UNCONVENTIONAL APPROACH WITH THE LIKELIHOOD OF CORRELATION MATRICES

by

Myung Soon Song

B.S. Mathematics, Seoul National University, South Korea, 1996

M.A. Statistics, University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

the Department of Statistics in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF STATISTICS

This dissertation was presented

by

Myung Soon Song

It was defended on

May 19th 2011

and approved by

Professor Leon J. Gleser

Professor Satish Iyengar

Assistant Professor Yu Cheng

Professor Joel Greenhouse

Dissertation Advisors: Professor Leon J. Gleser,

UNCONVENTIONAL APPROACH WITH THE LIKELIHOOD OF CORRELATION MATRICES

Myung Soon Song, PhD

University of Pittsburgh, 2011

Numerical approximations are important research areas for dealing with complicated functional forms. Techniques for developing accurate and efficient calculation of combined likelihood functions in meta-analysis are studied. The first part of the thesis introduces a B-spline approximation for making a parsimonious model in the simplest case(2-dimensional case) of correlation structure. Inference about the correlation between vitamin C intake & vitamin C serum level is developed by using likelihood intervals and the MLE, along with comparison with conventional methods. The second part studies a multivariate numerical integration method for developing a better approximation of the likelihood for correlation matrices. Analyses for (1) intercorrelations among Math, Spatial and Verbal scores in an SAT exam and (2) intercorrelations among Cognitive Anxiety, Somatic Anxiety and Self Confidence from Competitive State Anxiety Inventory (CSAI-2) are explored. Algorithms to evaluate likelihood and to find the MLE is developed. Comparison with two conventional methods (joint asymptotic weighted average method & marginal asymptotic weighted average method) is shown.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0   INTRODUCTION

In many areas of statistics, reliably approximating the values of integrals is a problem of substantial concern. With development of computer and calculation methods, modern statisticians can study very complicated functional structures. They don't need to suffer from using difficult mathematics which give us usually elegant but not so practical forms due to computational problems. The problem of complicated functional form can be severe especially in the realm of multi-dimensional analysis. For example, how to combine information about population correlation matrices from many different independent studies is one of the hot questions in meta-analysis. Conventionally, asymptotic methods are used to tackle this problem. But these approaches are naïve and have some evident flaws. One of the most serious of these flaws is that we have to assume that the sample size from each study is sufficiently large to justify Central Limit approximations, an assumption which is violated in many situations.

In this paper, I develop approximation methods for calculating the combined likelihood of correlation matrices from different studies. Chapter 1 introduces a B-spline approximation method for developing a parsimonious model of the simplest 2-dimensional case of correlation structure, 2×2 correlation matrix. Inference about the correlation between vitamin C intake and vitamin C serum level is developed by using likelihood intervals and the MLE; The results are compared to conventional methods. Chapter 2 studies quadrature methods for developing a better approximation of the likelihood for $k \times k$ population correlation matrices. Analyses for (1) intercorrelations among Math, Spatial and Verbal scores in an SAT exam and (2) intercorrelations among Cognitive Anxiety, Somatic Anxiety and Self Confidence from Competitive State Anxiety Inventory (CSAI-2) are explored. Algorithms to evaluate likelihood, to find the MLE and to show likelihood regions are developed. Comparison

1

with two conventional methods (joint asymptotic weighted average method and marginal asymptotic weighted average method) are shown.

2.0   EXPLANATION OF SOME TOOLS USED

In this chapter, some tools used in the following chapters are explained briefly. If you know these already, you can skip this chapter.

2.1   B-SPLINE

In mathematics or statistics, a B-spline is a spline function having minimal support with respect to a given degree, smoothness, and domain partition.

The term B-spline was first introduced by Romanian mathematician Isaac Jacob Schoenberg and is short for basis spline. B-splines can be evaluated in a numerically stable way by the de Boor algorithm (de Boor [7]).

Hastie et al. [19] provides a quite rigorous definition of the B-spline. Let ξ_0 and ξ_{K+1} be two boundary knots, which typically define the domain over which a spline is evaluated. Now the augmented knot sequence τ satisfies:

1. $\xi_0 < \xi_1$ and $\xi_K < \xi_{K+1}$

2. $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_M \leq \xi_0$;

3. $\tau_{j+M} = \xi_j, \;\; j = 1, \cdots, K$;

4. $\xi_{K+1} \leq \tau_{K+M+1} \leq \tau_{K+M+2} \leq \cdots \leq \tau_{K+2M}$.

The actual values of these additional knots beyond the boundary are arbitrary, and it is customary to make them all the same and equal to ξ_0 and ξ_{K+1}, respectively.

Denote by $B_{i,m}(x)$ the ith B-splines basis function of order $m \leq M$ for the knot-sequence

τ. They are defined recursively in terms of divided differences as follows:

$$
B_{i,1} = \begin{cases} 1 & \text{if } \tau_i \leq x \leq \tau_{i+1} \\ 0 & \text{otherwise} \end{cases}
\tag{2.1}
$$

for $i = 1, \cdots, K + 2M - 1$.

$$
B_{i,m}(x) = \frac{x - \tau_i}{\tau_{i+m-1} - \tau_i} B_{i,m-1}(x) + \frac{\tau_{i+1} - x}{\tau_{i+m} - \tau_{i+1}} B_{i+1,m-1}(x)
\tag{2.2}
$$

for $i = 1, \cdots, K + 2M - m$.

Thus with $M = 4$, $B_{i,4}$, $i = 1, \cdots, K + 4$ are the $K + 4$ cubic B-spline basis functions for the knot sequence ξ. This recursion can be continued and will generate the B-spline basis for any order spline. In the statistical package R, cubic spline basis is used as default for B-spline approximation.

A fundamental theorem states that every spline function of a given degree, smoothness, and domain partition, can be represented as a linear combination of B-splines of that same degree and smoothness, and over that same partition.

Also, a B-spline curve of order K is in general C^{K-2} continuous (continuous up to $(K - 2)^{nd}$ derivative). For example, a cubic B-spline curve(a B-spline curve of order 4) is C^2 continuous. At a knot position the continuity is C^{K-M-1},where M is the multiplicity of that knot.

In this chapter the cubic B-spline is used for approximation. From now on, a B-spline refers to a cubic B-spline if there is no further explanation.

2.2 HIGHEST LIKELIHOOD REGIONS (HLR)

Many statistical methods involve summarizing a probability distribution by a region of the sample space covering a specified probability. But it is not always clear which region should be used. Suppose someone wishes to give a 90% prediction interval from a given distribution. Should he or she use the interval symmetric with respect to the mean or the median, the interval defined between the 5% and 95% quantiles, the interval of shortest length, or the interval that maximizes the probability of covering a given set? Hyndman [23] investigates

an approach to this question by suggesting highest density regions (HDR) from any given (possibly multivariate) density $f(x)$ which is bounded and continuous in x. He claims that the usual purpose in summarizing a probability distribution by a region of the sample space is to sketch a comparatively small set which contains most of the specified probability. The criteria he adapts are the following:

1. The region should occupy the smallest possible volume in the sample space.

2. Every point inside the region should have probability density at least as large as every point outside the region.

It follows immediately from the criteria and corresponding definition of HDR that the highest density region has the smallest possible volume in the sample space. Furthermore, the HDR has some virtues like:

1. The modes are contained in every HDR.

2. The HDR has a link to conventional wisdom. For example, in the case of a normal distribution an HDR coincides with the usual probability region symmetric with respect to the mean spanning the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles. This assertion is true for any symmetric unimodal distribution.

In this paper I will use the term Highest Likelihood Regions (HLR) rather than the Highest Density Regions (HDR). Both density functions and likelihood functions are non-negative on the corresponding domains. But there is a big difference between these two types of function. The area under the density curve defined on the domain should equal one, but the likelihood function has no such restriction. Of course the essential characteristics of HLR are inherited from those of HDR with no difficulty.

2.3   LIKELIHOOD STRIP

In many situations, it is needed to display a collection of likelihoods over a common interval. For example, in meta-analysis, comparing and possibly combining information about a parameter from a collection of studies can be a main interest. One way to do this is simply to

superimpose the likelihood curves. But such displays are very difficult to read if there are many curves. A second way is to display likelihood intervals (a point estimate with margin of error) for each likelihood function. Such intervals are sometimes misinterpreted by users as implying that all values within the interval are equally plausible. These approaches will be inadequate if the likelihood function is not approximately normal, as may occur with small sample sizes or nonlinear models.

The 'density strip' is a spectrum-like shaded monochrome strip whose darkness at a point is proportional to the probability density of the quantity at that point. Its use has been suggested by Jackson [24]. Density strips are useful for comparing distributions arising from parameter estimation, such as posterior distributions from Bayesian multiple regression or meta-analysis. The purpose of the density strip is to indicate the shape of a distribution, emphasizing the uncertainty surrounding the parameter estimate, rather than to allow the value of the density at every point to be accurately determined.

The concept of density strip will be adapted to likelihoods for the purpose of this paper, and will be referred to as 'likelihood strip' instead of 'density strip' throughout the paper.

3.0   CORRELATION FOR BIVARIATE CASE

3.1   INTRODUCTION

Numerical approximation methods for likelihood function of correlation coefficient are studied. The least squares approach is used to estimate a log-likelihood function by a function from a space of B-splines having desirable mathematical properties. The likelihood interval from the Highest Likelihood Regions (HLR) is used for further inference.

This approach can be easily extended to the realm of meta-analysis involving sample correlations from different studies by use of an approximated combined likelihood function. The sample correlations between vitamin C intake and serum level of vitamin C from many studies are used to illustrate application of this approach.

3.2   BACKGROUND

In many areas of research, it is useful to assess the relationship between continuous variables. Correlation coefficients have been used extensively as an index of the relationship between two normally distributed variables. Since the correlation coefficient is a scale-free measure of the relationship between variables, it is invariant under substitution of different but linearly equitable measures of the same construct. Therefore, the correlation coefficient is a natural candidate as an index of effect magnitude across studies in meta-analysis. (Hedges and Olkin [20])

How to combine the sample correlation coefficients from many independent studies having possibly different sample sizes has been an old question in meta-analysis.

Arguably, the most common process to deal with the issue above in meta-analysis is (P1) to calculate the sample correlations for individual studies, (P2) convert them to a common metric, and (P3) combine the results to obtain an average effect size. Why the process (P2) is needed instead of going straight from (P1) to (P3) may be worthy of explanation.

The sample correlation r was proposed by Pearson as an estimator of the population correlation ρ. The exact distribution of the sample correlation coefficient under the assumption of a bivariate normal distribution was first derived by Fisher [13], who obtained the distribution in a rather complicated form. Simpler forms of the distribution more suitable for computation are given by Hotelling [21]. These distributions will be shown later in this chapter. Due to the complexity of the exact distribution, the large sample distribution of a sample correlation r has been preferred. The asymptotic distribution of a sample correlation r is normal with mean ρ and variance $\frac{(1-\rho^2)^2}{n}$, where n is the sample size. Unfortunately, the variance of r in the approximation depends strongly on ρ, the unknown true value of the correlation. In order to stabilize the variation of r, Fisher [14] proposes the z-transformation

$$z = z(r) = \frac{1}{2}\log\frac{1+r}{1-r} = \tanh^{-1} r. \tag{3.1}$$

The corresponding transformation for ρ is

$$\zeta = \zeta(\rho) = \frac{1}{2}\log\frac{1+\rho}{1-\rho} = \tanh^{-1} \rho. \tag{3.2}$$

The z-transformation stabilizes the variance in the sense that z is approximately normally distributed with mean ζ and variance $\frac{1}{n}$ when n is large. A more accurate approximation to the distribution of z is obtained by setting the asymptotic variance equal to $\frac{1}{n-3}$ instead of $\frac{1}{n}$ for moderate values of n. Consequently $\sqrt{n-3}(z-\zeta)$ has, approximately the standard normal distribution:

$$\sqrt{n-3}(z-\zeta) \sim N(0,1) \tag{3.3}$$

For the conversion of metric (P2) described in the previous page, there are some alternative approaches(such as unbiased estimators, Kramer's t-transformation) beside Fisher's z-transformation. (Hedges and Olkin [20]) Among these, Fisher's z-transformation is the

most popular method, partly because of a simple distributional form. But Fisher's z-transformation produces an upward bias in the estimation of the correlation coefficients in the process (P3) described in the previous page. This upward bias is usually higher than the negligible downward bias produced by untransformed correlations.(Hunter [22]) Furthermore, Fisher's z-transformation is based on asymptotic theory, and may not work very well for small sample sizes.

Considering these problems, asking whether or not direct approaches from the process (P1) to the process (P3) are possible is still attractive. In the following sections, the use of log-likelihood functions and the B-spline approximations are suggested to answer this question.

3.3 APPROACH

3.3.1 Likelihood Function and Approximation

It is well known that likelihood functions play an important role in both the frequentist and Bayesian statistical paradigms. Many times uncertainty is taken into account by using the likelihood when studying a statistical problem. The concept of the likelihood is one of the best methods for unifying the demands of statistical modeling and inference. One of the very important advantages of likelihood functions is that they are most naturally represented, understood, and communicated graphically. In order to see what the data say, we look at graphs of likelihood functions.(Royall [35])

The distribution of the sample correlation r in a sample of size n from a bivariate normal distribution with correlation ρ was first obtained by Fisher[13] in the form:

$$\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\pi(n-3)!}\left|\frac{d^{n-2}}{dx^{n-2}}\left\{\frac{\cos^{-1}(-x)}{\sqrt{1-x^2}}\right\}\right|_{x=r\rho}. \tag{3.4}$$

Anderson [1] gives a different form of the density,

$$\frac{2^{n-3}(1-\rho^2)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-4)}}{\pi(n-3)!}\sum_{\alpha=0}^{\infty}\frac{(2\rho r)^\alpha}{\alpha!}\Gamma^2[\frac{1}{2}(n-1+\alpha)], \tag{3.5}$$

but the most commonly used form of the likelihood function of the population correlation coefficient based on a single sample correlation r from a sample size n is

$$L(\rho|r,n) = \frac{n-2}{\sqrt{2\pi}} \frac{\Gamma(n-1)}{\Gamma(n-\frac{1}{2})} (1-\rho^2)^{\frac{1}{2}(n-1)} (1-r^2)^{\frac{1}{2}(n-4)} (1-\rho r)^{-n+\frac{3}{2}} \qquad (3.6)$$

$$\times F\left(\frac{1}{2},\frac{1}{2};\ n-\frac{1}{2};\ \frac{1+\rho r}{2}\right),$$

where

$$F(a,b;\ c;\ x) = \sum_{j=0}^{\infty} \frac{\Gamma(a+j)}{\Gamma(a)} \frac{\Gamma(b+j)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+j)} \frac{x^j}{j!}, \qquad (3.7)$$

which stems from Hotelling [21] who made a comprehensive study of the distribution of r. The series in Eqn(3.6) converges faster than the ones in (3.4) or (3.5).

3.3.2 Numerical Evaluation Results

In this section, some numerical evaluation results are provided.

Fig 1 and Fig 2 show the graphs of the log-likelihood functions of the population correlaton (ρ) with various given sample correlations (r's) and a fixed sample size $n = 100$.

In Fig 1, the graphs of the log-likelihood of the population correlation ρ are displayed when the sample correlation r has one of the values $0.1, 0.2, \cdots, 0.9$. These functions are all uni-modal and the greater the r's, the more skewed the likelihood functions.

Fig 2 uses the same magnitudes of the sample correlations as in Fig 1 with opposite (negative) sign. The graphs in Fig 2 are mirror images of those in Fig 1.

Fig 3 displays the likelihood functions for ρ when the sample correlations r range from 0.1 to 0.9 in steps of 0.1, respectively with a sample size $n = 100$. The greater the magnitude of the sample correlation, the more the likelihood function concentrates around the peak or mode.

Fig 4, Fig 5 and Fig 6 display the log-likelihood functions of ρ and their B-spline approximations for fixed sample size $n = 100$ and various r (r equalling $-0.3, 0$ and 0.7) when one knot is used. In each figure, the solid line is for the log likelihood function and the dashed line is for the B-spline approximation.

10
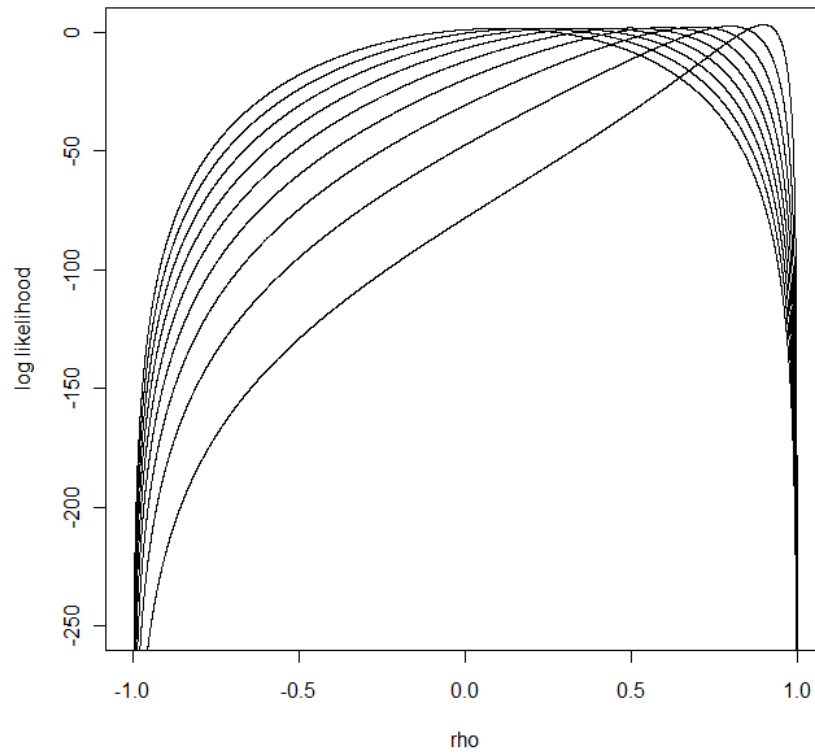
Figure 1: Log likelihood functions obtained from positive sample correlations (n=100)

11
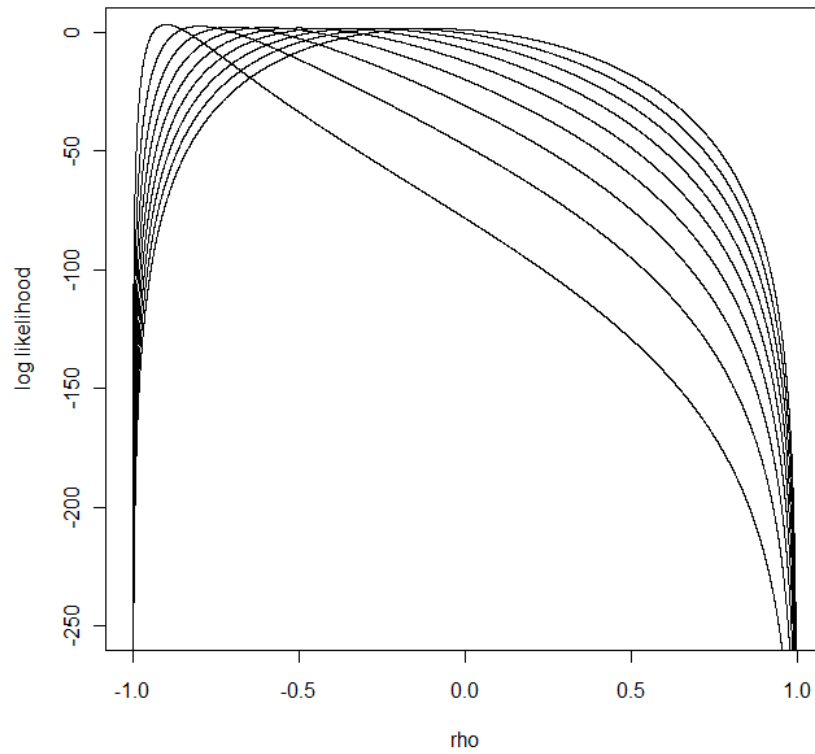
Figure 2: Log likelihood functions obtained from negative sample correlations (n=100)
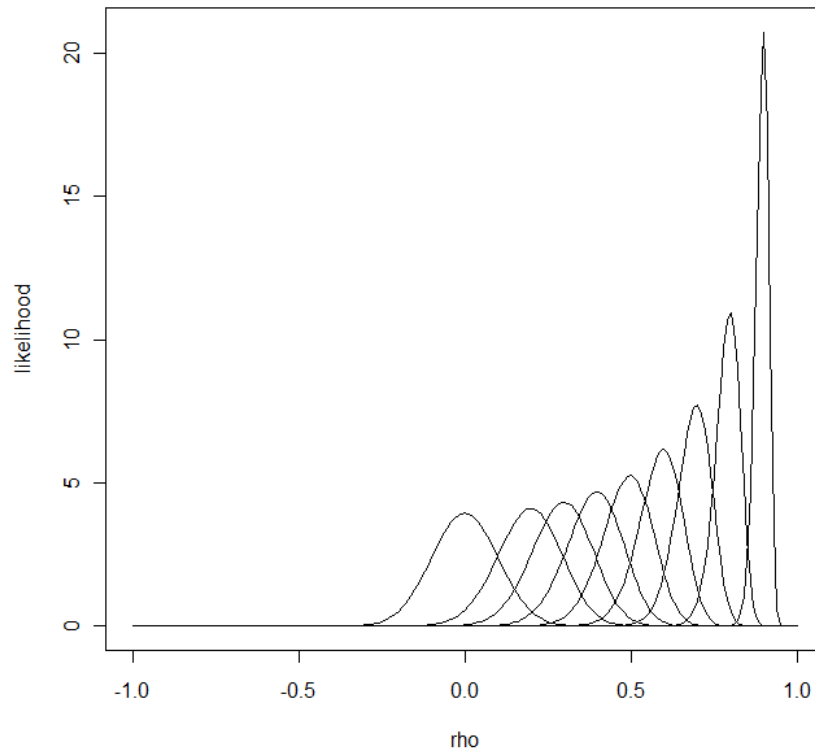
Figure 3: Likelihood functions obtained from postive sample correlations (n=100)
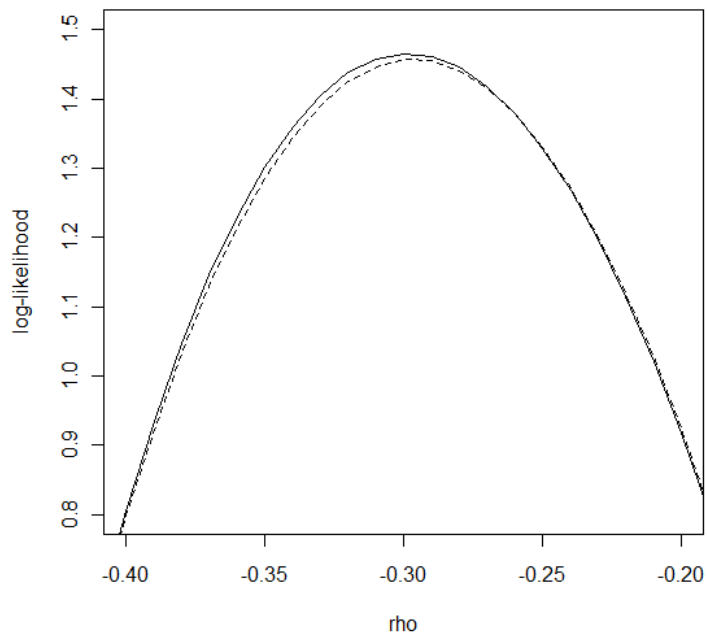
Figure 4: The log-likelihood(solid line) and its B-spline approximation(dashed line) when r=-0.3, n=100
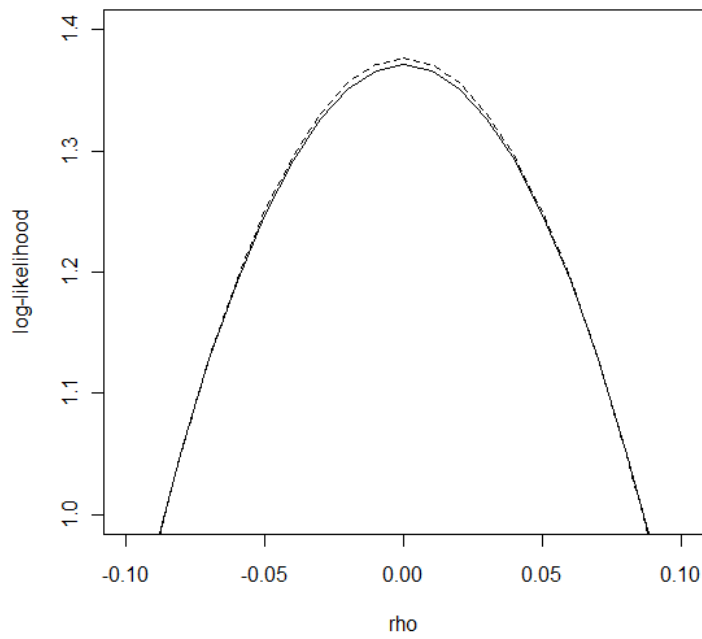
Figure 5: The log-likelihood(solid line) and its B-spline approximation(dashed line) when r=0, n=100
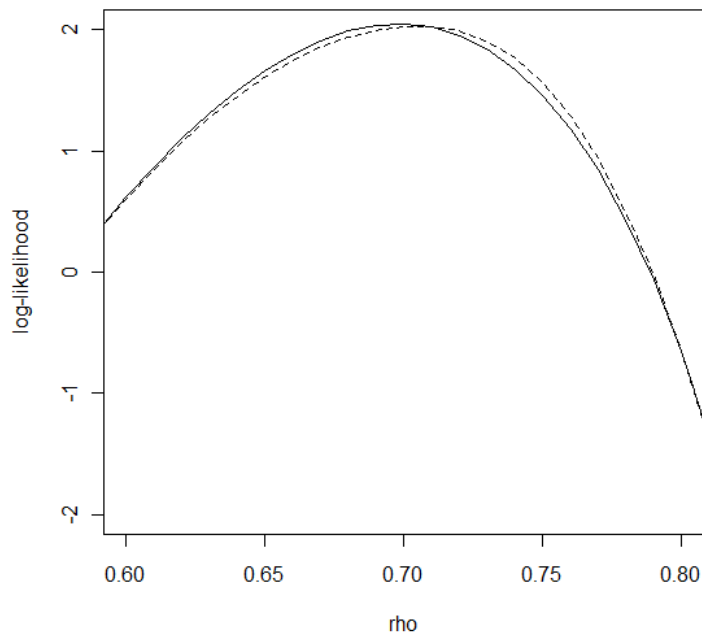
Figure 6: The log-likelihood(solid line) and its B-spline approximation(dashed line) when r=0.7, n=100

3.4   DATA - VITAMIN C

3.4.1   Notation and Convention

Let X and Y be random variables with bivariate normal distribution. Consider the situation in which each of k studies has examined correlations between X and Y. Let $r_{(i)}$ and $n_{(i)}$ be the sample correlation coefficient between X and Y and the sample size from the ith study for $i = 1, \cdots, k$, respectively.

3.4.2   Background

Recently, a great deal of attention has been given to fruit and vegetable consumption and their role in reducing rates of chronic diseases such as cancer, coronary heart disease (CHD), stroke, diabetes, and arthritis. It is suggested that the protective effect of fruits and vegetables is partly due to antioxidant nutrients like vitamin C and carotenoids which inhibit lipid per-oxidation and oxidative cell damage.(Steinmetz and Potter [37])

Vitamin C intake is measured by many different methods including Food Frequency Questionnaires (FFQ), Diet History Questionnaires (DHQ), 24 hr Dietary Recalls (DR) and Weight Records (WR). FFQ is one of the most commonly used tools in epidemiologic studies to assess long-term nutritional exposure. It is used to determine usual intakes of selected items from all major food groups.

Because fruits and also vegetables are the main source of dietary vitamin C, the serum level of vitamin C has been considered a good predictor of vitamin C intake from fruit and vegetable consumption. Therefore, significant changes of plasma vitamin C are expected by altering fruit and vegetable consumption.

Dehghan et al. [8] assess the association between vitamin C intake measured by dietary assessment methods and plasma level of vitamin C in epidemiological studies. The purposes of their study are: first, to investigate the strength of the correlation between plasma level of vitamin C as a biomarker and dietary vitamin C intake measured by methods commonly used for dietary assessment in epidemiological studies and, second, to explore whether the correlation between dietary vitamin C intake and plasma vitamin C varies between different

Study ID	Sample	$r_{(i)}$	$n_{(i)}$
1	Bingam et al.(1997)	0.35	127
2	EPIC group of Spain (1997)	0.61	40
3	Mckeowen et al. FFQ1(2001)	0.41	70
4	Mckeowen et al. FFQ2(2001)	0.39	70

Table 1: Four Studies used for Vitamin C example

Method for forming interval	Study 1	Study 2	Study 3	Study 4
Asymptotic	(0.20, 0.50)	(0.42, 0.80)	(0.22, 0.60)	(0.19, 0.59)
Exact Likelihood	(0.19, 0.49)	(0.37, 0.77)	(0.19, 0.58)	(0.17, 0.56)
B-spline Approx	(0.19, 0.49)	(0.37, 0.76)	(0.19, 0.58)	(0.18, 0.57)

Table 2: 95% C.I's and HLR likelihood intervals from Individual Studies for Comparison

dietary assessment methods.

In the following subsections, a baby version of a meta-analysis will be given based on the data from Dehghan et al. [8]. The analysis is done for just four studies of the correlation between dietary vitamin C measured by FFQ and plasma vitamin C for females.

In Table 2, 95% confidence intervals by using conventional asymptotic theory, 95% HLR likelihood intervals of the exact likelihood function and 95% HLR likelihood intervals obtained from the B-spline approximation to the log-likelihood function are compared, for each study. Table 2 shows differences between the classical asymptotic method and the two likelihood methods. Generally, the intervals from the asymptotic method tend to shift to the right compared to those of the likelihood methods for all studies. As we see, B-splines approximate the corresponding exact likelihood function extremely well.

3.4.3 Test of Homogeneity

If k studies investigate the same population correlation, combining the data from several studies to form a single estimated correlation is meaningful. Thus, we first need to determine whether or not the data obtained from several studies are consistent with the hypothesis of homogeneous correlations.

A formal hypothesis test can be made for the hypotheses:

$$
\begin{cases}
H_0: & \rho_1 = \rho_2 = \rho_3 = \rho_4 \\
H_a: & \text{Not all of } \rho_i\text{'s are the same}
\end{cases}
\tag{3.8}
$$

where ρ_i is a population correlation from the ith study for $i = 1, \cdots, 4$. This test uses the statistic

$$
Q = -2 \log \left[\frac{\sup_{H_0} \mathbf{L}(\boldsymbol{\rho}|\mathbf{r})}{\sup_{H_a} \mathbf{L}(\boldsymbol{\rho}|\mathbf{r})} \right]
\tag{3.9}
$$

where $\mathbf{L}(\boldsymbol{\rho}|\mathbf{r}) = \prod_{i=1}^{4} L(\rho_{(i)}|r_{(i)}) = \exp\left[\sum_{i=1}^{4} \log L(\rho_{(i)}|r_{(i)}) \right]$ is the combined likelihood from the four study.

Q has approximately a χ^2 distribution with $3(= 4 - 1)$ degrees of freedom if H_0 is true. A test of H_0 at the 100α % level of significance is given by rejecting H_0 if Q is greater than the $100(1 - \alpha)$ percentile of the χ^2 distribution with 3 degrees of freedom.

By using the statistical software R, we obtain $Q = 3.39$ which is less than 7.81, the 95% critical value of the χ^2 distribution with 3 degrees of freedom, so the hypothesis of homogeneous correlations is accepted at the 5% level of significance.

3.4.4 Comparison of Estimators and Visualization

When a series of k independent studies share a common population correlation coefficient ρ, it is natural to estimate ρ by pooling estimates from each of the studies. If the sample sizes of the studies differ, then the estimates from the larger studies will be more precise than the estimates from the smaller studies. In this case it is reasonable to give more weight to the

	Asymptotic	Exact Likelihood & B-spline Approx
Estimators	$\widetilde{r} = 0.43$	$\widehat{r} = 0.41$
95% C.I. or	C.I.	HLR Likelihood Interval
HLR Likelihood Interval	(0.34, 0.52)	(0.31, 0.50)

Table 3: Results from the Combined Study

more precise estimates when pooling. This leads to weighted estimators of the form:

$$r_w = w_1 r_{(1)} + \cdots + w_k r_{(k)} \qquad (3.10)$$

where $w_1, \cdots, w_k$ are nonnegative weights that sum to 1.(Hedges and Olkin [20])

The weights that minimize the variance of r_w give weight inversely proportional to the variance in each study. This is intuitively clear in that a smaller variance, i.e., more precision, should lead to larger weight. Consequently, if this weight is used, the asymptotic weighted average of correlation has the form:

$$\widetilde{r} = \frac{\sum_{m=1}^{k} \dfrac{n_{(m)}}{\left(1 - r_{(m)}^2\right)^2} r_{(m)}}{\sum_{m=1}^{k} \dfrac{n_{(m)}}{\left(1 - r_{(m)}^2\right)^2}} \qquad (3.11)$$

Table 3 summarizes the main results obtained from the statistical software R from this way of combining the four studies. It shows the MLE $\widehat{r}$ from the exact combined likelihood or its B-spline approximation and an asymptotic pooled estimate $\widetilde{r}$. Note that $\widehat{r}$ is less than $\widetilde{r}$ by 0.02. Also, in Table 3 appear the 95% asymptotic C.I. for ρ and the 95% HLR likelihood interval for ρ.

Fig 7 shows the combined likelihood(solid line) and its B-spline approximation(dashed line) from the four studies. The B-spline curve approximates the likelihood extremely well.

Fig 8 illustrates the likelihood strip(See Section 2.3) of the combined likelihood function (or its B-spline approximation) from the four studies.
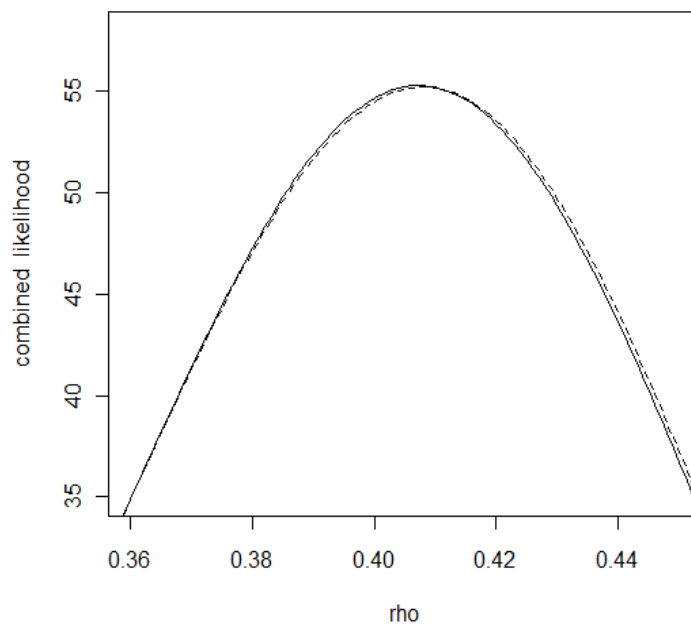
Figure 7: The Combined Likelihood(solid line) & its B-spline approximation(dashed line) from 5 Studies
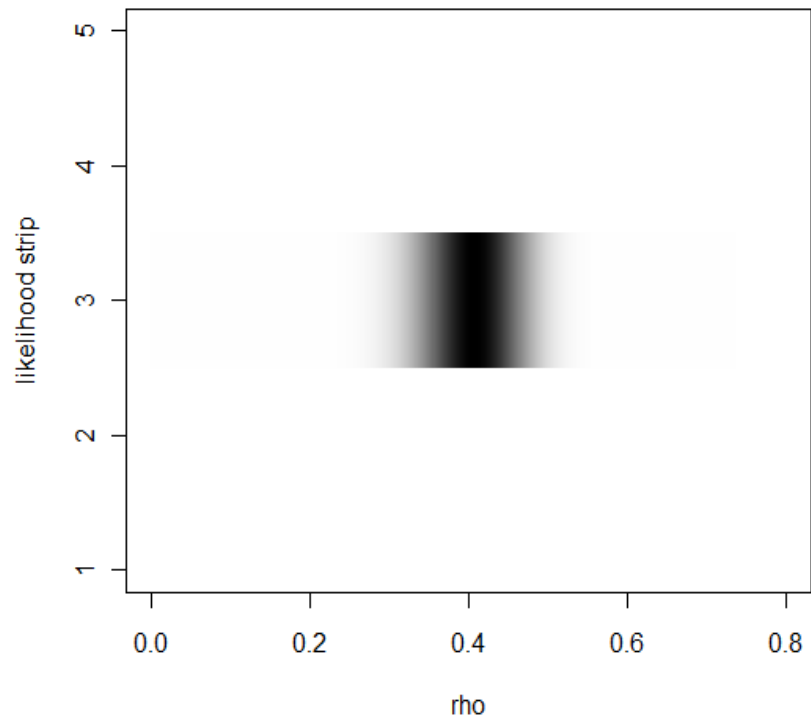
Figure 8: The Likelihood Strip of Combined Likelihood (and B-spline approximation) from 4 Studies

3.5 SUMMARY AND DISCUSSION

As we see in the previous sections, a B-spline generally gives a close approximation to the log-likelihood function. In particular, when the MLE of the combined likelihood (or log-likelihood) is not far from the origin, say between -0.8 and 0.8, a B-spline fits the exact likelihood function nearly perfectly.(See Fig.7) From Table 3, we can also observe that B-splines detect the right location of the MLE and also the likelihood interval from the exact combined likelihood, which are very important for inference.

Furthermore, just 2 or 3 inner knots, which correspond to degrees of freedom 5 or 6 when we use a cubic B-spline in the computer program R, are used for approximation in the previous sections. Consequently we can obtain a parsimonious model from the exact likelihood function.

But there is a limitation when applying a B-spline approximation. When the MLE of ρ is close to 1 in magnitude, the B-spline approximation fits less well, which is a general flaw of spline approximation when we deal with values close to the boundaries of domain of the approximated function. Fortunately, the vitamin C data do not suffer from this boundary issue, so the results are reliable.

4.0 CORRELATIONS FOR MULTIVARIATE CASE

4.1 INTRODUCTION

All the contents in the previous chapter were devoted to the bivariate case - the single correlation coefficient ρ. In this chapter, numerical methods are developed for approximating the combined likelihoods of correlation matrices estimated in different studies. As a good starting point for this more general multivariate case, it is quite natural to deal with the trivariate case.

There are many conventional ways to combine information from different studies. Using weighted averages are probably the most common choices. However, if the sample sizes are not sufficiently large, these methods are not so reliable.

We can start by deriving a needed form of the density function and from this obtain the likelihood function. The derivation of the density function is added in the Appendix. The distribution of the sample correlation matrix R in a sample of size N from a trivariate normal distribution with correlation matrix P has the form:

$$f(R|n, P) = \frac{|R|^{\frac{1}{2}(n-4)}|P|^n}{\pi^{\frac{3}{2}}} \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} \sum_{\gamma=0}^{\infty} \frac{(\rho_{12} - \rho_{13}\rho_{23})^{\alpha} (\rho_{13} - \rho_{12}\rho_{23})^{\beta} (\rho_{23} - \rho_{12}\rho_{13})^{\gamma}}{(1 - \rho_{12}^2)^{\frac{1}{2}(n+\beta+\gamma)} (1 - \rho_{13}^2)^{\frac{1}{2}(n+\alpha+\gamma)} (1 - \rho_{23}^2)^{\frac{1}{2}(n+\alpha+\beta)}}$$

$$\times \frac{2^{\alpha+\beta+\gamma} r_{12}^{\alpha} r_{13}^{\beta} r_{23}^{\gamma}}{\alpha!\beta!\gamma!} \times \frac{\Gamma[\frac{1}{2}(n + \alpha + \beta)]\Gamma[\frac{1}{2}(n + \alpha + \gamma)]\Gamma[\frac{1}{2}(n + \beta + \gamma)]}{\Gamma[\frac{1}{2}n]\Gamma[\frac{1}{2}(n - 1)]\Gamma[\frac{1}{2}(n - 2)]} \tag{4.1}$$

where $n = N - 1$, $R = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix}$, $P = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$, and
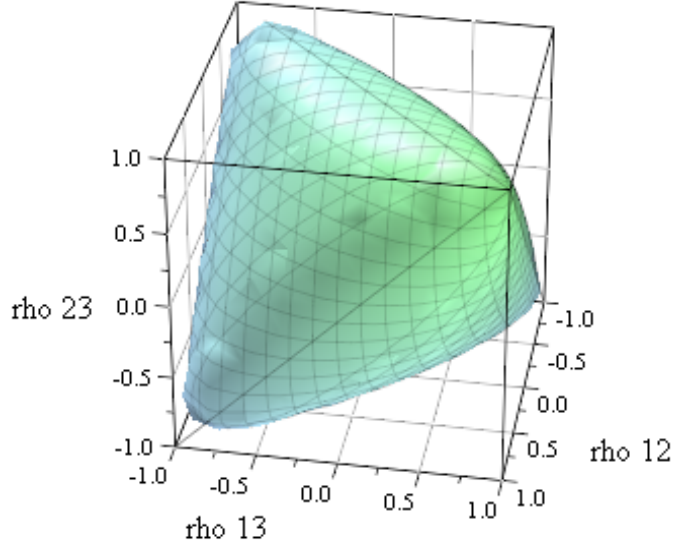
Figure 9: The Domain of the Likelihood of P in the Three Dimensional Case

$$|R| = 1 - (r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23}), \quad |P| = 1 - (\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23}).$$

The likelihood of the correlation matrix P when in a sample of size N we observe the sample correlation matrix R is proportional to (4.1). But Eqn(4.1) is not defined on the complete entire cubic domain $(-1, 1) \times (-1, 1) \times (-1, 1)$ because P is supposed to be positive definite. P is positive definite if and only if

$$|\rho_{ij}| < 1, \ 1 \le i < \ j \le 3 \text{ and } \rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23} < 1. \tag{4.2}$$

Fig 9 and Fig 10 illustrate the domain (4.2) of the likelihood function (4.1) in the trivariate case.

It turns out that representing the likelihood in terms of infinite series has serious drawbacks, both because of the mathematical complexity of the representation and also due to increasing difficulty in calculation of the likelihood as the dimensionality of the parameter space increases. In general, $\frac{p(p-1)}{2}$ combined infinite series are required for a $p \times p$ correlation
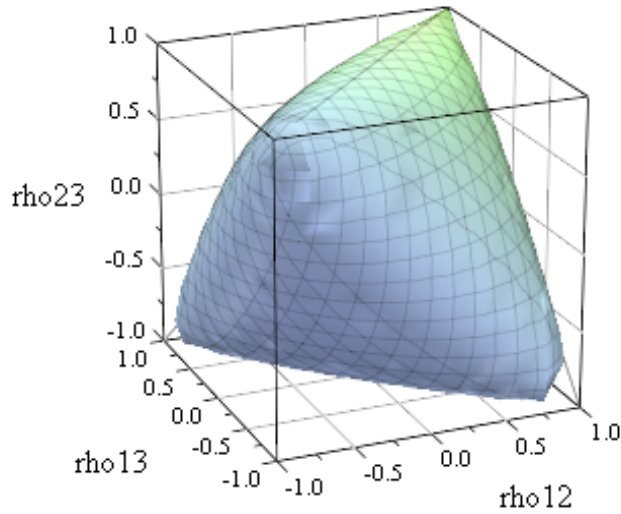
Figure 10: The Different Perspective of the Domain of the Likelihood of P in the Three Dimensional Case

matrix. Even in relatively low dimensions, computation can be difficult. For example, if you have 5 variables you need to use $\frac{5(5-1)}{2} = 10$ combined infinite series for the likelihood. Consequently, it was chosen to follow Fisher's [16] approach and represents the likelihood function as a multiple integral to which numerical integration can be applied in this paper. It is proved that just $p - 1$ multiple integrals need to be computed for the general p variable case. For example, $5 - 1 = 4$ multiple integrals are needed for the 5 dimensional case. When Fisher [16] suggested a multiple integral representation of the likelihood for the p-dimensional case, calculation was not easy, but we can now calculate a revised multiple integral form of the likelihood easily with advanced computers.

As an application of inference, intercorrelation among math, spatial and verbal scores in an SAT examination is explored and comparisons among various approaches are given. Also, a similar process will be conducted for intercorrelations among cognitive anxiety, somatic anxiety, and self-confidence from the area of sports psychology.

4.2 APPROACH - LIKELIHOOD AS A FORM OF MULTIPLE INTEGRAL

As mentioned in the introduction of this chapter, the likelihood function of the correlation matrix P obtained from a sample correlation matrix R from a sample of size N has a form of $\frac{p(p-1)}{2}$ combined infinite series. This form turns out to be computationally unpractical due to heavy burden of calculations - curse of dimensionality. But the computational problem can be solved by using a different form of the likelihood function based on a multiple integral. It is shown in Appendix that the likelihood function has a form of $p-1$ multiple integrals in p variable case:

$$L(P) = f(R|n, P) = C \int_0^1 \cdots \int_0^1 \frac{\left(\prod_{i=1}^p v_i\right)^{\frac{n}{2}-1} \cdot \prod_{k=2}^{p-1} b_k^{k-1}}{\left(\sum_{i=1}^p (P^{-1})_{ii} v_i + 2 \sum_{1 \le i < i \le p} (P^{-1})_{ij} R_{ij} \sqrt{v_i v_j}\right)^{\frac{np}{2}}} db_{p-1} \cdots db_1$$

(4.3)

where

$$C = \frac{|R|^{\frac{n-p-1}{2}} \Gamma(\frac{np}{2})}{\pi^{\frac{p(p-1)}{4}} |P|^{\frac{n}{2}} \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]} \quad ,$$

(4.4)

$$v_i = (1 - b_{i-1}) b_i \cdots b_{p-1} \ , \ \text{ for } i = 1, \cdots, p-1,$$

$$v_p = 1 - b_{p-1}, \ n = N - 1,$$

and where P_{ij}, R_{ij} and $(P^{-1})_{ii}$ are the elements from the ith row and j^{th} column in P, R and P^{-1}(the inverse matrix of P), respectively.

Fisher[16] uses a representation of the likelihood in terms of p multiple integrals in the p variable case. Its integration domain is $(0, \infty)$ for each coordinate. But this paper suggests a representation of the likelihood with $p-1$ multiple integrals in the p variable case with integration domain $(0, 1)$ for each coordinate. This representation has obvious calculational advantage over Fisher's representation.

In section 4.3, numerical integration methods are introduced to calculate the likelihood (4.3).

4.3 APPROACH - NUMERICAL INTEGRATION

Numerical integration is the study of how the numerical value of an integral can be obtained by using approximate computational methods. It is sometimes called quadrature. Basically, all numerical integration methods are based on adding up the value of the integrand at a sequence of points in the range of integration.

We can formulate the situation in which we have main interest:

$$I(f) = \int_{S^m} \cdots \int f(x_1, \cdots, x_m) \ dx_1 \cdots dx_m \approx \sum_{i=1}^{M} W_i \ f(y_{i,1}, \cdots, y_{i,m}) \qquad (4.5)$$

where R^m is a m-dimensional Euclidean space, S^m is a specified region in R^m, $f \colon R^m \to R$ is a common function. A vector $(y_{i,1}, \cdots, y_{i,m})$ is called a *point* of the formula. The scalar W_i is called the ith *coefficient* of the formula. We say that formula (4.5) has *degree r*(or *degree of exactness r*) if it is exact for all polynomials in $x_1, \cdots, x_m$ of degree r and there is at least one polynomial of degree $r + 1$ for which it is not exact.(Evans [10])

At this point, we need to mention a very important aspect of numerical integration - the assessment of error in an approximation. The only absolutely certain method to assess is to compare the approximation with the correct answer, which is not always possible in practice. With iterative methods, we have the natural method of examining the approximations at successive stages and stopping the iteration when the changes become small for a number of iterations. Perhaps the best way to be confident that we have accurately approximated a particular integral is to use very different methods and see if the results agree.(Evans and Swartz [11]) In the following sections, three main methods of numerical integration will be used: Gauss-Legendre quadrature, adaptive integration and the Monte Carlo method.

Among the many quadrature rules, one of the most commonly used rules is the Gauss-Legendre rule(Press et al. [33]). For the multi-dimensional case (4.5) here becomes:

$$I(f) = \int_0^1 \cdots \int_0^1 f(x_1, \cdots, x_m) \ dx_1 \cdots dx_m \approx \sum_{i_1=1}^{M} \cdots \sum_{i_m=1}^{M} W_{i_1} \cdots W_{i_m} f(y_{i,1}, \cdots, y_{i,m})$$
$$(4.6)$$

Even if an optimal Gauss-Legendre quadrature extension giving a degree of exactness of $3M + 1$ can be found(Kuonen [26]), you still do not know the accuracy in terms of correct decimal places. To get a prescribed accuracy you need adaptive integration, which keeps subdividing the domain of integration S^m until a specified error has been achieved.

Adaptive algorithms developed by Genz and Malik [18] operate by repeated subdivision of hyper-rectangular regions into smaller hyper-rectangles. In each subregion, the integral is estimated using a rule of degree seven, and an error estimate is obtained by comparison with a rule of degree five which uses a subset of the same points. These subdivisions are designed to dynamically concentrate the computational work into the subregions where the integrand is most irregular, and thus adapt to the behavior of the integrand. But one of the disadvantages of adaptive algorithms is their slow speed. This disadvantage can be considerably overcome by using Monte Carlo methods.

Monte Carlo (MC) methods can be loosely described as statistical simulation methods. We refer to Robert and Casella [34] and Tanner [39] for a comprehensive introduction. The classical MC method for approximating a multiple integral such as given in the left-hand side of (4.5), denoted by $I(f)$, is as follows. We choose M sets of points $\{y_{i,1}, \cdots, y_{i,m}\}$, $i = 1, \cdots, M$ at random, uniformly distributed in S^m. The integral is then estimated using $W_i = V/M$ in the right-hand side of (4.5),

$$I(f) \approx \widehat{I}(f) = \frac{V}{M} \sum_{i=1}^{M} f(y_{i,1}, \cdots, y_{i,m}) \tag{4.7}$$

where $V = I(1)$ is the m-dimensional volume of S^m. Thus we see that the basic MC method iteratively approximates a definite integral by uniformly sampling from the domain of integration,and averaging the function values at the samples. The integrand is treated as a random variable, and the sampling scheme yields an estimate of the mean of the random variable. Since $\widehat{I}(f)$ in (4.7) estimates $I(f)$, the absolute error in this mean can be evaluated by considering the corresponding standard error of the mean,

$$\epsilon = \left| I(f) - \widehat{I}(f) \right| \approx \frac{\sigma}{M^{1/2}} \tag{4.8}$$

where $\sigma^2 = V \cdot I(f^2) - I^2(f)$. If $\{y_{i,1}, \cdots, y_{i,m}\}$, $i = 1, \cdots, M$ are regarded as independent random variables then $\widehat{I}(f)$ is a random variable with mean $I(f)$ and variance σ^2/M, which can also be estimated from the random sample through

$$\frac{V}{M^2} \sum_{i=1}^{M} \left\{ f(y_{i,1}, \cdots, y_{i,m}) - \widehat{I}(f) \right\}^2. \tag{4.9}$$

The absolute error (4.8) has an average magnitude of $O(M^{-1/2})$.(Kuonen [26])

In section 4.4, it is sketched how to calculate maximum likelihood estimates(MLEs) and the corresponding likelihood regions by using the multiple integral representation (4.3) of the likelihood and the numerical integration methods introduced in this section.

4.4   HOW TO GET THE INFERENCE

Fig 11 illustrates how information from sample correlation matrices from k studies can be combined using calculation of the likelihood function at selected points in the parameter space to get inferential results. In this figure, R_i and N_i stand for the correlation matrix and the sample size from study i for $i = 1, \cdots, k$, respectively. Also, LR denotes a HLR likelihood region. This flow chart is helpful for readers to understand and use the likelihood function and its numerical integration when they try to apply the approach of this paper to their data.

Now, we want to explain briefly how to get the MLE step by step.

1) Fix an initial subdomain of the parameter space for calculating the likelihood:

The likelihood function (4.3) can be calculated for any values in the parameter space (4.2) by the numerical integration methods described in section 4.3. But many values of likelihood in the domain (4.2) are needed to find the MLE numerically. We do not need to use the whole domain (4.2). The sample correlations will suggest what subset is relevant for an initial stage.

2) Choose a collection of coarse grid points in the subdomain chosen in step 1).

3) Calculate the values of the combined likelihood (the multiplication of the likelihoods of the k studies) at the grid points from step 2).
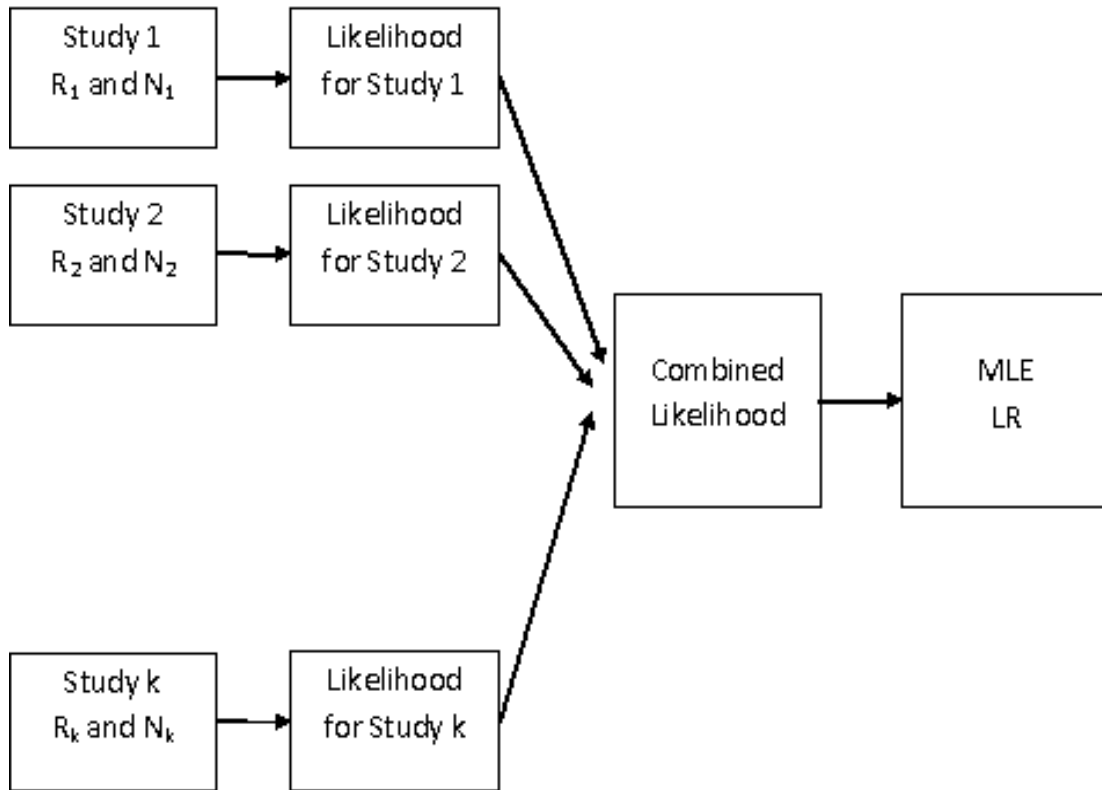
Figure 11: Flow Chart of Combining Correlation Matrices Process

4) Find the point (an initial candidate for the MLE) which achieves the maximum value of the calculated combined likelihoods from step 3).

5) Choose a smaller subdomain around the point from step 4) and a more refined set of grid points than in step 2).

6) Calculate the values of the combined likelihood at the grid points from step 5).

7) Find the point which attains the maximum value of the combined likelihood computed in step 6)

8) Repeat steps 5) - 7) until convergence occurs with desired accuracy.

The three numerical integration methods in section 4.3 can be used separately in the computations to check whether each calculation method yields the same result for the MLE.

Also, we can explain briefly how to visualize the $100(1-\alpha)\%$ HLR likelihood region for elements of the population correlation matrix in the following steps:

a) Fix a subdomain of the parameter space for calculating the likelihood:

The sample correlations given will suggest what subset is relevant. Outside of this subdomain, the likelihood should be practically zero so that a negligible contribution is made to the total volume under the surface defined by the likelihood.

b) Assign equidistant grid points in the subdomain chosen in step a):

The finer the grid points, the higher resolution the plot.

c) Calculate the values of the combined likelihood at the grid points from step b) and sort these points in descending order.

d) Sum these combined likelihood values starting with the largest and stop summing up if the ratio between the cumulative sum from the top to the total sum of all the combined likelihood values from step c) becomes $1-\alpha$.

e) Determine the grid points for which the combined likelihood contribute to the numerator of the ratio in step d).

f) Visualize the $100(1-\alpha)\%$ likelihood region of the population correlation matrix using the set of grid points from step e).

4.5   NOTATION AND CONVENTION

Let $X_1, \cdots, X_p$ be random variables with a multivariate normal distribution. Consider the situation in which each of k studies has independently observed correlations among these p variables. Each study provides $\frac{p(p-1)}{2}$ nonredundant correlations. Let $r_{(i)st}$ be the sample correlations between X_s and X_t in the ith study, and let $\mathbf{r}_{(i)} = \left(r_{(i)12}, \; r_{(i)13}, \cdots, \; r_{(i)(1-p)p} \right)'$ be the vector of the $\frac{p(p-1)}{2}$ nonredundant sample correlations from study i, $i = 1, \cdots, k$.

4.6   DATA - SAT

In this section, data from Friedman [17] are used. The data are from four studies of inter-relations among math, spatial, and verbal ability. Becker [3] uses the same data and some results in this section are compared with her results.

The correlation matrices among three variables in the ith study for $i = 1, 2, 3,$ and 4 are represented as:

$$
\begin{array}{c}
 \\
M \\
S \\
V
\end{array}
\begin{array}{ccc}
M & S & V \\
\end{array}
\begin{bmatrix}
1 & r_{(i)12} & r_{(i)13} \\
r_{(i)12} & 1 & r_{(i)23} \\
r_{(i)13} & r_{(i)23} & 1
\end{bmatrix}
$$

where M: Math, S: Spatial, and V: Verbal.

Also, these correlation matrices can be written as a vector $\mathbf{r}_{(i)}$. The relationships represented are

$$
\mathbf{r}_{(i)} = \begin{bmatrix} r_{(i)12} \\ r_{(i)13} \\ r_{(i)23} \end{bmatrix} \quad \begin{array}{l} \text{Math-Spatial} \\ \text{Math-Verbal} \\ \text{Spatial-Verbal} \end{array} \tag{4.10}
$$

The basic information about the four samples is given in Table 4.

The $\mathbf{r}_{(i)}$ vectors for the four studies are

$$
\mathbf{r}_{(1)} = \begin{bmatrix} 0.46 \\ 0.31 \\ 0.19 \end{bmatrix}, \; \mathbf{r}_{(2)} = \begin{bmatrix} 0.46 \\ 0.55 \\ 0.32 \end{bmatrix}, \; \mathbf{r}_{(3)} = \begin{bmatrix} 0.40 \\ 0.40 \\ 0.18 \end{bmatrix}, \; \text{and } \mathbf{r}_{(4)} = \begin{bmatrix} 0.27 \\ 0.57 \\ 0.22 \end{bmatrix}. \tag{4.11}
$$

Sample ID	Sample	Sample Size
1	Berry (1957)	$n_1 = 103$
2	Rosenberg (1981)	$n_2 = 69$
3	Weiner 1 (1984)	$n_3 = 69$
4	Weiner 2 (1984)	$n_4 = 70$

Table 4: Four Studies used for SAT example

4.6.1 Test of Homogeneous Correlation Matrices

Before combining the data from different independent studies to estimate a correlation matrix, it is important to check if the studies actually have the same population correlation matrix. It is useful to conduct a hypothesis test to decide whether the sample correlation matrices from the studies are consistent with a common population correlation matrix.

This test can be formalized as testing the hypotheses:

$$
\begin{cases}
H_0 : & P_1 = \cdots = P_k \\
H_a : & \text{At least two } P_i\text{'s are different}
\end{cases}
\tag{4.12}
$$

where P_i is the $p \times p$ population correlation matrix from the ith study, $i = 1, \cdots, k$. This test uses the statistic

$$
Q = -2 \log \left[\frac{\sup_{H_0} L(\mathbf{P}|\mathbf{R})}{\sup_{H_0 \cup H_a} L(\mathbf{P}|\mathbf{R})} \right]
\tag{4.13}
$$

where $L(\mathbf{P}|\mathbf{R}) = \prod_{i=1}^{k} L(P_i|R_i)$, $\mathbf{P} = \{P_1, \cdots, P_k\}$, $\mathbf{R} = \{R_1, \cdots, R_k\}$ and R_i is the $p \times p$ sample correlation matrix from the ith study, $i = 1, \cdots, k$.

Because Q has approximately a χ^2 distribution with $\frac{1}{2}p(p-1)(k-1)$ degrees of freedom if H_0 is true, a test of H_0 at the 100α % level of significance is given by rejecting H_0 if Q is greater than the $100(1-\alpha)$ percentile of the χ^2 distribution with $\frac{1}{2}p(p-1)(k-1)$ degrees of freedom.

In the SAT example, we obtain $Q = 9.28$ which is less than 9.41, the 40% critical value of the χ^2 distribution with $\frac{1}{2}3(3-1)(4-1) = 9$ degrees of freedom, so the hypothesis of

34

homogeneity of correlation matrices cannot be rejected even at the 40% level of significance.

Becker [3] tests the hypothesis (4.12) and has a similar result with $Q = 9.34$, but she uses a different statistic for Q (a quadratic form in the sample correlations). Usually, a likelihood ratio test such as (4.13) does not use an actual likelihood because the likelihood cannot be evaluated in many situations, but the actual likelihood is used for the homogeneity test of correlations in this paper. When the sample sizes in the studies are large enough to permit Central Limit approximations, the test used by Becker and the likelihood ratio test (4.13) yield similar results, as is the case here, but this is not guaranteed to be the case when the studies have small sample sizes.

4.6.2 Comparison of Estimators

Assuming that all of the four studies share a common population correlation matrix, i.e. $P_1 = \cdots = P_4$, pooling estimates from the studies to estimate the common correlation matrix is quite natural as a next step in the meta-analysis. Becker [3] suggests two conventional methods to get a pooled estimate of a correlation matrix. One is to separately calculate a simple weighted average of corresponding sample correlations across studies for each population correlation whereas the other is a generalized least squares approach. In this paper these estimators will be referred to as the Marginal Asymptotic Weighted Average(MAWA) and the Joint Asymptotic Weighted Average(JAWA), respectively. In this section, the MLE is compared with these conventional pooled estimates.

If we have k studies in which each study uses the same p variables, the MAWA has the form:

$$\widetilde{\boldsymbol{\rho}}_M = \left(\widetilde{\rho}_{M12}, \ \widetilde{\rho}_{M13}, \cdots, \ \widetilde{\rho}_{M1p}, \ \widetilde{\rho}_{M23}, \cdots, \ \widetilde{\rho}_{M(1-p)p} \right)' \tag{4.14}$$

where

$$\widetilde{\rho}_{Mij} = \frac{\displaystyle\sum_{m=1}^{k} \frac{n_m}{\left(1 - r_{(m)ij}^2\right)^2} r_{(m)ij}}{\displaystyle\sum_{m=1}^{k} \frac{n_m}{\left(1 - r_{(m)ij}^2\right)^2}}, \qquad i < j \tag{4.15}$$

35

and the JAWA has the form:

$$\widetilde{\boldsymbol{\rho}}_J = \left(\sum_{m=1}^{k} \widehat{\Gamma}_m^{-1}\right)^{-1} \cdot \sum_{m=1}^{k} \widehat{\Gamma}_m^{-1} r_{(m)} = \left(\widetilde{\rho}_{J12}\,,\ \widetilde{\rho}_{J13}\,,\cdots,\ \widetilde{\rho}_{J1p},\ \widetilde{\rho}_{J23},\cdots,\ \widetilde{\rho}_{J(1-p)p}\right)' \quad (4.16)$$

where

$$\widehat{\Gamma}_m = \begin{pmatrix} \widehat{Var}(r_{(m)12}) & \cdots & \widehat{Cov}\left(r_{(m)12}, r_{(m)(p-1)p}\right) \\ \cdots & \cdots & \cdots \\ \widehat{Cov}\left(r_{(m)12}, r_{(m)(p-1)p}\right) & \cdots & \widehat{Var}(r_{(m)(p-1)p}) \end{pmatrix} : \quad (4.17)$$

and where

$$\begin{cases} \widehat{Var}(r_{(m)ij}) = \dfrac{\left(1 - r_{(m)ij}^2\right)^2}{n_m} \\[4ex] \widehat{Cov}\left(r_{(m)ij}, r_{(m)ik}\right) \\ = \dfrac{1}{n_m}\left\{\tfrac{1}{2}r_{(m)ij}r_{(m)ik}\left(r_{(m)ij}^2 + r_{(m)ik}^2 + r_{(m)jk}^2 - 1\right) + r_{(m)jk}\left(1 - r_{(m)ij}^2 - r_{(m)ik}^2\right)\right\} \end{cases}$$

$$\text{for } i < j, \qquad i < k, \qquad j < k,\ \text{and } m = 1, 2, 3, \text{and } 4,$$

(4.18)

and $\widehat{\Gamma}_m$ is the associated asymptotic covariance matrix for each $\mathbf{r}_{(m)}$.(Olkin and Siotani [32])

Similarly, the MLE matrix can be written as a vector $\widehat{\boldsymbol{\rho}}$, where

$$\widehat{\boldsymbol{\rho}} = \left(\widehat{\rho}_{12}\,,\ \widehat{\rho}_{13}\,,\cdots,\ \widehat{\rho}_{1p}\,,\ \widehat{\rho}_{23}\,,\cdots,\ \widehat{\rho}_{(1-p)p}\right)'. \quad (4.19)$$

If you plug $\mathbf{r}_{(i)}$'s (4.11) into (4.17), you have

$$\widehat{\Gamma}_1 = \begin{bmatrix} 0.0060 & 0.0008 & 0.0020 \\ 0.0008 & 0.0079 & 0.0037 \\ 0.0020 & 0.0037 & 0.0090 \end{bmatrix}, \widehat{\Gamma}_2 = \begin{bmatrix} 0.0090 & 0.0016 & 0.0051 \\ 0.0016 & 0.0071 & 0.0035 \\ 0.0051 & 0.0035 & 0.0117 \end{bmatrix}$$

(4.20)

$$\widehat{\Gamma}_3 = \begin{bmatrix} 0.0102 & 0.0010 & 0.0043 \\ 0.0010 & 0.0102 & 0.0043 \\ 0.0043 & 0.0043 & 0.0136 \end{bmatrix}, \quad \widehat{\Gamma}_4 = \begin{bmatrix} 0.0122 & 0.0013 & 0.0070 \\ 0.0013 & 0.0065 & 0.0019 \\ 0.0070 & 0.0019 & 0.0129 \end{bmatrix}$$

	MAWA	JAWA	MLE

$$\widetilde{\boldsymbol{\rho}}_M = \begin{pmatrix} 0.413 \ (0.025) \\ 0.469 \ (0.056) \\ 0.227 \ (0.018) \end{pmatrix} \qquad \widetilde{\boldsymbol{\rho}}_J = \begin{pmatrix} 0.412 \ (0.023) \\ 0.472 \ (0.063) \\ 0.253 \ (0.134) \end{pmatrix} \qquad \widehat{\boldsymbol{\rho}} = \begin{pmatrix} 0.403 \\ 0.444 \\ 0.223 \end{pmatrix}$$

Table 5: MAWA, JAWA and MLE from the SAT data

which can be used for computing standard errors of, and getting asymptotic confidence intervals based upon, the MAWA and the JAWA.

To estimate a specific component ρ_{ij} of the correlation matrix P, the MAWA uses only corresponding component(marginal information) from each study whereas the JAWA uses all of the components(joint information). If all the covariances are zero in (4.17), you can easily check that the JAWA of (4.16) is exactly the same as the MAWA of (4.14).

Table 5 presents the MAWA, the JAWA, and the MLE obtained from the data (4.11). The MAWA and the JAWA in Table 5 are the same as those in Becker's paper[3]. For the MAWA and the JAWA, the number in parentheses represents the corresponding relative difference from the MLE. For example, $0.025 \left(= \left| \frac{0.413 - 0.403}{0.403} \right| \right)$ is the relative difference of $\widetilde{\rho}_{M12}$ from the MLE $\widehat{\rho}_{12}$. Most of the components of the MAWA and the JAWA are fairly close to the corresponding components of the MLE and their relative differences are quite small, but $\widetilde{\rho}_{J23}$ shows considerable relative difference from the MLE $\widehat{\rho}_{23}$.

Table 6 shows comparisons among the MAWA, the JAWA and the MLE obtained from the sample correlations (4.11) when we vary the sample sizes from the four studies. The first column 'Sample Size' shows various sample sizes. For example, $(10, 10, 10, 30)$ means that $n_1 = 10$, $n_2 = 10$, $n_3 = 10$, $n_4 = 30$. When sample sizes are as small as $(7, 5, 7, 5)$, the MAWA and the JAWA are not reliable because large sample theory approximations are known to be inaccurate and we can easily detect considerable difference between the conventional estimators and the MLE. When one study has a much bigger sample size than the other studies like $(10, 10, 10, 30)$,the MLE tends to give more weight to that study. In the third row, each component of $(10, 7, 7, 7)$ is one-tenth of its counter part of $(103, 69, 69, 70)$ which presents the real sample sizes used in the SAT data. The MAWA and the JAWA with the

sample sizes $(10, 7, 7, 7)$ are practically same as those with the sample sizes $(103, 69, 69, 70)$. (See Table 6) On the contrary, the MLE with sample sizes $(10, 7, 7, 7)$ is quite different from the MLE with the sample sizes $(103, 69, 69, 70)$.

4.6.3 HLR Likelihood Regions

In this subsection the HLR likelihood regions (The term Likelihood Regions will be used hereafter) obtained from the likelihood function are illustrated. Even if expression of a likelihood region with a closed mathematical form is impossible due to the complexity of the likelihood function, we can still visualize a likelihood region.

Fig 12 illustrates the 95% likelihood region of the combined likelihood function based on the sample correlation matrices in (4.11). Fig 13 and Fig 14 show different perspectives for the same likelihood region. All of these figures show that the likelihood region of the correlation matrix P has an egg-shaped(approximately elliptical) shape.

Also, it is quite useful to see projections from the original likelihood region for assessing pairs of population correlation coefficients. Fig 15 shows the three-dimensional 95% likelihood region from (4.11), its projections into the ρ_{12} & ρ_{13} plane, the ρ_{12} & ρ_{23} plane and the ρ_{13} & ρ_{23} plane, respectively, in clockwise manner from the upper-left corner. Similar but different graphics can be seen in Fig 16, in which the projections of 30%, 70% and 90% confidence regions for the elements of $P = [\rho_{ij}]$ are given.

Fig 17 demonstrates the likelihood regions with different credibilities - 99%, 70%, 50% and 30%, respectively

4.6.4 Visualization of Likelihoods

In the previous subsection, we put main emphasis on joint likelihood but in this subsection we deal with likelihoods of ρ_{12} and ρ_{13} when ρ_{23}'s are fixed.

In Fig 18, Fig 19 and Fig 20, we can see level plot, contour plot and likelihood plot of ρ_{12}, ρ_{13} when $\rho_{23} = 0.12$, $\rho_{23} = 0.22$ and $\rho_{23} = 0.32$, respectively. Those ρ_{23}'s are chosen to detect the behavior of the likelihood with respect to ρ_{12} and ρ_{13} when ρ_{23} changes around its MLE 0.22. (Table 5) In Fig 19, the global maximum of the likelihood is achieved when

38

Sample Size	MAWA	JAWA	MLE
$(7, 5, 7, 5)$	$\begin{pmatrix} 0.411 \\ 0.466 \\ 0.224 \end{pmatrix}$	$\begin{pmatrix} 0.410 \\ 0.469 \\ 0.248 \end{pmatrix}$	$\begin{pmatrix} 0.373 \\ 0.405 \\ 0.200 \end{pmatrix}$
$(10, 10, 10, 30)$	$\begin{pmatrix} 0.368 \\ 0.517 \\ 0.227 \end{pmatrix}$	$\begin{pmatrix} 0.368 \\ 0.519 \\ 0.255 \end{pmatrix}$	$\begin{pmatrix} 0.341 \\ 0.491 \\ 0.219 \end{pmatrix}$
$(10, 7, 7, 7)$	$\begin{pmatrix} 0.413 \\ 0.471 \\ 0.227 \end{pmatrix}$	$\begin{pmatrix} 0.411 \\ 0.473 \\ 0.253 \end{pmatrix}$	$\begin{pmatrix} 0.382 \\ 0.419 \\ 0.210 \end{pmatrix}$
$(400, 400, 400, 400)$	$\begin{pmatrix} 0.407 \\ 0.483 \\ 0.231 \end{pmatrix}$	$\begin{pmatrix} 0.406 \\ 0.485 \\ 0.255 \end{pmatrix}$	$\begin{pmatrix} 0.398 \\ 0.462 \\ 0.229 \end{pmatrix}$

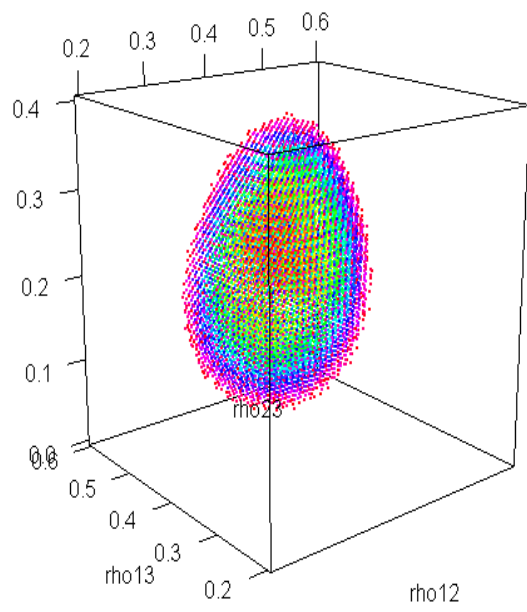Table 6: MAWA, JAWA and MLE with various sample sizes

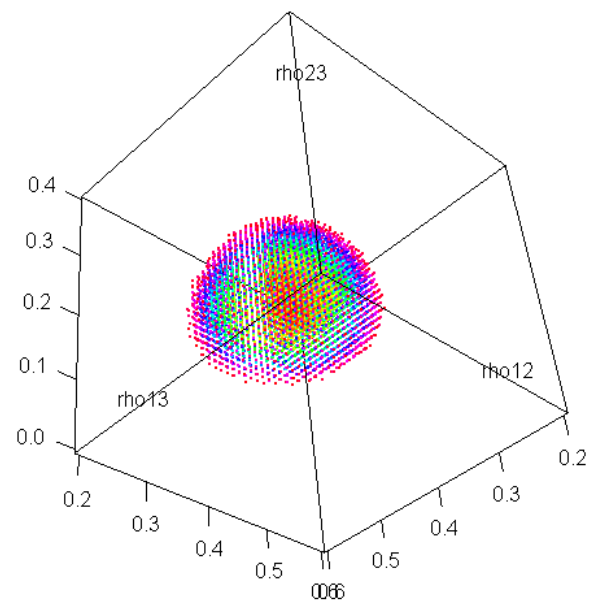Figure 12: 95% Likelihood Region of Three Correlations

40

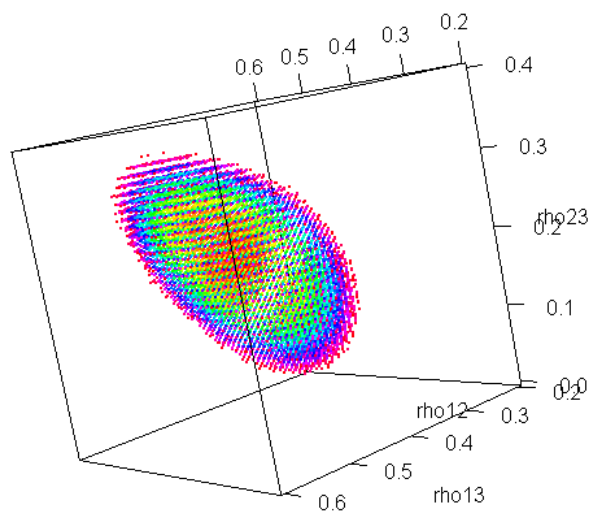Figure 13: 95% Likelihood Region of Three Correlations (Different Perspective I)

Figure 14: 95% Likelihood Region of Three Correlations (Different Perspective II)
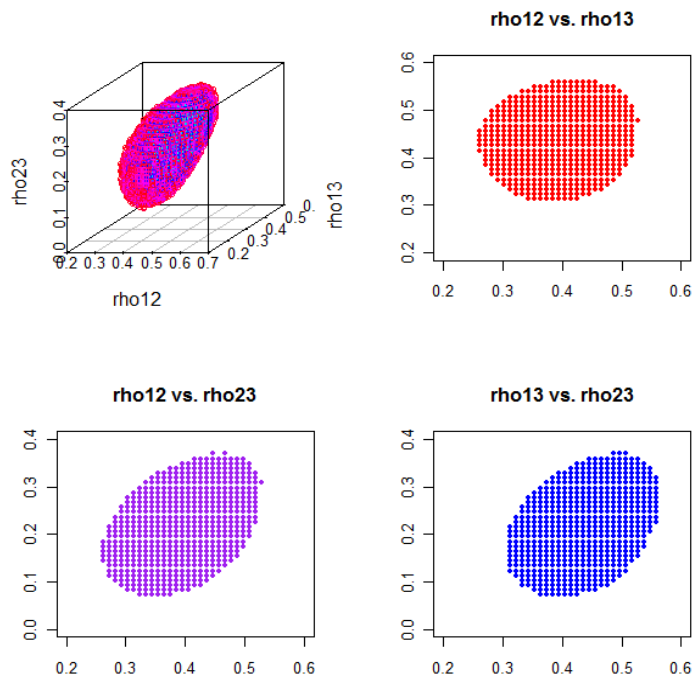
Figure 15: 95% Likelihood Region and its Two-Dimenional Projections
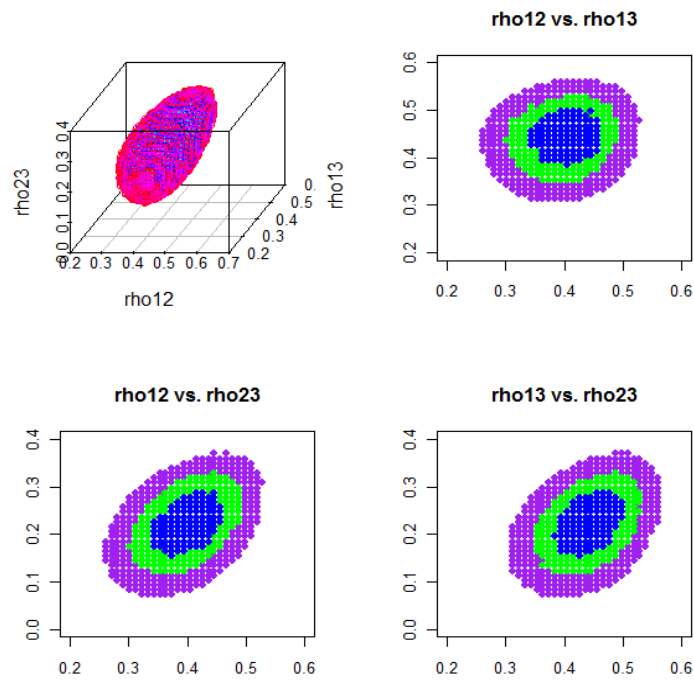
Figure 16: 95% Likelihood Region and Two-Dimensional Projections of 30%, 70% and 95% Likelihood Regions
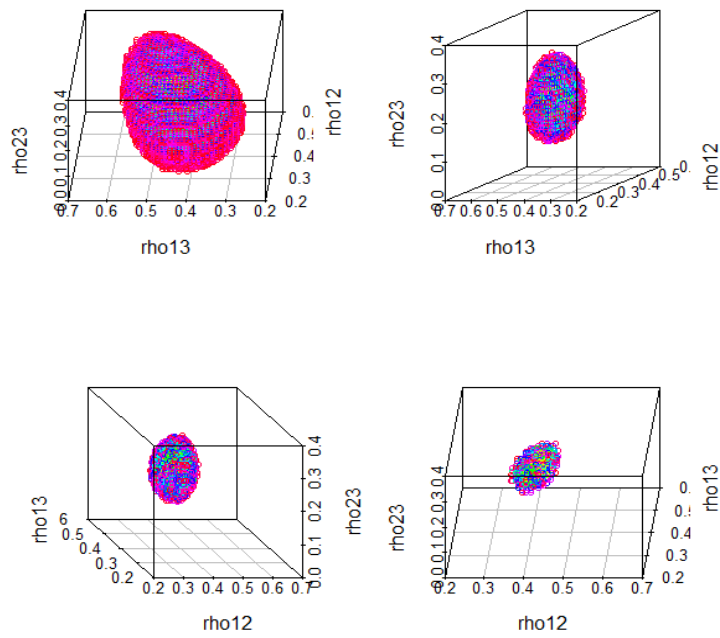
Figure 17: 99%, 70%, 50% and 30% Likelihood Regions

we take $\rho_{12} = 0.40$, $\rho_{13} = 0.44$ with the fixed value of $\rho_{23} = 0.22$, but the local maximums are achieved when we take $\rho_{12} = 0.36$, $\rho_{13} = 0.42$ with the fixed value of $\rho_{23} = 0.12$ (See Fig 18) and when we take $\rho_{12} = 0.44$, $\rho_{13} = 0.48$ with the fixed value of $\rho_{23} = 0.32$.(See Fig 20) The global maximum from Fig 19 is $374,239$, which is much greater than the local maximum $65,766$ from Fig 18 and the local maximum $70,518$ from Fig 20.

Fig 21 summarizes the results from Fig 18, Fig 19 and Fig 20.

4.7   DATA - ANXIETY

4.7.1   Background

An inherent aspect of competitive athletics is the need for athletes to meet the demands of competition and to perform well under pressure. Depending on how the athlete perceives the demands of competition, he or she may interpret pressure situations in a variety of ways. For example, they may be perceived as a natural part of athletic competition, or they may invoke heightened levels of stress. When in stressful and anxiety-provoking circumstances, some athletes have been observed to experience deficits in performance, even to the point of "choking." Thus, the relationship between anxiety and athletic performance has received considerable attention from researchers in the field of sport psychology.(Craft et al.[6])

The multidimensional approach to the study of sports anxiety (Martens et al.[29]) considers subcomponents of anxiety, specifically cognitive anxiety, somatic anxiety, and self-confidence. Cognitive anxiety is the mental component of anxiety and is caused by negative expectations about success or by negative self-evaluation. Somatic anxiety refers to the physiological and affective elements of the anxiety experience that develop directly from autonomic arousal. Self-confidence is the athlete's global perceptions of confidence.

In order to assess the multidimensional aspects of anxiety, Martens et al.[29] develop the Competitive State Anxiety Inventory-2 (CSAI-2). This 27-item measure has three subscales: cognitive anxiety, somatic anxiety, and self-confidence. Athletes are asked to indicate "how you feel right now" for each item on a 4-point Likert scale ranging from "not at all" to
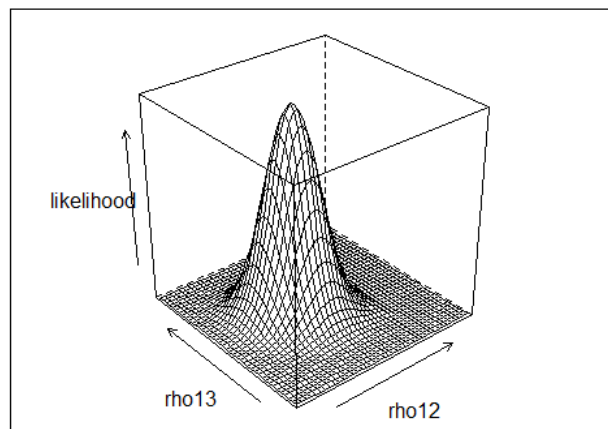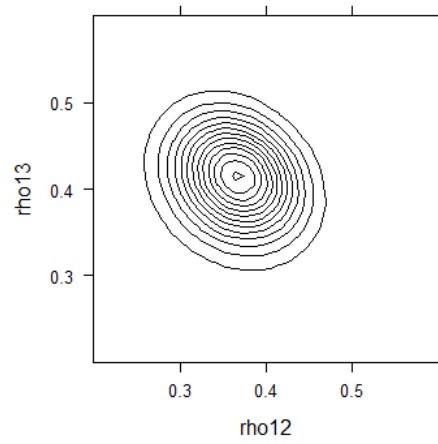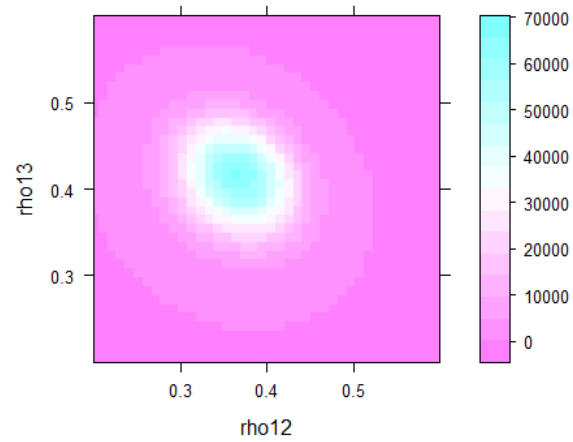
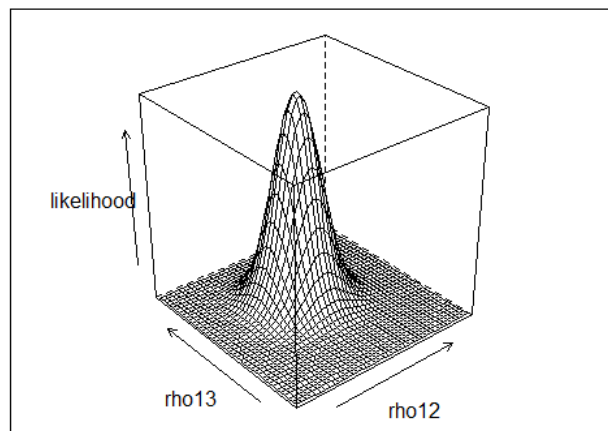Figure 18: Level plot, contour plot and conditional likelihood I
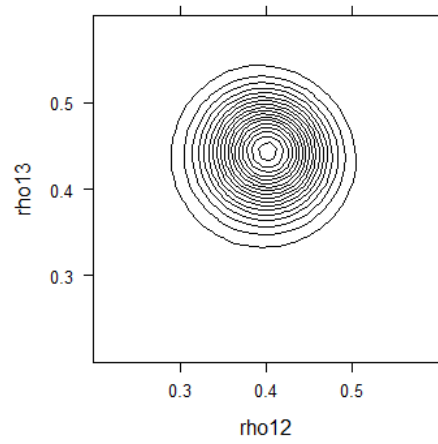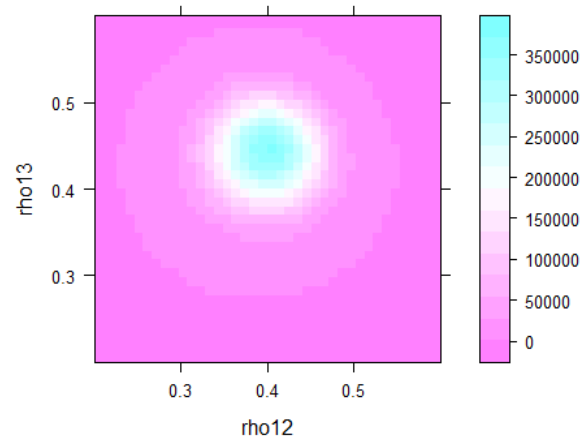
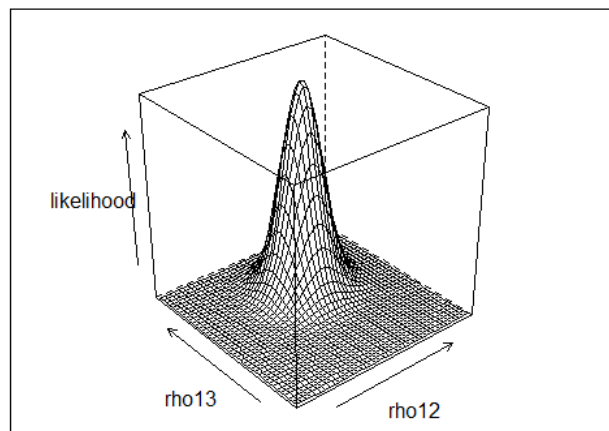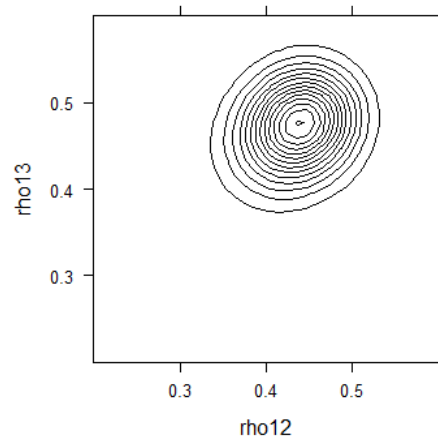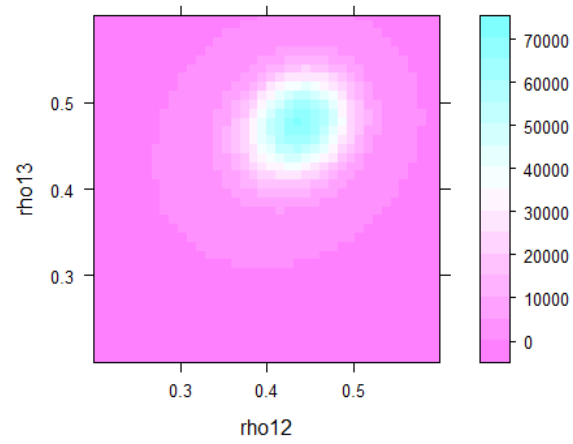Figure 19: Level plot, contour plot and conditional likelihood II
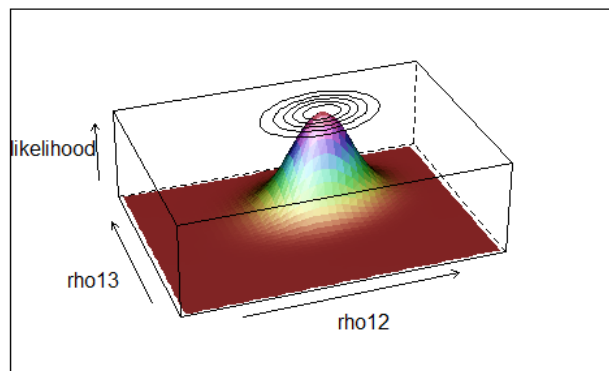
Figure 20: Level plot, contour plot and conditional likelihood III
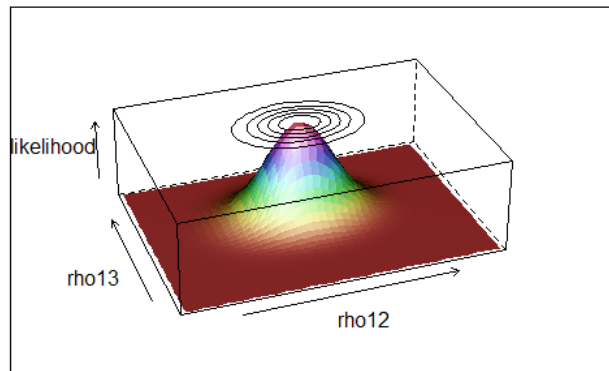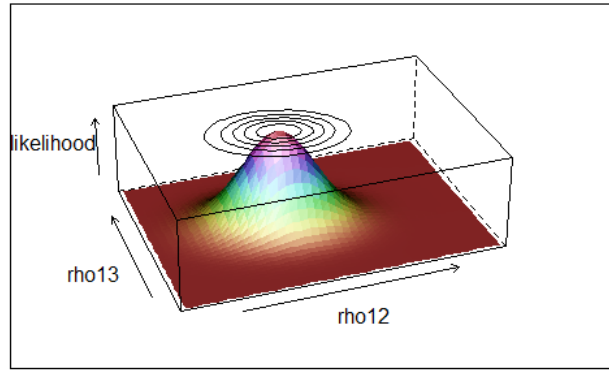
Figure 21: Three-dimensional view of Correlations and their corresponding conditional like-lihoods and contours

"very much so." Examples of the cognitive anxiety items include "I am concerned about this competition." and "I am concerned about choking under pressure." These items differ from the somatic anxiety statements such as "I feel nervous." or "I feel tense in my stomach." The self-confidence subscale includes items such as "I feel at ease." and "I'm confident I can meet the challenge." Each of the three subscales has 9 items, which are summed to get a score representing the level of intensity the athlete is feeling for each component of anxiety, and for self-confidence about performing.(Craft et al.[6])

In this section, I used the data from three studies which are Caruso et al.[4], Edwards and Hardy [9] and Maynard et al.[30]. These studies use the CSAI-2 to calculate correlations among cognitive anxiety, somatic anxiety, and self-confidence.

When I checked Craft et al.[6], there are 29 studies which can be used for meta-analysis, but I decided to use just three studies partly because there is a limitation for access to the specific data and partly because my goal is not to conduct rigorous meta-analysis but to show that my likelihood approach is applicable to meta-analytic application. To be notationally consistent, I use the same notations in this section as were used for the SAT data in section 4.6.

The correlations matrices among the three variables in the ith study for $i = 1, 2,$and 3 are represented as:

$$
\begin{array}{cc}
 & \begin{array}{ccc} \text{CA} & \text{SA} & \text{SC} \end{array} \\
\begin{array}{c} \text{CA} \\ \text{SA} \\ \text{SC} \end{array} &
\begin{bmatrix}
1 & r_{(i)12} & r_{(i)13} \\
r_{(i)12} & 1 & r_{(i)23} \\
r_{(i)13} & r_{(i)23} & 1
\end{bmatrix}
\end{array}
$$

where CA: Cognitive Anxiety, SA: Somatic Anxiety, and SC: Self-Confidence

Also, these correlation matrices can be written as a vector $\mathbf{r}_{(i)}$, the relationships represented are

$$
\mathbf{r}_{(i)} = \begin{bmatrix} r_{(i)12} \\ r_{(i)13} \\ r_{(i)23} \end{bmatrix} \quad \begin{array}{l} \text{CA-SA} \\ \text{CA-SC} \\ \text{SA-SC} \end{array} \tag{4.21}
$$

Basic information about three studies are given in Table 7.

51

Study ID	Study	Size
1	Caruso et al. (1990)	$n_1 = 24$
2	Edwards and Hardy(1996)	$n_2 = 45$
3	Maynard et al. (1995)	$n_3 = 24$

Table 7: Three Studies used for the Anxiety example

The $\mathbf{r}_{(i)}$ vectors for the three studies are:

$$\mathbf{r}_{(1)} = \begin{bmatrix} 0.42 \\ -0.42 \\ -0.48 \end{bmatrix}, \ \mathbf{r}_{(2)} = \begin{bmatrix} 0.47 \\ -0.37 \\ -0.50 \end{bmatrix}, \ \mathbf{r}_{(3)} = \begin{bmatrix} 0.67 \\ -0.36 \\ -0.72 \end{bmatrix}. \tag{4.22}$$

4.7.2 Test of Homogeneous Correlation Matrices

With the same purpose as the SAT example, we conduct the test of homogeneous correlation matrices.

In the Anxiety example, we have $Q = 6.38$ which is less than 7.23, the 30% critical value of the χ^2 distribution with $\frac{1}{2}3(3-1)(3-1) = 6$ degrees of freedom, so we conclude that the hypothesis of homogeneity of correlation matrices cannot be rejected even at the 30% level of significance.

4.7.3 Comparison of Estimators

Because all of the three studies appear to share a common population correlation matrix, i.e. $P_1 = P_2 = P_3$, it is reasonable to pool estimates from the studies to estimate the common correlation matrix.

Table 8 shows the MAWA, the JAWA, and the MLE from (4.22). For the MAWA and the JAWA estimators, each value inside parentheses represents the corresponding relative difference from the MLE. The components of the MAWA are not so different from their counterparts in the MLE and their relative differences are relatively small, but the JAWA is

52

	MAWA		JAWA		MLE

$$\widetilde{\boldsymbol{\rho}}_M = \begin{pmatrix} 0.544 \ (0.035) \\ -0.381 \ (0.007) \\ -0.598 \ (0.046) \end{pmatrix} \quad \widetilde{\boldsymbol{\rho}}_J = \begin{pmatrix} 0.605 \ (0.111) \\ -0.378 \ (0.004) \\ -0.653 \ (0.115) \end{pmatrix} \quad \widehat{\boldsymbol{\rho}} = \begin{pmatrix} 0.511 \\ -0.374 \\ -0.555 \end{pmatrix}$$

Table 8: MAWA, JAWA and MLE from the Anxiety data

quite different from the MLE. ($\mathbf{r}_{J12} = 0.605$ and $\mathbf{r}_{J23} = -0.653$ have the relative differences of 0.111 and 0.115, respectively)

4.7.4 HLR Likelihood Regions

Fig 22 shows the 95% likelihood region for the common correlation matrix P determined from the combined likelihood function obtained from the data in (4.22). Fig 23 and Fig 24 illustrate different perspectives for the same likelihood region. The likelihood region for the correlation matrix P looks like a skinny egg.

Fig 25 shows the three-dimensional 95% likelihood region and its projections into the ρ_{12} & ρ_{13} plane, the ρ_{12} & ρ_{23} plane and the ρ_{13} & ρ_{23} plane, proceeding clockwise from the upper-left corner. Fig 26 illustrates the two-dimensional projections of 30%, 70% and 95% likelihood regions for the element of P.

Lastly, Fig 27 demonstrates the likelihood regions with different credibilities with 99%, 70%, 50% and 30% clockwise from the upper-left corner, respectively.

4.7.5 Visualization of Likelihoods

We can see level plot, contour plot and likelihood plot of ρ_{12} and ρ_{13} when $\rho_{23} = -0.755$, $\rho_{13} = -0.555$ and $\rho_{23} = -0.355$, respectively, from Fig 28, Fig 29 and Fig 30, . Those ρ_{23}'s are picked because the MLE of r_{23}'s is -0.555 from Table 8 and we want to check the behavior of the "slices" of likelihood with respect to ρ_{12} and ρ_{13} when ρ_{23} changes around the MLE.

In Fig 29, the global maximum of the likelihood is achieved when we take $\rho_{12} = 0.511$, $\rho_{13} = -0.374$ with the fixed value of $\rho_{23} = -0.555$, but the local maximums are achieved
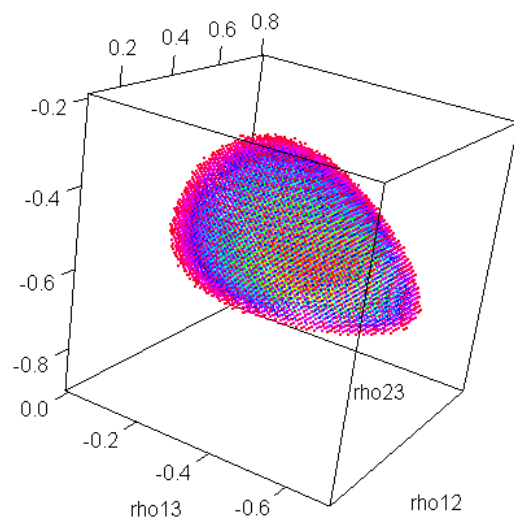
Figure 22: 95% Likelihood Region of Three Correlations
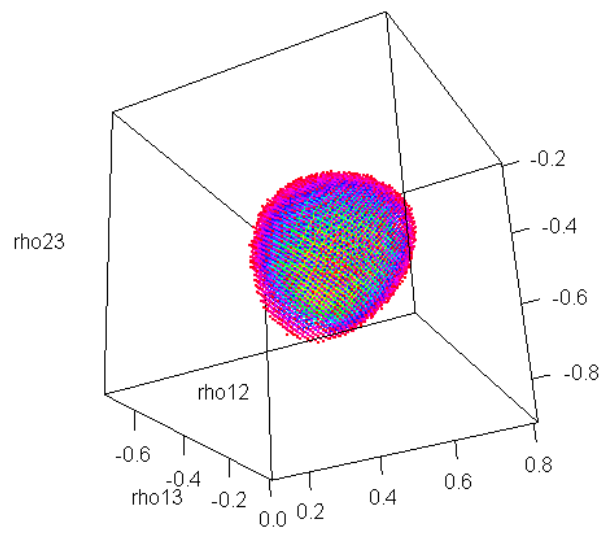
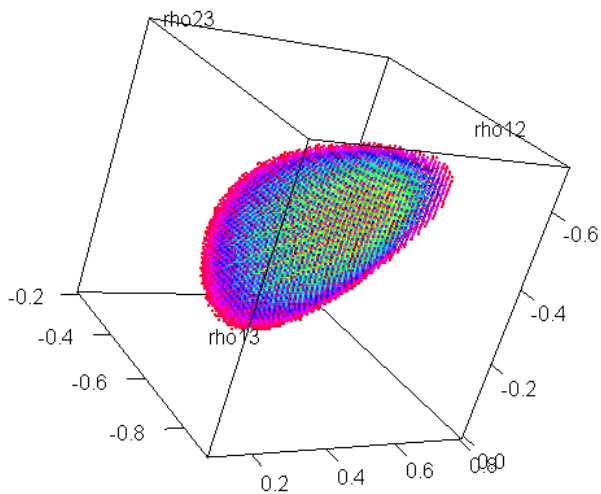Figure 23: 95% Likelihood Region of Three Correlations (Different Perspective I)

Figure 24: 95% Likelihood Region of Three Correlations (Different Perspective II)
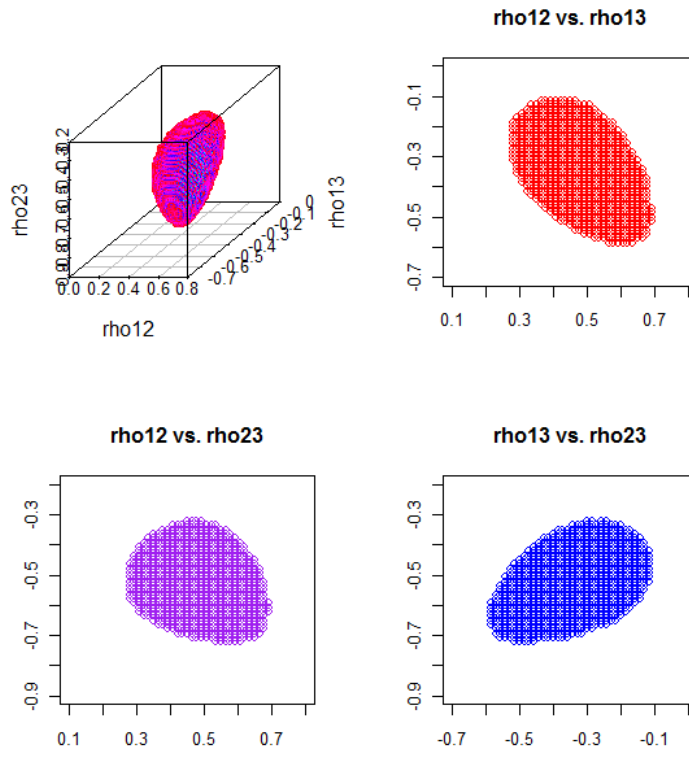
Figure 25: 95% Likelihood Region and its Two-Dimenional Projections
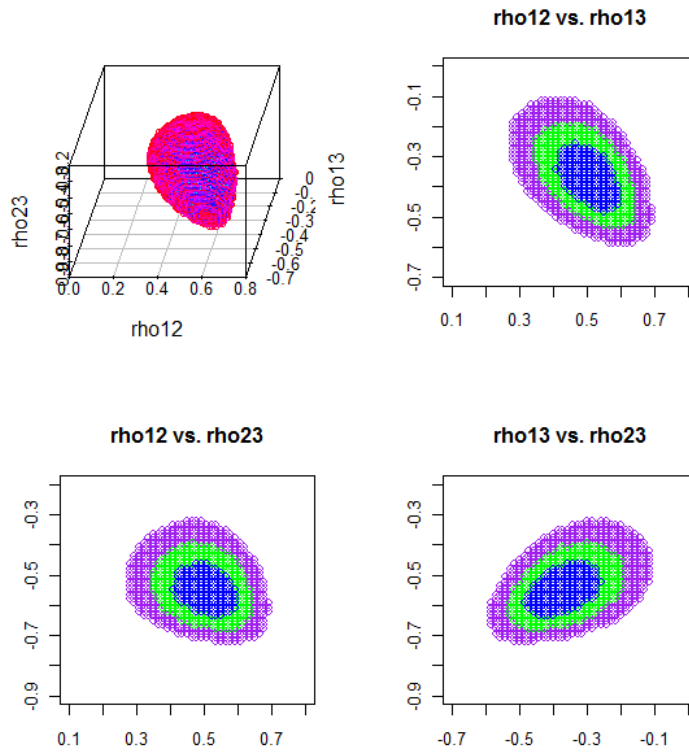
Figure 26: 95% Likelihood Region and Two-Dimensional Projections of 30%, 70% and 95% Likelihood Regions
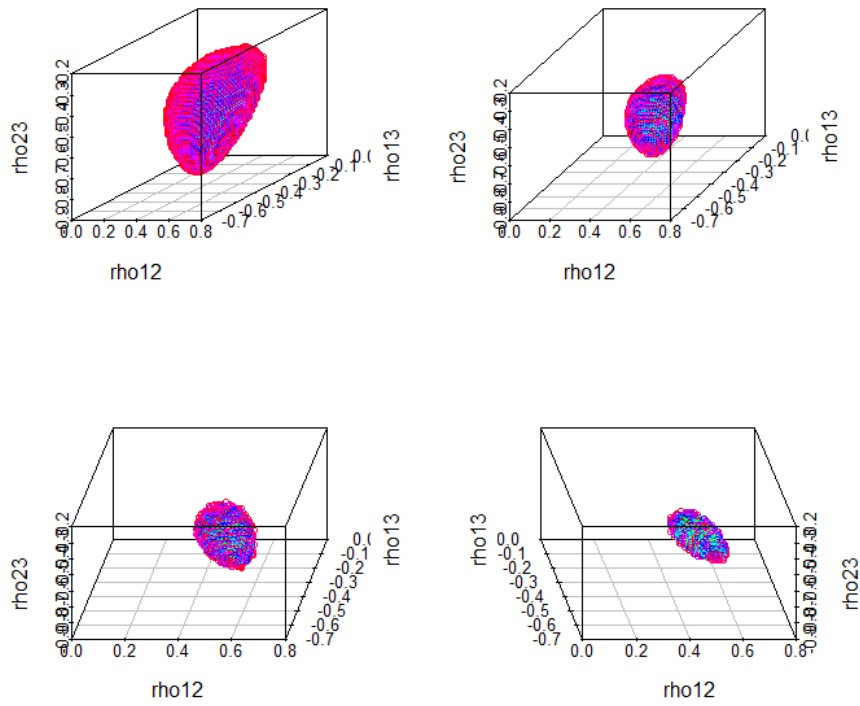
Figure 27: 99%, 70%, 50% and 30% Likelihood Regions

when we take $\rho_{12} = 0.586$, $\rho_{13} = -0.5$ with the fixed value of $\rho_{23} = -0.755$ and when we take $\rho_{12} = 0.471$, $\rho_{13} = -0.286$ with the fixed value of $\rho_{23} = -0.355$ in Fig 28 and Fig 30, respectively. The global maximum from Fig 29 is 1184.4 which is far greater than the local maximum 4.7 from Fig 28 and the local maximum 72.2 from Fig 30.

Fig 31 is a comprehensive summary of three different cases above, which practically combine contours and corresponding likelihoods.

4.8   SIMULATION

In this section, a simple simulation is conducted to compare the accuracies of the MAWA, the JAWA and the MLE in terms of Mean Squared Error (MSE hereafter).

The MSE of $\widetilde{P}$ of with respect to P is defined as:

$$MSE(\widetilde{P}) = E\left\{tr\left[(\widetilde{P} - P)'(\widetilde{P} - P)\right]\right\} = E\left[\sum_{1 \leq i < i \leq p} (\widetilde{\rho}_{ij} - \rho_{ij})^2\right] \qquad (4.23)$$

where $P = \left[\rho_{ij}\right]$ is the $p \times p$ population matrix and $\widetilde{P} = \left[\widetilde{\rho}_{ij}\right]$ is the estimator matrix of P. ($\widetilde{P}$ can be the MAWA or the JAWA or the MLE in this section.)

In the simulation, four different cases are assumed.   The basic information about the cases is given in Table 9.

Each case consists of five studies with the corresponding hypothesized three-dimensional population correlation matrix and the sample sizes. For example, case 3 has the population correlations $\rho_{12} = 0.7$, $\rho_{13} = 0.6$ and $\rho_{23} = 0.4$ and the five studies have the sample sizes $10, 10, 15, 15$ and 100, respectively.

In each case, the MAWA, the JAWA and the MLE is calculated 100 times and the corresponding MSEs are also calculated.

Table 10 shows the MSEs of the MAWA, the JAWA and the MLE for each case. The MLE shows consistently the best performance whereas the JAWA shows the worst performance. Case 2 shows that the three estimators work well when the sample sizes are large enough. Case 4 shows that all of the estimators have big errors when sample sizes are small, even
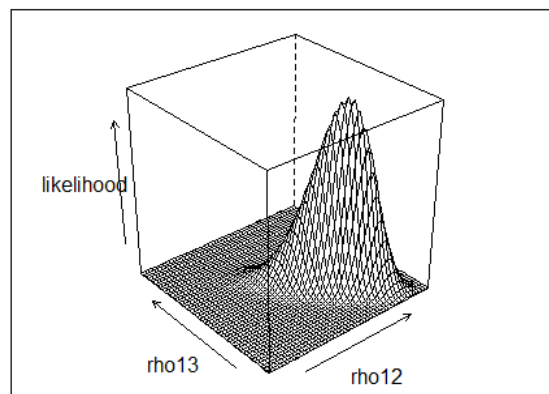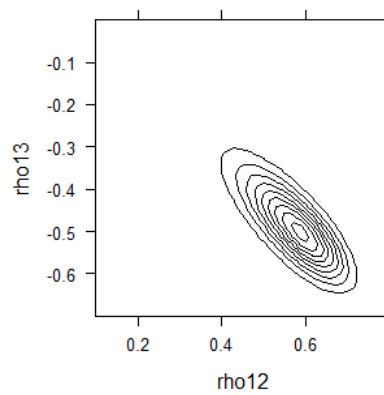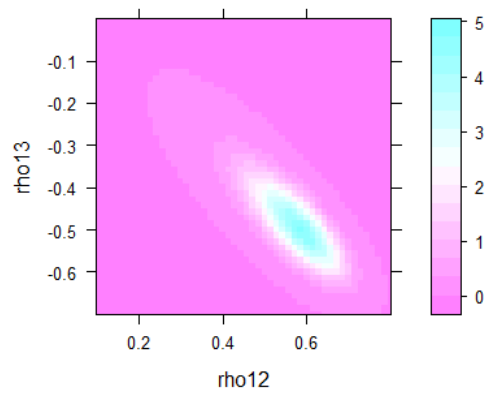
Figure 28: Level plot, contour plot and conditional likelihood I
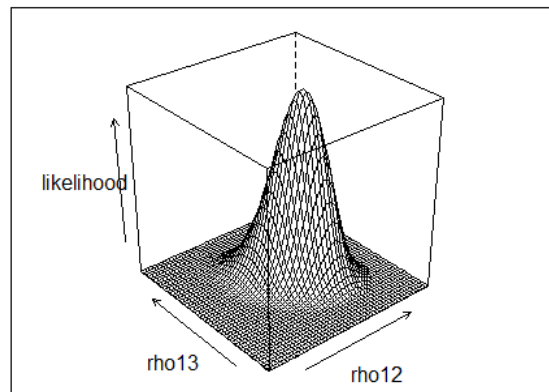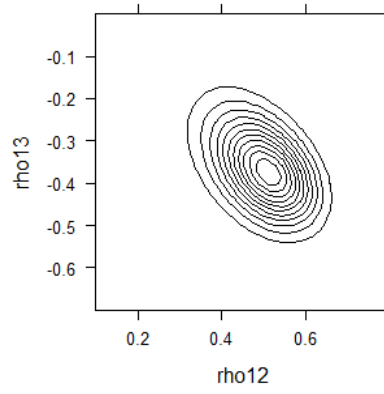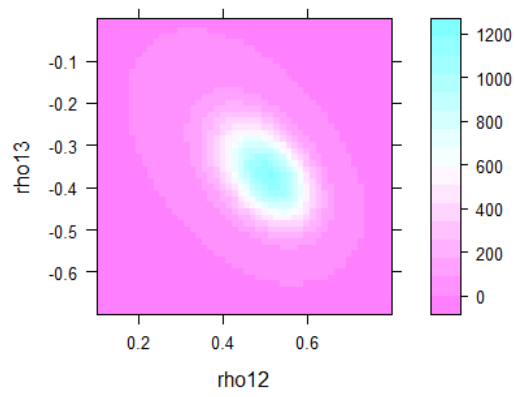
Figure 29: Level plot, contour plot and conditional likelihood II
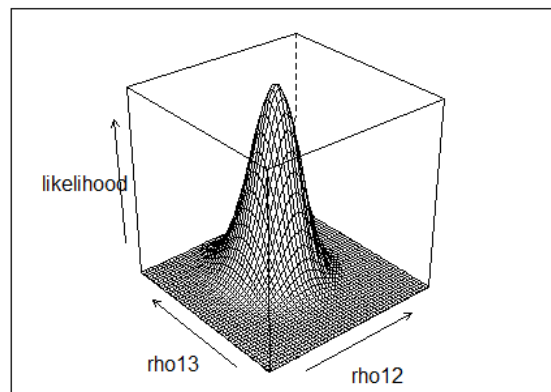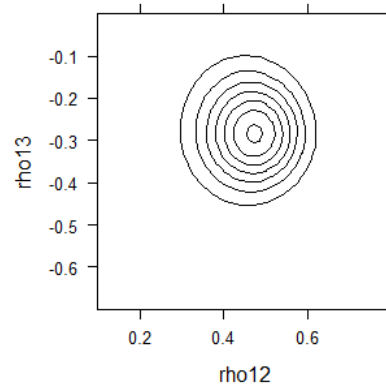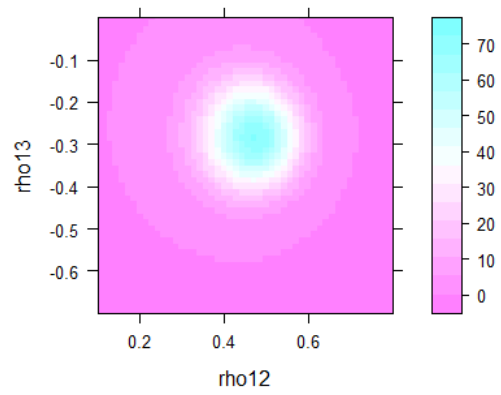
Figure 30: Level plot, contour plot and conditional likelihood III
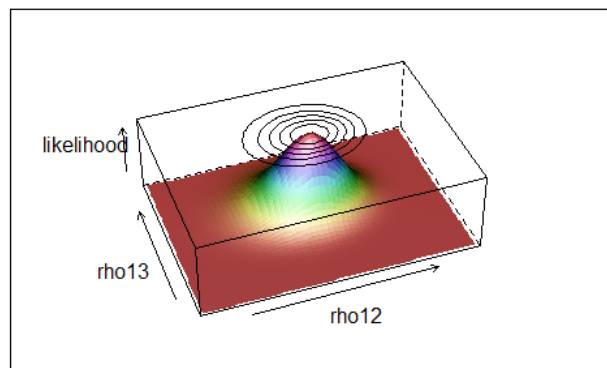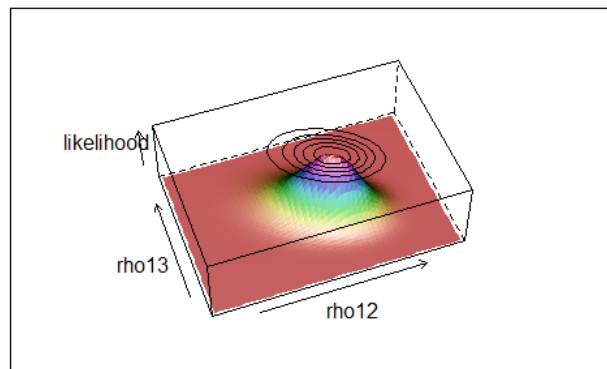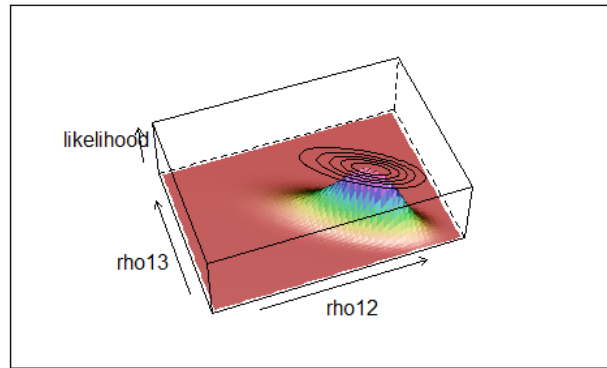
Figure 31: Three-dimensional view of Correlations and their corresponding conditional likelihoods and contours

Case	Population Correlations	Sample Sizes
1	$\rho_{12} = \rho_{13} = \rho_{23} = 0.7$	$n_1 = n_2 = n_3 = n_4 = n_5 = 10$
2	$\rho_{12} = \rho_{13} = \rho_{23} = 0.7$	$n_1 = n_2 = n_3 = n_4 = n_5 = 100$
3	$\rho_{12} = 0.7, \rho_{13} = 0.6, \rho_{23} = 0.4$	$n_1 = n_2 = 10, n_3 = n_4 = 15, n_5 = 100$
4	$\rho_{12} = 0.7, \rho_{13} = 0.6, \rho_{23} = 0.4$	$n_1 = n_2 = n_3 = 10, n_4 = n_5 = 15$

Table 9: Four Cases Used for Simulation

though the MLE is still the best among the three estimators. Interestingly enough, the MAWA is better than the JAWA for every case.

4.9   DISCUSSION

In meta-analysis, asymptotic normal approximation approaches have been used to combine information about population correlation matrices from many different independent studies. But these approaches are questionable when we do not have large enough sample sizes. To overcome this problem, the likelihood approach using numerical integration to calculate the likelihood is used in this paper. In relatively low dimensional spaces, say those for 3×3 or 4×4 correlation matrices, numerical integration works very well. But as dimension increases, the calculation becomes more difficult and cumbersome. Thus, the following are

Case	MAWA	JAWA	MLE
1	0.0487	0.0836	0.0194
2	0.0021	0.0025	0.0018
3	0.0236	0.0364	0.0111
4	0.0579	0.1026	0.0269

Table 10: The MSE's from the MAWA, the JAWA and the MLE

two limitations of the methods used in this paper:

1) Because it is really difficult to get the first and second derivatives in a form useful for Newton's method of root- finding, numerical methods using grid points to find the MLE and construct the confidence and the likelihood regions had to be used. Such an approach is computer intensive, especially when p is large.

2) The resolutions of our confidence and likelihood regions plots are not very high. It is sometimes hard to detect subtle changes of shape because when grid points are used to make the plots, there is a trade-off between high resolution and use of an economical number of grid points for calculation.

Despite these limitations, the approach used in this paper is quite promising. The main ideas can be summarized as: 1) Derivation of the likelihood function for the population correlation matrix as a certain integral; 2) Calculation of the likelihood with numerical integration based on wise choices of grid points to balance precision of resolution and computational effort; and 3) Use of graphs to visualize many aspects of the likelihood function. These three steps above can be applied in many situations provided that we have way to calculate the likelihood at individual particular points.

5.0 FUTURE WORK

5.1 BIVARIATE CASE

We need to improve the behavior of B-splines as approximations to the likelihood function for ρ when the sample correlation r is close to 1 in magnitude. B-spline approximations show extremely good performance when they are used for r in intermediate range - according to our experience, when $|r| < 0.8$. But they become worse as r becomes close to 1 in magnitude. It can be easily checked that the characteristics of likelihood function of ρ depend heavily on r; for example, if you choose $r = 0.9$, the likelihood of ρ will have a peak near $r = 0.9$, show heavy tail on left hand side and extremely steep slope on right hand side. Of course, many other explanations can be made for this symptom. Some improvement can be expected by applying other splines like natural cubic splines which show better behaviors at boundaries or by using non-equidistant grid points for problematic steep slope areas as a compensation. But it must be admitted that there are obvious limitations in improvement.

We also need to make a guide to tell users whether or not B-spline approximations are useful in various situations. Tables will be the optimal format for this purpose. With tables including many different choices of r and n, potential users may have a clue for their choices and have a chance to stimulate a good decision of using B-spline approximations.

5.2 MULTIVARIATE CASE

Sometimes it is impossible to get information about r_{12}, r_{13} and r_{23} from every study, then how to combine the information from all the studies is a natural question for the next step.

For example, suppose there are 10 studies having all of r_{12}, r_{13} and r_{23}, 5 studies having r_{12} and 5 studies having r_{13}, and it is desired to combine the information from these 20 studies. Under the assumption of the same population correlation matrices over all the studies, this goal can be achieved easily by multiplying the corresponding trivariate and bivariate likelihoods. Comparison between this approach and the large-sample approach described by Becker [3] will be of interest.

Expanded visualization ot the likelihood function can be considered. For example, the likelihood strip in one dimensional space can be generalized to the corresponding version in two dimensional space - 'likelihood contours' which may look like mono-colored or multi-colored spectrum contours. It can be even generalized further to three dimensional space by making 'likelihood cloud' with many different colors.

Finally, it is quite important to develop computing methods for the likelihood integral in the multivariate case. Among the three main methods used in multivariate case, the MC method is the fastest. But MC becomes slower as the dimension increases. Other methods need to developed to speed up getting the results

Beside the ideas and directions described above, many other good ideas will arise in the future because the topic in this paper is just at the starting point of research.

BIBLIOGRAPHY

[1] Anderson, T. W. (2003), *An introduction to multivariate statistical analysis* (3rd ed.), Hoboken: Wiley.

[2] Barrowman, N. J., and Myers, R. A. (2003), "Raindrop Plots: A New Way to Display Collections of Likelihoods and Distributions", *The American Statistician*, 57, 268-275.

[3] Becker, B. J. (1992), "Using Results From Replicated Studies to Estimate Linear Models", *Journal of Educational and Behavioral Statistics,* 17(4), 341-362

[4] Caruso, C. M., Dzewaltowski, D. A., Gill, D. L., McElroy, M. A. (1990), "Psychological and physiological changes in competitive state anxiety during noncompetition and competitive success and failure", *Journal of Sport and Exercise Psychology*, 12, 6-20.

[5] Cox, D. R., and Hinkley, D. V. (1974), *Theoretical Statistics*, London: Chapman & Hall.

[6] Craft, L. L., Magyar, T. M., Becker, B. J., Feltz, D. L. (2003), "The relationship between the Competitive State Anxiety Inventory-2 and sport performance: A meta-analysis", *Journal of Sport and Exercise Psychology,* 35, 44-65.

[7] de Boor, C. (2001), *A Practical Guide to Splines*, Verlag: Springer.

[8] Dehghan, M., Akhtar-Danesh, N., McMillan, C. R., and Thabane, L. (2007), "Is plasma vitamin C an appropriate biomarker of vitamin C intake? A systematic review and meta-analysis", *Nutrition Journal,* 6, 41-52.

[9] Edwards, T. and Hardy, L. (1996), "The interactive effects of intensity and direction of cognitive and somatic anxiety and self-confidence upon performance", *Journal of Sport and Exercise Psychology*, 18, 296-312.

[10] Evans, G. (1993), *Practical Numerical Integration*, New York: Wiley.

[11] Evans, M., and Swartz, T. (1995), "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems", *Statistical Science*, 10(4), 254-272

[12] Field, A. (2001), "Meta-Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed-and Random-Effects Methods." Psychological Methods 6(2), 161-180.

[13] Fisher, R. A. (1915), "Frequent Distribution of the Values of the Correlation Coefficients in Samples from an Indefinitely Large Population", *Biometrika,* 10, 507-521.

[14] Fisher, R. A. (1921), "On the probable error of a coefficient of correlation deduced from a small sample", *Metron,* 1, 3-32.

[15] Fisher, R. A. (1922), "On the Mathematical Foundations of Theoretical Statistics", *Philosophical Transactions of the Royal Society of London,* A, 222, 309-368.

[16] Fisher, R. A. (1962), "The simultaneous distribution of correlation coefficients", *Sankhya: The Indian Journal of Statistics,* A, 1-8.

[17] Friedman, L. (1994), "Meta-analytic contributions to the study of gender differences in mathematics–the relationship of mathematical and spatial skills", *International Journal of Educational Research,* 21, 4, 361-371.

[18] Genz, A., and Malik, A. A. (1980). "An adaptive algorithm for numerical integration over an N-dimensional rectangular region", *Journal of Computational and Applied Mathematics,* 6, 295-302.

[19] Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* New York: Springer.

[20] Hedges, L.V., and Olkin, I. (1985), *Statistical methods for meta-analysis,* Orlando: Academic Press.

[21] Hotelling, H. (1953), "New light on the correlation coefficient and its transforms", *Journal of the Royal Statistical Society,* B, 15, 193-232.

[22] Hunter, J. E., and Schmidt, F. L. (2004), *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

[23] Hyndman, R. J. (1996), "Computing and Graphing Highest Density Regions", *The American Statistician,* 50, 20-126.

[24] Jackson, Christopher. (2008), "Displaying Uncertainty With Shading", *The American Statistician,* 62(4), 340-347.

[25] Kooperberg, C., and Stone, C. J. (1991), "A study of logspline density estimation", *Computational Statistics & Data Analysis,* 12, 327-347.

[26] Kuonen, D. (2003), "Numerical integration in S-PLUS or R: A survey", Journal of Statistical Software, 8, 1-14.

[27] Kraemer, H. (1975), "On estimation and hypothesis testing problems for correlation coefficients", *Psychometrika* 40, 473-485.

[28] Ligges, U., and Maechler, M. (2003), "Scatterplot3d – an R Package for Visualizing Multivariate Data", *Journal of Statistical Software* 8(11), 1–20.

[29] Martens, R., Vealey, R. S., Burton, D. (1990) *Competitive anxiety in sport*, Human Kinetics.

[30] Maynard, I. W., Smith, M. J., Warwick-Evans, L. (1995), "The effects of a cognitive intervention strategy on competitive state anxiety and performance in semiprofessional soccer players", *Journal of Sport and Exercise Psychology*, 17. 428-446.

[31] Olkin, I., and Pratt, J. W. (1958), "Unbiased estimation of certain correlation coefficients", *Annals of Mathematical Statistics* 29, 201-211.

[32] Olkin, I., and Siotani, M. (1976), "Asymptotic distribution of functions of a correlation matrix", In S.Ideka (Ed.), *Essays in Probability and Statistics* 235-251. Tokyo: Sinko Tsusho.

[33] Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P. (2001). *Numerical Recipes: The Art of Scientific Computing*, Cambridge: Cambridge University Press.

[34] Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods,* New York: Springer.

[35] Royall, R. M. (1997), *Statistical Evidence: A Likelihood Paradigm*, New York: Chapman & Hall.

[36] Schott, J. R. (2005), *Matrix analysis for statistics*, New York: Wiley

[37] Steinmetz, K. A., and Potter. J. D. (1991), "Vegetables, fruit, and cancer. II. Mechanisms", *Cancer Causes Control*, 2, 427-442

[38] Stone, C. J. (1990), "Large-sample inference for log-spline models", *Annals of Statistics*, 18, 717-741.

[39] Tanner, M. A. (1996), *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions* (3rd ed.), New York: Springer.

[40] Venables, W. N., and Ripley, B. D. (1997), *Modern Applied Statistics with S-Plus*, Berlin: Springer.

[41] Wood, S. N. (2006), *Generalized additive models : an introduction with R*, Boca Raton: Chapman & Hall.

APPENDIX A

MATHEMATICAL DERIVATIONS

In this chapter, the mathematical derivations of the density functions or likelihood function are shown.

A.1 DERIVATION OF THE DENSITY OF R IN TRIVARIATE CASE

We will derive the density function of R when P is given.

where

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix} \text{ is the sample correlation matrix} \tag{A.1}$$

and

$$P = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \text{ is the population correlation matrix} \tag{A.2}$$

We can start from the distribution of S

where

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{bmatrix} \text{ is the scatter matrix} \tag{A.3}$$

The density of S is

$$\frac{|S|^{\frac{1}{2}(n-4)}e^{-\frac{1}{2}tr(P^{-1}S)}}{2^{\frac{3}{2}n}\pi^{\frac{3}{2}}|P|^{\frac{1}{2}n}\Gamma\left[\frac{1}{2}n\right]\Gamma[\frac{1}{2}(n-1)]\Gamma[\frac{1}{2}(n-2)]}$$

(A.4)

where $n = N - 1$ and N is sample size.

It can be easily calculated that

$$P^{-1} = \frac{1}{|P|}\begin{bmatrix} 1 - \rho_{23}^2 & \rho_{13}\rho_{23} - \rho_{12} & \rho_{12}\rho_{23} - \rho_{13} \\ \rho_{13}\rho_{23} - \rho_{12} & 1 - \rho_{13}^2 & \rho_{12}\rho_{13} - \rho_{23} \\ \rho_{12}\rho_{23} - \rho_{13} & \rho_{12}\rho_{23} - \rho_{23} & 1 - \rho_{12}^2 \end{bmatrix}$$

(A.5)

where

$$|P| = 1 - (\rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2 - 2\rho_{12}\rho_{13}\rho_{23})$$

(A.6)

and also

$$P^{-1}S = P^{-1}D_S^{\frac{1}{2}}RD_S^{\frac{1}{2}}$$

(A.7)

and

$$D_S^{\frac{1}{2}} = \begin{bmatrix} \sqrt{s_{11}} & 0 & 0 \\ 0 & \sqrt{s_{22}} & 0 \\ 0 & 0 & \sqrt{s_{33}} \end{bmatrix}$$

(A.8)

so

$$tr(P^{-1}S) = \frac{1}{|P|}[(1 - \rho_{23}^2)s_{11} + (1 - \rho_{13}^2)s_{22} + (1 - \rho_{12}^2)s_{33}$$

$$- 2\{r_{12}\left(\rho_{12} - \rho_{13}\rho_{23}\right)\sqrt{s_{11}s_{22}} + r_{13}\left(\rho_{13} - \rho_{12}\rho_{23}\right)\sqrt{s_{11}s_{33}}$$

(A.9)

$$+ r_{23}\left(\rho_{23} - \rho_{12}\rho_{13}\right)\sqrt{s_{22}s_{33}}\}]$$

The density of S can be rewritten as

$$C_1 \cdot s_{11}^{\frac{1}{2}(n-4)} s_{22}^{\frac{1}{2}(n-4)} s_{22}^{\frac{1}{2}(n-4)}$$

$$\exp[-\frac{1}{2|P|}\{(1-\rho_{23}^2)s_{11} + (1-\rho_{13}^2)s_{22} + (1-\rho_{12}^2)s_{33}\}$$

$$+\frac{1}{|P|}\{r_{12}\left(\rho_{12}-\rho_{13}\rho_{23}\right)\sqrt{s_{11}s_{22}} + r_{13}\left(\rho_{13}-\rho_{12}\rho_{23}\right)\sqrt{s_{11}s_{33}}$$

$$+r_{23}\left(\rho_{23}-\rho_{12}\rho_{13}\right)\sqrt{s_{22}s_{33}}\}] \tag{A.10}$$

where

$$C_1 = \frac{|R|^{\frac{1}{2}(n-4)}}{2^{\frac{3}{2}}\pi^{\frac{3}{2}}|P|^{\frac{1}{2}n}\Gamma[\frac{1}{2}n]\Gamma[\frac{1}{2}(n-1)]\Gamma[\frac{1}{2}(n-2)]} \tag{A.11}$$

Since

$$\frac{\partial s_{ij}}{\partial r_{ij}} = \sqrt{s_{ii}}\sqrt{s_{jj}}, \text{ for } i,\ j = 1,2,3 \text{ and } i < j \tag{A.12}$$

the density of s_{11}, s_{22}, $s_{33,}$ r_{12}, r_{13}, and r_{23} is

$$C_1 \cdot s_{11}^{\frac{1}{2}n-1} s_{22}^{\frac{1}{2}n-1} s_{22}^{\frac{1}{2}n-1}$$

$$\exp[-\frac{1}{2|P|}\{(1-\rho_{23}^2)s_{11} + (1-\rho_{13}^2)s_{22} + (1-\rho_{12}^2)s_{33}\}$$

$$+\frac{1}{|P|}\{r_{12}\left(\rho_{12}-\rho_{13}\rho_{23}\right)\sqrt{s_{11}s_{22}} + r_{13}\left(\rho_{13}-\rho_{12}\rho_{23}\right)\sqrt{s_{11}s_{33}}$$

$$+r_{23}\left(\rho_{23}-\rho_{12}\rho_{13}\right)\sqrt{s_{22}s_{33}}\}] \tag{A.13}$$

We know that

$$\exp\left\{\frac{r_{ij}\left(\rho_{ij}-\rho_{ik}\rho_{jk}\right)\sqrt{s_{ii}s_{jj}}}{|P|}\right\} = \sum_{\delta=0}^{\infty}\frac{\{r_{ij}\left(\rho_{ij}-\rho_{ik}\rho_{jk}\right)\sqrt{s_{ii}s_{jj}}\}^{\delta}}{\delta!|P|^{\delta}} \tag{A.14}$$

$$\text{for } i,\ j,\ k = 1,2,3 \text{ and } i < j,\ i \neq k,\ j \neq k$$

Now the density of s_{11}, s_{22}, s_{33}, r_{12}, r_{13}, and r_{23} can be written as

$$C_1 \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} \sum_{\gamma=0}^{\infty} \frac{r_{12}^{\alpha} r_{13}^{\beta} r_{23}^{\gamma}}{\alpha! \beta! \gamma! |P|^{\alpha+\beta+\gamma}}$$

$$\left(\rho_{12} - \rho_{13}\rho_{23} \right)^{\alpha} \left(\rho_{13} - \rho_{12}\rho_{23} \right)^{\beta} \left(\rho_{23} - \rho_{12}\rho_{13} \right)^{\gamma}$$

$$\exp\left[-\frac{(1-\rho_{23}^2)s_{11}}{2|P|} \right] s_{11}^{\frac{1}{2}(n+\alpha+\beta)-1} \exp\left[-\frac{(1-\rho_{13}^2)s_{22}}{2|P|} \right] s_{22}^{\frac{1}{2}(n+\alpha+\gamma)-1}$$

$$\exp\left[-\frac{(1-\rho_{12}^2)s_{33}}{2|P|} \right] s_{11}^{\frac{1}{2}(n+\beta+\gamma)-1}$$

Since

$$\int_0^{\infty} \exp\left[-\frac{(1-\rho_{ij}^2)s_{kk}}{2|P|} \right] s_{kk}^{\frac{1}{2}(n+\delta+\eta)-1} ds_{kk}$$

$$= \Gamma\left[\tfrac{1}{2}(n+\delta+\eta) \right] \left[\frac{2|P|}{(1-\rho_{ij}^2)} \right]^{\frac{1}{2}(n+\delta+\eta)} \tag{A.16}$$

for i, j, $k = 1,2,3$ and $i < j$, $i \neq k$, $j \neq k$

and the term-by-term integration of (A.16) with respect to s_{11}, s_{22} and s_{33} are permissible, the density of R is

$$C_1 \sum_{\alpha=0}^{\infty} \sum_{\beta=0}^{\infty} \sum_{\gamma=0}^{\infty} \frac{r_{12}^{\alpha} r_{13}^{\beta} r_{23}^{\gamma}}{\alpha! \beta! \gamma! |P|^{\alpha+\beta+\gamma}} 2^{\frac{3}{2}n+\alpha+\beta+\gamma} |P|^{\frac{3}{2}n+\alpha+\beta+\gamma}$$

$$\left(\rho_{12} - \rho_{13}\rho_{23} \right)^{\alpha} \left(\rho_{13} - \rho_{12}\rho_{23} \right)^{\beta} \left(\rho_{23} - \rho_{12}\rho_{13} \right)^{\gamma}$$

$$\Gamma\left[\tfrac{1}{2}(n+\alpha+\beta) \right] \Gamma\left[\tfrac{1}{2}(n+\alpha+\gamma) \right] \Gamma\left[\tfrac{1}{2}(n+\beta+\gamma) \right]$$

$$\left[\frac{1}{1-\rho_{23}^2} \right]^{\frac{1}{2}(n+\alpha+\beta)} \left[\frac{1}{1-\rho_{13}^2} \right]^{\frac{1}{2}(n+\alpha+r)} \left[\frac{1}{1-\rho_{12}^2} \right]^{\frac{1}{2}(n+\beta+\gamma)}$$

(A.15)

(A.17)

75

If we input (A.11) to (A.17) and rearrange the terms, the density of R can be expressed as

$$\frac{|R|^{\frac{1}{2}(n-4)}|P|^n}{\pi^{\frac{3}{2}}}\sum_{\alpha=0}^{\infty}\sum_{\beta=0}^{\infty}\sum_{\gamma=0}^{\infty}\frac{2^{\alpha+\beta+\gamma}r_{12}^\alpha r_{13}^\beta r_{23}^\gamma}{\alpha!\beta!\gamma!}$$

$$\frac{(\rho_{12}-\rho_{13}\rho_{23})^\alpha(\rho_{13}-\rho_{12}\rho_{23})^\beta(\rho_{23}-\rho_{12}\rho_{13})^\gamma}{(1-\rho_{23}^2)^{\frac{1}{2}(n+\alpha+\beta)}(1-\rho_{13}^2)^{\frac{1}{2}(n+\alpha+\gamma)}(1-\rho_{12}^2)^{\frac{1}{2}(n+\beta+\gamma)}} \tag{A.18}$$

$$\frac{\Gamma\left[\frac{1}{2}(n+\alpha+\beta)\right]\Gamma\left[\frac{1}{2}(n+\alpha+\gamma)\right]\Gamma\left[\frac{1}{2}(n+\beta+\gamma)\right]}{\Gamma\left[\frac{1}{2}n\right]\Gamma\left[\frac{1}{2}(n-1)\right]\Gamma\left[\frac{1}{2}(n-2)\right]}$$

This type of derivation can be extended to general p dimensional case. We can easily detect that if we have a p dimensional case ,we have to deal with $\frac{p(p-1)}{2}$ combined infinite series if I use a classical derivation of the likelihood function from Wishart distribution above.

A.2   DERIVATION OF THE LIKELIHOOD FUNCTION OF CORRELATION MATRIX P

We will derive the likelihood function of correlation matrix P when sample correlation matrix R and sample size N are given.

Let $\mathbf{Z}_1, \cdots, \mathbf{Z}_N$ be independently distributed, each following $N(\mathbf{0}, P)$ where $Z_i : p \times 1$ for $i = 1, \cdots, N$,

The density of $S = \sum_{\alpha=1}^{N}\mathbf{Z}_\alpha\mathbf{Z}_\alpha' = (s_{ij})_{p\times p}$, where $s_{ii} \equiv s_i$ for convention, is

$$\frac{|S|^{\frac{1}{2}(n-p-1)}e^{-\frac{1}{2}tr(P^{-1}S)}}{2^{\frac{np}{2}}\pi^{\frac{p(p-1)}{4}}|P|^{\frac{n}{2}}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(n+1-i)\right]} \tag{A.19}$$

where $n \equiv N - 1$

for S positive definite, and 0 otherwise.(Anderson[1])

Here

$$|S| = \left|D_s^{\frac{1}{2}}RD_s^{\frac{1}{2}}\right| = |D_s||R| = |R|\prod_{i-1}^{p}s_i \text{ and } P^{-1}S = P^{-1}D_s^{\frac{1}{2}}RD_s^{\frac{1}{2}} \tag{A.20}$$

76

where

$$D_s = diag\,(s_{1,}\cdots,s_p) \tag{A.21}$$

Since

$$\frac{\partial s_{ij}}{\partial r_{ij}} = \sqrt{s_i}\sqrt{s_j},\ \text{ for } i,\ j = 1,\cdots,p \text{ and } i < j \tag{A.22}$$

If we conduct the transformation from S to $\{s_{1,}\cdots,s_p\}$ and R, then the transformation matrix is $\frac{p(p-1)}{2} \times \frac{p(p-1)}{2}$ and the Jacobian is

$$J = \prod_{i<j} s_i^{\frac{1}{2}} s_j^{\frac{1}{2}} = \prod_{i=1}^{p} s_i^{\frac{p-1}{2}} \tag{A.23}$$

So the density of $\{s_{1,}\cdots,s_p\}$ and R can be expressed as

$$C_1 \cdot \left(\prod_{i=1}^{p} s_i\right)^{\frac{n}{2}-1} e^{-\frac{1}{2}tr(P^{-1}D_s^{\frac{1}{2}}RD_s^{\frac{1}{2}})} \tag{A.24}$$

where

$$C_1 = \frac{|R|^{\frac{1}{2}(n-p-1)}}{2^{\frac{np}{2}}\pi^{\frac{p(p-1)}{4}}|P|^{\frac{n}{2}}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}\,(n+1-i)\right]} \tag{A.25}$$

If we define

$$b_i \equiv \begin{cases} \displaystyle\sum_{j=1}^{i} s_j \Big/ \sum_{k=1}^{i+1} s_k & \text{for } i = 1,\cdots,\ p-1 \\[4mm] \displaystyle\sum_{k=1}^{p} s_k & \text{for } i = p \end{cases} \tag{A.26}$$

then we have an equivalent expression with respect to s_i's:

$$s_i = \begin{cases} \displaystyle\prod_{k=1}^{p} b_k & \text{for } i = 1 \\[4mm] \displaystyle(1 - b_{i-1})\prod_{j=i}^{p} b_j & \text{for } i = 2,\cdots,\ p \end{cases} \tag{A.27}$$

or simply

$$s_k = (1 - b_{k-1})\prod_{j=k}^{p} b_j \ \text{ for } k = 1,\cdots,\ p \ \text{ with } b_0 \equiv 0 \tag{A.28}$$

Now we can easily derive the partial derivatives:

$$
\frac{\partial s_k}{\partial b_j} =
\begin{cases}
0 & \text{for } j < k-1 \\[2ex]
-\prod\limits_{j=k}^{p} b_j & \text{for } j = k-1 \\[2ex]
\frac{(1-b_k)}{b_j} \prod\limits_{j=k}^{p} b_j & \text{for } j > k-1
\end{cases}
\tag{A.29}
$$

For the transformation from $\{s_1, \cdots, s_p\}$ to $\{b_1, \cdots, b_p\}$, the corresponding Jacobian is:

$$
J = b_2 \cdots \cdot b_{p-1}^{p-2} \cdot b_p^{p-1} = \prod_{k=2}^{p} b_k^{k-1}
\tag{A.30}
$$

Let

$$
v_i \equiv
\begin{cases}
(1-b_{i-1}) \prod\limits_{k=i}^{p-1} b_k & \text{for } i = 1, \cdots, \ p-1 \\[2ex]
1-b_{p-1} & \text{for } i = p
\end{cases}
\tag{A.31}
$$

Then

$$
D_s^{\frac{1}{2}} = diag\left(\sqrt{v_1 b_p,}\cdots, \sqrt{v_p b_p}\right) = \sqrt{b_p} D_b^{\frac{1}{2}}
\tag{A.32}
$$

where

$$
D_b = diag\left(v_1, \cdots, v_p\right)
\tag{A.33}
$$

Since $tr(P^{-1} D_s^{\frac{1}{2}} R D_s^{\frac{1}{2}}) = b_p \cdot tr(D_b^{\frac{1}{2}} P^{-1} D_b^{\frac{1}{2}} R)$, the density of $\{b_1, \cdots, b_p\}$ and R for given P can be expressed as

$$
C_1 \cdot \left(\prod_{i=1}^{p} v_i\right)^{\frac{n}{2}-1} b_p^{\frac{np}{2}-1} e^{-\frac{1}{2} tr(D_b^{\frac{1}{2}} P^{-1} D_b^{\frac{1}{2}} R) b_p} \prod_{k=2}^{p-1} b_k^{k-1}
\tag{A.34}
$$

If b_p is integrated out from (A.34), then the density of $\{b_1, \cdots, b_{p-1}\}$ and R for given P is

$$
C_1 \cdot \left(\prod_{i=1}^{p} v_i\right)^{\frac{n}{2}-1} \frac{\Gamma\left[\frac{np}{2}\right] 2^{\frac{np}{2}}}{\left[tr(D_b^{\frac{1}{2}} P^{-1} D_b^{\frac{1}{2}} R)\right]^{\frac{np}{2}}} \prod_{k=2}^{p-1} b_k^{k-1}
\tag{A.35}
$$

By Schott[36], We know that

$$
tr(D_b^{\frac{1}{2}} P^{-1} D_b^{\frac{1}{2}} R) = (\mathbf{v}^{\frac{1}{2}})'(P^{-1} \odot R) \mathbf{v}^{\frac{1}{2}}
\tag{A.36}
$$

78

where

$$\odot \text{ is Hadamard product and } \mathbf{v} \equiv (v_1, \cdots, v_p)' : p \times 1 \tag{A.37}$$

As a result the likelihood of P with R (or the density of R for given P) can be expressed as

$$C \cdot \int_0^1 \cdots \int_0^1 \frac{\left(\prod_{i=1}^p v_i \right)^{\frac{n}{2}-1} \cdot \prod_{k=2}^{p-1} b_k^{k-1}}{\left[(\mathbf{v}^{\frac{1}{2}})'(P^{-1} \odot R)\mathbf{v}^{\frac{1}{2}} \right]^{\frac{np}{2}}} db_{p-1} \cdots db_1 \tag{A.38}$$

where

$$
\begin{aligned}
C &= C_1 \cdot \Gamma \left[\tfrac{np}{2} \right] 2^{\frac{np}{2}} \\[2mm]
&= \frac{|R|^{\frac{n-p-1}{2}} \Gamma \left[\tfrac{np}{2} \right]}{\pi^{\frac{p(p-1)}{4}} |P|^{\frac{n}{2}} \prod_{i=1}^p \Gamma \left[\tfrac{1}{2} (n+1-i) \right]}
\end{aligned}
\tag{A.39}
$$

Furthermore, we can get an alternative form of (A.38) with

$$
\begin{aligned}
(\mathbf{v}^{\frac{1}{2}})'(P^{-1} \odot R)\mathbf{v}^{\frac{1}{2}} &= \sum_{i=1}^p \sum_{j=1}^p v_i^{\frac{1}{2}}(P^{-1} \odot R)v_j^{\frac{1}{2}} \\[2mm]
&= \sum_{i=1}^p \sum_{j=1}^p (P^{-1})_{ij} R_{ij} \sqrt{v_i v_j} \\[2mm]
&= \sum_{i=1}^p \sum_{j=1}^p (P^{-1})_{ii} v_i + 2 \sum_{i=2}^p \sum_{j=1}^{i-1} (P^{-1})_{ij} R_{ij} \sqrt{v_i v_j}
\end{aligned}
\tag{A.40}
$$

which is

$$C \cdot \int_0^1 \cdots \int_0^1 \frac{\left(\prod_{i=1}^p v_i \right)^{\frac{n}{2}-1} \cdot \prod_{k=2}^{p-1} b_k^{k-1}}{\left[\sum_{i=1}^p \sum_{j=1}^p (P^{-1})_{ii} v_i + 2 \sum_{i=2}^p \sum_{j=1}^{i-1} (P^{-1})_{ij} R_{ij} \sqrt{v_i v_j} \right]^{\frac{np}{2}}} db_{p-1} \cdots db_1 \tag{A.41}$$

APPENDIX B

R CODES

B.1 FUNCTIONS FOR GETTING THE MLE FROM NUMERICAL INTEGRATION

B.1.1 Getting the MLE by GL & Grid & 1-5-2 rule (3-dimensional)

I used basic functions GL.YW, Outer2 and GL.integrate.2D which is renamed to gl3d for convention from Kuonen[26] for numerical integration in this paper.

I developed the codes for expression of optimized form of integrand and iterative algorithm for finding the MLE.

```
## First, run the Codes(GL.YW, Outer2 and gl3d) below


## GL.YW
GL.YW <- function(M, xrange=NULL, epsilon=NULL){
if (is.null(epsilon)) epsilon <- .Machine$double.eps
if (M%%2 ==1) stop("M needs to be an even number")
MM <- (M + 1) / 2
Y <- W <- numeric(M)
```

```
for (i in 1:floor(MM)) {
ok<-F
z <- cos( pi * (i-0.25)/(M + 0.5) )
while (ok == F) {
p1 <- 1.0
p2 <- 0.0
for (j in 1:M) {
p3 <- p2
p2 <- p1
p1 <- ((2.0*j - 1.0)*z*p2 - (j - 1.0)*p3)/j
}
pp <- M*(z*p1 - p2)/(z^2 - 1.0)
z1 <- z
z <- z - p1/pp
if (abs(z - z1) < epsilon) ok<-T
}
Y[i] <- -z
Y[M+1-i] <- z
W[i] <- 2.0 / (( 1 - z^2) * pp^2)
W[M+1-i] <- W[i]
}
if(!is.null(xrange)) {
xL<- (xrange[2]-xrange[1])/2.0
W <- xL * W
Y <- xL * Y + (xrange[1]+xrange[2])/2.0
}
cbind(Y, W)
}
```

Outer2

```
outer2 <- function (X, Y, FUN="*", ...){
no.nx <- is.null(nx <- dimnames(X <- as.array(X)))
dX <- dim(X)
no.ny <- is.null(ny <- dimnames(Y <- as.array(Y)))
dY <- dim(Y)
if (is.character(FUN) && FUN=="*") {
robj <- as.vector(X) %*% t(as.vector(Y))
dim(robj) <- c(dX, dY)
} else {
match.fun <- function(FUN) return(FUN)
FUN <- match.fun(FUN)
Y <- rep(Y, rep(length(X), length(Y)))
X <- rep(X, length.out = length(Y))
robj <- array(FUN(X, Y, ...), c(dX, dY))
}
if (no.nx) nx <- vector("list", length(dX))
else if (no.ny) ny <- vector("list", length(dY))
if (!(no.nx && no.ny)) dimnames(robj) <- c(nx, ny)
robj
}


## GL for 2-dim integral (2 = 3-1)
gl3d <- function(fct, low=c(0,0), upp=c(1,1), order=10) {
YW.list<-as.list(1:2)
for(i in 1:2) {
name.YW<-paste("GL.YW", abs(low[i]), abs(upp[i]), order, sep=".")
if (!exists(name.YW)) {
assign(name.YW, GL.YW(order,
xrange=c(low[i],upp[i])), immediate=T)
}
```

```
YW.list[[i]]<- get(name.YW)
}
fcteval <- outer2(YW.list[[1]][,1], YW.list[[2]][,1], fct)
sum(YW.list[[1]][,2] * apply(YW.list[[2]][,2] *
fcteval, 2, sum))
}


### Getting the MLE

mle.gl3d = function(rn,mm) {
# mm >= 2 : accuracy = 0.1^mm, ex)mm=2 means 0.01 accuracy
## Making integrand for integration w.r.t (0,1)*(0,1) for GL in 3-dim
integ.gl3d = function(b1,b2, rho12,rho13,rho23, rn) {
## rn: 4 by 1 vector
n = rn[4]-1
rho = c(rho12,rho13,rho23)
## Making Matrices P & R
P = vec2sm(rho, diag = F)
r = rn[1:3]
R = vec2sm(r, diag = F)
P[1,1]=P[2,2]=P[3,3]=1
R[1,1]=R[2,2]=R[3,3]=1
## Making coefficient
if(det(P)>0.01) { ## Can change to make it efficient
coeff1 = 0.5*(n-4)*log(det(R))- 0.5*n*log(det(P))
sum1 = 0
for (i in 1:3){
temp1 = lgamma(0.5*(n+1-i))
sum1 = sum1+temp1
sum1
```

```
}
coeff2 = lgamma(0.5*n*3)-sum1
coeff3 = -0.25*3*2*log(pi)
coeff = coeff1 + coeff2 + coeff3
## 3. Making integrand
## For numerator
v1=b1*b2
v2=(1-b1)*b2
v3=1-b2
num =(n/2-1)*(log(v1)+log(v2)+log(v3))+log(b2)
## For denominator
Pv = solve(P,diag(1,3))
H = Pv*R # * : Hadamard Product
comp1 = Pv[1,1]*v1 + Pv[2,2]*v2 + Pv[3,3]*v3
comp2 = 2*(H[1,2]*sqrt(v1*v2)+H[1,3]*sqrt(v1*v3)+H[2,3]*sqrt(v2*v3))
den = (n/2)*3*log(comp1 + comp2)
logint = num - den
integ = exp(coeff)*exp(logint)
integ
}
else {
integ = 0
integ
}
integ
}


# Getting the row of MLE by 1-5-2 rule
max.grid = function(rho1,rho2,rho3){
n1 = length(rho1)
```

```
n2 = length(rho2)

n3 = length(rho3)

mat1 = matrix(0,nrow=n1*n2*n3, ncol=4)

for (k1 in 1:n1){

for (k2 in 1:n2){

for (k3 in 1:n3){

comb = 1

for (i in 1:length(rn[,1])){

ind = gl3d(function(x,y) integ.gl3d(x,y,rho1[k1],rho2[k2],rho3[k3],rn[i,]),

low=c(0,0),upp=c(1,1),order=128)

comb = comb*ind

comb

}

mat1[(k1-1)*n2*n3+(k2-1)*n3+k3,] <-c(rho1[k1], rho2[k2], rho3[k3], comb)

}

}

}

mat1[mat1[,4]==max(mat1[,4])]

}

rho12 =seq(round(min(rn[,1]),1)-.1, round(max(rn[,1]),1)+.1, by=.1)

rho13 =seq(round(min(rn[,2]),1)-.1, round(max(rn[,2]),1)+.1, by=.1)

rho23 =seq(round(min(rn[,3]),1)-.1, round(max(rn[,3]),1)+.1, by=.1)

maxt =max.grid(rho12,rho13,rho23)

nn =mm-2

for (k in 0:nn){

rho12 =seq(maxt[1]-.1*.1^k, maxt[1]+.1*.1^k, by=.05*.1^k)

rho13 =seq(maxt[2]-.1*.1^k, maxt[2]+.1*.1^k, by=.05*.1^k)

rho23 =seq(maxt[3]-.1*.1^k, maxt[3]+.1*.1^k, by=.05*.1^k)

maxt =max.grid(rho12,rho13,rho23)

rho12 =seq(maxt[1]-.05*.1^k, maxt[1]+.05*.1^k, by=.02*.1^k)
```

rho13 =seq(maxt[2]-.05*.1^k, maxt[2]+.05*.1^k, by=.02*.1^k)

rho23 =seq(maxt[3]-.05*.1^k, maxt[3]+.05*.1^k, by=.02*.1^k)

maxt =max.grid(rho12,rho13,rho23)

rho12 =seq(maxt[1]-.02*.1^k, maxt[1]+.02*.1^k, by=.01*.1^k)

rho13 =seq(maxt[2]-.02*.1^k, maxt[2]+.02*.1^k, by=.01*.1^k)

rho23 =seq(maxt[3]-.02*.1^k, maxt[3]+.02*.1^k, by=.01*.1^k)

maxt =max.grid(rho12,rho13,rho23)

maxt

}

rho12 =seq(maxt[1]-.01*.1^nn, maxt[1]+.01*.1^nn, by=.01*.1^k)

rho13 =seq(maxt[2]-.01*.1^nn, maxt[2]+.01*.1^nn, by=.01*.1^k)

rho23 =seq(maxt[3]-.01*.1^nn, maxt[3]+.01*.1^nn, by=.01*.1^k)

maxt =max.grid(rho12,rho13,rho23)

maxt

}

2. Using SAT score Data

rn1 = c(.46, .31, .19, 103)

rn2 = c(.46, .55, .32, 69)

rn3 = c(.40, .40, .18, 69)

rn4 = c(.27, .57, .22, 70)

rn = rbind(rn1,rn2,rn3,rn4)

for (103,69,69,70) (Accuracy 0.001)

start = proc.time()[3]

mle.gl3d(rn,2)

proc.time()[3]-start

> **mle.gl3d(rn,2)**

[1] **0.40 0.44 0.22 372036.01**

> proc.time()[3]-start

elapsed

488.83

for (7,5,7,5) (0.001)

> mle.gl3d(rn,3)

[1] 0.373 0.405 0.200 0.1546279

for (10,10,10,30) (0.001)

> mle.gl3d(rn,3)

[1] 0.341 0.491 0.219 110.4775

for (10,7,7,7) (0.001)

> mle.gl3d(rn,3)

[1] 0.382 0.419 0.210 2.117855

for (400,400,400,400) (0.01)

> mle.gl3d(rn,3)

[1] 0.398 0.462 0.229 53.43076

B.1.2   Getting the MLE by ADAPTIVE Algorithm & 1-5-2 rule

mle.ad3d <- function(rn,mm) {

    # mm >= 2 : accuracy = 0.1^mm, ex)mm=2 means 0.01 accuracy

    ## Making integrand for integration w.r.t (0,1)*(0,1) for MC in 3-dim

    integ.ad3d = function(b, rho12,rho13,rho23, rn) {

    ## rn: 4 by 1 vector

    n = rn[4]-1

    rho = c(rho12,rho13,rho23)

    ## 1. Making Matrices P & R

    P = vec2sm(rho, diag = F)

    r = rn[1:3]

    R = vec2sm(r, diag = F)

    P[1,1]=P[2,2]=P[3,3]=1

    R[1,1]=R[2,2]=R[3,3]=1

```
## 2. Making coefficient
if(det(P)>0.01) { ## Can change to make it efficient
coeff1 = 0.5*(n-4)*log(det(R))- 0.5*n*log(det(P))
sum1 = 0
for (i in 1:3){
temp1 = lgamma(0.5*(n+1-i))
sum1 = sum1+temp1
sum1
}
coeff2 = lgamma(0.5*n*3)-sum1
coeff3 = -0.25*3*2*log(pi)
coeff = coeff1 + coeff2 + coeff3
## 3. Making integrand
## For numerator
b1 = b[1]
b2 = b[2]
v1=b1*b2
v2=(1-b1)*b2
v3=1-b2
num =(n/2-1)*(log(v1)+log(v2)+log(v3))+log(b2)
## For denominator
Pv = solve(P,diag(1,3))
H = Pv*R # * : Hadamard Product
comp1 = Pv[1,1]*v1 + Pv[2,2]*v2 + Pv[3,3]*v3
comp2 = 2*(H[1,2]*sqrt(v1*v2)+H[1,3]*sqrt(v1*v3)+H[2,3]*sqrt(v2*v3))
den = (n/2)*3*log(comp1 + comp2)
logint = num - den
integ = exp(coeff)*exp(logint)
integ
}
```

```
else {

integ = 0

integ

}

integ

}


# Getting the row of the MLE by 1-5-2 rule

max.grid <- function(rho1,rho2,rho3){

n1 = length(rho1)

n2 = length(rho2)

n3 = length(rho3)

mat1 = matrix(0,nrow=n1*n2*n3, ncol=4)

for (k1 in 1:n1){

for (k2 in 1:n2){

for (k3 in 1:n3){

comb <- 1

for (i in 1:length(rn[,1])){

ind <- adapt(2, lo=c(0,0), up=c(1,1), functn = function(b) integ.ad3d(b,

rho1[k1],rho2[k2],rho3[k3],rn[i,]), eps=0.0001)

comb <- comb*(ind$value)

comb

}

mat1[(k1-1)*n2*n3+(k2-1)*n3+k3,] <-c(rho1[k1], rho2[k2], rho3[k3], comb)

}

}

}

mat1[mat1[,4]==max(mat1[,4])]

}

rho12 =seq(round(min(rn[,1]),1)-.1, round(max(rn[,1]),1)+.1, by=.1)
```

```
rho13 =seq(round(min(rn[,2]),1)-.1, round(max(rn[,2]),1)+.1, by=.1)

rho23 =seq(round(min(rn[,3]),1)-.1, round(max(rn[,3]),1)+.1, by=.1)

maxt =max.grid(rho12,rho13,rho23)

nn =mm-2

for (k in 0:nn){

rho12 =seq(maxt[1]-.1*.1^k, maxt[1]+.1*.1^k, by=.05*.1^k)

rho13 =seq(maxt[2]-.1*.1^k, maxt[2]+.1*.1^k, by=.05*.1^k)

rho23 =seq(maxt[3]-.1*.1^k, maxt[3]+.1*.1^k, by=.05*.1^k)

maxt =max.grid(rho12,rho13,rho23)

rho12 =seq(maxt[1]-.05*.1^k, maxt[1]+.05*.1^k, by=.02*.1^k)

rho13 =seq(maxt[2]-.05*.1^k, maxt[2]+.05*.1^k, by=.02*.1^k)

rho23 =seq(maxt[3]-.05*.1^k, maxt[3]+.05*.1^k, by=.02*.1^k)

maxt =max.grid(rho12,rho13,rho23)

rho12 =seq(maxt[1]-.02*.1^k, maxt[1]+.02*.1^k, by=.01*.1^k)

rho13 =seq(maxt[2]-.02*.1^k, maxt[2]+.02*.1^k, by=.01*.1^k)

rho23 =seq(maxt[3]-.02*.1^k, maxt[3]+.02*.1^k, by=.01*.1^k)

maxt =max.grid(rho12,rho13,rho23)

maxt

}

rho12 =seq(maxt[1]-.01*.1^nn, maxt[1]+.01*.1^nn, by=.01*.1^k)

rho13 =seq(maxt[2]-.01*.1^nn, maxt[2]+.01*.1^nn, by=.01*.1^k)

rho23 =seq(maxt[3]-.01*.1^nn, maxt[3]+.01*.1^nn, by=.01*.1^k)

maxt =max.grid(rho12,rho13,rho23)

maxt

}

————————-

## Getting the MLE

rn1 <- c(.46, .31, .19, 103)

rn2 <- c(.46, .55, .32, 69)

rn3 <- c(.40, .40, .18, 69)
```

rn4 <- c(.27, .57, .22, 70)

rn <- rbind(rn1,rn2,rn3,rn4)

library(corpcor)

> **mle.ad3d(rn,2)**

[1] 0.40 0.44 0.22 369249.06

> proc.time()[3]-start

elapsed

4055.84

B.1.3   Getting the MLE by Monte Carlo Method & 1-5-2 rule

mle.mc3d <- function(rn,mm,np) {

mm >= 2 : accuracy ex) mm = 2 means 0.01 accuracy

np : number of points used in specific axis

Monte Carlo Methods for 3-dim

3 dimensional (for p=3)

mc3d = function(ftn, low=c(0,0), upp=c(1,1),npoints=100) {

x = runif(n=npoints, min=low[1], max=upp[1])

y = runif(n=npoints, min=low[2], max=upp[2])

ftn.tmp = ftn(x, y)

v.tmp = diff(c(low[1], upp[1]))* diff(c(low[2], upp[2]))

mean.approx = mean(ftn.tmp) * v.tmp

mean.approx

}

Making integrand for integration w.r.t (0,1)*(0,1) for MC in 3-dim

integ.mc3d = function(b1,b2, rho12,rho13,rho23, rn) {

rn: 4 by 1 vector

n = rn[4]-1

91

rho = c(rho12,rho13,rho23)

1. Making Matrices P & R

P = vec2sm(rho, diag = F)

r = rn[1:3]

R = vec2sm(r, diag = F)

P[1,1]=P[2,2]=P[3,3]=1

R[1,1]=R[2,2]=R[3,3]=1

2. Making coefficient

if(det(P)>0.01) { ## Can change to make it efficient

coeff1 = 0.5*(n-4)*log(det(R))- 0.5*n*log(det(P))

sum1 = 0

for (i in 1:3){

temp1 = lgamma(0.5*(n+1-i))

sum1 = sum1+temp1

sum1

}

coeff2 = lgamma(0.5*n*3)-sum1

coeff3 = -0.25*3*2*log(pi)

coeff = coeff1 + coeff2 + coeff3

3. Making integrand

For numerator

v1=b1*b2

v2=(1-b1)*b2

v3=1-b2

num =(n/2-1)*(log(v1)+log(v2)+log(v3))+log(b2)

For denominator

Pv = solve(P,diag(1,3))

H = Pv*R # * : Hadamard Product

comp1 = Pv[1,1]*v1 + Pv[2,2]*v2 + Pv[3,3]*v3

comp2 = 2*(H[1,2]*sqrt(v1*v2)+H[1,3]*sqrt(v1*v3)+H[2,3]*sqrt(v2*v3))

```
den = (n/2)*3*log(comp1 + comp2)

logint = num - den

integ = exp(coeff)*exp(logint)

integ

}

else {

integ = 0

integ

}

integ

}


# Getting the row of the MLE by 1-5-2 rule

max.grid <- function(rho12,rho13,rho23){

n1 = length(rho12)

n2 = length(rho13)

n3 = length(rho23)

mat1 = matrix(0,nrow=n1*n2*n3, ncol=4)

for (k1 in 1:n1){

for (k2 in 1:n2){

for (k3 in 1:n3){

comb <- 1

for (i in 1:length(rn[,1])){

ind = mc3d(function(x,y) integ.mc3d(x,y,rho12[k1],rho13[k2],rho23[k3], rn[i,]),

npoints=np)

comb <- comb*ind

comb

}

mat1[(k1-1)*n2*n3+(k2-1)*n3+k3,] <-c(rho12[k1], rho13[k2], rho23[k3], comb)

}
```

```
}

}

mat1[mat1[,4]==max(mat1[,4])]

}

rho12 =seq(round(min(rn[,1]),1)-.1, round(max(rn[,1]),1)+.1, by=.1)

rho13 =seq(round(min(rn[,2]),1)-.1, round(max(rn[,2]),1)+.1, by=.1)

rho23 =seq(round(min(rn[,3]),1)-.1, round(max(rn[,3]),1)+.1, by=.1)

maxt =max.grid(rho12,rho13,rho23)

nn =mm-2

for (k in 0:nn){

rho12 =seq(maxt[1]-.1*.1^k, maxt[1]+.1*.1^k, by=.05*.1^k)

rho13 =seq(maxt[2]-.1*.1^k, maxt[2]+.1*.1^k, by=.05*.1^k)

rho23 =seq(maxt[3]-.1*.1^k, maxt[3]+.1*.1^k, by=.05*.1^k)

maxt =max.grid(rho12,rho13,rho23)

rho12 =seq(maxt[1]-.05*.1^k, maxt[1]+.05*.1^k, by=.02*.1^k)

rho13 =seq(maxt[2]-.05*.1^k, maxt[2]+.05*.1^k, by=.02*.1^k)

rho23 =seq(maxt[3]-.05*.1^k, maxt[3]+.05*.1^k, by=.02*.1^k)

maxt =max.grid(rho12,rho13,rho23)

rho12 =seq(maxt[1]-.02*.1^k, maxt[1]+.02*.1^k, by=.01*.1^k)

rho13 =seq(maxt[2]-.02*.1^k, maxt[2]+.02*.1^k, by=.01*.1^k)

rho23 =seq(maxt[3]-.02*.1^k, maxt[3]+.02*.1^k, by=.01*.1^k)

maxt =max.grid(rho12,rho13,rho23)

maxt

}

rho12 =seq(maxt[1]-.01*.1^nn, maxt[1]+.01*.1^nn, by=.01*.1^k)

rho13 =seq(maxt[2]-.01*.1^nn, maxt[2]+.01*.1^nn, by=.01*.1^k)

rho23 =seq(maxt[3]-.01*.1^nn, maxt[3]+.01*.1^nn, by=.01*.1^k)

maxt =max.grid(rho12,rho13,rho23)

maxt

}
```

```
——————————-
## Getting the MLE
rn1 <- c(.46, .31, .19, 103)
rn2 <- c(.46, .55, .32, 69)
rn3 <- c(.40, .40, .18, 69)
rn4 <- c(.27, .57, .22, 70)
rn <- rbind(rn1,rn2,rn3,rn4)
library(corpcor)
start =proc.time()[3]
mle.mc3d(rn,2,100)
proc.time()[3]-start
> mle.mc3d(rn,2,100)
[1] 0.40 0.45 0.18 2298211.37
> proc.time()[3]-start
elapsed
2.25
> mle.mc3d(rn,2,1000)
[1] 0.43 0.45 0.24 633359.08
> proc.time()[3]-start
elapsed
4.71
> mle.mc3d(rn,2,5000)
[1] 0.38 0.42 0.23 446156.21
> proc.time()[3]-start
elapsed
17.24
> mle.mc3d(rn,2,10000)
[1] 0.40 0.44 0.23 427133.47
> proc.time()[3]-start
elapsed
```

34.23

> **mle.mc3d(rn,2,100000)**

[1] **0.40 0.44 0.23 390936.23**

> proc.time()[3]-start

elapsed

449.56

B.2   FUNCTIONS FOR VISUALIZATION FROM NUMERICAL INTEGRATION

B.2.1   Getting the Likelihood Region by Monte Carlo(MC) Method

```
## Monte Carlo Method
    ## For Integration
    mc3d = function(ftn, low=c(0,0), upp=c(1,1),npoints=100) {
    x = runif(n=npoints, min=low[1], max=upp[1])
    y = runif(n=npoints, min=low[2], max=upp[2])
    ftn.tmp = ftn(x, y)
    v.tmp = diff(c(low[1], upp[1]))* diff(c(low[2], upp[2]))
    mean.approx = mean(ftn.tmp) * v.tmp
    mean.approx
    }
    ## Making Integrand in (0,1)*(0,1)
    integ.mc3d = function(b1,b2, rho12,rho13,rho23, rn) {
    ## rn: 4 by 1 vector
    n = rn[4]-1
```

```
rho = c(rho12,rho13,rho23)
## 1. Making Matrice P & R ## P: Population R: Sample
P = vec2sm(rho, diag = F)
r = rn[1:3]
R = vec2sm(r, diag = F)
P[1,1]=P[2,2]=P[3,3]=1
R[1,1]=R[2,2]=R[3,3]=1
## 2. Making Coefficient
if(det(P)>0) { ## Can change to make it efficient
coeff1 = 0.5*(n-4)*log(det(R))- 0.5*n*log(det(P))
sum1 = 0
for (i in 1:3){
temp1 = lgamma(0.5*(n+1-i))
sum1 = sum1+temp1
sum1
}
coeff2 = lgamma(0.5*n*3)-sum1
coeff3 = -0.25*3*2*log(pi)
coeff = coeff1 + coeff2 + coeff3
## 3. Making Core of Integrand
## For Numerator
v1=b1*b2
v2=(1-b1)*b2
v3=1-b2
num =(n/2-1)*(log(v1)+log(v2)+log(v3))+log(b2)
## For Denominator
Pv = solve(P,diag(1,3))
H = Pv*R                    # * : Hadamard Product
comp1 = Pv[1,1]*v1 + Pv[2,2]*v2 + Pv[3,3]*v3
comp2 = 2*(H[1,2]*sqrt(v1*v2)+H[1,3]*sqrt(v1*v3)+H[2,3]*sqrt(v2*v3))
```

```
den = (n/2)*3*log(comp1 + comp2)

logint = num - den

integ = exp(coeff)*exp(logint)

integ

}

else {

integ = 0

integ

}

integ

}

# Getting Grid Matrix (of Rho's and Likelihood)

cr.grid <- function(rho12,rho13,rho23,rn){ ####### changed

n1 = length(rho12)

n2 = length(rho13)

n3 = length(rho23)

mat1 = matrix(0,nrow=n1*n2*n3, ncol=4)

for (k1 in 1:n1){

for (k2 in 1:n2){

for (k3 in 1:n3){

comb <- 1

for (i in 1:length(rn[,1])){

ind = mc3d(function(x,y) integ.mc3d(x,y,rho12[k1],rho13[k2],rho23[k3], rn[i,]),

npoints=10000)                ## originally np

comb <- comb*ind

comb

}

mat1[(k1-1)*n2*n3+(k2-1)*n3+k3,] <-c(rho12[k1], rho13[k2], rho23[k3], comb)

}

}
```

```
}

mat1

}


## Previous Information

> mle.mc3d(rn,2,10000)

[1] 0.40 0.44 0.23 427133.47

> proc.time()[3]-start

elapsed

34.23

library(corpcor)

rn1 <- c(.46, .31, .19, 103)

rn2 <- c(.46, .55, .32, 69)

rn3 <- c(.40, .40, .18, 69)

rn4 <- c(.27, .57, .22, 70)

rn <- rbind(rn1,rn2,rn3,rn4)

## Matrix for Rho's and Likelihood (50*50*50=125,000)

rho12 =seq(0.2, 0.6, length.out=50)

rho13 =seq(0.2, 0.6, length.out=50)

rho23 =seq(0.0, 0.4, length.out=50)

start =proc.time()[3]

Sat50 =cr.grid(rho12,rho13,rho23,rn)

proc.time()[3]-start

write.table(Sat40, file="Sat50.dat")


### Opening pre-existed Matrix in R

Sat50 = as.matrix(read.table("Sat50.dat"))

## Matrix sorted

B1 = Sat50[order(Sat40[,4], decreasing = T),]

B2 = B1[B1[,4]>0,]
```

Matrix with probability

t1 = cumsum(B2[,4])

t2 = sum(B2)

t3 = t1/t2

B2[,4] = t3

30%, 50%, 70%, 95% and 99% likelihood region matrix

D30 = B2[B2[,4]<0.30,]

D50 = B2[B2[,4]<0.50,]

D70 = B2[B2[,4]<0.70,]

D95 = B2[B2[,4]<0.95,]

D99 = B2[B2[,4]<0.99,]

B.2.2   Making Plots

95% CR

library(rgl)

plot3d(D95[,1], D95[,2], D95[,3], col=rainbow(length(D95[,1])), size=2.5,

xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4),

xlab="rho12", ylab="rho13", zlab="rho23")

snapshot3d("ConfR95_sat_1.png")

snapshot3d("ConfR95_sat_2.png")

snapshot3d("ConfR95_sat_3.png")

95% CR and Projections

par(mfrow=c(2,2))

library(scatterplot3d)

scatterplot3d(D95[,1],D95[,2],D95[,3],

xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4),

```
xlab="rho12", ylab="rho13", zlab="rho23", pch=1,
angle=30,color=rainbow(length(D95[,1])))
plot(D95[,1],D95[,2],xlim=c(0.2,0.6), ylim=c(0.2,0.6), pch=20,
main="rho12 vs. rho13", xlab="", ylab="", col="red")
plot(D95[,1],D95[,3],xlim=c(0.2,0.6), ylim=c(0,0.4), pch=20,
main="rho12 vs. rho23", xlab="", ylab="", col="purple")
plot(D95[,2],D95[,3],xlim=c(0.2,0.6), ylim=c(0,0.4), pch=20,
main="rho13 vs. rho23", xlab="", ylab="", col="blue")


## 95% CR and 50%, 70%, 95% Projections
par(mfrow=c(2,2))
scatterplot3d(D95[,1],D95[,2],D95[,3],
xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4),pch=1,
xlab="rho12", ylab="rho13", zlab="rho23",
angle=50,color= rainbow(length(D95[,1])))
plot(D95[,1],D95[,2],xlim=c(0.2,0.6), ylim=c(0.2,0.6),pch=19,
main="rho12 vs. rho13", xlab="", ylab="", col="purple")
points(D70[,1],D70[,2], col="green", pch=19)
points(D30[,1],D30[,2], col="blue", pch=19)
plot(D95[,1],D95[,3],xlim=c(0.2,0.6), ylim=c(0,0.4), pch=19,
main="rho12 vs. rho23", xlab="", ylab="", col="purple")
points(D70[,1],D70[,3], col="green", pch=19)
points(D30[,1],D30[,3], col="blue", pch=19)
plot(D95[,2],D95[,3],xlim=c(0.2,0.6), ylim=c(0,0.4),pch=19,
main="rho13 vs. rho23", xlab="", ylab="", col="purple")
points(D70[,2],D70[,3], col="green", pch=19)
points(D30[,2],D30[,3], col="blue", pch=19)


## 30%, 50%, 70%, 99% CR's
par(mfrow=c(2,2))
```

```
scatterplot3d(D99[,1],D99[,2],D99[,3],
xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4),pch=1,
xlab="rho12", ylab="rho13", zlab="rho23",
angle=260,color= rainbow(length(D99[,1])))
scatterplot3d(D70[,1],D70[,2],D70[,3],
xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4), pch=1,
xlab="rho12", ylab="rho13", zlab="rho23",
angle=200,color= rainbow(length(D70[,1])))
scatterplot3d(D50[,1],D50[,2],D50[,3],
xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4), pch=1,
xlab="rho12", ylab="rho13", zlab="rho23",
angle=140,color= rainbow(length(D50[,1])))
scatterplot3d(D30[,1],D30[,2],D30[,3],
xlim=c(0.2,0.6),ylim=c(0.2,0.6),zlim=c(0,0.4), pch=1,
xlab="rho12", ylab="rho13", zlab="rho23",
angle=80,color= rainbow(length(D30[,1])))
```