STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE DISTRIBUTIONS
BASED ON INDUCED ORDER STATISTICS

by

Frank E. Harrell, Jr.

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1230

MAY 1979
STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE DISTRIBUTIONS

BASED ON INDUCED ORDER STATISTICS

by

Frank E. Harrell, Jr.

A Dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics

Chapel Hill

1979

Approved by:

[Signatures]

Adviser

Reader

Reader
This research is concerned with methods of multivariate analysis applicable to life testing problems with concomitant variables observable only for failing units. With this setup the author attempts to provide:

1) estimates of parameters from censored bivariate normal observations begun by Watterson [1959],
2) the likelihood ratio test of independence based on censored bivariate normal observations,
3) methods for inference based on censored multivariate normal distributions especially by studying multiple and partial correlations,
4) parameter estimates and likelihood ratio tests of independence for other selected bivariate distributions, and
5) locally most powerful rank tests of independence for general censored bivariate distributions.

These methods also have applications in inference from biased samples resulting from a selection process.
ACKNOWLEDGMENTS

The author wishes to express much appreciation to his adviser, Professor P. K. Sen, who suggested the topic of this dissertation, and who was constantly available for counsel and guidance. The insight provided by Professor Sen was invaluable throughout the research process. The other members of the committee, Professors C. E. Davis, K. L. Lee, D. Quade, and O. D. Williams, also provided helpful suggestions and ideas which went into the dissertation.

The author also wishes to thank his wife, Marielouise Worden Harrell, for her patience and assistance in his graduate studies, and his parents, Mr. and Mrs. Frank E. Harrell, Sr. for their constant encouragement and understanding throughout his studies.

Finally, the author wishes to acknowledge the support provided by the Lipids Research Clinics Program through the National Heart, Lung and Blood Institute, contract NIH-NHLBI-71-2243 from the National Institutes of Health.

FEH
TABLE OF CONTENTS

LIST OF TABLES .............................................. vii
LIST OF FIGURES ............................................. viii

CHAPTER

I. INTRODUCTION AND REVIEW OF LITERATURE ............. 1
  1.1. Introduction ........................................... 1
  1.2. Notation and Definitions .............................. 2
  1.3. Methods for Analyzing Censored Survival Data When All Concomitant Data are Available .............. 4
  1.4. Framework for Survival Analysis with Incomplete Concomitant Data .................................. 11
  1.5. Censored Bivariate Normal Distribution ............ 13
  1.6. Summary ................................................ 19

II. ESTIMATION OF PARAMETERS OF CENSORED BIVARIATE NORMAL DISTRIBUTIONS AND TEST OF INDEPENDENCE BASED ON INDUCED ORDER STATISTICS .......................... 20
  2.1. Introduction ........................................... 20
  2.2. Estimation of $\sigma^2_x$ using only Induced Order Statistics ........................................ 20
  2.3. Cramér-Rao Lower Bounds for Variances of
Unbiased Estimators ........................................ 24
2.4. Maximum Likelihood Estimation ..................... 33
2.5. Test of $H_0 : \rho = 0$ ................................. 41
  2.5.1. Likelihood Ratio Test ............................. 41
  2.5.2. Distribution of Test Statistic under
          Censoring ....................................... 44

III. STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE
    NORMAL DISTRIBUTIONS BASED ON INDUCED ORDER

STATISTICS ................................................. 53
  3.1. Introduction ........................................ 53
  3.2. Maximum Likelihood Estimation .................... 53
  3.3. Inference about Independence of $\tilde{X}$ and $Y$ ... 62
        3.3.1. Estimation of Multiple Correlation
                Coefficient .................................. 62
        3.3.2. Likelihood Ratio Test of Independence ....... 64
        3.3.3. Distribution of Test Statistic under
                Censoring .................................... 66
  3.4. Inference about Independence of $\tilde{X}^{(2)}$ and $Y$
       given $\tilde{X}^{(1)}$ .................................. 72
        3.4.1. Estimation of Partial Correlation
                Coefficients .................................
        3.4.2. Distributions of Test Statistic under
                Censoring .................................... 75
3.5. Inference about Independence of $X^{(1)}$ and
$(X^{(2)}, Y)$ ......................................................... 83

3.5.1. Estimation of Canonical Correlations . 83
3.5.2. Null Distribution of a Test Statistic. 84

IV. STATISTICAL INFERENCE FOR OTHER MULTIVARIATE
    DISTRIBUTIONS ............................................. 86

4.1. Introduction ............................................... 86
4.2. Inference when Order Statistics are from any
    Parent Distribution but Concomitant Variables
    are Conditionally Normally Distributed .... 87
    4.2.1. Cramèr-Rao Lower Bounds ................. 87
    4.2.2. Estimation ............................................. 91
    4.2.3. Likelihood Ratio Test of Independence 94
4.3. Example: Order Statistics from Extreme Value
    Parent Distribution ....................................... 94
4.4. Example: Order Statistics from Exponential
    Parent Distribution ....................................... 99

V. DISTRIBUTION-FREE TESTS ON INDEPENDENCE BASED ON
    INDUCED ORDER STATISTICS .............................. 111

5.1. Introduction ............................................... 111
5.2. Locally Most Powerful Rank Test of
    Independence ............................................... 111
5.3 Empirical Power Studies ................................... 118
5.4 Asymptotic Relative Efficiencies of Linear
    Rank Tests ................................................. 122
5.5. Hoeffding's General Test of Independence as applied to Censored Bivariate Samples \(\ldots\) 124

VI. SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH \(\ldots\) 131

6.1. Summary \(\ldots\) 131

6.2. Suggestions for further Research \(\ldots\) 132

BIBLIOGRAPHY \(\ldots\) 133
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Cramér-Rao Lower Bounds for Variances of Unbiased Estimators of the Parameters of the Bivariate Normal Distribution based on Order Statistics and Induced Order Statistics</td>
</tr>
<tr>
<td>2.2</td>
<td>Estimated Moments of Modified Maximum Likelihood Estimators of the Bivariate Normal Parameters</td>
</tr>
<tr>
<td>2.3</td>
<td>Estimated Power of 5% r—Test for Censored Data *</td>
</tr>
<tr>
<td>4.1</td>
<td>Estimated Power of 5% Test for Censored Data from the Extreme Value Distribution</td>
</tr>
<tr>
<td>4.2</td>
<td>Estimated Power of 5% Test for Censored Data under the Log-Linear Model</td>
</tr>
<tr>
<td>5.1</td>
<td>Estimated Power of 5% Normal Scores Test for Censored Data from the Bivariate Normal Distribution</td>
</tr>
<tr>
<td>5.2</td>
<td>Estimated Power of 5% Normal Scores Test for Censored Data from the Extreme Value/Normal Distribution</td>
</tr>
<tr>
<td>5.3</td>
<td>Estimated Power of 5% Hoeffding D Test for Censored Bivariate Normal Distribution</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure

4.1 Graph of Marginal Density of $X_i$ when $Y_i$ has an

Exponential Distribution .......................... 106
CHAPTER I
INTRODUCTION AND REVIEW OF LITERATURE

1.1. Introduction

In life testing studies, life length or time to failure of an item is the major response variable. Many parametric statistical methods are now available for analyzing the relationship between life length and a set of associated variables when life length measurements are censored in some way and the associated variables are observed for all n censored or uncensored units. The literature review included describes many of these methods as background for investigating the problem of concomitant data being available only for units that actually fail.

Selection problems can also be placed into this setting. For example, suppose that 1000 students take a college's entrance exam and the highest 100 scorers enter the college. Once there they take a second exam. The problem may be to estimate the population mean score of the second type of exam or the population correlation between the two exams given this biased sample, where the population of interest consists of all high school graduates.
1.2 Notation and Definitions

In most life testing studies, observations are accumulated sequentially in time and are censored, i.e., only a subset of the set of ordered life lengths is observed. Suppose that \( n \) items in an experiment have life lengths \( Y_1, \ldots, Y_n \) and let

\[
Y_{n,1} \leq Y_{n,2} \leq \cdots \leq Y_{n,n}
\]

be the order statistics corresponding to \( Y_1, \ldots, Y_n \) with \( Y_{n,j} \leq Y_{n,j+1} \leq \cdots \leq Y_{n,k} \) being observable, \( 1 \leq j \leq k \leq n \). If \( j > 1, k = n \) we have censoring on the left; for \( j = 1, k < n \) we have censoring on the right, the typical type for survival studies.

When \( j \) or \( k \) is a random variable, the censoring is said to be of type I; an example is an experiment with a pre-specified duration and \( k \) is the number of failures or deaths occurring during that time and \( j = 1 \). If \( k \) is pre-specified, the censoring is of type II and the duration of the experiment is thus a random variable. In multi-stage censoring schemes or staggered entry experiments (called progressive censoring by some authors - see for example Cohen [1965]), subjects are withdrawn from the study without having failed at each of a number of stages (or equivalently they enter the study at different times), but the statistical analysis is not performed until the end of the study. In true progressive censoring, statistical tests are made throughout the experiment to possibly terminate it early with rejection of the hypothesis of no effect of a treatment on life length.
Most published statistical methods as well as those proposed here are designed for analyzing type II right-censored data in which the smallest \( k \) out of \( n \) order statistics of life length are observed for fixed \( k \). However, for both type I and type II censoring maximum likelihood estimation procedures are the same and asymptotically the estimators for either type will have equivalent properties. When \( k \) is a random variable, the distribution of derived estimators depends on the distribution of \( k \) and hence may be more complicated.

Let \( X_1, \ldots, X_n \) denote concomitant variates corresponding to failure times \( Y_1, \ldots, Y_n \) so that \( (X_1, Y_1), \ldots, (X_n, Y_n) \) are independent and identically distributed random vectors (i.i.d.r.v.). Define

\[
X_n[i] = X_j \text{ if } Y_{n,i} = Y_j \text{ for } i, j = 1, \ldots, n.
\]

(1.2.1)

Bhattacharya [1974] has termed the \( X_n[i] \) as the induced order statistics while David and Galambos [1974] have named these as the concomitants of order statistics. We will call \( X_n[1], \ldots, X_n[k], \ 1 \leq k \leq n \) the first \( k \) of \( n \) induced order statistics. Note that these variates are not necessarily in any order. When concomitant data are observable only for failing units then only the first \( k \) of \( n \) induced order statistics are observable.
1.3. Methods for Analyzing Censored Survival Data When

All Concomitant Data are Available.

Assume that the conditional distribution function (d.f.) of $Y_i$ given $X_i$ is of the form $F_{BX_i}(y)$ with probability density function (p.d.f.) $f_{BX_i}(y)$, for $i = 1, \ldots, n$; here, the $X_i$ may be $p (>1)$-vectors, $\beta$ is an unknown $p$-vector, and $BX_i$ is a vector product. The likelihood function of a type II right-censored sample (and given the $X_i$) is

$$\hat{L}_{n,k} = \prod_{i=1}^{k} f_{BX_i}^{BX_i}(Y_{n,i}) \prod_{i=k+1}^{n} [1-F_{BX_i}^{BX_i}(Y_{n,i})].$$  \hspace{1cm} (1.3.1)

Various authors have used (1.3.1) to draw statistical inference [on $\beta$ and other parameters associated with p.d.f. $f_{BX}(y)$] using the principle of maximum likelihood. See Zippin and Armitage [1966] for the exponential distribution, Glasser [1967] for the exponential hazard distribution, for example.

A good general description of the setup of the likelihood function is given by David and Moeschberger [1978] generalized to more than one cause of failure and multiple censoring points. Consider the case of type I multi-stage censoring in which an individual's failure is observed only if it occurs within some predetermined time interval, the length of which may vary from subject to subject.

In this case, subjects can enter the study at different times...
with the termination point being the same for all, so the times are adjusted to start at zero for each subject. Starting with \( n \) subjects, the \( i \)th subject is observed until censoring time \( B_i \). The lifetime for a subject, \( Y_i \), is known only if \( Y_i \leq B_i \). The likelihood of the sample is then

\[
\left\{ \prod_{j} f_{X_j}(Y_j) \right\} \left\{ \prod_{l} \left[ 1 - F_{X_l}(B_l) \right] \right\}, \tag{1.3.2}
\]

where the first product is over the known failures and the second is over the censored times. If the \( B_i \) are random as in random entry times to the study, the estimation procedure is the same but the properties of the estimators will depend on the distribution of the entry times. For type II censoring, it is assumed that all \( B_i \) are equal. A discussion of this also appears in Cohen [1965], aimed more at the case of multi-batch censoring without covariates.

Estimation of the parameter vector \( \beta \) is usually carried out by maximizing the likelihood in (1.3.2) by an iterative process and associated tests of hypotheses are performed by computing likelihood ratios and using large sample theory pertaining to the distribution of \( -2 \log L \) where \( L \) is the likelihood ratio. Halperin [1952] showed that the maximum likelihood estimate (MLE) of a single parameter with no covariates is in some sense optimal under type II right censoring with mild regularity conditions on the underlying density. He also gave an outline of the proof for the multiple
parameter case, the MLE is weakly consistent, efficient, and asymptotically multi-normally distributed. An additional paper on this subject is by Silvey [1961], which includes a discussion of the properties of the likelihood for dependent random variables.

Maximum likelihood estimators can seldom be derived explicitly. Because of this the properties of the estimators are difficult to study. Saw [1961] was able to study the moments of the MLE of the mean ($\mu$) and standard deviation ($\sigma$) derived from a type II right censored normal sample. He found the estimators to be badly biased for large proportion of censoring (small $k/n$). For $k/(n+1)=.2$, he derived

$$
\mu - E(\hat{\mu}) = 5.538\sigma/(n+1) + O(n+1)^{-2}
$$

(1.3.3)

and the amount of bias is worse for smaller $k/(n+1)$.

An explicit and asymptotically efficient method for deriving unbiased estimators was given by Plackett [1958] and formalized by Chan [1967]. Chan's formulation is for a general location-scale distribution and double censoring. Here the sample consists of $Y_{n,j}, Y_{n,j+1}, \ldots, Y_{n,k}$, $1<j<k<n$ from the parent density

$$
\theta_2^{-1} f((y-\theta_1)/\theta_2)
$$

where $\theta_1$ and $\theta_2$ are respectively the location and scale parameters. Let $Z_{n,i} = (Y_{n,i} - \theta_1)/\theta_2$ and $U_{n,i} = E(Z_{n,i})$. The log likelihood function is (Chan [1967])
\[
L = \log[n!/(j-l)!(n-k)!] - (k-j+1)\log \theta_2 + \sum_{i=j}^{k} \log f(Z_{n,i}) \\
+ (j-l) \log F(Z_{n,j}) + (n-k)\log [1-F(Z_{n,k})],
\]

(1.3.4)

where \( F(z) \) is the d.f. corresponding to the p.d.f. \( f(z) \). The strategy is to expand \( \partial L/\partial \theta_r \), \( r = 1, 2 \), in a Taylor series about \( Z_{n,i} = u_{n,i}, j \leq i \leq k \), and then to ignore terms which converge in probability to zero. Thus, aside from these remaining terms of order greater than one, which converge in probability to zero,

\[
n^{-1/2} \left( \frac{\partial L}{\partial \theta_r} \right) = n^{-1/2} \left( \frac{\partial L}{\partial \theta_r} \right)_{u_{n,j}, \ldots, u_{n,k}} \\
+ n^{-1/2} \gamma_2^{-1} L'(r)
\]

(1.3.5)

where

\[
L^{(1)} = -(j-1) \left[ (Y_{n,j} - \theta_1)/\theta_2 - u_{n,j} \right] (f_j'/f_j - f_j^2/p_j^2) \\
- \sum_{i=j}^{k} \left[ (Y_{n,i} - \theta_1)/\theta_2 - u_{n,i} \right] (f_i''/f_i - f_i'^2/f_i^2) \\
+ (n-k) \left[ (Y_{n,k} - \theta_1)/\theta_2 - u_{n,k} \right] (f_k'/q_k + f_k^2/q_k^2)
\]

(1.3.6)
\[ L^{(2)} = -(j-1) \left[ (Y_{n,j} - \theta_1)/\theta_2 - u_{n,j} \right] \left( f_j/p_j + u_{n,j} f_j'/p_j - u_{n,j} f_j^2/p_j^2 \right) \]
\[ - \sum_{i=j}^{k} \left[ (Y_{n,i} - \theta_1)/\theta_2 - u_{n,i} \right] \left( u_{n,i} f_i''/f_i - u_{n,i} f_i^2/f_i^2 + f_i'/f_i \right) \]
\[ + (n-k) \left[ (Y_{n,k} - \theta_1)/\theta_2 - u_{n,k} \right] \left( f_k/q_k + u_{n,k} f_k'/q_k \right) \]
\[ + u_{n,k} f_k^2/q_k^2, \]

\[ f_i = f(u_{n,i}), \ p_i = F(u_{n,i}), \ q_i = 1-p_i, \ f_i' = \frac{\partial f(x)}{\partial x} \Big|_{x=u_{n,i}} \]

and so forth.

The first term of (1.3.5) also converges in probability to zero so the estimation procedure is to solve for \( L^{(r)} = 0, \ r = 1,2. \)

Denote the estimators satisfying this requirement by \( \theta_1^* \) and \( \theta_2^*. \)

The fact that \( (\theta_1^*, \theta_2^*) \) are unbiased follows from the fact that \( E(L^{(r)}) = 0. \) Chan also shows that estimators derived from replacing \( u_{n,i} \) by \( F^{-1}(i/(n+1)) \) have the same asymptotic properties. The estimators \( (\theta_1^*, \theta_2^*) \) can be called linearized maximum likelihood estimators.

Saw [1962] developed a procedure similar to that of Chan for the multiple parameter case and for general likelihood functions in which the likelihood is expressed as a polynomial with a remainder term. This procedure can be used in many cases to derive simple
and efficient moment estimators. Saw applied this to right censored normal samples (See also Saw [1959]) to derive simple and efficient estimators of the mean and variance. These estimators require computation of only two expected values as opposed to Chan's estimators which involve calculation of $u_{n,j}, \ldots, u_{n,k}$.

Modified maximum likelihood techniques could perhaps be applied to the more complicated likelihood in (1.3.1) involving covariates. This technique might eliminate the need for iterative solutions and provide unbiased estimates of regression parameters. At any rate, there seems to be no rigorous proof of the optimality of MLE's derived from expressions such as (1.3.1) which involve covariates. Many authors in the literature who discuss survival analysis with covariates seem to assume that theorems concerning identically distributed observations apply.

There are other methods in the literature for analyzing censored survival data than the conventional parametric maximum likelihood technique. Cox [1972] developed a conditional likelihood model which could be called a mixed parametric/nonparametric approach. In his model, which allows for arbitrary right censoring, only the ranks of the failure times are used. The model assumes that for some (unspecified) monotonic transformation $d(y)$ of the failure times, $d(Y) - \tilde{\beta}X$ has an extreme value distribution. This implies that the ratio of hazard rates for observations $i$ and $j$ at each fixed failure time $y$ is $\exp(\tilde{\beta}X_i - \tilde{\beta}X_j)$. [The hazard rate is defined to be $f(y)/(1-F(y))]$. This mixed approach has been explored for
general distributions by Kalbfleisch [1978].

Brown, Hollander, and Korwar [1974] developed a nonparametric test of association between arbitrarily right censored failure time and one uncensored concomitant variable. The test is based on Kendall's tau and is easily applied. The data for the test consist of \( n \) independent pairs \((X_i, Y_i), \ldots, (X_n, Y_n)\) where \( Y_i \) is known only if \( Y_i \leq B_i \).

Let \( \delta_i = I\{Y_i \text{ uncensored}\} = I\{Y_i \leq B_i\} \) and \( Z_i = \min(Y_i, B_i) \), where \( I\{A\} \) denotes the indicator function for the set \( A \). Let

\[
    a_{ij} = \begin{cases} 
        1, & X_i > X_j \\
        0, & X_i = X_j \\
        -1, & X_i < X_j 
    \end{cases} \tag{1.3.9}
\]

\[
    b_{ij} = \begin{cases} 
        1 & \text{if it can be inferred that } Y_i > Y_j \\
        0 & \text{if } Y_i = Y_j \text{ or if "uncertain"} \\
        -1 & \text{if it can be inferred that } Y_i < Y_j 
    \end{cases} \tag{1.3.10}
\]

so that

\[
    b_{ij} = \begin{cases} 
        1, & (\delta_i = 1 \text{ and } Z_i > Z_j) \text{ or } (\delta_i = 0, \delta_j = 1 \text{ and } Z_i = Z_j) \\
        -1, & (\delta_i = 1 \text{ and } Z_i < Z_j) \text{ or } (\delta_i = 1, \delta_j = 0 \text{ and } Z_i = Z_j) \\
        0, & \text{otherwise} 
    \end{cases} \tag{1.3.11}
\]

The test statistic for independence is

\[
    S = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij} \tag{1.3.12}
\]
and the test can be performed by either finding the permutation distribution of S or by using a normal approximation to its null distribution.

Brown et al. [1974] proposed a modification of this test which uses the Kaplan-Meier [1958] empirical survival curve to gain more information in the "uncertain" case $b_{ij} = 0$. This test is more complicated and may make little improvement when much censoring is present. In fact, in the example in their paper, in which little censoring is present, the significance levels for the more complicated test are generally larger than for the simple test in (1.3.12). When $X_1$ is a binary variable, (1.3.12) reduces to the modified two-sample test of Gehan [1965].

Majumdar and Sen [1978] have considered generalized linear rank statistics for progressive censoring with staggered entry. Miller [1976] estimated parameters of a linear regression on failure times by means of weighted regression using weights constructed from the Kaplan-Meier curve. A general theory of rank analysis of covariance under progressive censoring when there is no staggered entry has been developed by Sen [1977]. The asymptotic theory of multiple likelihood ratio tests in progressive censoring has been studied by Sen [1976] for the single parameter case.

1.4. Framework for Survival Analysis with Incomplete Concomitant Data

All methods in section 1.2 assume that though $Y_{n,1}, ..., Y_{n,k}$ are
only observable, the entire set \((X_1, \ldots, X_n)\) [or equivalently \(X_{n[1]}, \ldots, X_{n[n]}\)] is given prior to experimentation so that (1.3.1) is properly defined. In many other life testing problems, especially those arising in toxicological or other invasive or destructive studies, \(X_{n[i]}\) is observable only when \(Y_{n,i}\) is observable (for \(i \leq n\)), so that the second factor on the right hand side of (1.3.1) is no longer deterministic. This is particularly true if observing \(X_i\) necessitates the failure of the \(i\)th unit, so that for the surviving \([n-k]\) units, the \(X_i\) are not observable.

As an example, suppose that an experiment is performed in which a toxic substance is injected in constant amounts at regular time intervals in mice until death occurs. At death, the amount of the toxic substance taken up by the kidneys is measured to test the hypothesis that death occurs when a certain amount is precipitated there. For mice alive at the end of the study, the toxic substance may have been dissipated by means other than by the kidneys' action. Suppose that for time or cost reasons, the experimenter, having started with 30 mice, had decided to stop the experiment, after the first 10 deaths occur. He may want to study the relationship between survival time and absorption of the toxic substance by the kidney, the latter being measurable only at the termination of life and removal of a kidney.

As another example, suppose that in an experiment designed to study the relationship between arteriosclerosis and life length,
rhesus monkeys are fed an atherogenic diet. Autopsies are performed on the first 20 monkeys to die and the amount of fatty plaque deposited on the walls of the aorta is measured. The experimenter may wish to quantify the strength of relationship between this measurement and life length. Other examples of this type arise in selection problems, particularly in educational testing.

The necessity of observing concomitant data for all individuals is not required if we make the assumption that \((X_i, Y_i)\) has a multivariate distribution of known form. Suppose that \((X_i, Y_i)\) has p.d.f. \(f(x, y)\) and the conditional density of \(X_i\) given \(Y_i = y_i\) is \(g_{y_i}(x)\).

By an appeal to Lemma 1 of Bhattacharya [1974], we conclude that given \(Y_{n,1}, \ldots, Y_{n,k}, X_{n[1]}, \ldots, X_{n[k]}\) are (conditionally) independently distributed and the conditional p.d.f. of \(X_{n[j]}\) given the \(Y_{n,i}\), \(1 \leq i \leq k\) is given by

\[
g_{y_{n,j}}(x) \text{ for } j = 1, \ldots, k \text{ and every } 1 \leq k \leq n. \tag{1.4.1}
\]

This result provides the basis for much of the statistical inference based on induced order statistics.

1.5. Censored Bivariate Normal Distribution

Watterson [1959] considered the case where \((X_i, Y_i)\) have a bivariate normal distribution. Denote the p.d.f. of \((X_i, Y_i)\) by

\[
\phi_\theta(x, y) \text{ with } \theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho), \mu_x, \mu_y \text{ are the means, } \sigma_x^2, \sigma_y^2
\]
the variances, and \( \rho \) the correlation coefficient of \( X \) and \( Y \). By (1.4.1), given \( Y_{n,1}, \ldots, Y_{n,k}, X_{n[1]}, \ldots, X_{n[k]} \) are independently normally distributed with means

\[
\mu_X + \rho \sigma_X (Y_{n,i} - \mu_Y) / \sigma_Y
\]

and variances

\[
\sigma_X^2 (1 - \rho^2)
\]

for \( j=1, \ldots, k \). Let then \( \sim X = (X_{n[1]}, \ldots, X_{n[k]})' \), \( \sim Y = (Y_{n,1}, \ldots, Y_{n,k})' \) and \( \sim Z = \sigma_Y^{-1}(Y - \mu_Y) \). The joint (conditional) p.d.f. of \( \sim X \) given \( \sim Y \) is then

\[
[\sigma_X^2 (1 - \rho^2) 2\pi]^{-k/2} \exp\left\{ - \frac{1}{2 \sigma_X^2 (1 - \rho^2)} [X - \mu_X - \rho \sigma_X Z]' [X - \mu_X - \rho \sigma_X Z] \right\},
\]

which is the \( k \)-variate normal density with mean vector

\[
\mu_X + \rho \sigma_X Z
\]

and covariance matrix

\[
\sigma_X^2 (1 - \rho^2) I_k
\]
which implies

\[ E(X X'|Y) = \sigma_x^2 (1 - \rho^2) \xi_k + [E(X|Y)] [E(X|Y)]' \]  

(1.5.6)

Thus if we let

\[ \Xi = \nu = (\nu_{n,1}, \ldots, \nu_{n,k}) \] and \[ E(Z - \Xi)(Z - \Xi)' = \Sigma = (\Sigma_{i,j}), \]  

(1.5.7)

then, from (1.5.4), (1.5.5) and (1.5.6) we obtain

\[ E(X) = \mu + \rho \sigma \nu \] and \[ E(X - \mu)(X - \mu)' = \sigma_x^2 (1 - \rho^2) \xi_k + \rho^2 \sigma_x^2 \nu. \]  

(1.5.8)

For \( k \leq n \geq 20 \), \( \nu \) and \( \Sigma \) are tabulated in Sarhan and Greenberg [1962, pp. 193-205] and several approximations for these moments for higher values of \( n \) are in David [1970, pp. 65-7]. Harter [1961] suggests that a good approximation for \( \nu_{n,i} \) is

\[ \nu_{n,i} = \Phi^{-1} \left( \frac{(i - \alpha_n)/(n+1 - 2\alpha_n)}{\alpha_n = 3/8}, \ 1 \leq n \leq 50, \ \alpha_n = .4, \ n > 50, \right. \]  

(1.5.9)

where \( \Phi^{-1} \) denotes the standard univariate normal d.f. The \( \nu_{n,i} \) can also be approximated with high accuracy by numerical integration or by using the power series of Saw [1958] for select \( k/(n+1) \) not
too near 0 or 1. The result of Saw [1959, section 6.] may be used to interpolate the power series for values of \( k/(n+1) \) not tabulated. From (1.5.8) we obtain

\[
E(X_n[i] - EX_n[i])(X_n[j] - EX_n[j]) = \rho^2 \sigma^2_{x, ij}, \quad i \neq j
\]

\[
= \sigma^2_x(1 - \rho^2) + \rho^2 \sigma^2_{x, ii}, \quad i = j
\]  
\[(1.5.10)\]

and using (1.5.4),

\[
E(X_n[i] - EX_n[i])(Y_n,j - EY_n,j) = \rho \sigma_x \sigma_y \nu_{n, ij}
\]

\[
= \nu_{n, ij} \text{Cov}(X_i, Y_i).
\]  
\[(1.5.11)\]

All of these results can be found in Watterson [1959].

From (1.5.7) and the definition of \( \Xi \),

\[
E(\Xi) = \mu_y \Xi + \sigma_y \Xi \quad \text{and} \quad E(\Xi - \mu_y \Xi)(\Xi - \mu_y \Xi)' = \sigma^2_{y, \Xi}.
\]  
\[(1.5.12)\]

Let \( \alpha \) denote a \( k \times 1 \) vector of constants. Then

\[
E(\alpha'Y) = \mu_y \alpha' \Xi + \sigma_y \alpha' \Xi, \quad V(\alpha'Y) = \sigma^2_{y, \Xi} \Xi \alpha
\]  
\[(1.5.13)\]

from which the best linear unbiased estimators (BLUE's) of \( \mu_y \) and...
\(\sigma_y\) can be determined. However, if \(\tilde{a}'\tilde{a}\) is minimized instead of \(\tilde{a}'\tilde{a}\), as in Gupta [1952], the estimators are much simpler. The result is

\[
\tilde{\mu}_y = \tilde{a}' \tilde{y} , \quad \tilde{a} = k^{-1} - \frac{(u-u_1)}{k} \tilde{u}_k / \sum \frac{(u_{n,i} - \tilde{u}_k)^2}{m=1}
\] (1.5.14)

\[
\tilde{\sigma}_y = \tilde{b}' \tilde{y} , \quad \tilde{b} = (u-u_k) / \sum \frac{(u_{n,i} - \tilde{u}_k)^2}{m=1}
\] (1.5.15)

where

\[
u_k = k^{-1} \sum \frac{u_{n,i}}{i=1}
\] (1.5.16)

For censoring on either side, Sarhan and Greenberg [1962, pp. 208-11] computed the minimum efficiency for the simpler estimators with respect to the BLUE's to be over 84% for both the mean and standard deviation for \(n < 15\) and in general the efficiencies are well over 90%.

Now consider linear estimation using \(\tilde{X}\):

\[
E(\tilde{a}'\tilde{X}) = \tilde{\mu}_x \tilde{a}' \tilde{1} + \rho \tilde{\sigma}_x \tilde{a}' \tilde{u}
\] (1.5.17)

\[
V(\tilde{a}'\tilde{X}) = \tilde{\sigma}_x^2 [\tilde{a}'\tilde{a} + \rho^2 \tilde{a}'(\tilde{y} - \tilde{L}_k)\tilde{a}]
\] (1.5.18)

To get an unbiased estimator of \(\tilde{\mu}_x\) we require \(\tilde{a}' \tilde{1} = 1, \tilde{a}' \tilde{u} = 0\).
Minimizing the variance is a problem since it depends on the unknown parameter $\rho$. If we minimize for $\rho = 0$, we have the estimators

$$
\mu_X = \tilde{\alpha}'X \quad \text{and} \quad \phi X = \tilde{\beta}'X
$$

(1.5.19)

where $\tilde{\alpha}$ and $\tilde{\beta}$ are defined in (1.5.14) and (1.5.15) respectively.

Alternatively, minimization can be done for $\rho = 1$ yielding the same vectors as in the BLUE's of $\mu_y$ and $\sigma_y$. For $n = 10$ and all possible ways of left and/or right censoring, Watterson [1959] compared the relative efficiencies of the two for the worst possible value of $\rho^2$ (0 or 1) and found the simpler estimators in (1.5.19) to be often better than the others and never much worse. Note that from (1.5.17), no linear estimator in $X$ exists for $\rho$ or $\sigma_X$.

Estimation of $\rho$ and $\sigma_X$ will be the subject of much of the work in later chapters. Estimation of $\mu_X$ and $\sigma_X$ will perhaps be of more interest in selection problems while estimation of $\rho$ and testing $\rho = 0$ will be of interest in both selection problems and life testing studies.

We note that from (1.5.3) and (1.5.7) that the basic assumption of bivariate normality can be roughly checked by examining the conditional normality of $\tilde{X}$ by usual regression methods and by examining a plot of $Y_{n,i}$ vs. $u_{n,i}$, $i = 1, \ldots, k$ for linearity.

David and Galambos [1974] also discuss this setup. They give asymptotic results for the expected rank of $X_n[i]$ in $X_1, \ldots, X_n$.
for \(1/n \to \lambda\) (0 < \(\lambda\) < 1) as \(n \to \infty\). They also show that for fixed \(k\), \(X_n[1] - E(X_n[1]), \ldots, X_n[k] - E(X_n[k])\) are asymptotically independently normally distributed with mean 0 and variance \(\sigma^2_x(1-\rho^2)\) as \(n \to \infty\) for \(|\rho| < 1\). For our purposes, however, we will need \(k = pn\) with \(n \to \infty\), \(0 < p \leq 1\), \(p\) fixed.

1.6. Summary

Chapter I provides the background necessary for analyzing censored data when concomitant information is available only as induced order statistics for uncensored observations. In chapters to follow, estimators for the remaining parameters, \(\rho\) and \(\sigma_x\), of the bivariate normal distribution will be developed and a test of independence will be studied. These methods will be extended for the multivariate normal distribution thereby providing many tools for statistical inference for censored data that are parallel to those commonly used for complete data. Next, some other bivariate distributions will be considered. Finally, some nonparametric tests of independence will be studied and compared to parametric tests by both empirical power calculations and asymptotic results.
CHAPTER II

ESTIMATION OF PARAMETERS OF CENSORED BIVARIATE NORMAL DISTRIBUTIONS AND TEST OF INDEPENDENCE

BASED ON INDUCED ORDER STATISTICS

2.1. Introduction.

Suppose that \((X_i, Y_i)\) follows a bivariate normal distribution with parameters \(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2,\) and \(\rho\). Estimators of \(\mu_x\) and \(\rho \sigma_x\) using \(X_{n[1]} \ldots X_{n[k]}\) have been developed by Watterson [1959] while the estimation of \(\mu_y\) and \(\sigma_y\) using the order statistics \(Y_{n,1} \ldots Y_{n,k}\) can be made as in Sarhan and Greenberg [1962]. However, the estimation of \(\sigma_x\) and \(\rho\) poses certain difficulties and constitutes a major objective of this chapter. In deriving estimators of \(\sigma_x\) and \(\rho\) we also derive an estimator of \(\mu_x\) using both order statistics and induced order statistics which is slightly better than Watterson's. We will also show that the classical Pearsonian correlation coefficient has desirable properties as a test statistic for testing \(H_0: \rho = 0\) and we will estimate its power function.

2.2. Estimation of \(\sigma_x^2\) using only Induced Order Statistics.

Sometimes one may be interested in estimating population
parameters from censored data even when the ordered variates, \( \tilde{Y} \), are not observable. This problem arises often in selection processes. For example, suppose in a class of 100 students, the 20 tallest are determined by having the students order themselves by increasing height. These 20 students are asked to do high-jumps to determine their maximum jumping capability. Later, the athletic instructor wishes to estimate the population variability of jumping capacity, not just for the tallest students, using only the observed jumps and the ranks of students' heights out of 100.

As we have seen in section 1.5, the population standard deviation of \( X_i \), \( \sigma_x \), is not estimable by any linear function of \( \tilde{X} \). It is natural to try a quadratic estimator of the form \( \tilde{X}'A\tilde{X} \). The variance of such an estimator would depend on the unknown parameter \( \rho \) so the variance cannot be minimized for all \( \rho \). Following the Watterson approach, we might minimize the variance of \( \tilde{X}'A\tilde{X} \) when \( \rho = 0 \), which amounts to minimizing \( \text{tr}\tilde{A}^2 \) subject to \( E(\tilde{X}'A\tilde{X}) = \sigma_x^2 \). The authors have found by simulations that this estimator, which requires the solution of \( k \) simultaneous equations, has little to gain over a much simpler estimator presented below. Let

\[
\tilde{\sigma}_x^2 = \sum_{i=1}^{k} c_i (X_{n[i]} - \tilde{\mu}_X)^2
\]

(2.2.1)

where \( \tilde{\mu}_X \) is Watterson's estimator, \( \sum_{i=1}^{k} \tilde{\alpha}_i X_{n[i]} \), defined in (1.5.19) and the \( c_i \) are real constants. Now
\[ E(X_{n[i]} - \bar{\mu}_x)^2 = V(X_{n[i]}) + V(\bar{\mu}_x) + \rho^2 \sigma^2_x u^2_{n,i} + 2E \{X_{n[i]}(u_x - \bar{\mu}_x)\} \]

and

\[ V(X_{n[i]}) = \sigma^2_x(1 - \rho^2 + \rho^2 \nu_{n,i,i}), \quad V(\bar{\mu}_x) = \sigma^2_x \{(1 - \rho^2)\bar{\alpha} \bar{\alpha} + \rho^2 \bar{\nu}_x\}, \]

\[ \bar{\alpha} = \bar{u}_k / \sum_{m} (u_{n,m} - \bar{u}_k)^2 \]

where \( u_{n,i}, \nu_{n,i,i} \), and \( \bar{u}_k \) are defined in (1.5.7) and (1.5.16).

Also,

\[
E(X_{n[i]} \bar{\mu}_x) = E(\sum_{j=1}^{k} \bar{\alpha}_j X_{n[i]} X_{n[j]} + \bar{\alpha}_i E(X_{n[i]}^2) \\
= \sum_{j \neq i} E(\alpha_j X_{n[i]} X_{n[j]} + \mu_x + \rho \mu_x \sigma_x \{u_{n,i} + u_{n,j}\}) \\
+ \bar{\alpha}_i E(X_{n[i]}^2) \\
= \rho^2 \sigma^2_x \sum_{j=1}^{k} \bar{\alpha}_j \nu_{n,i,j} + \mu_x + \rho \mu_x \sigma_x \{u_{n,i} + \bar{\alpha}_i \sigma_x \}\]

and

\[ E[X_{n[i]}(\mu_x - \bar{\mu}_x)] = -\rho^2 \sigma^2_x \sum_{j=1}^{k} \bar{\alpha}_j \nu_{n,i,j} - \bar{\alpha}_i \sigma_x^2(1 - \rho^2). \]

Therefore,
\[ E(\tilde{X}_n[i] - \tilde{\mu}_x)^2 = \sigma_x^2(a_i + \rho b_i) \]  

(2.2.2)

where

\[ a_i = \frac{(k-1)}{k} + \sum_{m=1}^{k} \frac{1 + 2(u_{n,i} - u_n)}{\Sigma (u_{n,m} - u_k)^2} \]

(2.2.3)

\[ b_i = -a_i - 2 \sum_{j=1}^{k} \tilde{\alpha}_j \nu_{n,ij} + \sum_{r=1}^{k} \sum_{s=1}^{k} \tilde{\alpha}_r \tilde{\alpha}_s \nu_{n,rs} + \nu_{n,ii} + u_{n,i}^2 \]

for \( i = 1, \ldots, k \). Let \( \tilde{a} = (a_1, \ldots, a_k)' \), \( \tilde{b} = (b_1, \ldots, b_k)' \), \( \tilde{c} = (c_1, \ldots, c_k)' \). Thus, \( E(\tilde{\sigma}_x^2) = \sigma_x^2(a'\tilde{c} + \rho b'\tilde{c}) \) and for an unbiased estimator, we require \( \tilde{a}'\tilde{c} = 1 \) and \( \tilde{b}'\tilde{c} = 0 \) and, motivated by Watterson, we minimize \( \tilde{c}'\tilde{c} \) which minimizes \( V(\tilde{\sigma}_x^2) \) when \( \rho = 0 \), considering \( \tilde{\mu}_x \) given. The result is

\[ c_i = \{b_i a'_\tilde{c} - a_i b'_\tilde{c}\}/\{(a'\tilde{b})^2 - (a'_a)(b'_b)\}, 1 \leq i \leq k. \]  

(2.2.4)

An estimator of \( \rho \) can also be obtained by considering the Watterson estimator of \( \rho \sigma_x \) (based on \( X \) alone) and dividing the same by the square root of \( \sigma_x^2 \) in (2.2.1). However, such an estimator does not depend on \( Y \), and hence, when the latter is given, may not be a very efficient one. For this reason, in the next section, we proceed to study the efficiency of estimators of \( \Theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho) \) based on \((X,Y)\).
2.3 Cramér-Rao Lower Bounds for Variances of Unbiased Estimators

Denote the parent bivariate normal density of \((X_i, Y_i)\) by

\[
\phi_\theta(x, y) \text{ where } \theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho), \quad \mu_x, \mu_y \text{ stand for the means,}
\]

\[
\sigma_x^2, \sigma_y^2 \text{ for the variances, and } \rho \text{ for the correlation coefficient of } X_i \text{ and } Y_i. \text{ By (1.5.3) the joint (conditional) pdf of } \tilde{X}_i \text{ given } \tilde{Y}_i \text{ is}
\]

\[
\prod_{i=1}^{k} \phi_\theta(x_i - \mu_x, \rho \sigma_x \sigma_y^{-1}(y_i - \mu_y)) / \sigma_x \sqrt{1-\rho^2} / \sigma_y \sqrt{1-\rho^2} \tag{2.3.1}
\]

\[
= \prod_{i=1}^{k} \phi_\theta(x_i, y_i) / \phi_1((y_{n-k}, \mu_y) / \sigma_y)
\]

where \(\phi_1\) is the standard univariate normal density. The joint pdf for the first \(k\) of \(n\) order statistics is

\[
\prod_{i=1}^{[k]} \frac{\phi_1((y_i - \mu_y) / \sigma_y)}{[1 - \phi_1((y_k - \mu_y) / \sigma_y)]^{n-k} \sigma_y^{-k}} \tag{2.3.2}
\]

\[
y_1 \leq y_2 \leq \cdots \leq y_k
\]

where \(\phi_1\) is the standard univariate normal df and \(n[k] = n \cdots (n-k+1)\). Multiplying (2.3.1) and (2.3.2), we obtain that the joint pdf of \(\tilde{X}_i\) and \(\tilde{Y}_i\) is equal to

\[
\prod_{i=1}^{[k]} \frac{\phi_\theta(x_i, y_i)}{[1 - \phi_1((y_k - \mu_y) / \sigma_y)]^{n-k}} \tag{2.3.3}
\]

\[
y_1 \leq y_2 \leq \cdots \leq y_k
\]

Let \(L_{n,k}\) denote the log likelihood function based on (2.3.3) aside
from constant terms. Then letting \( T = (X - \mu_x) / \sigma_x \),
\[ Z = (Y - \mu_y) / \sigma_y, \quad Z_{n,k} = (Y_{n,k} - \mu_y) / \sigma_y \]
we have

\[
L_{n,k} = (n-k) \log[1 - \phi_1(Z_{n,k})] - k \log \sigma_x - k \log \sigma_y - (k/2) \log(1 - \rho^2)
- \{(1 - \rho^2)^{-1/2}\} \{ T' T - 2 \rho T' Z + Z' Z \}. \tag{2.3.4}
\]

Then we have

\[
\frac{\partial}{\partial \mu_x} L_{n,k} = \frac{l'}{\sigma_x} (T - \rho Z) / \sigma_x (1 - \rho^2) \tag{2.3.5}
\]

\[
\frac{\partial}{\partial \sigma_x} L_{n,k} = -k / \sigma_x + (T' T - \rho T' Z) / \sigma_x (1 - \rho^2)
\]

\[
\frac{\partial}{\partial \rho} L_{n,k} = (1 - \rho^2)^{-2} \left[ k \rho (1 - \rho^2) - \rho T' T - \rho Z' Z + (1 + \rho^2) T' Z \right]
\]

\[
\frac{\partial}{\partial \mu_y} L_{n,k} = (n-k) \sigma_y^{-1} \phi_1(Z_{n,k}) / [1 - \phi_1(Z_{n,k})]
- (1 - \rho^2)^{-1} \sigma_y^{-1} \frac{l'}{\rho T - Z}
\]

\[
\frac{\partial}{\partial \sigma_y} L_{n,k} = (n-k) \sigma_y^{-1} Z_{n,k} \phi_1(Z_{n,k}) / [1 - \phi_1(Z_{n,k})]
- \sigma_y^{-1} (1 - \rho^2)^{-1} (\rho T' Z - Z' Z) - k \sigma_y^{-1}.
\]

With notations in (1.5.7), we let \( u = (u_{n,1}, \ldots, u_{n,k})' \) and
\( \tilde{u} = \left( w_1, \ldots, w_k \right)' \) where \( w_i = v_{n,ii} + u_{n,i}^2 \). Also let
\[ \psi(n,k; a,b,c,d) = E\{ \phi_1(Z_{n,k})^a \phi_1(Z_{n,k})^{-b} [1-\phi_1(Z_{n,k})]^{-c} z_{n,k}^d \} \quad (2.3.6) \]

where \( a, b, c, d \) are non-negative numbers and the pdf of \( Z_{n,k} \) is

\[ \left[ \frac{n}{(k-1)!} \right] \phi_1(z)^{k-1} [1-\phi_1(z)]^{n-k} \phi_1(z), \quad -\infty < z < \infty \quad (2.3.7) \]

for \( k = 1, \ldots, n \). Noting that from (1.5.8)

\[ E(T) = \rho \mu, \quad E(T' T) = k(1-\rho^2) + \rho^2 1' \omega, \]
\[ E(T'Z) = \rho 1' \omega, \quad E(Z'Z) = 1' \omega, \quad (2.3.8) \]

we find the following results after differentiating each side of each equation in (2.3.5) with respect to \( \mu_x, \mu_y, \sigma_x, \sigma_y \), and \( \rho \) and taking expected values of the negations of the resulting second partial derivatives:

\[ \left( \frac{\partial^2}{\partial \mu_x^2} \right) L_{n,k} = -k/\sigma_x^2(1-\rho^2) \]

\[ I_{11} = E\{ -(\partial^2/\partial \mu_x^2) L_{n,k} \} = k/\sigma_x^2(1-\rho^2) \quad (2.3.9) \]

\[ \left( \frac{\partial^2}{\partial \mu_x \partial \sigma_x} \right) L_{n,k} = -1' (2 T - \rho Z)/\sigma_x^2(1-\rho^2) \]

\[ I_{13} = E\{ -(\partial^2/\partial \mu_x \partial \sigma_x) L_{n,k} \} = \rho 1' \mu/\sigma_x^2(1-\rho^2) \quad (2.3.10) \]
\[
\left(\frac{\partial^2}{\partial u_x \partial u_p}\right)_{n,k} = \frac{1}{1} \left(2\rho T - (1+\rho^2)z\right)/\sigma_x (1-\rho^2)^2
\]

\[
I_{15} = E\{-(\partial^2/\partial u_x \partial u_p)\} = \frac{1}{1} u/\sigma_x (1-\rho^2) \tag{2.3.11}
\]

\[
\left(\frac{\partial^2}{\partial u_x \partial u_y}\right)_{n,k} = k\rho/\sigma_x \sigma_y (1-\rho^2)
\]

\[
I_{12} = E\{-(\partial^2/\partial u_x \partial u_y)\} = -k\rho/\sigma_x \sigma_y (1-\rho^2) \tag{2.3.12}
\]

\[
\left(\frac{\partial^2}{\partial u_y \partial u_y}\right)_{n,k} = \rho \frac{1}{1} u/\sigma_x \sigma_y (1-\rho^2)
\]

\[
I_{14} = E\{-(\partial^2/\partial u_y \partial u_y)\} = -\rho \frac{1}{1} u/\sigma_x \sigma_y (1-\rho^2) \tag{2.3.13}
\]

\[
\left(\frac{\partial^2}{\partial u_x^2}\right)_{n,k} = k/\sigma_x^2 - \left(3\frac{T^'T - 2\rho T^'Z}{\sigma_x^2}\right)/\sigma_x (1-\rho^2)
\]

\[
I_{33} = E\{-(\partial^2/\partial u_x^2)\} = \{2k(1-\rho^2) + \rho \frac{2}{1} \frac{1}{1} \frac{w}{\sigma_x^2}(1-\rho^2)\} \tag{2.3.14}
\]

\[
\left(\frac{\partial^2}{\partial u_x \partial u_y}\right)_{n,k} = \rho \frac{1}{1} \frac{1}{T}/\sigma_x \sigma_y (1-\rho^2)
\]

\[
I_{23} = E\{-(\partial^2/\partial u_x \partial u_y)\} = -\rho \frac{2}{1} \frac{1}{1} u/\sigma_x \sigma_y (1-\rho^2) \tag{2.3.15}
\]

\[
\left(\frac{\partial^2}{\partial u_x \partial u} \right)_{n,k} = \{2\rho \frac{1}{T}T - (1+\rho^2)\frac{T^'Z}{\sigma_x(1-\rho^2)}\}/\sigma_x (1-\rho^2)^2
\]

\[
I_{35} = E\{-(\partial^2/\partial u_x \partial u)\} = \rho (\frac{1}{1} \frac{1}{w} - 2k)/\sigma_x (1-\rho^2) \tag{2.3.16}
\]
\[(\frac{\partial^2}{\partial \sigma_x \partial \sigma_y}) L_{n,k} = \rho T'Z/\sigma_y (1-\rho^2)\]

\[I_{34} = E \{- (\frac{\partial^2}{\partial \sigma_x \partial \sigma_y}) L_{n,k} \} = - \rho^2 \frac{1'}{y'_{-x}} \frac{w}{\sigma_x \sigma_y} (1-\rho^2) \quad (2.3.17)\]

\[(\frac{\partial^2}{\partial \rho^2}) L_{n,k} = \{k(1+\rho^2)(1-\rho^2) - (1+3\rho^2) (T'T + Z'Z) \]

\[+ 2\rho(3 + \rho^2) T'Z \}/(1-\rho^2)^3\]

\[I_{55} = E\{ - (\frac{\partial^2}{\partial \rho^2}) L_{n,k} \} = \{2k\rho^2 + (1-\rho^2) \frac{1'}{z} \frac{w}{\sigma_y} (1-\rho^2) \quad (2.3.18)\]

\[(\frac{\partial^2}{\partial \sigma \partial \mu_y}) L_{n,k} = 1' \{z \rho \sigma - (1+\rho^2) T'Z \}/\sigma_y (1-\rho^2)\]

\[I_{52} = E\{ - (\frac{\partial^2}{\partial \rho \sigma_y}) L_{n,k} \} = - \rho \frac{1'}{y} \frac{w}{\sigma_y} (1-\rho^2) \quad (2.3.19)\]

\[(\frac{\partial^2}{\partial \sigma \sigma_y}) L_{n,k} = \{2z' \frac{\sigma_z}{z} - (1+\rho^2) T'Z \}/\sigma_y (1-\rho^2)^2\]

\[I_{54} = E\{ - (\frac{\partial^2}{\partial \sigma \sigma_y}) L_{n,k} \} = - \rho \frac{1'}{y} \frac{w}{\sigma_y} (1-\rho^2) \quad (2.3.20)\]

\[(\frac{\partial^2}{\partial \mu_y^2}) L_{n,k} = -(n-k) \sigma_y^{-2} \phi_1(Z_{n,k}) [1-\phi_1(Z_{n,k})]^{-1}\]

\[\{\phi_1(Z_{n,k}) [1-\phi_1(Z_{n,k})]^{-1} - Z_{n,k}\} - k/\sigma_y^2 (1-\rho^2)\]

\[I_{22} = E \{- (\frac{\partial^2}{\partial \mu_y^2}) L_{n,k} \} = \{(n-k)[\psi(n,k; 2,0,2,0)-\psi(n,k; 1,0,1,1)]\]

\[+ k(1-\rho^2)^{-1}/\sigma_y^2 \quad (2.3.21)\]

\[2 2^1 2^2\]
\[ \frac{\partial^2}{\partial y \partial \sigma_y} L_{n,k} = -(n-k) \phi_1(Z,n,k) [1-\phi_1(Z,n,k)]^{-1} \{ 1-z^2 + \phi_1(Z,n,k) [1-\phi_1(Z,n,k)]^{-1} Z_{n,k} \} / \sigma_y^2 \]

\[ - \frac{1'}{(\rho \tau - 2 \tau)} / \sigma_y^2 (1-\rho^2) \]

\[ I_{24} = E\left\{ -(\partial^2/\partial y \partial \sigma_y) \right\} = \{ (n-k) [\psi(n,k;1,0,1,0) - \psi(n,k;1,0,1,2)] \right. \]

\[ + \left. \psi(n,k;2,0,2,1)] + \frac{1'}{\tau} (2-\rho^2)/(1-\rho^2) \} / \sigma_y^2 \]

\[ \frac{\partial^2}{\partial \sigma_y^2} L_{n,k} = -(n-k) Z_n,k \phi_1(Z,n,k) [1-\phi_1(Z,n,k)]^{-1} \{ 2Z_n,k^2 + Z_n,k \phi_1(Z,n,k) [1-\phi_1(Z,n,k)]^{-1} \} / \sigma_y^2 \]

\[ + \left. (2\rho \tau' - 3 \tau' \tau) / \sigma_y^2 (1-\rho^2) + k / \sigma_y^2 \right\} \]

\[ I_{44} = E\left\{ -(\partial^2/\partial \sigma_y^2) L_{n,k} \right\} = \{ (n-k) [2\psi(n,k;1,0,1,1) - (n,k;1,0,1,3)] \right. \]

\[ + \left. \psi(n,k;2,0,2,2)] + \frac{1'}{\tau} (3-2\rho^2)/(1-\rho^2) \} \right. \]

\[ \left. - k \} / \sigma_y^2 \right\} \] (2.3.23)

and where \( I_{j\xi} = I_{\xi j} \) for every \( j,\xi = 1,\ldots,5 \).

Des Raj (1953), who discussed maximum likelihood estimation from bivariate samples when \( Y \) is truncated instead of censored, developed equations similar to these. The expected values of the second partial derivatives are different in that case.

The \( \psi \)-functions, which include all moments of single order statistics, can be evaluated for given \( n,k \) by numerical integration to almost any desired precision (12 significant digits were used for
the computations in this paper). To complete the evaluation of the Fisher information matrix \( \mathbf{I} = (I_{ij}) \), we need to evaluate \( l'w \) and \( l'u \). For this as in Saw [1958], we note that given \( Z_{n,k} \) and disregarding the ordering of \( (Z_{n,1}, \ldots, Z_{n,k-1}) \), the joint distribution of \( (Z_{n,1}, \ldots, Z_{n,k-1}) \) is reducible to that of \( k-1 \) independent random variables from a univariate normal d.f., truncated from above by \( Z_{n,k} \). Denote \( (Z_{n,1}, \ldots, Z_{n,k-1}) \) in random order by \( (S_1, \ldots, S_{k-1}) \), so that for \( 1 \leq i \leq k-1 \),

\[
E(S_i | Z_{n,k}) = \frac{-\phi_i(Z_{n,k})/\phi_1(Z_{n,k})}{E(S_i | Z_{n,k})} = 1 - Z_{n,k} \frac{\phi_1(Z_{n,k})}{\phi_1(Z_{n,k})} \]

and, in general, for \( r \geq 0 \),

\[
E(S_i^{r+2} | Z_{n,k}) = (r+1)E(S_i^r | Z_{n,k}) - Z_{n,k} \frac{\phi_1(Z_{n,k})}{\phi_1(Z_{n,k})} \]

Hence

\[
l'u = \sum_{i=1}^{k-1} E(Z_{n,i}) = E(Z_{n,k}) + \sum_{i=1}^{k-1} E(S_i) \]

\[
= E(Z_{n,k}) + \sum_{i=1}^{k-1} E\{E(S_i | Z_{n,k})\} \]

\[
= E(Z_{n,k}) + \sum_{i=1}^{k-1} E\{-\phi_i(Z_{n,k})/\phi_1(Z_{n,k})\} \]

\[
= \psi(n,k;0,0,0,1) - (k-1)\psi(n,k;1,1,0,0) \]

(2.3.26)
\[
\omega = \sum_{i=1}^{k} E(Z_{n,i}^2) = E(Z_{n,k}^2) + \sum_{i=1}^{k-1} E(S_{i}^2)
\]

\[
= E(Z_{n,k}^2) + \sum_{i=1}^{k-1} E(E(S_{i}^2 | Z_{n,k})]
\]

\[
= E(Z_{n,k}^2) + \sum_{i=1}^{k-1} E(1-Z_{n,k} \phi_1(Z_{n,k})/\phi_1(Z_{n,k}))
\]

\[
= \psi(n,k;0,0,0,2) + (k-1)[1-\psi(n,k;1,1,0,1)].
\]

(2.3.27)

Note that in Saw's notation, \(\psi(n,k; a,a,o,b)\) is \(\psi(k/(n+1),n:a,b)\).

Let \(A_{22}, A_{24}, \text{ and } A_{44}\) denote respectively the linear combinations of \(\psi\)-functions in (2.3.21), (2.3.22) and (2.3.23). For three selected sample sizes, the values for use in determining the information matrix, \(I\) are:

<table>
<thead>
<tr>
<th>(n)</th>
<th>(k)</th>
<th>(A_{22})</th>
<th>(A_{24})</th>
<th>(A_{44})</th>
<th>(1'u)</th>
<th>(1'w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0.618616</td>
<td>0.746412</td>
<td>0.0471597</td>
<td>-7.67491</td>
<td>10.0000</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.412067</td>
<td>-0.00905918</td>
<td>-0.276115</td>
<td>-27.7132</td>
<td>43.0867</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>0.189672</td>
<td>-0.203353</td>
<td>0.157445</td>
<td>-102.906</td>
<td>218.888</td>
</tr>
</tbody>
</table>

The Cramèr-Rao lower bounds for variances of unbiased estimators are presented in the following table for three sample sizes, five values of \(\rho\) and \(\sigma_x = \sigma_y = 1\). These bounds are the diagonal elements of \(I^{-1}\) and also represent approximate variances of the maximum likelihood estimators, \(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\beta}\) (i.e., solutions to
(2.3.5)) as predicted by asymptotic theory.

TABLE 2.1

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>( V[\hat{\mu}_x] )</th>
<th>( V[\hat{\mu}_y] )</th>
<th>( V[\hat{\sigma}_x] )</th>
<th>( V[\hat{\sigma}_y] )</th>
<th>( V[\hat{\rho}] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0.2433</td>
<td>0.0761</td>
<td>0.0500</td>
<td>0.0601</td>
<td>0.2433</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2417</td>
<td>0.0761</td>
<td>0.0514</td>
<td>0.0601</td>
<td>0.2372</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2283</td>
<td>0.0761</td>
<td>0.0618</td>
<td>0.0601</td>
<td>0.1916</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2015</td>
<td>0.0761</td>
<td>0.0775</td>
<td>0.0601</td>
<td>0.1181</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1078</td>
<td>0.0761</td>
<td>0.0787</td>
<td>0.0601</td>
<td>0.0049</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.4598</td>
<td>0.0567</td>
<td>0.0250</td>
<td>0.0345</td>
<td>0.2134</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4557</td>
<td>0.0567</td>
<td>0.0266</td>
<td>0.0345</td>
<td>0.2077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4235</td>
<td>0.0567</td>
<td>0.0385</td>
<td>0.0345</td>
<td>0.1653</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3590</td>
<td>0.0567</td>
<td>0.0562</td>
<td>0.0345</td>
<td>0.0984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1333</td>
<td>0.0567</td>
<td>0.0564</td>
<td>0.0345</td>
<td>0.0032</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>0.6170</td>
<td>0.0508</td>
<td>0.0100</td>
<td>0.0155</td>
<td>0.1410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.6114</td>
<td>0.0508</td>
<td>0.0112</td>
<td>0.0155</td>
<td>0.1370</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5661</td>
<td>0.0508</td>
<td>0.0200</td>
<td>0.0155</td>
<td>0.1081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.4755</td>
<td>0.0508</td>
<td>0.0330</td>
<td>0.0155</td>
<td>0.0630</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1584</td>
<td>0.0508</td>
<td>0.0322</td>
<td>0.0155</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

It is interesting to note that for a large proportion of censoring (small \( k \)) \( \mu_x \) cannot be estimated well from induced order statistics,
yet $\sigma_x$ can be estimated with more accuracy than $\sigma_y$ for small $\rho$. Also, lower bounds for variance of estimators of $\mu_y$ and $\sigma_y$ are independent of $\rho$ as one would expect; $\mu_y$ and $\sigma_y$ can be estimated efficiently with a small number of order statistics whether or not induced order statistics are observable.

The difficulty in estimating $\mu_x$ stems from $\rho$ being unknown, so the selection process on $Y$ causes an unknown location bias in the sample of induced order statistics. For example, if $\rho$ were known to be zero, $\mu_x$ would be estimated by the sample mean with variance $0.02 \sigma^2_x$ at $n=1000, k=50$, instead of $0.6170 \sigma^2_x$.

2.4. Maximum Likelihood Estimation

An iterative method may be used to estimate $\theta$ by using (2.3.5). Des Raj [1953] set up the implicit solutions to the likelihood equations for the case of double truncation on $Y$ from a bivariate normal sample which is similar to our case relating to censoring. Assuming the theorem outlined by Halperin [1952] to be valid for our bivariate case, the maximum likelihood estimator (MLE) will be consistent, asymptotically normally distributed and asymptotically efficient. Note that since $Y_{n,1}, \ldots, Y_{n,k}$ are not independent, knowledge of the asymptotic properties of the MLE may demand additional regularity conditions [viz. Sen [1976]]. Nevertheless, for bivariate normal d.f.'s, optimal properties hold under quite general setups.

The complexity of the solutions to (2.3.5) makes the study of
the properties of the MLE difficult for any finite sample size. It will be shown in a more general setting in section 4.2, however, that the MLE of $\mu_y, \sigma_y$ is the usual MLE based on $Y_{n,1}, \ldots, Y_{n,k}$ alone. Nonetheless, the results obtained by Saw [1961] for estimating $\mu_y$ and $\sigma_y$ in this case implies that the MLE of $\Theta$ will have undesirable properties for small $k$ or $n$. In particular, the MLE may have considerable bias (relative to its standard error). The linearization of the log-likelihood function of $Y_{n,1}, \ldots, Y_{n,k}$ described in Chan [1967], which results in simple and efficient estimators of $\mu_y$ and $\sigma_y$, will solve this problem.

There are other good estimators of $\mu_y$ and $\sigma_y$ besides the ones considered by Chan. For instance, Saw [1959] developed an unbiased estimator of $\mu_y$ which is a linear combination of $Y_{n,k}$ and 
\[(k-1)^{-1} \sum_{i=1}^{k-1} Y_{n,i}\]
and has asymptotic efficiency (a function of $k/(n+1)$) $> 94\%$ for $k/(n+1)$ as low as $.25$. He also considered a linear combination of $\sum_{i=1}^{k-1} (Y_{n,i} - Y_{n,k})^2$ and $\{\sum_{i=1}^{k-1} (Y_{n,i} - Y_{n,k})\}^2$ as an unbiased estimator of $\sigma_y^2$ which is of $100\%$ asymptotic efficiency for all $k/(n+1)$. Linear unbiased estimators of $\sigma_y$ have also been developed (for example see Gupta [1952]).

Since for the estimation of $(\mu_y, \sigma_y)$, the induced order statistics do not contribute any information, it seems natural to estimate $(\mu_y, \sigma_y)$ by some efficient procedure based only on $\bar{Y} = (Y_{n,1}, \ldots, Y_{n,k})$. Denote the estimators by $(\mu_y^*, \sigma_y^*)$ and then consider only the first three equations in (2.3.5) wherein $\mu_y, \sigma_y$ are replaced by $\mu_y^*, \sigma_y^*$. 

Equating these functions to 0, we get three equations in three unknown parameters $(\mu_x, \sigma_x, \rho)$ and denote these solutions $(\mu_x^*, \sigma_x^*, \rho^*)$ as the modified MLE. For this purpose, we write

$$
\bar{x} = k^{-1} \sum_{i=1}^{k} x_{n[i]}, \quad \bar{y} = k^{-1} \sum_{i=1}^{k} y_{n[i]}, \quad S_x^2 = k^{-1} \sum_{i=1}^{k} (x_{n[i]} - \bar{x})^2, \\
S_y^2 = k^{-1} \sum_{i=1}^{k} (y_{n[i]} - \bar{y})^2 \text{ and } S_{xy} = k^{-1} \sum_{i=1}^{k} (x_{n[i]} - \bar{x})(y_{n[i]} - \bar{y}).
$$

(2.4.1)

Equating $(\partial/\partial \mu_x)_{n,k}$ in (2.3.5) to 0 we get

$$
(\bar{x} - \mu_x^*)/\sigma_x^* = \rho^*(\bar{y} - \mu_y^*)/\sigma_y^*.
$$

(2.4.2)

Equating $(\partial/\partial \sigma_x)_{n,k}$ to 0, we get

$$
\sigma_x^* = S_x^2 + \sigma_y^2 (\bar{x} - \mu_x^*)^2 / (\bar{y} - \mu_y^*)^2 - S_{xy} (\bar{x} - \mu_x^*) / (\bar{y} - \mu_y^*).
$$

(2.4.3)

Combining (2.4.2) and (2.4.3) we have

$$
\rho^* S_{xy}/\sigma_x^* \sigma_y^* - \rho^2 = S_x^2 / \sigma_x^2 - 1.
$$

(2.4.4)

Equating $(\partial/\partial \rho)_{n,k}$ to 0,

$$
\rho^* (1 - \rho^2) - \rho^* [\{S_x^2 + (\bar{x} - \mu_x^*)^2\} / \sigma_x^2 + \{S_y^2 + (\bar{y} - \mu_y^*)^2\} / \sigma_y^2]
$$
Using (2.4.2) this simplifies to

\[
\rho^* (1-\rho^*) - \rho^* \left( \frac{\sigma_{x}^2}{\sigma_{x}^2} + \frac{\sigma_{y}^2}{\sigma_{y}^2} \right) + (1-\rho^*) S_{xy} \frac{\sigma_{x}^2}{\sigma_{x}^2} \frac{\sigma_{y}^2}{\sigma_{y}^2} = 0. \tag{2.4.6}
\]

Substituting the identity in (2.4.4) into (2.4.6),

\[
\rho^* = \frac{\sigma_{x}^*}{\sigma_{y}^*} \frac{S_{xy}}{S_{xy}} \tag{2.4.7}
\]

The similarity of \( \rho^* \) to the Pearson product moment correlation should be noted, for (2.4.7) can be re-written

\[
\rho^* = \frac{S_{xy}}{S_{xy}} \left[ \frac{S_{xy}}{S_{xy}} \left( \frac{\sigma_{x}^*}{\sigma_{x}^*} \frac{\sigma_{y}^*}{\sigma_{y}^*} \right) \right]. \tag{2.4.8}
\]

From (2.4.4) and (2.4.7) we get

\[
\sigma_{x}^2 = S_{x}^2 + \left( \frac{S_{xy}^2}{S_{y}^2} \right) \left( \frac{\sigma_{y}^*}{\sigma_{y}^*} - 1 \right). \tag{2.4.9}
\]

Substituting (2.4.7) into (2.4.2),
Thus, having estimated $\mu_y, \sigma_y$ by $\mu_y^*, \sigma_y^*$, we may estimate $\mu_x, \sigma_x$ and $\rho$ by (2.4.10), (2.4.9), and (2.4.7), respectively. One nice feature of these estimators is that as $k/n \to 1$ (i.e., the amount of censoring decreases), the estimators approach the ordinary MLE, which are known to be efficient for complete samples of any size. Again, if $\mu_y^*$, $\sigma_y^*$ are MLE based on $Y_{n,1}, \ldots, Y_{n,k}$, the entire set of estimators is MLE.

Since $E(X|Y) = \mu_x + \rho \sigma_x (Y - \mu_y)/\sigma_y$ it follows that $E(X|Y)$ is an unbiased estimator of $\mu_x$ if $\mu_y^*$ is an unbiased estimator of $\mu_y$. Since $V(X|Y) = \sigma_x^2 (1-\rho^2)I_k$, we have:

$$E(\mu_x^*|Y) = \mu_x + \rho \sigma_x (\mu_y^* - \mu_y)/\sigma_y. \quad (2.4.12)$$

And so from (2.4.10),

$$E(\mu_x^*|Y) = \mu_x + \rho \sigma_x (\mu_y^* - \mu_y)/\sigma_y. \quad (2.4.12)$$

Therefore, $\mu_x^*$ is an unbiased estimator of $\mu_x$ if $\mu_y^*$ is an unbiased estimator of $\mu_y$. Since $V(\mu_x|Y) = \sigma_x^2 (1-\rho^2)I_k$, we have:

$$E(\mu_x^*|Y) = \mu_x + \rho \sigma_x (\mu_y^* - \mu_y)/\sigma_y. \quad (2.4.12)$$
Combining (2.4.13) and (2.4.11) we find

\[ \nu(S_{xy} | Y) = \sum_{i=1}^{k} [(Y_{ni} - \bar{Y})^2 / k]^2 \sigma_x^2 (1 - \rho^2) = \sigma_x^2 (1 - \rho^2) \sigma_y^2 / k. \tag{2.4.13} \]

Combining (2.4.13) and (2.4.11) we find

\[ E(S_{xy}^2 | Y) = \nu(S_{xy} | Y) + E(S_{xy} | Y) \nu(Y | Y) = \sigma_x^2 \sigma_y^2 \left( (1 - \rho^2) + \rho^2 S_y^2 / \sigma_y^2 \right) \tag{2.4.14} \]

It can also be shown that \( E(S_{xy}^2 | Y) = \sigma_x^2 \left[ k^{-1} (k-1)(1 - \rho^2) + \rho^2 S_y^2 / \sigma_y^2 \right] \).

Combining this with (2.4.14) and (2.4.9) we have

\[ E(\sigma_x^2) = \sigma_x^2 \left[ 1 + \rho^2 \left( E(\sigma_y^2) / \sigma_y^2 - 1 \right) \right. \]
\[ \left. + k^{-1} (1 - \rho^2) E(\sigma_y^2 / S_y^2 - 2) \right] \tag{2.4.15} \]

and if \( \sigma_y^2 \) is unbiased for \( \sigma_y^2 \),

\[ E(\sigma_x^2) = \sigma_x^2 \left[ 1 + k^{-1} (1 - \rho^2) \left( E(\sigma_y^2 / S_y^2 - 2) \right) \right]. \tag{2.4.16} \]

It can also be shown that

\[ \nu(\mu_x^*) = \sigma_x^2 (1 - \rho^2) \left[ 1 + E((\mu_y^* - \bar{Y})^2 / S_y^2) \right] / k + \rho^2 \sigma_x^2 \nu(\mu_y^*) / \sigma_y^2. \tag{2.4.17} \]

Note that from (2.4.11) if \( \sigma_y^2 \) is unbiased for \( \sigma_y^2 \) then an unbiased estimator for the population covariance of \( X \) and \( Y \), \( \rho \sigma_x \sigma_y \), is given by
In order to estimate the expected values of $\sigma_x^*$ and $\rho^*$ as well as the variances of $\mu_x^*$, $\sigma_x^*$, and $\rho^*$, for each of three different combinations of $(n,k)$ and five different values of $\rho$, 500 random sets of $(X,Y)$ were generated and the statistics were evaluated. $\mu_y^*$, $\sigma_y^*$ were taken as the estimators due to Gupta (1952) in (1.5.14) and (1.5.15). Also, for comparison, the variance of Watterson's estimator of $\mu_y$, $\tilde{\mu}_y$ defined by (1.5.19) was estimated. For $\mu_x^*$ and $\tilde{\mu}_x$ the $u_{n,1}$ in (1.5.14) were approximated by (1.5.9). For simplicity, we have taken $\sigma_x = \sigma_y = 1$. We have also estimated the trace efficiency of the modified MLE, which is the ratio of the sum of the five Cramèr-Rao lower bounds from Table 2.1 to the sum of the estimated variances of the five modified MLE.
Table 2.2. Estimated moments of modified maximum likelihood estimators

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>ρ</th>
<th>$V[\mu_x^*]$</th>
<th>$V[\bar{\mu}_x]$</th>
<th>$V[\mu_y^*]$</th>
<th>$E[\sigma_x^*]$</th>
<th>$V[\sigma_x^*]$</th>
<th>$V[\sigma_y^*]$</th>
<th>$E[\rho^*]$</th>
<th>$V[\rho^*]$</th>
<th>Trace Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0</td>
<td>.281</td>
<td>.297</td>
<td>.074</td>
<td>1.021</td>
<td>.077</td>
<td>.078</td>
<td>-.020</td>
<td>.236</td>
<td>.902</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.1</td>
<td>.287</td>
<td>.293</td>
<td>.092</td>
<td>.998</td>
<td>.074</td>
<td>.088</td>
<td>.060</td>
<td>.213</td>
<td>.884</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.281</td>
<td>.293</td>
<td>.082</td>
<td>1.043</td>
<td>.079</td>
<td>.079</td>
<td>.275</td>
<td>.196</td>
<td>.862</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.265</td>
<td>.276</td>
<td>.087</td>
<td>.998</td>
<td>.102</td>
<td>.086</td>
<td>.449</td>
<td>.172</td>
<td>.749</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.123</td>
<td>.126</td>
<td>.093</td>
<td>1.014</td>
<td>.106</td>
<td>.083</td>
<td>.881</td>
<td>.024</td>
<td>.764</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0</td>
<td>.495</td>
<td>.508</td>
<td>.076</td>
<td>1.038</td>
<td>.042</td>
<td>.045</td>
<td>-.044</td>
<td>.169</td>
<td>.955</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.1</td>
<td>.548</td>
<td>.564</td>
<td>.087</td>
<td>1.059</td>
<td>.047</td>
<td>.051</td>
<td>.080</td>
<td>.176</td>
<td>.859</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.487</td>
<td>.507</td>
<td>.090</td>
<td>1.054</td>
<td>.050</td>
<td>.049</td>
<td>.256</td>
<td>.152</td>
<td>.868</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.400</td>
<td>.420</td>
<td>.081</td>
<td>1.020</td>
<td>.057</td>
<td>.046</td>
<td>.442</td>
<td>.115</td>
<td>.865</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.202</td>
<td>.209</td>
<td>.087</td>
<td>1.014</td>
<td>.080</td>
<td>.048</td>
<td>.885</td>
<td>.009</td>
<td>.667</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>0</td>
<td>.663</td>
<td>.685</td>
<td>.078</td>
<td>1.045</td>
<td>.019</td>
<td>.022</td>
<td>-.016</td>
<td>.114</td>
<td>.931</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.1</td>
<td>.601</td>
<td>.616</td>
<td>.091</td>
<td>1.044</td>
<td>.018</td>
<td>.025</td>
<td>.092</td>
<td>.105</td>
<td>.983</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.624</td>
<td>.648</td>
<td>.094</td>
<td>1.048</td>
<td>.024</td>
<td>.024</td>
<td>.274</td>
<td>.103</td>
<td>.875</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.477</td>
<td>.508</td>
<td>.091</td>
<td>1.030</td>
<td>.032</td>
<td>.025</td>
<td>.462</td>
<td>.066</td>
<td>.923</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.187</td>
<td>.188</td>
<td>.082</td>
<td>.987</td>
<td>.038</td>
<td>.023</td>
<td>.887</td>
<td>.003</td>
<td>.777</td>
</tr>
</tbody>
</table>
From Table 2.2 we see that $\mu^*_x$ is consistently slightly better than $\tilde{\mu}_x$, although a small part of this may have to do with the approximation used for the $u_{ni}$. The efficiency of $\mu^*_x$ by simulation is between 66% and 102% according to the bounds in Table 2.1. As expected, $\mu^*_x$ has large variance for large censoring proportion. On the other hand, $\sigma_x$ is estimated as well as $\sigma_y$ in almost every case and has lower variance in some cases. Also, $\sigma^*_x$ appears to have small bias and is 53-103% efficient. The efficiency of $\rho^*$ is between 69% and 130% for the cases studied except when $\rho = .9$. The bias in $\rho^*$ is very low except for moderate $\rho$ ($\rho = .3$ or $.5$ resulted in maximum estimated bias of .06). Note that $V(\mu^*_x)$ would be even smaller if for example Saw's [1959] estimator of $\mu_y$ were used.

2.5. Test of $H_0: \rho = 0$

2.5.1. Likelihood Ratio Test

From the results of the previous section, if $(\mu^*_y, \sigma^*_y)$ are MLE of $(\mu_y, \sigma_y)$, they are MLE independent of any restriction on $\rho$. Thus the modified MLE of section 2.4 may be incorporated to prescribe a likelihood ratio test for $H_0: \rho = 0$. As before, $\mu^*_y, \sigma^*_y, \mu^*_x, \sigma^*_x, \rho^*$ are the estimators considered in section 2.4 except that now we require $\mu^*_y, \sigma^*_y$ to be true MLE. Over the parameter space restricted by $\rho = 0$, the parallel estimators are $\mu^*_y, \sigma^*_y, \tilde{\mu}_y, \tilde{\sigma}_y, 0$ where
\[ \bar{\mu}_x = \bar{x} \quad \text{and} \quad \bar{\sigma}_x^2 = S_x^2 \] are defined by (2.4.1) \hspace{1cm} (2.5.1.1)

From the definition of \( L_{n,k} \) in (2.3.4), the log ratio of the maximum likelihood under the null hypothesis and the maximum likelihood over the whole parameter space, multiplied by -2, results in a likelihood ratio statistic of

\[ \chi = k \log \left[ \frac{\sigma^*_{x^2}}{\sigma_x^2} (1 - \rho^*_{x^2}) \right] + \bar{\sigma}_x^* \Sigma (x_{n[i]} - \bar{\mu}_x^*)^2 \]

\[ + \sigma_{y}^* \Sigma (y_{n,i} - \bar{\mu}_y^*)^2 \]

\[ - (1 - \rho^*)^{-1} \Sigma (\sigma_{x}^* \Sigma (x_{n[i]} - \bar{\mu}_x^*)^2 - 2 \rho^* \sigma_{x}^* \sigma_{y}^* \Sigma (x_{n[i]} - \bar{\mu}_x^*)(y_{n,i} - \bar{\mu}_y^*) \]

\[ + \sigma_{y}^* (y_{n,i} - \bar{\mu}_y^*)^2 \]

where all summations are from 1 to \( k \). We now use the relation

\[ \Sigma (y_{n,i} - \bar{\mu}_y^*)^2 = k S_y^2 + k(\bar{y} - \bar{\mu}_y^*)^2 \] \hspace{1cm} (2.5.1.3)

and further, by (2.4.10),

\[ \Sigma (x_{n[i]} - \bar{\mu}_x^*)^2 = k S_x^2 + k S_{xy}^2 (\bar{y} - \bar{\mu}_y^*)^2 / S_y^2 \] \hspace{1cm} (2.5.1.4)

\[ \Sigma (x_{n[i]} - \bar{\mu}_x^*)(y_{n,i} - \bar{\mu}_y^*) = k S_{xy}^2 + k S_{xy}^2 (\bar{y} - \bar{\mu}_y^*)^2 / S_y^2 \] \hspace{1cm} (2.5.1.5)
and using (2.4.4) and (2.4.7),

\[ \sigma^2_y (1-\rho^2) = (S_{x'y}^2 - S_{xy}^2)/S_y^2 \]  

(2.5.1.6)

so the first term in (2.5.1.2) is

\[ -k \log (1-r^2) \text{ where } r^* = S_{xy}/S_x S_y \]  

(2.5.1.7)

and \( S_x, S_y, S_{xy} \) are defined by (2.4.1). Thus \( r^* \) is the ordinary product moment correlation of the set \( (X_n[i], Y_n, i), i=1, \ldots, k. \)

The remaining terms of \( \chi \) become

\[
k + \sigma^2_y \sum (Y_n, i - \mu^*_y)^2 - S_y^2 (S_{x'Y}^2 - S_{xy}^2)^{-1} \Sigma (X_n[i] - \mu^*_x)^2
\]

\[
+ 2 S_{xy} \{\sigma_x^* (1-\rho^2) S_y^2\}^{-1} \Sigma (X_n[i] - \mu^*_x) (Y_n, i - \mu^*_y)
\]

\[
- (1-\rho^2)^{-1} \sigma^2_y \sum (Y_n, i - \mu^*_y)^2
\]

Using (2.5.1.3) through (2.5.1.6), we get, on dividing by \( k \),

\[
1 + \sigma^2_y \{S_{y^2} + (\overline{Y} - \mu^*_y)^2\} - (S_{x'y}^2 - S_{xy}^2)^{-1} \{S_{x'y}^2 + S_{xy}^2 (\overline{Y} - \mu^*_y)^2 / S_y^2\}
\]

\[
+ 2 S_{xy} (S_{x'y}^2 - S_{xy}^2)^{-1} \{S_{xy} + S_{xy} (\overline{Y} - \mu^*_y)^2 / S_y^2\}
\]
which after simplification is seen to be zero. Hence the likelihood ratio statistic is given by (2.4.1.7) which is identical to the likelihood ratio statistic for a complete sample of size $k$. This suggests that an "optimum" test of independence using censored data would still be the same test as if the data were not censored and that we may use $r^*$ (or some function of $r^*$) as a test statistic. As with a complete sample it will be advantageous to consider as a test statistic

$$T^* = (k-2)r^2/(1-r^2) = \frac{S_{xy}^2}{(k-2)[S_x^2S_y^2 - S_{xy}^2]}. \quad (2.5.1.8)$$

Note that this test is identical to that one would obtain from the usual conditional regression test based on the conditional distribution in (2.3.2).

2.5.2. Distribution of Test Statistic under Censoring

A statistic for testing the hypothesis that $X$ and $Y$ are independent is $T^*$ in (2.5.1.8). Define

$$U^* = k \frac{S_x^2}{\sigma_x^2} - \sigma_y^{-2} \sum_{i=1}^{k} (Y_{ni} - \bar{Y}_{i})^2. \quad (2.5.2.1)$$

Let $F_\Delta (t;a,b)$ be the non-central F d.f. with degrees of freedom
(DF) \((a,b)\) and non-centrality parameter \(\Delta\) and \(G_{n,k}(u) = P\{U^* \leq u\}\) for \(u \geq 0\). Finally, let

\[
F^*_p(t; a, b) = \int_0^\infty F^2(t; a, b) d G_{n,k}(u), \quad t \geq 0.
\]

Then, we have the following

**Theorem 2.5.1.1:**

For every \(t \geq 0\) and given \(U^*\),

\[
P\{T^* \leq t \mid U^*\} = F_{\rho^2 (1-\rho^2)}^{-1}(t; 1, k-2).
\]

Hence \(P\{T^* \leq t\} = F^*_p(t; 1, k-2), \quad t \geq 0\). Under \(H_0: \rho = 0\),

\[
F^*_p(t; 1, k-2) = F_0(t; 1, k-2)
\]

is the central F d.f. with DF \((1, k-2)\).

**Proof:** From (1.5.3), given \(Y, X_n [1], \ldots, X_n [k]\) are (conditionally)

independently normally distributed with means \(\mu_x + \rho \sigma^{-1}(Y_{n,1} - \mu_y)\), \(1 \leq i \leq k\) and a common variance \(\sigma^2_x (1-\rho^2)\). Hence \(k S_{xy}\) is condi-

tionally normally distributed with mean

\[
(Y - \overline{Y})' \{\mu_x + \rho \sigma^{-1}(Y - \mu_y)\} = k \rho \sigma_s^2 / \sigma_y = \rho \sigma x y u^*
\]

and variance

\[
\sigma^2_x (1-\rho^2) (Y - \overline{Y})' \overline{I}_k (Y - \overline{Y}) = \sigma^2_s^2 (1-\rho^2) u^*.
\]
Therefore

\[ k^2 \frac{S_{xy}^2}{\sigma_{xy}^2} / \{ \sigma_x^2 \sigma_y^2 (1-\rho^2)^2 U^* \} \sim \chi_1^2 (\rho^2 U^* (1-\rho^2)^{-1}), \]  

(2.5.2.4)

where \( \chi_p^2(\Delta) \) stands for the non-central chi-square d.f. with \( p \) DF and non-centrality parameter \( \Delta \). Further,

\[ k^2 (S_{xy}^2 - S_{xy}^2) / \{ \sigma_x^2 \sigma_y^2 (1-\rho^2)^2 U^* \} \]

\[ = \left[ \sigma_x^2 U^* (\tilde{I}_{k-1} - \tilde{1} \tilde{1}') \tilde{x} - \{ \tilde{x}' (\tilde{Y} - \tilde{\mu} \tilde{1}) \}^2 \right] / \{ \sigma_x^2 \sigma_y^2 (1-\rho^2)^2 U^* \} \]

\[ = \tilde{x}' \left[ \sigma_x^2 U^* (\tilde{I}_{k-1} - \tilde{1} \tilde{1}') - (\tilde{Y} - \tilde{\mu} \tilde{1}) (\tilde{Y} - \tilde{\mu} \tilde{1})' \right] \tilde{x} / \{ \sigma_x^2 \sigma_y^2 (1-\rho^2)^2 U^* \} \]

\[ = \tilde{U}' A \tilde{U} \]

(2.5.2.5)

where \( \tilde{U} = \{ \sigma_x^2 \sigma_y^2 (1-\rho^2)^2 U^* \}^{-1/2} \tilde{x} \)

(2.5.2.6)

\[ A = \sigma_y^2 U^* (\tilde{I}_k - \tilde{1} \tilde{1}') = (\tilde{Y} - \tilde{\mu} \tilde{1}) (\tilde{Y} - \tilde{\mu} \tilde{1})' \]

(2.5.2.7)

and given \( \tilde{x}, \tilde{U} \) is conditionally normally distributed with mean vector

\[ \tilde{y} = \{ \sigma_x^2 \sigma_y^2 (1-\rho^2)^2 U^* \}^{-1/2} \{ \mu_x \tilde{1} + \rho \sigma_x \sigma_y^{-1} (\tilde{Y} - \tilde{\mu} \tilde{1}) \} \]

(2.5.2.8)
and dispersion matrix \( k^{-1}S^{-2}I_k \). Further, it can be shown that \( k^{-1}S^{-2}A \) is an idempotent matrix and that the rank of \( \hat{A} = k-2 \).

The non-centrality parameter, \( \gamma' A \gamma \), is zero. Thus given \( \bar{Y} \),

\[
\frac{k(\bar{S}^2 - S^2)}{\chi^2_{k-2}(0)}
\]

which implies its unconditional distribution is also \( \chi^2_{k-2}(0) \).

Finally, noting that \( (\bar{Y} - \bar{Y})' A = 0 \), we obtain from (2.5.2.5) and (2.5.2.6) that given \( \bar{Y} \), \( S^2 \) and \( \{S^2 - S^2\} \) are conditionally independent. Hence (2.5.2.3) follows from (2.5.1.8), (2.5.2.4), and (2.5.2.9), and the fact that this conditional d.f. depends on \( \bar{Y} \) through \( U^* \) alone. Finally, by (2.5.2.3), \( P(T^*_t) = E[P(T^*_t|U^*)] = F_{\rho}(t; 1, k-2) \). Under \( H_0: \rho = 0 \), this is equal to \( F_0(t; 1, k-2) \) for all \( U^* \) and hence, \( F_{\rho} = F_0 \). Q.E.D.

From Theorem 2.5.2.1, \( T^* \) has the same conditional distribution as the statistic from an uncensored sample and under \( H_0: \rho = 0 \), it has the classical (variance ratio) distribution with DF (1, k-2) and the test can be made without any difficulty. Let \( F^*_\alpha \) be the upper 100 \( \alpha \)% point of \( F_0(t; 1, k-2) \) i.e., \( F_0(F^*_\alpha; 1, k-2) = 1 - \alpha \). Then the power of the test based on \( T^* \) (or \( r^* \)) and corresponding to the level of significance \( \alpha (0 < \alpha < 1) \) is given by

\[
1 - F^*_\rho(F^*_\alpha; 1, k-2).
\]
In general, for $k < n$, $G_{n,k}$ the d.f. of $U^*$ is quite complicated, and hence, analytical solutions for (2.5.2.10) are difficult to obtain. However, one can obtain one or two moment approximations for it by computing the moments of $U^*$ by the formulae of Saw [1958]. Unfortunately these formulae contain additional errors not reported in the Corrigenda to Saw [1958]. Let $\psi_{ab} = \psi(n,k; a,a,0,b)$, defined by (2.3.6). Then, noting that $U^* = \left( Z - \bar{Z} \right)' \left( Z - \bar{Z} \right)$ where $Z = (Z_{n,1}, \ldots, Z_{n,k})'$, $\bar{Z} = \frac{1}{k} \sum_{i=1}^{k} Z_{n,i}$, with $Z_{n,i} = \frac{(Y_{n,i} - \mu_y)}{\sigma_y}$, 1 $\leq i \leq k$ relate to the standard normal d.f., we find after deriving Saw's formulae that

$$E(U^*) = k^{-1}(k-1)\left[(k-1) - (k-3)\psi_{11} + \psi_{22}(k-2)\psi_{20}\right] = kE(S^2/\sigma_y^2)$$

and that the coefficient of $\psi_{13}$ in Saw's expression (3.6) concerning $E(U^{*2})$ should be $-(3k^2 - 14k + 15)$ instead of $(5k^2 - 10k + 1)$. These findings have been confirmed by Professor Saw in a personal communication. Note that for a complete sample ($k = n$), $E(U^*) = k-1$, $V(U^*) = 2(k-1)$.

Letting $\zeta = \rho^2(1-\rho^2)^{-1}$ and using the properties of the F-distribution,

$$E(T^* | U^*) = (1 + \zeta U^*)(k-2)/(k-4) \quad (2.5.2.11)$$

$$E(T^{*2} | U^*) = (3 + 6\zeta U^* + \zeta^2 U^{*2})(k-2)^2/(k-4)(k-6) \quad (2.5.2.12)$$
which imply that

\[ E(T^*) = \{1 + \zeta E(U^*)\} \frac{(k-2)}{(k-4)} \]  \hspace{1cm} (2.5.2.13)

\[ E(T^{*2}) = \{3 + 6 \zeta E(U^*) + \zeta^2 E(U^{*2})\} \frac{(k-2)^2}{(k-4)(k-6)}. \]  \hspace{1cm} (2.5.2.14)

For the one moment approximation, we suppose that \( T^* \sim F_{1,k-2}(\lambda) \)
where \( F_{a,b}(\Delta) \) has the d.f. \( F_\Delta(t; a, b) \) defined after (2.5.2.1).

Thus we require \( \lambda = \zeta E(U^*) \) and replace the distribution of \( T^* \) with
\( F_{1,k-2}(\zeta E(U^*)) \). For the two moment approximation, we replace the
distribution of \( T^* \) by \( aF_{b,k-2}(0) \) and equate its moments to (2.5.2.13),
(2.5.2.14). Thus

\[ E(T^*) = a\frac{(k-2)}{(k-4)} \equiv \{1 + \zeta E(U^*)\} \frac{(k-2)}{(k-4)} \text{ so} \]

\[ a = 1 + \zeta E(U^*) \]  \hspace{1cm} (2.5.2.15)

\[ E(T^*) = a^2\frac{(k-2)^2(b+2)}{b(k-4)(k-6)} \text{ which implies} \]

\[ b = 2a^2/\{\zeta^2 V(U^*) + 4 \zeta E(U^*) + 2\}. \]  \hspace{1cm} (2.5.2.16)

Note that when \( \rho = 0, a = b = 1 \) and \( \lambda = 0 \) so that both approxima-
tions give the correct null d.f. Also, we limit ourselves to \( k > 4 \).
for the one-moment approximation or \( k > 6 \) for the two moment approximation.

For the different \( n,k \) the moments of \( U^* \) are as follows:

<table>
<thead>
<tr>
<th>( n )</th>
<th>( k )</th>
<th>( \mathbb{E} \left( U^* \right) )</th>
<th>( \mathbb{V} \left( U^* \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>3.43246</td>
<td>3.87797</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>4.22029</td>
<td>3.60135</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>6.79179</td>
<td>4.41258</td>
</tr>
</tbody>
</table>

The resulting \( (a,b) \) values are:

<table>
<thead>
<tr>
<th>( n=20 )</th>
<th>( k=10 )</th>
<th>( n=100 )</th>
<th>( k=20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>( a )</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.035</td>
<td>1.043</td>
<td>1.003</td>
</tr>
<tr>
<td></td>
<td>1.339</td>
<td>1.417</td>
<td>1.057</td>
</tr>
<tr>
<td></td>
<td>2.144</td>
<td>2.407</td>
<td>1.312</td>
</tr>
<tr>
<td></td>
<td>15.633</td>
<td>18.992</td>
<td>3.731</td>
</tr>
<tr>
<td>( b )</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.001</td>
<td>1.002</td>
<td>1.004</td>
</tr>
<tr>
<td></td>
<td>1.057</td>
<td>1.085</td>
<td>1.182</td>
</tr>
<tr>
<td></td>
<td>1.312</td>
<td>1.443</td>
<td>1.845</td>
</tr>
</tbody>
</table>

For each \( n,k,\rho \) the power of the \( r^* \)-test at the 5% level under censoring was estimated by generating 2000 \( k \times 1 \) random \( Y \) vectors. The exact upper 5% critical values were used and the power was
calculated by averaging 2000 conditional non-central F probabilities to estimate the unconditional distribution of $U^*$ as in (2.5.2.2).
In order to compare the power of the test using censored data to that using a complete bivariate sample, the complete sample size $n^*$ required to achieve the same simulated power using the Fisher-Yates $Z$-test is given.

TABLE 2.3
ESTIMATED POWER OF 5% $r^*$-TEST FOR CENSORED DATA

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$\rho$</th>
<th>Power from One Moment Approximation</th>
<th>Power from Two Moment Approximation</th>
<th>Simulated Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>.1</td>
<td>.0531</td>
<td>.0531</td>
<td>.0532</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.0810</td>
<td>.0802</td>
<td>.0814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.1570</td>
<td>.1522</td>
<td>.1581</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.9160</td>
<td>.8331</td>
<td>.8256</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.1</td>
<td>.0544</td>
<td>.0544</td>
<td>.0544</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.0939</td>
<td>.0921</td>
<td>.0942</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.2025</td>
<td>.1929</td>
<td>.2026</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.9796</td>
<td>.9468</td>
<td>.9349</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>.1</td>
<td>.0576</td>
<td>.0575</td>
<td>.0576</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.1266</td>
<td>.1219</td>
<td>.1273</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.3140</td>
<td>.2967</td>
<td>.3140</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.9995</td>
<td>.9985</td>
<td>.9956</td>
</tr>
</tbody>
</table>
The two-moment approximation is better for $\rho = .9$, but the one-moment approximation is better overall in estimating the power function. This is fortunate because the one-moment formula is easier to calculate since only the first moment of $U^*$ need be computed. Note that a complete sample of about one-fifth the size ($k$) of the observed censored sample would achieve the same power when $n = 1000$ in which a high proportion of observations are censored. However, because of considerations of cost of failures or controlling the duration of the experiment, this type of sacrifice may be necessary.
CHAPTER III

STATISTICAL INFERENCE FOR CENSORED MULTIVARIATE NORMAL DISTRIBUTIONS BASED ON INDUCED ORDER STATISTICS

3.1. Introduction.

Let \( (X_1,Y_1), \ldots, (X_n,Y_n) \) be independent and identically distributed random \((p+1)\)-vectors where \( X_1 \) is a \( p \)-vector. There is no difficulty in (1.2.1) in defining \( X_{n[i]} \) as a \( 1 \times p \) induced order statistic where \( X_{n[i]} = X_j \) if \( Y_{n,i} = Y_j \) for \( i,j=1,\ldots,n \). In this chapter we will extend the results of Chapter 2 by supposing that \( (X_1,Y_1) \) has a \((p+1)\)-variate normal d.f. with density \( \phi_\bar{x}(x,y) \) where \( \bar{x} \) denotes the mean, variance, and covariance parameters. Maximum likelihood estimation as in section 2.4 will be used to estimate \( \bar{x} \) and a likelihood ratio test of \( H_0: \bar{x} \) and \( Y \) are independent will be developed as in section 2.5.1. Other tools of statistical inference commonly used in complete data problems, namely tests of partial and canonical correlation, will be developed for censored multivariate normal data.

3.2. Maximum Likelihood Estimation

By (1.4.1), the joint (conditional) pdf of
\( \bar{X} = (X'_{n[1]}, X'_{n[2]}, \ldots, X'_{n[k]})' \) given \( \bar{Y} = (Y_{n[1]}, \ldots, Y_{n,[k]})' \) is

\[
\prod_{i=1}^{k} \phi_{n[i]}(x_i) = \prod_{i=1}^{k} \left\{ \frac{\phi_{n[i]}(y_{n,i})}{\phi_{n[i]}((Y_{n,i} - \mu_y)/\sigma_y)} \right\} \]

(3.2.1)

where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \) and \( \phi_{n,i}(x_i) \) is the conditional density of \( x_i \) given \( y_{n,i} \). Multiplying (3.2.1) by (2.3.2) we obtain the joint pdf of \( \bar{X} \) and \( \bar{Y} \),

\[
n_{[k]} \prod_{i=1}^{k} \phi_{n,i}(x_i, y_i) \left[ 1 - \phi_{n,i}((y_{k} - \mu_y)/\sigma_y) \right]^{n-k},
\]

(3.2.2)

\( y_1 \leq y_2 \leq \ldots \leq y_k \)

where \( n_{[k]} \) is defined in (2.3.2). The log likelihood function based on (3.2.2) is, aside from constant terms,

\[
L_{n,k} = (n-k) \log \left[ 1 - \phi_{n,k}((Y_{n,k} - \mu_y)/\sigma_y) \right] - (k/2) \log |\Sigma|
\]

(3.2.3)

\[
- \frac{1}{2} \sum_{i=1}^{k} (y_i - \mu_y)' \Sigma^{-1} (y_i - \mu_y)
\]

where \( y_i = (x_{n[i]}, y_{n,i}) \), \( \Sigma \) is the dispersion matrix of \((x_i, y_i)\), and \( \mu = (\mu_x, \mu_y) \) where \( \mu_x = E(x_i) \), \( \mu_y = E(y_i) \). Partition \( \Sigma \) as

\[
\Sigma = \begin{bmatrix}
\Sigma_{xx} & \Sigma_{xy} \\
\Sigma_{yx} & \Sigma_{yy}
\end{bmatrix}
\]

(3.2.4)
where $V_{xx}$ is the $p \times p$ dispersion matrix of $X_i$, $V_{xy}$ is the $p \times 1$ covariance vector of $X_i$ and $Y_i$, and $\sigma^2_y$ is the (scalar) variance of $Y_i$. Following section 2.4 we estimate $\mu_y$ and $\sigma^2_y$ using the vector $Y$ only; denote these estimators by $\hat{\mu}_y$ and $\hat{\sigma}^2_y$ respectively. For simplicity, the * will be dropped for the present. So we must estimate $\mu_x$, $V_{xx}$, and $V_{xy}$ and we consider the non-fixed terms of $L_{n,k}$, when multiplied by $-2$ to be:

$$k \log |\Sigma| + \sum_{i=1}^{k} (\mu_i - \hat{\mu})' \Sigma^{-1} (\mu_i - \hat{\mu}) \quad (3.2.5)$$

Now define

$$\overline{X} = k^{-1} \sum_{i=1}^{k} X_{n[i]}, \quad \overline{Y} = k^{-1} \sum_{i=1}^{k} Y_{n[i]}, \quad \overline{\Sigma} = (\overline{X}, \overline{Y})$$

$$S_{xx} = k^{-1} \sum_{i=1}^{k} (X_{n[i]} - \overline{X}) (X_{n[i]} - \overline{X})', \quad S_{xy} = k^{-1} \sum_{i=1}^{k} (X_{n[i]} - \overline{X}) (Y_{n[i]} - \overline{Y})$$

$$S_{yy} = k^{-1} \sum_{i=1}^{k} (Y_{n[i]} - \overline{Y})^2 \quad (3.2.6)$$

$$\Sigma = \begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy}' & S_{yy} \end{bmatrix}$$

After dividing by $k$, (3.2.5) can be written

$$\log |\Sigma| + \text{tr} \Sigma^{-1} S + (\mu - \overline{\mu})' \Sigma^{-1} (\mu - \overline{\mu}) \quad (3.2.7)$$
We find that \(|\psi| = \sigma_y^2 |A^{-1}|\) and

\[
\psi^{-1} = \begin{bmatrix}
    \tilde{A} & -\sigma_y^2 A V_{xy} \\
    -\sigma_y^2 V_{xy}' A & \sigma_y^2 + \sigma_y^4 V_{xy}' V_{xy}
\end{bmatrix}
\]  \hspace{1cm} (3.2.8)

\[
\text{tr} \psi^{-1} S = \sigma_y^2 S_{yy} - 2\sigma_y^2 S_{xy}' A V_{xy} + \sigma_y^4 S_{yy} V_{xy}' A V_{xy} + \text{tr} A S_{xx}
\]  \hspace{1cm} (3.2.9)

\[
(\mu - \bar{\mu})' \psi^{-1} (\mu - \bar{\mu}) = (\mu - \bar{\mu})' A(\mu - \bar{\mu}) - 2\sigma_y^2 (\mu - \bar{\mu})' A V_{xy}' (\mu - \bar{\mu})
\]

\[
+ \sigma_y^2 (\mu - \bar{\mu})^2 + \sigma_y^4 (\mu - \bar{\mu}) V_{xy}' A V_{xy} (\mu - \bar{\mu})
\]  \hspace{1cm} (3.2.10)

where the symmetric matrix \( A \) is defined to be

\[
(\psi_{xx} - \sigma_y^2 V_{xy} V_{xy}')^{-1}
\]  \hspace{1cm} (3.2.11)

Therefore, ignoring fixed terms, (3.2.7) implies

\[
L_{n,k}^{\alpha L} = \log |A^{-1}| \sim -2\sigma_y^2 S_{xy}' A V_{xy} + \sigma_y^4 S_{yy} V_{xy}' A V_{xy} + \text{tr} A S_{xx}
\]

\[
+ (\mu - \bar{\mu})' A(\mu - \bar{\mu}) - 2\sigma_y^2 (\mu - \bar{\mu})' A V_{xy}' (\mu - \bar{\mu})
\]

\[
+ \sigma_y^4 (\mu - \bar{\mu})^2 V_{xy}' A V_{xy}
\]  \hspace{1cm} (3.2.12)
\[ \frac{\partial L}{\partial \mu_x} = 2 \Delta \left( \mu_x - \overline{x} \right) - 2 \sigma_{y}^{-2} \Delta_{V_{xy}} \left( \mu_y - \overline{y} \right) \] (3.2.13)

so

\[ \mu_x - \overline{x} = \sigma_{y}^{-2} \Delta_{V_{xy}} \left( \mu_y - \overline{y} \right). \] (3.2.14)

Let \( \nu_{ij} \) denote an element of \( V_{xx} \), \( i,j = 1, \ldots, p \). Also let

\[ \Delta_{ij} = \frac{\partial V_{xx}}{\partial \nu_{ij}}. \] \( \Delta_{ij} \) is all zeros except for elements \((i,j)\) and \((k,i)\), which are unity. Using

\[ \frac{\partial \Delta}{\partial \nu_{ij}} = -A \Delta_{ij} A \] (3.2.15)

\[ \log |A^{-1}|/\partial \nu_{ij} = \text{tr} A \Delta_{ij} \] (3.2.16)

we have

\[ \frac{\partial L}{\partial \nu_{ij}} = \text{tr} A \Delta_{ij} + 2 \sigma_{y}^{-2} S^t A \Delta_{ij} A V_{xy} - \sigma_{y}^{-4} S V' A \Delta_{ij} A V_{xy} \]

\[ - \text{tr} A \Delta_{ij} A S_{xx} - (\mu_x - \overline{x})' A \Delta_{ij} A (\mu_x - \overline{x}) \]

\[ + 2 \sigma_{y}^{-2} (\mu_x - \overline{x})' A \Delta_{ij} A V_{xy} (\mu_y - \overline{y}) - \sigma_{y}^{-4} (\mu_y - \overline{y})^2 V' A \Delta_{ij} A V_{xy} \]

\[ = \text{tr} \left[ \{ A + 2 \sigma_{y}^{-2} A V_{xy} S^t A \} - \sigma_{y}^{-4} S V' A \Delta_{ij} A V_{xy} \right] \]
\[ - A(\mu_x - \overline{X})(\mu_x - \overline{X})'A + 2\sigma_y^{-2}(\mu_y - \overline{Y})AV_{xy}(\mu_x - \overline{X})'A \]

\[ - \sigma_y^{-4}(\mu_y - \overline{Y})^2A_{xy}V'_{xy}A_{ij} \]

\[ = tr(D \Delta_{ij}) \quad (3.2.17) \]

Since this trace is zero for every \( i, j = 1, \ldots, p \) we conclude by the Lemma of Smith [1978] that \( D \) must be zero. Using (3.2.14) this implies

\[ A + 2\sigma_y^{-2}A_{xy}S'_{xy}A - \sigma_y^{-4}S_{yy}A_{xy}V'_{xy}A_{xx}A = 0_{p \times p} \quad (3.2.18) \]

Postmultiplying by \( \sigma_y^4A^{-1} \) yields

\[ \sigma_y^4I_p + A - 2\sigma_y^{-2}A_{xy}S'_{xy} - \sigma_y^{-4}S_{yy}A_{xy}V'_{xy} - \sigma_y^{-4}S_{xx} = 0 \quad (3.2.19) \]

or

\[ \sigma_y^4[S_{yy}V_{xy}V'_{xy} + \sigma_y^4S_{xy} - 2\sigma_y^{-2}S_{xy}V'_{xy}]^{-1} = A \quad (3.2.20) \]

\[ \sigma_y^{-4}[S_{yy}V_{xy}V'_{xy} + \sigma_y^4S_{xy} - 2\sigma_y^{-2}S_{xy}V'_{xy}] = A^{-1} \]

\[ V_{xy} - \sigma_y^{-2}V_{xy}V'_{xy} \quad (3.2.21) \]
Therefore,

\[ V_{xx} = S_{xx} - 2 \sigma_y^{-2} S_{xy} + \sigma_y^{-4} (S_{yy} + \sigma_y^2) V' \]  \hspace{1cm} (3.2.22)

Next we need to find the derivative of (3.2.12) with respect to \( V_{xy} \).

We have

\[ \log |A^{-1}|/\partial V_{xy} = (\partial/\partial V_{xy}) \text{tr} \{ B(-\sigma_y^{-2}) V_{xy} V' \} \bigg|_{B=A} \]

\[ = - \sigma_y^{-2} (\partial/\partial V_{xy}) V' B V' \bigg|_{B=A} = -2\sigma_y^{-2} A V \]  \hspace{1cm} (3.2.23)

\[ (\partial/\partial V_{xy}) S'_{xy} A V = (\partial/\partial V_{xy}) S'_{xy} (-B)(-\sigma_y^{-2} V_{xy} V') B Q \bigg|_{B=A, Q=V} \]

\[ + (\partial/\partial V_{xy}) S'_{xy} B V \bigg|_{B=A} \]

\[ = (\partial/\partial V_{xy}) \sigma_y^{-2} V' B Q S'_{xy} B V \bigg|_{B=A, Q=V} + A S \]

\[ = 2 \sigma_y^{-2} A V' S'_{xy} A V + A S \]  \hspace{1cm} (3.2.24)

Similarly

\[ (\partial/\partial V_{xy}) V' A V = 2 \sigma_y^{-2} A V' A V + 2 A V \]  \hspace{1cm} (3.2.25)

\[ (\partial/\partial V_{xy}) \text{tr} A S_{xx} = 2 \sigma_y^{-2} A S_{xx} A V \]  \hspace{1cm} (3.2.26)
\[
(\partial/\partial_{\bar{V}_{xy}})(\mu_{x, \bar{X}} \bar{A} (\mu_{x, \bar{X}})) = 2 \sigma_{y}^{-2} \bar{A} (\mu_{y, \bar{X}})(\mu_{x, \bar{X}}) \bar{A} V_{xy}
\]
\[
= 2 \sigma_{y}^{-4} (\mu_{y, \bar{X}})^{2} V_{xy} V'_{xy} A V_{xy}
\]

(3.2.27)

using (3.2.14).

\[
(\partial/\partial_{\bar{V}_{xy}})(\mu_{x, \bar{X}}) \bar{A} V_{xy} = 2 \sigma_{y}^{-2} \bar{A} V_{xy} (\mu_{x, \bar{X}}) \bar{A} V_{xy} + A(\mu_{x, \bar{X}})
\]
\[
= 2 \sigma_{y}^{-4} (\mu_{y, \bar{X}}) V_{xy} V'_{xy} A V_{xy}
\]
\[
+ \sigma_{y}^{-2} (\mu_{y, \bar{X}}) A V_{xy}
\]

(3.2.28)

using (3.2.14). Combining (3.2.23) through (3.2.28) and premultiplying by \( \frac{1}{2} \sigma_{A}^{-1} \),

\[
\frac{\partial L}{\partial \bar{V}_{xy}} \alpha (\sigma_{y}^{2} S_{yy} - \sigma_{y}^{4} V_{xy}) - \sigma_{y}^{4} S_{y-xy}
\]
\[
+ [\sigma_{y}^{4} S_{y-xx} + S_{yy} V_{xy} V'_{xy} - 2\sigma_{y}^{2} V_{xy} S'_{xy}] A V_{xy} = 0.
\]

(3.2.29)

By (3.2.20) the expression in brackets is \( \sigma_{A}^{4} \bar{A}^{-1} \). Solving for \( \bar{V}_{xy} \)

and adding the * notation to denote the estimators we have

\[
\bar{V}_{xy}^{*} = \sigma_{y}^{*2} S_{-1}^{y} S_{yy} \bar{S}_{xy}
\]

(3.2.30)
Substituting (3.2.30) into (3.2.14),

\[ \mu_x^* = \bar{X} = (\mu_y^* - \bar{Y})S_{yy}^{-1}S_{xy} \quad (3.2.31) \]

Substituting (3.2.30) into (3.2.22),

\[ \nu_x^* = \bar{S} = S_{xx} - S_{yy}^{-1}(\sigma_y^2 - S_{xy})S_{xy} \quad (3.2.32) \]

These modified MLE's are direct extensions of those for the bivariate case given respectively in (2.4.18), (2.4.10), and (2.4.9).

Using (3.2.1),

\[ E_{X|n[i]} = \mu_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)S_{xy} \quad (3.2.33) \]

which implies

\[ E(X|Y) = \mu_x + \sigma_y^{-2}(\bar{Y} - \mu_y)S_{xy} \quad (3.2.34) \]

and

\[ E(S_{xy}|Y) = k^{-1} E \left[ \sum_{i=1}^{k} X_{n[i]} (Y_{n,i} - \bar{Y}) Y_{n,i} \right] \]

\[ = \sigma_y^{-2} S_{yy} S_{xy} \quad (3.2.35) \]
So from (3.2.31), (3.2.34), and (3.2.35),

$$E(\mu^*_x | Y) = \mu_x + \sigma_{y}^{-2}(\mu^*_y - \mu_y)\sigma_{xy}$$  \hspace{1cm} (3.2.36)

so $\mu^*_x$ is unbiased for $\mu_x$ if $\mu^*_y$ is unbiased for $\mu_y$. From (3.2.30) and (3.2.35),

$$E(\sigma_{xy}^* | Y) = \sigma_{y}^{*2} \sigma_y^{-2} \sigma_{xy}$$  \hspace{1cm} (3.2.37)

so $\sigma_{xy}^*$ is unbiased for $\sigma_{xy}$ if $\sigma_{y}^{*2}$ is unbiased for $\sigma_y^{2}$.

By the invariance property of maximum likelihood estimators we may take as the modified MLE of the vector of correlations between $X$ and $Y$ to be

$$C_{xy}^* = \sigma_{y}^{*} S^{-1} \text{Diag}(\sigma_{xy}^*)^{-1/2} S$$  \hspace{1cm} (3.2.38)

using (3.2.30) and (3.2.32).

3.3. Inference about Independence of $X$ and $Y$

3.3.1. Estimation of Multiple Correlation Coefficient

With the setup in (3.2.4), the square of the population multiple correlation coefficient between $X_i$ and $Y_i$ is

$$\rho^2 = \sigma_{y}^{-2} \sigma_{xy}^{*2} \sigma_{xx}^{-1} \sigma_{xy}^{-1}$$  \hspace{1cm} (3.3.1.1)
The modified maximum likelihood estimator of $\rho^2$ is

$$\rho^2 = \sigma_y^* V_{xy}^* V_{xx}^{-1} V_{xy}^*,$$  \hspace{1cm} (3.3.1.2)

where $\sigma_y^*$ is defined in section 2.4 and $V_{xy}^*$ and $V_{xx}^*$ are defined respectively in (3.2.30) and (3.2.32). Expanding $V_{xy}^*$ and $V_{xx}^*$ this is

$$\rho^2 = \sigma_y^{*2} S_{yy}^{-2} S_{xy}^* \left[ S_{xx}^{-1} + S_{yy}^{-1} (\sigma_y - S_{yy}^{-1} S_{xy} S_{xy}^*)^{-1} S_{yy}^{-1} \right].$$  \hspace{1cm} (3.3.1.3)

Now using for example, Rao [1965, p. 29] if $P$ is a non-singular matrix, $W$ a column vector, $s$ a scaler, and $M = W'P^{-1}W$, then

$$\left( P + s W W' \right)^{-1} = P^{-1} - s (1 + sM)^{-1} P^{-1} W W' P^{-1} \hspace{1cm} (3.3.1.4)$$

which implies further that

$$W'(P + s W W')^{-1} W = M/(1 + sM). \hspace{1cm} (3.3.1.5)$$

Using this, (3.3.1.3) becomes

$$\rho^2 = \sigma_y^{*2} S_{yy}^{-1} S_{xy}^* \left\{ S_{yy}^{-1} + (\sigma_y - S_{yy}^{-1} S_{xy} S_{xy}^*)^{-1} S_{yy}^{-1} \right\}^{-1}.$$

$$\hspace{1cm} (3.3.1.6)$$
or, with \( c = S_{yy} \sigma^*_y \),

\[
\rho^2* = R^2* \{ c^2 + (1-c)R^2* \}^{-1}
\]

(3.3.1.7)

where \( R^2* \) is the ordinary sample squared multiple correlation coefficient defined by

\[
R^2* = S_{yy}^{-1} S_{xy} S_{xx}^{-1} S_{xy}.
\]

(3.3.1.8)

Note that for a complete sample, \( c \) approaches unity since in that case both \( S_{yy} \) and \( \sigma^*_y \) are consistent estimates of the same quantity.

3.3.2. **Likelihood Ratio Test of Independence**

As in section (2.5.1), unrestricted parameter estimates for the \((p+1)\)-variate censored normal distribution are

\[
\theta_* = (\mu^*_y, \sigma^*_y; \mu^*_x, V^*_xx, V^*_xy).
\]

(3.3.2.1)

\( X_i \) and \( Y_i \) are independent if and only if \( V^*_xy = 0 \). The estimators over the parameter space restricted by \( V^*_xy = 0 \) are

\[
\tilde{\theta}_* = (\mu^*_y, \sigma^*_y, \mu^*_x, V^*_xx, 0)
\]

(3.3.2.2)
where $\tilde{\mu}_x = \bar{X}$ and $\tilde{V}_{xx} = S_{xx}$ as defined by (3.2.6).

Now under $H_0: V_{xy} = 0$, $A$ in (3.2.11) evaluated at $\theta = \tilde{\theta}^*$ becomes $S_{xx}^{-1}$. Evaluating $A$ at $\tilde{\theta}^*$ using (3.2.30) through (3.2.32) we obtain

$$A^*_\sim = (S^{-1}_\sim S_{xy} S_{xx} S')^{-1}$$

with the quantities involved defined in (3.2.6). Using these results along with (3.2.12), the maximum log likelihood under $H_0$ is proportional to

$$p + \log |S_{xx}|$$

and the unrestricted maximum is

$$\log |A^*_{xx}| - S_{yy}^{-1} A^*_x S_{xy} + \text{tr} A^*_x S_{xx}.$$
\[ \text{tr} \left( \left( P + s \, W \, W' \right)^{-1} P \right) = p - s \, M / (1 + s \, M) \]  

(3.3.2.8)

with the notation as in section 3.3.1 where \( P \) is a non-singular \( p \times p \) matrix and \( M = W' \, P^{-1} \, W \). So (3.3.2.7), (3.3.1.5), and (3.3.2.8) imply that (3.3.2.6) equals

\[ \log |S'_{xx}| + \log \left( 1 - S_{yy}^{-1} \, S'_{xy} \, S_{xx}^{-1} \, S_{xy} \right) + p. \]  

(3.3.2.9)

Subtracting the restricted maximum, we obtain as the likelihood ratio criterion

\[ \log (1 - R^2), \quad R^2 = S_{yy}^{-1} \, S_{xy} \, S_{xx}^{-1} \, S_{xy} \]  

(3.3.2.10)

and \( R^2 \) is the ordinary squared multiple correlation coefficient.

Thus the test for independence in multivariate normal complete samples is optimal for testing independence with censored multivariate normal samples in the sense that this test statistic is the likelihood ratio criterion.

### 3.3.3. Distribution of Test Statistic under Censoring

We take as a test statistic for testing \( H_0: V_{xy} = 0 \)

\[ T = \frac{R^2 / p}{(1 - R^2) / (k - 1 - p)}, \]  

(3.3.3.1)
where $R^2$ is defined in (3.3.1.8) and $p$ is the dimension of $X_n[1]$. Let $F_\Delta(t; a, b)$ denote the non-central F d.f. as in section 2.5.2. Let $\rho$ denote the population multiple correlation coefficient of $X$ and $Y$ in (3.3.1.1) and

\[ U^* = k \frac{\sigma_y^2}{y} S_{yy} = \frac{\sigma_y^2}{y} \sum_{i=1}^{k} (Y_{n, i} - \overline{Y})^2 \]  

(3.3.3.2)

\[ \lambda = \frac{U^* \rho^2}{(1-\rho)^2} \]  

(3.3.3.3)

\[ G_{n,k}(u) = P\{U^* \leq u\}, \ u \geq 0 \]  

(3.3.3.4)

Then with $F^*_\rho(t; a, b)$ defined in (2.5.2.2) we have the following

**Theorem 3.3.3.1:**

For every $t \geq 0$

\[ P\{T^* \leq t \mid U^*\} = F_\lambda(t; p, k-1-p), \ t \geq 0. \]  

(3.3.3.5)

Hence $P\{T^* \leq t\} = F^*_\rho(t; p, k-1-p), \ t \geq 0$. Under $H_0: V_{xy} = 0$, $F^*(t; p, k-1-p) = F_0(t; p, k-1-p)$ is the central F d.f. with DF $(p, k-1-p)$.

**Proof:** By (3.3.1.5) with $F = S_{xx}, \ W = S_{xy}, \ s = S^{-1}_{yy}$, 

\[ \overline{Y} \]
From (3.2.1), (3.2.4), and (3.2.33), conditional on $Y_{n,i}$, $X_n[i]$ has a $p$-variate normal distribution with mean

$$\mu_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)\Sigma_{xy}$$

and variance matrix

$$\Sigma_{xx} - \sigma_y^{-2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}$$

and the $X_n[i]$, $1 \leq i \leq k$ are independent. So $S_{xy}$ is conditionally $p$-variate normal with mean

$$k^{-1} \sum_{i=1}^{k} \{ \mu_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)\Sigma_{xy} \} (Y_{n,i} - \bar{Y}) = \sigma_y^{-2} \Sigma_{yy}^{-1} \Sigma_{xy}$$

and variance

$$k^{-2} \sum_{i=1}^{k} \{ \Sigma_{xx} - \sigma_y^{-2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy} \} (Y_{n,i} - \bar{Y})^2 = k^{-1} \Sigma_{yy}^{-1} \Sigma_{xx} - \sigma_y^{-2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}.$$  

(3.3.3.10)

Let $L$ be any non-zero $p$-vector and $Q_i = X_n[i]L$, $1 \leq i \leq k$. Conditional on $Y_{n,i}$, $Q_i$ is univariate normal with

$$E(Q_i) = L'\mu_x + \sigma_y^{-2}(Y_{n,i} - \mu_y)L' \Sigma_{xy}$$

(3.3.3.11)
V(Q₁) = L'(Vₓₓ − σ₂ −²ᵧᵧ) \cdot L \tag{3.3.3.12}

and Q₁, ..., Qₖ are independent. Define \( Q = (Q₁, ..., Qₖ)' \). Conditional on \( Y \), \( Q \) is \( k \)-variate normal with moments

\[
\begin{align*}
E(Q) &= L' \muₓ L + \sigma⁻²ᵧᵧ Vᵧᵧ (Y - \muᵧ) L \\
V(Q) &= L'(Vₓₓ − σ⁻²ᵧᵧ) \cdot L \cdot Iₖ
\end{align*}
\tag{3.3.3.13, 3.3.3.14}
\]

Let \( H = Iₖ − k⁻¹ X'X \) so that \( H² = H \). Then \( Sₓₓ = k⁻¹ X'HX \) and \( Sₓᵧ = k⁻¹ X'HY \) with \( Xₓk \times p \), \( Y_k × 1 \) defined in section 3.2. Note that \( Q = X'L \). Let \( E = k(Sₓₓ Sₓᵧ Sₓ��) \). Then

\[
L'E L = L'X'[H − k⁻¹ S⁻¹ₓₓ Sₓᵧ Sₓ成型]X L = Q'JQ \tag{3.3.3.15}
\]

Using (3.3.3.14) and the fact that \( J \) is idempotent,

\[
[J V(Q)]² = [L'(Vₓₓ − σ⁻²ᵧᵧ) L]² \tag{3.3.3.16}
\]

The noncentrality parameter of \( L'E L \) is \( \mu'J \mu \) where \( \mu \) is given by (3.3.3.13). This is seen to be zero. Since rank \( (J) = k-2 \) it follows that

\[
L'E L = L'(Vₓₓ − σ⁻²ᵧᵧ) L \cdot \chi²_{k-2} \tag{3.3.3.17}
\]

Since this is true for every \( L \) we have by Rao [1965, p. 452]

\[
k(Sₓₓ Sₓᵧ Sₓ成型) \cdot W(k-2, \frac{Vₓₓ − σ⁻²ᵧᵧ Vᵧᵧ成型}{\chi²_{k-2}}) \tag{3.3.3.18}
\]

where \( W(p, m, \Sigma) \) denotes the central Wishart distribution of dimension \( p, m \) degrees of freedom, and matrix parameter \( \Sigma \). Since this
distribution does not depend on \( Y \), the unconditional distribution of \( E \) is the central Wishart distribution.

Now \( L' S_{xy} \sim \kappa^{-1} L' X' H ~ Y \sim \kappa^{-1} Q'(H Y) \). Based on (3.3.3.13) and (3.3.3.14), \( L' S_{xy} \) and \( L' E L \) are conditionally independent since

\[
y' H \{ L' (V - \sigma^{-2} V^{-1} V' )L \} \sim (H - k^{-1} S^{-1} yy) H YY' H \sim \zeta.
\]

Since the conditional independence holds for any \( L, E \) and \( S_{xy} \) are conditionally independent by Rao [1965, p. 453]. Finally, from (3.3.3.9) and (3.3.3.10), \( k^{1/2} S_{xy} \) is conditionally \( p \)-variate normal with mean and variance given respectively by

\[
k^{1/2} \sigma^{-2} S_{yy}^{-1} \text{ and } S_{yy}(V - \sigma^{-2} V^{-1} V').
\]

We now invoke the theorem on the distribution of Hotelling's \( T^2 \) statistic (see Rao [1965, p. 458]). In Rao's notation we let

\[
c = S_{yy}^{-1}, \quad d = S_{xy}, \quad S = S_{xx} - S_{xy} S_{yy} S_{xy}, \quad \xi = V_{xx} - \sigma^{-2} V_{xy} V', \quad \delta = k^{1/2} \sigma^{-2} S_{yy}^{-1} V_{xy}.
\]

Then

\[
(k-1-p) c \sim \zeta \sim d/p = T^*
\]

has conditional d.f. \( F_{\lambda}(t; p, k-1-p) \) with non-centrality parameter.
\[ \lambda = c \delta \Sigma^{-1} \delta = S^{-1} k \sigma^2 \Sigma^{-1} y' \Sigma^{-1} y = k S \sigma^2 y' \Sigma^{-1} x y = k S \sigma^2 y' \Sigma^{-1} x y. \]

Using (3.3.1.5),

\[ \lambda = U^* \rho^2 / (1 - \rho^2) \quad (3.3.3.20) \]

with \( \rho^2 \) and \( U^* \) defined respectively in (3.3.1.1) and (3.3.3.2).

Equation (3.3.3.5) follows from the fact that this conditional d.f. depends on \( Y \) only through \( U^* \). Finally by (3.3.3.5),

\[ P(T^* \leq t) = E\{P(T^* \leq t|U^*)\} = F^*_\rho(t; p, k-1-p). \]

Under \( H_0: \Sigma_{xy} = 0 \), this is equal to \( F_0(t; p, k-1-p) \) for all \( U^* \) and in that case \( F^*_\rho = F_0 \). Q.E.D.

From Theorem 3.3.3.1, \( T^* \) has the same conditional distribution as the statistic from an uncensored sample and under \( H_0: \Sigma_{xy} = 0 \), it has the classical F distribution with DF \( (p, k-1-p) \). Note that other authors, e.g. Anderson [1958, pp. 89-96], obtain the distribution of the sample multiple correlation coefficient by conditioning on \( X_1 \). In our case this would not lead to a normal conditional distribution of \( Y_1 \) so instead we have had to condition on \( Y_1 \).

The complicated unconditional non-null distribution of the complete sample multiple correlation coefficient can be found in Anderson [1958, pp. 93-5]. As in section 2.5.2, exact power calculations for this test are difficult to obtain. Here we may also take a one-moment approximation to the distribution of \( T^* \) using (3.3.3.20) by taking
\[ \lambda = \frac{\rho^2}{1-\rho^2} E(U^*). \]  

(3.3.3.21)

\( E(U^*) \) can be calculated as in section 2.5.1. The estimated power of the test of size \( \alpha \) is

\[ 1 - F_\lambda (F^*_\alpha; p, k-l-p) \]  

(3.3.3.22)

where \( F_\alpha (F^*_\alpha; p, k-l-p) = 1-\alpha \). This provides a much simpler method for higher moment power approximations in the complete sample case (\( n-k \)) than does the result obtained by conditioning on \( X_1 \) (see Anderson [1958, p. 95]) since the non-centrality parameter is a function only of the univariate \( Y_i \), \( 1 \leq i \leq n \).

3.4. Inference about Independence of \( X^{(2)} \) and \( Y \) given \( X^{(1)} \)

3.4.1. Estimation of Partial Correlation Coefficients

We make the following partitions:

\[ X_n[1] = (X^{(1)}_{n[1]} l \times p_1 \quad X^{(2)}_{n[1]} l \times p_2) \quad p_1 + p_2 = p \]

\[ V_{xx} = \begin{bmatrix} V_{11} P_1 \times P_1 & V_{12} P_1 \times P_2 \\ V_{12} P_1 \times P_2 & V_{22} P_2 \times P_2 \end{bmatrix} \quad V_{xy} = \begin{bmatrix} V_{1y} P_1 \times P_1 \\ V_{2y} P_2 \times P_2 \end{bmatrix} \]  

(3.4.1.1)

and similarly for \( V^*_x \), \( V^*_y \), \( S_{xx} \), and \( S_{xy} \)
Let $d = \frac{s_{yy}^2}{s_{yy}^2} (\sigma_y^2 - s_{yy})$.  \hfill (3.4.1.2)

Then from (3.2.32) and (3.2.30),

$$V_{xx}^* = \begin{bmatrix} S_{11} + d S_{1y} S_{1y}' & S_{12} + d S_{1y} S_{2y}' \\ S_{21} + d S_{2y} S_{1y}' & S_{22} + d S_{2y} S_{2y}' \end{bmatrix} \hfill (3.4.1.3)$$

$$V_{xy}^* = \sigma_y^2 S_{yy}^{-1} \begin{bmatrix} S_{1y} \\ S_{2y} \end{bmatrix} \hfill (3.4.1.4)$$

The conditional covariance matrix of $(X_{12}^{(2)}, Y_{1})$ given $X_{1}^{(1)}$ is given by

$$W = \begin{bmatrix} V_{22} - V_{12}^{-1} V_{12}^{-1} & V_{2y} - V_{1y}^{-1} V_{1y}^{-1} \\ V_{2y} - V_{1y}^{-1} V_{1y}^{-1} & \sigma_y^2 - V_{1y}^{-1} V_{1y}^{-1} \end{bmatrix} = \begin{bmatrix} W_{22} & W_{2y} \\ W_{2y} & W_{yy} \end{bmatrix}. \hfill (3.4.1.5)$$

Using (3.4.1.3) and (3.3.1.4),

$$V_{11}^{-1} = S_{11}^{-1} - d(1 + d S_{1y} S_{1y}^{-1} S_{1y}^{-1})^{-1} S_{1y}^{-1} S_{1y} S_{1y}^{-1} \hfill (3.4.1.6)$$

which along with (3.4.1.3) and (3.4.1.4) provides the modified MLE of $W$.  

Thus the modified MLE of the partial correlation vector of \( \mathbf{x}^{(2)} \) and \( \mathbf{y} \) given \( \mathbf{x}^{(1)} \) is

\[
 \mathbf{\hat{W}}^* = \sigma_y^2 \frac{(1 - S^{-1}_{yy} S^{-1}_{yl} S^{-1}_{yl})}{(1 + d S^{-1}_{ly} S^{-1}_{yl})} \quad (3.4.1.7)
\]

\[
 \mathbf{\hat{W}}_{22} = S_{22} - S^{-1}_{ly} S^{-1}_{yl} + d(1 + d S^{-1}_{ly} S^{-1}_{yl})^{-1}.
\]

\[
 (S_{2y} - S^{-1}_{ly} S^{-1}_{yl})(S_{2y} - S^{-1}_{ly} S^{-1}_{yl})'.
\]

\[
 \mathbf{\hat{W}}_{2y} = \sigma_y S^{-1}_{yy} (S_{2y} - S^{-1}_{ly} S^{-1}_{yl})/(1 + d S^{-1}_{ly} S^{-1}_{yl}).
\]

The multiple (or "multiple partial") correlation between \( \mathbf{x}^{(2)} \) and \( \mathbf{y} \) given \( \mathbf{x}^{(1)} \) is defined in an analogous manner to the multiple correlation between \( \mathbf{x} \) and \( \mathbf{y} \) (3.3.1.1) except that the covariance matrix is replaced by \( \mathbf{\hat{W}} \). This reduces to

\[
 \rho_{y2|1}^2 = (\rho^2 - \rho_{y1}^2)/(1 - \rho_{y1}^2)
\]

where \( \rho \) is the multiple correlation coefficient between \( \mathbf{x} \) and \( \mathbf{y} \) defined by (3.3.1.1) and \( \rho_{y1} \) is the multiple correlation coefficient.
between $X^{(1)}$ and $Y$. Using (3.3.1.7) the modified MLE of $\rho_{y1}^2$ is

$$\rho_{y1}^{2*} = R_{y1}^{2*} \{c^2 + (1-c)R_{y1}^{2*}\}^{-1}$$  (3.4.1.12)

where $R_{y1}^*$ is the ordinary sample multiple correlation coefficient between $X^{(1)}$ and $Y$,

$$R_{y1}^{2*} = \frac{S_{yy}^{-1} S_{yl}^{-1} S_{yl}^{-1} S_{ly}^{-1}}{c, S_{yy} - c R_{y1} - c R_{y1}^{-1} R_{y1} R_{y1}^{-1}}.$$  (3.4.1.13)

Combining (3.4.1.12) and (3.3.1.7),

$$\rho_{y2|1}^{2*} = \frac{(R_{y1}^{2*} - R_{y1}^{2*})}{((c-R_{y1}^{2*})[c+(c-1)R_{y1}^{2*}])}.$$  (3.4.1.14)

where as before $c = S_{yy}^{-1} S_{y}^{-2}$ is a "correction factor for censoring".

3.4.2. Distribution of Test Statistic under Censoring.

A statistic for testing the independence of $X^{(2)}$ and $Y$ given $X^{(1)}$, $H_0: V_{2y} - V_{12} V_{11}^{-1} V_{1y} = 0$, or equivalently $H_0: \rho_{y2|1}^2 = 0$ is

$$S^* = \{R_{y2|1}^{*2}/p_2\}/\{(1-R_{y2|1}^{*2})/(k-1-p)\}.$$  (3.4.2.1)

with the sample squared multiple partial correlation coefficient defined by
and its components defined in (3.3.1.8) and (3.4.1.13). Let \( \rho_{y1}^2 \) and \( \rho_{y2|1}^2 \) denote the population correlations defined in section 3.4.1 and

\[
R_{y2|1}^* = \frac{(R_{y2}^* - R_{y1}^*)}{(1-R_{y1}^*)} \quad (3.4.2.2)
\]

Then with \( F_\Delta(t;a,b) \) denoting the non-central F d.f. as before and \( F_\rho(t;a,b) \) defined in (2.5.2.2), we have the following

**Theorem 3.4.2.1:**

For every \( t > 0 \) and given \( P^* \)

\[
P\{S^* \leq t | P^* \} = F_\lambda(t; p_2, k-1-p), \quad t \geq 0. \quad (3.4.2.5)
\]

Hence \( P\{S^* \leq t \} = F_{\rho_{y2|1}^*}(t; p_2, k-1-p), \quad t \geq 0. \quad \text{Under } H_0: \rho_{y2|1} = 0, \)

\( F_{\rho_{y2|1}}(t; p_2, k-1-p) = F_0(t; p_2, k-1-p) \) is the central F d.f. with D.F. \( (p_2, k-1-p) \).

**Proof:** Equation (3.3.1.5) and the matrix partitions given in (3.4.1.5)
lead us to the relation

\[ R_{y_2}^{*2} = \frac{1}{1 - R_{y_2}^{*2}} = S_{yy}^{-1} (1 - R_{y_1}^{*2})^{-1} \sim \bar{A} \sim w, \]  

\hspace{1cm} (3.4.2.6)

where

\[ \bar{w} = S_{2y} - S_{12} S_{11}^{-1} S_{1y}, \text{ and} \]

\[ \bar{A} = S_{22} - S_{12} S_{11}^{-1} S_{12}. \]

\[ S_{yy}^{-1} (1 - R_{y_1}^{*2})^{-1} (S_{2y} - S_{12} S_{11} S_{1y}) (S_{2y} - S_{12} S_{11} S_{1y})'. \]

\hspace{1cm} (3.4.2.7)

Making partitions as in section 3.4.1 and letting

\[ \mu_x = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \]

\[ \bar{X}^{(1)} = \left( \bar{x}_{n[1]}^{(1)}, \ldots, \bar{x}_{n[k]}^{(1)} \right)' , \bar{X}^{(2)} = \left( \bar{x}_{n[1]}^{(2)}, \ldots, \bar{x}_{n[k]}^{(2)} \right)' . \]

\[ \bar{X}^{(1)} = k^{-1} \sum_{i=1}^{k} \bar{x}_{n[i]}^{(1)} , \bar{X}^{(2)} = k^{-1} \sum_{i=1}^{k} \bar{x}_{n[i]}^{(2)} . \]

\[ \bar{H} = \bar{I}_k - k^{-1} \bar{I} \bar{I}'. \]

then we have
Now conditional on $Y_n,i, X_i' \sim_{n[i]}$ is p-variate normal with

$$E(X_{n[i]}' | Y_n,i) = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \sigma_y^{-2}(Y_{n,i} - \mu_y) \begin{pmatrix} V_{1y} \\ V_{2y} \end{pmatrix}$$

(3.4.2.10)

$$V(X_{n[i]}' | Y_n,i) = \begin{pmatrix} V_{11} - \sigma_y^{-2}V_{1y}V_{1y}' & V_{12} - \sigma_y^{-2}V_{1y}V_{2y}' \\ V_{12} - \sigma_y^{-2}V_{2y}V_{1y}' & V_{22} - \sigma_y^{-2}V_{2y}V_{2y}' \end{pmatrix}.$$

(3.4.2.11)

So conditional on $Y_n,i$ and $X_i' \sim_{n[i]}$, $X_{n[i]}' \sim_{n[i]}$ is p-variate normal with

$$E(X_{n[i]}(2) | Y_n,i, X_{n[i]}(1)) = \frac{1}{\sigma_y^2} \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \frac{1}{\sigma_y^2} \begin{pmatrix} V_{1y} \\ V_{2y} \end{pmatrix}$$

and

$$V(X_{n[i]}(2) | Y_n,i, X_{n[i]}(1)) = \begin{pmatrix} V_{11} - \sigma_y^{-2}V_{1y}V_{1y}' & 0 \\ 0 & V_{22} - \sigma_y^{-2}V_{2y}V_{2y}' \end{pmatrix}.$$
which, using (3.3.1.4) is

$$
\Sigma = (V_{22} - \Sigma_{12}' V_{11}^{-1} V_{12}') - \sigma_y^{-2} (1 - \rho_{1y}^2)^{-1} (V_{2y} - \Sigma_{12}' V_{11}^{-1} V_{12}') (V_{2y} - \Sigma_{12}' V_{11}^{-1} V_{12}')'
$$

(3.4.2.12)

and \( X_{n[i]}^{(2)}, 1 \leq i \leq k, \) are conditionally independent. We have

$$
\Sigma = \sum_{i=1}^{k} X_{n[i]}^{(2)} - \sum_{i=1}^{k} (X_{n[i]}^{(1)} - \bar{X}_{n[i]})(X_{n[i]}^{(1)} - \bar{X}_{n[i]})' = \Sigma_{yy} (1 - \rho_{y1}^2)^{-1} (V_{2y} - \Sigma_{12}' V_{11}^{-1} V_{12}')
$$

(3.4.2.13)

and because of the conditional independence of the \( X_{n[i]}^{(2)}, \)

$$
E(\Sigma | X_{n[i]}^{(1)} Y) = \Sigma_{yy} (1 - \rho_{y1}^2)^{-1} (V_{2y} - \Sigma_{12}' V_{11}^{-1} V_{12}')
$$

(3.4.2.14)

Let \( \Sigma \) be any non-zero \( p_2 \)-vector and \( Q_{1} = X_{n[i]}^{(2)} L, \) \( 1 \leq i \leq k. \)

Conditional on \( Y_{n,i}, X_{n[i]}^{(1)} \), \( Q_{1} \) is univariate normal and

$$
E(Q_{1} | Y_{n,i}, X_{n[i]}^{(1)}) = L' \mu_{2} + \sigma_{y}^{-2} (Y_{n,i} - \mu_{y}) L' V_{2y} +
$$

$$
[X_{n[i]}^{(1)} - (\mu_{1} + \sigma_{y}^{-2} (Y_{n,i} - \mu_{y}) V_{1y})' (V_{1y}^{-1} - \sigma_{y}^{-2} V_{1y} V_{1y}^{-1})^{-1} \cdot
$$

$$
(V_{12} - \sigma_{y}^{-2} V_{1y} V_{1y}^{-1}) L
$$

(3.4.2.15)
\begin{equation}
V(Q_1|Y_{n,1}, X^{(1)}_{n[1]}) = L' \Sigma L
\end{equation}

(3.4.2.15)

and \( Q_1, \ldots, Q_k \) are conditionally independent. Let
\( Q = (Q_1, \ldots, Q_k)' \sim \gamma^{(2)} \). It follows that \( Q \) is \( k \)-variate normal

\[ E(Q|X^{(1)}, Y) = (L' \mu_2) + \Sigma_y (L' \Sigma_{2y})(Y - \mu_Y) + \]

\[ + [X^{(1)} - (\mu_1 + \Sigma_y (Y_{n,1} - \mu_Y) \Sigma_{1y})'] (\Sigma_{11} - \Sigma_y \Sigma_{1y} \Sigma_{1y})^{-1} (\Sigma_{12} - \Sigma_y \Sigma_{1y} \Sigma_{2y}) \]

\[ V(Q|X^{(1)}, Y) = L' \Sigma L \quad L_{k}. \quad \text{(3.4.2.16)} \]

Let \( E = k \). Then

\[ L'E L = k^{-1} L' X^{(2)} [k H - HX^{(1)} S^{-1} X^{(1)}]' H^{-1} \]

\[ = H S_{11}^{-1} Y' H^{-1} + H X^{(1)} S_{11}^{-1} S_{1y} S_{1y}^{-1} X^{(1)}' H \]

\[ = (X^{(2)} L)' J (X^{(2)} L) = Q' J Q. \quad \text{(3.4.2.17)} \]

It can be verified that \( J \) is idempotent. Therefore,
\[ \left[ J \mathbf{v}(Q|X^{(1)}, Y) \right]^2 = (L' \Sigma L)^2 J. \quad (3.4.2.18) \]

The non-centrality parameter, \( \mu' \mathbf{J} \mu \) with \( \mu = E(Q|X^{(1)}, Y) \) can be shown to be zero. Note that rank (\( J \)) = \( k-2-p_1 \). It follows that conditionally

\[ L' \Sigma L \sim L' \Sigma L \cdot \chi^2_{k-2-p_1}. \quad (3.4.2.19) \]

Since this is true for every \( L \),

\[ E = k \mathbf{A} \sim W_{p_2}(k-2-p_1, \Sigma) \quad (3.4.2.20) \]

where as before \( W_{p_2} \) (*) denotes the central Wishart distribution. We note in passing that since the parameters of this conditional distribution do not depend on \( X^{(1)} \) or \( Y \), \( E \) also has an unconditional Wishart distribution. This result is also a direct consequence of (3.3.3.18) and properties of the Wishart distribution. However, for our purposes we needed to verify that the conditional distribution (given \( Y \) and \( X^{(1)} \)) is also a Wishart distribution.

To demonstrate that \( \mathbf{w} \) and \( E \) are conditionally independent we see that

\[ L' \mathbf{w} = k^{-1}L'X^{(2)}(\mathbf{H}^{-1} \mathbf{X}^{(1)}S^{-1}_{11} \mathbf{S}_{1y}) = k^{-1}Q'(\mathbf{H}^{-1} \mathbf{X}^{(1)}S^{-1}_{11} \mathbf{S}_{1y}) \]

where \( Q'(\mathbf{H}^{-1} \mathbf{X}^{(1)}S^{-1}_{11} \mathbf{S}_{1y}) \) is a Wishart distribution.
and

\[(Y'X - S' \Sigma^{-1} X_l)' (L' \Sigma L) I_k J = 0. \quad (3.4.2.21)\]

Then since \(L'w\) and \(L'E\) are conditionally independent for every \(L, w\) and \(E\) are conditionally independent. From (3.4.2.16),
\(d = k^{1/2} \tilde{w}\) is conditionally \(p_2\)-variate normal with

\[
\begin{align*}
\tilde{\delta} &= E(d|\tilde{X}^{(1)}_w, Y) = k^{1/2} \sigma^2_y (1-R_{yl})(1-\rho_{yl})^{-1} (v_{2y} - v' v_{1y}) (3.4.2.22) \\
V(d|\tilde{X}^{(1)}_w, Y) &= S_{yy} (1-R_{yl})^{-1}.
\end{align*}
\]

We again use Rao's Theorem concerning Hotelling's \(T^2\) distribution [1965, p. 458] with \(S = E_s, c = S^{-1} y (1-R_{yl})^{-1},\) and \(\tilde{\delta}\) and \(d\) given as above. Thus

\[
S^{-1} (1-R_{yl})^{-1} w \sim \tilde{A}^{-1} w \sim (k-1-p)/p_2 = \tilde{s}^*
\]

has conditional d.f. \(F_\lambda(t; p_2, k-1-p)\) with \(\lambda = c \tilde{\delta} \Sigma^{-1} \tilde{\delta}\) given by (3.4.2.3) using (3.3.1.5). Equation (3.4.2.5) follows from the fact that the conditional d.f. depends on \(Y\) and \(X^{(1)}_w\) only through \(\rho^*\). This implies

\[
P(S^* \leq t) = E[P(S^* \leq t|\rho^*]) = F^*_{\rho y_2|1}(t; p_2, k-1-p).
\]

Under \(H_0: \rho_{y_2} = 0, F^*_{\rho y_2|1}\) is equal to \(F_0(t; p_2, k-1-p)\) for all \(\rho^*\) and in that case \(F^*_{\rho y_2|1} = F_0\). Q.E.D.
From Theorem 3.4.2.1, \( S^* \) has the classical null distribution. Power approximation to the \( S^* \) test will be more difficult than in the preceding sections. For a complete sample some simplification may be made, for in that case \( R_{y1}^{*2} \) is a consistent estimator of \( \rho_{y1}^2 \) in (3.4.2.3).

3.5. Inference about Independence of \( \tilde{X}^{(1)} \) and \( \tilde{(X^{(2)}, Y)} \)

3.5.1. Estimation of Canonical Correlations

Let the random variable \( \tilde{X} \) be partitioned as in section 3.4.1. We wish to test \( H_0: \tilde{X}^{(1)} \) is independent of \( \tilde{(X^{(2)}, Y)} \) where \( Y \) may be subject to censoring. Let the covariance matrix of \( \tilde{(X^{(1)'}, X^{(2)'}, Y)} \) be partitioned as

\[
\begin{pmatrix}
\begin{pmatrix}
\sigma_{11}^2 & \sigma_{12}^2 & \sigma_{1y}^2 \\
\sigma_{12}^2 & \sigma_{22}^2 & \sigma_{2y}^2 \\
\sigma_{1y}^2 & \sigma_{2y}^2 & \sigma_y^2
\end{pmatrix}
\end{pmatrix}
\] (3.5.1)

The hypothesis of interest is then \( H_0: (\sigma_{12}^2, \sigma_{1y}^2) = 0 \). Let

\( \bar{T} = (\tilde{X}^{(2)}, Y) \) and

\[
V(T) = \begin{pmatrix}
\sigma_{22} & \sigma_{2y} \\
\sigma_{2y} & \sigma_y^2
\end{pmatrix} = \Sigma_T.
\] (3.5.1.2)
Also let $\Sigma_{1T} = (V_{12} V_{1y})$ and $V_{11}, \Sigma_T, \Sigma_{11}$ denote modified MLEs of the corresponding parameters as given in section 3.2. Population canonical correlations, the largest of which is the maximum correlation between any linear combination of $T$ with any linear combination of $X^{(1)}$, can be estimated by finding eigenvalues of

$$V_{11}^{-1} \Sigma_T^{-1} \Sigma_{1T}^{*-1} \Sigma_1^{*}. \quad (3.5.1.3)$$

3.5.2. Null Distribution of a Test Statistic

From (3.4.2.10) conditional on $Y_{n,i}, X_{n[i]}^{(1)}$ is $\mathcal{P}_1$-variate normal with

$$E(X_{n[i]}^{(1)} | Y_{n,i}) = \mu_1 + \sigma_y^{-2} (Y_{n,i} - \mu_y) V_{1y}$$

$$V(X_{n[i]}^{(1)} | Y_{n,i}) = \Sigma_{11}^{-1} - \sigma_y^{-2} V_{1y} V_{1y}' . \quad (3.5.2.1)$$

Under $H_0$: $\Sigma_{1T} = 0$, $X_{n[i]}^{(1)}', 1 \leq i \leq k$ is (unconditionally) normal with parameters $\mu_1$ and $\Sigma_{11}$ and $X_{n[i]}^{(1)}$ is independent of $X_{n[i]}^{(2)}$ as well as of $Y_{n,i}$. Thus by Anderson [1958 , p. 324] , eigenvalues of

$$S_{11}^{-1} S_{1T} S_{T}^{-1} S_{1T}^{*} , S_{1T} = (S_{12} S_{1y}) \quad (3.5.2.2)$$

have the classical null distribution and the test of $H_0$: $\Sigma_{1T} = 0$
can be carried out without difficulty. The matrices involved are ordinary sample sums of products matrices. The non-null distribution of the eigenvalues will be exceedingly complex.
CHAPTER IV
STATISTICAL INFERENCE FOR OTHER MULTIVARIATE DISTRIBUTIONS

4.1. Introduction.
Suppose that \((X_i,Y_i)\) has a bivariate d.f. such that \(Y_i\) has marginal density \(g_\lambda(y)\) with at least the first two moments existing where \(\lambda\) is a \(1 \times \ell\) vector. Suppose also that, conditional on \(Y_i\), \(X_i\) has a normal distribution with linear regression and homoscedasticity. Define

\[
E(Y_i) = \mu_y, \quad V(Y_i) = \sigma_y^2, \quad E(X_i) = \mu_x, \quad V(X_i) = \sigma_x^2
\]

\[
E(X_i|Y_i) = \mu_x + \beta(Y_i - \mu_y), \quad V(X_i|Y_i) = \delta^2.
\]

It follows that

\[
\text{Cov}(X_i,Y_i) = \beta \sigma_y^2, \quad \text{Corr}(X_i,Y_i) = \rho = \frac{\beta \sigma_y}{\sigma_x},
\]

\[
\sigma_x^2 = \delta^2 + \beta^2 \sigma_y^2.
\]
The information matrix of $\hat{\theta} = (\mu, \delta, \beta, \lambda)$ and the large sample covariance matrix of $\hat{\theta}$ will be calculated. Joint MLE or modified MLE of $\hat{\theta}$ will be derived with the MLE of $\delta, \beta$ being independent of $g_\lambda$. These estimators will be used to estimate $\rho$ and $\sigma_x$. The likelihood ratio test of $H_0: \beta = 0$ will be displayed. The test statistic does not depend on $g_\lambda$. The properties of the test will be studied for $Y_1$ having an exponential or extreme value distribution.

4.2. Inference when Order Statistics are from any Parent Distribution but Concomitant Variables are Conditionally Normally Distributed

4.2.1. Cramér-Rao Lower Bounds

The joint density of $Y_{n,1}, \ldots, Y_{n,k}$ is

$$n[k] \prod_{i=1}^{k} g_\lambda(y_i) [1 - G_\lambda(y_k)]^{n-k},$$

(4.2.1.1)

$$y_1 \leq y_2 \leq \cdots \leq y_k$$

where $G_\lambda$ is the d.f. corresponding to the density $g_\lambda$. The conditional density of $X_{n[1]}, \ldots, X_{n[k]}$ given $Y_{n,1}, \ldots, Y_{n,k}$ is, by Lemma 1 of Bhattacharya [1974],

$$\prod_{i=1}^{k} \{f((x_i - \mu_x - \beta(Y_{n,i} - \mu_y))/\delta)/\delta\}$$

(4.2.1.2)
and $X_n[1], \ldots, X_n[k]$ are conditionally independent. As before, $\phi_1$ denotes the univariate normal density. Denote the log likelihood function of $Y_{n1}, \ldots, Y_{nk}$ by

$$L_{n,k} = \log n[k] + \sum_{i=1}^{k} \log \phi(Y_{n,i}) + (n-k) \log [1 - \phi(Y_{n,k})].$$

(4.2.1.3)

The total log likelihood function of $(X_n[i], Y_{n,i}, 1 \leq i \leq k)$ is then, aside from constant terms,

$$L_{n,k} = -k \log \delta - (2\delta^2)^{-1} \sum_{i=1}^{k} [X_n[i] - \mu_x - \beta(Y_{n,i} - \mu_y)]^2 + L_{n,k}^T.$$

(4.2.1.4)

Let $\alpha$ denote the $1 \times (k-1)$ parameter vector in $\lambda$ aside from $\mu_y$. Then letting $T = X - \mu_x, Z = (Y - \mu_y)$,

$$\frac{\partial}{\partial \mu_y} L_{n,k} = \frac{\partial}{\partial \mu_y} L_n^T / \delta^2$$

(4.2.1.5)

$$\frac{\partial}{\partial \delta} L_{n,k} = -k \delta^{-1} + \delta^{-3} (T - \beta Z)'(T - \beta Z)$$

(4.2.1.6)

$$\frac{\partial}{\partial \beta} L_{n,k} = (T - \beta Z)' T / \delta^2$$

(4.2.1.7)

$$\frac{\partial}{\partial \mu_y} L_{n,k} = -\beta \frac{\partial}{\partial \mu_y} L_n^T / \delta^2 + \frac{\partial}{\partial \mu_y} L_{n,k}^T$$

(4.2.1.8)

$$\frac{\partial}{\partial \alpha} L_{n,k} = (\partial / \partial \alpha) L_n^T, L_{n,k}^T$$

(4.2.1.9)
Now let the information matrix derived from $L_{n,k}^Y$ be

$$I^Y = \begin{bmatrix} I_{\mu_y} & I_{\mu_y \alpha} \\ I_{\mu_y \alpha} & I_{\alpha} \end{bmatrix}$$  \hspace{1cm} (4.2.1.10)$$

and let $u_{n,1} = E(Y_{n,1} - \mu_y)/\sigma_y$, $w_1 = E(Y_{n,1} - \mu_y)^2/\sigma_y^2$  \hspace{1cm} (4.2.1.11)
and $u = (u_{n,1}, \ldots, u_{n,k})'$, $w = (w_1, \ldots, w_k)'$. Using the fact that $E(T) = \beta E(Z)$, the information matrix $I$ for $\theta = (\mu_x, \delta, \beta, \mu_y, \alpha)$ is:

$$I_{\mu_x} = \begin{bmatrix} \frac{k}{\delta^2} & 0 & \sigma_y & 1' u/\delta^2 & -k\beta/\delta^2 \\ 0 & 2k/\delta^2 & 0 & 0 & 0 \\ \sigma_y & 1' w/\delta^2 & -\beta \sigma_y & 1' u/\delta^2 & 0 \\ 1' u & +k\beta^2/\delta^2 & 1' u & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha \end{bmatrix}$$  \hspace{1cm} (4.2.1.12)

The inverse of the upper left $3 \times 3$ submatrix of $I$ is

$$\begin{bmatrix} \delta^2 1' w/kd & 0 & -\delta^2 1' u/kd \sigma_y \\ 0 & \delta^2/2k & 0 \\ -\delta^2 1' u/kd \sigma_y & 0 & \delta^2/\delta^2 \sigma_y^2 \end{bmatrix}, \hspace{0.5cm} d = (1' w - (1' u)^2/k).$$  \hspace{1cm} (4.2.1.13)
This yields $I^{-1} =$

$$
\begin{bmatrix}
\mu_x & \delta & \beta & \mu_y & \gamma \\
\delta^2 u' / k_d + \beta^2 f & 0 & -\delta^2 u' / k_d \sigma_y & \beta f & -\beta f I^{-1}_{\mu_y \alpha} \\
\delta^2 / 2k & 0 & 0 & 0 & 0 \\
\delta^2 / d \sigma^2_y & 0 & 0 & 0 & 0 \\
\end{bmatrix}
$$

(4.2.1.14)

where $f = (I_{\mu_y} - \hat{\alpha} I_{\mu_y \alpha} I^{-1}_{\alpha} - I_{\mu_y} I^{-1}_{\alpha})^{-1}$, $I^{-1}_{\alpha}$ is $l \times l$, and $I_{\alpha}$ is $(l-1) \times (l-1)$.

When $g_\lambda \lambda$ is a one parameter distribution such as the exponential, $\alpha$ is not defined. For this case we obtain the large sample covariance matrix

$$
\begin{bmatrix}
\mu_x & \delta & \beta & \mu_y \\
\delta^2 u' / k_d + \beta^2 / I_{\mu_y} & 0 & -\delta^2 u' / k_d \sigma_y & \beta / I_{\mu_y} \\
\delta^2 / 2k & 0 & 0 & 0 \\
\delta^2 / d \sigma^2_y & 0 & 0 & 0 \\
\end{bmatrix}
$$

(4.2.1.15)
and \( \sigma_y \) is a function of \( \mu_y \). To place this within the framework of section 2.3, Cramér-Rao lower bounds on the parameters \( \rho \) and \( \sigma_x \) which are derived from \( \beta, \delta, \) and \( \sigma_y \), can be calculated from (4.2.1.14) by standard large-sample methods.

### 4.2.2. Estimation.

As in previous chapters, we will place \( \ast \) by a parameter to denote an estimator which is derived from maximum likelihood estimators of \( \lambda \) or by another efficient method such as linearization of \( L^y_{n,k} \) (Chan [1967]). A \( \wedge \) over a parameter will denote a true MLE.

To estimate \( \mu_x, \delta, \beta, \) we set partial derivatives of \( L^y_{n,k} \) in (4.2.1.5) through (4.2.1.7) to zero. Define \( \bar{x}, \bar{y}, S^2_x, S^2_y, \) and \( S_{xy} \) as in (2.4.1). Dropping \( \ast \) and \( \wedge \) for the moment we find

\[
\bar{x} - \mu_x = \beta(\bar{y} - \mu_y) \quad \text{from (4.2.1.5)},
\]

\[
\delta^2 = S^2_x + \beta^2 S^2_y - 2\delta S_{xy} \quad \text{from (4.2.1.6)},
\]

and

\[
\beta = S_{xy}/S^2_y, \quad \text{from (4.2.1.7)}.
\]

Note that \( \beta \) is the usual regression estimator. Since it is MLE not depending on \( L^y_{n,k} \) we write

\[
\hat{\beta} = S_{xy}/S^2_y. \quad (4.2.2.4)
\]
Substituting \( \hat{\beta} \) into (4.2.2.2) we find

\[
\hat{\sigma}^2 = S_x^2 - \frac{S_{xy}^2}{S_y^2},
\]  

(4.2.2.5)

the usual residual squared error estimator from regression. Using (4.2.2.1), (4.2.1.8) becomes

\[
(\partial/\partial \mu_y) I_{n,k} = (\partial/\partial \mu_y) I_{n,k}^Y
\]  

(4.2.2.6)

which along with (4.2.1.9) implies that \( \lambda^* = (\mu^*_y, \sigma^*_y) \) are MLE or other efficient estimators based on \( I_{n,k}^Y \) alone. From (4.2.2.1) we complete the estimation by

\[
\mu^*_x = \bar{x} + (S_{xy}/S_y^2)(\mu^*_y - \bar{y})
\]  

(4.2.2.7)

which, of course, is the same as (2.4.10). For the derived parameters we obtain, using (4.1.2),

\[
\sigma^*_x = S_x^2 + (S_{xy}^2/S_y^4)(\sigma^*_y - S_y^2)
\]  

(4.2.2.8)

\[
\rho^* = \frac{\sigma^*_y S_{xy}}{\sigma^*_x S_y}
\]

as in (2.4.9) and (2.4.7). If \( \mu^*_y, \sigma^*_y \) are MLE it follows that \( \sigma^*_x \) and \( \rho^* \) are also. This statement can be proved by writing
The transformation is one-to-one. Formulas for the expected values of estimators in section 2.4 are also valid here. In particular (2.4.11) and (2.4.13) imply

\[ E(\hat{\beta}) = \beta, \quad V(\hat{\beta} | \gamma) = \delta^2 / k \sigma_y^2 \]  \hspace{1cm} (4.2.2.9)

so

\[ V(\hat{\beta}) = (\delta^2 / k) E(S_y^{-2}). \]

This exact variance should be compared with the large sample variance from (4.2.1.14),

\[ \delta^2 / (\bar{\gamma}^2 - (\bar{\gamma})^2 / k) \sigma_y^2 \]  \hspace{1cm} (4.2.2.10)

which for a complete sample equals

\[ \delta^2 / k \sigma_y^2. \]  \hspace{1cm} (4.2.2.11)

It can also be shown that

\[ E(\hat{\beta}^2) = (k-2)k^{-1} \delta^2. \]  \hspace{1cm} (4.2.2.12)
4.2.3. Likelihood Ratio Test of Independence

By (4.2.2.6), estimation of $\lambda$ is unaffected by restrictions on $\beta$. Hence the log likelihood ratio depends only on the conditional normality of $X$. Thus the argument in section 2.5.1 applies and the likelihood ratio statistic is a function of the product moment correlation $r^*$ and we use the test statistic $T^*$ in (2.5.1.8). Theorem 2.5.2.1 used only the premise of conditional normality so conditional on $S^2_y, T^*$ has an $F$ d.f. with DF$(1, k-2)$ and non-centrality parameter $(\rho^2/(1-\rho^2))k\sigma^2_y/\sigma^2_y$. The unconditional distribution of $T^*$ then depends only on $g_{\lambda}(y)$ and under $H_0: \beta = 0, \rho = 0$ and $T^*$ has an unconditional $F$ d.f.

4.3. Example: Order Statistics from Extreme Value Parent Distribution

Let $Y_1, \ldots, Y_n$ denote random variables from the extreme value type I distribution with d.f.

$$G_{\lambda}(y) = \exp\{-e^{-(y-\xi)/\theta}\}, \quad -\infty < y < \infty, \quad \theta > 0. \quad (4.3.1)$$

This distribution is a useful model for life-testing. An interpretation of its parameters is given by the moments,

$$E(Y_1) = \xi + \gamma \theta \quad \text{(4.3.2)}$$

$$V(Y_1) = \theta^2 \pi^2/6$$
where $\gamma$ is Euler's constant (.57721...). The location parameter $\xi$ can be efficiently estimated by linear combinations of order statistics while the scale parameter $\theta$ cannot; see Johnson and Kotz [1970, p. 284] for a summary of efficiencies of various estimators. Harter and Moore [1968] showed that while the MLE of $\xi$ and $\theta$ have considerable bias, the MLE of $\xi$ based on a censored sample is still comparable to the best linear unbiased estimator in terms of mean squared error and the MLE of $\theta$ is superior to the best linear unbiased estimator.

Let

$$\Gamma_w(a) = \int_0^w u^{a-1} e^{-u} du$$

(4.3.3)

denote the incomplete gamma function and $\Gamma'(w) = (\partial \Gamma_w(t)/\partial t)|_{t=a},$ etc. Let

$$\gamma'(a) = \Gamma'(a) - \Gamma'_\log p(a), \quad \gamma''(a) = \Gamma''(a) - \Gamma''\log p(a),$$

(4.3.4)

where $p = k/n$ is the proportion of order statistics observed. $\gamma'(a), \gamma''(a),$ etc. are related to incomplete polygamma functions.

Specializing the results of Harter and Moore [1968] to the case of censoring on the right, we find as elements of the information matrix,
\[
\lim_{n \to \infty} \left( -\frac{\theta^2}{n} \right) \mathbb{E} \left( \frac{\partial^2 L_{n,k}^Y}{\partial \xi^2} \right) = p - p \log p = a
\]

\[
\lim_{n \to \infty} \left( -\frac{\theta^2}{n} \right) \mathbb{E} \left( \frac{\partial^2 L_{n,k}^Y}{\partial \xi \partial \theta} \right) =
- [\gamma'(2) + p \log p + (1-p)^{-1} p \log p] = b
\]

(4.3.5)

\[
\lim_{n \to \infty} \left( -\frac{\theta^2}{n} \right) \mathbb{E} \left( \frac{\partial^2 L_{n,k}^Y}{\partial \theta^2} \right) = \gamma''(2) + 2[\gamma'(2) - \gamma'(1)] - p
+ p \log p \{2 + q + (1-p)^{-1} q \log p\} = c
\]

where \( q = \log(-\log p) \) and \( L_{n,k}^Y \) is the log likelihood function based on the order statistics as defined in (4.2.1.3). So the "asymptotic information matrix" of \( \xi, \theta \) is

\[
\frac{n}{\theta^2} \begin{bmatrix}
  a & b \\
  b & c
\end{bmatrix}
\]

(4.3.6)

To draw inferences from the induced order statistics \( X \) we need to identify the transformed parameters

\[
\mu_Y = \mathbb{E}(Y_i) = \xi + \gamma \theta
\]

(4.3.7)

\[
\sigma_Y = [\mathbb{V}(Y_i)]^{1/2} = \theta \pi / \sqrt{\theta}
\]
since the parameters associated with $X$ are defined by conditional moments of $X$. The large sample covariance matrix of the MLE of $\mu_y$ and $\sigma_y (\hat{\xi} + \gamma \hat{\theta}$ and $\hat{\pi}/\sqrt{\delta}$ respectively) is

$$
\begin{pmatrix}
(\sigma^2_y/n) & (6/\pi^2)(\gamma^2 a + c - 2 \gamma b) & (\sqrt{\delta}/\pi)(\gamma a - b) \\
(\sqrt{\delta}/\pi)(\gamma a - b) & a
\end{pmatrix}
$$

(4.3.8)

with corresponding information matrix

$$
I^Y = (n/\sigma^2_y) \begin{pmatrix}
(\pi^2/6)a & (\pi/\sqrt{\delta})(b - \gamma a) \\
(\pi/\sqrt{\delta})(b - \gamma a) & \gamma^2 a + c - 2 \gamma b
\end{pmatrix}
$$

(4.3.9)

Using this, the asymptotic joint lower bounds for variances of estimators of $(\mu_x, \delta, \beta, \mu_y, \sigma_y)$ can be readily obtained using (4.2.1.14). For example we find that the lower bound for the variance of $\mu^*_x$ from (4.2.1.14) is

$$
\delta^2 \frac{1}{2} \frac{\omega/kd + (\sigma^2_y/n)(6/\pi^2)\beta^2(\gamma^2 a + c - 2 \gamma b)}{(ac - b^2 - 4 \gamma ab)}
$$

(4.3.10)

where $\omega$ and $d$ in (4.2.1.13) can be computed using the first two moments of extreme value order statistics (see Johnson and Kotz [1970, p. 279]).

To study the test of independence of $X_i$ and $Y_i$, or $H_0: \beta = 0$ we need to study the distribution of
\[ U^* = \sigma_y^{-2} \sum_{i=1}^{k} \left( Y_{n,i} - \bar{Y} \right)^2 = (6/\pi^2) \theta^{-2} \sum_{i=1}^{k} \left( Y_{n,i} - \bar{Y} \right)^2 \]  
\[ = (6/\pi^2) \sum_{i=1}^{k} (Z_{n,i} - \overline{Z})^2 \]  
(4.3.11)

where \( Z_{n,i}, 1 \leq i \leq k \) are order statistics from the standard extreme value distribution (\( \xi = 0, \theta = 1 \)). \( E(U^*) \) is most easily approximated by simulation. For this purpose, 2000 \( Z \) vectors were generated for each \( n, k \) combination. The average values of \( U^* \) are as follows:

| \( n \) | \( k \) | \( E(U^*) \)  
|---|---|---|
| 20 | 10 | 1.518  
| 100 | 20 | 1.250  
| 1000 | 50 | 1.419  
(4.3.12)

To estimate the unconditional power of the test, the conditional power was computed for each simulated value of \( U^* \) and each value of \( \rho \) by using the noncentral \( F \) distribution. These 2000 powers were then averaged to estimate the unconditional distribution of \( U^* \). For the one moment power approximations, the corresponding non-central \( F \) probabilities were evaluated using \( E(U^*) \) in (4.3.12). The results are in Table 4.1.
<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>ρ</th>
<th>Power</th>
<th>Power from One Moment Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>.1</td>
<td>.0514</td>
<td>.0514</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.0636</td>
<td>.0636</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.0967</td>
<td>.0965</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.5686</td>
<td>.6079</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.1</td>
<td>.0513</td>
<td>.0513</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.0628</td>
<td>.0628</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.0940</td>
<td>.0939</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.5677</td>
<td>.5889</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>.1</td>
<td>.0516</td>
<td>.0516</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.0656</td>
<td>.0656</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.1035</td>
<td>.1035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.6581</td>
<td>.6738</td>
</tr>
</tbody>
</table>

The power approximation is apparently satisfactory.

4.4. **Example: Order Statistics from Exponential Parent Distribution**

Let $Y_i$ have density function

$$g_\lambda(y) = \sigma_y^{-1} \exp(-y/\sigma_y)$$  \hspace{1cm} (4.4.1)
so that \( \mu_y \equiv \sigma_y \). This implies

\[
L_{n,k} = \log n^{[k]} - k \log \sigma_y - \sigma_y^{-1} \sum_{i=1}^{k} Y_{n,i} - (n-k)\sigma_y^{-1} Y_{n,k} \quad (4.4.2)
\]

\[
\hat{\sigma}_y = k^{-1}(\sum_{i=1}^{k} Y_{n,i} + (n-k)Y_{n,k}) - \frac{1}{k} + k^{-1}(n-k)Y_{n,k} \quad (4.4.3)
\]

The MLE \( \hat{\sigma}_y \) has several desirable properties, among them being that it is the best linear unbiased estimator for \( \sigma_y \) (Sarhan and Greenberg [1962, p. 354]). Its variance is \( \sigma_y^2/k \), which is also the reciprocal of

\[
I_{\sigma_y} = I_{\mu_y} = k \sigma_y^{-2} \quad (4.4.4)
\]

Using \( \hat{\sigma}_y \) we can estimate \( \mu_x \), \( \sigma_x \), and \( \rho \) with (4.2.2.7) and (4.2.2.8) and using (2.4.15), (2.4.16), and (2.4.17), we find that

\[
\begin{align*}
\mu_x^* &= \bar{X} - (n-k)S_{xy} Y_{n,k} / k S_y^2, \quad E(\mu_x^*) = \mu_x \\
V(\mu_x^*) &= \sigma_x^2 / k + k^{-3}(n-k)^2 \rho^2 S_y^2 E(Y_{n,k} / S_y^2)
\end{align*}
\]

(note that the last term may be large for large amount of censoring)

and

\[
E(\sigma_x^2) = \sigma_x^2(1 + \rho^2/k) + k^{-1}\sigma_y^2[E(\sigma_y^2 / S_y^2)^2 - 2]
\]
since

\[ E(\hat{\sigma}_y^2) = k^{-1}(k+1)\sigma_y^2. \] (4.4.6)

From Sarhan and Greenberg [1962, p. 343], the order statistics have the following moments, with \( Z_{n,i} = \frac{Y_{n,i}}{\sigma_y} \):

\[
\begin{align*}
  a_{n,i} &= E(Z_{n,i}) = \sum_{l=1}^{i} (n-l+1)^{-1} \\
  b_{n,i} &= V(Z_{n,i}) = \text{Cov}(Z_{n,i}, Z_{n,j}) = \sum_{l=1}^{i} (n-l+1)^{-2}, \ i \leq j.
\end{align*}
\] (4.4.7)

So

\[
\begin{align*}
  l'u &= \sum_{i=1}^{k} E(Z_{n,i}^{-1}) = -(n-k)a_{n,k}, \text{ using } (4.4.3) \\
  l'w &= \sum_{i=1}^{k} (E(Z_{n,i}^{-1}))^2 = \sum_{i=1}^{k} (b_{n,i} + a_{n,i}^2 - 2a_{n,i} + 1)
\end{align*}
\] (4.4.8)

and

\[
\begin{align*}
  \sum_{i=1}^{k} b_{n,i} &= \sum_{i=1}^{k} \sum_{l=1}^{i} (n-l+1)^{-1} = \sum_{i=1}^{k} \frac{(k-i+1)/(n-i+1)^2}{i=1} \\
  &= \sum_{i=1}^{k} \{(n-i+1) + [(k-i+1)-(n-i+1)]/(n-i+1)^2} \\
  &= a_{n,k} - (n-k)b_{n,k}
\end{align*}
\] (4.4.9)
\[ \sum_{i=1}^{k} a_{n,i}^{2} = \sum_{i=1}^{k} (k-i+1)(n-i+1)^{-2} + 2 \sum_{i>l} (k-i+1)(n-i+1)^{-1}(n-l+1)^{-1} \]

\[ = a_{n,k} - (n-k)b_{n,k} \quad (4.4.10) \]

The rightmost term is

\[ 2 \sum_{i>l} (n-i+1)/[(k-i+1)-(n-i+1)](n-i+1)^{-1}(n-l+1)^{-1} \]

whose first term is

\[ 2 \sum_{i>l} (n-l+1)^{-1} -2(n-k) \sum_{i>l} (n-i+1)^{-1}(n-l+1)^{-1} \]

and second term is

\[ (n-k)(b_{n,k} - a_{n,k}^{2}) \]

so (4.4.10) equals

\[ 2k - [1 + 2(n-k)]a_{n,k} - (n-k)a_{n,k}^{2} \quad (4.4.11) \]

implying
With these quantities, the large sample covariance matrix in (4.2.1.15) can readily be evaluated.

An indication of the power of the \( r^* \) correlation test is given by the expected value of

\[
U^* = \sum_{i=1}^{k} Z_{n,i}^2 - k \left( \sum_{i=1}^{k} Z_{n,i} \right)^2. \quad (4.4.14)
\]

By combining (4.4.9) and (4.4.11), the expected value of the first term is

\[
E \left( \sum_{i=1}^{k} Z_{n,i}^2 \right) = 2k - 2(n-k)a_{n,k} - (n-k)(b_{n,k} + a_{n,k}^2). \quad (4.4.15)
\]

Using the fact that \( V \left[ \sum_{i=1}^{k} Z_{n,i} + (n-k)Z_{n,k} \right] = k \) we obtain

\[
E \left( \sum_{i=1}^{k} Z_{n,i}^2 \right)^2 = k + (E \sum_{i=1}^{k} Z_{n,i}^2)^2 - (n-k)^2 b_{n,k}
- 2(n-k) \text{Cov}(\sum_{i=1}^{k} Z_{n,i}, Z_{n,k}). \quad (4.4.16)
\]

The second term is obtained from (4.3.15). For the covariance term,
\[
\text{Cov}(\sum_{i=1}^{k} Z_{n,i}, Z_{n,k}) = \text{E}(\sum_{i=1}^{k} Z_{n,i} Z_{n,k}) - \text{E}(Z_{n,k}) \sum \text{E}(Z_{n,i})
\]

\[
= \sum_{i=1}^{k} \text{Cov}(Z_{n,i}, Z_{n,k}) = \sum_{i=1}^{k} b_{n,i} \quad \text{(since } i \leq k)\]

\[
= a_{n,k} - (n-k)b_{n,k} \quad \text{by } (4.4.9),
\]

(4.4.17)

Thus we obtain

\[
\text{E}\{(\sum_{i=1}^{k} Z_{n,i}^2)\} = k^2 + k-2(n-k)(k+1)a_{n,k} + (n-k)^2(b_{n,k} + a_{n,k}^2) \quad (4.4.18)
\]

and using (4.4.15),

\[
\text{E}(U^*) = k-1 + k^{-1}(n-k)[2a_{n,k} - n(b_{n,k} + a_{n,k}^2)]. \quad (4.4.19)
\]

For large amounts of censoring, \(E(U^*)\) is small in comparison with the normal distribution (for \(n = 1000, k = 50, E(U^*) = 6.79\) for normal distribution, 0.011 for exponential). Now since \(a_{n,k} = -\log (1-p), b_{n,k} = \frac{1}{n-k} - \frac{1}{n} = p(1-p)/n\) where \(p = k/n\),

\[
(k-1)^{-1}E(U^*) \approx 1-p^{-2}(1-p) \log^2 (1-p) \quad (4.4.20)
\]

for large \(n\) and fixed \(p\). This quantity does not exceed .5 until \(p > .95\). Since \(U^*\) relates to the non-centrality parameter of the conditional \(F\) distribution, power characteristics of the test of
H₀: β = 0 will be inferior with this exponential model with any substantial amount of censoring. Where the exponential distribution is most dense (in its left tail), there is little variation in the smallest order statistics. Either the linear regression model or the assumption of homoscedastic variances for this setup is unreasonable. Thus while the extreme value distribution model performed quite well, this theory may be of limited utility in some non-normal models.

It is interesting to consider the marginal density of X in this model where Y has the exponential distribution. Taking σₓ = σᵧ = 1 and μₓ = 0 for simplicity, X has the standard normal density when ρ = 0, and when ρ ≠ 0 it has density

\[
f(x) = \int_0^\infty [2\pi(1-\rho^2)]^{-1/2} \exp\left\{-\frac{(x-\rho(y-1))^2}{2(1-\rho^2)}\right\}\exp(-y)dy
\]

where \(\Phi(\cdot)\) is the standard normal d.f. This density closely resembles the normal density for moderate ρ and becomes more skewed as ρ becomes large in absolute value; however its first two moments remain unchanged. A graph of \(f(x)\) is shown in Figure 4.1.
Figure 4.1. Graph of Marginal Density of $X_i$ when $Y_i$ has an Exponential Distribution
A model in which the marginal distribution of \( X \) should depend on \( \rho \) is one in which both \( X \) and \( Y \) are reflections of another variable with a strength of association \( \rho \).

The problems with this exponential model can be remedied if instead of linear regression holding the regression on \( X \) is log-linear, i.e.,

\[
E(X_i | Y_i) = \mu_X + \beta(T_i - \mu_Y)
\]

where \( T_i = \log Y_i \). As far as estimation of parameters associated with \( X \) is concerned, we could use the extreme value distribution for \( Y \) as discussed earlier. This is a generalization of the log linear model since \( -T_i \) has the extreme value distribution with parameters \( \xi = -\log \sigma_y \), \( \theta = 1 \). However, the present discussion will be for the log linear model with \( Y \) having an exponential distribution.

Using properties of the extreme value distribution,

\[
E(T_i) = \log \sigma_y - \gamma
\]

\[
V(T_i) = \pi^2/6
\]

and since \( 2k \hat{\sigma_y}/\sigma_y \sim \chi^2_{2k} \) (Epstein and Sobel [1953] in Sarhan and Greenberg [1962, p. 362]) it follows that the moment generating function of \( \log \hat{\sigma_y} \) is
\[ m(t) = (\sigma_y/k)^t \Gamma(t+k)/\Gamma(k). \]  

(4.4.24)

Therefore,

\[ E(\log \hat{\sigma}_y) = m'(0) = \log \sigma_y + \psi(k) - \log k \]  

(4.4.25)

\[ V(\log \hat{\sigma}_y) = m''(0) - m'(0)^2 = \psi'(k) \]

where

\[ \psi(k) = \frac{\Gamma'(k)}{\Gamma(k)} = -\gamma + \sum_{i=1}^{k-1} \frac{1}{i} \]

and

(4.4.26)

\[ \psi'(k) = \frac{\pi^2}{6} - \sum_{i=1}^{k-1} \frac{1}{i^2} + \frac{1}{k-1}. \]

Note that \( \lim_{k \to \infty} \psi(k) - \log k = \lim_{k \to \infty} \psi'(k) = 0. \) An unbiased estimator of \( \mu_T = E(T_1) \) is

\[ \mu_T^* = \log \hat{\sigma}_y + \log k - \sum_{i=1}^{k-1} \frac{1}{i} \]  

(4.4.27)

with variance \( \psi'(k). \) From Johnson and Kotz [1970, p. 285], if \( k = n, \) \( \log \hat{\sigma}_y \) is the MLE of \( E(T_1) \) and for any \( k,n, \) \( V(\mu_T^*) \) is equal
to the variance of the MLE of $E(T^*_1)$ for a complete sample of size $k$.

From $\mu^*_T$ an unbiased estimator of $\mu^*_x$ can be produced as well as simple estimators of $\rho$ and $\sigma^*_x$ since $V(T^*_1)$ is a known constant. All parameters are estimated in the manner described previously with $Y$ replaced by $T = \log Y$, e.g.

$$\hat{\beta} = \frac{S_{XT}}{S^2_T}$$

$$\mu^*_x = \bar{X} + (S_{XT}/S^2_T)(\bar{T} - \mu^*_T)$$

$$\sigma^*_x = S^2_x + (S_{XT}^2/S^4_T)(\pi^2/6 - S^2_T)$$

$$\rho^* = (\pi/\sqrt{6})S_{XT}/\sigma^*_x S^*_T.$$

The power function of the test of $\beta = 0$ for this model is different from that of section 4.3 since here we are sampling from the opposite tail of the distribution of $Y^*_1$. The average values of $U^*$ here are:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$E(U^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>5.653</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>12.244</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>30.545</td>
</tr>
</tbody>
</table>

The power estimates calculated as in table 4.1 are shown in table 4.2.
The one moment approximation is again satisfactory. By comparison with table 2.3 the power characteristics of this model, when it holds, are greatly superior to those of the bivariate normal model under censoring.

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>ρ</th>
<th>Power</th>
<th>Power from One Moment Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>.1</td>
<td>.0560</td>
<td>.0560</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.1102</td>
<td>.1100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.2486</td>
<td>.2569</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.9079</td>
<td>.9957</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.1</td>
<td>.0629</td>
<td>.0628</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.1804</td>
<td>.1808</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.4516</td>
<td>.4810</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.9953</td>
<td>.9999</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>.1</td>
<td>.0847</td>
<td>.0846</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.3910</td>
<td>.3988</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.8289</td>
<td>.8783</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
5.1. Introduction.

In this chapter we will derive locally most powerful rank tests for independence of $X_i$ and $Y_i$ when $(X_i, Y_i)$ are from a general bivariate distribution. The tests will be explored in more detail for the case when $X_i$ is conditionally normally distributed or when linear regression-type alternatives are appropriate. Empirical power studies will be made for $Y_i$ having a normal or extreme value distribution. An asymptotic study will then be made of relative efficiencies of rank tests of independence with regards to local regression alternatives. Finally, an omnibus nonparametric test of independence due to Hoeffding [1948] will be studied for the case of censored bivariate samples.

5.2. Locally Most Powerful Rank Test of Independence

Suppose that $Y_i$ has density $g(y)$ with d.f. $G(y)$ and $f_0(x|y)$ is the conditional density of $X_i$ given $Y_i = y$. We assume that $\theta$ is a parameter such that $X_i$ and $Y_i$ are independent if $\theta = 0$ and
that $\theta$ is not a parameter of $g(y)$. When $\theta = 0$ we write
\[ f_0(x|y) = f_0(x). \]
With $X$ and $Y$ defined as in section 1.4, the joint density of $X$ and $Y$ is

\[
\begin{align*}
n[k] & \prod_{i=1}^{k} f_\theta(x_i|y_i)g(y_i) \left[1-G(y_k)\right]^{n-k}, \quad y_1 \leq y_2 \leq \cdots \leq y_k. \\
\end{align*}
\]

(5.2.1)

The density when $X_i$ and $Y_i$ are independent is

\[
\begin{align*}
n[k] & \prod_{i=1}^{k} f_0(x_i)g(y_i) \left[1-G(y_k)\right]^{n-k}, \quad y_1 \leq y_2 \leq \cdots \leq y_k. \\
\end{align*}
\]

(5.2.2)

The likelihood ratio then is

\[
L_{n,k} = \prod_{i=1}^{k} \left\{ \frac{f_\theta(X_n[i]|Y_n,i)}{f_0(X_n[i])} \right\}.
\]

(5.2.3)

It can easily be seen that

\[
Q_{n,k} = (\partial L_{n,k}/\partial \theta) \bigg|_{\theta=0} = \sum_{i=1}^{k} \partial \log f_\theta(X_n[i]|Y_n,i)/\partial \theta \bigg|_{\theta=0}.
\]

(5.2.4)

Let $S_{n,i}$ be the rank of $Y_{n,i}$ among $Y_1, \ldots, Y_n$ ($S_{n,i} = i$), and $S_n = (S_{n,1}, \ldots, S_{n,k})$. For continuous distributions, the probability of a tie is zero and ranks are well defined with probability 1.

For the induced order statistics, the rank of $X_{n[i]}$ among $X_1, \ldots, X_n$ is unknown. Define $R_{k,i}$ to be the rank of $X_{n[i]}$ among $X_{n[1]}, \ldots, X_{n[k]}$ and $R_k = (R_{k,1}, \ldots, R_{k,k})$. Then following, for example,
Cox and Hinkley [1974, p. 188], the locally most powerful rank test (l.m.p.r.t.) for testing $H_0: \theta = 0$ using only $R_k$ and $S_n$ is based on the statistic

$$T_{n,k} = E(Q_n| R_k, S_n, \theta = 0)$$

$$= \sum_{i=1}^{k} E\left(\log f_\theta(X_n[i]| Y_{n,i})/\theta| R_k, i, S_{n,i}, \theta = 0\right)$$

$$= \sum_{i=1}^{k} c_{n,k}(i, R_k, i). \quad (5.2.5)$$

Computation of $c_{n,k}(i, R_k, i)$ and therefore $T_{n,k}$ is facilitated by using the fact that under $H_0$, $X$ represents a complete random sample of size $k$ from $f_\theta(x)$. In that case $X_n[R_1], \ldots, X_n[R_k]$ are the $k$ order statistics of the $k$ i.i.d.r.v. $X_n[i], 1 \leq i \leq k$ where we have abbreviated $R_{k,i}$ by $R_i$. Now under $H_0$: $X$ and $Y$ independent, $R_k$ can take on any of the $k!$ permutations of the integers $1, \ldots, k$. This fact gives rise to an exact permutation test for $H_0: \theta = 0$.

Hoeffding [1951] showed that, under $H_0$,

$$E(T_{n,k}) = k^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} c_{n,k}(i, j), \text{ and}$$

$$V(T_{n,k}) = (k-1)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} d_{n,k}^2(i, j), \quad (5.2.6)$$
where

\[ d_{n,k}(i,j) = c_{n,k}(i,j) - k^{-1} \sum_{g=1}^{k} c_{n,k}(g,j) - k^{-1} \sum_{h=1}^{k} c_{n,k}(i,h) + k^{-2} \sum_{g=1}^{k} \sum_{h=1}^{k} c_{n,k}(g,h). \]  

(5.2.7)

He also showed that under suitable conditions on the \( c_{n,k}(i,j) \),

\[ V(T_{n,k})^{-1/2} (T_{n,k} - E T_{n,k}) \]

converges in distribution to the standard normal distribution under \( H_0 \) as \( k \to \infty \). [Note that we have introduced the dependence of \( c_{n,k}(i,j) \) on the additional parameter \( n \) which was not the case in Hoeffding [1951], however, the same theory still applies.] Motoo (1957) weakened the conditions on the \( c_{n,k}(i,j) \) to the following:

\[ \lim_{k \to \infty} \frac{1}{\epsilon} \left| \frac{d_{n,k}(i,j)}{d_{n,k}} \right|^2 = 0 \]  

(5.2.8)

where \( d_{n,k} = k^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} d_{n,k}(i,j)^2 \).

For many densities \( f_\theta \), the quantity \( c_{n,k}(i,j) \) will factor into a product \( a_{n,k}(i) b_{n,k}(j) \), so that the resulting test statistic is a linear rank statistic, and the resulting test is easy to apply. There are distributions, however, for which such factorization is not possible. An example is the Gumbel bivariate exponential distribution.
Assume now the more restricted model \( f_\theta(x|y) = [2\pi(1-\theta^2)]^{-1/2} \exp\{-x-\theta y\}^2/2(1-\theta^2) \) which does yield linear rank statistics. Location and scale parameters can be ignored because the rank tests to be considered are location and scale invariant. For this model,

\[
Q_{n,k} = \sum_{i=1}^{k} X_{n[i]} Y_{n,i}.
\]  

(5.2.9)

Define \( u_{k,i} = E(X_{k,i}) \) where \( X_{k,i} \) is the \( i \)th order statistic out of \( k \) from \( f_\theta(x) = (2\pi)^{-1/2} \exp(-x^2/2) \). Also define \( t_{n,i} = E(Y_{n,i}) \) where as before \( Y_{n,i} \) is the \( i \)th order statistic out of \( n \) from the parent distribution \( g(y) \). Then

\[
E(X_{n[i]}|R_i, \theta = 0) = u_{k,R_i},
\]

(5.2.10)

\[
E(Y_{n,i}|S_i) = t_{n,i},
\]

and since \( X_{n[i]} \) and \( Y_{n,i} \) are independent when \( \theta = 0 \),

\[
T_{n,k} = E(Q_{n,k}|R_k S_n, \theta = 0) = \sum_{i=1}^{k} u_{k,R_i} t_{n,i}.
\]

(5.2.11)

This statistic is useful not only for the conditional normal model but also for general monotonic association alternatives. As in Ghosh and Sen [1971], one can also define "mixed rank statistics" which are asymptotically efficient. These are
The statistic $M_{n,k}^Y$ might be preferable since the scores $u_{n,i}$ are more dependent on $g(y)$ for large amounts of censoring.

Since under $H_0$ all permutations of $R_i, 1 \leq i \leq k$ among the integers $1, \ldots, k$ are equally likely, $T_{n,k}$ provides a non-parametric test of independence. By Theorem II.3.1.c of Hájek and Šidák [1967, p. 61], under $H_0$,

$$E(T_{n,k}) = k \overline{u} \overline{t}$$

and

$$V(T_{n,k}) = \sigma_u^2 \sum_{i=1}^{k} \left( t_{n,i} - \overline{t} \right)^2,$$

where

$$\overline{u} = k^{-1} \sum_{i=1}^{k} u_{k,i}, \quad \overline{t} = k^{-1} \sum_{i=1}^{k} t_{n,i},$$

and

$$\sigma_u^2 = (k-1)^{-1} \sum_{i=1}^{k} \left( u_{k,i} - \overline{u} \right)^2.$$
These results hold true for all $t_{n,i}$ and $u_{k,i}$ whether or not these scores are defined as in (5.2.7). Assume that $\sum_{i=1}^{k} (t_{n,i} - \bar{t})^2 > 0$ and that there exists a square integrable function $\tau(u)$ such that

$$\lim_{k \to \infty} \int_0^1 [u_k, l + [v_k] - \tau(v)]^2 dv = 0. \quad (5.2.15)$$

Hájek and Šidák [1967] have shown (Theorem V.1.6.a) that under $H_0$, $T_{n,k}$ is asymptotically normal for $k \to \infty$ with mean and variance given in (5.2.13). The asymptotic variance of $T_{n,k}$ can also be used:

$$\lim_{k \to \infty} V(T_{n,k}) = \left[ \sum_{i=1}^{k} (t_{n,i} - \bar{t})^2 \right] \int_0^1 [\tau(u) - \bar{\tau}]^2 du, \quad (5.2.16)$$

where $\bar{\tau} = \int_0^1 \tau(u) du$. So for large $k$, critical values from the standard normal distribution can be compared to

$$Z_{n,k} = (k-1)^{1/2} r_{n,k} \quad (5.2.17)$$

where $r_{n,k} = \left( \sum_{i=1}^{k} u_{k,i} R_{i} (t_{n,i} - \bar{u} \bar{t}) / \left[ \sum_{i=1}^{k} (u_{k,i} - \bar{u})^2 \sum_{i=1}^{k} (t_{n,i} - \bar{t})^2 \right] \right)^{1/2} \quad (5.2.18)$

can be used as a rank measure of correlation. Another approximation is to treat
(k-2)^{1/2} \frac{r_{n,k}}{(1-r_{n,k}^2)^{1/2}} \quad (5.2.19)

as an approximate \( t \) random variable with \( k-2 \) D.F. This approximation is accurate for small sample sizes in many situations (e.g. logistic or normal generating distributions).

The paper by Shirahata [1975] also deals with the \( \lambda.m.p.r.t. \) under censoring but in the case where \( X \) and \( Y \) are both time-censored, i.e. \( X_i \) is observed if it is one of the \( n_1 \) smallest \( X_i \) and \( Y_i \) is observed if it is one of the \( n_2 \) smallest \( Y_i \).

### 5.3. Empirical Power Studies

In this section we study power properties of the rank tests of independence discussed in section 5.2, assuming the conditional normality model for \( X \) given \( Y \) and for \( g(y) \) being the normal or extreme value distribution. We write \( \rho \) instead of \( \theta \) to be consistent with discussions of parametric tests. The power is estimated here by generating 1000 values of \( r_{n,k} \) for each \( n,k,\rho \), estimating the 95\textsuperscript{th} percentile of \( |r_{n,k}| \) for \( \rho = 0 \) and computing the proportion of \( |r_{n,k}| \) greater than this critical value when \( \rho \neq 0 \). Blom scores defined in (1.4.9) are employed for \( u_{k,i} \) and for \( t_{n,i} \) when \( g(y) \) is the normal density. The empirical critical values are:
When \( X \) and \( Y \) are truly from a bivariate normal distribution and normal scores are used for \( u_{k,i} \) and \( t_{n,i} \), the following power estimates result:

<table>
<thead>
<tr>
<th>( n )</th>
<th>( k )</th>
<th>Empirical Critical Value</th>
<th>Critical Value Estimated from (5.2.19)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>.620</td>
<td>.632</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.457</td>
<td>.444</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>.274</td>
<td>.279</td>
</tr>
</tbody>
</table>

TABLE 5.1

ESTIMATED POWER OF 5\% NORMAL SCORES TEST FOR CENSORED DATA FROM THE BIVARIATE NORMAL DISTRIBUTION

<table>
<thead>
<tr>
<th>( n )</th>
<th>( k )</th>
<th>( \rho )</th>
<th>Estimated Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>.1</td>
<td>.061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.142</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.743</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.1</td>
<td>.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.895</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>.1</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.319</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.993</td>
</tr>
</tbody>
</table>
Except for $n = 100 \; k = 20 \; \rho = .9$ these results compare quite favorably to the power of the best parametric test for the bivariate normal case as found in table 2.3. The normal scores test statistic has a null distribution that is independent of $f_0(x|y)$ and $g(y)$ in the sense that it has a permutation distribution depending only on the assumption that all permutations of $R_k$ are equally likely. Therefore, the normal scores test may be preferred over the parametric test which assumes that either $f_0(x)$ or $g(y)$ is normal before its null distribution is known. The normal scores test may also be appropriate in the presence of non-linear regression or heteroscedasticity.

In order to examine power properties of the normal scores test when the scores are not optimal we now assume the extreme value/normal model of section 4.3 but continue to use normal scores for $t_{n,i}$ and $u_{k,i}$. The power estimates are found in table 5.2.
TABLE 5.2
ESTIMATED POWER OF 5% NORMAL SCORES TEST FOR CENSORED DATA FROM THE EXTREME VALUE/NORMAL DISTRIBUTION

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$\rho$</th>
<th>Estimated Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>.1</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.499</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>.1</td>
<td>.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.513</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>.1</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.3</td>
<td>.089</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5</td>
<td>.106</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.9</td>
<td>.652</td>
</tr>
</tbody>
</table>

By comparing those results with table 4.1, we see that very little power is lost over the parametric test even though only the scores for $X$ are optimal. The normal scores test of independence may perform well for many bivariate distributions.
5.4. Asymptotic Relative Efficiencies of Linear Rank Tests

Assume that optimum scores for $R_k^x$ are generated asymptotically from a square integrable function $\tau_1(u)$ as in (5.2.15) and likewise that scores for $R_k^y$ are generated by $\psi_1(u)$ satisfying

$$\lim_{n \to \infty} \int_0^1 [t_{n,1}^{-1} [un] - \psi_1(u)]^2 \, du = 0$$

(5.4.1)

where $k = pn$ and $p \in (0,1]$ is fixed. We assume that the linear rank statistic from $\tau_1(u)$ and $\psi_1(u)$ is locally most powerful for testing $H_0: \theta = 0$ and is asymptotically efficient relative to the best parametric test in the Pitman (local alternatives) sense.

Now let $\tau_2(u)$ and $\psi_2(u)$ be other corresponding score functions which are not necessarily optimal. Define

$$\rho_x^2 = \frac{\int_0^1 \tau_1(u)\tau_2(u) \, du - \bar{\tau}_1 \bar{\tau}_2}{\int_0^1 [\tau_1(u) - \bar{\tau}_1]^2 \, du \int_0^1 [\tau_2(u) - \bar{\tau}_2]^2 \, du}$$

(5.4.2)

and

$$\rho_y^2 = \frac{\int_0^1 \psi_1(u)\psi_2(u) \, du - \bar{\psi}_1 \bar{\psi}_2}{\int_0^1 [\psi_1(u) - \bar{\psi}_1]^2 \, du \int_0^1 [\psi_2(u) - \bar{\psi}_2]^2 \, du}$$

(5.4.3)

where $\bar{\tau}_i = \int_0^1 \tau_i(u) \, du$, $i=1,2$
and

\[ \bar{\psi}_i = p^{-1} \int_0^p \psi_i(u) \, du, \quad i = 1, 2. \]  

(5.4.4)

Note that \( \rho_x^2 \) and \( \rho_y^2 \) can be interpreted as squared correlation coefficients between corresponding "correct" and "incorrect" scores.

We now argue heuristically based on the results of Ghosh and Sen [1971] and Basu [1967] that the asymptotic relative efficiency of the linear rank test based on \((\tau_2, \psi_2)\) to that based on \((\tau_1, \psi_1)\) is given by

\[ e_{12} = \rho_x^2 \rho_y^2. \]  

(5.4.5)

For example, choosing correct X-scores \((\tau_1(u) = \tau_2(u))\) yields \( \rho_x^2 = 1 \). However, choosing Wilcoxon scores for X \((\tau_2(u) = u)\) when the optimal scores are normal scores \((\tau_1(u) = \phi^{-1}(u))\) yields \( \rho_x^2 = 3/\pi \). If this same mistake is made in choosing scores for Y \((\psi_1(u) = \phi^{-1}(u), \quad \psi_2(u) = u)\), the following component quantities for \( \rho_y^2 \) are:

\[ \bar{\psi}_1 = p^{-1} \int_0^p \phi^{-1}(u) \, du = p^{-1} \int_{-\infty}^a (2\pi)^{-1/2} y^{-1/2} e^{-y^2/2} \, dy = -\phi(a)/p \]  

(5.4.6)

where \( a = \phi^{-1}(p), \quad \phi(u) = (2\pi)^{-1/2} e^{-u^2/2} \).
The squared correlation of $Y$-score functions has respective values .636, .902, and .955 (3/π) when $p = .00057$, .5, and 1 respectively. Thus for example very little efficiency is lost in using the Spearman correlation coefficient (which incorporates Wilcoxon scores) instead of the normal scores correlation for censored bivariate normal data.

5.5. **Hoeffding's General Test of Independence as applied to Censored Bivariate Samples**

Suppose that two random variables have a d.f. $H(x,y)$ with continuous joint and marginal density functions. Let
where

$$D(x,y) = H(x,y) - H(x,\infty)H(\infty, y). \quad (5.5.2)$$

The "D" test based on a random sample \{\((X_i, Y_i)\), 1 \leq i \leq n\} was devised by Hoeffding in 1948. It depends only on the ranks of the \(X\) and \(Y\) vectors and tests \(H_0: \Delta = 0\), being consistent against all alternatives to independence. The test can also be applied to a censored bivariate sample \{\((X_{n[i]}, Y_{n[i]})\), 1 \leq i \leq n\} since under the null hypothesis all permutations of the ranks \(R_k\) are equally likely (fixing the ranks \(S_n\)) due to \{\(X_{n[i]}\), 1 \leq i \leq k\} representing a random sample of size \(k\) under \(H_0\).

Hoeffding derived an unbiased estimator of \(\Delta\) based on a U-statistic. Alternatively \(\Delta\) can be estimated with the sample bivariate empirical distribution:

$$\hat{\Delta} = \int_0^\infty \int_0^\infty D_n^2(x,y) dH_n(x,y) \quad (5.5.3)$$

where

$$H_n(x,y) = n^{-1} \sum_{i=1}^{n} I\{X_i < x, Y_i < y\}, \quad (5.5.4)$$

$$D_n(x,y) = H_n(x,y) - H_n(x,\infty)H_n(\infty, y),$$
and \( \mathbb{1}_A \) is the indicator function for the set \( A \). Blum, Kiefer, and Rosenblatt [1961] developed another test of this type in the spirit of Kolmogorov-Smirnov tests. Their test statistic is

\[
B_n = \sup_{x,y} D_n(x,y).
\]  

(5.5.5)

For a censored bivariate sample the following information is available:

\[
F_{n,k}(x) = k^{-1} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}} c_i
\]  

(5.5.6)

\[
= k^{-1} \sum_{i=1}^{k} \mathbb{1}_{\{X_{n[i]} \leq x\}},
\]

\[
G_{n,k}(y) = n^{-1} \sum_{i=1}^{k} \mathbb{1}_{\{Y_{n[i]} \leq y\}},
\]  

(5.5.7)

\[
H_{n,k}(x,y) = n^{-1} \sum_{i=1}^{k} \mathbb{1}_{\{X_{n[i]} \leq x\}} \mathbb{1}_{\{Y_{n,i} \leq y\}},
\]  

(5.5.8)

\[
D_{n,k}(x,y) = H_{n,k}(x,y) - F_{n,k}(x) G_{n,k}(y)
\]  

(5.5.9)

where
When \( Y_{n,k} > y \) the following simplifications hold:

\[
G_{n,k}(y) = n^{-1} \sum_{i=1}^{n} I\{Y_i \leq y\} \quad (5.5.11)
\]

and

\[
H_{n,k}(x,y) = n^{-1} \sum_{i=1}^{n} I\{X_i \leq x\} I\{Y_i \leq y\}. \quad (5.5.12)
\]

These sample distribution functions have the following properties:

\[
E[G_{n,k}(y) | Y_{n,k} \geq y] = G(y), \quad (5.5.13)
\]

and

\[
E[H_{n,k}(x,y) | Y_{n,k} \geq y] = H(x,y). \quad (5.5.14)
\]

and if \( \theta \) denotes the dependence parameter as in section 5.2,

\[
E[F_{n,k}^0(x) | \theta = 0] = F(x) \quad (5.5.15)
\]

and
E[D_{n,k}(x,y)|Y_{n,k} > y, \Theta = 0] = 0 \hspace{1cm} (5.5.16)

since \hspace{0.5cm} E[c_i] = k/n. \hspace{1cm} (5.5.17)

Here \hspace{0.5cm} F(x) and \hspace{0.5cm} G(y) are respectively the marginal d.f. of \hspace{0.5cm} X_i and \hspace{0.5cm} Y_i. A statistic for testing independence is

\[ \hat{\Delta}_{n,k} = \int_{-\infty}^{\infty} \int_{-\infty}^{Y_{n,k}} D_{n,k}^2(x,y)dH_{n,k}(x,y). \hspace{1cm} (5.5.18) \]

An alternative test statistic analogous to (5.5.5) is \[ B_{n,k} = \sup_{-\infty < x < \infty, \infty < y < Y_{n,k}} D_{n,k}(x,y). \]

These statistics should perhaps be studied in more detail. However, Hoeffding's \( D \) test can be carried out without difficulty. It would be interesting to see how such an omnibus test of independence performs with respect to a simple test directed at simple alternatives.

To this end, 1000 censored bivariate normal random vectors were generated for different \( n,k, \) and \( \rho. \) The empirical upper 95th percentile when \( \rho = 0 \) was used as a critical value since the null distribution of \( D \) is tabulated only for very small sample sizes. The resulting critical values are:
Power is estimated by computing the proportion of observed $D$ statistics greater than $D_{.95}$. The results are in table 5.3.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$D_{.95}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0.00503</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.00218</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>0.00067</td>
</tr>
</tbody>
</table>

**TABLE 5.3**

**ESTIMATED POWER OF 5% HOEFFDING D TEST FOR CENSORED BIVARIATE NORMAL DISTRIBUTION**

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k$</th>
<th>$\rho$</th>
<th>Estimated Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10</td>
<td>0.1</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9</td>
<td>0.573</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.1</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9</td>
<td>0.782</td>
</tr>
<tr>
<td>1000</td>
<td>50</td>
<td>0.1</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.236</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9</td>
<td>0.970</td>
</tr>
</tbody>
</table>
By comparison with table 2.3 the power is reduced from that of the likelihood ratio test but is not devastated. The power reduction is the smallest for the case $n = 1000 \, k = 50$ in which a high proportion of censoring is present.
CHAPTER VI
SUMMARY AND SUGGESTIONS FOR FURTHER RESEARCH

6.1. **Summary.**

Many tools of classical multivariate theory have been extended for use in analyzing censored multivariate samples consisting of order statistics of one variate and induced order statistics of other variates. Maximum likelihood and simplified estimators have been developed for means, variances, covariances, correlation coefficients, and multiple, partial and canonical correlation coefficients. Statistics for testing total or partial independence of random vectors have also been derived. Linear rank tests for independence in the bivariate case have also been proposed and Hoeffding's general test of independence in this case has been studied.

Properties of estimators have been studied by deriving Cramèr-Rao lower bounds for dispersion and by simulations. Power properties of some of the proposed test procedures have been studied by deriving conditional power functions and by making empirical power calculations. The rank tests for independence were shown to have many advantages over the parametric tests since the power
characteristics of the two types of tests are similar while the rank tests make fewer assumptions.

6.2. **Suggestions for Further Research**

There are several open areas remaining in this research. For one, the unconditional distribution of test statistics such as (2.5.1.8) could be derived if the distribution of the sample variance calculated from a censored normal sample could be obtained. This may not turn out to be a tractable problem. Also, it would be advantageous to prove formally that the partial correlation test in section 3.4.2 is actually the likelihood ratio test [the author could find no formal proof of this even for the complete sample case].

Another question worth considering is how much information regarding tests of independence is lost when only induced order statistics are observed and not the entire vector of concomitant data. This question could be studied by comparing the power characteristics of the rank test proposed in section 5.2 with those of the test proposed by, for example, Brown *et al.* [1974].


