_{1,n}(z) = \Phi(z) + \frac{\phi(z)}{\sqrt{m}} \left(A_1 z^2 + B_1 \right) \tag{3.55}$$

with

$$A_1 = \frac{2p_1 - 1}{3\sigma_{p_1}} + \frac{\sigma_{p_1} f_1'(a|x_0)}{2f_1(a|x_0)^2}; B_1 = \frac{mp_1 - [mp_1]}{\sigma_{p_1}} + \frac{p_1 + 1}{3\sigma_{p_1}} \tag{3.56}$$

and

$$G_{0,n}(z) = \Phi(z) + \frac{\phi(z)}{\sqrt{n}} \left(A_0 z^2 + B_0 \right) \tag{3.57}$$

with

$$A_0 = \frac{2q_0 - 1}{3\sigma_{q_0}} + \frac{\sigma_{q_0} f_X'(x_0)}{2f_X(x_0)^2}; B_0 = \frac{nq_0 - [nq_0]}{\sigma_{q_0}} + \frac{q_0 + 1}{3\sigma_{q_0}}. \tag{3.58}$$

By Theorem 3.3.3 the second order approximation to the distribution of $V_{s:m}$ is then given by:

$$\begin{aligned}
&\mathbf{P}\left(\frac{V_{s:m} - a}{b_n} \le v \right) \\
&\approx \int_{\mathbb{R}} \Phi(v - g_1(x_0)z)\phi(z)dz + \int_{\mathbb{R}} \frac{\phi(v - g_1(x_0)z)}{\sqrt{m}} (A_1 z^2 + B_1)\phi(z)dz \\
&\quad + \int_{\mathbb{R}} \frac{\Phi(v - g_1(x_0)z)}{\sqrt{n}} [-A_0 z^3 + (2A_0 - B_0)z]\phi(z)dz \\
&= \Phi\left(\frac{v}{\sqrt{1 + g_1^2(x_0)}} \right) + \int_{\mathbb{R}} \frac{\phi(v - g_1(x_0)z)}{\sqrt{[np_0]}} (A_1 z^2 + B_1)\phi(z)dz \\
&\quad + \int_{\mathbb{R}} \frac{\Phi(v - g_1(x_0)z)}{\sqrt{n}} [-A_0 z^3 + (2A_0 - B_0)z]\phi(z)dz \\
&:= H_n(v).
\end{aligned} \tag{3.59}$$

Similar to the G_n approximation given by (3.50) for the cdf of the sample quantile $X_{r:n}$, the approximation $H_n(v)$ in (3.59) is not necessarily a legitimate cdf. But based on simulation studies on the bivariate normal distribution, we observe that the approximation by (3.59) is monotonically increasing in v for values between -3 and 3 when p_0 and p_1 are not very close to 0. Since the last two terms there are of order $o(n^{-1/2})$, $H_n(v)$ will eventually be a monotonically function for all v.

Example 3.3.5. We now illustrate the improvement in the accuracy achieved by the approximation given by (3.59) with the standard bivariate normal distribution with correlation $\rho = 0.5$. Figure 3.10 gives the cdf of normalized $V_{s:m}$ with $n = 50, m = 20, s = 6$, along with cdfs by the normal approximation and the approximation by (3.59). Figure 3.11 gives the plot of the absolute errors of normal approximation and the approximation by (3.59) to the distribution of normalized $V_{s:m}$. Again we observe significant improvement in the approximation achieved by (3.59), and the higher order approximation given by (3.59) is uniformly better than the normal approximation.

3.4 Asymptotic Distribution of $W_{t:n-m}$ in the Quantile Case

In this section we will derive the asymptotic distribution of $W_{t:n-m}$ in the *quantile case*, i.e., for the case where $t = [(n-m)p_2]$, $m = [np_0]$ for p_0 and p_2 such that $0 < p_i < 1, i = 0, 2$, and $n \to \infty$.

From Lemma 3.2.1 we know that given $X_{n-m+1:n} = x$, $W_{t:n-m}$ behaves the same as the tth order statistic of a random sample of size $n-m$ from the distribution with cdf $F_2(\cdot|x)$. So the arguments used in Section 3.3 can be used to establish the limiting distribution of $W_{t:n-m}$ in the quantile case, as well as the rate of convergence to the resulting limiting distribution.

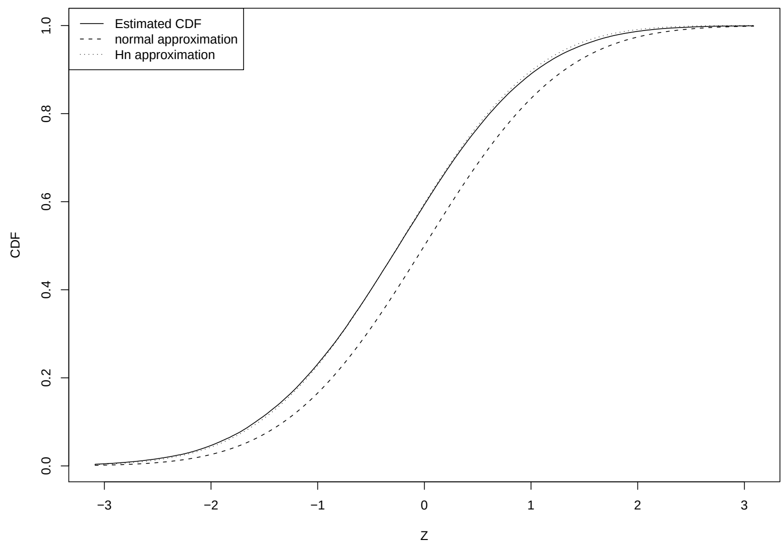

Figure 3.10: Normal Approximation and the Approximation given by (3.59) to the Distribution of Normalized $V_{s:m}$ for the Bivariate Normal Population when $n = 50, m = 20, s = 6$.

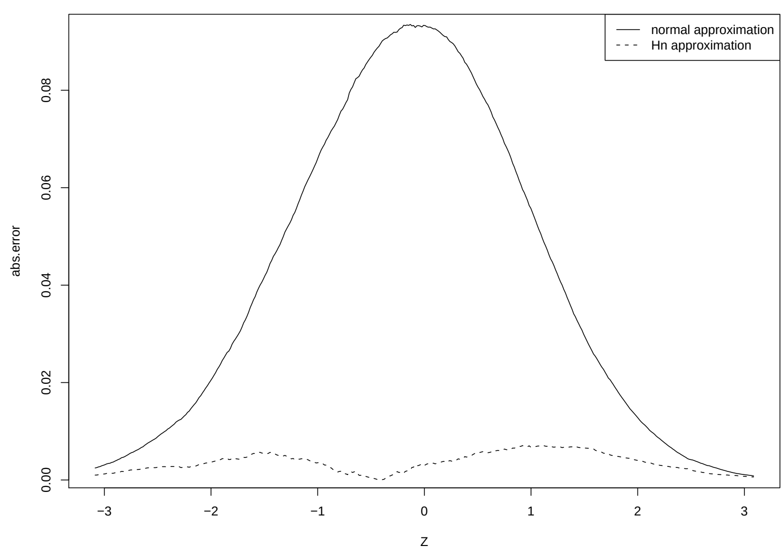

Figure 3.11: Absolute Errors of the Normal Approximation and Approximation by (3.59) to the Distribution of Normalized $V_{s:m}$ with $n = 50, m = 20, s = 6$ for the Bivariate Normal Population.

First we have the following result regarding the asymptotic distribution of $W_{t:n-m}$ in the quantile case.

Theorem 3.4.1. *Let $(X_i, Y_i), i = 1, 2, \ldots, n$, be a random sample from the absolutely continuous bivariate distribution $F(x, y)$, and we assume the joint density $f(x, y)$ is continuous in both arguments. Suppose $m = [np_0]$ and $t = [(n-m)p_2]$, $0 < p_i < 1, i = 0, 2$, and $n \to \infty$. Let*

$$x_0 = F_X^{-1}(q_0), \tag{3.60}$$

with $q_0 = 1 - p_0$, and

$$c = F_2^{-1}(p_2|x_0), \tag{3.61}$$

and assume that $f(x_0, c) > 0$. Then

$$\frac{W_{t:n-m} - c}{d_n} \xrightarrow{\mathcal{L}} Z_1 + g_2(x_0)Z_2 \tag{3.62}$$

where

$$d_n = \left[\frac{\sqrt{nq_0}f_2(c|x_0)}{\sqrt{p_2q_2}}\right]^{-1}, \tag{3.63}$$

Z_1 and Z_2 are independent standard normal variables, and

$$g_2(x_0) = \frac{\sqrt{p_0q_0}}{f_X(x_0)}\left[\frac{\sqrt{q_0}f_2(c|x_0)}{\sqrt{p_2q_2}}\right]\frac{\partial F_2^{-1}(p_2|x)}{\partial x}|_{x=x_0}. \tag{3.64}$$

Remark 3.4.1. (3.62) is equivalent to

$$\sqrt{n}(W_{t:n-m} - c) \xrightarrow{\mathcal{L}} dZ_1 + h_2(x_0)Z_2 \tag{3.65}$$

with

$$d = \frac{\sqrt{p_2q_2}}{\sqrt{q_0}f_2(c|x_0)} \tag{3.66}$$

and

$$\begin{aligned}
h_2(x_0) &= \frac{\sqrt{p_0 q_0}}{f_X(x_0)} \cdot \frac{\partial F_2^{-1}(p_1|x)}{\partial x}|_{x=x_0} \\
&= \frac{\sqrt{p_0 q_0}}{f_X(x_0)} \cdot \frac{p_2 f_X(x_0) - \int_{-\infty}^{c} f(x_0, v) dv}{\int_{-\infty}^{x_0} f(u, c) du} \\
&= \frac{\sqrt{p_0}(F_2(c|x_0) - F_3(c|x_0))}{\sqrt{q_0} f_2(c|x_0)}. \qquad (3.67)
\end{aligned}$$

Example 3.4.1. With the bivariate normal population given by (3.19), the expressions for c, d, and h_2 are given below:

(i) $c = F_2^{-1}(p_2|x_0)$ is the value satisfying the following equation:

$$\int_{-\infty}^{c} \int_{x_0}^{\infty} f(u, v) du dv - p_1 p_0 = 0 \qquad (3.68)$$

with $x_0 = \Phi^{-1}(1 - p_0)$. Equation (3.68) can be solved numerically using Newton-Raphson method.

(ii)

$$d = \frac{\sqrt{p_2 q_2}}{\sqrt{q_0} f_2(c|x_0)} = \frac{\sqrt{p_2 q_2 q_0}}{\Phi\left(\frac{x_0 - \rho c}{\sqrt{1-\rho^2}}\right) \phi(c)}. \qquad (3.69)$$

(iii)

$$\begin{aligned}
h_2(x_0) &= \frac{\sqrt{p_0 q_0}}{f_X(x_0)} \cdot \frac{\partial F_2^{-1}(p_2|x)}{\partial x}|_{x=x_0} && (3.70) \\
&= \frac{\sqrt{p_0 q_0}}{\phi(x_0)} \cdot \frac{p_2 \phi(x_0) - \Phi\left(\frac{c - \rho x_0}{\sqrt{1-\rho^2}}\right) \phi(x_0)}{\Phi\left(\frac{x_0 - \rho c}{\sqrt{1-\rho^2}}\right) \phi(c)} && (3.71) \\
&= \frac{\sqrt{p_0 q_0}\left(p_2 - \Phi\left(\frac{c - \rho x_0}{\sqrt{1-\rho^2}}\right)\right)}{\Phi\left(\frac{x_0 - \rho c}{\sqrt{1-\rho^2}}\right) \phi(c)}. && (3.72)
\end{aligned}$$

For the rate of convergence of the distribution of $W_{t:n-m}$, we have the following result:

$$\sup_w \left| \mathbf{P}\left(\frac{W_{t:n-m}-c}{d_n} \le w\right) - \Phi\left(\frac{w}{\sqrt{1+g_2^2(x_0)}}\right)\right| \le O(n^{-1/2}) \tag{3.73}$$

or equivalently

$$\sup_w \left| \mathbf{P}\left(\sqrt{n}(W_{t:n-m}-c) \le w\right) - \Phi\left(\frac{w}{\sqrt{d^2+h_2^2(x_0)}}\right)\right| \le O(n^{-1/2}). \tag{3.74}$$

And the second order approximation to the distribution of $W_{t:n-m}$ is given by:

$$\begin{aligned}
&\mathbf{P}\left(\frac{W_{t:n-m}-c}{d_n} \le w\right) \\
&\approx \int_{\mathbb{R}} \Phi(w-g_2(x_0)z)\phi(z)dz + \int_{\mathbb{R}} \frac{\phi(w-g_2(x_0)z)}{\sqrt{n-m}}(A_2z^2+B_2)\phi(z)dz \\
&\quad + \int_{\mathbb{R}} \frac{\Phi(w-g_2(x_0)z)}{\sqrt{n}}[-A_0z^3+(2A_0-B_0)z]\phi(z)dz \\
&=\Phi\left(\frac{w}{\sqrt{1+g_2^2(x_0)}}\right) + \int_{\mathbb{R}} \frac{\phi(w-g_2(x_0)z)}{\sqrt{n-m}}(A_2z^2+B_2)\phi(z)dz \\
&\quad + \int_{\mathbb{R}} \frac{\Phi(w-g_2(x_0)z)}{\sqrt{n}}[-A_0z^3+(2A_0-B_0)z]\phi(z)dz
\end{aligned} \tag{3.75}$$

where

$$A_2 = \frac{2p_2-1}{3\sigma_{p_2}} + \frac{\sigma_{p_2} f_2'(c|x_0)}{2f_2(c|x_0)^2}; B_2 = \frac{(n-m)p_2 - [(n-m)p_2]}{\sigma_{p_2}} + \frac{p_2+1}{3\sigma_{p_2}} \tag{3.76}$$

and A_0, B_0 are given by (3.58).

Remark 3.4.2. A comparison of (3.59) and (3.75) reveals the common role played by the features of f_X at x_0 through A_0, B_0, and the symmetric roles played by the conditional cdfs F_1 and F_2 through g_1, A_1, B_1, and g_2, A_2, B_2, respectively.

CHAPTER 4

JOINT DISTRIBUTION OF ORDER STATISTICS OF SUBSETS OF CONCOMITANTS OF ORDER STATISTICS

In Chapter 3, we studied the marginal distributions of $V_{s:m}$ and $W_{t:n-m}$ separately. Here we will focus on the joint distribution of $(V_{s:m}, W_{t:n-m})$. Joshi and Nagaraja (1995) studied the joint distribution of $(V_{s:m}, W_{t:n-m})$ for the special case of $s = m$ and $t = n - m$. We extend their results by considering general s and t. Some interesting applications are also provided for our results.

In Section 4.1, we obtain the joint distribution of $(V_{s:m}, W_{t:n-m})$ for the finite sample case using a conditioning argument. In Section 4.2 the asymptotic distribution of $(V_{s:m}, W_{t:n-m})$ is obtained for the quantile case under appropriate regularity conditions. The asymptotic distribution of $(V_{s:m}, W_{t:n-m})$ in the extremal case is derived in Section 4.3. In Section 4.4 we apply the results to approximate the probability that at least t of $\{Y_{[i:n]} : n - m + 1 \leq i \leq n\}$ are among the top k of all the Y sample values for the bivariate normal population. The results are also used to study the power of identifying the disease-susceptible gene in two-stage designs for gene-disease association studies as discussed in Satagopan et al. (2002) and Satagopan et al. (2004).

4.1 Finite-Sample Joint Distribution of $V_{s:m}$ and $W_{t:n-m}$

The finite sample distribution of $(V_{s:m}, W_{t:n-m})$ can be derived using conditioning argument similar to that used in deriving the marginal distributions of $V_{s:m}$ and $W_{t:n-m}$ in Section 3.2. We have the following lemma regarding the conditional distribution of $(V_{s:m}, W_{t:n-m})$ given the value of $X_{n-m:n}$.

Lemma 4.1.1. *Given $X_{n-m:n} = x$, $V_{s:m}$ behaves like the sth order statistic of a random sample of size m from the cdf $F_1(\cdot|x)$; $W_{t:n-m}$ behaves the same 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 $F_3(\cdot|x)$. Moreover, $V_{s:m}$ and $W_{t:n-m}$ are conditionally independent given the value of $X_{n-m:n}$.*

Remark 4.1.1.

- The results of Lemma 4.1.1 can be proved using exactly the same arguments as those used in the proof of Lemma 3.2.1. So the proof will be omitted here.

- Lemma 3.2.1 deals with the conditional behavior of $V_{s:m}$ and $W_{t:n-m}$ separately for different conditioning events (i.e., given the values of $X_{n-m:n}$ and $X_{n-m+1:n}$, respectively). In contrast, Lemma 4.1.1 considers the joint behavior of $(V_{s:m}, W_{t:n-m})$ conditional on the value of $X_{n-m:n}$, which is very important in deriving the joint distribution of $(V_{s:m}, W_{t:n-m})$.

Using Lemma 4.1.1, the joint cdf of $(V_{s:m}, W_{t:n-m})$ can be derived as follows:

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

with

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

and

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

Note that the conditional distribution of $W_{t:n-m}$ given $X_{n-m:n} = x$ is that of an order statistic when there is a single outlier (Arnold and Balakrishnan, 1989, p109).

4.2 Asymptotic Distribution of $(V_{s:m}, W_{t:n-m})$ in the Quantile Case

4.2.1 Main Results

To derive the asymptotic joint distribution of $(V_{s:m}, W_{t:n-m})$ for the quantile case, we first need the following lemma about the limiting distribution of a central order statistic for a random sample with a single outlier.

Lemma 4.2.1. *Let $X_1, \ldots, X_{n-1}$ be a random sample from an absolutely continuous distribution with cdf $F_X(x)$, and Y be an independent random variable with cdf $F_Y(y)$. Let*

W_m be the mth order statistic of $(X_1, \ldots, X_{n-1}, Y)$, $1 \leq m \leq n$. If $m = [np]$ as $n \to \infty$, then we have

$$\sqrt{n}(W_m - x_0) \xrightarrow{\mathcal{L}} N\left(0, \frac{pq}{[f_X(x_0)]^2}\right)$$

where

$$q = 1 - p; \qquad x_0 = F_X^{-1}(p).$$

Proof. Let $X_{m:n-1}$ be the mth order statistic of X sample values. Then we have:

$$\sqrt{n}(X_{m:n-1} - x_0) \xrightarrow{\mathcal{L}} N\left(0, \frac{pq}{[f_X(x_0)]^2}\right).$$

If we can show that

$$\sqrt{n}\,|W_m - X_{m:n-1}| \xrightarrow{P} 0$$

then the desired result will follow from the Slutsky's Theorem.

Notice

$$\begin{aligned} & |W_m - X_{m:n-1}| \\ = \;& |X_{m-1:n-1} - X_{m:n-1}|\, I_{\{Y \leq X_{m-1:n-1}\}} + |Y - X_{m:n-1}|\, I_{\{X_{m-1:n-1} \leq Y < X_{m:n-1}\}} \\ \leq \;& 2\,|X_{m-1:n-1} - X_{m:n-1}| \end{aligned}$$

It suffices to show that

$$\sqrt{n}(X_{m:n-1} - X_{m-1:n-1}) \xrightarrow{P} 0$$

By Bloch and Gastwirth (1968) we have

$$(n-1) f_X(x_0)(X_{m:n-1} - X_{m-1:n-1}) \xrightarrow{\mathcal{L}} N(1, 1)$$

so the desired result follows. $\square$

Then the main results regarding the joint limiting distribution of $(V_{s:m}, W_{t:n-m})$ in the quantile case can be summarized by the following theorem:

Theorem 4.2.1. *Let* $(V_{s:m}, W_{t:n-m})$ *be defined as before. Suppose* $m = [np_0]$, $s = [mp_1]$, *and* $t = [(n-m)p_2]$, $0 < p_i < 1, i = 0, 1, 2,$ *as* $n \to \infty$. *Assume that the joint density* $f(x, y)$ *is continuous in both arguments, and is positive over the entire support set. Then*

$$\begin{aligned} F_{V_{s:m},W_{t:n-m}}(a + b_n v, c + d_n w) &\to \int \Phi(v - g_1(x_0)z) \cdot \Phi(w - g_2(x_0)z) d\Phi(z) \\ &= \mathbf{E}\{\Phi(v - g_1(x_0)Z) \cdot \Phi(w - g_2(x_0)Z)\} \end{aligned}$$

where x_0, a, b_n, c, d_n, g_1, *and* g_2 *are defined as in (3.4), (3.5), (3.7), (3.61), (3.63), (3.8), and (3.64), respectively, and* Z *is a standard normal random variable.*

Proof. Notice

$$\begin{aligned} & F_{V_{s:m},W_{t:n-m}}(a + b_n v, c + d_n w) \\ = \ & \mathbf{P}(V_{s:m} \le a + b_n v, W_{t:n-m} \le c + d_n w) \\ = \ & \int \mathbf{P}(V_{s:m} \le a + b_n v, W_{t:n-m} \le c + d_n w | Z_n = z) F_{Z_n}(dz) \\ & \text{(where } Z_n \text{ is defined as in Theorem 3.3.1)} \\ = \ & \int \mathbf{P}(V_{s:m} \le a + b_n v | Z_n = z) \mathbf{P}(W_{t:n-m} \le c + d_n w | Z_n = z) F_{Z_n}(dz) \\ & \text{(since } V_{s:m} \text{ and } W_{t:n-m} \text{ are conditionally independent).} \end{aligned}$$

From the proof of Theorem 3.3.1, we have for any $z \in \mathbb{R}$,

$$\mathbf{P}(V_{s:m} \le a + b_n v | Z_n = z) \to \Phi(v - g_1(x_0)z).$$

Using the same arguments as in the proof of Theorem 3.3.1, combined with the result given by Lemma 4.2.1, we can easily establish that

$$\mathbf{P}(W_{t:n-m} \le c + d_n w | Z_n = z) \to \Phi(w - g_2(x_0)z)$$

as $n \to \infty$.

So the desired result easily follows from a convergence result in Royden (1968, Proposition 18). $\square$

Remark 4.2.1. Notice the results of Theorem 4.2.1 can also be expressed as:

$$\left(\frac{V_{s:m} - a}{b_n}, \frac{W_{t:n-m} - c}{d_n}\right) \xrightarrow{\mathcal{L}} (g_1(x_0)Z_1 + Z_2, g_2(x_0)Z_1 + Z_3)$$

or

$$\sqrt{n}\,(V_{s:m} - a, W_{t:n-m} - c) \xrightarrow{\mathcal{L}} (bZ_2 + h_1(x_0)Z_1, dZ_3 + h_2(x_0)Z_1) \tag{4.1}$$

where Z_1, Z_2, and Z_3 are i.i.d standard normal random variables, b, d, h_1, and h_2 are defined as in (3.15), (3.66), (3.18) and (3.67) respectively. So we have

$$\sqrt{n}\,(V_{s:m} - a, W_{t:n-m} - c) \xrightarrow{\mathcal{L}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} h_1^2(x_0) + b^2 & h_1(x_0)h_2(x_0) \\ h_1(x_0)h_2(x_0) & h_2^2(x_0) + d^2 \end{bmatrix}\right). \tag{4.2}$$

4.2.2 Rate of Convergence

From the proof of Theorem 4.2.1 we know that

$$\begin{aligned} &\mathbf{P}(V_{s:m} \le a + b_n v, W_{t:n-m} \le c + d_n w) \\ =&\mathbf{E}\left(\mathbf{P}(V_{s:m} \le a + b_n v | Z_n)\mathbf{P}(W_{t:n-m} \le c + d_n w | Z_n)\right). \end{aligned} \tag{4.3}$$

Similar arguments as in the proof of Theorem 3.3.2 can establish that for any $z \in \mathbb{R}$:

$$|\mathbf{P}(V_{s:m} \le a + b_n v | Z_n = z) - \Phi(v - g_1(x_0)z)| \le O(n^{-1/2})$$

and

$$|\mathbf{P}(W_{t:n-m} \le c + d_n w | Z_n = z) - \Phi(w - g_2(x_0)z)| \le O(n^{-1/2})$$

which in turn implies that

$$\begin{aligned}|\mathbf{E}\left(\mathbf{P}(V_{s:m} \leq a + b_n v|Z_n)\mathbf{P}(W_{t:n-m} \leq c + d_n w|Z_n)\right) \\ - \mathbf{E}\{\Phi(v - g_1(x_0)Z_n) \cdot \Phi(w - g_2(x_0)Z_n)\}| \leq O(n^{-1/2}). \qquad (4.4)\end{aligned}$$

So it follows from Lemma 3.3.2 that:

$$\begin{aligned}|\mathbf{P}(V_{s:m} \leq a + b_n v, W_{t:n-m} \leq c + d_n w) \\ - \mathbf{E}\{\Phi(v - g_1(x_0)Z) \cdot \Phi(w - g_2(x_0)Z)\}| \leq O(n^{-1/2}), \qquad (4.5)\end{aligned}$$

or equivalently

$$\left|\mathbf{P}\left(\sqrt{n}(V_{s:m} - a) \leq v, \sqrt{n}(W_{t:n-m} - c) \leq w\right) - F\left(v, w\right)\right| \leq O(n^{-1/2}) \qquad (4.6)$$

where $F(x, y)$ is the cdf of a bivariate normal distribution given in (4.2). So similar to those in the marginal distributions, the rate of convergence in the joint distribution of $(V_{s:m}, W_{t:n-m})$ is also of order $n^{-1/2}$.

Example 4.2.1 (Bivariate Normal Distribution)**.** As done in Chapter 3, here we want to examine the rate of convergence in the distribution of $(V_{s:m}, W_{t:n-m})$ numerically for the standard bivariate normal distribution. Using Monte Carlo simulations we estimate the cdf of the joint distribution of normalized $(V_{s:m}, W_{t:n-m})$ for a grid bounded by 2 standard deviations around their means respectively. We also calculate the Normal approximation as given by (4.2). Then we calculate the maximum absolute error of the normal approximation over the entire grid. We carry out these calculations for $n = 100, 400, 900, 1600, 2500$ and 10000, and we set $\rho = 0.4$, $p_0 = 0.5$, $p_1 = p_2 = 0.3$. Figure 4.1 gives the scatter plot of the negative logarithm of the maximum absolute errors vs. the logarithm of sample sizes. Very similar to the Figure 3.4, the points in the plot is very close to a straight line with slope approximately 0.5, suggesting that the rate of convergence in the distribution of $(V_{s:m}, W_{t:n-m})$ is of order $\sqrt{n}$ as implied by (4.6).

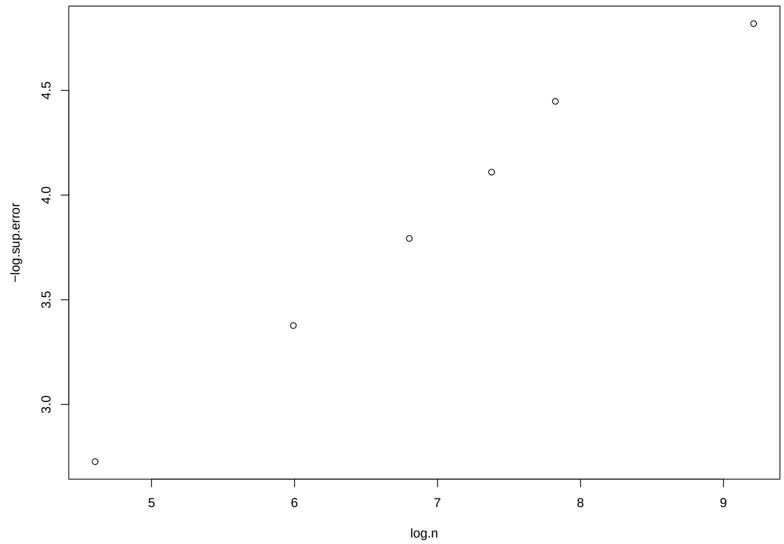

Figure 4.1: Plot of the Logarithm of the Maximum Absolute Errors of the Normal Approximation to the Distribution of $(V_{s:m}, W_{t:n-m})$ vs. the Logarithm of Sample Sizes with $\rho = 0.4$, $p_0 = 0.5$, $p_1 = p_2 = 0.3$ ($m = 0.5n$, $s = 0.15n$ and $t = 0.15n$).

4.3 Asymptotic Distribution of $(V_{s:m}, W_{t:n-m})$ in the Extremal Case

In this section we will study the asymptotic distribution of $(V_{m-s+1:m}, W_{n-m-t+1:n-m})$ when s and t are fixed as n and m approach infinity. The development follows that of Joshi and Nagaraja (1995). The main result regarding the joint limiting distribution of $(V_{m-s+1:m}, W_{n-m-t+1:n-m})$ is given by the following theorem:

Theorem 4.3.1. *Let* $m = [np_0]$*, as* $n \to \infty$*, with* $0 < p_0 < 1$*, while* s *and* t *are kept fixed. Let* $V_{m-s+1:m}$ *and* $W_{n-m-t+1:n-m}$ *be defined as before. Assume:*

(i) The marginal density of X, f_X, is continuous at $x_0 = F_X^{-1}(q_0)$, where $q_0 = 1 - p_0$, and $f_X(x_0) > 0$;

(ii) the density of the conditional distribution of Y given $X = x$, $f_3(y|x)$, is continuous at x_0 for all real y;

(iii) there exist constants $a_n,\ b_n > 0$, and $c_n,\ d_n > 0$, such that as $n \to \infty$

$$[F_1(a_n + b_n y|x_0)]^n \to G_1(y)$$

$$[F_2(c_n + d_n y|x_0)]^n \to G_2(y)$$

for all real y, where G_1 and G_2 are some nondegenerate cdf's;

(iv) for all real c and y we have:

$$\mathbf{P}(x_0 \leq X \leq x_0 + c/\sqrt{n}, Y > a_n + b_n y) = o(1/n)$$

$$\mathbf{P}(x_0 \leq X \leq x_0 + c/\sqrt{n}, Y > c_n + d_n y) = o(1/n).$$

Then we have

$$F_{V_{m-s+1:m}, W_{n-m-t+1:n-m}}(a_n + b_n v, c_n + d_n w) \to G_{1,s}(v) G_{2,t}(w)$$

where

$$G_{1,s}(v) = [G_1(v)]^{p_0} \sum_{i=0}^{s-1} \frac{[-p_0 \log G_1(v)]^i}{i!}$$
$$G_{2,t}(w) = [G_2(w)]^{q_0} \sum_{j=0}^{t-1} \frac{[-q_0 \log G_2(w)]^j}{j!}$$

Proof. Notice

$$\begin{aligned} & F_{V_{m-s+1:m}, W_{n-m-t+1:n-m}}(a_n + b_n v, c_n + d_n w) \\ = \ & E[h_1(a_n + b_n v, X_{n-m:n}) h_2(c_n + d_n v, X_{n-m:n})] \\ = \ & E[h_1(a_n + b_n v, x_0 + c_0 Z_n/\sqrt{n}) h_2(c_n + d_n v, x_0 + c_0 Z_n/\sqrt{n})] \end{aligned}$$

where c_0 and Z_n are defined in (3.11) and

$$h_1(v, x) = \mathbf{P}(V_{m-s+1:m} \leq v | X_{n-m:n} = x)$$

$$h_2(w, x) = \mathbf{P}(W_{n-m-t+1:n-m} \leq w | X_{n-m:n} = x).$$

In view of the arguments presented in the proof of Result 2 in Nagaraja and David (1994), it is enough to show for all real z:

$$h_1(a_n + b_n v, x_0 + c_0 z/\sqrt{n}) \rightarrow G_{1,s}(v) \tag{4.7}$$

$$h_2(c_n + d_n w, x_0 + c_0 z/\sqrt{n}) \rightarrow G_{2,t}(w) \tag{4.8}$$

Note

$$\begin{aligned} & h_1(a_n + b_n v, x_0 + c_0 z/\sqrt{n}) && (4.9) \\ = \; & \mathbf{P}(V_{m-s+1:m} \leq a_n + b_n v | X_{n-m:n} = x_0 + c_0 z/\sqrt{n}) && (4.10) \\ = \; & \sum_{j=m-s+1}^{m} \binom{m}{j} [F_1^{(n)}(z)]^j [1 - F_1^{(n)}(z)]^{m-j} && (4.11) \end{aligned}$$

where $F_1^{(n)}(z) = F_1(a_n + b_n v | x_0 + c_0 z/\sqrt{n})$.

While by assumption (ii) and (iii), it can be shown that

$$m[1 - F_1^{(n)}(z)] \sim p_0 n[1 - F_1^{(n)}(z)] \rightarrow -p_0 \log G_1(v) \tag{4.12}$$

Applying Poisson approximation to Binomial probabilities, from (4.11) and (4.12), we obtain (4.7).

For (4.8), note

$$
\begin{aligned}
& h_2(c_n + d_n w, x_0 + c_0 z/\sqrt{n}) \\
=& \mathbf{P}(W_{n-m-t+1:n-m} \le c_n + d_n w | X_{n-m:n} = x_0 + c_0 z/\sqrt{n}) \\
=& \sum_{j=n-m-t+1}^{n-m-1} \binom{n-m-1}{j} [F_2^{(n)}(z)]^j [1 - F_2^{(n)}(z)]^{n-m-1-j} \\
& + \binom{n-m-1}{n-m-t} [F_2^{(n)}(z)]^{n-m-t} [1 - F_2^{(n)}(z)]^{t-1} \\
& F_3(c_n + d_n w | x_0 + c_0 z/\sqrt{n}) \qquad (4.13)
\end{aligned}
$$

where $F_2^{(n)}(z) = F_2(c_n + d_n w | x_0 + c_0 z/\sqrt{n})$.

From the proof of Theorem 1 in Joshi and Nagaraja (1995), we know that

$$F_3(c_n + d_n w | x_0 + c_0 z/\sqrt{n}) \to 1, \text{ as } n \to \infty$$

Similarly we can prove that

$$(n-m-1)[1 - F_2^{(n)}(z)] \to -q_0 \log G_2(w)$$

and hence (4.8) follows from the application of the Poisson approximation to the Binomial cdf in (4.13). □

Remark 4.3.1.

- Assumption (ii) requires the continuity of the function $f_3(y|x)$ as a function of x at the value x_0 for all real y, which can be guaranteed by the continuity of the joint pdf $f(x, y)$ at x_0; see Joshi and Nagaraja (1995);

- Assumption (iii) implies that the distributions F_1 and F_2 are in the domain of attraction of G_1 and G_2 respectively, i.e., $F_1 \in D(G_1)$ and $F_2 \in D(G_2)$. Notice that the tail behavior of F_1 and F_2 are much messier than that of F_Y. So it is usually hard to verify this assumption. However if we can verify that $F_Y \in D(G)$ as $n \to \infty$ such that $[F_Y(a_n + b_n y)]^n \to G(y)$ for some non-degenerate distribution G, and F_1 is tail-equivalent to F_Y in the sense that there exists a finite positive β_1, such that

$$\lim_{y\to\infty} \frac{1 - F_Y(y)}{1 - F_1(y|x_0)} = \beta_1$$

then by Resnick (1987), we have:

$$[F_1(a_n + b_n y|x_0)]^n \to G_1(y)$$

with $G_1(y) = G(a + by)$, where a and b are some constants depending on G. Further it can be shown that the tail-equivalence of F_1 and F_Y combined with the assumption that $\beta_1 > p_0$ implies the tail-equivalence of F_2 and F_Y. So we have a similar result for F_2; see Joshi and Nagaraja (1995) for details.

- For assumption (iv) to hold, as discussed in Joshi and Nagaraja (1995), a sufficient condition is that:

$$\Delta_1 \bar{F}(x, a_n + b_n y) = o(1/\sqrt{n})$$

and

$$\Delta_1 \bar{F}(x, c_n + d_n y) = o(1/\sqrt{n})$$

hold uniformly in x in the neighborhood of x_0, where $\Delta_1 \bar{F}(x, y)$ is the first partial derivative of $F(x, y)$ with respect to x.

Example 4.3.1 (Bivariate Normal Distribution)**.** For the bivariate normal distribution given by (3.19), it can be shown that (Nagaraja and David, 1994):

$$\lim_{y\to\infty} \frac{1 - F_Y(y)}{1 - F_1(y|x_0)} = p_0.$$

So F_1 and F_Y are tail-equivalent, and by the well-known fact that $[F_Y(a_n + b_n y)]^n \to \Lambda(y)$ as $n \to \infty$ with

$$a_n = \sqrt{2\log n} - \frac{1}{2}\frac{\log(4\pi\log n)}{\sqrt{2\log n}} \tag{4.14}$$

$$b_n = \frac{1}{\sqrt{2\log n}} \tag{4.15}$$

where $\Lambda(y) = \exp\{-e^{-y}\}$ is the Gumbel extreme value cdf, we have

$$[F_1(a_n + b_n y)]^n \to \Lambda(\log p_0 + y) = \exp\{-e^{-y}/p_0\}$$

Notice $\beta_1 = p_0$, so F_2 and F_Y are not tail-equivalent. But as in the proof of Theorem 3 of Joshi and Nagaraja (1995), we do have $F_2 \in D(\Lambda)$ with

$$c_n = \rho x_0 + \theta\left[\sqrt{2\log n} - \frac{\log(4\pi\log n)}{\sqrt{2\log n}}\right] - \theta\frac{(x_0^2/2) + \log(q_0\rho/\theta)}{\sqrt{2\log n}} \tag{4.16}$$

$$d_n = \theta b_n \tag{4.17}$$

where $\theta = \sqrt{1-\rho^2}$.

As verified in Joshi and Nagaraja (1995), the assumption (iv) of Theorem 4.3.1 also holds. So by Theorem 4.3.1, we have

$$\begin{aligned} & F_{V_{m-s+1:m}, W_{n-m-t+1:n-m}}(a_n + b_n v, c_n + d_n w) \\ \to \quad & \exp\{-[e^{-v} + q_0 e^{-w}]\} \sum_{i=0}^{s-1} \frac{e^{-iv}}{i!} \sum_{j=0}^{t-1} \frac{q_0^j e^{-jw}}{j!} \end{aligned}$$

as $n \to \infty$, where a_n, b_n, c_n and d_n are given by (4.14), (4.15), (4.16), and (4.17), respectively.

4.4 Applications

4.4.1 Selection Through an Associated Characteristic

In selection procedures, items or subjects may be chosen on the basis of their X characteristic, and an associated characteristic Y that is hard to measure or that can be observed only later, may be of interest. For example, X may be the score of a candidate on a screening test, and Y is the measure of the final performance of the candidate; or X could be the score assigned by a particular search engine (like Google, Yahoo etc.), and Y is the score assigned by the user based on his needs.

Yeo and David (1984) considered the problem of choosing the best k objects out of a group of n on the basis of auxiliary measurements X_i, while the measurements of primary interest, Y_i, are not available. They were interested in the probability that the k subjects with the largest Y-values are among the m subjects with the largest X-values.

Here we consider a more general situation. Define the event of our interest, E, to be *"at least s of* $\{Y_{[i:n]} : n-m+1 \leq i \leq n\}$ *are among* $\{Y_{i:n}, n-k+1 \leq i \leq n\}$*"*, i.e., of the m objects with largest X-values, at least s are included in the set of k objects with largest Y-values, where $s \leq k \leq m$. Notice with $s = k$, the problem reduces to that of Yeo and David (1984). Next we will derive an expression for the probability of the event E using the results in previous sections.

Expression for the Probability of the Event of Interest

Notice that the event E can be expressed in terms of the order statistics of subsets of concomitants as

$$E = \{V_{m-s+1:m} > W_{n-m-k+s:n-m}\},$$

and so we have

$$\begin{aligned}
\mathbf{P}(E) &= \mathbf{P}(V_{m-s+1:m} > W_{n-m-k+s:n-m}) \\
&= \int_x \mathbf{P}(V_{m-s+1:m} > W_{n-m-k+s:n-m} | X_{n-m:n} = x) f_{n-m:n}(x) dx \\
&= \int_x \left[\int_v \mathbf{P}(W_{n-m-k+s:n-m} < v | X_{n-m:n} = x) f_{V_{m-s+1:m}}(v) dv \right] f_{n-m:n}(x) dx,
\end{aligned}$$

where

$$
\begin{aligned}
&\mathbf{P}(W_{n-m-k+s:n-m} < v | X_{n-m:n} = x) \\
&= \sum_{i=n-m-k+s}^{n-m-1} \binom{n-m-1}{i} [F_2(v|x)]^m [1 - F_2(v|x)]^{n-m-1-i} \\
&\quad + \binom{n-m-1}{n-m-k+s-1} [F_2(v|x)]^{n-m-k+s-1} [1 - F_2(v|x)]^{k-s} F_3(v|x),
\end{aligned}
$$

$$f_{V_{m-s+1:m}}(v) = \frac{m!}{(m-s)!(s-1)!} [F_1(v|x)]^{m-s} [1 - F_1(v|x)]^{s-1} f_1(v|x),$$

and

$$f_{n-m:n}(x) = \frac{n!}{(n-m-1)!m!} [F_X(x)]^{n-m-1} [1 - F_X(x)]^m f_X(x).$$

Let

$$\theta_1(x, v) = \mathbf{P}(X \le x, Y \le v); \qquad \theta_2(x, v) = \mathbf{P}(X \le x, Y > v);$$

$$\theta_3(x, v) = \mathbf{P}(X > x, Y \le v); \qquad \theta_4(x, v) = \mathbf{P}(X > x, Y > v);$$

and

$$
\begin{aligned}
g(x, v) &= \int_x^\infty f_{X,Y}(u, v) du = (1 - F_X(x)) f_1(v|x); \\
h(x, v) &= \int_{-\infty}^v f_{X,Y}(x, w) dw = f_X(x) F_3(v|x).
\end{aligned}
$$

Then $\mathbf{P}(E)$ can be expressed as

$$
\begin{aligned}
&\sum_{i=n-m-k+s}^{n-m-1} \frac{n!}{i!(n-m-i-1)!(m-s)!(s-1)!} \times \\
&\quad \int_x \int_v \theta_1^i(x, v) \theta_2^{n-m-i-1}(x, v) \theta_3^{m-s}(x, v) \theta_4^{s-1}(x, v) g(x, v) f_X(x) dv dx \\
&+ \frac{n!}{(k-s)!(n-m-k+s-1)!(m-s)!(s-1)!} \times \\
&\int_x \int_v \theta_1^{n-m-k+s-1}(x, v) \theta_2^{k-s}(x, v) \theta_3^{m-s}(x, v) \theta_4^{s-1}(x, v) g(x, v) h(x, v) dv dx. \qquad (4.18)
\end{aligned}
$$

Remark 4.4.1.

- If X and Y are independent, $(Y_{[1:n]}, \ldots, Y_{[n-m:n]})$ and $(Y_{[n-m+1:n]}, \ldots, Y_{[n:n]})$ become two independent random samples from the distribution F_Y. So we have

$$\mathbf{P}(E) \quad = \quad \mathbf{P}(U_{m-s+1:m} > U^*_{n-m-k+s:n-m})$$

where $U_{m-s+1:m}$ is the $(m-s+1)$st order statistic of a random sample of size m from the standard uniform distribution, and $U^*_{n-m-k+s:n-m}$ is the $(n-m-k+s)$th order statistic of a random sample of size $n-m$ from the standard uniform distribution. This probability can be readily computed as:

$$\begin{aligned}\mathbf{P}(E) =& \mathbf{P}(U_{m-s+1:m} > U^*_{n-m-k+s:n-m}) \\ =& [\binom{n}{m}]^{-1} \sum_{i=n-m-k+s}^{n-m} \binom{i+m-s+1}{i} \binom{n-i-m+s-1}{s-1} \\ =& [\binom{n}{k}]^{-1} \sum_{i=s}^{m} \binom{n-m}{k-i} \binom{m}{i}. \end{aligned} \tag{4.19}$$

Notice (4.19) is just a special case of the probability addressed by Olkin and Stephens (1993) in which these authors studied the probability that *exactly* t of the selected subsample of size s are in the top k of the entire list for independent but nonidentically distributed normal populations.

- Intuitively, the probability $\mathbf{P}(E)$ will only depend on the dependence structure between X and Y arising from the distribution $F(x, y)$. Actually as shown in the following theorem, the probability $\mathbf{P}(E)$ depends on the cdf $F(x, y)$ only through the associated copula function, which can be thought of as the intrinsic measure of dependence between X and Y.

Theorem 4.4.1. *Suppose* $(X_i, Y_i), i = 1, \ldots, n$, *is a random sample from the absolutely continuous bivariate distribution* $F(x, y)$. *The probability that at least* s *of*

$\{Y_{[i]}, n-m+1 \leq i \leq n\}$ are among $\{Y_{(i)}, n-k+1 \leq i \leq n\}$ depends on the $F(x,y)$ only through the copula function associated with F.

Proof. Let $C_F(x,y)$ be the copula function associated with F, i.e.,

$$F(x,y) = C_F(F_X(x), F_Y(y)) \quad \forall x, y \in \mathbb{R}$$

where F_Y and F_Y are the marginal distributions for X and Y, respectively.

Note

$$\begin{aligned}\mathbf{P}(E) =& \mathbf{P}(\text{at least } s \text{ of } \{Y_{[i]}, n-m+1 \leq i \leq n\} \\ & \text{are among } \{Y_{(i)}, n-k+1 \leq i \leq n\}) \\ =& \mathbf{P}(\text{at least } s \text{ of } \{F_Y(Y_{[i]}), n-m+1 \leq i \leq n\} \\ & \text{are among } \{F_Y(Y_{(i)}), n-k+1 \leq i \leq n\})\end{aligned}$$

It can be easily shown that the joint cdf for $(F_X(X), F_Y(Y))$ is just the copula function $C_F(x,y)$ associated with F. So we have:

$$((F_X(X_1), F_Y(Y_1)), \ldots, (F_X(X_n), F_Y(Y_n))) \stackrel{\mathrm{d}}{=} ((X_1^*, Y_1^*), \ldots, (X_n^*, Y_n^*))$$

where $(X_i^*, Y_i^*), i = 1, \ldots, n$ is a random sample from the distribution with cdf C_F (the copula function associated with F) and support $[0,1] \times [0,1]$.

Also notice that

$$((F_X(X_{(1)}), F_Y(Y_{[1]})), \ldots, (F_X(X_{(n)}), F_Y(Y_{[n]}))) \stackrel{\mathrm{d}}{=} ((X_{(1)}^*, Y_{[1]}^*), \ldots, (X_{(n)}^*, Y_{[n]}^*)).$$

Thus $\mathbf{P}(E)$ can be expressed as

$$\begin{aligned}\mathbf{P}(E) =& \mathbf{P}(\text{at least } s \text{ of } \{Y_{[i]}^*, n-m+1 \leq i \leq n\} \\ & \text{are among } \{Y_{(i)}^*, n-k+1 \leq i \leq n\}),\end{aligned}$$

which depends only on C_F. So the desired result follows. $\square$

A Large Sample Approximation

We see that the formula (4.18) for the desired probability $\mathbf{P}(E)$ is quite complicated, especially the double integrals involved. And it is not practical to apply the formula to calculate $\mathbf{P}(E)$ for large values of n. But we can use the results of Theorem 4.2.1 to obtain a large sample approximation to that probability. By Theorem 4.2.1, we have:

$$\left(\frac{V_{s:m}-a}{b_n}, \frac{W_{t:n-m}-c}{d_n}\right) \xrightarrow{\mathcal{L}} (g_1(x_0)Z_1 + Z_2, g_2(x_0)Z_1 + Z_3)$$

where a, b_n, c, d_n, x_0, g_1 and g_2 are defined the same as in Theorem 4.2.1; Z_1, Z_2 and Z_3 are i.i.d standard normal variables.

So for sufficiently large n, we have:

$$(V_{s:m}, W_{t:n-m}) \overset{\text{approx}}{\sim} BVN\left(\begin{bmatrix} a \\ c \end{bmatrix}, \begin{bmatrix} \frac{h_1^2(x_0)+b^2}{n} & \frac{h_1(x_0)h_2(x_0)}{n} \\ \frac{h_1(x_0)h_2(x_0)}{n} & \frac{h_2^2(x_0)+d^2}{n} \end{bmatrix}\right)$$

where $h_1(x_0) = \sqrt{n}b_n g_1(x_0)$, $b = \sqrt{n}b_n$, $h_2(x_0) = \sqrt{n}d_n g_2(x_0)$, and $d = \sqrt{n}d_n$. Thus it follows that

$$\mathbf{P}(V_{s:m} - W_{t:n-m} > 0) \tag{4.20}$$

$$\approx \quad 1 - \Phi\left(\frac{\sqrt{n}(c-a)}{\sqrt{h_1^2(x_0) + b^2 + h_2^2(x_0) + d^2 - 2h_1(x_0)h_2(x_0)}}\right). \tag{4.21}$$

By (4.21) a large sample approximation to the probability of our interest $\mathbf{P}(E)$ can be readily written as:

$$\begin{aligned}\mathbf{P}(E) &= \mathbf{P}(V_{m-s+1:m} > W_{n-m-k+s:n-m}) \\ &\approx 1 - \Phi\left(\frac{\sqrt{n}(\tilde{c}-\tilde{a})}{\sqrt{h_1^2(\tilde{x_0}) + \tilde{b}^2 + h_2^2(\tilde{x_0}) + \tilde{d}^2 - 2h_1(\tilde{x_0})h_2(\tilde{x_0})}}\right),\end{aligned} \tag{4.22}$$

where $\tilde{x_0}, \tilde{a}, \tilde{b}, \tilde{c}$ and $\tilde{d}$ are, respectively, the x_0, a, b, c and d defined in Theorem 4.2.1, obtained by setting $p_0 = [m/n]$, $p_1 = [(m-s+1)/m]$, and $p_2 = [(n-m-k+s)/(n-m)]$.

ρ	Case 1 ($n = 800$, $m = 100$,$k = 30$ and $s = 20$)		Case 2 ($n = 500$, $m = 50$, $k = 20$ and $s = 10$)	
	Simulation	Approximation	Simulation	Approximation
0.4	0.001	0.001	0.068	0.083
0.6	0.187	0.190	0.592	0.629
0.8	0.985	0.983	0.996	0.996

Table 4.1: The Desired Probability $\mathbf{P}(E)$ for the Standard Bivariate Normal Distribution

Example 4.4.1 (Bivariate Normal Distribution)**.** Suppose the distribution we are sampling from is bivariate normal given by:

$$(X, Y) \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

We will consider the following two cases: Case 1: $n = 800$, $m = 100$, $k = 30$ and $s = 20$; Case 2: $n = 500$, $m = 50$, $k = 20$ and $s = 10$. For each case, we calculate the $\mathbf{P}(E)$ with ρ being 0.4, 0.6 and 0.8. We calculate the "Exact" values for $\mathbf{P}(E)$ by crude Monte Carlo estimation, as well as the large sample approximation using formula (4.22). The results are given in Table 4.1. We observe that the large sample approximations for Case 1 work fairly well for different values of ρ; while in Case 2 we do see significant discrepancies between the large sample approximations and exact values for $\mathbf{P}(E)$, especially when the correlation between X and Y is small.

4.4.2 Power of Two-Stage Designs for Gene-Disease Association Studies

In gene-disease association studies with a large number of candidate markers, genotyping all the markers on all samples would be inefficient in resource utilization. Satagopan et al. (2002) and Satagopan et al. (2004) proposed a two-stage design which is shown to be more cost effective, while providing power[4] closer to that of one-stage designs. In their design all markers are evaluated on a fraction of available subjects at the first stage, and only the most promising markers selected at stage one are evaluated on the remaining subjects.

Suppose there are m candidate markers, and n available subjects in the sample. The two-stage design proposed by Satagopan et al. (2002) and Satagopan et al. (2004) works in the following way. At stage one, all the m markers are evaluated using only n_1 ($n_1 \leq n$) subjects, and let X_i be the resulting test statistic (for example the chi-square test statistic based on a 2×2 table) for marker i. We rank the m markers based on the test statistics X_i, and select the top k markers to go on to the second stage. In stage two, we evaluate the k markers using the remaining n_2 ($n_1 + n_2 = n$) subjects, and construct the test statistics Y_i based on the outcomes of both stages. Notice that at stage two we only obtain the test statistics for the markers with highest X values, which can be expressed as $(Y_{[m-k+1:m]}, \ldots, Y_{[m:m]})$. The setting is similar to the usual selection problems, but one big difference is that the random vectors (X_i, Y_i) are from one of the two bivariate distributions: one is corresponding to the disease-susceptibility markers, and the other corresponding to the null markers.

In Satagopan et al. (2002) and Satagopan et al. (2004), the power function for the proposed two-stage design is derived using a conditioning argument, and the authors are interested in the problem of how to optimally choose k and n_1 such that the two-stage design has a large gain in cost reduction while keeping the power close to that of the one-stage design involving the same number of subjects. Here we will derive an expression for the power of the two-stage design using concomitants of order statistics, as well as the results we obtained in the previous section.

[4]Here we define the power to be the probability that the true gene is selected at the end of the study.

As argued in Satagopan et al. (2002) and Satagopan et al. (2004), the test statistic for association computed from n independent subjects has an asymptotic normal distribution, $N(n\mu, n\sigma^2)$, where μ and σ^2 are asymptotic mean and variance respectively. The asymptotic mean μ will be zero if there is no association. Without loss of generality we can assume that $\sigma = 1$ by appropriate scaling of the statistic.

Let X^* be the test statistic from n_1 subjects obtained in stage 1 and let Y^* be the test statistic from the combined $n(= n_1 + n_2)$ subjects from stages 1 and 2 for the true gene. Similarly let X_i and Y_i be the corresponding test statistics for null gene i, $i = 1, \ldots, m-1$. If we assume that the gene outcomes are independent within a subject, then we have (X^*, Y^*) and (X_i, Y_i)'s are independently distributed as

$$(X^*, Y^*)^T \sim N\left((n_1\mu, n\mu)^T, \mathbf{\Sigma}\right)$$

$$(X_i, Y_i)^T \sim N\left((0, 0)^T, \mathbf{\Sigma}\right)$$

where

$$\mathbf{\Sigma} = \begin{pmatrix} n_1 & n_1 \\ n_1 & n \end{pmatrix}$$

The event of our interest, E, is that the true gene is selected at the end of the study, and can be expressed as:

$$E = \{X^* > X_{m-k:m-1}, Y^* > \max(Y_{[i:m-1]}, i = m-k+1, \ldots, m-1)\}$$

where $X_{i:m-1}$ is the ith order statistic of X_i, $i = 1, \ldots, m-1$, and $Y_{[i:m-1]}$ is the concomitant associated with $X_{i:m-1}$. So the power, P^*, can be derived as following:

$$\begin{aligned} P^* &= \mathbf{P}\{X^* > X_{m-k:m-1}, Y^* > V_{k-1:k-1}\} && (4.23) \\ &= \int_x \int_y \mathbf{P}(X_{m-k:m-1} < x, V_{k-1:k-1} < y) dF^*(x, y) && (4.24) \end{aligned}$$

where $V_{k-1:k-1}$ is the maximum of concomitants subset $(Y_{[i:m-1]}, i = m-k+1, \ldots, m-1)$, and $F^*(x, y)$ is the cdf for bivariate normal (X^*, Y^*).

Notice the integrand, $\mathbf{P}(X_{m-k:m-1} < x, V_{k-1:k-1} < y)$, in (4.24) can be expressed as

$$
\begin{aligned}
& \mathbf{P}(X_{m-k:m-1} < x, V_{k-1:k-1} < y) && (4.25)\\
= & \int_{-\infty}^{x} \mathbf{P}(V_{k-1:k-1} < y | X_{m-k:m-1} = u) f_{m-k:m-1}(u) du && (4.26)\\
= & \int_{-\infty}^{x} [F_1(y|u)]^{k-1} f_{m-k:m-1}(u) du && (4.27)
\end{aligned}
$$

where $f_{m-k:m-1}(u)$ is the pdf of the order statistic $X_{m-k:m-1}$; $F_1(\cdot|u)$ is the cdf of Y given $X > u$. In deriving (4.27), we use the fact that given $X_{m-k:m-1} = u$, $V_{k-1:k-1}$ behaves the same as the maximum of a random sample of size $k-1$ from the distribution $F_1(\cdot|u)$ as stated in Lemma 3.2.1.

CHAPTER 5

ESTIMATION OF THE REGRESSION FUNCTION AT A SELECTED QUANTILE OF THE EXPLANATORY VARIABLE

5.1 Introduction

Suppose (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d observations from some bivariate cdf $F(x, y)$. We want to estimate $\mathbf{E}[Y|X = F_X^{-1}(p)]$ for a given p, i.e., we want to estimate the conditional expectation of Y given X is at its given population quantile. This can be classified as a regression problem. But unlike the usual regression problem, the evaluation point $F_X^{-1}(p)$ is itself an unknown parameter. Such kind of problem can arise from the sensitivity analysis of Value-at-Risk (VaR), a popular risk measure based on the quantile of profit-and-loss distribution (Gourieroux et al., 2000; Mausser, 2001).

VaR is a popular and synthetic measure of risk widely used by financial institutions. It quantifies the risk of financial institutions' portfolios using the lower quantile of the distribution of the loss-and-profit of portfolios for a given time frame. For example, a 95% 1-week VaR of a given portfolio is just the 5% quantile of the loss-and-profit distribution of the portfolio over one week, and can be interpreted as the level of 1-week losses for the portfolio which will only be exceeded on average once every 20 weeks. The resulting VaR measure can be utilized by financial institutions to set up the capital reserve for losses of portfolios. For more detailed introduction please refer to Jorion (1997).

Recently there has been growing interest in studying the sensitivity of VaR with respect to a change of the portfolio positions, see for example, Gourieroux et al. (2000), Mausser (2001), and references therein. Suppose a portfolio is made up of n financial assets whose positions are given by a_i, $i = 1, \dots, n$. Then the value of the portfolio at time t is given by $V_t(\boldsymbol{a}) = \sum_{i=1}^{n} a_i p_{i,t}$ with $p_{i,t}$ being the price of financial asset i at time t. Then the level α VaR of the portfolio at time t is a constant, $\mathrm{VaR}_t(\boldsymbol{a}, \alpha)$, that satisfies the condition

$$\mathbf{P}_t(V_{t+1}(\boldsymbol{a}) - V_t(\boldsymbol{a}) + \mathrm{VaR}_t(\boldsymbol{a}, \alpha) < 0) = \alpha, \tag{5.1}$$

where $\mathbf{P}_t$ refers to the conditional distribution of future asset prices $p_{i,t+1}$ given all the information at time t. Notice by (5.1), $-\mathrm{VaR}_t(\boldsymbol{a}, \alpha)$ is just the lower α quantile of the distribution of the portfolio's loss-and-profit $(V_{t+1}(\boldsymbol{a}) - V_t(\boldsymbol{a}))$. The sensitivity of VaR is represented by the partial derivative of VaR with respect to portfolio allocations $\boldsymbol{a}$, that is

$$\frac{\partial \mathrm{VaR}_t(\boldsymbol{a}, \alpha)}{\partial \boldsymbol{a}}. \tag{5.2}$$

Gorieroux et al. (2000) showed that the partial derivative of (5.2) is given by:

$$\frac{\partial \mathrm{VaR}_t(\boldsymbol{a}, \alpha)}{\partial \boldsymbol{a}} = -\mathbf{E}[\boldsymbol{y_{t+1}} | \boldsymbol{a}' \boldsymbol{y_{t+1}} = -\mathrm{VaR}_t(\boldsymbol{a}, \alpha)] \tag{5.3}$$

where $\boldsymbol{y_{t+1}} = \boldsymbol{p_{t+1}} - \boldsymbol{p_t}$. Notice that the right-hand side of (5.3) is just the conditional expectation of price change given that the portfolio is at loss of the level equal to VaR at time t. Upon taking the loss-and-profit of the portfolio as the random variable X, and the price change, $y_{i,t+1}$, as the random variable Y, it follows that the sensitivity of level α VaR with respect to the financial asset i is just $\mathbf{E}(Y | X = F_X^{-1}(\alpha))$. So the problem of estimating the sensitivity of VaR reduces to the problem of estimation of the regression function given that the predictor is at a specified population quantile.

Since the problem can be thought of as a regression problem, a natural way to proceed is given by the following 2-step procedure:

Step 1: Obtain an estimator, $\hat{\xi}_p$, of the X population quantile $\xi_p = F_X^{-1}(p)$ based on the X sample. This can be done by using the sample quantile, or other quantile estimators proposed in the literature, for example, the Harrell-Davis estimator (Harrell and Davis, 1982), the Kaigh-Lachenbruch estimator (Kaigh and Lachenbruch, 1982), and the kernel quantile estimators (Yang, 1981; Sheather and Marron, 1990). Parrish (1990), and Dielman et al. (1994) compared the performance of different versions of quantile estimators by simulation studies.

Step 2: After we obtain the quantile estimator $\hat{\xi}_p$, we can plug it in to estimate the conditional mean of Y given X is $\hat{\xi}_p$ to yield the estimator, $\hat{m}(\hat{\xi}_p)$, for the quantity $\mathbf{E}[Y|X = F_X^{-1}(p)]$. To estimate the conditional mean function $m(\cdot)$, we can use the nonparametric regression methods suggested in the literature, for example, the kernel regression estimators (Nadaraya, 1964; Watson, 1964; Priestley and Chao, 1972; Gasser and Müller, 1979), nearest neighbor regression estimator (Benedetti, 1977; Stone, 1977), or the kernel-weighted local regression estimator proposed more recently (Fan, 1992, 1993; Ruppert and Wand, 1994). See monographs Müller and Muller (1988), Härdle (1990), and Fan and Gijbels (1996) for general introductions of these nonparametric regression methods.

The above 2-step procedure essentially combines the quantile estimation and the nonparametric regression estimation to yield the estimator of $\mathbf{E}[Y|X = F_X^{-1}(p)]$. Instead of using the above two-step procedure, we will consider two other types of estimators based on the concomitants of order statistics. This is the main purpose of this Chapter.

The first class of estimators is a kernel-type estimator, firstly proposed and studied by Yang (1981). Here we formalize the idea behind the estimator proposed by Yang (1981), and make extension to allow for other versions of kernel-type estimator based on concomitants of order statistics.

The second estimator is based on Mausser (2001), who studied the sensitivity of VaR for a portfolio with respect to its constituent positions. We generalize the estimator of the marginal VaR to have it being an estimator of $\mathbf{E}[Y|X = F_X^{-1}(p)]$, and we argue that the resulting estimator is essentially a bootstrap estimator of $\mathbf{E}(m(X_{k:n}))$ with $k = [(n+1)p]$, and $m(x)$ being the mean regression function $\mathbf{E}(Y|X = x)$, which in turn converges to $\mathbf{E}[Y|X = F_X^{-1}(p)]$ as $n \to \infty$.

In Section 5.2 and 5.3, the above two types of estimators are examined in detail, and their asymptotic properties are also studied. In Section 5.4 finite sample properties of these estimators are compared using Monte Carlo simulations.

5.2 Kernel-Type Estimators Based on Concomitants of Order Statistics

5.2.1 Motivation

Yang (1981) proposed the following class of L-estimators based on concomitants of order statistics for estimating $\mathbf{E}(Y|X = F_X^{-1}(p))$:

$$\hat{M}_Y = n^{-1} \sum_{i=1}^{n} h_n^{-1} K\left(\frac{i/n - p}{h_n}\right) Y_{[i:n]} \tag{5.4}$$

where K is some kernel function. Yang studied the estimator (5.4) as a linear combination of concomitants of order statistics, rather than a kernel smoothing estimator. But as we argue in the following, (5.4) can be thought of as a kernel smoothing estimator, where smoothing takes place in the transformed space $F_X(X)$, rather than the original X space. By doing this, other versions of kernel-type estimators can also be made available for estimating $\mathbf{E}(Y|X = F_X^{-1}(p))$.

To elaborate, we can think in the following way. Notice the quantity of our interest can be expressed as:

$$\mathbf{E}(Y|X = F_X^{-1}(p)) = \mathbf{E}(Y|F_X(X) = p). \tag{5.5}$$

With the assumption that X is continuous, $F_X(X)$ will be a uniform random variable in $(0, 1)$. So we can estimate the desired quantity using the transformed data (U_i, Y_i), with $U_i = F_X(X_i)$, $i = 1, \ldots, n$. The problem with the above argument is that the cdf $F_X(\cdot)$ is unknown, but we can estimate it with the empirical cdf. So finally our problem becomes estimating $\mathbf{E}(Y|F_X(X) = p)$ using the "data" (U_i^*, Y_i), with $U_i^* = F_n(X_i)$, $i = 1, \ldots, n$. After ordering the variate U_i^*, our transformed data can also be expressed as $(i/n, Y_{[i:n]}), i = 1, \ldots, n$, which corresponds to the so-called *equally-spaced fixed design* (Chu and Marron, 1991).

It can be easily seen that the estimator given by (5.4) can be derived by applying the Priestley-Chao kernel smoothing estimator (Priestley and Chao, 1972) to the transformed data. We can also apply the Gasser-Müller (Gasser and Müller, 1979) estimator to the transformed data to yield the following kernel-type estimator:

$$\hat{m}_{GM} = \sum_{i=1}^{n} \left[\int_{\frac{i-1/2}{n}}^{\frac{i+1/2}{n}} h_n^{-1} K\left(\frac{p-t}{h_n}\right) dt \right] Y_{[i:n]}. \tag{5.6}$$

5.2.2 Asymptotic Properties of the Kernel-type Estimators

In this subsection we will study the asymptotic properties of the kernel-type estimators based on concomitants of order statistics discussed above.

Treating the estimator $\hat{M}_Y$ given by (5.4) as a linear combination of concomitants of order statistics, Yang (1981) established several important results regarding the asymptotic properties of the estimator using Hájek's projection lemma (Hájek, 1968). Here we list these results without proofs.

Yang (1981) proved that under mild regularity assumptions the estimator $\hat{M}_Y$ is mean square consistent as given by the following proposition (Yang, 1981, Corollary 1):

Proposition 5.2.1. *Suppose the following conditions are satisfied:*

- $\mathbf{E}(Y^2) < \infty$*;*
- *the mean regression function* $m(x)$ *is a right continuous function of bounded variation in any finite interval;*
- *the kernel function* K *satisfies a Lipschitz condition in the sense that there exists a constant* M *such that* $|K(x_1) - K(x_2)| \leq M|x_1 - x_2|$ *for all* x_1 *and* x_2*; and* $|tK(t)| \to 0$ *as* $|t| \to \infty$*;*
- $h(n) \to 0$*, and* $n^{1/4}h(n) \to \infty$ *as* $n \to \infty$*.*

Then for any p_0 *at which* $m(\xi_p) = \mathbf{E}(Y|X = F_X^{-1}(p))$ *as a function of* p *is continuous,*

$$\lim_{n\to\infty} \mathbf{E}[\hat{M}_Y - m(\xi_{p_0})]^2 = 0. \tag{5.7}$$

With additional assumptions, Yang (1981) proved that the following result (Corollary 2 of Yang, 1981)) regarding the convergence rate of the bias of $\hat{M}_Y$.

Proposition 5.2.2. *Suppose the assumptions of Proposition 5.2.1 hold, and the following assumptions are also satisfied:*

- $\int_{-\infty}^{\infty} tK(t)dt = 0$*;*
- $\int_{-\infty}^{\infty} t^2 |K(t)|\, dt < \infty$ *and* $|t^3K(t)| \to 0$ *as* $|t| \to \infty$*;*

- *$K''(z)$ exists, satisfies a Lipschitz condition, and $\int_{-\infty}^{\infty} |K''(t)|\, dt < \infty$, $tK''(t) \to 0$ as $|t| \to \infty$;*

- *The second derivative of $g(p) = m(F_X^{-1}(p))$ exists and is continuous at p_0.*

Then we have

$$\lim_{n\to\infty} [\mathbf{E}(\hat{M}_Y) - m(\xi_{p_0})]/h(n)^2 = \frac{g''(p_0)}{2} \int_{-\infty}^{\infty} t^2 K(t)dt. \tag{5.8}$$

Next we will show that under mild assumptions $\hat{M}_Y$ and $\hat{M}_{GM}$ are asymptotically equivalent. So the asymptotic properties of $\hat{M}_Y$ will also hold for $\hat{M}_{GM}$.

Theorem 5.2.1. *Suppose the following assumptions hold:*

- *$\mathbf{E}(Y^2) < \infty$;*

- *$m(x)$ is right continuous function of bounded variation in any finite interval;*

- *the support of the kernel function K is compact in the sense that $K(x) = 0$ for $|x| > a$ where a is some finite positive real number;*

- *K satisfies a Lipschitz condition, i.e., there exists some constants $M > 0$ such that $|K(x_1) - K(x_2)| \le M\,|x_1 - x_2|$ for all x_1 and x_2 in the support of K.*

Then for sequence a_n such that $a_n \to \infty$ and $a_n/(h(n)n) \to 0$ as $n \to \infty$, we have

$$\mathbf{E}[\hat{M}_Y - \hat{M}_{GM}]^2 = o\left(\frac{1}{nh(n)^2 a_n^2}\right). \tag{5.9}$$

Proof. Note that

$$\begin{aligned} &\hat{M}_Y - \hat{M}_{GM} \\ = \;& \textstyle\sum_{i=1}^n \left[(nh(n))^{-1} K\left(\frac{i/n-p}{h(n)}\right) - h(n)^{-1} \int_{\frac{i-1/2}{n}}^{\frac{i+1/2}{n}} K\left(\frac{t-p}{h(n)}\right) dt\right] Y_{[i:n]} \\ = \;& [a_n h(n) n]^{-1} \textstyle\sum_{i=1}^n J_n(i/n) Y_{[i:n]}, \end{aligned} \tag{5.10}$$

where

$$J_n(u) = a_n \left[K\left(\frac{u-p}{h(n)}\right) - n\int_{u-\frac{1}{2n}}^{u+\frac{1}{2n}} K\left(\frac{t-p}{h(n)}\right) dt \right]. \tag{5.11}$$

There exists an interior point u_n in $(u+\frac{1}{2n}, u-\frac{1}{2n})$ such that

$$n\int_{u-\frac{1}{2n}}^{u+\frac{1}{2n}} K\left(\frac{t-p}{h(n)}\right) dt = K\left(\frac{u_n-p}{h(n)}\right). \tag{5.12}$$

So by the assumption on K we have

$$|J_n(u)| \leq a_n M \frac{|u-u_n|}{h(n)} \leq \frac{a_n M}{nh(n)}. \tag{5.13}$$

Hence by the assumption on a_n, we have $J_n(u) \to 0$ uniformly in n and p as $n \to \infty$.

Let $S_n = n^{-1}\sum_{i=1}^{n} J_n(i/n)Y_{[i:n]}$. Then from Theorem 1 of Yang (1981) and Remark 2 therein, we have

$$\lim_{n\to\infty} n^{1/2}\mathbf{E}(S_n) = 0, \tag{5.14}$$

and

$$\lim_{n\to\infty} n\text{Var}(S_n) = 0. \tag{5.15}$$

And notice

$$\mathbf{E}[\hat{M}_Y - \hat{M}_{GM}]^2 = [a_n h(n)]^{-2}\left(\text{Var}(S_n) + [\mathbf{E}(S_n)]^2\right), \tag{5.16}$$

so the desired result follows. □

5.2.3 Adjustment for the Boundary Effect

Sometimes we are interested in the conditional mean of Y evaluated at the upper tail of the X distribution, which means that the p in $\mathbf{E}(Y|X = F_X^{-1}(p))$ is very close to 0 or 1, say 0.05, as in the case of the estimation of Marginal VaR's. Then the kernel-type estimators given by (5.4) or (5.6) will be subject to larger bias due to the so-called *boundary effect*.

See Gasser and Müller (1979) and Rice (1984b) for more detailed discussions of boundary effects for the kernel smoothing method). Here we will use the method suggested by Rice (1984b) to modify the kernel estimators to reduce their boundary biases.

Suppose p is within a bandwidth of the lower boundary 0, i.e. $p = \tau h$ for some $\tau < 1$ with h being the chosen bandwidth. According to Rice's method (see also Section 4.4 of Härdle, 1990), the following estimator based on linear combination of two kernel smoothing estimators with different bandwidths can be used to reduce the boundary bias

$$\tilde{m}_h(p) = (1 - R)\hat{m}_{h,\tau}(p) + R\hat{m}_{\alpha h,\tau}(p) \tag{5.17}$$

where

$$R = \frac{\omega_K(1,\tau)/\omega_K(0,\tau)}{\alpha\omega_K(1,\tau/\alpha)/\omega_K(0,\tau/\alpha) - \omega_K(1,\tau)/\omega_K(0,\tau)} \tag{5.18}$$

with

$$\omega_K(k,\tau) = \int_{-1}^{\tau} u^k K(u)du, \qquad k = 0, 1.$$

Here K is the symmetric kernel function with support on $[-1, 1]$; $\hat{m}_{h,\tau}(\cdot)$ and $\hat{m}_{\alpha h,\tau}(\cdot)$ are the kernel estimators using the kernel $K_\tau(\cdot) = K(\cdot)/\omega_K(0,\tau)$ with bandwidth h and αh, respectively. For the constant α, Rice recommends choosing $\alpha = 2 - \tau$.

Remarks:

1. When $\tau \geq 1$, we have $\omega_K(0,\tau) = 1$ and $\omega_K(1,\tau) = 0$. Then R will be equal to 0, and the estimator given by (5.17) will reduce to the estimator given by (5.4) or (5.6), implying that no adjustment is needed for the interior point.

2. As pointed out by Härdle (1990), the modified kernel estimator can also be obtained by using the so-called "boundary kernel" given by

$$\tilde{K}_\tau(u) = (1 - R)K(u) - (R/\alpha)K(u/\alpha). \tag{5.19}$$

3. For p near the upper boundary 1, i.e., $p = 1 - \tau h$ for some $\tau < 1$, the modification is totally analogous, and the resulting estimator will still be given by (5.17), but with

$$\omega_K(k, \tau) = \int_{-\tau}^{1} u^k K(u) du, \qquad k = 0, 1.$$

5.2.4 Choice of Kernel and Bandwidth

As in all kernel smoothing methods, the problem of choosing appropriate kernel function and bandwidth is to be addressed before applying the estimator given by (5.4) or (5.6). Here we will discuss this issue very briefly without going further into the theoretical details. In the next section, we will examine this issue empirically with simulation studies.

It has been established in both theoretical and empirical settings that the choice of kernel functions in kernel smoothing methods is much less important than the choice of bandwidth, since one can always make the difference between two kernel smoothing estimates using two different kernels almost negligible by appropriately rescaling the bandwidths. A detailed discussion of this fact is available in Marron and Nolan (1988). So more emphasis has been placed on the issue of bandwidth selection in the kernel smoothing literature. In our discussion we will also focus on the bandwidth selection only, and choose the commonly used Epanechnikov kernel given by

$$K(u) = \frac{3}{4}(1 - u^2) I(|u| \leq 1) \tag{5.20}$$

in the empirical studies in the next section.

As noted above, the choice of bandwidth is of great importance in the kernel smoothing estimators $\hat{M}_Y$ and $\hat{M}_{GM}$. Both the bias and the variance of these estimators will depend on the bandwidth, which can be demonstrated as follows. From Proposition 5.2.2, we know that the asymptotic bias of the estimator $\hat{M}_Y$ is $O(h_n^2)$, which implies that the smaller the bandwidth h_n, the smaller the bias of this estimator. As shown in Yang (1981), the asymptotic variance of $\hat{M}_Y$ is $o(n^{-1}h_n^{-2})$. This implies that the smaller the bandwidth h_n, the larger the variance of the estimator $\hat{M}_Y$. So the choice of bandwidth turns out to be a balance of the trade-off between the bias and variance, which is common to smoothing parameter selection in all kernel smoothing methods.

As pointed out in Yang (1981), one possible approach to the problem is to choose the bandwidth h_n that minimizes the MSE, $\mathbf{E}(\hat{M}_Y - \mathbf{E}(Y|X = F_X^{-1}(p))^2$, of $\hat{M}_Y$. But that will be a difficult task in practice even when the distribution of X is known (Yang, 1981). Since the estimators $\hat{M}_Y$ and $\hat{M}_{GM}$, as we argued in Subsection 5.2.1, are essentially kernel smoothing regression estimators, we might borrow the existing methods of bandwidth selection for the kernel smoothing regression in the literature.

Various methods for choosing the smoothing parameter h_n in the setting of kernel smoothing regression have been proposed in the literature. Among them are Rice (1984a), Härdle et al. (1988), Vieu (1991), Brockmann et al. (1993), and Herrmann (1997). Herrmann (2000) gave an overview and a comparison of important and popular bandwidth selection methods in the context of kernel regression.

Since our primary interest is to estimate the conditional expectation of Y evaluated at a specified population quantile of X, rather than the whole regression function, more appropriate is the local bandwidth selection method to our problem. Other justification for using local bandwidth selection method are available in Vieu (1991) and Brockmann et al. (1993). As a result, we suggest to use the local adaptive bandwidth selection method proposed by Brockmann et al. (1993) for choosing h_n in the estimators $\hat{M}_Y$ and $\hat{M}_{GM}$. The method is essentially a local modification of the global iterative plug-in bandwidth selection method proposed by Gasser et al. (1991). It consists of the following iterative scheme (see Brockmann et al. (1993), for more details of this local bandwidth selection method), where we take $M = \int K^2(t)dt$, and $\mu_2 = \int K(t)t^2 dt$.

1. Let $\hat{h}_0 = n^{-1}$.

2. Iterate
$$\hat{h}_i = \left(\frac{\hat{\sigma}^2 M \int v(t)dt}{n\mu_2^2 \int v(t)\hat{m''}(t; \hat{h}_{i-1}n^{1/10}))^2 dt} \right)^{1/5} \quad \text{for} \quad i = 1, \ldots, 8. \tag{5.21}$$
where v is a weight function that is assumed to be twice continuously differentiable in the support $[\delta, 1-\delta]$ for some $\delta > 0$; $\hat{\sigma}^2$ is the estimator of σ^2 proposed by Gasser et al. (1986), and $\hat{m''}(\cdot)$ is the estimator of the second derivative of the regression function suggested by Gasser and Mueller (1984).

3. Iterate
$$\hat{h}_i(p) = \left(\frac{\hat{\sigma}^2 M}{\frac{n\mu_2^2}{\hat{h}_{i-1}(p)} \int K\left(u - p\hat{h}_{i-1}(p)\right) (\hat{m''}(u; \hat{h}_{i-1}(p)n^{1/10}))^2 du} \right)^{1/5} \quad \text{for} \quad i = 9, 10. \tag{5.22}$$

4. Let $\hat{h}(p) = \hat{h}_{10}(p)$ be the estimator for the optimal local bandwidth.

5.3 A Bootstrap Estimator Based on Concomitants of Order Statistics

5.3.1 Introduction to the Estimator

To estimate the marginal level p VaR of a position in a portfolio, which is defined to be the partial derivative of level p VaR with respect to the given position, Mausser (2001) proposed the following estimator:

$$\hat{M}_{HD} = \sum_{i=1}^{n} w_i Y_{[i:n]}, \tag{5.23}$$

where

$$
\begin{aligned}
w_i =& \frac{1}{\beta((n+1)p,(n+1)(1-p))} \int_{(i-1)/n}^{i/n} y^{(n+1)p-1}(1-y)^{(n+1)(1-p)-1} dy \\
=& I_{i/n}\{(n+1)p,(n+1)(1-p)\} - I_{(i-1)/n}\{(n+1)p,(n+1)(1-p)\}
\end{aligned}
\tag{5.24}
$$

with $I_x(a,b)$ being the incomplete beta function.

Since the marginal VaR of a position is just the conditional expected loss of the position given that the portfolio is at a loss of the VaR level as shown in Gorieroux et al. (2000), we might think of (5.23) as an estimator of the conditional mean of the response variable (which is the loss of the specific position here) evaluated at a given quantile of the predictor variable (which is the loss-and-profit of the portfolio here).

The motivation for (5.23) to be an estimator of $\mathbf{E}(Y|X = F_X^{-1}(p))$ is rather vague. But as argued below, we can think of the estimator given by (5.23) as a "Bootstrap" estimator of $\mathbf{E}(Y_{[k:n]}) = \mathbf{E}(m(X_{k:n}))$ with $k = [(n+1)p]$, and $m(x)$ being the regression function $\mathbf{E}(Y|X = x)$.

Notice

$$
\begin{aligned}
&\mathbf{E}(Y_{[k:n]}) = \mathbf{E}[\mathbf{E}(Y_{[k:n]}|X_{k:n})] \\
=&\mathbf{E}(m(X_{k:n})) \\
=&\int m(x) f_{r:n}(x) dx \\
=&\int m(x) \frac{n!}{(k-1)!(n-k)!}[F_X(x)]^{k-1}[1-F_X(x)]^{n-k} dF_X(x) \\
=&\frac{1}{\beta(k,n-k+1)} \int m(F_X^{-1}(u)) u^{k-1}(1-u)^{n-k} du.
\end{aligned}
\tag{5.25}
$$

Now if we substitute $F_X^{-1}(\cdot)$ by the sample quantile function, $\tilde{F}_X^{-1}(\cdot)$, which is defined to be

$$
\tilde{F}_X^{-1}(p) = \inf\{x|F_n(x) \geq p\} \tag{5.26}
$$

with $F_n(x) = \sum I(X_i \leq x)/n$ being the sample cdf, we get the following estimator of $\mathbf{E}(Y_{[k:n]})$ for $k = [(n+1)p]$:

$$\sum_{i=1}^{n} w_i m(X_{i:n}) \tag{5.27}$$

Finally if we approximate $m(X_{i:n})$ by $Y_{[i:n]}$, then the estimator given by (5.27) will reduce to the proposed "bootstrap" estimator given by (5.23). As a result, the estimator $\hat{M}_{HD}$ given by (5.23) can be thought of as a bootstrap estimator of $\mathbf{E}(Y_{[k:n]})$ with $k = [(n+1)p]$.

Notice $X_{k:n} \xrightarrow{P} \xi_X(p)$ as $n \to \infty$ with $k = [(n+1)p]$. So under appropriate conditions on the regression function m, we will have $\mathbf{E}(Y_{[k:n]}) \to m(\xi_X(p))$, for example m is a bounded function which is continuous at $\xi_X(p)$, or more generally m is continuous and $m(X_{k:n})$ is uniformly integrable, see more detailed discussions in Vaart (1998, Section 2.5).

5.3.2 Asymptotic Equivalence of the Bootstrap Estimator and the Kernel-type Estimators

In this subsection we will show that in large samples the estimator $\hat{M}_{HD}$ is essentially the same as the kernel-type estimator $\hat{M}_{GM}$ for appropriate choice of kernel function K.

We need the following Lemma due to Sheather and Marron (1990).

Lemma 5.3.1. *Let $q = 1 - p$ where $0 < p < 1$, and $m = n + O(1)$. Then as $n \to \infty$,*

$$\frac{\Gamma(np+mq)}{\Gamma(np)\Gamma(mq)} x^{np-1}(1-x)^{mq-1} \approx [2\pi pq/n]^{-1/2} \exp[-\frac{n(x-p)^2}{2pq}], \tag{5.28}$$

in the sense that

$$\frac{\Gamma(np+mq)}{\Gamma(np)\Gamma(mq)} [p + (pq/n)^{1/2}y]^{np-1}[q - (pq/n)^{1/2}y]^{mq-1}(pq/n)^{1/2} = \phi(y) + O(n^{-1/2}), \tag{5.29}$$

where $\phi(\cdot)$ is the pdf of the standard normal distribution.

Remark 5.3.1. Note that the result of Lemma 5.3.1 is directly related to the convergence of the pdf of a central order statistic from the standard uniform distribution, see Reiss (1989, Section 4.7) for more discussions about expansions of densities of central order statistics.

Then we have the following Theorem regarding the asymptotical equivalence of $\hat{M}_{HD}$ and $\hat{M}_{GM}$.

Theorem 5.3.1. *The estimator $\hat{M}_{HD}$ is asymptotically equivalent to the estimator $\hat{M}_{GM}$ with K being the standard normal density and $h(n) = [pq/(n+1)]^{1/2}$.*

Proof. Let $h(n) = [pq/(n+1)]^{1/2}$. Notice by Lemma 5.3.1 we can express the weights w_i in the estimator $\hat{M}_{HD}$ as:

$$\begin{aligned} w_i =& \frac{1}{\beta((n+1)p, (n+1)(1-p))} \int_{(i-1)/n}^{i/n} u^{(n+1)p-1}(1-u)^{(n+1)(1-p)-1} du \\ =& \int_{(i-1)/n}^{i/n} \left[(2\pi h_n^2)^{-1/2} \exp\left\{ -\frac{(u-p)^2}{2h_n^2} \right\} + O(n^{-1/2}) \right] du \\ =& \int_{(i-1)/n}^{i/n} h_n^{-1} \phi\left(\frac{u-p}{h_n} \right) du + O(n^{-3/2}). \end{aligned} \tag{5.30}$$

So we have

$$\begin{aligned} \hat{M}_{HD} =& \sum_{i=1}^{n} \int_{(i-1)/n}^{i/n} h_n^{-1} \phi\left(\frac{u-p}{h_n} \right) du + O(n^{-1/2}) \\ =& \hat{M}_{GM} + O(n^{-1/2}) \end{aligned} \tag{5.31}$$

where $\hat{M}_{GM}$ is the kernel estimator given by (5.6) using the gaussian kernel and bandwidth $h_n = [pq/(n+1)]^{1/2}$. This proves the theorem. $\square$

5.4 Numerical Examples

In this section we will study the finite sample behavior of the estimators discussed in Section 5.2 and 5.3 using simulations. The following three data generating processes are used for the simulation studies, where ε and X are independent.

(i) A random sample of size n is simulated from the model

$$Y = X + X\sin(\pi X) + \varepsilon \tag{5.32}$$

with $\varepsilon \sim N(0, 0.25)$, $X \sim N(0, 1)$.

(ii) A random sample of size n is simulated from the model

$$Y = X + X\sin(\pi X) + \varepsilon \tag{5.33}$$

with $\varepsilon \sim N(0, 1)$, $X \sim N(0, 1)$.

(iii) A random sample of size n is simulated from the model

$$Y = X + X\sin(\pi X) + \varepsilon \tag{5.34}$$

with $\varepsilon \sim N(0, 0.25)$, $X \sim 0.25N(-1.5, 1.5^2) + 0.75N(0, 1)$.

All these three data generating processes share the same mean regression function, but they differ from each other in terms of the magnitude of the noise inherent in the observed data, and the distribution of the covariate X. In particular, the data generating process (ii) has larger error term variance than processes (i) and (iii), implying greater noise in the observed data for process (ii); the distribution of covariate X in process (iii) is a mixture of $N(-1.5, 1.5^2)$ and $N(0, 1)$ which has heavier left tails than the standard normal in (i) and (ii).

5.4.1 Comparison of the Estimators by MSE

For each data generating process, we simulate random samples of size $n = 100, 200, 400$ and 900 respectively, and then calculate the estimators $\hat{M}_Y$, $\hat{M}_{GM}$, and $\hat{M}_{HD}$ of $\mathbf{E}(Y|X = \xi_X(p))$ for $p = 0.1, \ldots, 0.9$. For the kernel smoothing estimators $\hat{M}_{GM}$ and $\hat{M}_{HD}$, we used the R function "lokerns" to calculate the data adaptive bandwidth. The MSE's of these estimators for each data generating process are estimated based on 10,000 simulations.

Figure 5.1 gives the estimated MSE's of the three estimators for the data generating process (i). We observe that these three estimators perform essentially the same in terms of the MSE, especially for the two kernel smoothing estimators of which we see very little difference. But as the sample size increases, the kernel smoothing estimators perform a little better than the bootstrap estimator $\hat{M}_{HD}$. Also in the plots, we see that all these estimators have worse performance near the boundary region than in the interior region.

To examine the performance of these estimators more closely, we also compare the biases and standard deviations of these estimators.[5] Figure 5.2 and 5.3 give the biases and standard deviations of the three estimators for data generating process (i), respectively, and Figure 5.4 gives the plot of the ratio of the square of the bias to the MSE for these three estimators. Very similar to the plots for the MSE, we observe no significant difference between these three estimators in terms of the bias and standard deviation. Also for all these three estimators, the bias is dominated by the variance in terms of the contribution to the MSE, and the improvement in the bias is larger than that for the variance with the increase of sample size.

[5]It is well-known that the MSE can be decomposed into the variance and the square of the bias. By examining the bias and standard deviation we can know more about the contributions of these two parts to the MSE.

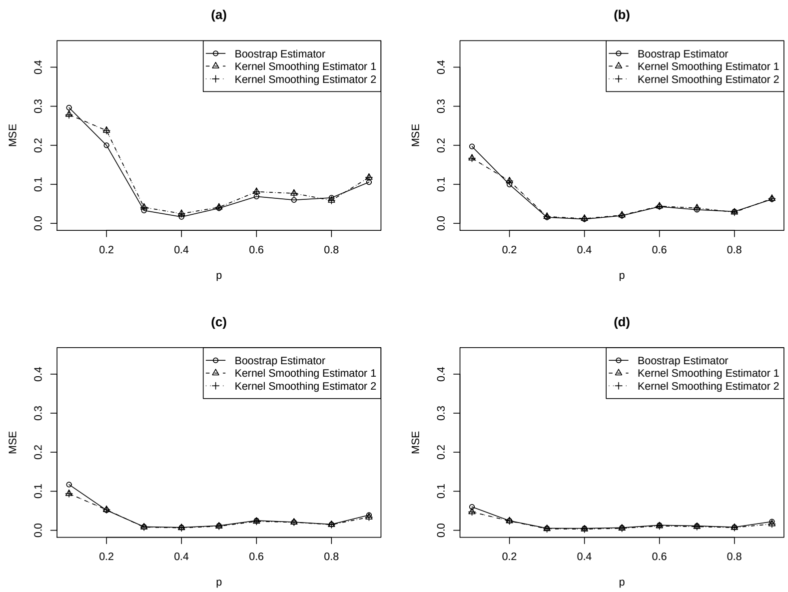

Figure 5.1: MSE of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (i): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

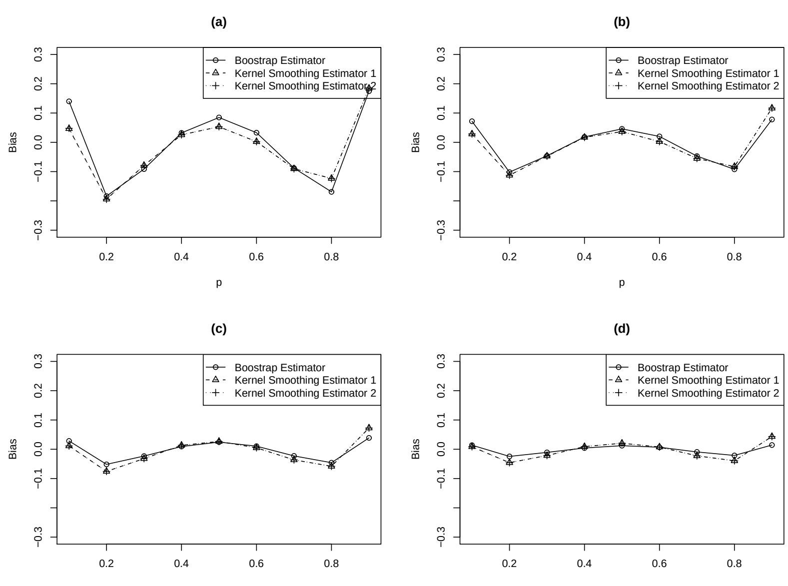

Figure 5.2: Bias of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (i): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

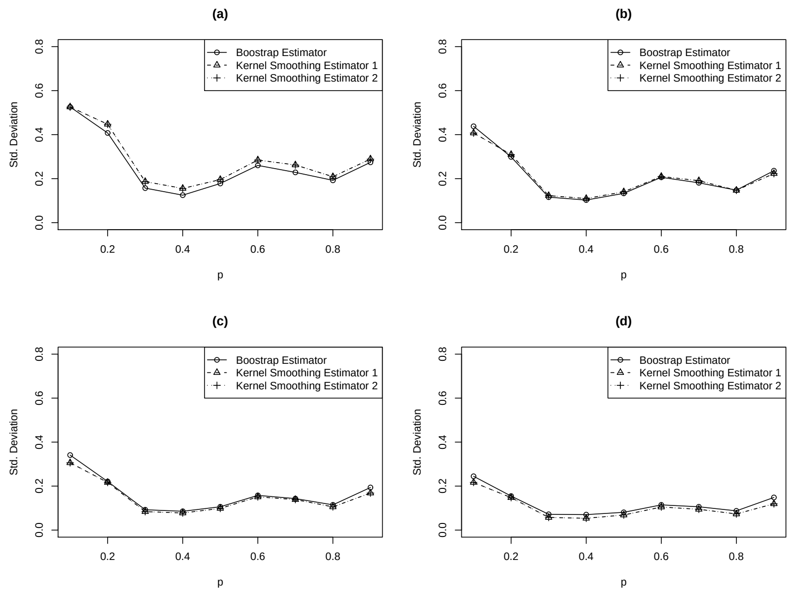

Figure 5.3: Standard Deviation of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (i): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

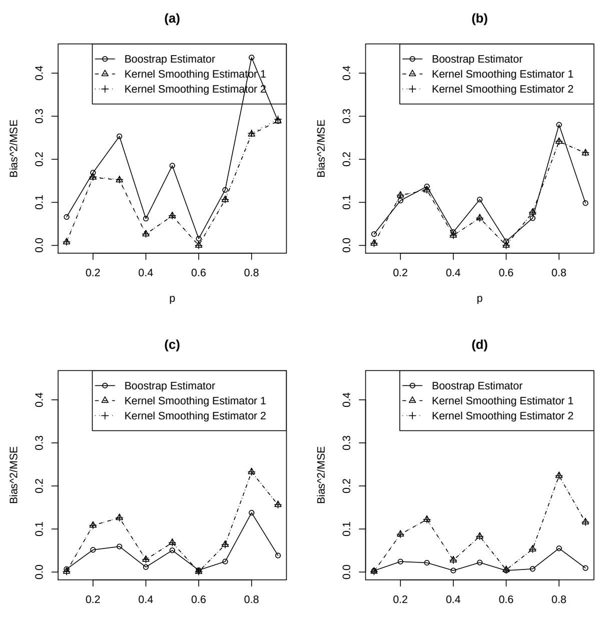

Figure 5.4: The Ratio of the Square of the Bias to the MSE for the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (i): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

The plots of MSE, bias, standard deviation and the ratio of the square of the bias to the MSE of the three estimators for the data generating process (ii), which produces more noisy data than the process (i), are given in Figure 5.5, 5.6, 5.7 and 5.8, respectively. From the plots we see again that the two kernel smoothing estimators are almost the same in the performance. For small sample size $n = 100$, the bootstrap estimator $\hat{M}_{HD}$ performs a little bit better than the two kernel smoothing estimators. But with the increase of the sample size, the two kernel smoothing estimators outperform the bootstrap estimator, and the discrepancy increases as the sample size increases. Similar to what we observe in data generating process (i), the variance dominates the bias in terms of contribution to the MSE, especially for the bootstrap estimator.

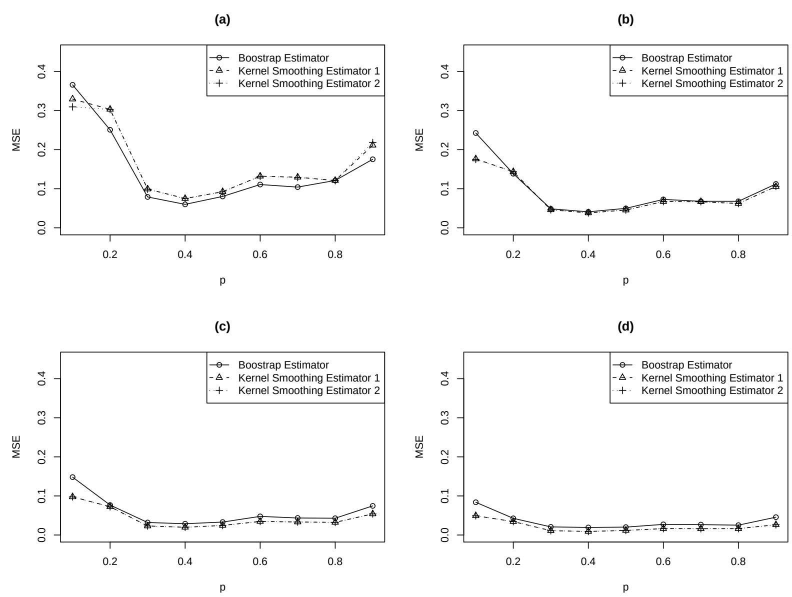

Figure 5.5: MSE of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (ii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

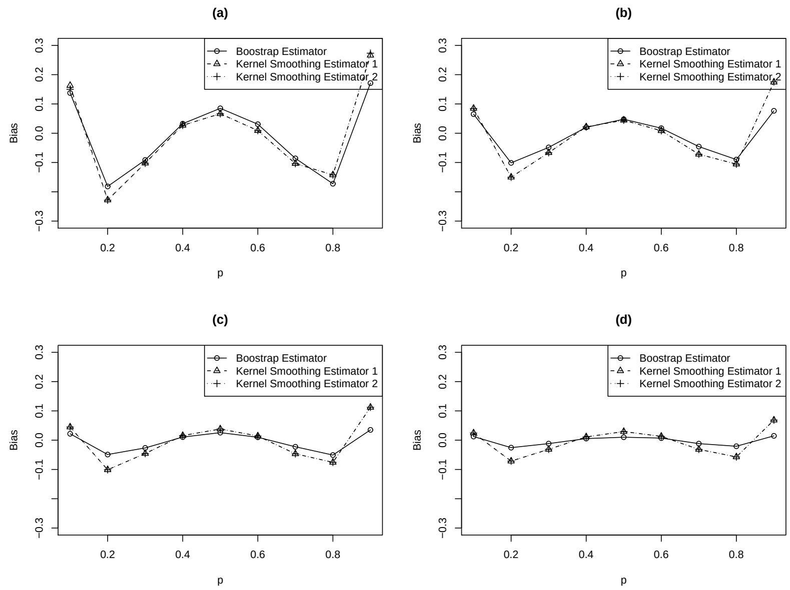

Figure 5.6: Bias of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (ii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

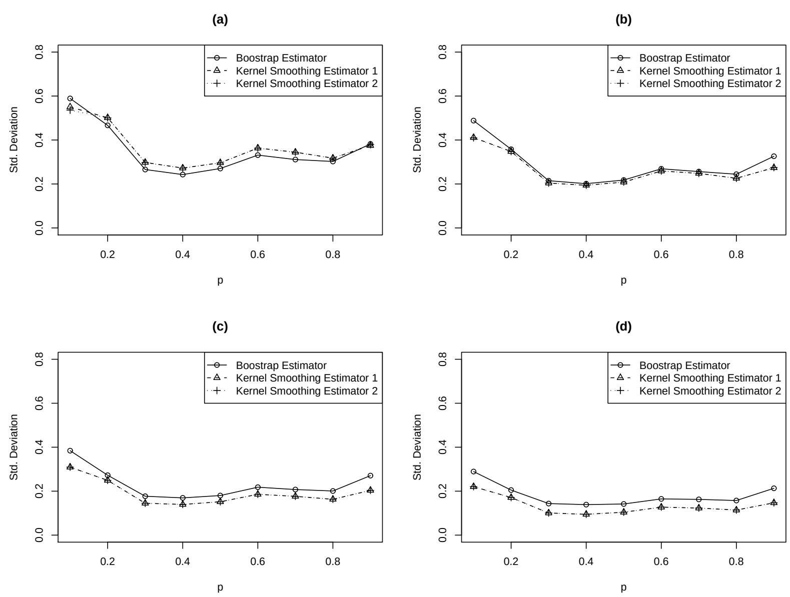

Figure 5.7: Standard Deviation of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (ii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

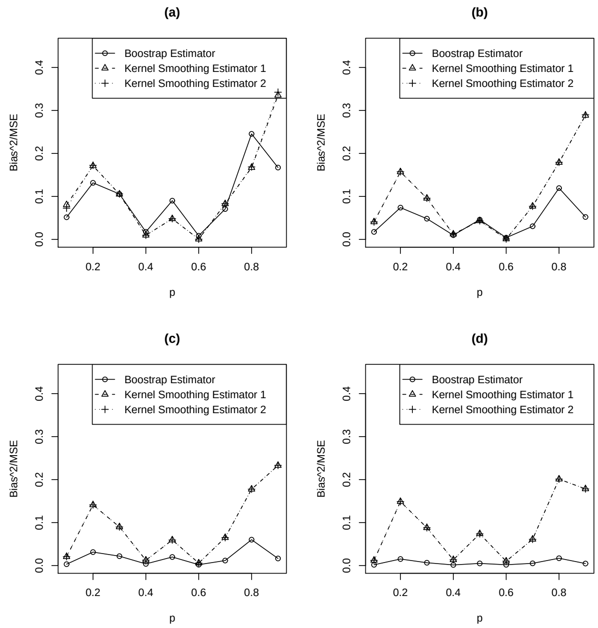

Figure 5.8: The Ratio of the Square of the Bias to the MSE for the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (ii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

Figure 5.9, 5.10, 5.11 and 5.12 present the estimated MSE, bias, standard deviation and the ratio of the square of the bias to the MSE of the three estimators for the data generating process (iii), in which the distribution of the covariate X has heavier left tails. We observe that the bootstrap estimator $\hat{M}_{HD}$ performs a little better than the other two kernel smoothing estimators for the sample size 100. But with the increase of sample size, the difference between these estimators diminishes. Again we observe that for all the three estimators the variance dominates the bias in its contribution to the MSE, and with the increase of sample size, the bias decreases much faster than the standard deviation.

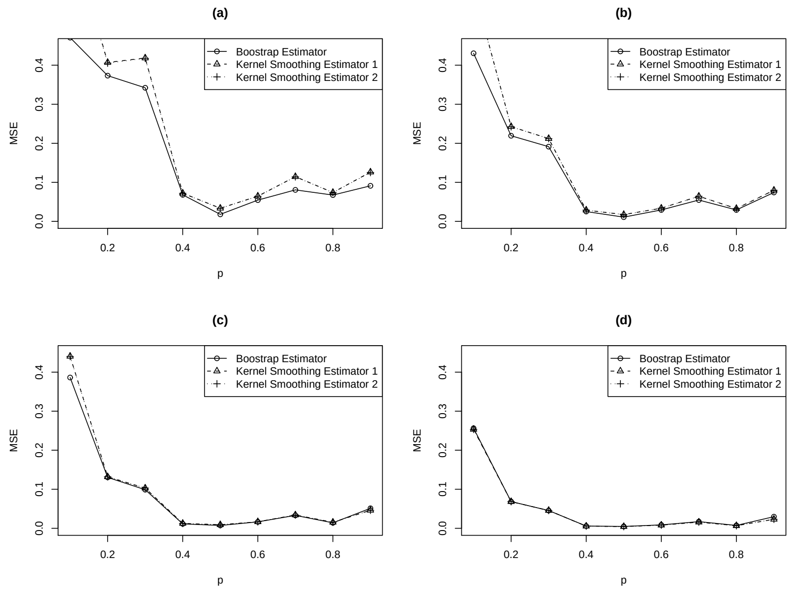

Figure 5.9: MSE of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (iii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

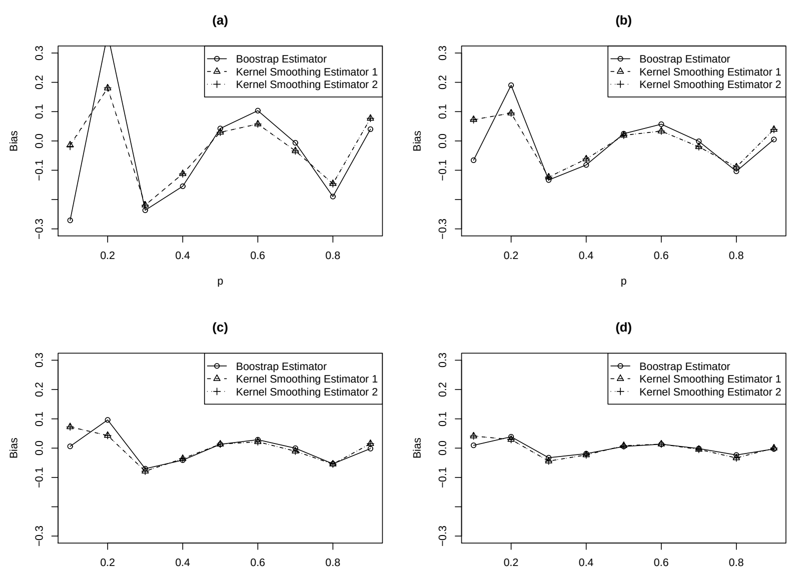

Figure 5.10: Bias of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (iii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

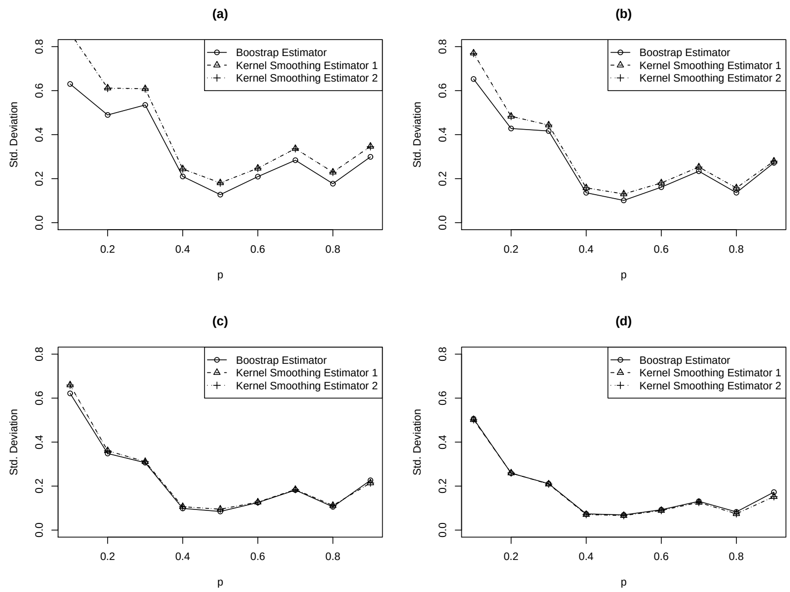

Figure 5.11: Standard Deviation of the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (iii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

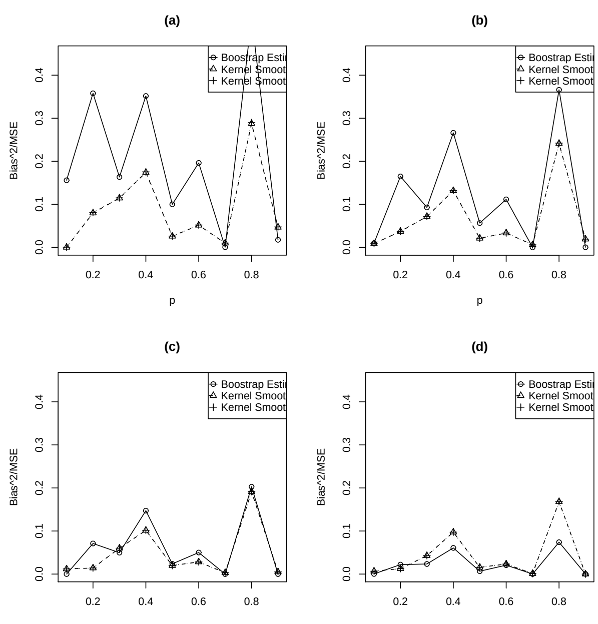

Figure 5.12: The Ratio of the Square of the Bias to the MSE for the Estimators of $\mathbf{E}(Y|X = \xi_X(p))$ for the Data Generating Process (iii): (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$

5.4.2 Bandwidth Selection

In Section 5.2.4, we suggested the use of the local adaptive bandwidth selection method proposed by Brockmann et al. (1993) for the choice of bandwidth in the kernel smoothing estimators $\hat{M}_Y$ and $\hat{M}_{GM}$. Here we will study the performance of this data-driven iterative bandwidth selection method numerically.

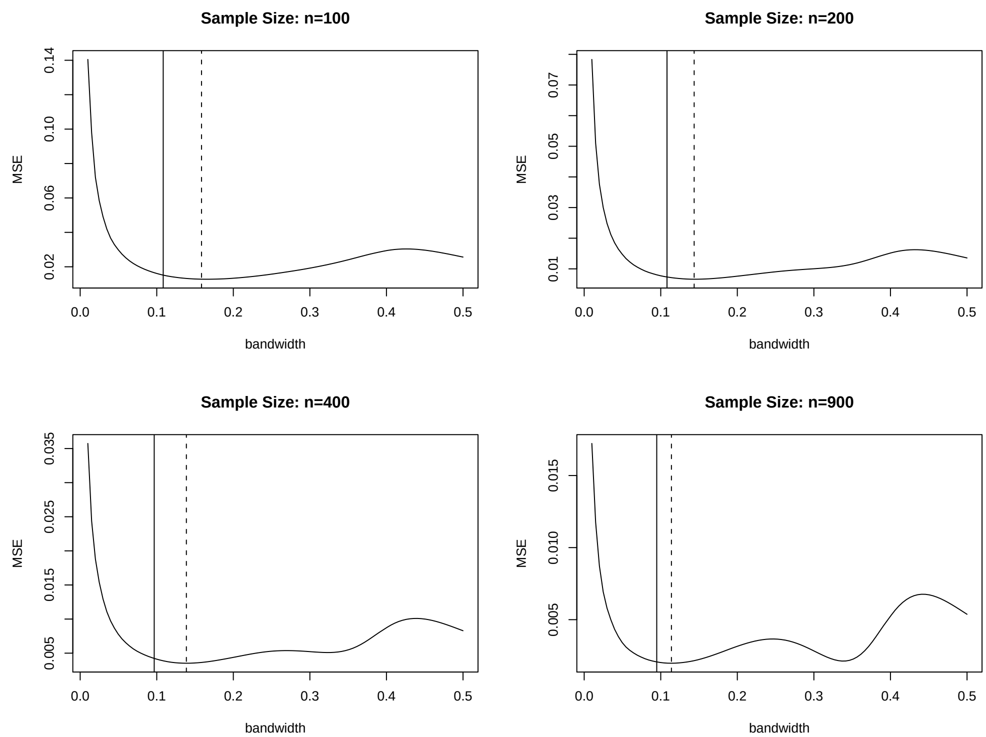

Figure 5.13: MSE's of the Estimator $\hat{M}_{GM}$ of $\mathbf{E}(Y|X = \xi_X(p))$ with $p = 0.4$ for different bandwidth values: (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$. Solid (dashed) vertical line locates the estimated (actual) optimal bandwidth.

Figure 5.13 gives the estimated MSE's of the estimator $\hat{M}_{GM}$ of the quantity $\mathbf{E}(Y|X = \xi_X(p))$ with $p = 0.4$ for different bandwidth values based on 1000 simulations. The data are generated according to the process (i) given by (5.32), and we do these simulations for sample sizes $n = 100$, 200, 400 and 900. In these plots, the solid lines give the estimated optimal bandwidth values by the local adaptive method, and the dashed lines give the actual optimal bandwidths based on our simulations. We observe that the estimated optimal bandwidths by the local adaptive method are very close to the actual optimal bandwidths, suggesting good performance of the local adaptive bandwidth selection method for the kernel smoothing estimators $\hat{M}_Y$ and $\hat{M}_{GM}$. We carry out these simulations for $p = 0.8$, which is close to the upper boundary 1. Figure 5.14 gives the corresponding results. Again we see that the estimated optimal bandwidths are close to the actual optimal bandwidths based on simulations.

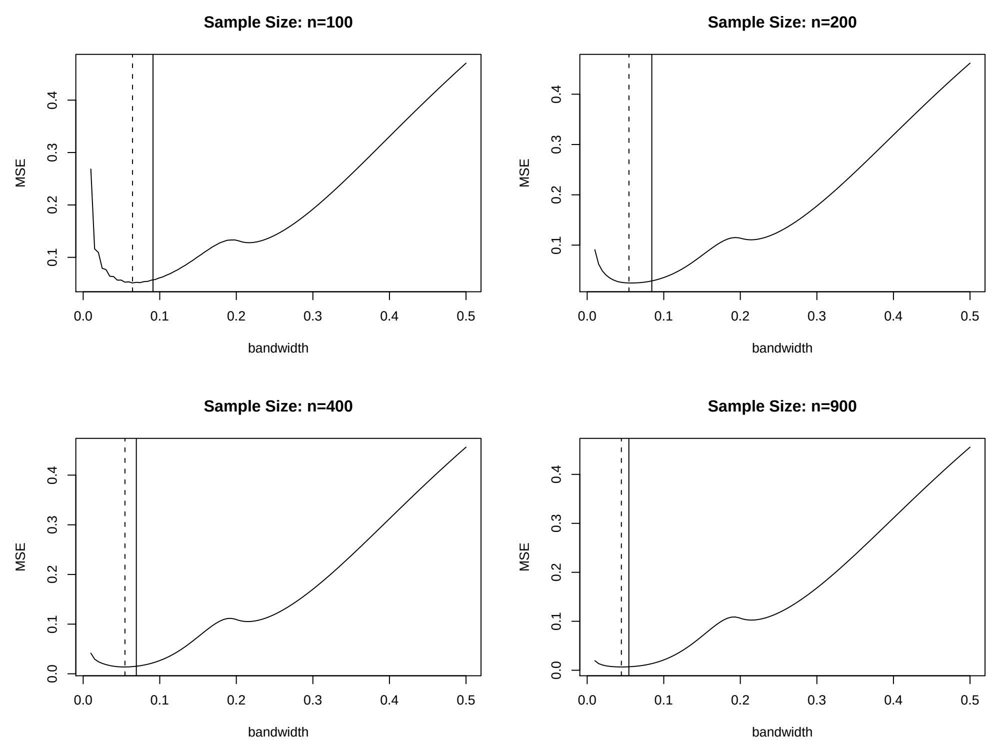

Figure 5.14: MSE's of the Estimator, $\hat{M}_{GM}$, of $\mathbf{E}(Y|X = \xi_X(p))$ with $p = 0.8$ for different bandwidth values: (a) $n = 100$; (b) $n = 200$; (c) $n = 400$; (d) $n = 900$. Solid (dashed) vertical line locates the estimated (actual) optimal bandwidth.

5.4.3 Boundary Adjustment

As pointed out in Subsection 5.2.3, the kernel smoothing estimators $\hat{M}_Y$ and $\hat{M}_{GM}$ are subject to larger bias at the lower or upper boundaries, which can also be seen in Figure 5.1. We suggested the use of the jackknife estimator proposed by Rice (1984b) to reduce the boundary bias. In this subsection we will examine this boundary adjustment numerically.

We consider the lower boundary behavior of these estimators. In particular, we estimate the MSE's of the estimator $\hat{M}_{GM}$ for $\mathbf{E}(Y|X = \xi_X(0.05))$ with and without boundary adjustment given by (5.17). The simulated data is generated according to the data generating process (i), and we choose the bandwidth to be 0.1 for the estimator $\hat{M}_{GM}$. The MSE's are estimated based on 1000 simulations. The MSE of the estimator $\hat{M}_{HD}$ is also estimated. Table 5.4.3 gives the results for different sample sizes.

n	$\hat{M}_{GM}$ without adjustment	$\hat{M}_{GM}$ with adjustment	$\hat{M}_{HD}$
100	1.592	1.557	0.551
200	1.441	1.416	0.205
400	1.394	1.373	0.069
900	1.332	1.311	0.022

Table 5.1: MSE's of the Estimators for $\mathbf{E}(Y|X = \xi_X(p))$ with $p = 0.05$

From Table 5.4.3 we observe that there is some improvement for the estimator $\hat{M}_{GM}$ with the boundary adjustment as given by (5.17). But the gain in the MSE is not very significant. Also we notice that the bootstrap estimator $\hat{M}_{HD}$ gives much better performance than the kernel smoothing estimator even with the boundary adjustment.

CHAPTER 6

SUMMARY AND FUTURE WORK

In the previous chapters we have studied several topics related to concomitants of order statistics. Here we summarize our findings, and discuss some future work related to these studies.

In Chapter 2 we studied the finite sample and asymptotic distributions of concomitants of order statistics for dependent samples. This extends the available results on the distribution theory of concomitants of order statistics in the literature, which usually assumed i.i.d or independent samples. In deriving these distributions, we model the dependence structure in X sample values by an equally correlated multivariate normal model, and the dependence of Y on X through the usual general regression model. A possible application of these results is to quantify theoretically the inferential biases associated with induced ordering in the F-tests of financial asset pricing models.[6]

In Chapters 3 and 4, we studied the distribution of order statistics of concomitants subsets associated with higher and lower order statistics, namely $V_{s:m}$ and $W_{t:n-m}$ in our notation. The motivation for studying these order statistics stems from an event of interest in selection procedures.

[6]See Lo and Mckinlay (1990) for more details about the problem of data-snooping when the data used for testing the financial asset pricing model is obtained with induced ordering.

In Chapter 3, the marginal distribution of these order statistics are studied. Using conditioning argument, we first derived the finite-sample distribution of $V_{s:m}$ and $W_{t:n-m}$, which turns out to be quite complicated for practical applications. Then a large sample approximation to these distribution is considered. Under appropriate assumptions, we established that the limiting distributions for $V_{s:m}$ and $W_{t:n-m}$ are both normal in the quantile case. The rates of convergence in these distributions are also derived, and are shown to be of order $n^{-1/2}$. Based on the results from Reiss (1989), we propose a higher order expansion to the cdfs of these order statistics, which achieves much better performance than the normal approximation even for moderate sample size. The results are illustrated with some numerical examples.

In Chapter 4 we extended the results in Chapter 3 by deriving the joint distribution of $(V_{s:m}, W_{t:n-m})$. With some minor adjustment to the arguments used in Chapter 3, we first derived the finite-sample joint distribution of $(V_{s:m}, W_{t:n-m})$; then we derived the asymptotic distribution of these order statistics in the quantile case, which turns out to be multivariate normal. Similar to the case of the marginal distribution, the rate of convergence in the joint distribution is shown to be of order $n^{-1/2}$. Using arguments similar to those in Joshi and Nagaraja (1995), we derived the asymptotic joint distribution of $(V_{s:m}, W_{t:n-m})$ in the extremal case under appropriate conditions. Unlike in the quantile case, $V_{s:m}$ and $W_{t:n-m}$ are asymptotically independent. We illustrated these results with a bivariate normal example.

We then discussed two applications of the results about the joint distribution of $(V_{s:m}, W_{t:n-m})$. The first application is about calculating the probability of an event of interest in selection procedures, which is essentially a generalization of the problem discussed in Yeo and David (1984). Although it is quite difficult to calculate the exact probability of the event of interest, we provide a simple formula for approximating the desired probability based on the asymptotic joint distribution of $(V_{s:m}, W_{t:n-m})$. The approximation is studied with some numerical examples. The other application is about calculating the power of a two-stage design for gene-disease association studies proposed by Satagopan et al. (2002) and Satagopan et al. (2004). We demonstrated that the desired (asymptotic) power function can be compactly expressed using order statistics of concomitant-subsets.

Along this line of research on order statistics of concomitant-subsets, there are two possible areas of work to consider in the future. Although we derived a higher order approximation to the marginal distribution of $V_{s:m}$ and $W_{t:n-m}$, the corresponding approximation to the joint distribution has not been derived yet. While this would be very helpful to approximating probabilities involved in selection procedure problems as discussed in Section 4.4.1. We have seen from the numerical examples that the normal approximation to the desired probability is not satisfactory even for sample size $n = 500$ in some cases.

Another challenging task is to derive the rate of convergence in the distribution of order statistics of concomitant-subsets in the extremal case, which in turn determines how useful the corresponding asymptotic distribution would be in applications. Deriving the rate of convergence in the distribution of extremal order statistics is itself a very difficult task, which is shown to depend on the second-order behavior of the population distribution (Smith, 1982; Haan, 2006). Establishing the rate of convergence in the distribution of order statistics of concomitant-subsets in the extremal case is much more complicated since we need to deal with the much messier distributions $F_1(\cdot|x_0)$ and $F_2(\cdot|x_0)$.

Motivated by the problem of estimating the marginal VaR of a position in a portfolio in financial risk management applications, we studied the problem of estimating the conditional mean of Y variable given that the X variable is at specific quantiles of its distribution in Chapter 5. The problem is very similar to the usual bivariate regression except that the evaluation point for the explanatory variable, its population quantile, is itself unknown. A natural approach to the problem is to use a 2-step procedure: the population quantile of X is first estimated; then the conditional mean of Y is estimated at the estimated X quantile. But here we propose two estimators based on the concomitants of order statistics, which circumvent the step for estimating the population quantile of X.

The first class of estimators, generalized from Yang (1981), can be thought of as a kernel smoothing estimator applied to the transformed data. We studied the asymptotic properties of these kernel smoothing estimators. It is shown that under mild conditions, these estimators are mean square consistent. The issues of bandwidth selection and boundary effect adjustment for these kernel smoothing estimators are also discussed. We recommended to use the local adaptive bandwidth selection method proposed by Brockmann et al. (1993) from the consideration that our primary interest is to estimate the conditional expectation of Y evaluated at a specified population quantile of X rather than the whole regression function. Simulation studies showed that this data-driven adaptive bandwidth selection method performs well for these kernel smoothing estimators in identifying the optimal bandwidth. To alleviate the boundary effect for these kernel smoothing estimators, we suggest the use of the method proposed by Rice (1984b). While simulation studies showed that the improvement in the MSE by adopting the given boundary adjustment is not very significant. One possible reason is that by applying the empirical cdf on the X values, the resulting transformed data are no longer independent. So the expansion of MSE in deriving the adjustment by Rice (1984b) do not directly carry over to our case. More rigorous study of how the MSE of our kernel smoothing estimator behaves in boundary regions is thus needed for finding more appropriate boundary adjustment.

The second estimator is motivated by Mausser (2001). We argued that the estimator can be thought of as a bootstrap estimator of a quantity which in turn approaches the parameter of our interest as the sample size goes to infinity. Under mild conditions we established that the estimator is asymptotically equivalent to the kernel smoothing estimators discussed above. So the nice asymptotic properties of the kernel smoothing estimators carry over to the bootstrap estimator.

The finite-sample behavior of the above two estimators were studied by simulations. We observed that although most of the time the performance of these estimators is very close to each other, the kernel smoothing estimator performs a little better than the bootstrap estimator for noisy data, while the bootstrap estimator performs better than the kernel smoothing estimator when the distribution of X is heavy-tailed. We also noticed that with the increase of sample size, the bias of these estimators diminishes much faster than the variance. Simulation studies also showed that in the boundary region the bootstrap estimator performs much better than the kernel smoothing estimators, even those with boundary adjustment.

BIBLIOGRAPHY

Arnold, B. and Balakrishnan, N. (1989), *Relations, Bounds and Approximations for Order Statistics*, Springer, New York.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992), *A First Course in Order Statistics*, John Wiley & Sons, New York.

Balakrishnan, N. (1993), "Multivariate Normal Distribution and Multivariate Order Statistics Induced by Ordering Linear Combinations," *Statistics & Probability Letters*, 17, 343–350.

Balasubramanian, K. and Beg, M. I. (1998), "Concomitant of Order Statistics in Gumbel's Bivariate Exponential Distribution," *Sankhyā, Series B*, 60, 399–406.

Barnett, V., Green, P. J., and Robinson, A. (1976), "Concomitants and Correlation Estimates," *Biometrika*, 63, 323–328.

Benedetti, J. K. (1977), "On the Nonparametric Estimation of Regression Functions," *Journal of the Royal Statistical Society, Series B: Methodological*, 39, 248–253.

Bhattacharya, P. K. (1974), "Convergence of Sample Paths of Normalized Sums of Induced Order Statistics," *The Annals of Statistics*, 2, 1034–1039.

— (1984), "Induced Order Statistics: Theory and Applications," in *Handbook of Statistics Volume 4: Nonparametric Methods*, eds. Krishnaiah, P. R. and Sen, P. K., Elsevier, Amsterdam, pp. 383–403.

Bloch, D. A. and Gastwirth, J. L. (1968), "On a Simple Estimate of the Reciprocal of the Density Function," *The Annals of Mathematical Statistics*, 39, 1083–1085.

Brockmann, M., Gasser, T., and Herrmann, E. (1993), "Locally Adaptive Bandwidth Choice for Kernel Regression Estimators," *Journal of the American Statistical Association*, 88, 1302–1309.

Chen, H., Huang, M.-N. L., and Huang, W.-J. (1996), "Estimation of the Location of the Maximum of a Regression Function Using Extreme Order Statistics," *Journal of Multivariate Analysis*, 57, 191–214.

Chen, Z., Bai, Z., and Sinha, B. K. (2004), *Ranked Set Sampling : Theory and Applications*, Springer, New York.

Chu, C.-K. and Marron, J. S. (1991), "Choosing a Kernel Regression Estimator (Disc: P419-436)," *Statistical Science*, 6, 404–419.

Chu, S.-J., Huang, W.-J., and Chen, H. (1999), "A Study of Asymptotic Distributions of Concomitants of Certain Order Statistics," *Statistica Sinica*, 9, 811–830.

David, H. A. (1973), "Concomitants of Order Statistics," *Bulletin of International Statistical Institute*, 45, 295–300.

— (1994), "Concomitants of Extreme Order Statistics," in *Extreme Value Theory and Applications. Proceedings of the Conference on Extreme Value Theory and Applications, Volume 1*, eds. Galambos, J., Lechner, J., and Simiu, E., Kluwer Academic Publishers, Boston, pp. 211–224.

David, H. A. and Galambos, J. (1974), "The Asymptotic Theory of Concomitants of Order Statistics," *Journal of Applied Probability*, 11, 762–770.

David, H. A. and Levine, D. N. (1972), "Ranked Set Sampling in the Presence of Judgement Error," *Biometrics*, 28, 553–555.

David, H. A. and Nagaraja, H. (2003), *Order Statistics, Third Edition*, John Wiley & Sons, Hoboken, NJ.

David, H. A. and Nagaraja, H. N. (1998), "Concomitants of Order Statistics," in *Order Statistics: Theory & Methods*, eds. Balakrishnan, N. and Rao, C. R., Elsevier, Amsterdam, pp. 487–513.

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," *The Annals of Statistics*, 5, 216–223.

Dielman, T., Lowry, C., and Pfaffenberger, R. (1994), "A Comparison of Quantile Estimators," *Communications in Statistics: Simulation and Computation*, 23, 355–371.

Egorov, V. A. and Nevzorov, V. B. (1984), "Rate of Convergence to the Normal Law of Sums of Induced Order Statistics," *Journal of Soviet Mathematics (New York)*, 25, 1139–1146.

Eryilmaz, S. (2005), "Concomitants in a Sequence of Independent Nonidentically Distributed Random Vectors," *Communications in Statistics: Theory and Methods*, 34, 1925–1933.

Fan, J. (1992), "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

— (1993), "Local Linear Regression Smoothers and Their Minimax Efficiencies," *The Annals of Statistics*, 21, 196–216.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modeling and Its Application-Theory and Methodologies*, Chapman and Hall, New York.

Ferguson, T. S. (1996), *A Course in Large Sample Theory*, Chapman and Hall, New York.

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*, John Wiley & Sons, New York.

Gasser, T., Kohler, W., Koehler, W., Kneip, A., and Köhler, W. (1991), "A Flexible and Fast Method for Automatic Smoothing," *Journal of the American Statistical Association*, 86, 643–652.

Gasser, T. and Mueller, H.-G. (1984), "Estimating Regression Functions and Their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics*, 11, 171–185.

Gasser, T. and Müller, H. (1979), "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation*, eds. Gasser, T. and Rosenblatt, M., Springer-Verlag, Heidelberg, pp. 23–67.

Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986), "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika*, 73, 625–633.

Ghosh, J. K. (1971), "A New Proof of the Bahadur Representation of Quantiles and an Application," *The Annals of Mathematical Statistics*, 42, 1957–1961.

Gill, P. S., Tiku, M. L., and Vaughan, D. C. (1990), "Inference Problems in Life Testing under Multivariate Normality," *Journal of Applied Statistics*, 17, 133–147.

Gourieroux, C., Laurent, J., and Scaillet, O. (2000), "Sensitivity analysis of Values at Risk," *Journal of Empirical Finance*, 7, 225–245.

Haan, L.D., F. A. (2006), *Extreme Value Theory: an Introduction*, Springer, New York.

Hájek, J. (1968), "Asymptotic Normality of Simple Linear Rank Statistics under Alternatives," *The Annals of Mathematical Statistics*, 39, 325–346.

Hall, P. (1979), "A Note on a Paper of Barnes and Tucker," *Journal of the London Mathematical Society.*, 19, 170–174.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, England.

Härdle, W., Hall, P., and Marron, J. S. (1988), "How Far Are Automatically Chosen Regression Smoothing Parameters from Their Optimum?" *Journal of the American Statistical Association*, 83, 86–95.

Harrell, F. E. and Davis, C. E. (1982), "A New Distribution-free Quantile Estimator," *Biometrika*, 69, 635–640.

Harrell, F. E. and Sen, P. K. (1979), "Statistical Inference for Censored Bivariate Normal Distributions Based on Induced Order Statistics," *Biometrika*, 66, 293–298.

Herrmann, E. (1997), "Local Bandwidth Choice in Kernel Regression Estimation," *Journal of Computational and Graphical Statistics*, 6, 35–54.

— (2000), "Variance Estimation and Bandwidth Selection for Kernel Regression," in *Smoothing and Regression: Approaches, Computation, and Application*, ed. Schimek, M. G., John Wiley & Sons, New York, pp. 71–107.

Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, Irwin, Chicago.

Joshi, S. N. and Nagaraja, H. N. (1995), "Joint Distribution of Maxima of Concomitants of Subsets of Order Statistics," *Bernoulli*, 1, 245–255.

Kaigh, W. D. and Lachenbruch, P. A. (1982), "A Generalized Quantile Estimator," *Communications in Statistics: Theory and Methods*, 11, 2217–2238.

Kaufmann, E. and Reiss, R.-D. (1992), "On Conditional Distributions of Nearest Neighbors," *Journal of Multivariate Analysis*, 42, 67–76.

Lo, A. and Mckinlay, A. (1990), "Data-Snooping Biases in Tests of Financial Asset Pricing Models," *Review of Financial Studies*, 3, 431–467.

Marron, J. S. and Nolan, D. (1988), "Canonical Kernels for Density Estimation," *Statistics & Probability Letters*, 7, 195–199.

Mausser, H. (2001), "Calculating Quantile-based Risk Analytics with L-estimators," *ALGO Research Quarterly*, 4, 33–47.

McIntyre, G. A. (1952), "A Method of Unbiased Selective Sampling Using Ranked Sets," *Australian Journal of Agricultural Research*, 3, 385–390.

Müller, H.-G. and Muller, H.-G. (1988), *Nonparametric Regression Analysis of Longitudinal Data*, Springer-Verlag, Berlin.

Nadaraya, E. A. (1964), "On Estimating Regression," *Theory of Probability and Its Applications*, 9, 141–142.

Nagaraja, H. N. (1982), "Some Nondegenerate Limit Laws for the Selection Differential," *The Annals of Statistics*, 10, 1306–1310.

Nagaraja, H. N. and David, H. A. (1994), "Distribution of the Maximum of Concomitants of Selected Order Statistics," *The Annals of Statistics*, 22, 478–494.

Olkin, I. and Stephens, M. A. (1993), “On Making the Shortlist for the Selection of Candidates,” *International Statistical Review*, 61, 477–486.

Parrish, R. S. (1990), “Comparison of Quantile Estimators in Normal Sampling,” *Biometrics*, 46, 247–257.

Priestley, M. B. and Chao, M. T. (1972), “Non-parametric Function Fitting,” *Journal of the Royal Statistical Society, Series B: Methodological*, 34, 385–392.

Reiss, R.-D. (1989), *Approximate Distributions of Order Statistics: with Applications to Nonparametric Statistics*, Springer-Verlag Inc.

Resnick, S. I. (1987), *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag Inc.

Rice, J. (1984a), “Bandwidth Choice for Nonparametric Regression,” *The Annals of Statistics*, 12, 1215–1230.

— (1984b), “Boundary Modification for Kernel Regression,” *Communications in Statistics: Theory and Methods*, 13, 893–900.

Royden, H. (1968), *Real Analysis, Second Edition*, Macmillan, New York.

Ruppert, D. and Wand, M. P. (1994), “Multivariate Locally Weighted Least Squares Regression,” *The Annals of Statistics*, 22, 1346–1370.

Satagopan, J. M., Venkatraman, E. S., and Begg, C. B. (2004), “Two-stage Designs for Gene-Disease Association Studies with Sample Size Constraints,” *Biometrics*, 60, 589–597.

Satagopan, J. M., Verbel, D. A., Venkatraman, E. S., Offit, K. E., and Begg, C. B. (2002), “Two-stage Designs for Gene-disease Association Studies,” *Biometrics*, 58, 163–170.

Sheather, S. J. and Marron, J. S. (1990), “Kernel Quantile Estimators,” *Journal of the American Statistical Association*, 85, 410–416.

Smith, R. L. (1982), “Uniform Rates of Convergence in Extreme-value Theory,” *Advances in Applied Probability*, 14, 600–622.

Sondhauss, U. (1994), “Asymptotische Eigenschaften Intermediärer Ordnungsstatistiken und ihrer Konkomitanten.” Ph.D. thesis, Department of Statistics, Dortmund University, Germany.

Song, R., Buchberger, S. G., and Deddens, J. A. (1992), “Moments of Variables Summing to Normal Order Statistics,” *Statistics & Probability Letters*, 15, 203–208.

Song, R. and Deddens, J. A. (1993), “A Note on Moments of Variables Summing to Normal Order Statistics,” *Statistics & Probability Letters*, 17, 337–341.

Spruill, N. L. and Gastwirth, J. L. (1982), "On the Estimation of the Correlation Coefficient from Grouped Data," *Journal of the American Statistical Association*, 77, 614–620.

Stokes, S. L. (1977), "Ranked Set Sampling with Concomitant Variables," *Communications in Statistics-Theory and Methods*, 6, 1207–1212.

Stone, C. J. (1977), "Consistent Nonparametric Regression," *The Annals of Statistics*, 5, 595–620.

Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.

Vaughan, R. J. and Venables, W. N. (1972), "Permanent Expressions for Order Statistics Densities," *Journal of the Royal Statistical Society, Series B: Methodological*, 34, 308–310.

Vieu, P. (1991), "Nonparametric Regression: Optimal Local Bandwidth Choice," *Journal of the Royal Statistical Society, Series B: Methodological*, 53, 453–464.

Watson, G. (1964), "Smooth Regression Analysis," *Sankhyā, Series A*, 26, 359–372.

Wolfe, D. A. (2004), "Ranked Set Sampling: An Approach to More Efficient Data Collection," *Statistical Science*, 19, 636–643.

Yang, S. S. (1977), "General Distribution Theory of the Concomitants of Order Statistics," *The Annals of Statistics*, 5, 996–1002.

— (1981), "Linear Functions of Concomitants of Order Statistics with Application to Nonparametric Estimation of a Regression Function," *Journal of the American Statistical Association*, 76, 658–662.

Yeo, W. B. and David, H. A. (1984), "Selection through an Associated Characteristic, with Applications to the Random Effects Model," *Journal of the American Statistical Association*, 79, 399–405.