INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
On multiple comparisons for complex experimental designs

Soong, Wen-Chang, Ph.D.
The Ohio State University, 1989
ON MULTIPLE COMPARISONS FOR COMPLEX EXPERIMENTAL DESIGNS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of the Ohio State University

By

Wen-Chang Soong, B. S.

**** *

The Ohio State University
1989

Dissertation Committee:
Prof. Jason C. Hsu
Prof. William I. Notz
Prof. Dennis K. Pearl
Prof. Elizabeth A. Stasny

Approved by

Adviser

Department of Statistics
To My Wife and Son
ACKNOWLEDGMENTS

I express sincere appreciation to Dr. Jason C. Hsu for his guidance, insight and encouragement throughout the research. Thanks go to the other members of my advisory committee, Dr. William I. Notz, Dr. Dennis K. Pearl and Dr. Elizabeth A. Stasny, for their suggestions and comments. Gratitude is expressed to Department of Statistics of the Ohio State University for the support. To my wife, Pao-Shieh, I offer sincere thanks for your encouragement, your unshakable faith in me and your willingness to endure with me the vicissitudes of my endeavors. To my son, Jin-Yuh, I thank you for understanding my frequent absences.
VITA

November 2, 1955 ................................................. Born in Janghwa, Taiwan
1978 ................................................................. B. S., Nat'l Chung-Hsing Univ.,
Taipei, Taiwan
1984 - Present .................................................... Teaching, Research Associate and
Statistical Consultant, Dept. of Statistics,
Ohio State Univ., Columbus, Ohio

FIELDS OF STUDY

Multiple Comparison Procedures
Categorical Data Analysis
Statistical Computing
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
</tr>
<tr>
<td>VITA</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
</tr>
<tr>
<td>CHAPTER</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>I INTRODUCTION</td>
</tr>
<tr>
<td>Definitions</td>
</tr>
<tr>
<td>Nature of the Problem</td>
</tr>
<tr>
<td>Goal of the Study</td>
</tr>
<tr>
<td>Examples</td>
</tr>
<tr>
<td>Existing Results for the One-factor Structures</td>
</tr>
<tr>
<td>Extended Results for the Two-factor Structures</td>
</tr>
<tr>
<td>Computations</td>
</tr>
<tr>
<td>Applications</td>
</tr>
<tr>
<td>Outlines for the Other Chapters</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>II ONE-FACTOR STRUCTURES</td>
</tr>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Goal of the Chapter</td>
</tr>
<tr>
<td>Definitions and Preliminaries</td>
</tr>
<tr>
<td>Singular One-negative-factor Correlation with n=2</td>
</tr>
<tr>
<td>Sign invariance in evaluating two-sided probabilities</td>
</tr>
</tbody>
</table>

v
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 95% Confidence Intervals for Blood Lead Data</td>
<td>42</td>
</tr>
<tr>
<td>3.1 Critical Values given by Various Methods for Starch Data</td>
<td>66</td>
</tr>
<tr>
<td>3.2 Huber's Data</td>
<td>70</td>
</tr>
<tr>
<td>4.1 Ratios of the lengths of Bonferroni to Pearson Intervals</td>
<td>76</td>
</tr>
<tr>
<td>4.2 Observed Frequencies for Bank Data</td>
<td>79</td>
</tr>
<tr>
<td>4.3 Critical Values given by Various Methods for Bank Data</td>
<td>79</td>
</tr>
<tr>
<td>4.4 95% Confidence Intervals computed by Various Methods</td>
<td>80</td>
</tr>
<tr>
<td>for Bank Data</td>
<td></td>
</tr>
<tr>
<td>4.5 95% Confidence Intervals (in %) computed by Various Methods</td>
<td>81</td>
</tr>
<tr>
<td>for Bank data: Multiple Comparisons with a Control</td>
<td></td>
</tr>
<tr>
<td>4.6 Case Records of Eye-Testing of Women (Aged 30-39) Employees</td>
<td>83</td>
</tr>
<tr>
<td>In Royal Ordnance Factories</td>
<td></td>
</tr>
<tr>
<td>4.7 Studentized Residuals for Vision Data under the</td>
<td>83</td>
</tr>
<tr>
<td>Complete Symmetry Model</td>
<td></td>
</tr>
<tr>
<td>4.8 95% Confidence Intervals of all Symmetrical Pair Differences of</td>
<td>85</td>
</tr>
<tr>
<td>Off-diagonal Cell Probabilities in Vision Data</td>
<td></td>
</tr>
<tr>
<td>4.9 Observed Cell Counts(n), Estimated Expected Cell Counts(m),</td>
<td>86</td>
</tr>
<tr>
<td>and Adjusted Residuals(r) for Dowdall's Data</td>
<td></td>
</tr>
<tr>
<td>4.10 95% Confidence Intervals of Logits for Dowdall's Data</td>
<td>88</td>
</tr>
<tr>
<td>4.11 Two-way table for Psychiatric Therapy v.s. Psychoses</td>
<td>89</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Plot of the Real Part of Erfc(x+iy), x = 0 to 7 by 0.2 and y = 0 to 7 by 0.2</td>
<td>29</td>
</tr>
<tr>
<td>2.2 Plot of the Imaginary Part of Erfc(x+iy), x = 0 to 7 by 0.2 and y = 0 to 7 by 0.2</td>
<td>30</td>
</tr>
<tr>
<td>2.3 Plot of the Integrand g(y) with d*=1 and ρ=-0.5</td>
<td>32</td>
</tr>
<tr>
<td>2.4 Plot of the Integrand G=g(y)<em>10^9 with d</em>=1 and ρ=-0.5</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Plot of the Integrand g(y) with d*=1 and ρ=-1</td>
<td>33</td>
</tr>
<tr>
<td>2.6 Plot of the Integrand G=g(y)<em>100 with d</em>=1 and ρ=-1</td>
<td>33</td>
</tr>
<tr>
<td>2.7 Plot of the Integrands g(y) and f(z): d*=1 and ρ=-0.5</td>
<td>35</td>
</tr>
<tr>
<td>2.8 Timing on 3 Integration Methods: a: Romberg’s, b: Stoer’s and c: Gauss-Legendre’s</td>
<td>37</td>
</tr>
<tr>
<td>2.9 Actual Probabilities of Sidak’s and Slepian’s Bounds with a Non-singular One-negative-factor structure</td>
<td>39</td>
</tr>
<tr>
<td>3.1 Scatter Plot for Huber’s Data: b: Quadratic Fit, a(c): Linear Fit with (without) obs. 6</td>
<td>71</td>
</tr>
<tr>
<td>3.2 Plot of Residuals v.s. x: without obs. 6 for Huber’s Data</td>
<td>71</td>
</tr>
<tr>
<td>3.3 Plot of Residuals v.s. x: Linear Fit for Huber’s Data</td>
<td>72</td>
</tr>
<tr>
<td>3.4 Plot of Residuals v.s. x: Quadratic Fit for Huber’s Data</td>
<td>72</td>
</tr>
<tr>
<td>3.5 Q-Q Plot of Residuals: Quadratic Fit for Huber’s Data</td>
<td>73</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Consider $n$ independent normal populations with a common unknown variance $\sigma^2$ and respective means $\mu_i$, $i = 1, \ldots, n$. Suppose rectangular simultaneous confidence intervals for a set of parameters $\beta_j$, $j = 1, \ldots, r$, are of interest. For example, the $\beta_j$'s can be the population means themselves or the pairwise differences of the population means. Depending on the correlation structure of the estimators of the $\beta_j$'s, the evaluation of the exact critical value $q$, needed in constructing rectangular simultaneous confidence intervals for the $\beta_j$'s, can be either easily computable as in the comparisons of the population means, or extremely difficult to compute using current computer facilities, as in the all pairwise comparisons of the population means with $n > 3$. In this dissertation, new techniques for computing the exact critical value $q$ are proposed, when the correlation matrix of the estimators of the $\beta_j$'s can be decomposed into a one-factor structure $D \pm \lambda \lambda'$ or a two-factor structure $D \pm \lambda \lambda' \pm \gamma \gamma'$, where $\lambda$ and $\gamma$ are known vectors of constants, and $D$ is a diagonal matrix.

In terms of calculating probabilities with a known critical value $q$, the proposed method involves at most triple integrals by extending the univariate normal distribution to the complex plane for structures with negative factors in the decomposition. The techniques can be applied to obtain better confidence intervals in a variety of problems; in particular, in the area of categorical data analysis, where most of the existing comparison methods are too conservative in the sense that the confidence intervals are too wide.
In this chapter, the motivation of this dissertation will be discussed in terms of three examples. Existing results in this area will be reviewed in this chapter as well. We start with detailed definitions for some structures of a correlation matrix.

**Definitions**

A correlation matrix $\Sigma$ is said to have a two-factor structure if it can be decomposed as $\Sigma = D \pm \lambda \lambda' \pm \gamma \gamma'$, where $D$ is a diagonal matrix, and $\lambda$ and $\gamma$ are vectors of constants. In particular, if either $\lambda$ or $\gamma$, but not both, is zero, then $\Sigma$ is said to have a one-factor structure. Furthermore, the one-factor correlation matrix with a positive sign, or a negative sign in the decomposition is said to have one-positive-factor, or one-negative-factor structure respectively. The two-factor correlation matrix with two positive signs, one positive and one negative sign, or two negative signs in the decomposition is said to have two-positive-factor, one-positive-one-negative-factor, or two-negative-factor structure respectively.

**Nature of the Problem**

The motivation of this study can be well explained by one of the simplest cases of the General Linear Model (GLM), namely, the one-way design model, in which simultaneous confidence intervals for a set of parameters are of the primary interest. Depending on the underlying correlation structure of the estimators of the parameters, the evaluation of the exact critical value, needed in constructing the simultaneous confidence intervals of the desired confidence level, can be either easily computable or extremely difficult to compute.

Consider the model $Y_{jh} = \mu_j + e_{jh}$; $j = 1, \ldots, n$; $h = 1, \ldots, m_j$, where $\mu_1, \ldots, \mu_n$ are the treatment effects, and $e_{jh}$ are iid $N(0, \sigma^2)$ with an unknown $\sigma^2$. Let $N = \sum_{j=1}^n m_j$, $v = N - n$, $\bar{Y}_{j} = \sum_{h=1}^{m_j} Y_{jh} / m_j$, and $S^2 = \sum_{j=1}^{n} \sum_{h=1}^{m_j} (Y_{jh} - \bar{Y}_{j})^2 / v$. Let $\beta' = (\beta_1, \ldots, \beta_r)$ be a set of
parameters and let $\hat{\beta}$ be the Best Linear Unbiased Estimate (BLUE) of $\beta$. Let $\hat{\sigma}_{\hat{\beta}_j}$ be an estimator of the standard error of $\hat{\beta}_j$.

Suppose confidence intervals for the $\beta_j$'s are of interest. In particular, rectangular simultaneous confidence intervals on the $\beta_j$'s are desired. Then, we have to solve for the critical value $q$ in the following equation:

\begin{equation}
(1.0) \quad P\{ \beta_j \in \hat{\beta}_j \pm q\hat{\sigma}_{\hat{\beta}_j}; \ j = 1, \ldots, r \} = 1 - \alpha.
\end{equation}

**Goal of the Study**

The main purpose of this dissertation is to propose new techniques for computing the exact critical value $q$ and to apply the new techniques to obtain better confidence intervals in a variety of problems.

**Examples:**

1. Suppose $\beta_j = \mu_j$, $j = 1, \ldots, n$, that is, treatment effects are of interest. Then, the left hand side (LHS) of (1.0) equals $P\{ | X_j / U | \leq q; j = 1, \ldots, n \}$, where $X_j = (\bar{Y}_{j+} - \mu_j) / (\sigma\sqrt{\hat{m}_j})$ and $U = S/\sigma$ is such that $vU^2$ is distributed as a Chi-square with $v$ degrees of freedom. We note that $X = (X_1, \ldots, X_n)$ is multivariate normal with zero means and correlation matrix equal to the identity matrix. Hence, the LHS of (1.0) can be expressed as $\prod_{\delta=1}^{\infty} \prod_{j=1}^{n} [\Phi(qu)-\Phi(-qu)] \ d\Gamma(u)$, where $\Phi(\cdot)$ and $\Gamma(\cdot)$ are the cumulative distribution functions of a normal random variable and $U$ respectively. References are Tukey (1953) and Roy and Bose (1953). The critical value $q$ can be computed very fast and accurately in this case; fast enough to conduct interactive data analysis on a Macintosh microcomputer and accurate at least up to the 5th decimal place using 48 points Gauss-Legendre quadrature for the integration with respect to $\Gamma(\cdot)$. 
2. Suppose \( \beta_j = \mu_j - \mu_n, j = 1, \ldots, n-1 \), that is, treatments versus control effects are of interest. Then, the LHS of (1.0) equals \( P\{ |X_j / U| \leq q; j = 1, \ldots, n - 1 \} \), where \( X_j = \left[ (\bar{Y}_{j+} - \bar{Y}_{n+}) - (\mu_j - \mu_n) / (\sigma \sqrt{1/m_j + 1/m_n}) \right] \) and \( U \) is defined as that in Example 1.

Now, \( X = (X_1, \ldots, X_{n-1}) \) is multivariate normal with zero means and a correlation matrix \( \Sigma = D + \lambda \lambda' \), where \( \lambda = (\lambda_1, \ldots, \lambda_{n-1}) \) with \( \lambda_j = \sqrt{m_j/(m_j+m_n)} \) and \( D \) is diagonal with elements \( d_{jj} = 1 - \lambda_j^2 \). Thus, \( X \) is equal in distribution to \( (\sqrt{d_{11}} \ Z_1 - \lambda_1 \ Z_n, \ldots, \sqrt{d_{(n-1) \ (n-1)}} \ Z_{n-1} - \lambda_{n-1} \ Z_n) \), where \( Z_1, \ldots, Z_n \) are iid \( N(0,1) \) random variables.

Conditioning on the common random variables \( Z_0 \) and \( U \), the LHS of (1.0) becomes:

\[
\int_0^\infty \int_0^\infty \prod_{j=1}^{n-1} \left\{ \Phi[(qu+\lambda_j z)/\sqrt{d_{jj}}] - \Phi[(-qu+\lambda_j z)/\sqrt{d_{jj}}] \right\} \ d\Phi(z) \ d\Gamma(u),
\]

References are Dunnett (1955 and 1964). Here, the critical value \( q \) still can be evaluated very fast and accurately.

The expressions for the \( X_j \)'s in terms of a set of independently distributed standard normal random variables will be referred to as normal representations.

3. Suppose the \( \beta_j \)'s are the pairwise differences: \( \mu_i - \mu_h, i < h \), that is, all pairwise comparisons on the \( \mu_i \)'s are of interest. Hayter (1984) proved that

\[
P\{ |(\bar{Y}_{i+} - \bar{Y}_{h+}) - (\mu_i - \mu_h)| \leq q_{nv}^{\alpha} \sqrt{1/m_i + 1/m_h}, i < h \} \geq 1 - \alpha,
\]

where the critical value \( q_{nv}^{\alpha} \) is the upper \([100(\alpha)]^{th}\) percentage point of the studentized range distribution with parameters \( n \) and \( v \). That is, Tukey-Kramer's intervals are conservative. In the case where all the \( m_j \)'s are identical, the inequality in (1.1) becomes an equality, otherwise the inequality is a strict inequality except for \( n = 2 \). That is, \( q_{nv}^{\alpha} \) is the exact critical value for the equal sample size case. When sample sizes are unequal, the computation of the exact critical value \( q \) becomes extremely difficult because of the complicated region of integrations, introduced by the terms \( \sqrt{1/m_i + 1/m_h}, i < h \).
Consider the special case where \( n=3 \). Let
\[
X_1 = \frac{\bar{Y}_1 - \bar{Y}_2 + (\mu_1 - \mu_2)}{\sigma \sqrt{1/m_1 + 1/m_2}},
\]
\[
X_2 = \frac{\bar{Y}_1 - \bar{Y}_3 + (\mu_1 - \mu_3)}{\sigma \sqrt{1/m_1 + 1/m_3}},
\]
and
\[
X_3 = \frac{\bar{Y}_2 - \bar{Y}_3 + (\mu_2 - \mu_3)}{\sigma \sqrt{1/m_2 + 1/m_3}}.
\]
Let \( T_j = X_j \). Then, \((X_1, X_2, X_3)\) is singular multivariate normal with zero means and correlation matrix
\[
\Sigma = D - \lambda \lambda',
\]
where \( \lambda' = \left[ \sqrt{m_3/(m_1 + m_2)}, \sqrt{m_2/(m_1 + m_3)}, \sqrt{m_1/(m_2 + m_3)} \right] \), and
\( D \) is a diagonal matrix that makes \( \Sigma \) a correlation matrix. It will be shown that the LHS of (1.1) can be written as
\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{3} \left[ \Phi\left( qu - i \bar{\lambda}_j z / \sqrt{1 + \lambda_j^2} \right) - \Phi\left( -qu - i \bar{\lambda}_j z / \sqrt{1 + \lambda_j^2} \right) \right] \Phi(z) d\Phi(z) d\Gamma(u),
\]
distribution function with complex argument. The critical value is still attainable in this case; although the CPU time necessary to calculate it is much longer than that required in the procedures in the previous examples to achieve the same accuracy.

**Existing Results for the One-factor Structures**

For the one-positive-factor correlation structures, Dunnett and Sobel (1955) have derived an expression for the one-sided multivariate normal probabilities with correlations
\( \rho_{jh} = \rho \geq 0 \). Their results are also valid for \( \rho_{jh} = \lambda_j \lambda_h \) with \( 0 \leq \lambda_j < 1 \).

For the one-negative-factor correlation structures, Steck and Owen (1962) used the notion of normal distribution function with complex argument to show that Dunnett and Sobel's results hold for any \( \rho > -1/(n-1) \). Bland and Owen (1966) proved the validity of the same results in the singular case \( \rho = -1/(n-1) \). In deriving a procedure for detecting the studentized extreme deviation from the sample mean of a single random sample, Nair (1948) has tabulated values for the probability that the extreme deviation is less than a constant. We note that the underlying correlation matrix of Nair's procedure has a singular one-negative-factor structure.
Nelson (1981, 1988) proved similar results with \( \rho_{ji} = -\lambda_j \lambda_h \) (0 \( \leq \lambda_j \lambda_h \leq 1 \)) in both the singular and non-singular case. Nelson's (1988) proof of the singular case result relies on induction. However, he did not give an analytical proof for \( n = 2 \). Instead, he provided some numerical justifications. We shall show analytically the validity of this special case in Chapter II under the assumption that the elements of the factor pattern \( \lambda_j \)'s are of the same sign. We shall justify that the assumption comes naturally in our problem.

For general known correlations, Hsu (1989) proposed the use of the one-factor or two-factor structure in factor analysis to approximate the original correlation and computed the approximate coverage probabilities accordingly. He has shown extensive simulation results as well as applications to regression, one-way design with a covariate, and unbalanced two-way designs. The probabilities computed using the one-factor structure were found to be very close to the exact coverage probabilities.

Extended Results for the Two-factor Structures

In this thesis, we extended the results for the one-factor structures to two-factor cases. The extension follows the principle that introducing one more factor to the correlation structure results in adding another integral which can be independently integrated with respect to a standard normal density function.

If the correlation matrix has a two-positive-factor or one-positive-one-negative-factor structure, this extension principle guides a formal proof of the desired results with the aid of the normal representations. However, there exist no simple normal representations for the two-negative-factor case. To show the results in this case, we turn to analytic proofs, which are the major subjects of Chapter III.

Unlike the non-singular one-negative-factor case, where the result holds without imposing any other assumptions except the natural restriction that the determinant of the underlying correlation matrix \( \Sigma \), denoted by \( |\Sigma| \), is non-negative, our result for the non-
singular two-negative-factor case is valid only under the additional restrictions that
\( |D - \lambda'\lambda| \geq 0, |D - \gamma\gamma'| \geq 0 \) and not both \( |D - \lambda'\lambda| \) and \( |D - \gamma\gamma'| \) equal to zero. For the singular two-negative-factor case, our result is valid if the elements in both \( \lambda \) and \( \beta \) are all of the same sign.

Computations

The evaluation of the exact coverage probability in (1.0) involves the computation of a product integrand, which is a function of the normal distribution function with real argument in the one-positive-factor case and with complex argument in the one-negative-factor case. To avoid additional integrations for the product integrand, we need to have an accurate approximation for the normal distribution function with real or complex argument.

In the one-positive-factor case, the normal distribution function with real argument is bounded by one and, hence, it can be relatively easily approximated. For example, Cody (1969) developed a rational Chebyshev approximation for the error function with real argument, which can be transformed to evaluate the normal distribution function with real argument. The relative error of Cody's approximation is at most \( 6(10^{-19}) \). Thus, the exact coverage probability can be easily evaluated. Gaussian quadratures generally give quite satisfactory results for this case, unless many of the correlations are very large (> 0.9), in which case we have to use other numerical methods such as adaptive quadrature routines.

In the one-negative-factor case, the normal distribution function with complex argument can grow without a limit and, hence, its approximation is relatively difficult. Nelson (1982a, 1982b) has derived one approximation by adapting the rational Chebyshev approximation for the normal distribution function with real argument given by Abramowitz and Stegun (1972). However, the maximal absolute error of Nelson's approximation is only about \( 6(10^{-5}) \), for arguments with imaginary parts less than 1.5.
Note that the restriction is only on the imaginary part, because the approximation is less accurate at points with high magnitudes in the imaginary part.

Gautschi (1969, 1970) has derived an approximation for a complex function, known as the complex-valued error function, for all arguments in the first quadrant, which can be transformed to evaluate the normal distribution function with complex argument for all arguments in the entire complex plane. The maximal relative error of Gautschi's approximation is $10^{-10}$. All our computations involving the normal distribution function with complex argument are based on Gautschi's approximation.

Even though the complex-valued product integrand can be approximated accurately, some difficulties remain in the evaluation of the exact coverage probability. First, the complex-valued product integrand is unbounded. As a remedy, we absorb part or all of the standard normal density function from the outer integral to make the integrand bounded. Second, the real part of the complex-valued product integrand usually oscillates around zero with a low convergence rate, especially in the singular case and when $n = 2$. We should note that the imaginary part of the complex-valued product integrand is an odd function and that the inner integral is known to be zero for an odd function and, hence, the imaginary part of the complex-valued product integrand can be ignored. To achieve a relatively high accuracy in the computation of the exact coverage probability, we turned to an adaptive quadrature method proposed by Stoer and Bulirsch (1967), which not only is a good numerical method for oscillating integrands but also utilizes a rational extrapolation algorithm to accelerate the convergence. We note that the Stoer and Bulirsch quadrature is more accurate than the Gaussian quadrature in the one-positive-factor case when the correlations are very large.

The computation of the exact coverage probability in the two-factor case is not easy, especially in the singular two-negative-factor case. The first difficulty is with the inner double integration in which we are integrating with respect to two independent standard
normal density functions. If we iteratively apply an adaptive quadrature or use two adaptive quadrature routines for a double integration, it is not trivial to set up the two tolerance parameters involved so as to control the overall tolerance (see Fritsch, Kahaner and Lyness (1981)).

The other problem in the two-factor case is in the CPU time. Even on vector machines, such as the Cray X-MP, computing a triple integral is time-consuming. We note that the inner double integral is bounded after integrations. Hence, the outer integral, integrated with respect to the density function of the square-root of a Chi-square random variable, can be evaluated using the Gauss-Legendre quadrature by truncating the tail of the density which has negligible probability.

In order to handle the difficulties involved in solving the triple integral, we suggest that Hsu's (1989) one-factor approximation be used whenever it is appropriate. Alternatively, if the correlations are very small in absolute value (maybe < 0.2) and a lower accuracy is affordable (for example, computing the p-values in hypothesis testing), then Sidak's inequality can be applied to get a conservative bound on the coverage probability. Because of the nature of the restriction on the negative-factor correlation structures, Sidak's approximation is reasonably good provided that the number of parameters of interest is large (maybe > 7 in the negative equal correlation case).

To evaluate the critical value for the construction of the desired confidence intervals, we need a root-finding algorithm. The algorithm that we applied is a modified secant method, known as the Illinois method. One of the motivations of this method is to reduce the number of iterations needed for convergence by preventing the retention of an end point.
Applications

The areas of application of the results for one-factor or two-factor correlation structures are numerous:

1. One of the most important applications of the results for one-positive-factor structures is in the treatments-versus-control multiple comparison method by Dunnett (1955). Examples of this application can be found in the textbooks of multiple comparison procedures, for example, Hochberg and Tamhane (1987), or in the multiple comparison sections of experimental design textbooks, for example, Fleiss (1986).

2. The results for the one-negative-factor structures can be applied in the analysis of means procedures with unequal sample sizes where one is interested in the comparisons of the differences of the sample means from their grand mean (see Nelson (1988)).

3. When the detection of the studentized extreme deviate from the sample mean of a single random sample is desired, the underlying correlation has a one-negative-factor structure (see Nair (1948)).

4. In the one-way model based on rank statistics, suppose the design matrix includes the vector of ones and is of full rank. The results for the one-negative-factor structures can be applied to compute the p-value for the test that a set of un-redundant parameters is equal to fixed values (to be discussed in Chapter II).

5. Other applications of the results for one-negative-factor structures can be found in the area of categorical data analysis. For example, in the multinomial sampling model, if one is interested in comparing cell probabilities, then the underlying correlation matrix has a one-negative-factor structure. We shall discuss these applications in Chapter IV.

6. In the one-way design with one covariate and a common slope, if one is interested in comparing the treatments-versus-control effects, the underlying correlation matrix has a two-positive-factor structure (see Hsu (1989) and Chapter III).
7. In the discrete multinomial sampling model, if one is interested in comparing the treatments-versus-control effects (say, all cell probabilities minus the last cell probability), the underlying correlation matrix has a one-positive-one-negative-factor structure.

8. In the simple linear regression model, if one is interested in the detection of a single outlier based on some standardized residuals, the underlying correlations can have a two-negative-factor structure under certain restrictions on the standardization methods (Chapter III).

Outlines for the other chapters

In Chapter II, we shall discuss the results for the one-factor correlation structures. Rationale for the definition of the normal distribution function with complex argument will be given through the notion of the integration along a path. An analytic proof for the singular one-negative case with \( n = 2 \) will be presented. Numerical considerations for and applications of the one-factor case will be examined as well.

In Chapter III, the results for the two-factor correlation structures will be explored. Analytic proofs for the non-singular and singular two-negative-factor structures will be given under certain restrictions on the correlations. Numerical considerations for and applications of the two-factor case will also be examined.

In Chapter IV, we shall present the applications of the one-factor or two-factor results to three discrete sampling models, multinomial, binomial and Poisson sampling models. Data analyses as well as simulation studies will be given. We note that all the comparison procedures discussed in this chapter are based on large-sample normal approximations.

In Chapter V, we shall draw some conclusions and present some suggestions for future studies.
CHAPTER II

ONE-FACTOR STRUCTURES

Introduction

In simultaneous statistical inferences, one often has to evaluate one-sided or two-sided multivariate normal or multivariate t probabilities. For small dimensions (≤ 3), one can compute the probabilities to great accuracy for general correlations. For example, Plackett (1954) and Steck (1958) have introduced some reduction formulae for evaluating trivariate normal probabilities. Nevertheless, none of the existing numerical methods is totally satisfactory (in terms of speed and accuracy) for all correlations when dimensions are large (≥ 4).

One of the numerical methods for multi-dimensional integrations involves integrating iteratively one variable after another. However, the number of evaluations of the integrand grows exponentially with the dimensionality. Even on computers with vector processing and/or multitasking capability, computing the probability in this manner is too slow for interactive data analysis. As a result, the majority of the literature for evaluating one-sided or two-sided multivariate normal or multivariate t probabilities requires constraints on the correlations and/or limits of the integrals.

Let $X_1, \ldots, X_n$ be random variables with a multivariate normal distribution such that $E(X_j) = 0$, $E(X_j^2) = 1$ and $E(X_jX_k) = \rho_{jk}$, for all $j \neq k$. One of the most interesting special correlation structures from both a practical and theoretical point of view is $\rho_{jk} = \pm \alpha_j \alpha_k$, a one-factor structure. For evaluating the one-sided multivariate normal
probability, Dunnett and Sobel (1955) have derived an expression which is valid for all
\( \rho_{jk} = \alpha_j \alpha_k \) with \( 0 \leq \alpha_j < 1 \). Steck and Owen (1962) showed that Dunnett and Sobel's
results hold for any \( \rho > -1/(n-1) \). Bland and Owen (1966) proved the same results in the
singular case \( \rho = -1/(n-1) \). Nelson (1981,1988) proved similar results for \( \rho_{jk} = -\alpha_j \alpha_k \)
\( 0 \leq \alpha_j \alpha_k \leq 1 \) in non-singular case.

For simplicity, we shall concentrate on the one-sided multivariate normal model.
Similar results can be applied to the two-sided and the multivariate t model. Let us denote
\( F_n(h | \alpha) = P\{ X_j \leq h_j ; j = 1, \cdots, n \} \), where \( X=(X_1, \cdots, X_n) \) is multivariate normal with
zero means and correlation matrix \( R= D \pm \alpha \alpha' \), with \( \alpha'=(\alpha_1, \cdots, \alpha_n) \), \( D= \text{diag}(d_j^2 ; j= 1, \cdots, n) \), and \( d_j^2 = 1 \mp \alpha_j^2 \). The main result of the previous authors is
\[
F_n(h | \alpha) = \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left\{ \left[ h_j - c \alpha_j y \right] / d_j \right\} \, d\Phi(y), \tag{2.0}
\]
where \( c = 1 \) (or \( i \)) for positive (or negative) one-factor structure.

**Goal of the Chapter**

The main purpose of this chapter is to provide a proof of (2.0) for the special case
where \( n = 2 \), for which Nelson (1988) did not have an analytic proof, and to develop a
numerical algorithm for evaluating the exact coverage probability in (2.0). Some
applications of the results for one-factor structures will be discussed as well. We shall
provide a rationale for the definition of the normal distribution function with complex
argument through the notion of integration along a path. Let us start with some definitions
and preliminaries in the area of integration along a path.

**Definitions and Preliminaries**

The derivation of the expression for \( F_n(h | \alpha) \) in cases having negative-sign
correlation structures relies on the definition of the normal distribution function with
complex argument, introduced by Steck and Owen (1962). In order to define the normal distribution function with complex argument we must first review the concept of integral of a function along a path. Let us recall that a curve $\gamma$ with parameter interval $[a, b]$ ($-\infty < a < b < \infty$) is defined to be a continuous function such that $\gamma : [a, b] \rightarrow \mathbb{C}$, a complex set, with the initial point $\gamma(a)$ and the final point $\gamma(b)$. Define $\gamma^*$ to be the image set of $\gamma([a, b]) = \{\gamma(t) : t \in [a, b]\}$. Then $\gamma^*$ is a compact subset of $\mathbb{C}$, since $[a, b]$ is a compact interval on the real line.

**Definition 1:** A curve $\gamma : [a, b] \rightarrow \mathbb{C}$ is said to be smooth if it has a continuous derivative on $[a, b]$, in the sense that both the real and the imaginary parts are continuously differentiable.

**Definition 2:** A path is defined to be the join of finitely many smooth curves.

**Definition 3:** A path is said to be closed on $[a, b]$, if $\gamma(a) = \gamma(b)$.

Let $\gamma$ be a path on $[a, b]$. There exist points $a = t_0 < t_1 < \cdots < t_n = b$ such that $\gamma$ is a continuous differentiable function on $[t_k, t_{k+1}]$, although, at the $t_k$'s, $\gamma'$ might not exist. Let $f : \gamma^* \rightarrow \mathbb{C}$ be continuous. Then $(f \circ \gamma)'$ is piecewise continuous, and hence integrable on $[a, b]$.

**Definition 4:** The integral of a function $f$ along a path $\gamma$ is defined to be

$$\int_{\gamma} f(z) \, dz = \int_{-\infty}^{\infty} f(\gamma(t)) \, \gamma'(t) \, dt.$$  

Along with the notion of an integral along a path, there are several well-known facts that can be found in most of the textbooks of complex variable analysis, for example Priestley (1985).
Fact 1: Suppose \( \gamma \) is a path on \([a,b]\), and \( \hat{\gamma} \) is another path on \([\hat{a},\hat{b}]\) such that \( \hat{\gamma} = g \circ \varphi \),

where \( \varphi \) is a real function mapping the parameter interval \([\hat{a},\hat{b}]\) of \( \hat{\gamma} \) onto \([a,b]\) and having a positive continuous derivative. That is, for every \( t \in [\hat{a},\hat{b}] \), \( \dot{\gamma}(t) = \gamma'(\varphi(t))\varphi'(t) \).

Then \( \int_\gamma f(z) \, dz = \int_{\hat{\gamma}} f(z) \, dz \), where \( f : \gamma^* \rightarrow \mathbb{C} \) is continuous.

Implication: This fact implies that \( \int_\gamma f(z) \, dz \) depends on \( \gamma^* \) and the direction in which it is traced, not on any particular parameterization, under quite mild conditions.

Fact 2: Let \( \gamma \) be a join of paths \( \gamma_1, \ldots, \gamma_n \) and let \( f : \gamma^* \rightarrow \mathbb{C} \) be continuous.

Then, \( \int f(z) \, dz = \sum_{j=1}^n \int f(z) \, dz \).

Fact 3: If \( \gamma \) is closed in \( G \), a convex region, then \( \int f(z) \, dz = 0 \), where \( f \) is analytic in \( G \), i.e. \( f \) is continuously differentiable at every point of \( G \).

Definition 5: Let \( z = x + iy \) be a complex variable and let \( \gamma \) be a path defined by \( \gamma(t) = t + iy; \) where \(-\infty < t \leq x \) and \( y \) is a fixed constant. Clearly, \( \gamma'(t) = 1 \). One can define \( \Phi(z) \) as an integral along \( \gamma \) by the virtue of Fact 1. That is,

\[
\Phi(z) = (2\pi)^{1/2} \exp \left( \frac{y^2}{2} \right) \int_{-\infty}^{x} \exp(-it\,y) \exp(-t^2/2) \, dt.
\]

Justifications:

1. \( \exp(-z^2/2) = \exp(-\gamma^2(t)/2); \gamma^* \rightarrow \mathbb{C} \) is continuous for all \( t \) such that \(-\infty < t \leq x \).
2. \( \Phi(z) \), as defined, is analytic in the sense that both its real part and imaginary part are analytic.

Let \( g \) be an integrable function on the real line. A Fourier transform of \( g \), denoted by \( \hat{g} \), is defined as \( \hat{g}(s) = \int_{-\infty}^{\infty} g(x) \exp(-isx) \, dx \). If \( g \) is a legitimate probability density function,
then is the characteristic function of . The following lemma is a well-known result, which can be found in most textbooks of probability theory or complex variable analysis (for example, Chung (1974) p.148 or Priestley (1985) p.148). We shall use the result in Theorem 1 below to show that the probability of a lower quadrant can be written as a sum of a constant and the associated probability of a upper quadrant for correlation matrices with singular one-negative-factor structure and . The result will also be used to derive an expression for a two-dimensional Fourier transform Chapter III.

Lemma 1: Suppose , the standard normal probability density function. Then, .

Singular One-negative-factor Correlation structure with \( n=2 \)

In the case of the singular one-negative-factor structure, Nelson (1988) proved that equation (2.0) is valid for \( n=k+1 \) if the equation holds for \( n=k \). The following theorem provides an analytic proof of the same result for \( n=2 \), which together with Nelson's proof establishes by induction the important result that equation (2.0) is in fact valid for all singular one-negative-factor cases. Note that we have imposed a technical assumption that the elements of the factor pattern are all of the same sign. Without loss of generality, we assume that the elements are all positive. Later in this chapter, we shall justify that this assumption comes naturally for the problem of evaluating two-sided coverage probabilities, and, hence for constructing simultaneous confidence intervals.

Suppose \( X_1 \) and \( X_2 \) are bivariate normal with zero means and unit variances and with correlation equal to -1, then the probability of a lower quadrant can be written as a two-sided univariate probability as follows: \( P(X_1 \leq h_1^*, X_2 \leq h_2^*) = P(X_1 \leq h_1^*, -X_1 \leq h_2^*) = P(-h_2^* \leq X_1 \leq h_1^*), \) for real constants \( h_1^* \) and \( h_2^* \). We shall show that the two-sided univariate probability can be evaluated by the expression in (2.0) with \( n=2 \) and \( c=i \).
Theorem 1: Let $X_1$ and $X_2$ be bivariate normal with $E(X_1) = E(X_2) = 0$, $V(X_1) = V(X_2) = 1$, and $E(X_1X_2) = -1$. Suppose that $\alpha_1^*$ and $\alpha_2^*$ are positive and that $\alpha_1^* \alpha_2^* = 1$.

Then, for real constants $h_1^*$ and $h_2^*$,

\begin{equation}
(2.1) \int_0^\infty \Phi \left[ \frac{h_1^* - i \alpha_1^* y}{\sqrt{1 + \alpha_1^*}} \right] \Phi \left[ \frac{h_2^* - i \alpha_2^* y}{\sqrt{1 + \alpha_2^*}} \right] d\Phi(y) = P(-h_2^* \leq X_1 \leq h_1^*).
\end{equation}

For simplicity, let us denote $\alpha_j = \alpha_j^*/(1+\alpha_j^*)^{1/2}$ and $h_j = h_j^*/(1+\alpha_j^*)^{1/2}$, for $j = 1, 2$.

We have $\alpha_1^2 + \alpha_2^2 = 1$, $h_1^* = h_1^* \alpha_2$ and $h_2^* = h_2^* \alpha_1$. Further, let us denote

$I_0 = \int \int \int f_0(s,t,y) ds \, dt \, dy$, where $f_0(s,t,y) = \exp[-i(\alpha_1 t + \alpha_2 s)y] \exp(-s^2/2) \exp(-t^2/2)$.

The left hand side of equation (2.1), denoted by $R$, can be written as

\[ R = \frac{1}{2\pi \sqrt{\pi}} \int_0^\infty \int_0^\infty \cos \left[ (\alpha_1 t + \alpha_2 s) y \right] e^{-t^2/2} e^{-s^2/2} ds \, dt \, dy. \]

The proof of theorem 1 depends on where the point $(h_1^*, h_2^*)$ is located in the two dimensional plane of $X_1$ and $X_2$. We will present our proof in three cases, which correspond to the three regions: $h_1^* > -h_2^*$, $h_1^* < -h_2^*$ and $h_1^* = h_2^*$. We will show that $R = \Phi(h_1^*) - \Phi(-h_2^*)$, if $h_1^* > -h_2^*$ and $R = 0$, otherwise. We note $h_1^* = -h_2^*$ implies $\alpha_1 h_1 = -\alpha_2 h_2$.

Proof: Case 1: $h_1^* > -h_2^*$. For $j \in \{1, 2\}$, let us write

\begin{equation}
(2.2) \int_{-\infty}^{h_j} e^{i\alpha_j s y} e^{-s^2/2} ds = \int_{-\infty}^{\infty} e^{i\alpha_j s y} e^{-s^2/2} ds - \int_{h_j}^{\infty} e^{i\alpha_j s y} e^{-s^2/2} ds.
\end{equation}

Using this equation with $j = 2$, we have $R = T_1 + T_2$, where
\[ T_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1 e^{i\alpha_1 ty} e^{-\frac{t^2}{2}} e^{-\alpha_2^2 y^2/2} \, dt \, dy \]

\begin{equation}
(2.3) \quad T_1 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1 e^{-t^2/2} \, dt \, dy \int_{-\infty}^{\infty} e^{i\alpha_1 ty} e^{-\alpha_2^2 y^2/2} \, dy \, dt
\end{equation}

\[ T_2 = -\frac{1}{2\pi\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_1 e^{i(\alpha_1 t + \alpha_2 s) y} e^{-t^2/2} e^{-s^2/2} \, ds \, dt \, dy. \]

Note the first term on the right hand side of (2.2) is a Fourier transform of \( \exp(-s^2/2) \) and is equal to \( (2\pi)^{1/2} \exp(-\alpha_1^2 y^2/2) \). The inner integral of (2.3) is a Fourier transform of \( \exp(-\alpha_2^2 y^2/2) \) and is equal to \( (2\pi)^{1/2} \alpha_2^{-1} \exp[-\alpha_1^2 t^2/(2\alpha_2^2)] \). The exchange of the order of integration in \( T_1 \) is legitimate because the integrand is integrable.

Using equation (2.2) with \( j = 1 \), we have \( T_2 = T_3 + T_4 \), where \( T_3 = -\Phi(-h_2^*) \) and

\[ T_4 = \frac{1}{\pi\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos[(\alpha_1 t + \alpha_2 s)y] e^{-t^2/2} e^{-s^2/2} \, ds \, dt \, dy. \]

Thus, it remains to show that \( T_4 = 0 \).

The proof that \( T_4 = 0 \) is divided into cases according to which quadrant \( (h_1^*, h_2^*) \) is in. First, suppose that \( (h_1^*, h_2^*) \) is in the first quadrant excluding the origin. Then, we know at least one of \( h_1^* \) and \( h_2^* \) must be positive, say \( h_2^* > 0 \). The proof can be done by exchanging the roles of \( t \) and \( s \) and, hence, \( h_1^* \) and \( h_2^* \), if \( h_1^* > 0 \) and \( h_2^* \geq 0 \). Now, let us write

\begin{equation}
(2.4) \quad I(b) = \int_{h_1}^{b} \int_{h_2}^{b} \cos[(\alpha_1 t + \alpha_2 s)y] e^{-t^2/2} e^{-s^2/2} \, ds \, dt.
\end{equation}

When integrating with respect to \( y \) first, we obtain

\[ I(b) = \int_{h_1}^{b} e^{-t^2/2} \int_{h_2}^{b} \frac{e^{-s^2/2}}{\alpha_1 t + \alpha_2 s} \sin[(\alpha_1 t + \alpha_2 s) b] \, ds \, dt. \]
Since \( \exp(-s^2/2) / (\alpha_1 t + \alpha_2 s) \) is a decreasing function of \( s \) on \([h_2, a]\) for any \( a > h_2 \) and any fixed \( t \geq 0 \), then by the second mean value theorem, there exists a point \( s_0 \) in \([h_2, a]\) such that

\[
\int_{h_2}^{a} \frac{e^{-s^2/2}}{\alpha_1 t + \alpha_2 s} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds = \frac{e^{-h_2^2/2}}{\alpha_1 t + \alpha_2 h_2} \int_{h_2}^{s_0} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds + \frac{e^{-a^2/2}}{\alpha_1 t + \alpha_2 s_0} \int_{s_0}^{a} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds.
\]

Now, with a change of variable by letting \( w(s) = (\alpha_1 t + \alpha_2 s) b \), we see that

\[
\int_{h_2}^{s_0} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds \leq \frac{1}{\alpha_2 b} \left\{ \| \cos[w(s_0)] \| + \| \cos[w(h_2)] \| \right\} \leq \frac{2}{\alpha_2 b}.
\]

Similarly, we can have

\[
\int_{s_0}^{a} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds \leq \frac{2}{\alpha_2 b}. \quad \text{Thus,}
\]

\[
\int_{h_2}^{a} \frac{e^{-s^2/2}}{\alpha_1 t + \alpha_2 s} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds \leq \frac{e^{-h_2^2/2}}{\alpha_1 t + \alpha_2 h_2} \cdot \frac{2}{\alpha_2 b} + \frac{e^{-a^2/2}}{\alpha_1 t + \alpha_2 a} \cdot \frac{2}{\alpha_2 b}.
\]

Letting \( a \to \infty \), we have

\[
\int_{h_2}^{\infty} \frac{e^{-s^2/2}}{\alpha_1 t + \alpha_2 s} \sin \left( (\alpha_1 t + \alpha_2 s) b \right) ds \leq \frac{e^{-h_2^2/2}}{\alpha_1 t + \alpha_2 h_2} \cdot \frac{2}{\alpha_2 b}.
\]

Hence,

\[
|I(b)| \leq \frac{2}{\alpha_2 b} e^{-h_2^2/2} \int_{h_1}^{\infty} \frac{e^{-t^2/2}}{\alpha_1 t + \alpha_2 h_2} dt.
\]

But,

\[
\int_{h_1}^{\infty} \frac{e^{-t^2/2}}{\alpha_1 t + \alpha_2 h_2} dt = \frac{1}{\alpha_1} \int_{h_1/\sqrt{x}}^{\infty} \frac{e^{-t^2}}{t + x} dt < \infty, \text{ where } x = \frac{\alpha_2 h_2}{\alpha_1 \sqrt{2}} > 0
\]

(see Goodwin and Staton (1948), and Abramowitz and Stegun (1965) equation 7.4.10).

We have \( |I(b)| \to 0 \), as \( b \to \infty \).

Therefore, \( T_4 = \frac{1}{\pi \sqrt{2} \pi} \lim_{b \to \infty} I(b) = 0 \).
Suppose that \((h^*_1, h^*_2)\) is in the fourth quadrant with \(h^*_2 < 0\) and \(h^*_1 > -h^*_2\). (The proof is similar if \((h^*_1, h^*_2)\) is in the second quadrant.) Now \(T_1, T_2\) and \(T_3\) are as before and

\[
T_4 = \frac{1}{\pi \sqrt{2\pi}} \int_0^\infty \int_0^\infty \cos \left[ (\alpha_1 t + \alpha_2 s)y \right] e^{-t^2/2} e^{-s^2/2} \, ds \, dt \, dy
\]

\[
+ \frac{1}{\pi \sqrt{2\pi}} \int_0^\infty \int_{h_1}^{h_2} \cos \left[ (\alpha_1 t + \alpha_2 s)y \right] e^{-t^2/2} e^{-s^2/2} \, ds \, dt \, dy.
\]

The first term of this expression has been proved to be zero. Thus, it remains to prove that the second term is also equal to zero. Let \(I(b)\) be as defined in (2.4) except that the bounds of \(s\) are now from \(h_2\) to 0. After exchanging the order of integration, we have

\[
I(b) = \int_0^{h_2} e^{-s^2/2} \int_0^\infty \frac{e^{-t^2/2}}{\alpha_1 t + \alpha_2 s} \sin \left[ (\alpha_1 t + \alpha_2 s) b \right] \, dt \, ds.
\]

Since \(\alpha_1 h_1 + \alpha_2 h_2 > 0\), \(\exp(-t^2/2) / (\alpha_1 t + \alpha_2 s)\) is a decreasing function of \(t\) on \([h_1, a]\) for any \(a > h_1\) and any fixed \(s\) in \([h_2, 0]\). By the second mean value theorem, in the same manner as before, we have

\[
\left| \int_0^\infty \frac{e^{-t^2/2}}{\alpha_1 t + \alpha_2 s} \sin \left[ (\alpha_1 t + \alpha_2 s) b \right] \, dt \right| \leq \frac{e^{-h_2^2/2}}{\alpha_1 h_1 + \alpha_2 s} \cdot \frac{2}{\alpha_1 b}.
\]

Hence, \(|I(b)| \leq \frac{2}{\alpha_1 b} e^{-h_2^2/2} \int_0^{h_2} \frac{e^{-s^2/2}}{\alpha_2 s + \alpha_1 h_1} \, ds \to 0\), as \(b \to \infty\).

Thus, \(T_4 = 0\).

Case 2: \(h^*_1 < -h^*_2\). The proof that \(R = 0\), in this case, is similar to the proof that \(T_4 = 0\) in Case 1 and hence is omitted.

Case 3: \(h^*_1 = -h^*_2\). Let us first discuss the case \(h^*_1 = h^*_2 = 0\). Let \(I_0\) and \(f_0(s,t,y)\) be as before and, for a fixed constant \(\partial > 0\), let us introduce \(I_\partial\) and \(f_\partial(s,t,y)\), where

\[
f_\partial(s,t,y) = \exp[-i(\alpha_1 t + \alpha_2 s - \partial^{1/4})y] \exp(-\partial y^2) \exp(-s^2/2) \exp(-t^2/2)\] and
Our proof consists of showing that \( \lim_{\delta \to 0} I_0 = I_0 \) and \( \lim_{\delta \to 0} I_\delta = 0 \), where, in this case,

\[
I_0 = 2 \int_0^\infty \int_0^\infty \int_0^\infty f_\delta(s,t,y) \, ds \, dt \, dy.
\]

By integrating with respect to \( y \) first, in order to use the known result in Lemma 1, we obtain

\[
I_\delta = \sqrt{\pi} \int_0^\infty \int_0^\infty \frac{1}{\sqrt{\gamma}} \int_0^\infty \int_0^\infty e^{-(\alpha_1 t + \alpha_2 s - \gamma^{1/4})^2/(4\gamma)} \, e^{-s^2/2} \, e^{-t^2/2} \, ds \, dt.
\]

So,

\[
|I_\delta| \leq \frac{1}{\sqrt{\pi}} \int_0^\infty \int_0^\infty e^{-s^2/2} \, e^{-t^2/2} \, ds \, dt \to 0, \quad \gamma \to 0.
\]

Let \( I(b) \) now be as defined in (2.4) and take \( h_1 = h_2 = 0 \). We have \( I_0 = \lim_{b \to \infty} I(b) \).

But,

\[
|I(b)| \leq \int_0^\infty |\sin(\alpha_1 t b)| \, e^{-t^2/2} \, e^{-s^2/2} \, ds \, dt \left( \int_0^\infty \frac{e^{-s^2/2}}{\alpha_1 t + \alpha_2 s} \, \cos(\alpha_2 s b) \, ds \right) dt
\]

\[
+ \int_0^\infty |\sin(\alpha_2 s b)| \, e^{-s^2/2} \, ds \, dt \left( \int_0^\infty \frac{e^{-t^2/2}}{\alpha_1 t + \alpha_2 s} \, \cos(\alpha_1 t b) \, dt \right) ds.
\]

Applying the second mean value theorem, we have

\[
\left| \int_0^\infty \frac{e^{-s^2/2}}{\alpha_1 t + \alpha_2 s} \, \cos(\alpha_2 s b) \, ds \right| \leq \frac{1}{\gamma t} \frac{1}{\alpha_1 b},
\]

and

\[
\left| \int_0^\infty \frac{e^{-t^2/2}}{\alpha_1 t + \alpha_2 s} \, \cos(\alpha_1 t b) \, dt \right| \leq \frac{1}{\gamma s} \frac{1}{\alpha_2 b}.
\]

Using the fact that \( \left| \frac{\sin(x)}{x} \right| \leq 1 \), we have \( I(b) \) \( \leq \sqrt{2\pi} \frac{1}{\alpha_2 + 1/\alpha_1} \).

Thus, \( I_0 \) is finite and so is \( \int_0^\infty \int_0^\infty |f_\delta(s,t,y)| \, ds \, dt \, dy \). Lebesgue's dominated convergence theorem implies that \( \lim_{\delta \to 0} I_\delta = I_0 \). Therefore, \( I_0 = 0 \).
Now, suppose that $h_1^* = -h_2^*$ and $h_1^* > 0$. (The case that $h_1^* = -h_2^*$ and $h_2^* > 0$ can be proved similarly.) Let $f_2(s,t,y)$ and $f_0(s,t,y)$ be as defined in the case that $h_1^* = h_2^* = 0$.

Then, because that both $\alpha_1$ and $\alpha_2$ are positive and that $\alpha_1 h_1 + \alpha_2 h_2 = 0$, we can write

\[ R = \frac{1}{\pi \sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(s,t,y) \, ds \, dy = \frac{1}{\pi \sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_2(s,t,y) \, ds \, dy. \]

Let $I_2$ and $I_0$ be as defined before except that the bounds of $s$ (and $t$) should be from $h_2$ (and $h_1$) to $\infty$ respectively. As before, we may show that

\[ |I_2| \leq \sqrt{\frac{\pi}{2}} e^{\frac{1}{2}} \int_{h_1}^{\infty} \int_{h_2}^{\infty} e^{-s^2/2} e^{-t^2/2} \, ds \, dt \longrightarrow 0, \quad \text{as} \quad \partial \longrightarrow 0. \]

Let $I(b)$ be as defined in (2.4). Then, because $\alpha_1 h_1 + \alpha_2 h_2 = 0$,

\[ I(b) = \int_{\infty}^{\infty} \int_{\infty}^{\infty} e^{-(t+h_1)^2/2-(s+h_2)^2/2} \sin \left(\frac{(\alpha_1 t + \alpha_2 s) b}{\alpha_1 t + \alpha_2 s}\right) \, ds \, dt. \]

We may write $|I(b)| \leq T_5 + T_6$, where

\[ T_5 = \int_{0}^{\infty} \sin(\alpha_2 s b) \left| e^{-(s+h_2)^2/2} \int_{0}^{\infty} e^{-(t+h_1)^2/2} \cos(\alpha_1 t b) \, dt \right| \, ds, \quad \text{and} \]

\[ T_6 = \int_{0}^{\infty} \cos(\alpha_2 s b) \left| e^{-(s+h_2)^2/2} \int_{0}^{\infty} e^{-(t+h_1)^2/2} \sin(\alpha_1 t b) \, dt \right| \, ds. \]

Since $h_1$ is positive, using the second mean value theorem, we have

\[ \left| \int_{0}^{\infty} \frac{e^{-(t+h_1)^2/2}}{\alpha_1 t + \alpha_2 s} \cos(\alpha_1 t b) \, dt \right| \leq \frac{e^{-h_1^2/2}}{\alpha_1 \alpha_2 b s}. \]

Using the fact that $\left| \frac{\sin(x)}{x} \right| \leq 1$, we have $T_5 \leq \frac{\sqrt{2\pi}}{\alpha_1} e^{-h_1^2/2}$.

Now, by definition,

\[ \int_{0}^{\infty} \frac{e^{-(t+h_1)^2/2}}{\alpha_1 t + \alpha_2 s} \sin(\alpha_1 t b) \, dt = \lim_{a \to \infty} \int_{0}^{a} \frac{e^{-(t+h_1)^2/2}}{\alpha_1 t + \alpha_2 s} \sin(\alpha_1 t b) \, dt. \]

Using the second mean value theorem, we have

\[ \left| \int_{0}^{\infty} \frac{e^{-(t+h_1)^2/2}}{\alpha_1 t + \alpha_2 s} \sin(\alpha_1 t b) \, dt \right| \leq \frac{e^{-h_1^2/2}}{\alpha_1 \alpha_2 b s}. \]
Thus,

\[ T_6 \leq \frac{e^{-h^2/2}}{\alpha_1 b} \left| \int_0^\infty \frac{e^{-(s+h_2)^2/2}}{\alpha_2 s} \cos(\alpha_2 sb) ds \right|. \]

But,

\[
\int_0^\infty \frac{e^{-(s+h_2)^2/2}}{\alpha_2 s} \cos(\alpha_2 sb) ds = \int_0^\infty \frac{e^{-(s+h_2)^2/2}}{\alpha_2^2 bs} d\sin(\alpha_2 sb)
\]

\[ = -\frac{e^{-(h_2)^2/2}}{\alpha_2} + \frac{1}{\alpha_2} \int_0^\infty \sin(\alpha_2 sb) (s+h_2) e^{-(s+h_2)^2/2} ds, \text{ using the fact that } \lim_{x \to 0} \frac{\sin(x)}{x} = 1. \]

We have,

\[
\left| \int_0^\infty \frac{e^{-(s+h_2)^2/2}}{\alpha_2} \cos(\alpha_2 sb) ds \right| \leq \frac{e^{-h_2^2/2}}{\alpha_2} + \frac{1}{\alpha_2} \int_0^\infty |s+h_2| e^{-(s+h_2)^2/2} ds < \infty.
\]

This implies that \( T_6 \to 0 \), as \( b \to \infty \).

Hence, for any \( b > 0 \), \( |I(b)| \leq \frac{\sqrt{2\pi}}{\alpha_1} e^{-h^2/2} \), a fixed constant.

Thus, \( |I_0| \) is finite and so is \( \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(s,t,y)| ds dt dy. \)

Lebesgue's dominated convergence theorem implies that \( \lim_{\theta \to 0} I_0 = I_0 \).

Therefore, \( I_0 = 0. \)

QED

Once we established that equation (2.0) is valid for all correlation matrices with one-factor structure, two useful corollaries follow easily. The first corollary extends the result in (2.0), which works for the one-sided multivariate normal model, to the one-sided multivariate \( t \) case. The second corollary in turn extends the result in the first corollary to the two-sided multivariate \( t \) case.
Corollary 1: Let $X = (X_1, \cdots, X_n)$ be multivariate normal with zero means and a one-factor correlation matrix $\Sigma = D \pm \alpha \alpha'$. Further, suppose $S$ is independent of the $X_j$'s and $\nu S^2$ has a Chi-square distribution with $\nu$ degrees of freedom. For all $j = 1, \cdots, n$, define $T_j = X_j / S$. Then, $P(T_1 \leq h_1, \cdots, T_n \leq h_n)$ can be expressed as:

$$
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left( \frac{sh_j - c \alpha_j y}{\sqrt{1 - c^2 \alpha_j^2}} \right) d\Phi(y) d\Gamma(s),
$$

where $\Phi(\cdot)$ is the distribution function of a standard normal random variable, $\Gamma(\cdot)$ is the distribution function of $S$, and $c = 1$ (or $i$) for positive (or negative) one-factor structure.

Proof: Conditioning on $S$, $P(T_1 \leq h_1, \cdots, T_n \leq h_n)$ becomes

$$
\int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left( \frac{sh_j - c \alpha_j y}{\sqrt{1 - c^2 \alpha_j^2}} \right) d\Phi(y) d\Gamma(s),
$$

The result follows by applying the main result in (2.0) to the integrand of the last equation.

QED

Corollary 2: Let $X$, $S$ and $T$ be as defined in Corollary 1. Then,

$$
P(T_1 \leq h_1, \cdots, T_n \leq h_n) = \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left( \frac{sh_j - c \alpha_j y}{\sqrt{1 - c^2 \alpha_j^2}} \right) d\Phi(y) d\Gamma(s),
$$

where $\Phi$, $\Gamma$ and $c$ are as defined in Corollary 1.

The Proof is trivial and is omitted.
Sign invariance in evaluating two-sided coverage probabilities

The assumption necessary for the proof in the case of a one-negative-factor structure with \( n = 2 \) ruled out cases where the components of the one-factor structure might not be all of the same sign. However, for the problem of evaluating two-sided coverage probabilities, the assumption comes naturally because:

1. the collection of all correlation matrices with one-positive-factor structure is disjoint from those correlation matrices with one-negative-factor structure, and

2. the problem is invariant with respect to \( G = \{ D : D = \text{diag}[\pm 1, \pm 1, \ldots, \pm 1], D \Sigma D \geq 0 \} \)

for a correlation matrix \( \Sigma \) with one-factor structure.

A detailed justification of these properties follows.

Let \( S^+ = \{ \Sigma \geq 0 : \Sigma = D + \alpha \alpha', D \text{ is a diagonal matrix, and } \alpha \text{ is a vector of constants such that not all its components are zero} \} \), and let \( S^- = \{ \Sigma \geq 0 : \Sigma = D - \alpha \alpha', D \text{ is a diagonal matrix and } \alpha \text{ is a vector of constants such that not all its components are zero} \} \). We should note that \( \Sigma \) is diagonal and, hence, does not have a one-factor structure if all components of \( \alpha \) are zero. In the next proposition, we shall show that \( S^+ \) and \( S^- \) are disjoint.

**Proposition 1:** The events \( S^+ \) and \( S^- \) are disjoint.

**Proof:** The case where \( n = 2 \) is trivial. In cases where \( n \geq 3 \), we shall show that there is a contradiction to the restriction on the components of the one-factor structure if \( S^+ \) and \( S^- \) are not disjoint.

Suppose there exists a correlation matrix \( \Sigma \) belongs to both \( S^+ \) and \( S^- \). That is, \( \Sigma \) can be written as \( \Sigma = D + \alpha \alpha' \), and \( \Sigma = D^* - \alpha^* \alpha^{*'} \). Let \( 1 \leq i < j < k \leq n \), we have \( \alpha_i \alpha_j = -\alpha^*_i \alpha^*_j \), \( \alpha_i \alpha_k = -\alpha^*_i \alpha^*_k \), and \( \alpha_j \alpha_k = -\alpha^*_j \alpha^*_k \). Thus, \( \alpha_i^2 \alpha_j \alpha_k = \alpha_i^2 \alpha_j^* \alpha_k^* = -\alpha_i^2 \alpha_j \alpha_k \).

Hence, \( \alpha_i^2 = -\alpha_i^{*2} \). Similarly, we have \( \alpha_j^2 = -\alpha_j^{*2} \) and \( \alpha_k^2 = -\alpha_k^{*2} \). Since \( i, j \) and \( k \) are arbitrary, we have \( \alpha_j^2 = -\alpha_j^{*2} \), \( j = 1, \cdots, n \), which contradicts the restriction that not all the components of \( \alpha \) and \( \alpha^* \) are zero.

QED
The problem of computing \( P\{|X_j| \leq q, j=1,\cdots, n\} \), where \((X_1,\cdots, X_1)\) is multivariate normal with zero means and a correlation matrix \( \Sigma \) of one-factor structure, is invariant for \( \Sigma \) with respect to \( G \) in the sense that the coverage probability evaluated with respect to \( \Sigma \) is identical to that evaluated with respect to \( D\Sigma D \), where \( D \) belongs to \( G \). The reason is that \( P\{|X_j| \leq q, j=1,\cdots, n\} = P\{|Y_j| \leq q, j=1,\cdots, n\} \), where \((Y_1,\cdots, Y_1)' = D(X_1,\cdots, X_1)'\).

The invariance of the problem together with Proposition 1 implies that we can assume without loss of generality that all the components of the one-factor structure are positive. Hence, the assumption needed in the proof of the singular one-negative-factor structure with \( n=2 \) comes naturally. This implication is also useful in the proof of the following proposition, which says that any correlation matrix \( \Sigma \) can be decomposed into a one-factor structure, if the dimension \( n \) is 3 or less.

**Proposition 2:** For the problem of computing two-sided coverage probabilities, any known correlation matrix \( \Sigma \) can be decomposed into a one-factor structure, provided that \( n \leq 3 \).

**Proof:** The case where \( n=2 \) is trivial. In the case that \( n=3 \), let the known upper off-diagonal elements of \( \Sigma \) be \( \rho_{12}, \rho_{13}, \text{ and } \rho_{23} \). Since the problem is invariant and the events \( S^+ \) and \( S^- \) are disjoint, we can assume without loss of generality that all the \( \rho_{jh} \)'s are of the same sign. We shall show that \( \Sigma \) can be decomposed into a one-factor structure say, \( \Sigma = D \pm \alpha \alpha' \) with \( \alpha' = (\alpha_1, \alpha_2, \alpha_3) \), \( D = \text{diag}(d_1^2, d_2^2, d_3^2) \), and \( d_j^2 = 1 - \alpha_j^2 \), by actually deriving solutions for the \( \alpha_j \)'s in terms of the \( \rho_{jh} \)'s.

In the case that all \( \rho_{jh} \)'s are positive, let \( \theta_j = \log(\alpha_j) \), \( j=1, 2, 3 \), and let \( r_{jh} = \log(\rho_{jh}) \), \( 1 \leq j < h \leq 3 \). Letting \( \rho_{12} = \alpha_1 \alpha_2, \rho_{13} = \alpha_1 \alpha_3, \text{ and } \rho_{23} = \alpha_2 \alpha_3 \), we have \( \theta_1 + \theta_2 = r_{12}, \theta_1 + \theta_3 = r_{13} \text{ and } \theta_2 + \theta_3 = r_{23} \), which forms a linear system: \( A\theta = r \), where \( A = (\alpha_1, \alpha_2, \alpha_3), \theta' = (\theta_1, \theta_2, \theta_3), r' = (r_{12}, r_{13}, r_{23}), a_1' = (1, 1, 0), a_2' = (1, 0, 1), \text{ and } a_3' = (0, 1, 1) \). Since
a_1, a_2, a_3 are linearly independent, the linear system has a set of unique solutions for the 
θ_j's and, hence, the unique solutions for the α_j's are exp(θ_j), j = 1, 2, 3.

In the case that all ρ_jh's are negative, we can let -ρ_{12} = α_1 α_2, -ρ_{13} = α_1 α_3, -ρ_{23} = 
α_2 α_3, and derive the set of unique solutions for the θ_j's. The corresponding unique 
solution for α_j is - exp(θ_j), j = 1, 2, 3.

QED

Computations

The evaluation of the probability statement P\{X_j ≤ h_j; j = 1, ⋯, n\}, where X =
(X_1, ⋯, X_n) is multivariate normal with zero means and a correlation matrix with one-
negative-factor structure, involves the computation of a product integrand, the product of
n normal distribution functions with complex argument as shown in equation 2.0. If we
can find an algorithm that is efficient and accurate enough to evaluate the product
integrand, then F_n can be regarded as a single integral integrated with respect to the
standard normal density. Unlike the case of the positive-sign correlation structure, where
the product integrand is bounded by 1, the product integrand can grow without a limit in
the negative-sign case. As a remedy, we may absorb a part or all of the standard normal
density into the product integrand to make the integrand bounded. Even so, the behavior
of the scaled product integrand is still different for different correlation structures and/or
different bounds, h*'s. As a result, we have considered using different integration
methods for different correlation structures.

In general, we found the Gauss-Legendre quadrature to be efficient for the
computation of F_n with non-singular correlation structure. In the singular case, we have
turned to adaptive quadratures such as the Romberg integration method (see Bauer,
Rutishauser, and Stiefel (1963)) or the Stoer and Bulirsch (1967) integration method.
With equal accuracy, the latter is about three times faster than the former on the average in the application to our problem.

For the computation of a single normal distribution function with complex argument, \( \Phi(z) \), where \( z=x + iy \), if we can compute \( \Phi(z) \) in the first quadrant, then we can compute \( \Phi(z) \) for any \( z \) in the entire extended complex plane. To see this, let us write \( z = r(a + ib) \), where \( r, a, \) and \( b \) are real numbers such that \( r \) is non-negative, and that not both \( a \) and \( b \) are zero. For example, we can write \( z \) in its polar form such that \( r=|z|, a=\cos(\theta) \) and \( b=\sin(\theta) \), where \( \theta \) is the argument function of \( z \), that is, \( \theta=\arctan(y/x) \). Then, \( a^2 + b^2 = 1 \).

Define \( c=a+ib \) and \( c^2=\alpha+i\beta \), that is \( \alpha=a^2-b^2 \) and \( \beta=2ab \). Then,

\[
\Phi(z) = (2\pi)^{-1/2} \left[ a \int_{-\infty}^{r} \cos(\alpha s^2) \exp(-\alpha s^2/2) ds + b \int_{-\infty}^{r} \sin(\alpha s^2) \exp(-\alpha s^2/2) ds \right] + i \left[ b \int_{-\infty}^{r} \cos(\alpha s^2) \exp(-\alpha s^2/2) ds - a \int_{-\infty}^{r} \sin(\alpha s^2) \exp(-\alpha s^2/2) ds \right]
\]

We see that the real part of \( \Phi(z) \) is odd in \( a \) but even in \( b \) and that the imaginary part is odd in \( b \) but even in \( a \). Therefore, we can restrict ourselves to the computation of \( \Phi(z) \) in the first quadrant, without loss of generality. To compute \( \Phi(z) \), for \( z \) in the other quadrants, we can use the following properties:

1. \( \Phi(\bar{z}) = \text{Real} \, \Phi(z) - i \, \text{Imag} \, \Phi(z) \);
2. \( \Phi(-z) = 1 - \text{Real} \, \Phi(z) + i \, \text{Imag} \, \Phi(z) \), and
3. \( \Phi(-z) = 1 - \text{Real} \, \Phi(z) - i \, \text{Imag} \, \Phi(z) \).

The notation \( \bar{z} \) represents the conjugate complement of \( z \); \( \text{Real} \, \Phi(z) \) is the real part of \( \Phi(z) \) and \( \text{Imag} \, \Phi(z) \) is the imaginary part of \( \Phi(z) \).

Let us introduce a closely related complex function called the complementary error function with complex argument, denoted by \( \text{Erfc}(z) \). As an intermediate, \( \text{Erfc} \) plays an important role connecting \( \Phi(z) \) with an exponentially scaled complex error function, denoted by \( W(z) \). For any \( z \) in the complex plane, \( \text{Erfc}(z) \) is defined as
2(\pi)^{-1/2} \int_0^\infty \exp(-t^2) \, dt and W(z) is defined as \exp(- z^2) \text{Erfc}(-iz). We note that "Error Function with Complex Argument" usually refers to the exponentially scaled complex error function.

It can be shown that \( \Phi(z) = 1 - \text{Erfc}(z/\sqrt{2})/2 \), for all \( z \) that lie in the closed right hand half of the complex plane, where the real part of \( z \) is greater than or equal to zero. First, the result holds for all non-negative real \( z \), by a simple change of variable. Hence the result holds for all \( z \) in the right hand half plane, by analytic continuation.(see Chapter 2 of Lebedev (1972) and Markushevich (1984)) Therefore, a computation algorithm, valid for \( \text{Erfc}(z) \) in the right hand half plane, should be valid for \( \Phi(z) \) in the first quadrant. We note that the relationship between \( \Phi \) and \( W \) is \( \Phi(z) = 1 - \exp(- z^2/2)W(iz/\sqrt{2})/2, \) for all \( z \) in the first quadrant.

**FIGURE 2.1:** A 3-D Plot of the Real Part of \( \text{Erfc}(x+iy) \)

\( x = 0 \) to 7 by 0.2 and \( y = 0 \) to 7 by 0.2
FIGURE 2.2: A 3-D Plot of the Imaginary Part of \text{Erfc}(x+iy)

As a function of \(x\) and \(y\), if \(x\) is small then \(\text{Erfc}(z)\) can increase without a limit as \(y\) increases. For example, the absolute values of both the real and imaginary parts are about \(10^{19}\) when \(z = 7i\) (see Figure 2.1 and 2.2). Thus, calculations are more reliable if we consider the transformed function \(W(z)\), which remains bounded (by 1) for all \(z\) in the first quadrant.

In the integral representation, \(W(z)\) can be written as \((i/\pi) \int_{-\infty}^{\infty} \exp(-t^2)/(z-t)dt\), for all \(z\) in the open left hand half plane, where the imaginary part of \(z\) is greater than zero. Notice that the integral is of the Gauss-Hermite integral form and is analytic for all \(z\) excluding the real line. Hence it can be expanded as a descending power series of \(z\). Gautschi (1970) has derived an efficient approximation of \(W(z)\) for any \(z\) in the first quadrant. The IMSL library has adapted Gautschi's algorithm to have a controlled maximal absolute
error of 10^{-10}. All of our computational results are based on the subroutine Cerfe/Zerfe for exponentially scaled complex complement error function in IMSL.

As an example, we consider the computation of $F_2$ with $h_1 = h_2 = d^*$, a non-negative constant. The reason why we choose $n=2$ is twofold. First, there are numerous methods available for the calculation of bivariate normal distribution (for example, Donnelly (1973)) and, hence, we can check the accuracy of our proposed method against known methods. Second, in the singular one-negative-factor case, the computation of $F_2$ is less accurate than the computation of $F_n$ with $n \geq 3$ because of the nature of the restriction on the negative-factor correlation structures. Hence, we must check whether our proposed method works for $n=2$. Let $\mathbb{E}(X_1X_2) = \rho$, where $-1 \leq \rho \leq 0$. In the special case that $\alpha_1 = \alpha_2 = \alpha = [\rho/(1-\rho)]^{1/2}$ and $d = d^*/(1-\rho)^{1/2}$, $F_2$ becomes

$$\int_0^{\sqrt{2}/\pi} \text{Real} \left\{ \left( \Phi(-i\alpha y + d) \right)^2 \exp(-y^2/2) \right\} dy.$$

As an example in the non-singular case where $\rho \neq -1$, we take $\rho = -0.5$ so that $\alpha = 1/\sqrt{3}$ and $d = d^*/\sqrt{3}$. Then $F_2 = \int_0^{\sqrt{2}/\pi} g(y) dy$, where

$$g(y) = \left[ \exp(-y^2/6) - \exp(-d^*/2) \right] \left[ \left( \cos(\sqrt{2} d^* y/3) + i \sin(\sqrt{2} d^* y/3) \right) W(iz) \right]^2 \exp(-y^2/6).$$

In the singular case where $\rho = -1$, we have $\alpha_1 = \alpha_2 = \alpha = 1/\sqrt{2}$ and $d = d^*/\sqrt{2}$. The integrand $g(y)$ becomes $\left[ \exp(-y^2/4) \exp(-d^*/4) \right] \left[ \left( \cos(d^* y/2) + i \sin(d^* y/2) \right) W(iz) \right]^2$.

Frequently (not for the case that $d^*=0$ and $\rho=-1$), the integrand is oscillating around zero and tends toward zero rapidly in an exponential fashion. (see Figures 2.3 through 2.6) The magnitude of the tail of the integrand is of the order of $\exp[-y \log(10)]$ in the non-singular case and $\exp[-3 \log(y)]$ in the singular case. We see that the integrand of the non-singular case approaches zero as $y$ goes to $\infty$ much faster than in the singular case.
FIGURE 2.3: Plot of the Integrand $g(y)$ with $d^*=1$ and $\rho=-0.5$

FIGURE 2.4: Plot of the Integrand $G=g(y)\times10^9$ with $d^*=1$ and $\rho=-0.5$
FIGURE 2.5: Plot of the Integrand $g(y)$ with $d^* = 1$ $\rho = -1$

FIGURE 2.6: Plot of the Integrand $G = g(y) \times 100$ with $d^* = 1$ $\rho = -1$
In the non-singular case, if one is willing to use Gauss-Legendre quadrature, one can truncate an infinite interval to obtain a finite interval by ignoring the tail of the integrand. Rigorous application of this truncation method requires that one can estimate the magnitude of the tail area. In the calculation of \( F_2 \), the absolute value of the integrand is bounded by \( \exp(-y^2/6) \) (or by \( \exp\{-[1 - \tau^2]y^2/2\} \), where \( \tau \) is the sum of the squares of \( \alpha_j \) for general \( n \)). By a simple change variable, the tail area of \( g(y) \) can be transformed to be the tail of the error function with real argument, which is estimable.

Alternatively, one can split the integral into two parts; one is a finite integral ranging from zero to a positive constant, say \( b \); the other is an infinite integral ranging from \( b \) to \( \infty \). By a simple change variable such as \( z = 1/y \), one can transform the second integral to a finite integral. There are two disadvantages in this application. First, the integrand of the transformed integral is no longer the same as the original one. Second, the transformed integral might have singularities at the end points.

We note that the integral in the non-singular case is of the Gauss-Hermite integral form with weight function \( \exp(-y^2/6) \). With a change of variable of \( z = y/\sqrt{6} \), we have

\[
F_2 = \sqrt{2\pi} \int_{-\infty}^{\infty} g(y)dy = \sqrt{2\pi} \int_{-\infty}^{\infty} f(z)\exp(-z^2)dz,
\]

where \( f(z) = \sqrt{6}h(\sqrt{6}z) \) and \( g(y) = h(y)\exp(y^2/6) \).

The integrand \( f \) has a similar behavior to that of \( g \); both oscillating around zero and approaching zero at the same speed. However, the period of oscillation of \( f \) is shorter than that of \( g \), which is an indication that \( g \) can be approximated by a lower degree of polynomial (see Figure 2.7). In fact, with the same number of points, Gauss-Legendre provided a better approximation than did Gauss-Hermite. The superiority of the Gauss-Legendre approximation is even more evident as \( \rho \) approaches -1 (shown in Figure 2.7).
Because of the complicated behavior of the integrand in the singular case, we need some special treatments of the approximation of $F_2$. First, we need some techniques that can help accelerate the convergence. Second, we need some special integrator that can take care of the oscillating behavior of the integrand.

According to Davis and Rabinowitz (1984), the Trapezoidal rule or its variants generally give very good results for periodic functions. We note that Romberg's (Bauer, Rutishauser, and Stiefel (1963)) and Stoer and Bulirsch's (1962, 1967) methods are variants of the Trapezoidal rule. There are various techniques to speed up the convergence. Among others, Wynn (1956) has implemented an $\varepsilon$-algorithm for Aitken's $\Delta^2$ method. Romberg has described a recursive algorithm known as Romberg integration method to implement Richardson's extrapolation procedure to the Euler-Maclaurin sum formula. Stoer and Bulirsch (1962, 1967) have developed a rational extrapolation method, which is quite successful in the computation of $F_2$. In terms of speed, Stoer and
Bulirsch's method is about 3 times faster than on the average Romberg's. Therefore, we concluded that Stoer and Bulirsch's method is an efficient method for the singular case.

As far as execution time is concerned, the computation of the non-singular $F_2$ is marginally faster than that of singular case on a Cray X-MP/48 computer at the Ohio State University, if one uses an appropriate integration method such as Stoer and Bulirsch's. Part of the reason might be that the number of evaluations of the integrand in the singular case is slightly larger than in the non-singular case. The other reason is that the program for the non-singular case can be easily written so that quite a few of the innermost loops are vectorizable on vector machines.

The number of treatments, $n$, turns out to have little influence on the execution time. In fact, the execution time is linear in $n$ for the non-singular case. This is also true for the singular case, if $n$ is not too small ($n \geq 3$) and an appropriate integrator is used. Figure 2.8 provides some timing information on both singular and non-singular cases; the correlations are $\rho_{jk} = -1/[2(n-1)]$ and $\rho_{jk} = -1/(n-1)$ for non-singular and singular cases respectively. We note that the number of evaluations of the integrand (and hence the execution time) increases, if the correlations, $\rho_{jk}$, are large in absolute value. The execution time decreases after $n = 2$ since the correlations, $\rho_{jk}$, can not be too large in absolute value because of the nature of the restriction on the negative correlation structure.

For the computation of the common critical value $d^*$ for a fixed confidence level $1 - \alpha$, say 0.95, we need a root-finding method to compute a root, say $d_*^*$, of the equation $F_n(d^*) - (1 - \alpha) = 0$. The algorithm that we used is a modified regula falsi (or modified secant) method, also known as the Illinois method (see Conte and DeBoor (1980)). The motivation for the Illinois method is to reduce the number of iterations needed for convergence, because the method prevents the retention of an end point.
The speed of a root-finding algorithm is quite dependent on the initial estimates as well as on some tuning parameters, such as the error tolerance and the maximal number of iterations allowable. By using some well-known inequalities, we have derived an upper bound and a lower bound for $d^*$, which turn out to be good initial estimates. We shall state without proof two inequalities needed in deriving the bounds for $d^*$. References can be found in Tong (1980) and Marshall and Olkin (1979).

**Slepian's inequality:** Let $Y$ be multivariate normal with zero mean vector and correlation matrix $\Sigma$. Let $S$ and $T$ be two positive semi-definite correlation matrices such that every element in $S$ is not less than the corresponding element in $T$. Then

$$
P\left[\bigcap_{j=1}^{n} \{ Y_j \leq h_j \} \mid \Sigma = S \right] \geq P\left[\bigcap_{j=1}^{n} \{ Y_j \leq h_j \} \mid \Sigma = T \right].$$
Sidak's inequality: Let $Y$ be the same as defined above except that $\Sigma(\lambda) = \{\rho_{jh}\}$ depends on a real number $\lambda \in [0,1]$ such that $\rho_{jh} = \lambda \tau_{jh}$, for all $j \neq h$, where $T = \{\tau_{jh}\}$ is a positive semi-definite matrix. If the $h_j$'s are positive, then

$$\Pr\left[ \bigcap_{j=1}^{n} \{ |Y_j| \leq h_j \} \mid \Sigma = \Sigma(\lambda) \right]$$

is nondecreasing in $\lambda$.

If we can find the $h_j$'s such that the confidence level is $1 - \alpha$ for the probability statement in Sidak's inequality with $\lambda = 0$, then the same $h_j$'s would guarantee a larger confidence level for the probability statement with $\lambda \neq 0$. In the case where all $h_j$'s are identical, the common upper critical bound is the $[.5 + .5(1-\alpha)^{1/n}]^{1/2}$ percentile of a standard normal distribution.

On the other hand, if $S$ is an identity matrix and $T$ has a negative correlation structure, then, Slepian's inequality implies

$$\prod_{j=1}^{n} \Pr\{ Y_j \leq h_j \} \geq \Pr\left[ \bigcap_{j=1}^{n} \{ Y_j \leq h_j \} \right] \geq \Pr\left[ \bigcap_{j=1}^{n} \{ |Y_j| \leq h_j \} \right].$$

The last inequality holds because $\{ Y_j \leq h_j \} \supseteq \{ |Y_j| \leq h_j \}$, provided that the $h_j$'s are positive. If we can find $h_j$'s such that the leftmost term of the above inequalities is equal to the desired confidence level $1 - \alpha$, then the same $h_j$'s would guarantee that the rightmost term would not be larger than $1 - \alpha$. In the case where all the $h_j$'s are identical, the common lower critical bound is equal to the $[(1-\alpha)^{1/n}]^{1/2}$ percentile of a standard normal distribution.

For some special negative correlation structures, the upper bounds derived by Sidak's inequality are close to the true critical value. For instance, if we have a negative equal correlation structure and the number of treatments is not small (maybe $n \geq 7$), then the upper bound is close to the true critical value. In terms of probability, the closeness is at least in 3 significant figures. Figure 2.9 shows the values of some numerical calculations of the actual probabilities obtained by Sidak's and Slepian's bounds with $n = 2, 3, 4, 5, 6, 8, 10, 15, 20, 1-\alpha = 0.95$, and the correlations $\rho_{jk} = -1/[2(n-1)]$. Note that Sidak's bound always outperforms Slepian's bound.
Applications

Several applications of equation (2.0) were mentioned in Chapter I. In the next section, we shall present, in detail, two of these applications in the General Linear Model. One assumes normal errors and the other does not. An example of constructing the exact confidence intervals will be illustrated for the application with the normal assumption.

Unbalanced One-way Design (with 3 treatments)

Consider the model \( Y_{jk} = \mu_j + e_{jk} \); \( j = 1, 2, \ldots, n \); \( k = 1, 2, \ldots, m_j \), where \( \mu_1, \ldots, \mu_n \) are the treatment effects, and \( e_{jk} \) are iid \( N(0,\sigma^2) \) with an unknown \( \sigma^2 \). Let 
\[
N = \sum_{j=1}^{n} m_j, \quad v = N - n, \quad \bar{Y}_{j+} = \sum_{k=1}^{m_j} Y_{jk}/n_j, \quad S^2 = \sum_{j=1}^{n} \sum_{k=1}^{m_j} (Y_{jk} - \bar{Y}_{j+})^2 / v.
\]
Suppose that the
comparisons of the pairwise treatment differences are of interest. Let $q_{n\alpha}$ be the upper $[100(\alpha)]^{th}$ percentile of the studentized range distribution with parameters $n$ and $v$.

Hayter (1984) proved that $P\{ |(\bar{Y}_{j+} - \bar{Y}_{h+}) - (\mu_j - \mu_h)| \leq q_{n\alpha} s \sqrt{1/m_j+1/m_h} , j \neq h \} \geq 1 - \alpha$. That is, Tukey-Kramer’s intervals are conservative. However, the inequality becomes an equality when $n$ is 4 or less. Uusipaikka (1985) and Spurrier and Isham (1985) have developed procedures to evaluate the exact coverage probabilities for $n = 3$ or 4 and $n = 3$ respectively. Spurrier and Isham (1985) have tabulated (in their Table 1) the critical values at $\alpha = 0.01, 0.05$ and 0.10 and for most of the practical sample sizes $m_1, m_2$, and $m_3$.

We shall show that the underlying correlation matrix for this problem has a one-negative-factor structure for $n = 3$. Thus, we propose that the double integral in (2.6) be used to compute the exact coverage probability. We have checked quite a few critical values computed by Spurrier and Isham (1985). The same accuracy (4 significant figures in terms of computing a critical value) can be achieved by choosing suitable tuning parameters. For example, the tolerance of the function evaluation in the Illinois method is set to 0.000005, the maximal error of evaluating coverage probabilities, and the tolerance of the half width of the output interval is set to 0.0005, to guarantee that refinements on the output, $q_{n\alpha}$, will be continued until the current $q_{n\alpha}$ agrees with the $q_{n\alpha}$ returned from the previous step to 4 significant figures.

To show that the underlying correlation matrix has a one-negative-factor structure for $n = 3$, let $X_1 = [\bar{Y}_{1+} - \bar{Y}_{2+} - (\mu_1 - \mu_2)] / [\sigma \sqrt{1/m_1+1/m_2}] , \ X_2 = [\bar{Y}_{1+} - \bar{Y}_{3+} - (\mu_1 - \mu_3)] / [\sigma \sqrt{1/m_1+1/m_3}]$, and $X_3 = [\bar{Y}_{2+} - \bar{Y}_{3+} - (\mu_2 - \mu_3)] / [\sigma \sqrt{1/m_2+1/m_3}]$. Let $T_j = X_j / [S / \sigma]$. Then, $(X_1, X_2, X_3)$ is multivariate normal with zero means and correlation matrix $\Sigma = D - \alpha \alpha'$, where $\alpha' = (\sqrt{m_3/(m_1+m_2)}, \sqrt{m_2/(m_1+m_3)}, \sqrt{m_1/(m_2+m_3)})$, and $D$ is a diagonal matrix that makes $\Sigma$ a correlation matrix.
We note that any one of the $T_j$ can be expressed in terms of the others. For example, $T_3$ can be written as

$$T_3 = \left( \frac{1}{m_1 + m_3} T_1 - \frac{1}{m_2 + m_3} T_2 \right) / \left( \frac{1}{m_2 + m_3} T_2 \right).$$

Now, $T_1$ and $T_2$ have a bivariate $t$ distribution with $v$ degrees of freedom and with the associated correlation $\rho$ equals $\sqrt{m_2 m_3} / \sqrt{(m_1 + m_2)(m_1 + m_3)}$. To find the critical value $d$ with an exact coverage probability $1 - a$, Spurrier and Isham (1985) have proposed solving for $d$ in the following equation:

$$d u(t_1) \int_{-d}^{u(t_1)} f(t_1, t_2; \rho, v) \, dt_1 \, dt_2 = 1 - a,$$

where $l(t_1) = \max\{-d, \frac{\sqrt{1/m_1 + 1/m_3} T_2 - \sqrt{1/m_2 + 1/m_3} d}{\sqrt{1/m_2 + 1/m_3}}\}$ and $u(t_1) = \max\{d, \frac{\sqrt{1/m_1 + 1/m_3} T_2 + \sqrt{1/m_2 + 1/m_3} d}{\sqrt{1/m_2 + 1/m_3}}\}$ and $f(t_1, t_2; \rho, v)$ is the density function of $T_1$ and $T_2$. Alternatively, since $X_j$ and $S$ are independent and $v [S/\sigma]^2$ is distributed as a Chi-square with $v$ degrees of freedom, Corollary 2 implies that one can solve the following equation for the exact critical value $d$:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi \left[ \frac{sd - i \alpha_j y}{\sqrt{1 + \alpha_j^2}} \right] - \Phi \left[ \frac{-sd - i \alpha_j y}{\sqrt{1 + \alpha_j^2}} \right] d\Phi(y) \, d\Gamma(s) = 1 - a,$$

where $\Gamma(s)$ is the distribution function of $S / \sigma$ and $\Phi(y)$ is the standard normal distribution function.

There are two major differences between Spurrier and Isham's and our proposed method. First, the integrand in (2.8), involving a product of 3 complex distribution functions, looks more complicated in structure than the integrand in (2.7). However, we have utilized the most efficient existing approximation of $W(z)$ (and hence of $\Phi[z]$) to avoid the complexity. Second, the regions to be integrated are quite different. The integrations in (2.8) involve improper integrals; on the other hand, the integrals in (2.7) do not have a regular region for which standard integration formulas are known.

Spurrier and Isham (1985) proposed the use of 64-point Gaussian quadrature to iteratively evaluate integrals (2.7). The accuracy, in terms of the critical value, is to four
significant figures (see their Table1). To compute integrals (2.8), we propose that the inner integral be evaluated by Stoer and Bulirsch’s adaptive quadrature and that the outer integral be evaluated by Gauss-Legendre quadrature. Because the inner integral is finite (bounded by 1), the outer integral can be truncated by ignoring the tail of the integral. We checked quite a few of the exact critical values in Table 1 of Spurrier and Isham (1985) and found that the same accuracy can be achieved by using appropriate tuning parameters as specified before. Computing the critical value by (2.8) takes about 8 seconds on a Cray X-MP/48 computer at the Ohio State University with vectorization turned off, provided the tuning parameters were set such that the accuracy is of three significant figures.

Blood lead example

The data in the Pittsburgh lead-based paint study (Urban (1976) and on p. 330 of Morrison (1983)) includes the observations on blood lead level (μg/100ml) of children in the dwellings of the study area and their father’s education level in 3 groups: elementary school, high school, and some college. A one-way design was suggested and the following statistics were computed: \( m = (29, 101, 53), \bar{Y}_1 = 25.66, \bar{Y}_2 = 21.47, \bar{Y}_3 = 20.08 \), and \( s = 6.114191 \). To determine which groups differ in average blood lead level, an all pairwise comparison procedure was applied. Table 2.1 gives the 95% simultaneous confidence intervals obtained by various methods.

<table>
<thead>
<tr>
<th>Education</th>
<th>Exact CI</th>
<th>Scheffe's CI</th>
<th>Sidak's CI</th>
<th>Tukey's CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elem.-High</td>
<td>4.19±3.03</td>
<td>4.19±3.18</td>
<td>4.19±3.10</td>
<td>4.19±3.04</td>
</tr>
<tr>
<td>Elem.-College</td>
<td>5.58±3.32</td>
<td>5.58±3.49</td>
<td>5.58±3.40</td>
<td>5.58±3.34</td>
</tr>
<tr>
<td>High -College</td>
<td>1.39±2.44</td>
<td>1.39±2.56</td>
<td>1.39±2.50</td>
<td>1.39±2.45</td>
</tr>
</tbody>
</table>
In terms of simultaneous tests whether the pair differences of the sample means are zero, all of the methods, indicated in Table 2.1, have the same conclusion that the data does provide significant evidence against the null hypothesis of no pairwise differences in average blood lead level. However, the exact confidence intervals are shorter than those obtained by any other methods, although Tukey’s intervals are close to the exact intervals.

Linear Model Based on Ranks

Consider the general linear model \( Y = X \beta + e \); where \( Y \) is an \( N \times 1 \) vector of observations, \( X = (1, Z) \) is a known design matrix with an \( N \times p \) matrix \( Z \), \( \alpha \) is the grand mean, \( \beta' = (\beta_1, \ldots, \beta_p) \) is a vector of parameters, and \( e \) is an \( N \times 1 \) vector of iid errors from a common distribution \( G \in \Omega_0 = \{ G: G \) is absolutely continuous and \( G(0) = 1/2, \) uniquely\}. For convenience, let us assume that the column vectors of \( Z \) are centered, that is \( Z'1 = 0 \), which results in uncorrelated estimates of \( \alpha \) and \( \beta \). If the measure of dispersion is even and location free, then one can use either centered or uncentered \( Z \) without altering the results (see Hettmansperger (1984)).

Suppose \( \beta_0 \) is the true value of \( \beta \), let \( S(T) = [S_1(T), \ldots, S_p(T)] \), where \( T = (T_1, \ldots, T_N) \) with \( T_i = \text{Rank}(Y_i - z_i \beta) \), the rank of \( Y_i - z_i \beta \) among \( \{ Y_h - z_h \beta, h=1,\ldots, N \} \), and \( S_j(T) = \sum_{i=1}^N z_{ij} a(T_i) \), the regression rank statistics with scores \( a(1) \leq \cdots \leq a(N) \) such that \( a(k) + a(N-k+1) = 0 \). Note that \( S(T) / \sqrt{N} \) is asymptotically multivariate normal with zero means and a covariance matrix \( \Sigma \), to be defined, provided that the following assumptions hold (see Hettmansperger (1984) p.237):

1. \( a(i) = \sqrt{12} [i / (N+1) - 1 / 2] \), the Wilcoxon scores.
2. \( X \) has full column rank, \( p+1 \).
3. \( G \in \Omega_0 \) has finite Fisher information.
Consider the one-way design. For all $j=1,\ldots,k$, and $h=1,\ldots,n_j$, let $Y_{jh}$ be independent and follow the distribution $G(y - \mu_j)$ with $G \in \Omega_0$. To formulate the problem as a linear model with the non-singular design matrix described above, we let $Y' = (Y_{11}, \ldots, Y_{1n_1}, \ldots, Y_{k1}, \ldots, Y_{kn_k})$ and let $e'$ be the corresponding vector of errors. Furthermore, let $N = \sum_{j=1}^{k} n_j$, $\beta' = (\beta_2, \ldots, \beta_k)$, $X = (1, z_2, \ldots, z_k)$, an $N \times k$ matrix, where $z_j' = (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)$, with the ones in the $[[((\sum n_i) + 1)^{\frac{1}{2}}]_1$ through $[[((\sum n_i) + n_j)^{\frac{1}{2}})]$ components, and $Y = X \left[ \begin{array}{c} \beta' \\ e \end{array} \right]$. The parameters $\alpha$ and $\beta$ are related to the $\mu_j$'s as follows: $\mu_1 = \alpha$, $\mu_j = \alpha + \beta_j$, $j = 2, \ldots, k$. Then, we have $Z'Z/N \rightarrow \Sigma = D - \lambda \lambda'$, where $D = \text{diag}(\lambda)$ with $\lambda = (\lambda_2, \ldots, \lambda_k)$ such that $n_j/N \rightarrow \lambda_j$, $0 < \lambda_j < 1$, as $n_j$ (and, hence, $N$) $\rightarrow \infty$. That is, the covariance matrix of the regression rank statistics $S_j$, $j = 2, \ldots, k$, has a non-singular one-negative-factor structure. Therefore, the results for one-factor correlation structures can be applied to compute the p-value under the null hypothesis $H_0: \beta = \beta_0$ with a rejection rule of rejecting $H_0$, if $\max \{ S_j/\sqrt{\lambda_j(1-\lambda_j)}$ , $j = 2, \ldots, k \}$ is significantly large.
CHAPTER III

TWO-FACTOR STRUCTURES

Introduction

Suppose one is interested in the evaluation of multivariate normal distributions, in which the underlying correlation matrix has a two-factor structure. Without loss of generality, let us assume that the random vector \( X \) has already been standardized so that \( X \) follows a multivariate normal distribution with zero means and a correlation matrix \( \Sigma = D \pm \alpha^*\alpha^* \pm \beta^*\beta^* \). It is natural to ask if one can extend the results for one-factor correlation structures to this two-factor case. One possible extension follows the principle that adding one more factor to the correlation structure results in adding another integral independently integrated with respect to a standard normal density function.

In the two-positive-factor or one-positive-one-negative-factor case, our results can be proved using the technique of normal representations and the results for one-factor correlation structures, under the assumptions described as follows. The results for the two-positive-factor and non-singular one-positive-one-negative-factor case hold without any additional assumptions except the natural restriction that the correlation matrix is positive semi-definite. The results for the singular one-positive-one-negative-factor case need the additional assumption that the factor patterns of the factor with a negative sign in the decomposition of \( \Sigma \) are of the same sign.

There exists, however, no simple normal representation for the two-negative-factor case. To obtain the results in this case, we turn to analytic proofs, which become the
major topic of this chapter. In this case, we found that it is very difficult to derive an expression for the computation of a multivariate normal probability without imposing more assumptions in addition to those mentioned in the two-positive-factor and one-positive-one-negative-factor case. We shall study these assumptions later in this chapter.

**Normal representations**

Suppose an $n$-dimensional random vector $X$ is multivariate normal with zero means and a correlation matrix $\Sigma$, and $Y$ is multivariate normal with zero means and a correlation matrix $R$, where $\Sigma = R + \beta\beta^\top$. Let $Z$ be a standard normal random variable independent of $Y$. Then, $X$ and $Y + Z\beta^\top$ are equal in distribution because both $X$ and $Y + Z\beta^\top$ are multivariate normal with the same vector of means and variance-covariance matrix. We say that the random vector $Y + Z\beta^\top$ is a normal representation of $X$.

In terms of normal representations, adding a positive factor in correlation structure is essentially the same as introducing another independent normal random variable. Therefore, conditioning on the added normal variable, the underlying correlation is reduced to a one-factor structure. In terms of computation, the effect of the added independent normal random variable is the introduction of an additional integral integrated with respect to a standard normal density function.

In particular, suppose $R$ is a diagonal matrix so that the $Y_j$ are independent. Then, $\Sigma$ has a one-positive-factor structure. If $R$ has a one-factor structure, say $R = D \pm \alpha^\top \alpha^*$, then $\Sigma$ has either a two-positive-factor or one-positive-one-negative-factor structure. Hence, we can write that $P(X_j \leq h_j, j=1, \ldots, n) = P(Y_j + Z\beta_j^\top \leq h_j, j=1, \ldots, n).$ Conditioning on $Z$ and using the one-factor result in (2.0), the last probability statement can be expressed as

$$
\int_{\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi(\{(h_j - z_j\beta_j^\top - c\alpha_j^\top y)/(1 + \alpha_j^2 - c^2\beta_j^2)^{1/2}\} \Phi(y) \ d\Phi(z),
$$

(3.0)
where \( c = 1 \) (or \( i \)) for a positive (or negative) one-factor structure in \( \mathbb{R} \). Therefore, results for two-positive-factor or one-positive-one-negative-factor case are established by virtue of the results for one-factor structures.

**Goal of the Chapter**

The main purpose of this chapter is to examine the conditions under which the computation of a multivariate normal probability is possible in the two-negative-factor case. Some numerical considerations of and applications for the two-factor case will be explored as well.

The major results for the two-negative-factor case are presented in three theorems. Theorem 2 presents an expression for computing the probability of a lower quadrant, provided that the underlying correlation matrix is non-singular. Theorem 3 presents an expression for calculating the probability of the same lower quadrant as in Theorem 2, for any correlation matrix (singular or non-singular). Theorem 4, which is proved by induction, shows that the result in Theorem 2 can be applied to the singular case.

**Two-negative-factor Structures**

Let us concentrate on computing coverage probabilities based on the two-negative-factor correlation structure, that is, the correlation matrix \( \Sigma = D - \alpha^*\alpha^{**} - \beta^*\beta^{**} \). To prove our results in this case, we need to assume that if we ignore one of the two factors, the remaining matrix is still positive semi-definite. In the case where either \( \alpha^* \) or \( \beta^* \) is identically zero, so that \( \Sigma \) has a one-factor structure, the assumption is required. This assumption restricts the values of the factor patterns so that the off-diagonal elements of \( \Sigma \) are not too large in absolute value to assure that the determinant of \( \Sigma \), denoted by \( |\Sigma| \), is non-negative. We shall refer to this assumption as the nature restriction on negative
correlation structures. The other technical assumption, needed in Theorem 2, is that either

\[ |D - \alpha^* \alpha^*| \neq 0 \] or \[ |D - \beta^* \beta^*| \neq 0 \] or both.

The derivation of the expression for computing the probability of a lower quadrant in
the non-singular case is completed in steps as follows. First, the joint probability density
function of a multivariate normal random vector (after standardized) can be written as a
two-dimension Fourier transform (to be defined shortly) of two independent univariate
normal random variables. Then, by exchanging the order of integration, the multi­
dimensional integrations with respect to the integrand of the two-dimension Fourier
transform can be written as a double integral. This double integral is integrated with
respect to the distribution functions of two independent normal random variables and with
respect to a product integrand, the product of n normal distribution functions with
complex argument.

Lemma 2 provides a result for a special case of a two-dimension Fourier transform.
Lemma 4, using the result in Lemma 3, gives the inverse and the determinant of a
correlation matrix with a two-negative-factor structure.

Now, let us now define a two-dimension Fourier transform of an integrable function.
Let \( g \) be an integrable function on \( \mathbb{R}^2 \). A two-dimension Fourier transform of \( g \), denoted
by \( \hat{g} \), is defined as

\[
\hat{g}(s,t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) \exp[-i(sx + ty)] \, dx \, dy.
\]

**Lemma 2:** Suppose \( g(x,y) = (2\pi)^{-1} \exp[-(ax^2 - 2bxy + cy^2) / 2]. \) Then,

\[
\hat{g}(s,t) = a^{-1/2}(c-b^2/a)^{-1/2} \exp[-s^2/(2a)] \exp[-(t+sb/a)^2/(2(c-b^2/a))], \text{ provided } b^2 \neq a \, c.
\]

**Proof:** By definition,

\[
2\pi \hat{g}(s,t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[-(ax^2 - 2bxy + cy^2)/2] \exp[-i(sx + ty)] \, dx \, dy
\]

\[
= \int_{-\infty}^{\infty} \exp(-cy^2/2 - ity) \int_{-\infty}^{\infty} \exp(-ax^2/2 + byx - isx) \, dx \, dy.
\]
Now, \[ \int_{-\infty}^{\infty} \exp(-ax^2/2 + byx - isx) \, dx \]

\[ = \exp[b^2y^2/(2a)] \int_{-\infty}^{\infty} \exp[-a(x-by/a)^2/2 - isx] \, dx \]

\[ = \sqrt{1/a} \exp[b^2y^2/(2a)] \int_{-\infty}^{\infty} \exp[-z^2/2 - is(z/\sqrt{a}+by/a)] \, dz; \text{ change variables: } z=\sqrt{a}(x-by/a) \]

\[ = \sqrt{2\pi/a} \exp[-s^2/(2a)] \exp[b^2y^2/(2a)] \exp(-isby/a), \text{ by Lemma 1.} \]

Hence,

\[ 2\pi \hat{g}(s,t) = \sqrt{2\pi/a} \exp[-s^2/(2a)] \int_{-\infty}^{\infty} \exp[-(c-b^2/a)y^2/2 - i(t+sb/a)y] \, dy. \]

But,

\[ \int_{-\infty}^{\infty} \exp[-(c-b^2/a)y^2/2 - i(t+sb/a)y] \, dy \]

\[ = (c-b^2/a)^{-1/2} \int_{-\infty}^{\infty} \exp[-z^2/2 - i(t+sb/a)(c-b^2/a)^{-1/2}] \, dz; \text{ change variables: } z=\sqrt{c-b^2/acy} \]

\[ = (2\pi)^{-1/2}(c-b^2/a)^{-1/2}\exp[-(t+sb/a)^2/[2(c-b^2/a)]], \text{ by Lemma 1.} \]

Therefore,

\[ \hat{g}(s,t) = a^{-1/2}(c-b^2/a)^{-1/2} \exp[-s^2/(2a)] \exp[-(t+sb/a)^2/[2(c-b^2/a)]]. \]

QED

Lemma 3: (Grabill 1983) Suppose that \( \Sigma \) is an \( n \) by \( n \) matrix which can be written as \( \Sigma = D + \gamma ab' \), where \( D = \text{diag}\{d_1^2, \ldots, d_n^2\} \), a non-singular diagonal matrix, and \( \gamma \) is a scalar such that \( \gamma = -\left[ \sum_{j=1}^{n} (a_jb_j/d_j^2) \right]^{-1} \). Then,

\[ |\Sigma| = \left[ 1 + \gamma \sum_{j=1}^{n} (a_jb_j/d_j^2) \right] \prod_{j=1}^{n} d_j^2, \]

and

\[ \Sigma^{-1} = D^{-1} + \gamma^* a^* b^*, \]

where \( \gamma^* = -\gamma \left[ 1 + \gamma \sum_{j=1}^{n} (a_jb_j/d_j^2) \right]^{-1}, a^* = D^{-1} a, \text{ and } b^* = D^{-1} b. \)

Proof: See Theorem 8.3.3 and Theorem 8.4.3 of Grabill (1983).

QED
Lemma 4: Suppose that $\Sigma$ is an $n$ by $n$ matrix and can be written as $\Sigma = D - \alpha^*\alpha^*$ - $\beta^*\beta^*$, where $D = \text{diag}(d_1^2, \ldots, d_n^2)$ and $d_j^2 = 1 + \alpha_j^2 + \beta_j^2$. If $\sum_{j=1}^{n} \alpha_j^2 \neq 1$, then

$$|\Sigma| = c_1^{-2} c_2^{-2} \prod_{j=1}^{n} d_j^2,$$

$$\Sigma^{-1} = D^{-1} + c_1^2 \tilde{\alpha}\tilde{\alpha}' + c_2^2 \tilde{\beta}\tilde{\beta}'$$

where $\alpha = D^{-1/2}\alpha^*$, $\beta = D^{-1/2}\beta^*$, $\tilde{\alpha} = D^{-1}\alpha^*$, $\tilde{\beta} = (D - \alpha^*\alpha^*)^{-1}\beta^*$,

$$c_1^{-2} = 1 - \sum_{j=1}^{n} \alpha_j^2 \quad \text{and} \quad c_2^{-2} = 1 - \sum_{j=1}^{n} \beta_j^2 - c_1^2 \left( \sum_{j=1}^{n} \alpha_j \beta_j \right)^2.$$

Proof: We shall use the following two facts without proof. References can be found in multivariate analysis textbooks (for example, Mardia, Kent, and Bibby (1979) p. 457-9).

1. If $A^{-1}$ exists, then

$$\left( A + B C D \right)^{-1} = A^{-1} - A^{-1}B\left( C^{-1} + D A^{-1}B \right)^{-1}D A^{-1}, \text{ and}$$

$$|A + ab'| = |A| (1 + a' A^{-1} b).$$

Now, let us write $\Sigma = [(D - \alpha^*\alpha^*) - \beta^*]$, we may obtain that

$$|\Sigma| = |D - \alpha^*\alpha^*| \left( 1 - \beta^* (D - \alpha^*\alpha^*)^{-1} \beta^* \right), \text{ and}$$

$$\Sigma^{-1} = (D - \alpha^*\alpha^*)^{-1} + (D - \alpha^*\alpha^*)^{-1} \beta^* \left( 1 - \beta^* (D - \alpha^*\alpha^*)^{-1} \beta^* \right)^{-1} \beta^* (D - \alpha^*\alpha^*)^{-1}.$$

Applying Lemma 3, we have that

$$(D - \alpha^*\alpha^*)^{-1} = D^{-1} + c_1^2 \tilde{\alpha}\tilde{\alpha}', \text{ and}$$

$$|D - \alpha^*\alpha^*| = c_1^{-2} \prod_{j=1}^{n} d_j^2.$$

Hence,

$$1 - \beta^* (D - \alpha^*\alpha^*)^{-1} \beta^* = 1 - \beta^* D^{-1} \beta^* - c_1^2 \beta^* \tilde{\alpha}\tilde{\alpha}' \beta^* = c_2^{-2}.$$

Therefore,

$$|\Sigma| = c_1^{-2} c_2^{-2} \prod_{j=1}^{n} d_j^2, \text{ and} \Sigma^{-1} = D^{-1} + c_1^2 \tilde{\alpha}\tilde{\alpha}' + c_2^2 \tilde{\beta}\tilde{\beta}'.$$

QED
For simplicity, let us denote \( \alpha^2_j = \sum_{j=1}^{n} \alpha_j^2 \), \( \beta^2_j = \sum_{j=1}^{n} \beta_j^2 \) and \( (\alpha \beta)_j^2 = (\sum_{j=1}^{n} \alpha_j \beta_j)^2 \).

Remarks:
1. A sufficient condition for \( |\Sigma| = 0 \) in the two-negative-factor structure, under the assumption that either \( \alpha^2_j < 1 \) or \( \beta^2_j < 1 \) or both, is that
   \[
   (1 - \alpha^2_j) (1 - \beta^2_j) = (\alpha \beta)^2.
   \]
   For example, if \( \alpha^2_j < 1 \) then \( |\Sigma| = c_1^{-2} c_2^{-2} \prod_{j=1}^{n} d_j^2 = 0 \), if \( c_2^{-2} = 0 \).
2. A sufficient condition for \( |\Sigma| = 0 \) with \( \Sigma = \Sigma - \alpha^2 \alpha + \beta^2 \beta^* \), is that, for any \( \alpha \) and \( \beta \),
   \[
   (\alpha^2 - 1) (1 + \beta^2) = (\alpha \beta)^2.
   \]

The following theorem presents an expression for computing the probability of a lower quadrant in the non-singular case. The proof of this result follows in the reverse order the steps that derived the expression. First, we shall show that the expression can be written as a multiple integrals involving a two-dimension Fourier transform. Then, by applying Lemma 2, we can show that the two-dimension Fourier transform is the joint probability density function of a multivariate normal random vector.

**Theorem 2:** Let \( X \) be multivariate normal with a zero mean vector and a two-negative-factor correlation matrix \( \Sigma \), that is, \( E(X_j^2) = 1 \), for all \( j = 1, \cdots, n \) and \( \rho_{jk} = E(X_j X_k) = -\alpha_j^* \alpha_k^* - \beta_j^* \beta_k^* \), for all \( j \neq k \). Let \( F_n(h^* l \alpha_j^*, \beta_j^*) = P(X_j \leq h_j^* ; j = 1, \cdots, n) \). If \( |\Sigma| \neq 0 \), then \( F_n(h^* l \alpha_j^*, \beta_j^*) \) equals

\[
\left( 3.1 \right) \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left[ \frac{h_j^* - i (\alpha_j^* y_1 + \beta_j^* y_2)}{\sqrt{1 + \alpha_j^* y_1 + \beta_j^* y_2}} \right] d\Phi(y_1) d\Phi(y_2).
\]

For simplicity, let \( d_j, \alpha_j, \beta_j, c_1, \) and \( c_2 \) be as defined in Lemma 4. In addition, let \( h_j = h_j^*/d_j \). Without loss of generality, let us assume \( \alpha^2_j < 1 \), \( (1 - \alpha^2_j) (1 - \beta^2_j) \neq (\alpha \beta)^2 \). (The case that \( \beta^2_j < 1 \), \( (1 - \alpha^2_j) (1 - \beta^2_j) \neq (\alpha \beta)^2 \) can be proved similarly by exchanging the roles of \( \alpha \) and \( \beta \).)
Proof: Let us write,

\[
\Phi \left[ \frac{h_{*} - i(\alpha_{j}^{*}y_{1} + \beta_{j}^{*}y_{2})}{\sqrt{1 + \alpha_{j}^{*2} + \beta_{j}^{*2}}} \right] = \Phi[h_{j} - i(\alpha_{j}y_{1} + \beta_{j}y_{2})].
\]

By definition,

\[
\Phi[h_{j} - i(\alpha_{j}y_{1} + \beta_{j}y_{2})] = \exp\left[ \frac{(\alpha_{j}y_{1} + \beta_{j}y_{2})^2}{2} \right] \int \exp\left[ it(\alpha_{j}y_{1} + \beta_{j}y_{2}) \right] d\Phi(t).
\]

Now, let us denote the expression in (3.1) by \( R_{n} \) and let

\[
G(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left[ \sum_{j=1}^{n} (\alpha_{j}y_{1} + \beta_{j}y_{2})^2 / 2 \right] \exp\left[ i \sum_{j=1}^{n} t_{j}(\alpha_{j}y_{1} + \beta_{j}y_{2}) \right] d\Phi(y_{1}) d\Phi(y_{2})
\]

\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ -\frac{1}{2} \left[ (1 - \sum_{j=1}^{n} \alpha_{j}^2 - 2(\sum_{j=1}^{n} \alpha_{j}\beta_{j})y_{1}y_{2} + (1 - \sum_{j=1}^{n} \beta_{j}^2) y_{2}^2 \right] \right\} \exp\left\{ i \left[ \sum_{j=1}^{n} t_{j}\alpha_{j} y_{1} + (\sum_{j=1}^{n} t_{j}\beta_{j}) y_{2} \right] \right\} dy_{1} dy_{2}
\]

\[
= c_{1}c_{2}\exp\left[ -\frac{c_{1}^2}{2}(\sum_{j=1}^{n} \alpha_{j} t_{j})^2 \right] \exp\left[ -\frac{c_{2}^2}{2}(\sum_{j=1}^{n} \beta_{j} t_{j} + c_{1}^2(\sum_{j=1}^{n} \alpha_{j} t_{j})(\sum_{j=1}^{n} \beta_{j})^2) \right], \text{ by Lemma 2.}
\]

Exchanging the order of integration, which is legitimate because \( G(t) \) is bounded, we obtain that

\[
R_{n} = \int_{-\infty}^{h_{1}} \cdots \int_{-\infty}^{h_{n}} G(t) \prod_{j=1}^{n} d\Phi(t_{j})
\]

Changing variables by letting \( t_{j} = u_{j} / d_{j} \), we have that

\[
\left\{ c_{1}^{-1} c_{2}^{-1} (2\pi)^{n/2} \prod_{j=1}^{n} d_{j} \right\} R_{n}
\]

\[
= \int_{-\infty}^{h_{1}} \cdots \int_{-\infty}^{h_{2}} \exp\left\{ -\frac{c_{1}^2}{2}(\sum_{j=1}^{n} \alpha_{j} u_{j} / d_{j})^2 - \frac{c_{2}^2}{2}(\sum_{j=1}^{n} \beta_{j} u_{j} / d_{j}) + c_{1}^2(\sum_{j=1}^{n} \alpha_{j} u_{j} / d_{j}) + c_{1}^2(\sum_{j=1}^{n} \alpha_{j} \beta_{j}) \right\}
\]

\[
\cdot \prod_{j=1}^{n} \exp\left\{ - (u_{j} / d_{j})^2 / 2 \right\} du_{1} \cdots du_{n}.
\]

\[
= \int_{-\infty}^{h_{1}} \cdots \int_{-\infty}^{h_{2}} \exp\left\{ - u^{*} u / 2 \right\} du_{1} \cdots du_{n}, \text{ because, by Lemma 4, we have}
\]
\[ u^* \Sigma^{-1} u \]
\[ = u' \{ D^{-1} + c_1^2 D^{-1} \alpha^* \alpha^* D^{-1} + c_2^2 (D^{-1} + c_1^2 D^{-1} \alpha^* \alpha^* D^{-1} - \beta^* \beta^* (D^{-1} + c_1^2 D^{-1} \alpha^* \alpha^* D^{-1})) \} u \]
\[ = u' D^{-1} u + c_1^2 (u' D^{-1} \alpha^*) (\alpha^* D^{-1} u) \]
\[ + c_2^2 \{(u' D^{-1} \beta^*) (\beta^* D^{-1} u) + 2c_1^2 (u' D^{-1} \beta^*) (\beta^* D^{-1} \alpha^*) (\alpha^* D^{-1} u) \}
\[ + c_1^4 \{(u' D^{-1} \alpha^*) (\alpha^* D^{-1} \beta^*) (\beta^* D^{-1} \alpha^*) (\alpha^* D^{-1} u) \}] \]
\[ = \sum_{j=1}^{n} \left( \frac{u_j}{d_j} \right)^2 + c_1^2 \left( \sum_{j=1}^{n} \alpha_j u_j/d_j \right)^2 \]
\[ + c_2^2 \left\{ \left( \sum_{j=1}^{n} \beta_j u_j/d_j \right)^2 + 2c_1^2 \left( \sum_{j=1}^{n} \alpha_j \beta_j \right) \left( \sum_{j=1}^{n} \alpha_j u_j/d_j \right) \right\} \]
\[ + c_1^4 \left\{ \left( \sum_{j=1}^{n} \alpha_j u_j/d_j \right) \left( \sum_{j=1}^{n} \alpha_j \beta_j \right) \right\}^2 \]
\[ = \sum_{j=1}^{n} \left( \frac{u_j}{d_j} \right)^2 + c_1^2 \left( \sum_{j=1}^{n} \alpha_j u_j/d_j \right)^2 \]
\[ + c_2^2 \left\{ \left( \sum_{j=1}^{n} \beta_j u_j/d_j \right) + c_1^2 \left( \sum_{j=1}^{n} \alpha_j u_j/d_j \right) \left( \sum_{j=1}^{n} \alpha_j \beta_j \right) \right\}^2 \]

Therefore,
\[ R_n = (2\pi)^{-n/2} (\Sigma)^{-1/2} \int_{-\infty}^{h_1^*} \cdots \int_{-\infty}^{h_2^*} \exp \{- u^* \Sigma^{-1} u / 2 \} \, du_1 \cdots du_n . \]

QED

For simplicity, let us drop the \(^*\) in \( h^*, \alpha^* \) and \( \beta^* \). We shall derive an expression for \( F_n(h | \alpha, \beta) \) in the next theorem, Theorem 3, to relate \( n \)-dimension distributions to \((n-1)\)-dimension distributions. The proof of this theorem does not require the non-singularity of the correlation matrix \( \Sigma \). This result will be used in the proof of Theorem 4.
Theorem 3: For all \( k \in \{1, \cdots, j-1, j+1, \cdots, n\} \), let the components of the vectors \( a, b \) and \( s \) be:

\[
a_k = \frac{\alpha_k \sqrt{1 + \alpha_j^2 + \beta_j^2}}{\sqrt{1 + \beta_j^2} \sqrt{1 - (\alpha_j \alpha_k + \beta_j \beta_k)^2}}.
\]

\[
b_k = \frac{\alpha_k (\alpha_j \beta_j) + \beta_k (1 + \beta_j^2)}{\sqrt{1 + \beta_j^2} \sqrt{1 - (\alpha_j \alpha_k + \beta_j \beta_k)^2}},
\]

and

\[
s_k = \frac{u(1 + \alpha_j \alpha_k + \beta_j \beta_k) + h_k - h_j}{\sqrt{1 - (\alpha_j \alpha_k + \beta_j \beta_k)^2}}.
\]

Then,

\[
F_n(h \mid \alpha, \beta) = \sum_{j=1}^{n} \int_{-\infty}^{h_j} F_{n-1}(s \mid a, b) \, d\Phi(u).
\]

Note that \( 1 + a_k^2 + b_k^2 = (1 + \alpha_k^2 + \beta_k^2) / [1 - (\alpha_j \alpha_k + \beta_j \beta_k)^2] \).

Proof: Let \( A_k \) be the event \( X_k \leq X_j + h_k - h_j \) and \( B_k \) be the event \( X_k \leq u + h_k - h_j \). By projecting the region \( \bigcap_{j=1}^{n} \{X_j \leq h_j\} \) onto the plane perpendicular to \((1, \cdots, 1)\), the region becomes the union of \( n \) disjoint areas specified by the associated events: \( \bigcap_{k \neq j} A_k \), for all \( j=1, \cdots, n \). Thus,

\[
F_n(h \mid \alpha, \beta) = \sum_{j=1}^{n} \prod_{k \neq j} \left[ \bigcap_{k \neq j} A_k \right] = \sum_{j=1}^{n} \left[ \bigcap_{k \neq j} B_k \mid X_j = u \right] \, d\Phi(u).
\]

Let \( Y = (X_1, \cdots, X_{j-1}, X_{j+1}, \cdots, X_n) \). Partition \( \Sigma \) into

\[
\begin{bmatrix}
1 & \sigma_{21} \\
\sigma_{21} & \Sigma_{22}
\end{bmatrix},
\]

where \( \Sigma_{22} = \text{Var}(Y) \) and \( \sigma_{21} = \text{Cor}(X_j, Y) \). Then, \( Y \mid X_j = u \) has a multivariate normal distribution with mean \( \mu_2 \) and covariance matrix \( \Sigma_{22:1} \), where \( \mu_2 = \sigma_{21} u \) and \( \Sigma_{22:1} = \Sigma_{22} - \sigma_{21} \sigma_{21}' \).

Explicitly, we may write

\[
E(X_k \mid X_j = u) = - (\alpha_j \alpha_k + \beta_j \beta_k) \, u,
\]

\[
\text{Var}(X_k \mid X_j = u) = 1 - (\alpha_j \alpha_k + \beta_j \beta_k)^2, \text{ and}
\]
Cor(X_m, X_k | X_j=u) = \{-\alpha_k \alpha_m (1+\alpha^2) \beta_k \beta_m (1+\beta^2) - \alpha_j \beta_j (\alpha_k \beta_m + \alpha_m \beta_k)\} \\
\cdot \{ [1 - (\alpha_j \alpha_m + \beta_j \beta_m)^2] [1 - (\alpha_j \alpha_k + \beta_j \beta_k)^2] \}^{-1/2} \\
= -a_k a_m - b_k b_m.

Hence,

\begin{align*}
F_n(h | \alpha, \beta) &= \sum_{j=1}^{n} \int_{-\infty}^{h_j} \mathbb{P} \left[ \bigcap_{k \neq j} \{ X_k \leq u + h_k - h_j | X_j = u \} \right] d\Phi(u) \\
&= \sum_{j=1}^{n} \int_{-\infty}^{h_j} \mathbb{P} \left[ \bigcap_{k \neq j} \{ X_k + u(\alpha_j \alpha_k + \beta_j \beta_k) \leq u(1+\alpha_j \alpha_k + \beta_j \beta_k) + h_k - h_j \} | X_j = u \} \right] d\Phi(u) \\
&= \sum_{j=1}^{n} \int_{-\infty}^{h_j} F_{n-1}(s|a,b)d\Phi(u); \text{ by standardizing the conditional r.v.'s } X_k | X_j = u's.)
\end{align*}

QED

To prove, by induction, that the expression in (3.1) is valid for the singular case, we need to show that equation (3.1) holds for n=2. The following lemma sketches a proof of this special case without the details which can be justified through similar arguments in Theorem 1.

**Lemma 5:** Let X_1 and X_2 be bivariate normal with E(X_1) = E(X_2) = 0, V(X_1) = V(X_2) = 1 and E(X_1 X_2) = -1. For j = 1, 2, let d_j^2 = 1 + \alpha_j^* \beta_j^*, \alpha_j = \alpha_j^* / d_j, \beta_j = \beta_j^* / d_j and h_j = h_j^* / d_j, where \alpha_j^* and \beta_j^* are positive real constants and where h_j^* is a real constant. Further suppose that not both \alpha_1^2 + \alpha_2^2 = 1 and \beta_1^2 + \beta_2^2 = 1 and that (1 - \alpha_1^2 - \alpha_2^2) (1 - \beta_1^2 - \beta_2^2) = (\alpha_1 \beta_1 + \alpha_2 \beta_2)^2. Without loss of generality, let \alpha_1^2 + \alpha_2^2 \neq 1.

Then,

\begin{align*}
(3.2) \int_{-\infty}^{h_1} \Phi(h_1 - i(\alpha_1 \alpha + \beta_1 y)) \Phi(h_2 - i(\alpha_2 x + \beta_2 y)) d\Phi(x) d\Phi(y) = \mathbb{P}(-h_2^* \leq X_1 \leq h_1^*).
\end{align*}
For simplicity let us denote

\[ f_0(s,t,x,y) = \exp \{-i[(\alpha_1 t + \alpha_2 s)x + (\beta_1 t + \beta_2 s)y]\} f_1(s,t,x,y), \]

\[ f_1(s,t,x,y) = f_2(x,y) \exp(-s^2/2) \exp(-t^2/2) \text{ and} \]

\[ f_2(x,y) = \exp\{-(1-\alpha_+^2)^{1/2}x - (1-\beta_+^2)^{1/2}y\}^2/2, \]

where \[ \alpha_+^2 = \alpha_1^2 + \alpha_2^2 \text{ and } \beta_+^2 = \beta_1^2 + \beta_2^2. \]

Furthermore, let the right hand side of equation (3.2) be denoted by \( R \), then

\[ R = (2\pi)^{-1} \int \int \int \int f_0(s,t,x,y) \, ds \, dt \, dx \, dy \]

\[ = \frac{1}{2\pi^2} \int \int \int \cos[(\alpha_1 t + \alpha_2 s)x + (\beta_1 t + \beta_2 s)y] f_1(s,t,x,y) \, ds \, dt \, dx \, dy. \]

The proof of this lemma follows closely that of Theorem 1. It is still easier to present the proof in three cases, corresponding to the three regions: \( h_1^* > -h_2^* \), \( h_1^* < -h_2^* \) and \( h_1^* = -h_2^* \). We need to show \( R = \Phi(h_1^*) - \Phi(-h_2^*) \), if \( h_1^* > -h_2^* \) and \( R = 0 \), otherwise.

**Sketch of the Proof:**

**Case 1:** \( h_1^* > -h_2^* \). Using equation (2.2) in Chapter II with \( j = 2 \), we have \( R = T_1 + T_2 \), where

\[ T_1 = \frac{1}{2\pi \sqrt{2\pi}} \int \int f_2(x,y) e^{i(\alpha_1 x - \beta_1 y)} e^{-t^2/2} e^{-(\alpha_2 x + \beta_2 y)^2/2} \, dt \, dx \, dy \]

\[ = \frac{1}{2\pi \sqrt{2\pi}} \int \int e^{-t^2/2} \int e^{i(\alpha_1 x - \beta_1 y)} e^{-(1-\alpha_1^2)x^2 - 2\alpha_1 \beta_1 xy - (1-\beta_1^2)y^2 - 2\alpha_1 \beta_1 xy)} \, dx \, dy \, dt \]

\[ = \frac{1}{\sqrt{\gamma_0 \gamma_1}} \int_{-\infty}^{\infty} e^{-\alpha_1^2/2} e^{-(\gamma_2/2) e^{\gamma_0 \gamma_1}} \, dt \]

\[ = \frac{1}{\sqrt{\gamma_0 \gamma_1}} \int_{-\infty}^{\infty} e^{-\gamma_2/2} \, dt \]

\[ = \Phi(h_1 \sqrt{2}) = \Phi(h_1^*) \text{ and} \]
By similar arguments, we may have that
\[ T_2 = T_3 + T_4, \]
where \( T_3 = \Phi(- h_2^2) \) and
\[ T_4 = \frac{1}{2\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cos[(\alpha_1 t + \alpha_2 s)x + (\beta_1 t + \beta_2 s)y] f_1(s,t,x,y) \, ds \, dt \, dx \, dy. \]

Thus, it remains to show \( T_4 = 0. \)

We shall further assume \((h_1^*, h_2^*)\) is in the first quadrant but excluding the origin. A similar proof can be applied to the cases when \((h_1^*, h_2^*)\) is in the second or fourth quadrants. Notice that \( \pi^2 T_4 \leq \lim_{b \to \infty} I(b) \, dy, \)
where
\[ I(b) = \int \int_{h_1 h_2} \int \int \cos[(\alpha_1 t + \alpha_2 s)x + (\beta_1 t + \beta_2 s)y] e^{-s^2/2} e^{-t^2/2} \, ds \, dt \, dx. \]

We have
\[ |I(b)| \leq \frac{2}{\alpha_2 b + \beta_2 y} e^{-h_2^2/2} \int_{h_1}^{\infty} e^{-t^2/2} \alpha_1 t + \alpha_2 h_2 \, dt \to 0, \text{ as } b \to \infty, \]
following closely the arguments in the proof of Theorem 1.

Thus, \( T_4 = 0. \)

**Case 2:** \( h_1^* < - h_2^* \). The proof that \( R = 0 \), in this case, is similar to the proof that \( T_4 = 0 \) in Case 1 and hence is omitted.

**Case 3:** \( h_1^* = - h_2^* \). Let us assume \( h_1^* = h_2^* = 0. \) The other cases can be proved similarly.

For a fixed constant \( \partial \geq 0 \), let \( f_0(s,t,x,y) \) be the same as defined earlier, in addition, let us introduce \( I_3, I_0, \) and \( f_3(s,t,x,y) \), where
\[ f_3(s,t,x,y) = \exp[-i(\alpha_1 t + \alpha_2 s)x] \exp(-\partial x^2) \exp[-i(\beta_1 t + \beta_2 s)y] \exp(-\partial y^2) \exp(-s^2/2) \exp(-t^2/2), \]
To prove the result that $I_0 = 0$, we need to show that $\lim_{\delta \to 0} I_\delta = I_0$ and $\lim_{\delta \to 0} I_\delta = 0$. Since

$$I_\delta = \frac{\pi^2}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\alpha_1 t + \alpha_2 s^2/(4\delta)} e^{-\beta_1 t + \beta_2 s^2/(4\delta)} e^{-s^2/2} e^{-t^2/2} \, ds \, dt,$$ 

we have

$$|I_\delta| \leq \frac{\pi^2}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-s^2/2} e^{-t^2/2} \, ds \, dt \to 0, \text{ as } \delta \to 0.$$

By similar arguments to those in Theorem 1, we may show that $|I_0|$ is finite and so is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_0(s, t, x, y) \, ds \, dt \, dx \, dy.$$

Lebesgue's dominated convergence theorem implies that $\lim_{\delta \to 0} I_\delta = I_0$. Therefore, $I_0 = 0$.

QED

Now, we will use the result in Theorem 3 to show that in fact the expression in (3.1) is valid for singular correlation structures.

Theorem 4: In the singular case $|\Sigma| = 0$, the expression in (3.1) still holds.

Proof: Lemma 5 has shown that the assertion is true for $n = 2$. By induction, we need to show that the assertion holds for $n$, assuming that it is true for $n-1$. Without loss of generality, let $h_j = h + \Delta h_j$, for all $j \in \{1, 2, \ldots, n\}$.

By Theorem 3,

$$F_n = \sum_{j=1}^{n} h_j^j \int F_{n-1} \left[ s(u) \mid a, b \right] d\Phi(u)$$

$$= \sum_{j=1}^{n} \int F_{n-1} \left[ s(u + \Delta h_j) \mid a, b \right] \frac{d\Phi(u + \Delta h_j)}{du} du$$

$$= \int \sum_{j=1}^{n} \int \prod_{k \neq j} \Phi(g_k(x_1, x_2)) \, d\Phi(x_1) \, d\Phi(x_2) \frac{d\Phi(u + \Delta h_j)}{du} du,$$
\[ g_k(x_1, x_2) = \frac{1}{\sqrt{1 + a_k^2 + b_k^2}} \{ s_k - i (a_k x_1 + b_k x_2) \} \]

\[ = \frac{1}{\sqrt{1 + \alpha_k^2 + \beta_k^2}} \left\{ \frac{\sqrt{1 + \alpha_k^2 + \beta_k^2} x_1 - i \left[ \frac{\beta_k \sqrt{1 + \beta_k^2} + \frac{\alpha_k \alpha_k \beta_k}{\sqrt{1 + \beta_k^2}} x_1 - i \left[ \frac{\beta_k \sqrt{1 + \beta_k^2} + \frac{\alpha_k \alpha_k \beta_k}{\sqrt{1 + \beta_k^2}} x_2 \right] \right] \right.\right. \]

\[ = \frac{1}{\sqrt{1 + \alpha_k^2 + \beta_k^2}} \left[ (u + \Delta h_k) - i \alpha_k y_1(x_1, x_2) - i \beta_k y_2(x_1, x_2) \right], \]

because Cor \( (X_k, X_j) = -\alpha_j \alpha_k - \beta_j \beta_k \) and Var\( (X_k) = \text{Var}(X_k) = 1 \), which imply \( \Delta h_j = -\Delta h_k (\alpha_j \alpha_k + \beta_j \beta_k) \), and, hence \( h_k - h_j = \Delta h_k - \Delta h_j = \Delta h_k - \left[ \Delta h_k (\alpha_j \alpha_k + \beta_j \beta_k) \right] = \Delta h_k (1 + \alpha_j \alpha_k + \beta_j \beta_k). \)

Changing variables by letting

\[ x_1 = \frac{\sqrt{1 + \beta_k^2}}{\sqrt{1 + \alpha_k^2 + \beta_k^2}} \left[ y_1 - i \frac{\alpha_j(u + \Delta h_j - i \beta_j y_2)}{1 + \beta_j^2} \right] \]

and

\[ x_2 = \frac{1}{\sqrt{1 + \beta_j^2}} \left[ y_2 - i \beta_j(u + \Delta h_j) \right], \]

we have that

\[ y_1 = \frac{\sqrt{1 + \alpha_j^2 + \beta_j^2}}{\sqrt{1 + \beta_j^2}} x_1 + \frac{\alpha_j \beta_j}{\sqrt{1 + \beta_j^2}} x_2 + i \alpha_j(u + \Delta h_j), \]

\[ y_2 = \sqrt{1 + \beta_j^2} x_2 + i \beta_j(u + \Delta h_j), \]

and that

\[ \frac{d}{du} \Phi(u + \Delta h_j) \frac{d}{d\Phi(x_1)} \frac{d}{d\Phi(x_2)} \]

\[ = \frac{d}{du} \Phi(u + \Delta h_j) \cdot \frac{d}{d\Phi} \left[ \frac{1}{\sqrt{1 + \beta_j^2}} \left[ y_2 - i \beta_j(u + \Delta h_j) \right] \right] \]

\[ \cdot \frac{d}{d\Phi} \left[ \frac{\sqrt{1 + \beta_j^2}}{\sqrt{1 + \alpha_j^2 + \beta_j^2}} \left[ y_1 - i \frac{\alpha_j(u + \Delta h_j - i \beta_j y_2)}{1 + \beta_j^2} \right] \right]. \]
\[
\begin{align*}
\frac{1}{\sqrt{2\pi}} & \exp\left[-(u + \Delta h_j)^2/2\right] \\
\times & \left\{ \frac{1}{\sqrt{2\pi\sqrt{1+\beta_j^2}}} \exp\left[-y_2^2/2\right] \exp\left[-\frac{1}{2(1+\beta_j^2)}(u+\Delta h_j-i\beta_j y_2)^2\right] \exp\left[(u+\Delta h_j)^2/2\right] dy_2 \right\} \\
& \cdot \exp\left[\frac{1}{2(1+\beta_j^2)}\left(1 + \frac{1}{\sqrt{1+\alpha_j^2+\beta_j^2}}\right)\right] \\
& \cdot \exp\left[\frac{1}{2(1+\alpha_j^2+\beta_j^2)}\left(\frac{1}{\sqrt{1+\alpha_j^2+\beta_j^2}}\right)^2\right] \\
& \cdot \exp\left\{ -\frac{1}{2} \left[ \frac{u+\Delta h_l-i(\alpha_l y_1+\beta_l y_2)}{\sqrt{1+\alpha_l^2+\beta_l^2}} \right] \right\} dy_1 dy_2 \\
& \cdot \exp\left\{ -\frac{1}{2} \left[ \frac{u+\Delta h_j-i(\alpha_j y_1+\beta_j y_2)}{\sqrt{1+\alpha_j^2+\beta_j^2}} \right] \right\} dy_1 dy_2 \\
& = \frac{d}{du}\left[ \frac{u + \Delta h_j - i (\alpha_j y_1 + \beta_j y_2)}{\sqrt{1 + \alpha_j^2 + \beta_j^2}} \right] d\Phi(y_1) d\Phi(y_2), \text{ denoted by } G(y_1, y_2 | \alpha, \beta). \\
\end{align*}
\]

Thus,

\[
F_n = \sum_{j=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{k \neq j} \Phi \left[ \frac{u + \Delta h_k - i (\alpha_k y_1 + \beta_k y_2)}{\sqrt{1 + \alpha_k^2 + \beta_k^2}} \right] G(y_1, y_2 | \alpha, \beta) du \\
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ \prod_{j=1}^{n} \Phi \left[ \frac{u + \Delta h_j - i (\alpha_j y_1 + \beta_j y_2)}{\sqrt{1 + \alpha_j^2 + \beta_j^2}} \right] \right\} \left\{ \prod_{j=1}^{n} \Phi \left[ \frac{u + \Delta h_j - i (\alpha_j y_1 + \beta_j y_2)}{\sqrt{1 + \alpha_j^2 + \beta_j^2}} \right] \right\} du d\Phi(y_1) d\Phi(y_2) \\
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left[ \frac{h + \Delta h_j - i (\alpha_j y_1 + \beta_j y_2)}{\sqrt{1 + \alpha_j^2 + \beta_j^2}} \right] d\Phi(y_1) d\Phi(y_2). \\
QED
\]
Once we established the results for two-factor structures, two useful corollaries follow easily. Similar to the corollaries derived for the one-factor case, Corollary 3 below will extend the result in (3.1), which works for the one-sided multivariate normal setting, to the one-sided multivariate \( t \) model, and Corollary 4 will in turn extend the result in Corollary 3 to the two-sided multivariate \( t \) case.

**Corollary 3:** Let \( X \) be multivariate normal with zero means and a two-factor correlation matrix \( \Sigma = D \pm \alpha \alpha' \pm \beta \beta' \). Further suppose that \( S \) is independent of the \( X_j \)'s and \( vS^2 \) has a Chi-square distribution with \( v \) degrees of freedom. For all \( j=1, \ldots, n \), let \( T_j = X_j / S \). Then, \( \Pr(T_1 \leq h_1, \ldots, T_n \leq h_n) \) can be expressed as

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left[ \frac{sh_i y - c_1 \alpha_i y_1 - c_2 \beta_i y_2}{\sqrt{1 - c_1^2 \alpha_i^2 - c_2^2 \beta_i^2}} \right] \Phi(y_1) \Phi(y_2) d\Gamma(s),
\]

where \( \Phi(\cdot) \) is defined as the distribution function of a standard normal random variable, \( \Gamma(\cdot) \) is the distribution function of \( S \), \( c_1 = 1 \) (or \( i \)) for positive (or negative) first factor and \( c_2 = 1 \) (or \( i \)) for positive (or negative) second factor.

The Proof is very similar to that of Corollary 1 and is omitted.

**Corollary 4:** Let \( X, S \) and \( T \) be the same as those defined in Corollary 3. Then, \( \Pr( |T_1| \leq h_1, \ldots, |T_n| \leq h_n ) \) can be expressed as

\[
(3.3) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{n} \Phi \left[ \frac{sh_i y - c_1 \alpha_i y_1 - c_2 \beta_i y_2}{\sqrt{1 - c_1^2 \alpha_i^2 - c_2^2 \beta_i^2}} \right] \Phi \left[ \frac{-sh_i y - c_1 \alpha_i y_1 - c_2 \beta_i y_2}{\sqrt{1 - c_1^2 \alpha_i^2 - c_2^2 \beta_i^2}} \right] \Phi(y_1) \Phi(y_2) d\Gamma(s),
\]

where \( \Phi, \Gamma, c_1 \) and \( c_2 \) are as defined in Corollary 3.

The Proof is trivial and is omitted.
Computations

Integrals (3.3) and (3.0) are essentially the same as integrals (2.6) and (2.0) respectively except for the additional outer integral integrated independently with respect to another standard normal density function. Hence, all the difficulties (such as oscillating integrand, slow convergence rate, and improper integrals in the negative-factor cases) involved in the evaluation of (2.6) remain in the computation of (3.3). The biggest concern is the addition of an integral in the evaluation of (3.3). Not only because it is time-consuming to have an additional integral but also it can be harmful to the accuracy, especially when the integrand is not well-behaved.

In the two-positive-factor case, the integrand is generally smooth and well-behaved in the sense that it approaches zero rapidly and its tail area is easily estimable so that the error can be controlled. Iteratively evaluating the triple integrals by Gaussian quadratures is feasible, although it is time-consuming on sequential machines. However, one can improve the execution time by coding the most time-consuming part of the integration so that vectorization can take place if vector processing capability is available. Hsu (1989) has introduced one possible improvement using vectorization, which can be at least three times faster than the procedure with the vector processing capability not used.

In the one-positive-one-negative-factor case, if we integrate with respect to the negative factor first, the outer double integrals can be iteratively evaluated by a Gaussian quadrature. Thus, in terms of computing probabilities, integrals (3.3) can be regarded as a sum of m parts of probabilities calculated on the one-negative-factor structure, where m is the number of points used in the Gaussian quadrature. Hence, computing probabilities in this way is about m times slower than in the one-negative-factor case. In principle, adaptive quadrature can be vectorized in an analogous fashion as in the Iterative Gaussian Quadrature method, that is, vectoring the product integrand. However, unless we have a
vectorizable code for the complex-valued error function, computing a critical value this way is too slow for interactive data analysis.

The most troublesome case in the evaluation of integrals (3.3) is the two-negative-factor case, especially when the correlation is singular and n equals two. Because of the oscillating and long-tailed integrand, the ranges of the inner double integral are usually wide, which means the number of the evaluations of the integrand can be very large if we iteratively apply an adaptive quadrature to handle the double integral. There are various ways of avoiding the iterative evaluation method in multiple integrations. Among others, quasi Monte Carlo method "may perhaps be superior to the other methods" when the dimensionality is high (maybe ≥ 5) (quoted from Stroud (1972) p.194). The iterative evaluation method is among the "other methods" referred to by Stroud (1972).

Another quite different approach to computing multivariate normal probabilities is to utilize the Fourier series expansion on the indicator function of the original range to be integrated, with which the new range become $\mathbb{R}^n$. (see Russell, Farrier and Howell (1985), they called the indicator function a window function) For the positive-factor correlations, this approach, which involves $\prod_{j=1}^n [2(m_j)]$ manipulations of trigonometric functions, might not be feasible when compared with other methods. Note that $m_j+1$ is the number of terms used in the series expansion of the window function; the higher the $m_j$, the better the accuracy. For the negative-factor correlations, this approach might be better than the iterative evaluation method, provided one can obtain the discrete Fourier series coefficients very fast. The reason is both the approximated window function and the oscillating integrand are functions of trigonometric functions, so it is nature to combine them together.

Suppose one is more concerned about CPU time and can afford a slightly lower accuracy (for instance, computing the p-values in hypothesis testing). Hsu (1989) has
proposed a one-factor approximation method, which has been shown to be quite satisfactory through a large scale simulation study in the GLM setting. The idea is to treat the known correlation matrix, say $R$, as a sample correlation matrix and use factor analysis methods to extract the one-factor pattern, say $f$. Then, $R_+ = D + ff'$, where $D$ is diagonal such that $R_+$ is a correlation matrix, is the closest one-factor approximation to $R$, provided a positive factor is closer to $R$ than a negative factor (The closeness is defined differently according to different factor analysis methods).

In the case that we know a negative factor is closer to $R$, Hsu (1989) has shown one possible way to get the negative factor by utilizing the factor analysis method. The procedure goes as follows. First, use the generalized least square method to get the uniqueness vector (say $u$) and the one-factor pattern (say $f$) on $R^{-1}$, the inverse of $R$. Then, $R_- = \text{the inverse of } [\text{diag}(u) + ff']$ is the proposed closest one-factor approximation of $R$. Note that $R_-$ might not be a correlation matrix, however, it is the matrix closest to $R^{-1}$ in the sense that $\| R^{-1} R_- - I \|$ is minimal, where $\| A - B \| = \left( \sum_{ij} (a_{ij} - b_{ij})^2 \right)^{1/2}$. In practice, if we do not know whether a positive or negative factor is closer to $R$, then choose $R_+$ if $\| R R_+^{-1} - I \| \leq \| R^{-1} R_- - I \|$, otherwise choose $R_-^{-1}$.

Applications

The applications of the results for two-factor structures were mentioned in Chapter I. For example, in the multinomial sampling model, if one is interested in comparing the treatments-versus-control effects (say, all cell probabilities minus the last cell probability) then the underlying correlation matrix has a one-positive-one-negative-factor structure. We will defer the discussion of these applications until Chapter IV. In the remaining sections, we shall show, in detail, the applications of equation (3.0) or (3.3) for computing multivariate normal or multivariate $t$ probabilities, when the underlying correlation matrix has a two-positive-factor or two-negative-factor structure.
One-way Design with One Covariate

Consider the model $Y_{jk} = \mu_j + \gamma(X_{jk} - \bar{X}_{++}) + e_{jk}$; $j = 1, \ldots, n$; $k = 1, \ldots, m_j$, where $\mu_1, \ldots, \mu_n$ are the treatment effects, $\gamma$ is the common slope to associate $X$ and $Y$, $\bar{X}_{++}$ is the grand average of $X_{jk}$, and $e_{jk}$ are iid $N(0, \sigma^2)$ with an unknown $\sigma^2$. Let

$$N = \sum_{j=1}^{n} m_j, \quad v = N - n - 1, \quad \bar{Y}_{++} = \sum_{j=1}^{n} Y_{jk}.$$ Then, the least square estimates of $\gamma$ and $\mu_j$, are

$$\hat{\gamma} = \frac{\sum_{j=1}^{n} \sum_{k=1}^{m_j} (Y_{jk} - \bar{Y}_{++})(X_{jk} - \bar{X}_{++})}{\sum_{j=1}^{n} \sum_{k=1}^{m_j} (X_{jk} - \bar{X}_{++})^2}, \quad \hat{\mu}_j = \bar{Y}_{++} - \hat{\gamma}(\bar{X}_{++} - \bar{X}_{++})$$

respectively. Let the unbiased estimate of $\sigma^2$ be

$$S^2 = \sum_{j=1}^{n} \sum_{k=1}^{m_j} [Y_{jk} - \hat{\mu}_j - \hat{\gamma}(X_{jk} - \bar{X}_{++})]^2 / v.$$ Suppose treatments versus control effects, say $\mu_1 - \mu_n, \ldots, \mu_{n-1} - \mu_n$, are of interest. Hsu (1989) has shown that the BLUEs, $\hat{\mu}_1 - \hat{\mu}_n, \ldots, \hat{\mu}_{n-1} - \hat{\mu}_n$, for those parameters have a two-factor decomposition, that is the correlation of those BLUEs has a positive two-factor structure. Let

$$\lambda_j = (\bar{X}_{++} - \bar{X}_{++}) / \left[ \sum_{j=1}^{n} \sum_{k=1}^{m_j} (X_{jk} - \bar{X}_{++})^2 \right]^{1/2}, \quad d_{jj} = 1/m_j + 1/m_n + (\lambda_j - \lambda_n)^2,$$

$\alpha_j = 1/\sqrt{m_n d_{jj}}$, and $\beta_j = (\lambda_j - \lambda_n)/\sqrt{d_{jj}}$. Then, the correlation matrix of $\sigma^2(\hat{\mu}_1 - \hat{\mu}_n, \ldots, \hat{\mu}_{n-1} - \hat{\mu}_n) = (D + \alpha \alpha' + \beta \beta')$. The goal is to compute the critical value $d$, such that

$$P(\| \hat{\mu}_j - \hat{\mu}_n \| \leq ds, j = 1, \ldots, n-1) = 1 - a, \text{ say } 0.95.$$

Starch example

The following example was originally studied by Scheffé (1959) and has also been analyzed by Hsu (1989). The $y$ variable is the breaking strength in grams and $x$ variable is the thickness in $10^{-4}$ inches on tests of 7 types of starch film. (Starch 1 = Canna, 2 = Sweet Potato, 3 = Corn, 4 = Rice, 5 = Dasheen, 6 = Wheat, 7 = Potato) A one-way design with one covariate is fitted to the data. The slope, associating $X$ and $Y$, is assumed to be identical for all starches.

Hsu (1989) has pointed out that the estimates, computed by the MEANS option of PROC GLM in SAS, do not take the covariate into account. Hence, the confidence
intervals computed by SAS are incorrect. The exact critical values were computed at α-level 0.01, 0.05 and 0.1. We found Hsu's one-factor critical values are close to the exact critical values to three significant figures. Other methods, such as Scheffé's method and Sidak's method, give critical values far from the exact values, especially Scheffé's method. (see Table 3.1)

TABLE 3.1: Critical Values by Various Methods for Starch Data

<table>
<thead>
<tr>
<th>alpha</th>
<th>Exact</th>
<th>1-factor</th>
<th>Sidak</th>
<th>Scheffe</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>2.26231</td>
<td>2.26424</td>
<td>2.42486</td>
<td>3.32609</td>
</tr>
<tr>
<td>0.05</td>
<td>2.55919</td>
<td>2.56063</td>
<td>2.69296</td>
<td>3.63812</td>
</tr>
<tr>
<td>0.01</td>
<td>3.15329</td>
<td>3.15406</td>
<td>3.24506</td>
<td>4.25630</td>
</tr>
</tbody>
</table>

Detection of a Single Outlier in the Simple Linear Regression Model

Consider the simple linear regression model \( Y_j = \gamma_0 + \gamma_1 x_j + e_j; j = 1, \ldots, n \), where the \( e_j \) are iid \( N(0, \sigma^2) \) with an unknown \( \sigma^2 \). Without loss of generality, let us assume the \( x_j \) are already centered and are not identical. Let \( \hat{\gamma}_0 \) and \( \hat{\gamma}_1 \) be the least square estimates of \( \gamma_0 \) and \( \gamma_1 \). One can estimate \( e_j \) by \( \hat{e}_j = Y_j - (\hat{\gamma}_0 + \hat{\gamma}_1 x_j) \), called the residual of the \( j \)th observation. To detect an outlier in this model, it is sensible to examine the relative sizes of \( \hat{e}_j \). That is, \( \hat{e}_j \) have to be standardized so that they have approximately unit variance.

Let \( Y' = (Y_1, Y_2, \ldots, Y_n) \), and let \( \hat{e}' \) be the corresponding vector of residuals. Let \( X' = \begin{bmatrix} 1 & x_1 & 1 & x_2 & \cdots & 1 & x_n \end{bmatrix} \) and \( H = [h_{jk}] = X(X'X)^{-1}X' \). Then, \( \hat{\epsilon} = (I_n - H)Y \), where \( I_n \) is the identity matrix. The \( j \)th diagonal element of \( H \) is called the leverage of the \( j \)th observation. As pointed out by Behnken and Draper (1972), the spread of \( h_{jj} \) reflects the nonhomogeneity of \( x_j \). Because \( E(\hat{\epsilon}) = 0 \) and \( V(\hat{\epsilon}) = \sigma^2 (I_n - H) \), \( h_{jj} \) is also related to \( V(\hat{\epsilon}_j) \) in the sense that the larger the leverage of \( \hat{e}_j \) the smaller the variance of \( \hat{\epsilon}_j \). We shall show
that \( \hat{e} \) is multivariate normal with zero means and a two-negative-factor correlation matrix. Therefore, the results for the two-negative-factor structures can be applied to compute the p-value in a hypothesis testing of rejecting the null hypothesis that the observation with the largest residual is an outlier, provided that \( \sigma \) is known or, if not, one can derive an estimator of \( \sigma \) independent of \( \hat{e}_j \).

Various methods have been proposed to rescale \( \hat{e}_j \). One of the most widely used methods is to divide \( \hat{e}_j \) by \( s \sqrt{1-h_{jj}} \) resulting in internally studentized residual, where \( s^2 \) equals the sum of squares of \( \hat{e}_j \) divided by \( n-2 \). The detection rule is to declare the observation that has the most extreme internally studentized residual, say \( r^* \), as an outlier if \( r^* > d \), where \( d \) is to be determined so that the investigator can control the desired error rate.

Let \( r_j \) be the internally studentized residual of the \( j \)-th observation. The distribution of \( r_1, \ldots, r_n \) is known as Inverted Student function, which has been studied independently by two authors. Stefansky (1972) has examined the joint distribution of \( r_{j1}, \ldots, r_{jk} \), any subset of \( r_1, \ldots, r_n \), under the assumption that the \( r_j \) have a common variance. Ellenberg (1973) has proved the same result for all cases (without the common variance assumption). Therefore, the critical value \( d \) can be exactly determined theoretically.

Without loss of generality, let \( r_{j1}, \ldots, r_{jk} \) be the first \( k \) of the \( r_j \)'s. Let \( M \) be the corresponding \( k \) by \( k \) principle minor of \( V(\hat{e}) \). Suppose \( M^{-1} \) exists. Then the joint density function of \( r_1, \ldots, r_k \) is

\[
\left[ \Gamma((v+k/2)/(2k)) \right]^{1/2} \prod_{j=1}^{k} \sqrt{m_{jj}} \left[ 1 - \sum_{j=1}^{k} \sum_{j'=1}^{h} \sqrt{m_{jj} m_{j'j'}} m_{j'j'} r_j r_{j'} \right]^{v-1} / \left[ \pi^{k/2} \Gamma(v) \right],
\]

provided \( \sum_{j=1}^{h} \sum_{j'=1}^{h} \sqrt{m_{jj} m_{j'j'}} m_{j'j'} r_j r_{j'} \leq 1 \), where \( m_{jj} \) are the diagonal elements of \( M \), \( m_{j'j'} \) are the elements of \( M^{-1} \) and \( v = (n-2-k)/2 \).
Direct computation of the critical value $d$ is not easy, especially when $n$ is not small. The reason is that evaluating the distribution function of $r_1, \ldots, r_n$ involves taking the inverse of an $n$ by $n$ matrix; and one has to deal with $n$ integrals. As a result, most of the existing methods for evaluating the exact critical value are based on either simulation results or Bonferroni type inequalities to get conservative bounds. For example, Tietjen, Moore and Beckman (1973) have empirically computed the critical values for various $n$. For each $n$, they generated a large number of sample from which the most extreme internally studentized residuals are computed and ordered. The percentiles of $r^*$ were approximated from the empirical sampling distribution.

Ellenberg (1973) suggested the application of the Bonferroni inequality to obtain upper and lower bound for the critical value. Prescott (1975) partially carried out Ellenberg's suggestion and calculated the critical values for those cases studied by Tietjen, Moore and Beckman (1973). Prescott concluded that the results he got are close to Tietjen, Moore and Beckman's simulation results. Lund (1975) extended Prescott's computation and constructed critical value tables, at significant levels 0.1, 0.05, 0.01, which are the most complete tables in the current literature.

Another equally popular standardization method is to divide $e_j$ by $\sigma(\hat{e}) \sqrt{1-h_{jj}}$, giving what is known as the externally studentized residuals, where $\sigma(\hat{e})^2$ is the sum of squares of the components of $\hat{e}$, without the $j^{th}$ observation, divided by $n - 3$. The externally studentized residual is equivalent to the internally studentized residual in the sense that one is a monotonic transform of the other. Let $t_j$ be the externally studentized residual of the $j^{th}$ observation. The relationship between $t_j$ and $r_j$ is $t_j = r_j \left[ \frac{(n-3)}{(n-2-r_j^2)} \right]^{1/2}$. The advantage of using the externally studentized residual is that each individual $t_j$ has a Student's $t$-distribution with $n-3$ degrees of freedom, because $\hat{e}_j$ is independent of $\sigma(\hat{e})$. We note that the detection rule is to declare the observation with the most extreme
externally studentized residual, say $t^*$, as an outlier if $t^* > d$, where $d$ is to be determined to have the desired error rate.

Let $c^2 = \sum x_j^2$. For all $j = 1, \ldots, n$, let $\alpha_j = 1 / \sqrt{n(1-h_{jj})}$ and $\beta_j = x_j / \sqrt{c^2(1-h_{jj})}$. Then, $V(\hat{e}) = \sigma^2(I_n - H) = \sigma^2(I_n - \mathbf{1}\mathbf{1}'/n - \mathbf{x}\mathbf{x}'/c^2)$. Explicitly, $V(\hat{e}_j) = \sigma^2(1 - h_{jj}) = \sigma^2(1 - 1/n - x_j^2/c^2)$ and $\text{Cov}(\hat{e}_j, \hat{e}_k) = \sigma^2(-1/n - x_jx_k/c^2)$. We have, $\text{Corr}(\hat{e}_j, \hat{e}_k) = -\alpha_j\alpha_k - \beta_j\beta_k$.

Therefore, $\hat{e}$ is multivariate normal with zero means and a two-negative-factor correlation matrix. The exact critical value or p-value can be computed by (3.0), provided the common variance $\sigma^2$ is known; or by an expression similar to (3.3), provided one can come up with another rescaling method that can replace $s$ or $\hat{e}_i$ by a common random variable independent of $\hat{e}_j$, for example, the square-root of the mean residual sum of squares obtained from the other independently fitted data.

**How accurate is the Bonferroni approximation?**

Cook and Prescott (1981) derived a procedure for computing the first order Bonferroni lower bound of the p-value for detecting an outlier in linear models. The lower bound together with the usual Bonferroni upper bound provides a tool to check the accuracy of the approximation. Their method involves evaluating of $1 + n(n-1)$ probabilities, each being an upper tail of an F-distribution.

For some special correlation structures, the number of probability statements to be evaluated can be reduced to a very small number even to zero in case that Bonferroni bound is exact. In the extreme case, the intersection of every pair of the events is empty. For instance, in Huber's data (to be analyzed), the Bonferroni bound is exact so that the number of probability statements need to be evaluated is 0. Generally speaking, if the residual correlations are not extremely large and / or the observed maximal studentized residual is large, the Bonferroni bound is reasonably close to the exact p-value.
Table 3.2 contains 6 pairs of x and y values, taken from Table 1.2 of Atkinson (1985), originally analyzed by Huber (1981). The data set is synthetic; the data points were generated by taking the line \( y = -2 - x \), adding a normal random error with mean zero and standard error 0.6 to the first 5 points and adding a gross error 12 to the last point (see Figure 3.1). Hence, the fit of a simple linear regression model to the first 5 data points is good, as it should be (\( R^2 = 0.9119 \) and nothing looks peculiar on the residual plot in Figure 3.2). However, if all 6 observations are included in the model, the resultant residuals suggest that a quadratic model in \( x \) should be fitted (see Figure 3.3). The quadratic fit is also good (\( R^2 = 0.952 \), the residuals appears to follow a random pattern and the normal quantile-quantile plot appears to close to a straight line; see Figures 3.4, 3.5).

We note the last observation does not have the largest residual (observation 1 does), because the last observation has the largest leverage, and high leverage forces the fitted line close to the observation. However, the last observation does have the highest external studentized residual. In terms of testing whether observation 6 is an outlier, the Bonferroni bound of the p-value is 0.0778, which is exact by Cook and Prescott's (1981) procedure. We see the outlier was not detected at level 0.05. This is one possible example of "abuse of diagnostic plots", cited by Atkinson (1985).
FIGURE 3.1: Scatter Plot for Huber's Data: b: Quadratic Fit
   a(c): Linear Fit with(without) Obs 6

FIGURE 3.2: Plot of Residuals vs X: without Obs 6: Huber's Data
FIGURE 3.3: Plot of Residuals vs X: Linear Fit: Huber's Data

FIGURE 3.4: Plot of Residuals vs X: Quadratic Fit: Huber's Data
FIGURE 3.5: Q-Q Plot of Residuals: Quadratic Fit: Huber's Data
CHAPTER IV

CATEGORICAL DATA

Introduction

In this chapter, multiple comparison procedures for three sampling models, multinomial, binomial, and Poisson, commonly used in categorical data analysis will be studied through examples drawn from the fields of medicine, biology and the social sciences. Whenever possible, a suitable log-linear model is fit to the data and a comparison of the parameters of interest is conducted. Simulation studies are also used where necessary. To incorporate in discrete data analysis our results for one-factor or two-factor correlation structures, all the comparison procedures to be discussed in this chapter are based on large-sample normal approximations. This is necessary because the distribution of the estimators of the parameters of interest based on small samples is too complicated to derive.

It is noteworthy that the majority of the existing comparison methods in discrete data analysis are based on Chi-square approximations, which are essentially Scheffé's projection method. For example, the procedures developed for comparing cell probabilities in a multinomial model by Gold (1963), Quesenberry and Hurst (1964) and Goodman (1965) are based on Chi-square approximations. Bhapkar and Somes's (1976) procedure for comparing marginal probabilities in the matched binomial model (to be studied later) is also based on a Chi-square approximation. Goodman's (1964)
method for comparing certain sets of un-redundant parameters in a log-linear model with Poisson error is based on a Chi-square approximation as well.

Goal of the Chapter

The main purpose of this chapter is to describe the comparison methods for comparing the set of the parameters of interest which provide narrower simultaneous confidence intervals than does Scheffé's projection method. In general, if the treatments-versus-control comparisons are of interest, then the confidence intervals can be made narrower by applying our results for one-factor or two-factor structures. If all pairwise comparisons are desired, then the Tukey-Kramer method can be applied to obtain shorter confidence intervals, provided that the variances of the estimates of the mean differences satisfy certain constraints (Hayter (1985)). We shall start our discussions with the comparison methods in the multinomial sampling model.

Multinomial Sampling Model

Let \( n_1, \ldots, n_k \) be the observed cell counts in a sample of size \( N = \sum_{j=1}^{k} n_j \) from a multinomial distribution with cell probabilities \( \pi_1, \ldots, \pi_k \), where \( \pi_j > 0 \) and \( \sum_{j=1}^{k} \pi_j = 1 \).

For all \( j = 1, \ldots, k \), suppose \( E(n_j) = N \pi_j \) is sufficiently large, or \( k \) is fairly large if some of the \( E(n_j) \) are small. Let \( p_j \) denote \( n_j / N \), the observed cell proportion in the \( j \)th cell. Then, 

\[
\sqrt{N}(p - \pi) \sim \mathcal{N}(0, \Sigma),
\]

where \( \Sigma = \text{diag}\{d_{11}, \ldots, d_{kk}\} \) with \( d_{jj} = 1 + \frac{1}{N} \). In practice, if the \( \pi_j \)'s are unknown then a consistent estimator of \( \Sigma \), denoted by \( \hat{\Sigma} \), may be obtained by replacing each \( \pi_j \) with \( p_j \). It is well-known that 

\[
N(p - \pi) \hat{\Sigma}^{-1}(p - \pi)
\]

has an asymptotic Chi-square distribution with \( k-1 \) degrees of freedom.
Suppose that comparisons of cell probabilities are of interest. Scheffé's projection method provides a set of simultaneous confidence intervals of the form:
\[ \pi_j \in \rho_j \pm \left( \chi^2_{k-1}(\alpha) \right)^{1/2} \frac{[\rho_j(1-\rho_j)/N]^{1/2}}{2}, \]
where \( \chi^2_{k-1}(\alpha) \) is the upper 100(\(\alpha\))th percentile of the Chi-square distribution with \( k - 1 \) degrees of freedom. Quesenberry and Hurst (1964) derived another set of confidence intervals for the \( \pi_j \)'s based on the Pearson's Chi-square statistic \[ \sum_{j=1}^{k} \left[ \left( n_j - N \pi_j \right)^2 / N \pi_j \right], \]
which is asymptotically distributed as a Chi-square with \( k - 1 \) degrees of freedom. The intervals proposed by Quesenberry and Hurst can be obtained by finding the two values of \( \pi_j \) which are solutions to the quadratic equation: \[ (\rho_j - \pi_j)^2 = \chi^2_{k-1}(\alpha) \pi_j (1-\pi_j) / N. \]
These intervals, which we will refer to as Pearson intervals, are of the form:
\[ \pi_j \in \left[ \chi^2_{k-1}(\alpha) / 2n_j \pm \left( \chi^2_{k-1}(\alpha) + 4n_j (N - n_j) / N \right)^{1/2} / \left( 2[N + \chi^2_{k-1}(\alpha)] \right) \right]. \]

Goodman (1965) proved that the above intervals can be improved upon by applying the Bonferroni method, that is, by replacing \( \chi^2_{k-1}(\alpha) \) with \( \chi^2_{1}(\alpha/k) \) in the above expression. The resulting intervals will be referred to as Bonferroni intervals. Goodman compared the lengths of Bonferroni and Pearson intervals and concluded that the ratio of the two lengths approaches \[ \left( \chi^2_{1}(\alpha/k) / \chi^2_{k-1}(\alpha) \right)^{1/2} \] in probability, as \( N \) goes to infinity.

**TABLE 4.1: Ratios of the Lengths of Bonferroni to Pearson Intervals**

<table>
<thead>
<tr>
<th>k</th>
<th>0.005</th>
<th>0.010</th>
<th>alpha 0.025</th>
<th>0.050</th>
<th>0.100</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1.077</td>
<td>1.090</td>
<td>1.114</td>
<td>1.144</td>
<td>1.192</td>
</tr>
<tr>
<td>3</td>
<td>0.966</td>
<td>0.967</td>
<td>0.971</td>
<td>0.978</td>
<td>0.992</td>
</tr>
<tr>
<td>4</td>
<td>0.901</td>
<td>0.898</td>
<td>0.894</td>
<td>0.894</td>
<td>0.897</td>
</tr>
<tr>
<td>5</td>
<td>0.854</td>
<td>0.848</td>
<td>0.841</td>
<td>0.836</td>
<td>0.834</td>
</tr>
<tr>
<td>6</td>
<td>0.817</td>
<td>0.809</td>
<td>0.800</td>
<td>0.793</td>
<td>0.788</td>
</tr>
<tr>
<td>7</td>
<td>0.786</td>
<td>0.778</td>
<td>0.767</td>
<td>0.758</td>
<td>0.751</td>
</tr>
<tr>
<td>8</td>
<td>0.760</td>
<td>0.751</td>
<td>0.739</td>
<td>0.729</td>
<td>0.721</td>
</tr>
<tr>
<td>9</td>
<td>0.737</td>
<td>0.728</td>
<td>0.714</td>
<td>0.704</td>
<td>0.695</td>
</tr>
<tr>
<td>10</td>
<td>0.717</td>
<td>0.707</td>
<td>0.693</td>
<td>0.682</td>
<td>0.672</td>
</tr>
<tr>
<td>15</td>
<td>0.641</td>
<td>0.630</td>
<td>0.615</td>
<td>0.603</td>
<td>0.591</td>
</tr>
<tr>
<td>20</td>
<td>0.590</td>
<td>0.579</td>
<td>0.563</td>
<td>0.551</td>
<td>0.538</td>
</tr>
<tr>
<td>25</td>
<td>0.551</td>
<td>0.540</td>
<td>0.525</td>
<td>0.512</td>
<td>0.500</td>
</tr>
<tr>
<td>30</td>
<td>0.520</td>
<td>0.510</td>
<td>0.494</td>
<td>0.482</td>
<td>0.470</td>
</tr>
</tbody>
</table>
Table 4.1 shows some of these ratios for $2 \leq k \leq 30$. The length of Bonferroni interval is only half the length of Pearson interval if $k$ is large (maybe $> 27$) and, except for $k=2$, Bonferroni intervals are shorter than Pearson intervals.

The existing methods for the comparison of cell probabilities require the assumption that the vector of estimates of cell probabilities is approximately multivariate normal in order to obtain Chi-square approximations. Since the correlation matrix $\Sigma$ of the multivariate normal distribution has a one-negative-factor structure, we propose that a set of narrower confidence intervals be obtained by using equation (2.0) to compute the asymptotic critical value, say $q_l$, and by replacing $\chi^2_{k-1}(\alpha)$ with $q_l^2$ in the expression of Pearson intervals. These proposed intervals will be referred to as Factor intervals. We note that Factor intervals are shorter than those based on Chi-square approximations because the critical value, $q_l$, is exact under large-sample normal assumption, while the critical values computed by Chi-square approximations are conservative. We should note that each $\pi_j$ is estimated by $p_j$ to get the factor pattern, since the $\pi_j$'s are unknown.

It can be shown that the ratio of the lengths of Factor intervals to Bonferroni intervals approaches $q_l / [\chi^2_1(\alpha/k)]^{1/2}$ in probability, as $N$ goes to infinity. Since the value of $q_l$ depends on the actual factor pattern $\lambda$, it is not a good practice to compute the ratios in general. However, the ratio is less than one unless $k$ is large (maybe $\geq 7$), in which case the ratio can be close to one because of the nature of the restriction on the negative correlation structures.

Gold (1963) derived a set of simultaneous confidence intervals for all linear combinations $c'\pi$, where $c$ is a vector of fixed constants. These intervals can be obtained by Scheffé's projection method. Suppose the comparisons of all cell probabilities minus the last cell probability, $\pi_k$, are of interest. Then, the intervals proposed by Gold are:

$$
\pi_j - \pi_k \in [p_j - p_k] \pm \left[\chi^2_{k-1}(\alpha)\right]^{1/2} \left\{\frac{[p_j + p_k - (p_j - p_k)^2]}{N}\right\}^{1/2}, \text{ for all } j=1,\ldots,k-1.
$$
Goodman (1965) applied the Bonferroni method to derive a set of narrower intervals by replacing \( \chi^2_{k-1}(\alpha) \) with \( \chi^2_{1}(\alpha/k) \) in the above expression.

Now, we shall show that the correlation matrix of \( p_j - p_k \), for all \( j = 1, \ldots, k-1 \), has a one-positive-one-negative structure and, hence, our two-factor results can be applied to improve upon the confidence intervals proposed by Goodman. For all \( j = 1, \ldots, k-1 \), let

\[
d_{jj} = \pi_j + \pi_k - (\pi_j - \pi_k)^2 \\
y_j = \frac{(p_j - p_k - (\pi_j - \pi_k))}{\sqrt{d_{jj}}}
\]

It can be shown that \( Y = (Y_1, \ldots, Y_{k-1}) \) is asymptotically multivariate normal with zero means and a correlation matrix \( \Sigma = D - \lambda \lambda' + \beta \beta' \), a non-singular one-negative-one-positive-factor structure, where

\[
D = \text{diag}(d_{jj}: j = 1, \ldots, k-1) \quad \lambda = (\lambda_1, \ldots, \lambda_{k-1}) \quad \text{with} \quad \lambda_j = (\pi_j - \pi_k) / \sqrt{d_{jj}} \quad \text{and} \quad \beta = (\beta_1, \ldots, \beta_{k-1}) \quad \text{with} \quad \beta_j = \sqrt{\pi_k} / \sqrt{d_{jj}}.
\]

We propose equation (3.0) be used to compute the asymptotic critical values, say \( l_1 q_2 \), and the confidence intervals be constructed by replacing \( \chi^2_{k-1}(\alpha) \) with \( l_1 q_2^2 \) in the expression of the intervals proposed by Gold.

**Bank Data**

The following data set (see Johnson and Wichern (1982)) is part of a large marketing survey in which a consultant for the Bank of Shorewood wants to know the proportion of all local savers who use its facilities as their primary savings bank. Three major banks, say A, B, and C, in the area are considered as its competitors. Each individual, contacted in the survey, responded to the following question: Which bank is your primary savings bank? Possible responses were:

Bank Shorewood, Bank A, Bank B, Bank C, Another bank, or No savings.

Those people who have no savings are ignored in the comparison. A sample of \( N=355 \) people with savings accounts produced the categorical frequencies in Table 4.2.
TABLE 4.2: Observed Frequencies of Bank Data

<table>
<thead>
<tr>
<th>Bank</th>
<th>S</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>105</td>
<td>119</td>
<td>56</td>
<td>25</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 4.3 shows the critical values computed by various methods at \( \alpha \)-levels 0.10, 0.05 and 0.01. We note that the critical values are the square roots of \( \chi^2_{k-1}(\alpha) \) and \( \chi^2_{1}(\alpha/k) \) for the Pearson and Bonferroni method respectively, and that Sidak critical values are derived by using Sidak's inequality. We found that Pearson critical values are much bigger than those computed by the other methods, and that Sidak critical values are reasonably close to Factor critical values, because the correlations among the estimates of cell probabilities are small. The correlation matrix of the observed cell proportions is displayed as follows.

\[
\begin{pmatrix}
1 & -0.307 & -0.195 & -0.288 & -0.460 \\
-0.307 & 1 & -0.119 & -0.175 & -0.280 \\
-0.195 & -0.119 & 1 & -0.111 & -0.178 \\
-0.288 & -0.175 & -0.111 & 1 & -0.262 \\
-0.460 & -0.280 & -0.178 & -0.262 & 1
\end{pmatrix}
\]

The largest 5 correlations fall in the interval \([-0.46, -0.26]\), while all the others fall in the interval \([-0.20, -0.11]\).

**TABLE 4.3: Critical Values given by Various Methods for Bank Data**

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Pearson</th>
<th>Bonfer.</th>
<th>Sidak</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>2.789</td>
<td>2.327</td>
<td>2.311</td>
<td>2.285</td>
</tr>
<tr>
<td>0.05</td>
<td>3.080</td>
<td>2.576</td>
<td>2.569</td>
<td>2.551</td>
</tr>
<tr>
<td>0.01</td>
<td>3.644</td>
<td>3.091</td>
<td>3.089</td>
<td>3.085</td>
</tr>
</tbody>
</table>
Table 4.4 displays the 95% simultaneous confidence intervals for all cell probabilities given by various methods. Except for the confidence intervals computed using Pearson's method, the confidence intervals are close to one another, because the standard error of the $p_j$'s are small (< 0.026). The ratio of the length of Factor intervals to Bonferroni intervals is about 0.99.

**TABLE 4.4: 95% Confidence Intervals computed by Various Methods for Bank Data**

<table>
<thead>
<tr>
<th>Bank</th>
<th>Pearson Lower</th>
<th>Pearson Upper</th>
<th>Bonferroni Lower</th>
<th>Bonferroni Upper</th>
<th>Sidak Lower</th>
<th>Sidak Upper</th>
<th>Factor Lower</th>
<th>Factor Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>0.227</td>
<td>0.375</td>
<td>0.238</td>
<td>0.361</td>
<td>0.238</td>
<td>0.361</td>
<td>0.238</td>
<td>0.361</td>
</tr>
<tr>
<td>A</td>
<td>0.263</td>
<td>0.416</td>
<td>0.274</td>
<td>0.402</td>
<td>0.274</td>
<td>0.402</td>
<td>0.275</td>
<td>0.402</td>
</tr>
<tr>
<td>B</td>
<td>0.107</td>
<td>0.226</td>
<td>0.114</td>
<td>0.214</td>
<td>0.114</td>
<td>0.214</td>
<td>0.115</td>
<td>0.213</td>
</tr>
<tr>
<td>C</td>
<td>0.039</td>
<td>0.124</td>
<td>0.043</td>
<td>0.114</td>
<td>0.043</td>
<td>0.114</td>
<td>0.043</td>
<td>0.113</td>
</tr>
<tr>
<td>D</td>
<td>0.093</td>
<td>0.207</td>
<td>0.100</td>
<td>0.195</td>
<td>0.100</td>
<td>0.195</td>
<td>0.100</td>
<td>0.194</td>
</tr>
</tbody>
</table>

Table 4.5 displays the 95% simultaneous confidence intervals for differences (in percent) between each cell probability and the first cell probability computed by various methods. The associated critical values are 3.0802, 2.4981, 2.4909 and 2.2448 for the methods indicated in Table 4.5 respectively. The critical value computed by the Factor method is substantially smaller than those computed by the other methods, partly due to the large correlations which are all in [0.58, 0.73]. The correlation matrix of the estimates of the differences between each cell probability and the first cell probability is displayed as follows.

\[
\begin{pmatrix}
1 & 0.576 & 0.684 & 0.592 \\
0.576 & 1 & 0.715 & 0.648 \\
0.684 & 0.715 & 1 & 0.723 \\
0.592 & 0.648 & 0.723 & 1
\end{pmatrix}
\]

The ratio of the length of Factor intervals to Sidak intervals is about 0.90.
TABLE 4.5: 95% Confidence Intervals (in %) computed by Various Methods for Bank Data: Multiple Comparisons with a Control

<table>
<thead>
<tr>
<th>Bank</th>
<th>Scheffe Lower</th>
<th>Scheffe Upper</th>
<th>Bonferroni Lower</th>
<th>Bonferroni Upper</th>
<th>Sidak Lower</th>
<th>Sidak Upper</th>
<th>Factor Lower</th>
<th>Factor Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - S</td>
<td>-2.404</td>
<td>2.483</td>
<td>-1.942</td>
<td>2.021</td>
<td>-1.937</td>
<td>2.016</td>
<td>-1.742</td>
<td>1.820</td>
</tr>
<tr>
<td>B - S</td>
<td>-2.168</td>
<td>1.892</td>
<td>-1.785</td>
<td>1.509</td>
<td>-1.780</td>
<td>1.504</td>
<td>-1.618</td>
<td>1.342</td>
</tr>
<tr>
<td>C - S</td>
<td>-1.955</td>
<td>1.505</td>
<td>-1.628</td>
<td>1.178</td>
<td>-1.624</td>
<td>1.174</td>
<td>-1.486</td>
<td>1.035</td>
</tr>
<tr>
<td>D - S</td>
<td>-2.134</td>
<td>1.824</td>
<td>-1.760</td>
<td>1.450</td>
<td>-1.755</td>
<td>1.445</td>
<td>-1.597</td>
<td>1.287</td>
</tr>
</tbody>
</table>

Contingency Tables with Multinomial Counts

In this section, we shall discuss multiple comparison procedures for three models in the analysis of contingency tables. For a square table, if one is interested in the comparisons of the differences between symmetric off-diagonal cell probabilities, for example, $\pi_{12} - \pi_{21}$, then the underlying asymptotic correlation matrix has a one-negative-factor structure. For a $2 \times k$ table, if one is interested in comparing the natural logarithms of the cross-product ratios of cell probabilities across either a column or row factor, then the underlying asymptotic correlation matrix has a one-positive-factor structure. Therefore, the results of Chapter II for one-factor structures can be applied in these two cases to derive a set of confidence intervals that are narrower than traditional intervals.

For a general $2 \times k$ table, if one is interested in comparing the natural logarithms of the cross-product ratios: $\theta_{ij} = \log \left[ \pi_{1i} \pi_{ij} / (\pi_{i1} \pi_{1j}) \right], 2 \leq i \leq r, 2 \leq j \leq k$, then the underlying asymptotic correlation matrix does not have an exact one-factor structure nor an exact two-factor structure. However, Hsu's (1989) one-factor approximation method can be applied to obtain approximate confidence intervals. In this case, simulation studies are necessary to check the accuracy of the approximation.
In two-way tables, there does not exist a regular log-linear representation for the model of marginal homogeneity (Bishop, Fienberg and Holland (1975)). However, in a square table, quasi symmetry plus marginal homogeneity implies symmetry. Therefore, the analysis suitable for the model of symmetry is also suitable for the model of marginal homogeneity, if we assume the model of quasi symmetry.

We shall start with a discussion of the comparison of the differences between symmetric off-diagonal cell probabilities in a square table.

**A Square Table**

For cross-classifications in square tables, one often wants to examine the symmetry among the attributes defining the classifications. For example, Table 4.6 contains observations of unaided distant vision from case records of eye tests on 7477 women, aged 30-39, employed in Royal Ordinance factories in Britain. (Stuart (1955), Bishop, Fienberg and Holland (1975) p.284) One of the most interesting questions is whether the data suggests any unbalances in eyesight symmetry. In terms of hypothesis testing, various authors have derived different testing procedures for assessing the fit of the model of symmetry. For example, McNemar's test, Pearson's $\chi^2$ test or Wilk's Likelihood-ratio $G^2$ test may all be used to assess the fit of the model. Here, we would like to construct simultaneous confidence intervals for the parameters: $\pi_{ij} - \pi_{ji}$, for all $i < j$, where $\pi_{ij}$ is the probability of the $(i,j)$ cell in a $4 \times 4$ table. The method can be applied to any square table.
Vision Data

TABLE 4.6: Case Records of Eye-Testing of Women (Aged 30-39) Employees in Royal Ordnance Factories

<table>
<thead>
<tr>
<th>Right Eye Grade (i)</th>
<th>Left Eye Grade (j)</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest (1)</td>
<td>1520</td>
<td>1976</td>
</tr>
<tr>
<td>Second (2)</td>
<td>234</td>
<td>2256</td>
</tr>
<tr>
<td>Third (3)</td>
<td>117</td>
<td>2456</td>
</tr>
<tr>
<td>Lowest (4)</td>
<td>36</td>
<td>789</td>
</tr>
<tr>
<td>Totals</td>
<td>1907</td>
<td>841</td>
</tr>
</tbody>
</table>

Let $n_{ij}$ be the observed frequency of the $(i,j)$ cell, and let $m_{ij}$ be $(n_{ij}+n_{ji})/2$, the MLE of the expected frequency of the $(i,j)$ cell, provided that there are no unbalances in eyesight symmetry. To check the adequacy of the model of symmetry, it is natural to examine the standardized residuals $(n_{ij}-m_{ij})/\sqrt{m_{ij}}$. Table 4.7 shows that most of the positive residuals are above the diagonal, and that most of the negative residuals are below the diagonal. This is a good indication that there might be some differences between the "right eye better" situation and the "left eye better" situation for these women.

TABLE 4.7: Standardized Residuals for Vision Data under the Complete Symmetry Model

<table>
<thead>
<tr>
<th>Right Eye Grade (i)</th>
<th>Left Eye Grade (j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest (1)</td>
<td>Highest (1)</td>
</tr>
<tr>
<td>Highest (1)</td>
<td>0</td>
</tr>
<tr>
<td>Second (2)</td>
<td>-1.012</td>
</tr>
<tr>
<td>Third (3)</td>
<td>-0.319</td>
</tr>
<tr>
<td>Lowest (4)</td>
<td>-2.100</td>
</tr>
<tr>
<td>Totals</td>
<td>1.012</td>
</tr>
<tr>
<td></td>
<td>0.319</td>
</tr>
<tr>
<td></td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Assuming a multinomial sampling model, Bishop, Fienberg and Holland (1975) concluded that the model of symmetry does not fit the data (Pearson's Chi-square $X^2 = 19.11$, Wilk's Likelihood-ratio $G^2 = 19.25$, p-value $\leq 0.0039$ and d.f. = 6) while the model of quasi-symmetry fits somewhat better ($X^2 = 7.26$, $G^2 = 7.27$, p-value $\leq 0.0640$ and d.f. = 3). We will assume that the quasi-symmetry model is appropriate and construct simultaneous confidence intervals for the parameters $\pi_{ij} - \pi_{ji}$ accordingly.

Let $p_{ij}$ be $n_{ij} / N$, where $N = \sum_{i=1}^{4} \sum_{j=1}^{4} n_{ij}$. Under the multinomial sampling model, the asymptotic covariance structure of $p_{ij} - p_{ji}$, for all $i < j$ is: $\text{Var}(p_{ij} - p_{ji}) = \frac{[\pi_{ij} - \pi_{ji}]^2}{N}$, denoted by $d_{ij}$, and $\text{Cov}(p_{ij} - p_{ji}, p_{lm} - p_{ml}) = - (\pi_{ij} - \pi_{ji})(\pi_{lm} - \pi_{ml}) / N$, where $i \neq l$, $j \neq m$. Therefore, the random vector with components: $[(p_{ij} - p_{ji}) - (\pi_{ij} - \pi_{ji})] / \sqrt{d_{ij}}$, for all $i < j$, is asymptotically multivariate normal with zero means and a correlation matrix with one-negative-factor structure. We should note that the unknown $\pi_{ij}$'s can be estimated by the MLEs of the expected frequencies under the quasi-symmetry model. These MLEs can be obtained by various methods, for instance, the Newton-Raphson method or the iterative proportional fitting algorithm (see Bishop, Fienberg and Holland (1975) pp. 289-292).

The 95% simultaneous confidence intervals for $\pi_{ij} - \pi_{ji}$, $i < j$, are shown in Table 4.8. The only interval that does not include zero is the interval for $\pi_{14} - \pi_{41}$, corresponding to the highest grade of vision versus the worst grade. This seems reasonable because the unbalance in eyesight symmetry is most likely to be detected among people who have a very good vision in one eye and very bad vision in the other. However, the lower bound of the confidence interval for $\pi_{14} - \pi_{41}$ is very close to zero, partly due to the small standard error ($\sqrt{d_{14}} = 0.00135$, critical value = 2.63104). In terms of testing the null hypothesis that $\pi_{ij} - \pi_{ji} = 0$, for all $i < j$, a rejection rule is to reject the null hypothesis if the maximum of $l[(p_{ij} - p_{ji}) - (\pi_{ij} - \pi_{ji})] / \sqrt{d_{ij}}$, denoted by $r^*$, is large. With the observed $r^* = 2.97094$, the p-value of this test is 0.01781, presenting evidence...
against the null hypothesis. We should note that the Bonferroni intervals are found to be close to the intervals given by our proposed method.

**Table 4.8:** 95% Confidence intervals for all symmetrical pair differences of off-diagonal cell probabilities in Vision Data

<table>
<thead>
<tr>
<th>Conf. Bound</th>
<th>Pair differences of cell probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(12) - (21)   (13) - (31)   (14) - (41)   (23) - (32)   (24) - (42)   (34) - (43)</td>
</tr>
<tr>
<td>Lower</td>
<td>-0.0036   -0.0045   0.0005   -0.0006   -0.0050   -0.0034</td>
</tr>
<tr>
<td>Upper</td>
<td>0.0121    0.0064    0.0076    0.0193    0.0039    0.0104</td>
</tr>
</tbody>
</table>

**A 2 x k Table**

Consider a 2 x k table, in which comparisons of the natural logarithm of the cross-product ratios are of interest. Let $\theta_j = \log \left[ \frac{\pi_{1j} \pi_{2j}}{\pi_{1j} \pi_{2j}} \right]$, $2 \leq j \leq k$, be the treatments-versus-control comparisons of the log odds ratios, $\log(\pi_{2j}/\pi_{1j})$, with the control, $\log(\pi_{11}/\pi_{21})$. Let $n_{ij}$ and $m_{ij}$ be the observed and the expected frequencies of the (i,j) cell respectively. Plackett (1962) noted that the log $n_{ij}$'s can be regarded as asymptotically (as N goes to infinity) uncorrelated with the variance $1/m_{ij}$, provided that the analysis is conducted on the basis of the contrasts among the log $n_{ij}$'s. This result can be justified by examining the information matrix of the log$n_{ij}$'s as follows.

Let $\pi = (\pi_1, \ldots, \pi_{2k-1})$ be the vector of the $\pi_{ij}$'s excluding $\pi_{2k}$, and let $n$ and $m$ be the corresponding vectors of the observed and expected cell frequencies respectively. The domain of the variation of the $\pi_{ij}$'s is a (2k-1)-dimensional simplex, say $(\pi: \pi_1 > 0, \ldots, \pi_{2k-1} > 0, \text{ and } \sum_{i=1}^{2k-1} \pi_i < 1)$, without loss of generality. Now, the covariance matrix of $n$, $\operatorname{Var}(n)$, equals $[\operatorname{diag}(\pi_1, \ldots, \pi_{2k-1}) - \pi \pi']/N$ and the information matrix of $\pi$, $I(\pi)$, equals $N \left[ \operatorname{diag}(1/\pi_1, \ldots, 1/\pi_{2k-1}) + 11'\pi_0 \right]$, where $\pi_0 = \pi_{2k}$ and $\pi_0$ can be regarded as a fixed constant. Let $\theta$ be a vector with components being the natural
logarithms of the components of \( \pi \). Then \( I(\theta) = \frac{\partial \pi'}{\partial \theta} I(\pi) \frac{\partial \pi}{\partial \theta} = \{ \tau_{ij} \} \), where \( \tau_{ii} = N (1/\pi_i + 1/\pi_0) \pi_i^2 \to m_i \) and \( \tau_{ij} = N \pi_i \pi_j / \pi_0 \to 0 \), as \( N \to \infty \). Therefore, the asymptotic covariance matrix of the log \( n_{ij} \)'s is \( \text{diag} \{ 1/m_1, \ldots, 1/m_{2k-1} \} \).

For all \( 2 \leq j \leq k \), let \( Y_j = \log \left[ n_{1j} n_{2j} / (n_{21} n_{1j}) \right] \). Using Plackett's result, we have \( \text{Var}(Y_j) = (1/\pi_{2j} + 1/\pi_{21} + 1/\pi_{1j} + 1/\pi_{11}) / N \) and \( \text{Cov}(Y_j, Y_h) = (1/\pi_{11} + 1/\pi_{21}) / N \). Let \( Y = (Y_2, \ldots, Y_k) \) and let \( \theta \) be the corresponding vector of the \( \theta_j \)'s. Then, \( Y \) is multivariate normal with mean \( \theta \) and a correlation matrix with a one-positive-factor structure. Therefore, the asymptotically simultaneous confidence intervals for the \( \theta_j \)’s can be obtained using our one-factor results. The following example illustrates this application.

### Dowdall's Data

<table>
<thead>
<tr>
<th>Attitude</th>
<th>Ethnic Origin</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Favorable (n)</td>
<td>29</td>
<td>78</td>
<td>56</td>
<td>43</td>
<td>53</td>
<td>43</td>
<td>36</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfavor.</td>
<td>7</td>
<td>47</td>
<td>29</td>
<td>29</td>
<td>32</td>
<td>30</td>
<td>22</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable (m)</td>
<td>23.4</td>
<td>79.7</td>
<td>54.4</td>
<td>46.2</td>
<td>54.4</td>
<td>46.8</td>
<td>37.3</td>
<td>41.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfavor.</td>
<td>13.6</td>
<td>46.3</td>
<td>31.6</td>
<td>26.8</td>
<td>31.6</td>
<td>27.2</td>
<td>21.7</td>
<td>24.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable (r)</td>
<td>2.14</td>
<td>-0.25</td>
<td>0.5</td>
<td>-0.7</td>
<td>-0.22</td>
<td>-0.85</td>
<td>-0.23</td>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfavor.</td>
<td>-2.14</td>
<td>0.25</td>
<td>-0.5</td>
<td>0.7</td>
<td>0.22</td>
<td>0.85</td>
<td>0.23</td>
<td>-0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The first portion of Table 4.9 gives the results of a survey conducted in Rhode Island in 1968 and 1969 by the Population Research Laboratory of Brown University to assess the effect of ethnic background on role attitudes of women aged 15 to 64 (see Dowdall
(1974) and Haberman (1978) pp.109-112). As a part of the study, the individuals interviewed were asked the question, "Do you think it is all right for a married women to have a job instead of only taking care of the house and the children while her husband provides for the family?" The responses to this question are presented in the first two rows of Table 4.9.

To reduce asymptotic estimation bias, Haldane (1955) and Anscombe (1956) suggested the addition of 0.5 to all of the $n_{ij}$'s (also see Gart and Zweifel (1967)). Haberman (1978) noted that the model of independence ($\pi_{ij} = \pi_{iv} \pi_{v+j}$, with $\pi_{iv}$ and $\pi_{v+j}$ being the sum of the $\pi_{ij}$'s over $j$ and $i$ respectively) fits the data set quite well ($X^2 = 6.03, G^2 = 6.48$, p-value $\approx 0.50$ and d.f. = 7). Under the independence model, the adjusted residual of the ($i,j$) cell without Anscombe's continuity correction is $r_{ij} = (n_{ij} - n_{i+} n_{+j}/N) / [n_{i+} n_{+j} (1 - n_{i+}/N)(1 - n_{+j}/N)/N]^{1/2}$, which is approximately standard normal (see Haberman (1973)). The adjusted residuals with Anscombe's continuity correction (see Table 4.9) show that responses of women of Portuguese origin may differ from those of other women surveyed ($r_{11} = 2.14$). So, it is natural to investigate whether this observation is true. We note that $r_{1j} + r_{2j} = 0$ for all $j$ and, hence, it suffices to confine attention to the $r_{1j}$'s.

Haberman (1978) remarked that the large value of "the observed adjusted residual for women of Portuguese origin is not that surprising." The reason, cited by Haberman (1978), is "because the expected number of ethnic origins $j$ for which $|r_{ij}| \geq 2.2$ is 0.11 ($P[|r_{11}| \geq 2.2 ] = 0.014$ and 8 [0.014] = 0.11)." We note that the $r_{ij}$'s computed by Haberman are different from those given in Table 4.9 because he did not use Anscombe's continuity correction. We shall confirm Haberman's conclusion by actually constructing the confidence intervals for the $\theta_j = \log [\pi_{1j} \pi_{2j}/(\pi_{21} \pi_{1j})]$, $2 \leq j \leq 8$, under the model of independence.
Table 4.10 displays the 95% simultaneous confidence intervals for the $\theta_{ij}$'s. Although all of the intervals include zero, the upper bounds of the intervals are at least five times larger in absolute value than the corresponding lower bounds, with the difference being especially large for the Irish versus Portuguese origins interval. Thus, women of Portuguese origin do appear to have stronger responses than women of the other origins, although the evidence is not statistically significant. One might conclude that a larger sample should be selected for further investigation. This analysis confirms and makes more apparent Haberman's conclusions stated earlier.

**TABLE 4.10: 95% Confidence Intervals of logits for Dowdall's Data**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper</td>
<td>1.850</td>
<td>1.753</td>
<td>2.042</td>
<td>1.904</td>
<td>2.072</td>
<td>1.988</td>
<td>1.856</td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>-0.116</td>
<td>-0.314</td>
<td>-0.080</td>
<td>-0.162</td>
<td>-0.044</td>
<td>-0.216</td>
<td>-0.302</td>
<td></td>
</tr>
</tbody>
</table>

**An $r \times k$ Table**

Consider a general two-way $r \times k$ table in which comparisons of the log cross-product ratios: $\theta_{ij} = \log \left[ \pi_{11} \pi_{ij} / (\pi_{1i} \pi_{1j}) \right]$, $2 \leq i \leq r$, $2 \leq j \leq k$, are of interest. The correlation matrix of the estimates of the $\theta_{ij}$'s does not have an exact one-factor nor a two-factor structure (see Hochberg and Tamhane (1987) p.278). However, one can approximate the correlation matrix by a one-factor structure matrix, and evaluate the critical value by the approximate correlation matrix accordingly (see Hsu (1989)).

To check the accuracy of the approximation, Hsu (1989) has developed a simulation procedure in which the techniques of Control Variates and Common Random Numbers are combined for variance reduction. The simulation method requires a known correlation matrix in order to simulate the actual distribution of the estimates of the
parameters. When applying this method to a $r \times k$ table, where the actual cell probabilities are usually unknown, we need to assume a model for a $r \times k$ table in order to estimate the cell probabilities and, hence, the correlations. Thus, the accuracy in this case is not as good as that with known correlations. We shall illustrate this application through the following example.

Psychiatric Therapy Example

Table 4.11 is taken from Table 2.4 of Haberman (1978), where records of 1442 psychotic patients in the New Haven area were checked and summarized in a two-way table with two polytomous factors: type of psychiatric therapy and diagnostic psychoses. Among other analysis, Haberman investigated the comparisons of a set of eight un-redundant log cross-product ratios on an individual basis. That is, the critical values of univariate normal distributions were applied to construct confidence intervals for the eight log cross-product ratios.

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Type of therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Organic therapy</td>
</tr>
<tr>
<td>Schizophrenic psychoses</td>
<td>344</td>
</tr>
<tr>
<td>Affective psychoses</td>
<td>102</td>
</tr>
<tr>
<td>Alcoholic psychoses</td>
<td>23</td>
</tr>
<tr>
<td>Organic psychoses</td>
<td>80</td>
</tr>
<tr>
<td>Senil psychoses</td>
<td>11</td>
</tr>
</tbody>
</table>

As a caution, Haberman noted that "there is a substantial probability that some parameter is not within its confidence interval." Haberman suggested that Goodman's (1964) method, under which the critical value is approximated by the first order Bonferroni inequality, be used to control the over-all error rate.
As an alternative, we shall apply Hsu (1989) one-factor approximation to obtain a set of shorter confidence intervals than those presented by Haberman. In this case, simulation studies were conducted to show that the critical values computed on the basis of Hsu's one-factor approximation method are reasonably close to the actual critical values, computed under the assumption that the estimated correlation matrices are close to the true correlation matrices.

To reduce asymptotic estimation bias, we have added 0.5 to the $n_{ij}$'s. Thus, the estimated log cross-product ratios, $\hat{\theta}_{ij}$, become log \[ \frac{(n_{11}+0.5)(n_{ij}+0.5)}{[(n_{11}+0.5)][(n_{ij}+0.5)]} \]. The schizophrenic diagnostic group may be used as the control group for the row factor, diagnostic group, arises since most patients are in this group. The selection of organic therapy as the control for the column factor, type of therapy, reflects the fact that choices, in the practice of psychiatric therapy, are often made between psychotherapy and organic therapy or between organic therapy and custodial care.

Since the true cell probabilities are unknown, the true correlation matrix has to be estimated. This is complicated by the fact that we do not know of a model which fits the data. As Haberman pointed out, the model of independence does not fit the data. However, the covariance matrix (and, hence, the correlation matrix) can be estimated by various methods (see Gart and Zweifel (1967)). Haberman (1978) used a set of estimates for the variances of the $\hat{\theta}_{ij}$'s: \[ \text{Var}(\hat{\theta}_{ij})=\frac{1/(n_{ij}+0.5)+1/(n_{ii}+0.5)+1/(n_{jj}+0.5)}{N} \]. Haberman's estimates of the covariances are \[ \text{Cov}(\hat{\theta}_{ij}, \hat{\theta}_{ij'})=\frac{1/(n_{11}+0.5)+\partial_{ii'}/(n_{11}+0.5)+\partial_{jj'}/(n_{1j}+0.5)}{N} \], where $\partial_{ii'}$ and $\partial_{jj'}$ are the Kronecker $\partial$'s. The estimations of the asymptotic standard deviations of the $\hat{\theta}_{ij}$'s for the psychiatric data set are given in parentheses in Table 4.12.
TABLE 4.12: Estimated Log Cross-Product Ratios for Psychiatric Data
Values in parentheses are Estimated Asymptotic Standard Error

<table>
<thead>
<tr>
<th>Diagnostic psychoses compared with Schizophrenic</th>
<th>Psychotherapy versus Organic therapy</th>
<th>Custodial care versus Organic therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective</td>
<td>-0.170 (0.2317)</td>
<td>0.902 (0.1373)</td>
</tr>
<tr>
<td>Alcoholic</td>
<td>1.767 (0.2726)</td>
<td>-0.241 (0.3112)</td>
</tr>
<tr>
<td>Organic</td>
<td>-0.376 (0.2736)</td>
<td>-0.169 (0.1766)</td>
</tr>
<tr>
<td>Senile</td>
<td>1.518 (0.3901)</td>
<td>2.405 (0.3155)</td>
</tr>
</tbody>
</table>

Using Hsu's one-factor approximation, critical values were computed at $\alpha$-levels 0.1, 0.05 and 0.01. To check the accuracy of these approximated critical values, 40000 pairs of $\max|\hat{\theta}_{ij} - \theta_{ij}|/\sqrt{d_{ij}}$ were generated based on the estimated correlation matrix and its one-factor approximation using Hsu's simulation algorithm. The results are shown in Table 4.13. We found that the unbiased estimates (see Hsu (1989)) of the $\alpha$-level under the estimated correlation matrix agree with the nominal $\alpha$-level up to the 3rd decimal place.

TABLE 4.13: Simulation results for Psychiatric Data

<table>
<thead>
<tr>
<th>Nominal Alpha</th>
<th>Critical Value</th>
<th>Est. of True Alpha</th>
<th>95% CI for Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>2.548244</td>
<td>0.10 - 0.00030</td>
<td>(0.0978, 0.1016)</td>
</tr>
<tr>
<td>0.05</td>
<td>2.716002</td>
<td>0.05 + 0.00035</td>
<td>(0.0488, 0.0519)</td>
</tr>
<tr>
<td>0.01</td>
<td>3.220968</td>
<td>0.01 + 0.00045</td>
<td>(0.0096, 0.0113)</td>
</tr>
</tbody>
</table>

The 95% simultaneous confidence intervals for the $\theta_{ij}$'s are displayed in Table 4.14. Although the proposed intervals are wider than those computed by Haberman on an individual basis, the conclusions are basically the same as those drawn by Haberman. First, there is no significant evidence against the hypothesis that the distribution of the
treatment types for organic psychoses does not differ from that for schizophrenic psychoses. Second, when contrasted with organic therapy, custodial care is more common in affective psychoses than in schizophrenic psychoses, and psychotherapy is more common in alcoholic psychoses than in schizophrenic psychoses. The main advantage of the proposed method over Haberman's method is the gain of the control of the over-all error rate.

**TABLE 4.14: 95% Confidence Intervals for the Log Cross-Product Ratios in Psychiatric Data**

<table>
<thead>
<tr>
<th>Diagnostic psychoses compared with Schizophrenic</th>
<th>Psychotherapy versus Organic therapy</th>
<th>Custodial care versus Organic therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective</td>
<td>(-0.799, 0.459)</td>
<td>(0.529, 1.275)</td>
</tr>
<tr>
<td>Alcoholic</td>
<td>(1.026, 2.507)</td>
<td>(-1.086, 0.604)</td>
</tr>
<tr>
<td>Organic</td>
<td>(-1.119, 0.367)</td>
<td>(-0.648, 0.311)</td>
</tr>
<tr>
<td>Senile</td>
<td>(0.458, 2.577)</td>
<td>(1.548, 3.262)</td>
</tr>
</tbody>
</table>

**Binomial Sampling Models**

Consider $k$ treatments with binary responses. Let $n_i$ be the number of observations taken on the $i^{th}$ treatment, of which $Y_i$ are positive. For the $i^{th}$ treatment, let $n_i$ and $p_i$ be the common probability of a positive response and the observed cell proportion respectively. Depending on whether the observations are matched or not, there are two considerations for this sampling model.

First, consider the case when the observations are not matched and the $Y_i$'s are independent. Then, the $Y_i$'s are independent binomial random variables with appropriate parameters, and the model is referred to as the independent binomial sampling model.
Here, we shall investigate the treatments-versus-control comparisons of the $\pi_i$'s. Comparison methods for some transformations of the $\pi_i$'s will be discussed as well.

The second sampling model arises if the $n_i$'s are all equal to a common value, say $n$, and the observations are matched. For example, suppose there are $n$ subjects each of whom was given $k$ different drugs with either a positive or a negative response to each drug. This is the so-called matched binomial sampling model. A statistical problem here is to test the marginal homogeneity of the $k$ treatments. As an alternative, we would like to construct simultaneous confidence intervals for the $\pi_i$'s.

**Independent Binomials**

Let $p = (p_1, \ldots, p_k)$, and let $\pi$ be the corresponding vector of cell probabilities. Let $n = \min(n_i)$ and suppose that as $n$ goes to $\infty$ each $n_i/n$ goes to $\lambda_i$, a constant. Under the independent binomial sampling model, $\sqrt{n}(p - \pi)$ is asymptotically multivariate normal with zero means and covariance matrix $\Sigma = \text{diag}(\pi_i(1-\pi_i)/\lambda_i)$, as $n \to \infty$. Suppose one is interested in the comparisons of $\pi_i - \pi_k$, $i = 1, \ldots, k-1$. Then, the underlying asymptotic correlation matrix has a one-positive-factor structure. Therefore, the results for one-factor structures can be applied to derive a set of confidence intervals that are narrower than the intervals derived by traditional methods.

**Lung Cancer Data**

The data in Table 4.15 are records from four retrospective studies cited by Dorn (1954) (also see Fleiss (1981)). In each study, the number of lung cancer patients and the number of smokers among lung cancer patients were recorded. To examine whether or not the proportions of smokers among lung cancer patients are different in different studies, Fleiss (1981) suggested a Chi-square test of the homogeneity of cell proportions. If cell proportions are found to differ significantly, then the technique of
partitioning Chi-square is applied to isolate sources of significance (see Fleiss (1981)).
As an alternative, we would like to construct simultaneous confidence intervals for all pairwise differences among cell proportions using Tukey-Kramer's comparison method.

Table 4.15: Smoking status among lung cancer patients in 4 studies

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of patients</th>
<th>No. of smokers</th>
<th>% of smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>86</td>
<td>83</td>
<td>0.965</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>90</td>
<td>0.968</td>
</tr>
<tr>
<td>3</td>
<td>136</td>
<td>129</td>
<td>0.949</td>
</tr>
<tr>
<td>4</td>
<td>82</td>
<td>70</td>
<td>0.854</td>
</tr>
<tr>
<td>Total</td>
<td>397</td>
<td>372</td>
<td>0.937</td>
</tr>
</tbody>
</table>

Fleiss (1981) has partitioned all observations into two groups. The first group contains observations from the first 3 studies and the second group contains observations from study 4. He concluded that "the proportion of smokers among the patients in study 4 differed from the proportions in studies 1 to 3," but that "there were no differences among the proportions in studies 1 to 3."

Table 4.16 displays the 95% simultaneous confidence intervals for all pairwise differences of cell proportions. We found that the only interval that does not include zero is the interval for $\pi_2 - \pi_4$ and that the lower bound of the interval for $\pi_1 - \pi_4$ is very close to zero. This not only confirms with Fleiss's conclusions but also identifies precisely which of the first 3 studies have cell proportions might differ from the cell proportion of the last study.
Table 4.16: The 95% simultaneous confidence intervals for all pairwise differences of cell proportions for lung cancer data

<table>
<thead>
<tr>
<th>Studies contrasted</th>
<th>Estimate</th>
<th>Stand. error</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 2</td>
<td>-0.003</td>
<td>0.027</td>
<td>-0.072</td>
<td>0.067</td>
</tr>
<tr>
<td>1 - 3</td>
<td>0.017</td>
<td>0.027</td>
<td>-0.054</td>
<td>0.087</td>
</tr>
<tr>
<td>1 - 4</td>
<td>0.111</td>
<td>0.044</td>
<td>-0.001</td>
<td>0.224</td>
</tr>
<tr>
<td>2 - 3</td>
<td>0.019</td>
<td>0.026</td>
<td>-0.048</td>
<td>0.087</td>
</tr>
<tr>
<td>2 - 4</td>
<td>0.114</td>
<td>0.043</td>
<td>0.003</td>
<td>0.225</td>
</tr>
<tr>
<td>3 - 4</td>
<td>0.095</td>
<td>0.043</td>
<td>-0.017</td>
<td>0.206</td>
</tr>
</tbody>
</table>

Now, let us consider the comparisons of some transformations of the $\pi_i$'s. Let $r = [2\sin^{-1}(\sqrt{p_1}), \ldots, 2\sin^{-1}(\sqrt{p_2})]$, and let $p$ be the corresponding vector with each $p_i$ being replaced by $\pi_i$. Let $n = \min(n_i)$ and suppose that as $n$ goes to $\infty$ each $n_i/n$ goes to $\lambda_i$, a constant. Then, $\sqrt{n}(r - p)$ is asymptotically multivariate normal with zero means and the correlation matrix $\Sigma = \text{diag}(1/\lambda_1, \ldots, 1/\lambda_k)$, as $n \to \infty$. The random vector $r$ is the variance-stabilizing transform of $p$.

Suppose that the comparisons of interest are $\sin^{-1}(\sqrt{\pi_i}) - \sin^{-1}(\sqrt{\pi_k})$, $i = 1, \ldots, k-1$. Then, the asymptotic correlation matrix of the estimators of the above parameters has a one-positive-factor structure. Therefore, a similar procedure for comparing the $(\pi_i - \pi_k)$'s can be applied to derive simultaneous confidence intervals for the transformed parameters. One of the nice properties of the variance-stabilizing transform is that the correlation matrix is known and, hence, need not be estimated. However, there does not exist a one-to-one correspondence between $\pi_i - \pi_k$ and the transformed parameter $\sin^{-1}(\sqrt{\pi_i}) - \sin^{-1}(\sqrt{\pi_k})$. Thus, one can not obtain the corresponding confidence intervals for the $\pi_i - \pi_k$ directly from those for the $\sin^{-1}(\sqrt{\pi_i}) - \sin^{-1}(\sqrt{\pi_k})$. 

For a complete discussion of comparison methods in the independent binomial sampling model, consider all pairwise comparisons of the $\pi_i$'s. Scheffé's projection method can be applied to derive a set of conservative confidence intervals. Alternatively, one can obtain a set of narrower intervals by using the Tukey-Kramer procedure. In addition to the assumptions for the independent binomial sampling model, suppose that each treatment corresponds to a certain dose-level of a common drug and that a linear logistic fit with different slopes for the logits is appropriate. Further suppose that one is interested in the treatments-versus-control comparisons of the slopes. Then, the underlying asymptotic correlation matrix has a one-positive-factor structure. Hence, the simultaneous confidence intervals can be derived using our one-factor results. We shall discuss in detail this linear logistic model later in this chapter.

**Matched Binomials**

Consider $k$ treatments each having $n$ observations with binary responses. Suppose the observations are matched according to certain characteristics. For example, in a controlled clinical-trial study, each of $n$ subjects may be given $k$ different diagnostic tests with either a positive or negative response on each test; or $N = nk$ subjects are grouped together in $n$ sets of $k$ patients, each patient in the same group having a similar characteristic, for instance, sex, age, etc., etc.

Let $\delta_j = 1$ (0), indicating a positive (negative) response to the $j^{th}$ treatment. We can summarize the observations into $2^k$ categories indexed by $(\delta_1, \delta_2, \cdots, \delta_k)$. For all $i=1, \cdots, k$, and $j = 1, \cdots, n$, let $X_{ij} = 1$ (0), if the $j^{th}$ subject has a positive (negative) response to the $i^{th}$ treatment, let $Y_i = \sum_{j=1}^{n} X_{ij}$, the number of positive responses to the $i^{th}$ treatment, and let $Y_{ir} = \sum_{j=1}^{n} X_{ij} X_{rj}$, the number of positive responses to both the $i^{th}$ and $r^{th}$ treatment. Further, let $\pi_i = P(X_{ij} = 1)$, the marginal probability of positive response to
the $i^{th}$ treatment, and $\pi_{ir} = P(X_{ij} = 1, X_{rj} = 1)$, the joint probability of positive responses to both the $i^{th}$ and $r^{th}$ treatment.

Let $p_i = Y_i / n$, then $\sqrt{n} (p_1, \ldots, p_k)$ is asymptotically multivariate normal with zero means and covariance matrix $\Sigma = \{\Sigma_{ij}\}$, where $\Sigma_{ii} = \pi_i (1 - \pi_i)$ and $\Sigma_{ij} = \pi_i \pi_j - \pi_i \pi_j$, $i \neq j$. Let $p_{ir} = Y_{ir} / n$, then $\Sigma$ can be consistently estimated by replacing $\pi_i$ and $\pi_j$ by $p_i$ and $p_j$ respectively in the expressions of $\Sigma_{ii}$ and $\Sigma_{ij}$. Thus, the correlation structure can be quite arbitrary except when $k = 3$, in which case the correlation matrix has a one-factor structure. In terms of testing for marginal homogeneity: $\pi_1 = \cdots = \pi_k$, Cochran (1950) proposed the Q-test procedure and Bishop, Fienberg and Holland (1975) suggested a conditional test procedure using a quasi-symmetry model. Both of the above testing procedures are based on Chi-square approximations. Here, we would like to construct simultaneous confidence intervals for the $\pi_i$'s by using exact one-factor results for $k = 3$ or by using Hsu's one-factor approximation results when $k > 3$.

**Drug Response Data**

To test for drug effect, forty-six subjects, each given drugs A, B, and C, were observed to have either a favorable or an unfavorable response to each drug. The results are summarized in Table 4.17 (see Table 8.3-3 of Bishop, Fienberg and Holland (1975)). The observed marginals are $p_1 = p_2 = 28$, and $p_3 = 16$, an indication that drugs A and B might be more effective than drug C. Both Cochran's Q-test and Bishop, Fienberg and Holland's conditional test show moderately strong evidence against the model of marginal homogeneity (the p-value is between 0.05 and 0.01). Bishop, Fienberg and Holland pointed out that the model of quasi-symmetry does not fit the data.
TABLE 4.17: Drug Responses Data

<table>
<thead>
<tr>
<th>Response to</th>
<th>Number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>A B C</td>
<td></td>
</tr>
<tr>
<td>+ + +</td>
<td>6</td>
</tr>
<tr>
<td>+ + -</td>
<td>16</td>
</tr>
<tr>
<td>+ - +</td>
<td>2</td>
</tr>
<tr>
<td>+ - -</td>
<td>4</td>
</tr>
<tr>
<td>- + +</td>
<td>2</td>
</tr>
<tr>
<td>- + -</td>
<td>4</td>
</tr>
<tr>
<td>- - +</td>
<td>6</td>
</tr>
<tr>
<td>- - -</td>
<td>6</td>
</tr>
</tbody>
</table>

Since \( k = 3 \), the asymptotic correlation matrix can be decomposed exactly into a one-positive-factor structure. Using our one-factor results, the 95% simultaneous confidence intervals for \( \pi_1, \pi_2, \pi_3 \) are \((0.438, 0.780), (0.438, 0.780), (0.181, 0.515)\) respectively. Since the upper bound of the interval for Drug C is not much bigger than the lower bound of the other intervals, it is natural to further investigate the differences: \( \pi_1 - \pi_3 \), and \( \pi_2 - \pi_3 \). The 95% simultaneous confidence intervals for \( \pi_1 - \pi_3 \), and \( \pi_2 - \pi_3 \) are both \((0.026, 0.496)\), which does not include zero. Therefore, the data does suggest that drugs A and B are significantly more effective than drug C.

For a complete discussion of comparison methods in the matched binomial sampling model, consider the comparisons of the contrasts \( c'\pi \), with \( \sum_{j=1}^{k} c_j = 0 \), are of interest. Bhapkar and Somes (1976) derived a set of asymptotically simultaneous confidence intervals for all contrasts \( c'\pi \) by using Scheffé's projection method. If one is interested in the all pairwise comparisons of the \( \pi_1 \)'s, it is not known whether the Tukey-Kramer
procedure can be applied to improve upon Scheffé's projection method except when \( k = 3 \) (see Hochberg and Tamhane (1987)).

**Linear Logistic Model**

Consider \( k \) treatments, each corresponds to certain dose levels, say \( x_{ij} \), \( i=1,\ldots,k \), and \( j=1,\ldots,g_i \), of a common drug. Let us assume that the responses are binary and that the observations are completely independent across different treatments and different dose levels of the common drug. At level \( x_{ij} \) of the \( i^\text{th} \) treatment, let \( \pi_{ij} \) be the probability of a positive response, \( n_{ij} \) be the total number of independent observations and \( Y_{ij} \) be the number of positive responses. Then the \( Y_{ij} \)'s are independent binomial random variables.

The log-likelihood of observing \( \{Y_{ij} | i=1,\ldots,k; j=1,\ldots,g_i \} \) given \( \{n_{ij} | i=1,\ldots,k; j=1,\ldots,g_i \} \) is

\[
\phi(\pi | Y=y) = \sum_{i=1}^{k} \sum_{j=1}^{g_i} \{y_{ij} \log(\pi_{ij}/(1-\pi_{ij})) + n_{ij} \log(1-\pi_{ij})\} + \text{constants}.
\]

Consider now a linear logistic model: \( \theta_{ij} = \log(\pi_{ij}/(1-\pi_{ij})) = \alpha_i + \beta_i x_{ij} \). Then, \( \phi(\alpha, \beta | Y=y) = \sum_{i=1}^{k} \sum_{j=1}^{g_i} \{y_{ij}(\alpha_i + \beta_i x_{ij}) + n_{ij} \log[1/(1+e^{\alpha_i + \beta_i x_{ij}})]\} \). Suppose the treatments-versus-control comparisons of the \( \beta_i \)'s are of interest. We shall show that the asymptotic correlation matrix of the weighted least square estimates (WLSE) of the \( \beta_i \)'s with matrix of weights \( W^{-1} = \text{diag}(\pi_{ij}(1-\pi_{ij})/n_{ij}) \) has a one-positive-factor structure. Hence, a set of narrower confidence intervals than the traditional confidence intervals can be obtained by applying the results for one-factor structures.

We note that the parameters \( \alpha_i \) and \( \beta_i \) can be estimated by various methods such as Fisher's method of scoring, a Newton-Raphson method with the second-order derivative matrix being replaced by its expected value, and Iterative re-Weighted Least Square method (IWLS). It has been noted (McCullagh and Nelder (1983) chapter 2 and section 4.3) that Fisher's method of scoring is equivalent to the IWLS method for some special systematic link functions including the linear logistic link function.
Let $\lambda = (\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_k)$, and let $\hat{\lambda}$ be the vector of the corresponding WLSE's. Let $n = \min(n_{ij})$ and suppose that as $n$ goes to $\infty$ each $n_{ij}/n$ goes to $\lambda_{ij}$, a constant. Then, $\sqrt{n}(\hat{\lambda} - \lambda)$ is asymptotically multivariate normal with zeros means and a covariance matrix $n(X'WX)^{-1}$, where $X$, of order $N \times 2k$, is the design matrix and $N = \sum_{i=1}^k \sum_{j=1}^n n_{ij}$. Explicitly, $\text{Var}(\hat{\alpha}_i) = m_{i0}/(m_{i0}m_{i2} - m_{i1}^2)$, $\text{Var}(\hat{\beta}_i) = m_{i0}/(m_{i0}m_{i2} - m_{i1}^2)$ and $\text{Cov}(\hat{\alpha}_i, \hat{\beta}_i) = -m_{i1}/(m_{i0}m_{i2} - m_{i1}^2)$, where $m_{i0} = \sum w_{ij}$, $m_{i1} = \sum w_{ij}x_{ij}$ and $m_{i2} = \sum w_{ij}x_{ij}^2$.

The asymptotic covariances of the other pairs $(\hat{\alpha}_i, \hat{\alpha}_h)$, $(\hat{\beta}_i, \hat{\beta}_h)$ and $(\hat{\alpha}_i, \hat{\beta}_h)$ are zero, for all $i \neq h$.

In practice, if the subjects who receive the same treatment at the same level have some common characteristic, say, sex, species, etc., then the variance of $Y_{ij}$ is usually larger than $n_{ij}x_{ij}(1-\pi_{ij})$. This phenomenon is known as over-dispersion. In such cases it is convenient to assume that $\text{Var}(Y_{ij}) = \sigma^2 n_{ij}\pi_{ij}(1-\pi_{ij})$, with $\sigma^2 > 1$, and then estimate $\sigma^2$ from the data. We note that, in principle, we might have $\sigma^2 < 1$, a situation of under-dispersion.

In the over-dispersion case, the estimates of $\alpha_i$ and $\beta_i$ can be obtained by the IWLS method with matrix of weights $W^{-1} = \text{diag}(\sigma^2\pi_{ij}(1-\pi_{ij})/n_{ij})$. Hence, $\sqrt{n}(\hat{\lambda} - \lambda)/\sigma$ is asymptotically multivariate normal with zero means and a covariance matrix $n(X'WX)^{-1}$.

The over-dispersion parameter can be estimated by $\hat{\sigma}^2 = T/v$, where

$$T = \sum_{i=1}^k \sum_{j=1}^n (Y_{ij}-n_{ij}\hat{\pi}_{ij})^2 / [n_{ij}\hat{\pi}_{ij}(1-\hat{\pi}_{ij})]$$

with $\hat{\pi}_{ij} = 1 / \{1+\exp[-(\hat{\alpha}_i+\hat{\beta}_j x_{ij})]\}$ and $v = N - 2k$.

It has been noted (McCullagh and Nelder (1983)) that $(N-2k)\hat{\sigma}^2 / \sigma^2$ is asymptotically distributed as a Chi-square random variable with $N-2k$ degrees of freedom and that $\hat{\lambda}$ and $\hat{\sigma}^2$ are asymptotically independent. Therefore, $\sqrt{n}(\hat{\lambda} - \lambda)/\hat{\sigma}$ has an asymptotic multivariate t distribution with a diagonal underlying correlation matrix.
Suppose the comparisons of interest are $\beta_i - \beta_k$, $i = 1, \ldots, k-1$. The asymptotic 95% simultaneous confidence intervals are of the form:

$$
\beta_i - \beta_k \in \hat{\beta}_i - \hat{\beta}_k \pm q^1 \sigma \sqrt{\text{Var}(\hat{\beta}_i) + \text{Var}(\hat{\beta}_k)}, \quad i = 1, \ldots, k-1,
$$

where $q^1$ is the critical value of a two-sided multivariate normal (or multivariate t) distribution with a one-positive-factor underlying correlation structure and an over-all confidence level of 0.95, for $\sigma^2 = (\text{or } >)1$.

**Grasshopper Data**

The data in Table 4.18, complied from Table 10.2 of McCullagh and Nelder (1983), are part of a data set involving challenging the melanoplus sanguinipes grasshopper with the insecticide carbofuran and the synergist piperonyl butoxide (PB) which enhances the toxicity of the insecticide. To examine the relationship between the number of grasshoppers killed, and the dose level of the insecticide and the synergist PB, McCullagh and Nelder assumed a logistic link function and binomial error terms with no over-dispersion and suggested fitting a non-linear model on the logits with two covariates: the natural logarithm of the dose of insecticide and the dose of the synergist PB. We note that they assumed a common slope for each of the covariates. Here, we would like to modify their model by assuming a fixed effect of each dose of the four synergist PB's (three treatments, corresponding to $z= 3.9, 19.5$ and 39, with a control, $z= 0$) and by fitting a non-linear model on the logits with a single covariate: log dose of insecticide. Since the sample proportions of the control group look different from those of the treatment groups, we shall assume different slopes for different treatments and construct simultaneous confidence intervals for the treatments-versus-control comparisons of the slopes.
The model that we propose is: \( \theta_{ij} = \alpha_i + \beta_i \log(x_{ij} - \tau) \), where \( x_{ij} \) is the dose level of the insecticide and \( \tau \) is a non-linear parameter to be estimated. If \( \tau \) were known, then a general linear model would result. To estimate \( \tau \), one can expand \( \log(x_{ij} - \tau) \) about an initial value \( \tau_0 \) to obtain a linear approximation: \( \log(x_{ij} - \tau) \approx \log(x_{ij} - \tau_0) + (x_{ij} - \tau_0)[1/(x_{ij} - \tau_0)] \). Thus, if we fit a linear model to the data: \( \theta_{ij} = \alpha_i + \beta_i \log(x_{ij} - \tau_0) + \gamma [1/(x_{ij} - \tau_0)] \), then \( \tau \) can be estimated by \( \tau_0 + \hat{\gamma}/\hat{\beta}_0 \), where \( \hat{\beta}_0 \) is the average of \( \hat{\beta}_i \). With an initial value \( \tau_0 = 1.0 \), the estimation process converges in five iterations to \( \tau = 1.5016 \) (with a -0.001 increment in \( \tau \) from the previous iteration); we take \( \tau = 1.50 \), for convenience.

To reduce the asymptotic estimation bias, we have added 0.5 to all of the cells (killed or not killed). The results from GLIM, a statistical computation package supported by the Royal Statistical Society, show that the fit is marginally acceptable at level 0.01 (deviance= 17.033, p-value= 0.017 and d.f.= 7), and that the slope of the control is
slightly smaller than those of the treatments. The 95% simultaneous confidence intervals are shown in Table 4.19. All three intervals contain zero. Thus, the slopes of the treatments are not significantly different from the slope of the control. The fit of the model with a common slope for the covariate is slightly better than the fit with different slopes (deviance = 19.315, p-value = 0.036 and d.f. = 10). One might conclude that a further investigation with larger sample sizes (especially, for the lower dose level of the insecticide) is necessary.

<table>
<thead>
<tr>
<th>Treatment Contrast</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 4</td>
<td>0.089</td>
<td>0.2037</td>
<td>(-0.396, 0.574)</td>
</tr>
<tr>
<td>2 - 4</td>
<td>0.282</td>
<td>0.2109</td>
<td>(-0.220, 0.788)</td>
</tr>
<tr>
<td>3 - 4</td>
<td>0.182</td>
<td>0.1980</td>
<td>(-0.289, 0.653)</td>
</tr>
</tbody>
</table>

For a complete discussion of comparison methods in a linear logistic model, consider the treatments-versus-control comparisons of the $\alpha_i + \beta_i x$, for some specified dose level $x$, are of interest. It can be shown that the asymptotic correlation matrix of the estimates $\hat{\alpha}_i + \hat{\beta}_i x - (\hat{\alpha}_k + \hat{\beta}_k x)$ has a one-positive-factor structure. Hence, a set of narrower confidence intervals than the traditional confidence intervals can be obtained by applying a procedure similar to the one used for the treatments-versus-control comparisons of the $\beta_i$. Suppose one is interested in all pairwise comparisons of the $\beta_i$'s or of the $(\alpha_i + \beta_i x)$'s. Reiner (1961) has derived confidence intervals for such comparisons using Scheffé's projection method. It has been noted (Hochberg and Tamhane (1987)) that the Tukey-Kramer procedure can be used to obtain a set of narrower confidence intervals than those derived using Scheffé's projection method.
Poisson Sampling Model

Consider n independent random variables representing counts of events in a Poisson or Poisson-like process, for example, number of suicides (per 100,000 residents) in a country or number of cosmic ray impacts (per second) registered by a cosmic ray detector. In practice, if there is evidence that the variances of the independent random samples are proportional to their means then an independent Poisson model is usually appropriate even though the observations might be measurements, not counts. Here, a statistical problem is to examine the variations of the means (or some transformations of the means) of the independent Poisson observations. In particular, we would like to construct simultaneous confidence intervals for parameters in a log-linear model for the log of the means.

Let $Y_i, i = 1, \ldots, n$, be a set of n independent Poisson random variables with respective means $\lambda_i$. The log-likelihood function of the observed $Y_i$'s is proportional to $\sum_{i=1}^{n} [y_i \log(\lambda_i) - \lambda_i]$. The MLE of $\lambda_i$ is $\hat{Y}_i$, a complete minimal sufficient statistic. In a log-linear model with Poisson error terms, the systematic component is usually of the logarithmic form: $\log(\lambda_i) = X_i' \beta, i = 1, \ldots, n$, where $X_i$ is a vector of explanatory variables (continuous or discrete) and $\beta$ is the vector of associated parameters of interest.

The asymptotic distribution of the random vector $[\log(Y_1) - \log(\lambda_1), \ldots, \log(Y_n) - \log(\lambda_n)]$ is multivariate normal with zeros means and covariance matrix $W^{-1} = \text{diag}(1/\lambda_1, \ldots, 1/\lambda_n)$ (see Bishop, Fienberg and Holland (1975) pp.486-500). The asymptotic variances $1/\lambda_i$ can be estimated by $1/\hat{Y}_i$, where $\hat{Y}_i$ is the fitted $Y_i$ under a similar sensible model, or $\hat{Y}_i$ can be just $Y_i$ if there is no suitable model which fits the data. Hence, the asymptotic distribution of the random vector $(\hat{\beta}_1 - \beta_1, \ldots, \hat{\beta}_n - \beta_n)$ is multivariate normal with zero means and variance-covariance matrix $(X'WX)^{-1}$, the $\hat{\beta}_i$'s are the WLS estimates of the $\beta_i$'s with the matrix of weights $W$. 
Suicide Data by Day of the Week

Table 4.20 contains counts of suicides in France by day of the week; originally studied by Durkheim (1951). Using no formal statistical analysis, Durkheim concluded that the suicide rate seems to decrease at the end of the week, starting on Friday. Haberman (1978) assumed an independent Poisson model and concluded that the model of constant suicide rates on the first four days of the week fits the data quite well. Haberman's model is: \( \log(\lambda_i) = \beta_i \), for \( 1 \leq i \leq 4 \) and \( \log(\lambda_i) = \beta_{i,3} \), for \( 5 \leq i \leq 7 \).

To check Durkheim's conclusion, we shall assume the above model and construct a set of simultaneous confidence intervals for \( \beta_i - \beta_1 \), \( i = 2, 3, 4 \).

**TABLE 4.20: Suicides by Day of the Week**

<table>
<thead>
<tr>
<th>Day index</th>
<th>No. of Suicides: ( Y )</th>
<th>Fitted ( Y )</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>1001</td>
<td>1012.75</td>
<td>15.200</td>
</tr>
<tr>
<td>Tuesday</td>
<td>1035</td>
<td>1012.75</td>
<td>15.713</td>
</tr>
<tr>
<td>Wednesday</td>
<td>982</td>
<td>1012.75</td>
<td>14.908</td>
</tr>
<tr>
<td>Thursday</td>
<td>1033</td>
<td>1012.75</td>
<td>15.682</td>
</tr>
<tr>
<td>Friday</td>
<td>905</td>
<td>905.00</td>
<td>13.739</td>
</tr>
<tr>
<td>Saturday</td>
<td>737</td>
<td>737.00</td>
<td>11.189</td>
</tr>
<tr>
<td>Sunday</td>
<td>894</td>
<td>894.00</td>
<td>13.572</td>
</tr>
</tbody>
</table>

It can be shown that \( (\hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2, \hat{\beta}_3 - \beta_3, \hat{\beta}_4 - \beta_4) \) is asymptotically multivariate normal with zeros means and covariance matrix = \( \text{diag}(1/\lambda_0, 1/\lambda_2, 1/\lambda_3, 1/\lambda_4) \), where \( \lambda_0 = \sum_{i=1}^{4} \lambda_i \). Hence, the asymptotic correlation matrix of \( (\hat{\beta}_2 - \hat{\beta}_1, \hat{\beta}_3 - \hat{\beta}_1, \hat{\beta}_4 - \hat{\beta}_1) \) has a one-positive-factor structure. The 95% simultaneous confidence intervals for \( \beta_i - \beta_1 \), \( i = 2, 3, 4 \), are \((-0.200, -0.025), (-0.413, -0.222)\) and \((-0.213, -0.037)\) respectively. The upper
bounds of all three intervals are less than zero. Therefore, one can conclude that the suicide rates are significantly smaller on the last three days than the first four days of the week. This confirms Durkheim's conclusion. The corresponding bounds for the \( \lambda_i / \lambda_1 \), \( i = 2, 3, 4 \), are (0.819, 0.975), (0.661, 0.801), and (0.808, 0.964) respectively. We note that the bounds for \( \lambda_i / \lambda_1 \) are \( \exp(L_i) \) and \( \exp(U_i) \), where \( L_i \) and \( U_i \) are the lower and upper bounds for \( \beta_1 - \beta_1 \) respectively.

**Suicide Data by Region**

Table 4.21 contains counts of suicides in the United States in 1970 by region of the country in which the suicides occurred. The suicide rates vary from region to region with a low rate of 8.5083 per 100,000 residents in the Middle Atlantic states and a high rate of 17.4756 per 100,000 residents in the Pacific states. Haberman (1978) concluded that the model of constant rates for the Central regions, including East North Central, West North Central, East South Central, and West Central regions, fits the data quite well.

**TABLE 4.21: Suicides by Region**

<table>
<thead>
<tr>
<th>Region</th>
<th>Index</th>
<th>Number of suicides: ( Y )</th>
<th>Estimated expected ( Y )</th>
<th>Enumerated population: ( z )</th>
<th>Rate per 100,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>New England</td>
<td>1</td>
<td>1119</td>
<td>1119.00</td>
<td>11841663</td>
<td>9.450</td>
</tr>
<tr>
<td>Middle Atlantic</td>
<td>2</td>
<td>3165</td>
<td>3165.00</td>
<td>37199040</td>
<td>8.508</td>
</tr>
<tr>
<td>E. N. Central</td>
<td>3</td>
<td>4308</td>
<td>4314.98</td>
<td>40252476</td>
<td>10.720</td>
</tr>
<tr>
<td>W. N. Central</td>
<td>4</td>
<td>1790</td>
<td>1749.38</td>
<td>16319187</td>
<td>10.720</td>
</tr>
<tr>
<td>S. Atlantic</td>
<td>5</td>
<td>3729</td>
<td>3729.00</td>
<td>30671337</td>
<td>12.158</td>
</tr>
<tr>
<td>E. S. Central</td>
<td>6</td>
<td>1335</td>
<td>1372.51</td>
<td>12803470</td>
<td>10.720</td>
</tr>
<tr>
<td>W. S Central</td>
<td>7</td>
<td>2075</td>
<td>2071.13</td>
<td>19320560</td>
<td>10.720</td>
</tr>
<tr>
<td>Mountain</td>
<td>8</td>
<td>1324</td>
<td>1324.00</td>
<td>8281562</td>
<td>15.987</td>
</tr>
<tr>
<td>Pacific</td>
<td>9</td>
<td>4635</td>
<td>4635.00</td>
<td>26522631</td>
<td>17.476</td>
</tr>
</tbody>
</table>
To determine the differences between the expected suicide rates in different divisions of the country, Haberman (1978) suggested the use of the Bonferroni method to construct a set of simultaneous confidence intervals for all pairwise differences of the expected suicide rates (see Goodman (1964)). We note that the form of the intervals suggested by Haberman is: \( \log(f_i / f_j) \in \log(\hat{\lambda}_i / \hat{\lambda}_j) \pm 3.1972 \left( \frac{1}{Y_i} + \frac{1}{Y_j} \right)^{1/2} \), where \( \hat{\lambda}_i = \log(Y_i / z_i) \) with \( z_i \) being the enumerated population of region \( i \) divided by 100,000 and \( f_i = \lambda_i / z_i \). That is, Haberman did not use a fitted model and, hence, the estimates of \( f_i \) and \( \lambda_i \) were directly computed on the basis of the observed \( Y_i \) and \( z_i \). As an alternative, we shall use a fitted model and construct confidence intervals by the Tukey-Kramer procedure.

Hayter (1985) has shown that the Tukey-Kramer procedure can be generalized to obtain a set of narrower confidence intervals, provided that the variance of the estimate of each pair difference, say \( \mu_i - \mu_j \), can be factored as a sum of two positive constants, say \( a_i + a_j \), for all \( i \neq j \) (see Hochberg and Tamhane (1987)). In our problem, the variance of \( \log(\hat{\lambda}_i / \hat{\lambda}_j) \) is \( 1/\lambda_i + 1/\lambda_j \), where \( \hat{\lambda}_i \) is the WLS estimate of \( \lambda_i \). Therefore, the Tukey-Kramer procedure can be applied and the form of the 95% simultaneous confidence intervals is: \( \log(f_i / f_j) \in \log(\hat{\lambda}_i / \hat{\lambda}_j) \pm 3.1042 \left( 1/\lambda_i + 1/\lambda_j \right)^{1/2} \), where \( \hat{\lambda}_i \) is the WLS estimate of \( \lambda_i \) under the assumed model.

The 95% simultaneous confidence intervals derived using the procedure described above are given in Table 4.22. The conclusions based on these intervals are the same as those drawn by Haberman based on Bonferroni intervals except that the lower bound of \( \log(f_6 / f_1) \), the comparison of East South Central and New England, is changed from a negative value to a positive one.
## TABLE 4.22: 95% Confidence Intervals for Relative Suicide Rates
for Different Regions: All Pairwise Multiple Comparison

<table>
<thead>
<tr>
<th>Region i</th>
<th>Region j</th>
<th>log(f[i]/f[j])</th>
<th>lower</th>
<th>upper</th>
<th>f[i]/f[j]</th>
<th>lower</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. Atlantic 2</td>
<td>New England 1</td>
<td>-0.213</td>
<td>0.003</td>
<td>0.808</td>
<td>1.003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.N. Central 3</td>
<td>New England 1</td>
<td>0.022</td>
<td>0.230</td>
<td>1.022</td>
<td>1.259</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.158</td>
<td>0.304</td>
<td>1.172</td>
<td>1.355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W.N. Central 4</td>
<td>New England 1</td>
<td>0.007</td>
<td>0.245</td>
<td>1.007</td>
<td>1.278</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.139</td>
<td>0.324</td>
<td>1.149</td>
<td>1.382</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.N. Central 3</td>
<td>-0.088</td>
<td>0.088</td>
<td>0.916</td>
<td>1.092</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. Atlantic 5</td>
<td>New England 1</td>
<td>0.146</td>
<td>0.358</td>
<td>1.157</td>
<td>1.430</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.282</td>
<td>0.432</td>
<td>1.326</td>
<td>1.540</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.N. Central 3</td>
<td>0.056</td>
<td>0.195</td>
<td>1.058</td>
<td>1.216</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.N. Central 4</td>
<td>0.036</td>
<td>0.216</td>
<td>1.037</td>
<td>1.241</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>New England 1</td>
<td>0.001</td>
<td>0.251</td>
<td>1.001</td>
<td>1.286</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.131</td>
<td>0.331</td>
<td>1.140</td>
<td>1.393</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.N. Central 3</td>
<td>-0.096</td>
<td>0.096</td>
<td>0.908</td>
<td>1.101</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.N. Central 4</td>
<td>-0.112</td>
<td>0.112</td>
<td>0.894</td>
<td>1.118</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Atlantic 5</td>
<td>-0.224</td>
<td>-0.028</td>
<td>0.799</td>
<td>0.973</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>New England 1</td>
<td>0.011</td>
<td>0.241</td>
<td>1.011</td>
<td>1.273</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.143</td>
<td>0.319</td>
<td>1.154</td>
<td>1.375</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.N. Central 3</td>
<td>-0.083</td>
<td>0.083</td>
<td>0.920</td>
<td>1.087</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.N. Central 4</td>
<td>-0.101</td>
<td>0.101</td>
<td>0.904</td>
<td>1.106</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Atlantic 5</td>
<td>-0.211</td>
<td>-0.041</td>
<td>0.810</td>
<td>0.960</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E. S. Central 6</td>
<td>-0.108</td>
<td>0.108</td>
<td>0.898</td>
<td>1.114</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>New England 1</td>
<td>0.340</td>
<td>0.652</td>
<td>1.491</td>
<td>1.919</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.529</td>
<td>0.732</td>
<td>1.698</td>
<td>2.080</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.N. Central 3</td>
<td>0.302</td>
<td>0.497</td>
<td>1.358</td>
<td>1.644</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.N. Central 4</td>
<td>0.287</td>
<td>0.513</td>
<td>1.332</td>
<td>1.670</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Atlantic 5</td>
<td>0.175</td>
<td>0.373</td>
<td>1.191</td>
<td>1.452</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E. S. Central 6</td>
<td>0.280</td>
<td>0.519</td>
<td>1.323</td>
<td>1.681</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.S. Central 7</td>
<td>0.290</td>
<td>0.509</td>
<td>1.337</td>
<td>1.664</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>New England 1</td>
<td>0.511</td>
<td>0.718</td>
<td>1.668</td>
<td>2.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M. Atlantic 2</td>
<td>0.648</td>
<td>0.791</td>
<td>1.912</td>
<td>2.206</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E.N. Central 3</td>
<td>0.423</td>
<td>0.554</td>
<td>1.527</td>
<td>1.741</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.N. Central 4</td>
<td>0.402</td>
<td>0.576</td>
<td>1.494</td>
<td>1.779</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Atlantic 5</td>
<td>0.295</td>
<td>0.431</td>
<td>1.343</td>
<td>1.539</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E. S. Central 6</td>
<td>0.393</td>
<td>0.584</td>
<td>1.482</td>
<td>1.793</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W.S. Central 7</td>
<td>0.407</td>
<td>0.571</td>
<td>1.502</td>
<td>1.770</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mountain 8</td>
<td>-0.008</td>
<td>0.186</td>
<td>0.992</td>
<td>1.204</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The general conclusions are as follows. First, the suicide rates in the Mountain and Pacific regions are generally higher than those in the other areas. Second, the New England and Middle Atlantic regions have the lowest suicide rates, when compared with the other regions. Third, except for the South Atlantic region, the expected suicide rate increases as one proceeds westward.

For a complete discussion of comparison methods in the Poisson sampling models, consider the comparisons of some transformations other than the logarithm transform of the means are of interest, for example, the variance-stabilizing transform \( \sqrt{\lambda_i} \). Then, it is well-known that \( 2(\sqrt{Y_1} - \sqrt{\lambda_1}, \ldots, \sqrt{Y_n} - \sqrt{\lambda_n}) \) is asymptotically multivariate normal with zero means and an correlation matrix equal to the identity matrix. Therefore, the one-factor results and the Tukey-Kramer method can be applied for the treatments-versus-control comparisons and all pairwise comparisons of the \( \sqrt{\lambda_i} \)’s respectively.
CHAPTER V

CONCLUSIONS

In this chapter, we shall conclude this thesis by drawing some conclusions and making some suggestions for future studies.

Conclusions on theories for one-factor structures
1. For the evaluation of one-sided coverage probabilities, the expressions in (2.0) and (2.6) are valid for all positive semi-definite correlation matrices with one-positive-factor structure or non-singular one-negative-factor structure.

2. For the singular one-negative-factor structures, the expressions in (2.0) and (2.6) hold for all positive semi-definite correlation matrices under the assumption that all the components of the factor pattern are of the same sign.

3. For the computation of two-sided coverage probabilities, the expressions in (2.0) and (2.6) work for all positive semi-definite correlation matrices with one-factor structure (positive or negative, and singular or non-singular).

4. For the purpose of constructing rectangular simultaneous confidence intervals, the critical value is obtained by solving an equation involving two-sided coverage probabilities. Therefore, the additional restriction on the singular one-negative-factor structure is not needed.

5. The collection of all correlation matrices with one-positive-factor structure is disjoint with that of all correlation matrices with one-negative-factor structure.
6. The maximal dimension of an arbitrary correlation matrix so that it can always be decomposed into a one-factor structure is three.

Conclusions on theories for two-factor structures

1. For the evaluation of one-sided coverage probabilities, the expressions in (3.0) and (3.3) are valid for all positive semi-definite correlation matrices with two-positive-factor structure.

2. For one-positive-one-negative-factor structures, the expressions in (3.0) and (3.3) hold for all positive semi-definite correlation matrices under the assumption that all the components of the factor pattern of the negative factor are of the same sign.

3. For the two-negative-factor structures, we need two additional assumptions concerning the factor patterns. First, if we ignore either one of the two factors, the determinant of the residual matrix is non-negative. Second, not both of the residual matrices have zero determinants. The first assumption restricts the values of the factor patterns so that the off-diagonal elements of the underlying correlation matrix, $\Sigma$, are not too large in absolute value to assure that $|\Sigma| \geq 0$. The second assumption is a technical assumption.

4. In the two-factor cases, unlike the one-factor cases, the additional assumptions cannot be ignored, when computing two-sided coverage probabilities.

Conclusions on computations

1. For one-positive-factor structures, the expression in (2.0) can be computed by a Gaussian quadrature, and the expression in (2.6) can be iteratively evaluated by a Gaussian quadrature or by two Gaussian quadratures.

2. For one-negative-factor structures, the expression in (2.0) and the inner integral in (2.6) can be suitably evaluated by an adaptive quadrature which is based on the
Trapezoidal rule or its variants, and the outer integral of (2.6) can be computed by a Gaussian quadrature. The adaptive quadrature routine by Stoer and Bulirsch (1967) is recommended.

3. For two-positive-factor structures, the expressions in (3.0) and (3.3) can be evaluated by iteratively applying Gaussian quadratures.

4. For one-positive-one-negative-factor and two-negative-factor structures, the inner double integrals can be suitably computed by iteratively applying adaptive quadratures, and the outer integral can be evaluated by a Gaussian quadrature.

5. As far as CPU time is concerned, the computations required for (3.0) and (3.3) are much slower than those required for (2.0) and (2.6) respectively, and the computations needed for negative-factor structures are much slower than those needed for positive-factor structures.

Open theoretical questions

1. It is not known whether or not the following sets are disjoint: the set of all correlation matrices with a two-positive-factor structure, the set of all correlation matrices with a one-positive-one-negative-factor structure and the set of all correlation matrices with a two-negative-factor structure.

2. The maximal dimension such that an arbitrary correlation matrix can always be decomposed into a two-factor structure is not known.

Future refinements on computations

1. To speed up the execution on machines with vector capacities for the computations in the negative-factor cases, we need to vectorize the product integrand, the most time-consuming part of the computation. In particular, we have to vectorize the codes for the complex-valued error function.
2. Another way to improve our current computation methods is to vectorize adaptive quadratures.

3. We might want to consider other totally different computational approaches such as a Quasi Monte Carlo method or utilization of the Fourier series expansion on the indicator function of the original range to be integrated (discussed by Russell, Farrier and Howell (1985)).

Future developments on applications

For the comparisons involving categorical data, the correlation matrix usually depends on the unknown parameters and, hence, it has to be estimated under certain appropriate models. However, if the assumed model fits the data set well, then the correlation matrix estimated under the fitted model is usually not too much different from the estimated correlation matrix based on the originally observed data. Therefore, the development of a unified comparison method is possible, if the correlation matrix is always estimated on the basis of the observed data.
LIST OF REFERENCES


[31] Goodwin, E. T. and Staton, J. (1948). Table of $\int_0^\infty \frac{e^{-u^2}}{u + x} du$. *Quarterly Journal of Mechanics and Applied Mathematics* 1, 319-326.


including a proof of the Tukey conjecture for unequal sample sizes.

Wiley, New York.

[38] Hsu, J. C. (1989). Simultaneous confidence intervals in the General Linear


Analysis.* Prentice-Hall, Enflwood Cliffs, New Jersey.

Marcel Dekker, New York.

[42] Lebedev, N. N. (1972). *Special Functions & Their Applications.*
Dover, New York.

*Technometrics* 17, 473-476.


Chapman and Hall, New York.


Unpublished manuscript.


[76] Uusipaikka, E. (1985). Exact simultaneous confidence intervals for multiple comparisons among three or four mean values.
*Journal of the American Statistical Association* 80, 196-201.