Objective Bayesian Analysis of Kullback-Liebler Divergence of two Multivariate Normal Distributions with Common Covariance Matrix and Star-shape Gaussian Graphical Model

Zhonggai LI

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Dongchu Sun, Co-chairman
Eric P. Smith, Co-chairman
John P. Morgan
Pang Du

June 18, 2008
Blacksburg, Virginia

Keywords: Objective Priors; Jeffreys' Priors; Reference Priors; Invariant Haar Prior; Fisher Information Matrix; Kullback-Liebler Divergence; Frequentist Matching; Multivariate Normal Distributions; Star-shape Gaussian Graphical Model; Monte Carlo

Objective Bayesian Analysis of Kullback-Liebler Divergence of two Multivariate Normal Distributions with Common Covariance Matrix and Star-shape Gaussian Graphical Model

Zhonggai Li

(ABSTRACT)

This dissertation consists of four independent but related parts, each in a Chapter. The first part is an introductory. It serves as the background introduction and offer preparations for later parts. The second part discusses two population multivariate normal distributions with common covariance matrix. The goal for this part is to derive objective/non-informative priors for the parameterizations and use these priors to build up constructive random posteriors of the Kullback-Liebler (KL) divergence of the two multivariate normal populations, which is proportional to the distance between the two means, weighted by the common precision matrix. We use the Cholesky decomposition for re-parameterization of the precision matrix. The KL divergence is a true distance measurement for divergence between the two multivariate normal populations with common covariance matrix. Frequentist properties of the Bayesian procedure using these objective priors are studied through analytical and numerical tools. The third part considers the star-shape Gaussian graphical model, which is a special case of undirected Gaussian graphical models. It is a multivariate normal distribution where the variables are grouped into one "global" group of variable set and several "local" groups of variable set. When conditioned on the global variable set, the local variable sets are independent of each other. We adopt the Cholesky decomposition for re-parametrization of precision matrix and derive Jeffreys' prior, reference prior, and invariant priors for new parameterizations. The frequentist properties of the Bayesian procedure using these objective priors are also studied. The last part concentrates on the discussion of objective Bayesian analysis for partial correlation coefficient and its application to multivariate Gaussian models.

Acknowledgments

At the completion of this dissertation, I would like to extend my biggest appreciation to my chief dissertation advisor Prof. Dongchu Sun for three main reasons. First, it is Dr. Sun that made it possible for me to continue my doctoral study at Virginia Tech. Second, it is Dr. Sun that leaded me into the area of Bayesian analysis, specially objective Bayesian analysis. Third, Dr. Sun's consistent support is very helpful in my job search process. Dr. Sun helped to change my life over the past three years.

I would like to thank my committee members: Professors. Eric P. Smith, John P. Morgan, and Pang Du for their intelligent comments and advice concerning my research. Michael C. Box helped me connect my laptop to the Linux server remotely for my simulation. Professors Eric P Smith and Jeffery B. Birch have been very supportive in making my graduate assistant assignments and over my job search. Professor William H. Woodall offered very good advice on my English writings. Dr. Pang Du, Inyoung Kim, and Dong-Yun Kim offered me a lot of information and advice related to job search.

I would like to thank all professors, staff, and graduate students in the Department of Statistics at Virginia Tech. They have made my study at Virginia Tech very rewarding.

My special thanks goes to my beloved wife Huihua Huang for taking over all household duties, which makes it possible for me to concentrate on my research and job search and for encouraging me when I was depressed occasionally.

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Background and Motivation of this Study

Statistics as a foundation for scientific reasoning is based on data information. It serves its role in two fundamental forms, statistical estimation and statistical hypothesis testing. As Cox (2005) stated, frequentist and Bayesian statistics are the two broad approaches taken to formal estimation and hypothesis testing, both with variants. Efron (2005) labeled the $19th$ century as Bayesian dominated, the $20th$ century as generally frequentist, and suggested that statistics in the $21st$ century will require a combination of Bayesian and frequentist ideas. Both frequentist and Bayesian statistics have their own strengths and weaknesses, for more details see Cox (2005), Little (2006), and Samaniego & Reneau (1994).

The debate between the Bayesian and frequentist perspective has been an ongoing issue for professional statisticians. For consumers of statistical analysis reports, the most confusing part of this debate is that most statistical problems can be addressed with either the frequentist or the Bayesian approach, and they frequently do not agree with each other as pointed out by Efron (2005). A lot of effort has been devoted to resolving this fundamental difficulties. Some examples include Berger et al (1997,1999,2003) and Casella & Berger (1987), who proposed statistical methods that unify Bayesian and frequentist results for some hypothesis testing problems; Good (1992) suggested a compromise between frequentist and Bayesian testing; Little (2006) proposed the calibrated Bayes approach that claims to take strength from both camps; and Samaniego & Reneau (1994) suggested a reconciliation of the Bayesian and frequentist approach for point estimation. More recently, Berger & Sun (2008), Sun & Berger (2007), Ghosh & Mukerjee (1993), and Ghosh & Mukerjee (1998) proposed matching priors to unify frequentist and Bayesian analysis.

Efron (2005) pointed out that classical statistics have mainly evolved in response to small problems, a few hundred data points at most and just a few parameters. The development of other scientific research has resulted in more statistical problems that are high-dimensional in the parameter space. Some involve huge data sets and some have only a small number of observations. The use of Bayesian analysis is becoming productive and popular in many of these situations. But soliciting the appropriate priors and avoiding the debate about subjectiveness of priors is still a challenge.

The frequentist/Bayesian debate and the clear potential of Bayesian analysis to handle difficult problems together form the main source of motivation for this thesis. In this study, we further research in Bayesian analysis for multivariate Gaussian models by deriving non-informative priors, and we compare the Bayesian outputs from these non-informative priors with frequentist counterparts. This approach has three main advantages: (a) It avoids the necessity of soliciting subjective and/or empirical prior information to build up a prior for Bayesian analysis, which could be difficult in many cases; (b) It offers a way of unifying Bayesian and frequentist analysis for some specific statistical problems; (c) It is capable of accommodating the strengths of Bayesian statistics for inference under an assumed model and the strengths of frequentist methods for model assessment (See Little (2006)).

The rest of this chapter offers a brief introduction to the area of of study, including commonly used terminologies and techniques. This is not meant to be a comprehensive or formal review. Rather, this provides basic information that will facilitate reading of later chapters.

1.2 Bayesian Analysis: Subjective, Empirical and Objective

Probability distributions for observable data are the core of Statistical Science. These probability distributions usually depend on unknown parameters, $\boldsymbol{\theta}$. In Bayesian analysis, probability is interpreted as "degree of belief" instead of the "limiting frequency" view of frequentist statistics. In Bayesian analysis, current knowledge about the model parameters, $\boldsymbol{\theta}$, is expressed by assuming a probability distribution on the parameters, called the "prior distribution", often written as

$$P(\boldsymbol{\theta}).$$

For an observed data set $\boldsymbol{y}$, the information about $\boldsymbol{\theta}$ is expressed through the likelihood function

$$\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}) = P(\boldsymbol{y} \mid \boldsymbol{\theta}).$$

Bayes' rule is then used to combine the prior information about model parameters with the likelihood of the observed data to get the "posterior distribution" of $\boldsymbol{\theta}$, which is mathematically expressed as

$$P(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{P(\theta) \times \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{\theta})}{\int P(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

Any inference about the model parameters, $\boldsymbol{\theta}$, is then based on the posterior distribution of $\boldsymbol{\theta}$. Frequentist statistics bases inference on the likelihood $\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y})$ alone.

Usually it is not necessary to evaluate the integral $\int P(\boldsymbol{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$. Instead, one can write

$$P(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto P(\theta) \times \mathcal{L}(\boldsymbol{y} \mid \boldsymbol{\theta})$$

and simply find the normalization constant to make the function become a probability density at the final calculation.

The above steps summarize the basic ideas of Bayesian analysis. Depending on how the prior distribution is selected, Bayesian analysis is further classified into subjective, empirical, and objective Bayesian analysis. When a prior is postulated, its distribution may have one or more unknown parameters, called hyperparameters and denoted by η, in which case the prior is written as $p(\boldsymbol{\theta} \mid \eta)$. If the hyperparameters η or $P(\boldsymbol{\theta})$ is specified by the Bayesian practitioner based on his/her (or even other's) personal "beliefs" about the distribution of $\boldsymbol{\theta}$, then this prior is a "subjective" prior and the analysis is called subjective Bayesian analysis. If the hyperparameters η or the prior $P(\boldsymbol{\theta})$ is determined through estimation from historical data (See Casella (1985)), then the prior is an "empirical" prior and the analysis is called empirical Bayesian analysis. Finally, if the prior $P(\boldsymbol{\theta})$ is specified through some "non-informative" prior generating algorithm that is free of the use of personal belief or historical data, then it is a non-informative prior, also called "objective prior", and the analysis is called objective Bayesian analysis. An objective prior typically still has a noticeable effect on the posterior distribution of some or all elements of $\boldsymbol{\theta}$, and thus the impact of non-informative priors should be carefully evaluated.

In this thesis, we only work with objective Bayesian analysis. Thus, the key to the study is deriving non-informative priors.

1.3 Introduction to Non-informative Priors

As stated in the previous section, after a probability model is specified for a statistical problem, Bayesian analysis begin with formalizing a prior distribution for the model parameters, $\boldsymbol{\theta}$. There are several prior-generating methods that are free of personal/subjective information or historical data. The priors derived through these methods are called non-informative priors. A prior distribution is preferably proper (that is, integrable), as is commonly seen for subjective priors and empirical priors. However, non-informative priors often are improper (not integrable). Fortunately, an improper prior does not necessary result in an improper posterior. Indeed, the Jeffreys's prior defined below actually yields a proper posterior in most cases. The reference prior is even better than Jeffreys' prior for multi-parameter cases. If use of an improper prior leads to a proper posterior, Bayesian analysis based on the posterior can be pursued, as will be seen from our study. Some procedures, such as hypothesis testing, may not allow improper priors though.

This subsection introduces three popular algorithms for generating non-informative priors: Jeffreys' prior/rule, reference prior, and Left- and right-Haar measure/prior. A thorough discussion on development and definition of various formal rules for non-informative priors can be found in Kass & Wasserman (1996).

The development of non-informative priors hales back to the uniform prior on the parameter p in the binomial distribution as proposed by Bayes (1763). This was popularized by Laplace (1982) as a constant density prior, with constant typically being chosen to be 1. The idea behind a constant prior is the "principle of insufficient reason," which requires a distribution on the finitely many events to be uniform unless there is some definite reason to consider one event more probable than another.

The problem with the constant prior is that it is not invariant with respect to re-parameterization. For example, if $\theta = \exp(\psi)$, the constant prior $\pi(\psi) \propto 1$ becomes the prior $\pi(\theta) \propto \frac{1}{\theta}$ which is no longer constant.

1.3.1 Jeffreys' Prior

To address the non-invariant problem of constant priors, Jeffreys (1961) proposed the Jeffreys' rule/prior. Jeffreys prior is derived by taking the square root of the determinant of Fisher information matrix $I(\boldsymbol{\theta})$, that is

$$\pi(\boldsymbol{\theta}) \propto \sqrt{det(I(\boldsymbol{\theta}))}. \tag{1}$$

This is the formal Jeffreys' prior, which is invariant to re-parameterization and is widely used.

In the one parameter case, Jeffreys' rule works very well. But for multi-parameter situations, inappropriate aspects of priors may accumulate across dimensions to detrimental effect. Examples of this show up in this study. Modifications have been proposed, such as the independent Jeffreys' rule, which partitions parameters into say two groups $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, and assumes they are independent to derive the Fisher information matrix and Jeffreys' prior. Jeffreys himself also recommended non-informative priors that differed from the formal Jeffreys' prior.

1.3.2 Reference Prior

As pointed out, the Jeffreys' prior has limitations that can be hard to overcome in multi-parameter cases. Bernardo (1979) proposed the idea of reference prior, which Berger & Bernardo (1992a) modified for multi-parameter problems. It is not possible to simply describe details of this algorithm; any person interested is strongly recommended to read the original paper by Berger & Bernardo (1992a, 1992b, 1992c) and Datta & Ghosh (1996). A copy of the algorithm is attached in the Appendix as well. However, the general idea is summarized below. The reference prior tries to modify Jeffreys' prior by reducing the dependence among parameters that is frequently induced by Jeffreys' rule. This is accomplished via two innovations: a notion of missing information and a stepwise procedure for handling nuisance parameters.

Let $\boldsymbol{Y}$ be a random vector with probability distribution $P(\boldsymbol{Y} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown model parameter vector. Let $\boldsymbol{z}_t = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_t)$ be t independent samples of $\boldsymbol{Y}$. The Kullback-Leibler (KL) distance is used to measure the difference between the posterior density and the prior density, as follows:

$$K_t(\pi(\boldsymbol{\theta} \mid \boldsymbol{z}_t), \pi(\boldsymbol{\theta})) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{z}_t) log \left(\frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{z}_t)}{\pi(\boldsymbol{\theta})} \right) d\boldsymbol{\theta}. \quad (2)$$

This is roughly the gain in information by the experiment. Let $K_t^\pi = E(K_t(\pi(\boldsymbol{\theta} \mid \boldsymbol{z}_t), \pi(\boldsymbol{\theta})))$, where the expectation is taken with respect to the marginal density of $\boldsymbol{z}_t = \int P(\boldsymbol{z}_t \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. The prior that maximizes K_t^π as $t \to \infty$ is "the" reference prior. The technical problem is that K_∞^π is usually infinite. The ordered-group reference prior algorithm given in Berger & Bernardo (1992a) is an approach to overcoming this problem. It employs asymptotic normality results satisfied under "regular" cases, where the support of data does not depend on the parameter.

For deriving a reference prior, one needs to partition the parameters $\boldsymbol{\theta}$ into m sub-groups such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m)$, each group has n_i parameters. The order of the sub-parameter groups represent the relative inferential importance of the groups, with the most interested one as the first group, $\boldsymbol{\theta}_1$. But Berger & Bernardo (1992a) pointed out, if there is no specific reason, treating every parameter as from an individual group is recommended. The priors derived are all one-at-a-time reference priors. There are $p!$ permutations for the p-dimensional case, so the possible number of distinct one-at-a-time reference priors can be very big. In some cases, all one-at-a-time reference priors are the same; one in our study has this property. Re-parameterization that provides better orthogonality among parameters and makes the Fisher information matrix closer to block-diagonal can be a very good way to reduce the complexity in deriving reference priors.

1.3.3 Left-Haar and Right-Haar Prior

This subsection summarizes the definitions and algorithms for deriving left- and right-Haar measures. The Haar prior we use in later chapters is also derived here. The main reference for this subsection is "Group invariance applications in statistics" by Eaton (1989).

Definition 1.1 *A **group** is a non-empty set G together with a binary operation $\circ$ such that the following conditions hold:*

(i) $g_1,\ g_2 \in G$ *implies* $g_1 \circ g_2 \in G$*;*

(ii) $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ *for* $g_1,\ g_2,\ g_3 \in G$*;*

(iii) *There exists a unique element* $e \in G$ *such that* $e \circ g = g \circ e = g$ *for* $g \in G$*;*

(iv) *For each $g \in G$, there exists a unique element $g^{-1} \in G$ such that $g \circ g^{-1} = g^{-1} \circ g = e$.*

The binary operation $\circ$ is usually suppressed. Here are the four group examples that we work on in later chapters.

Example 1.1 $G^a = \{g_u : \boldsymbol{u} \in \mathbb{R}^p\}$ *has elements* $g_u : \mathbb{R}^p \to \mathbb{R}^p$ *defined by* $g_u(\boldsymbol{x}) = \boldsymbol{u} + \boldsymbol{x},\ \boldsymbol{x} \in \mathbb{R}^p$.

Example 1.2 $G^b = \{g_\psi : \boldsymbol{\psi} \in G_T^+,$ *p×p lower triangle matrix with postive diagonal elements*$\}$ *has elements* $g_\psi :\ G_T^+ \to G_T^+$ *defined by* $g_\psi(\boldsymbol{V}) = \boldsymbol{\Psi V},\ \boldsymbol{V} \in G_T^+$.

Define a positive definite block diagonal lower triangle matrix as below.

$$\boldsymbol{G}^{\#} = \begin{pmatrix} \boldsymbol{G}_0 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{G}_{10} & \boldsymbol{G}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{G}_{20} & \boldsymbol{0} & \boldsymbol{G}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{G}_{k0} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{G}_k \end{pmatrix}, \tag{3}$$

where $\boldsymbol{G}_i$ is $p_i \times p_i$ lower triangle matrix with positive diagonal elements, and $\boldsymbol{G}_{i0}$ is a $p_i \times p_0$ matrix with real value elements.

Example 1.3 $G^c = \{g_\psi : \boldsymbol{\Psi} \in G_T^{\#},\ p \times p\}$ *has elements* $g_\psi :\ G_T^{\#} \to G_T^{\#}$ *defined by* $g_\psi(\boldsymbol{V}) = \boldsymbol{\Psi V},\ \boldsymbol{V} \in G_T^{\#}$.

Example 1.4 $G^d = G^a \times G^b$ *has elements* $g_{u,\psi} :\ \mathbb{R} \times G_T^+ \to \mathbb{R} \times G_T^+$ *defined by* $g_{u,\psi}(\boldsymbol{x}, \boldsymbol{V}) = (\boldsymbol{u} + \boldsymbol{x}, \boldsymbol{\Psi V}), \boldsymbol{x} \times \boldsymbol{V} \in \mathbb{R} \times G_T^+$.

Let G be a group and let the real vector space $K(G)$ be the set of all continuous function with compact support defined on G. Given $g \in G$, define the transformation L_g on $K(G)$ to $K(G)$ by

$$(L_g f)(x) = f(g^{-1}x), \tag{4}$$

for $g,\ x \in G$, and $f \in K(G)$.

Definition 1.2 *An integral J on $K(G)$ is a left-invariant integral if for all $f \in K(G)$,*

$$J(L_g f) = J(f) \quad for \quad g \in G. \tag{5}$$

If v_l is the measure corresponding to the left-invariant integral, then the measure v_l satisfies

$$\int_G f(g^{-1}x)v_l(dx) = \int f(x)v_l(dx) \quad g \in G. \tag{6}$$

And $v_l(dx)$ is called a left-Haar measure, which is called a left-Haar prior in statistical applications.

There is an important theorem on the existence and uniqueness (up to a positive constant) of left-invariant integrals.

Theorem 1.1 *On a group G, there exists a left-invariant integral (measure). If J_1 and J_2 are left-invariant integrals, then there exists a positive constant c such that $J_1 = cJ_2$.*

Now, let v_l be a left-Haar measure, for a fixed $g \in G$, define J_1 on $K(G)$ by

$$J_1(f) = \int f(xg^{-1})v_l(dx), \tag{7}$$

which is also left-invariant by definition. Therefore by Theorem 1.1, there is a positive constant which is denoted by $\Delta(g)$ such that $J_1 = \Delta(g)J$. $\Delta(g)$ is called the modulus of g. Then the integral J_1 defined by

$$J_1(f) = \int f(x)\Delta(x)v_l(dx) \tag{8}$$

is right-invariant and satisfies

$$\int f(x)\Delta(x^{-1})v_l(dx) = \int f(x^{-1})v_l(dx) \quad f \in K(G). \tag{9}$$

Thus relationship between left and right Haar measure is

$$v_r(dx) = \delta(x^{-1})v_l(dx) = \frac{1}{\delta(x)}v_l(dx), \tag{10}$$

where $v_r(dx)$ is the right-Haar measure.

Now, we are able to derive the left-Haar and right-Haar measures for the four Group examples given earlier in this section.

Example 1.1: The ordinary Lebesgue measure on $\mathbb{R}^p$ is both the left- and right-invariant Haar measure and the modulus is 1.

Example 1.2: The left-Haar and right-Haar measures are given by the following proposition.

Lemma 1.1 *Under the group G^b, for $\mathbf{\Psi} \in G_T^+$, the left- and right-Haar measures are*

$$\begin{aligned} v_l(\mathbf{\Psi}) &= \frac{d\mathbf{\Psi}}{\chi_(\mathbf{\Psi})} = \frac{d\mathbf{\Psi}}{\prod_i^p \psi_{ii}^i}, & (11) \\ v_r(\mathbf{\Psi}) &= \frac{d\mathbf{\Psi}}{\chi_1(\mathbf{\Psi})} = \frac{d\mathbf{\Psi}}{\prod_i^p \psi_{ii}^{n-(i-1)}}. & (12) \end{aligned}$$

Proof. For $\mathbf{\Psi} \in G_T^+$, consider

$$J(L_\psi f) = \int f(\mathbf{\Psi}^{-1}\boldsymbol{V})d\boldsymbol{V} \tag{13}$$

with $\boldsymbol{W} = \mathbf{\Psi}^{-1}\boldsymbol{V}$, so $\boldsymbol{V} = \mathbf{\Psi}\boldsymbol{W}$, the Jacobian of this transformation on $p(p+1)/2$ coordinate space is

$$\begin{aligned} \chi_0(\mathbf{\Psi}) &= \prod_i^p \psi_{ii}^i, \\ d\boldsymbol{V} &= \chi_0(\mathbf{\Psi})d\boldsymbol{W}, & (14) \\ J(L_g f) &= \chi_0(\mathbf{\Psi})J(f), & (15) \\ J(f) &= \frac{1}{\chi_0(\mathbf{\Psi})}J(L_g f), & (16) \end{aligned}$$

and,

$$v_l(d\boldsymbol{V}) = \frac{d\boldsymbol{V}}{\chi_0(\boldsymbol{V})} = \frac{d\boldsymbol{V}}{\prod_i^p \psi_{ii}^i} \tag{17}$$

Let $\boldsymbol{W}_r\mathbf{\Psi} = \boldsymbol{V}$ and the Jacobian from $\boldsymbol{V}$ into $\boldsymbol{W}_r$ is

$$\chi_1(\mathbf{\Psi}) = \prod_i^p \psi_{ii}^{n-(i-1)}. \tag{18}$$

Then

$$\begin{aligned}
\int f(\boldsymbol{V}\boldsymbol{\Psi}^{-1})v_l(d\boldsymbol{V}) &= \int f(\boldsymbol{W}_r)\frac{d\boldsymbol{V}}{\chi_0(\boldsymbol{V})} \\
&= \int f(\boldsymbol{W}_r)\frac{d(\boldsymbol{W}_r\boldsymbol{\Psi}^{-1})}{\chi_0(\boldsymbol{W}_r\boldsymbol{\Psi}^{-1})} \\
&= \frac{\chi_1(\boldsymbol{\Psi})}{\chi_0(\boldsymbol{\Psi})}\int f(\boldsymbol{W}_r)v_l(d\boldsymbol{W}_r), \\
\Delta(\boldsymbol{\Psi}) &= \frac{\chi_1(\boldsymbol{\Psi})}{\chi_0(\boldsymbol{\Psi})} \qquad (19)
\end{aligned}$$

The right-Haar measure for $\boldsymbol{V}$ is then given by

$$\begin{aligned}
v_r(\boldsymbol{V}) &= \Delta(\boldsymbol{V}^{-1})v_l(d\boldsymbol{V}) \\
v_r(\boldsymbol{V}) &= \frac{\chi_0(\boldsymbol{V})}{\chi_1(\boldsymbol{V})}\frac{d\boldsymbol{V}}{\chi_0(\boldsymbol{V})} \\
&= \frac{d\boldsymbol{V}}{\chi_1(\boldsymbol{V})}. \qquad (20)
\end{aligned}$$

Replacing the notation $\boldsymbol{V}$ by $\boldsymbol{\Psi}$, the Lemma is proved. $\square$ Example 1.3: Since the structure of Example 1.3 is similar to Example 1.2, the steps for deriving left- and right-Haar measures for $\boldsymbol{\Psi} \in G_T^{\#}$ are the same. An interesting finding that simplifies calculation for Haar measures is summarized in the following corollary.

Corollary 1.1 *For $\boldsymbol{\Psi} \in G_T^{\#}$, the left- and right-Haar measure are*

$$\begin{aligned}
v_l(\boldsymbol{\Psi}) &= \frac{d\boldsymbol{\Psi}}{\prod_i^p \psi_{ijj}^{a_{ij}}}, \qquad (21) \\
v_r(\boldsymbol{\Psi}) &= \frac{d\boldsymbol{\Psi}}{\chi_1(\boldsymbol{\Psi})} = \frac{d\boldsymbol{\Psi}}{\prod_i^p \psi_{ijj}^{b_{ij}}}, \qquad (22)
\end{aligned}$$

*where a_{ij} is the total number of non-zero elements of $\boldsymbol{\Psi}$ in the same **row** as ψ_{ijj} and b_{ij} is the total number of non-zero elements of $\boldsymbol{\Psi}$ in the same **column** as ψ_{ijj}.*

Example 1.4: This example is just the product of Example 1.1 and Example 1.2. The left- and right-Haar measures of this example are the product of the respective measures for Example 1.1 and Example 1.2.

Generally, right-Haar measure is preferred in practice. Our study offers further evidence for this statement.

1.4 Probability Matching Priors

A good reference on probability matching priors is Datta & Mukerjee (2004). One way to judge a "non-informative" prior is based on the idea that it will "let the data speak for itself." Frequentist statistics does not use prior information, which supports the notion that the information derived from frequentist analysis is only from the data observed. Therefore, one way to evaluate a "non-informative" prior is to compare the posterior result from the prior to the appropriate frequentist outcome. If the prior is non-informative, then the difference should be small.

As is customary in such comparisons, let $\tau(\boldsymbol{\theta})$ be a scale parameter or function of parameters, $\boldsymbol{\theta}$. The one-sided credible interval $(\tau_L, \tau_\alpha(x))$ for $\tau(\boldsymbol{\theta})$ is studied. Here τ_L is the lower bound of $\tau(\boldsymbol{\theta})$, and $\tau_\alpha(x)$ is the posterior quartile of $\tau(\boldsymbol{\theta})$, defined by

$$P(\tau(\boldsymbol{\theta}) < \tau_\alpha(x)|x) = \alpha. \tag{23}$$

The interesting quantity is the frequentist coverage achieved by the corresponding confidence interval

$$P(\tau(\boldsymbol{\theta}) < \tau_\alpha(x)|\boldsymbol{\theta}). \tag{24}$$

The closer this coverage to the nominal α, the more non-informative of the priors is judged to be. If the frequentist coverage (24) is exactly α, the prior used is called an **exactly frequentist matching prior.** There are other types of matching priors introduced in Datta & Mukerjee (2004).

1.5 Additional Terminology

After specifying a prior, the posterior density for the unknown parameter is proportional to the product of the prior and the likelihood function. If this product is integrable (even if is impossible to analytically integrate the function), the posterior is then called "proper", otherwise, it is called "improper". For a proper posterior, one may be interested in some marginal posterior densities or conditional densities of some parameters. In some cases, these posteriors for the recommended priors are essentially available in computational "closed form", allowing direct Monte Carlo simulation. These computational-ready representations of the posteriors are called "constructive posterior distributions." For example, a posterior of parameter ψ is written as $\psi = \sqrt{\chi_n^{2*}/s}$, where χ_n^{2*} is a standard chi-squared distribution with n degree of freedom and s is a statistic calculated from the observed data sample.

Chapter 2 KL Divergence of two Normal Distributions with Common Covariance Matrix

2.6 Introduction

One approach to compare two populations in Statistics is to apply some "distance" measure between the two density functions underlying the two populations. Kullback-Liebler (KL)(1951) Distance/Divergence is the most commonly used distance measurement in such comparisons. In this study, we discuss the comparison of two multivariate normal distributions with common but unknown covariance matrix using the KL divergence. The KL divergence for this model is equivalent to the divergence between the means, weighted by a common precision matrix, that is, it is also the Mahalanobis distance. It satisfies the three conditions to be a true "distance" metric. Recently, Sun & Berger (2007) considered objective inference for the parameters of the multivariate normal model with special focus on development of objective confidence or credible sets and evaluation of their frequentist matching properties for specific parameters. In this study, we follow Sun & Berger (2007) in decomposing the precision matrix into the product of a lower triangle matrix and its transpose through a type of Cholesky decomposition. Then, we further separate this lower triangle matrix into the product of a diagonal matrix and a unit lower triangle matrix. Jeffreys' priors, reference priors, and invariant priors for the new parameterizations are then derived. We find that all these priors belong to the same class of priors. Therefore, we derive the posterior distribution of KL divergence of the two multivariate normal distributions under this class of priors. Specially, we write the posteriors of KL divergence in constructive random posterior form, which shows how the posterior is constructed. It offers great convenience for direct Monte Carlo simulation, and is a powerful tool for verifying the frequentist matching property. The frequentist matching properties of the constructive random posterior of KL divergence are then studied through both analytical and numerical tools.

This chapter is arranged as follows. Section 2.7 introduces the model and its parameterizations. Section 2.8 derives various objective priors. Secion 2.9 derives the constructive random posteriors of KL divergence. Section 2.10 discusses the frequentist matching properties of KL divergence and its related components.

2.7 Model and Parameterizations

2.7.1 The KL divergence

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two observable p-dimensional random vectors from multivariate normal density functions $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, respectively. Denote $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1})$ and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2})$ as the observed random samples that are assumed to be independent and identically distributed (i.i.d.) as $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, respectively.
Define

$$\overline{\boldsymbol{x}} = \frac{1}{n_1}\sum_{i=1}^{n_1} \boldsymbol{x}_i, \tag{25}$$

$$\overline{\boldsymbol{y}} = \frac{1}{n_2}\sum_{j=1}^{n_2} \boldsymbol{y}_j, \tag{26}$$

$$\boldsymbol{S} = \sum_{i=1}^{n_1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})' + \sum_{j=1}^{n_2} (\boldsymbol{y}_j - \overline{\boldsymbol{y}})(\boldsymbol{y}_j - \overline{\boldsymbol{y}})'. \tag{27}$$

$(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}, \boldsymbol{S})$ are well known to be the sufficient statistics for the model parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, which is also given by Fact 2.5. The KL divergence between the two populations $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ is

$$\begin{aligned} K &= \int log\left(\frac{p_1(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma})}{p_2(\boldsymbol{x}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma})}\right) p_1(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) d\boldsymbol{x} \\ &= \int log\left(\frac{(2\pi)^{-\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\}}{(2\pi)^{-\frac{p}{2}}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\}}\right) p_1(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) d\boldsymbol{x} \\ &= \int \left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\right) p_1(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) d\boldsymbol{x}. \end{aligned} \tag{28}$$

After some algebra, the KL divergence is further reduced as

$$\begin{aligned} K &= \frac{1}{2}\int \left(\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) - \boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\right) p_1(\boldsymbol{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) d\boldsymbol{x} \\ &= \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2). \end{aligned} \tag{29}$$

In this case, the KL divergence is equivalent to the weighted difference between the means, thus it is also the Mahalanobis distance. It has three good properties that make it a true distance metric.

Fact 2.1 *The KL divergence of the two multivariate normal distributions with common covariance matrix satisfies the three conditions of a distance metric, that is,*

(a) *non-negativeness,*

(b) *symmetry,*

(c) *triangle inequality.*

Proof. The proof for part (a) and (b) of Fact 2.1 is straightforward from the KL expression given by (28), which is a quadratic form, and the fact that the precision matrix $\boldsymbol{\Sigma}^{-1}$ is a positive definite symmetric matrix.

To prove part (c), the triangle inequality, assume that there are three distinct populations $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$, $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, and $N_p(\boldsymbol{\mu}_3, \boldsymbol{\Sigma})$. Use a type of Cholesky decomposition for the covariance matrix $\boldsymbol{\Sigma}$ such that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi}'\boldsymbol{\Psi}$. Let vector $\boldsymbol{v}_i = \boldsymbol{\Psi}\boldsymbol{\mu}_i$, $i = 1, 2, 3$. Then the KL divergence between $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ and $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$ is $K_{ij} = \frac{1}{2}(\boldsymbol{v}_i - \boldsymbol{v}_j)'(\boldsymbol{v}_i - \boldsymbol{v}_j)$, $i, j = 1, 2, 3$. Define the vectors $\boldsymbol{z}_1 = \boldsymbol{v}_1 - \boldsymbol{v}_2$, $\boldsymbol{z}_2 = \boldsymbol{v}_1 - \boldsymbol{v}_3$, and $\boldsymbol{z}_3 = \boldsymbol{v}_3 - \boldsymbol{v}_2 = \boldsymbol{z}_1 - \boldsymbol{z}_2$. Then $K_{12} = \frac{1}{2}\|\boldsymbol{z}_1\|^2$, $K_{13} = \frac{1}{2}\|\boldsymbol{z}_2\|^2$, and $K_{23} = \frac{1}{2}\|\boldsymbol{z}_3\|^2$.
By Cauchy-Schwarz inequality,

$$\begin{aligned} \|\boldsymbol{z}_1 - \boldsymbol{z}_2\|^2 &= \|\boldsymbol{z}_1\|^2 + \|\boldsymbol{z}_2\|^2 - 2\boldsymbol{z}_1'\boldsymbol{z}_2 \\ &< (\|\boldsymbol{z}_1\| + \|\boldsymbol{z}_2\|)^2, \end{aligned}$$

that is ,

$$\|\boldsymbol{z}_3\| < \|\boldsymbol{z}_1\| + \|\boldsymbol{z}_2\| \text{ or } \|\boldsymbol{z}_3\| - \|\boldsymbol{z}_1\| < \|\boldsymbol{z}_2\|.$$

The role of $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, and $\boldsymbol{z}_3$ are interchangeable, thus the triangle inequality is proved. □

2.7.2 A Cholesky Decomposition

The original parameterization of the two normal distributions includes the common p dimensional covariance matrix $\boldsymbol{\Sigma}$, which is very inconvenient for manipulations such as taking an inverse. From Sun & Sun (2005), Berger & Sun (2008), and Sun & Berger (2007), it is most convenient to express the covariance matrix in terms of a lower-triangular matrix $\boldsymbol{\Psi}$ with positive diagonal elements through a type of Cholesky decomposition, such that

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi}'\boldsymbol{\Psi}. \tag{30}$$

Consequently we have

$$K = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Psi}'\boldsymbol{\Psi}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \tag{31}$$

Define $etr() = exp(tr())$, the likelihood function under the original parameterizations $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ is

$$\begin{aligned} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}|\boldsymbol{X}, \boldsymbol{Y}) \quad &= ((2\pi)^p|\boldsymbol{\Sigma}|)^{-\frac{n_1+n_2}{2}} etr\{-\frac{n_1}{2}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}}) \\ &\quad -\frac{n_2}{2}(\boldsymbol{\mu}_2 - \overline{\boldsymbol{y}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \overline{\boldsymbol{y}})\} etr\{-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\}. \end{aligned} \tag{32}$$

Under the parameterization $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Psi})$, the likelihood function is

$$\begin{aligned} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Psi}|\boldsymbol{X}, \boldsymbol{Y}) \quad &= ((2\pi)^{-\frac{p}{2}}|\boldsymbol{\Psi}|)^{n_1+n_2} etr\{-\frac{n_1}{2}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}})'\boldsymbol{\Psi}'\boldsymbol{\Psi}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}}) \\ &\quad -\frac{n_2}{2}(\boldsymbol{\mu}_2 - \overline{\boldsymbol{y}})'\boldsymbol{\Psi}'\boldsymbol{\Psi}(\boldsymbol{\mu}_2 - \overline{\boldsymbol{y}})\} etr\{-\frac{1}{2}\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S}\}. \end{aligned} \tag{33}$$

2.7.3 Alternative Parameterization

We further separate $\boldsymbol{\Psi}$ into the product of a diagonal matrix $\boldsymbol{\Xi}$ with positive diagonal elements and a unit lower triangle matrix $\boldsymbol{\Delta}$ as below,

$$\begin{aligned} \boldsymbol{\Psi} &= \begin{pmatrix} \psi_{11} & 0 & \cdots & 0 \\ \psi_{21} & \psi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{p1} & \psi_{p2} & \cdots & \psi_{pp} \end{pmatrix} = \begin{pmatrix} \psi_{11} & 0 & \cdots & 0 \\ \psi_{22}\delta_{21} & \psi_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{pp}\delta_{p1} & \psi_{pp}\delta_{p2} & \cdots & \psi_{pp} \end{pmatrix} \\ &= \boldsymbol{\Xi}\boldsymbol{\Delta}, \end{aligned}$$

that is,

$$\boldsymbol{\Xi} = diag(\psi_{11}, \psi_{22}, \ldots, \psi_{pp}); \tag{34}$$

$$\psi_{hl} = \psi_{hh}\delta_{hl}, \quad h = 2, \ldots, p, \quad l = 1, \ldots, h-1. \tag{35}$$

By Pourahmadi (1999), the h^{th} row below-diagonal entries of $\boldsymbol{\Delta}$ are the negatives of the linear least-squares predictor of x_h based on its predecessors $x_1, \ldots, x_{h-1}$ and the ψ_{hh}^{-2} are the prediction error variance $var(x_h - \hat{x}_h)$.

Define $\boldsymbol{\xi} = \boldsymbol{\Psi}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, then $K = \frac{1}{2}\boldsymbol{\xi}'\boldsymbol{\xi} = \frac{1}{2}\|\boldsymbol{\xi}\|$.

Under parameterizations $(\Xi, \boldsymbol{\Delta})$, we have

$$\begin{aligned} K &= \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), && (36) \\ \boldsymbol{\xi} &= \Xi\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). && (37) \end{aligned}$$

The likelihood function under this alternative parameterization is

$$\begin{aligned} L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}|\boldsymbol{X}, \boldsymbol{Y}) &= \left((2\pi)^{-\frac{p}{2}}|\Xi|\right)^{n_1+n_2} etr\{-\frac{n_1}{2}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}})'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}}) \\ &\quad -\frac{n_2}{2}(\boldsymbol{\mu}_2 - \overline{\boldsymbol{y}})'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{\mu}_2 - \overline{\boldsymbol{y}})\} etr\{-\frac{1}{2}\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\}. \qquad (38) \end{aligned}$$

For later use, we also introduce vectors

$$\begin{aligned} \boldsymbol{\psi} &= (\psi_{11}, \psi_{21}, \psi_{22}, \ldots, \psi_{p1}, \ldots, \psi_{pp})', && (39) \\ \boldsymbol{\tau} &= (\psi_{11}, \delta_{21}, \psi_{22}, \ldots, \delta_{p1}, \ldots, \delta_{p(p-1)}, \psi_{pp})'. && (40) \end{aligned}$$

The one-to-one transformation from vector $\boldsymbol{\psi}$ into vector $\boldsymbol{\tau}$ has the Jacobian matrix

$$J(\boldsymbol{\psi}, \boldsymbol{\tau}) = \frac{\partial(\boldsymbol{\psi})}{\partial(\boldsymbol{\tau})} = diag(\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_p), \qquad (41)$$

where, $\boldsymbol{\Lambda}_1 = 1$ and for $h = 2, \ldots, p$,

$$\boldsymbol{\Lambda}_h = \begin{pmatrix} \psi_{hh}\boldsymbol{I}_{h-1} & \boldsymbol{\delta}_{h,h-1} \\ \boldsymbol{0} & 1 \end{pmatrix} \text{ and } \boldsymbol{\delta}_{h,h-1} = \begin{pmatrix} \delta_{h1} \\ \vdots \\ \delta_{h(h-1)} \end{pmatrix}. \qquad (42)$$

2.8 Objective Priors

2.8.1 The Fisher Information Matrix

The core of objective Bayesian analysis is the derivation of objective/non-informative priors. In this section, we will derive the Jeffreys' prior and reference priors based on the parameterization $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ for the model. These three types of objective priors are most popularly used in objective Bayesian analysis literature. The Fisher information matrix is the launching site in the search for these objective priors. We first derive the Fisher information matrix of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$.

Fact 2.2 *The Fisher information matrix for* $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$ *is*

$$\boldsymbol{I}_\psi = diag\left(n_1 \boldsymbol{\Psi}'\boldsymbol{\Psi}, n_2 \boldsymbol{\Psi}'\boldsymbol{\Psi}, (n_1+n_2)\boldsymbol{\Lambda}_{\psi 1}, \ldots, (n_1+n_2)\boldsymbol{\Lambda}_{\psi p}\right), \tag{43}$$

where, for $h = 1, \ldots, p$,

$$\begin{aligned} \boldsymbol{\Lambda}_{\psi h} &= (\boldsymbol{I}_h \ \ \mathbf{0})\mathbf{Var}(\mathbf{x})(\mathbf{I_h} \ \ \mathbf{0})' + \frac{\mathbf{1}}{\psi^{\mathbf{2}}_{\mathbf{hh}}}\mathbf{e_h}\mathbf{e'_h}, & (44)\\ \boldsymbol{e}'_h &= (0, \ldots, 0, 1) \in \mathbb{R}^h, \textit{i.e. the } h^{th} \textit{ element is 1, all others are 0.} & (45) \end{aligned}$$

Proof. The log-likelihood function under parameterizations $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$ is

$$\begin{aligned} &\log L\left(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi} | \boldsymbol{X}, \boldsymbol{Y}\right) \\ = \ & -\frac{n_1+n_2}{2} p \log(2\pi) + (n_1+n_2)\log|\boldsymbol{\Psi}| \\ & -\frac{1}{2}\sum_{i=1}^{n_1}(\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\boldsymbol{\Psi}'\boldsymbol{\Psi}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) - \frac{1}{2}\sum_{j=1}^{n_2}(\boldsymbol{y}_j - \boldsymbol{\mu}_2)'\boldsymbol{\Psi}'\boldsymbol{\Psi}(\boldsymbol{y}_j - \boldsymbol{\mu}_2) \\ = \ & -\frac{n_1+n_2}{2} p \log(2\pi) + (n_1+n_2)\log\prod_{h=1}^{p}\psi_{hh} \\ & -\frac{1}{2}\sum_{i=1}^{n_1}\sum_{h=1}^{p}\left(\sum_{l=1}^{h}\psi_{hl}(x_{il} - \mu_{1l})\right)^2 - \frac{1}{2}\sum_{j=1}^{n_2}\sum_{h=1}^{p}\left(\sum_{l=1}^{h}\psi_{hl}(y_{jl} - \mu_{2l})\right)^2. \end{aligned} \tag{46}$$

The Fisher information matrix for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$ is

$$\begin{aligned} \boldsymbol{I}_\psi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi}) &= -E\left(\frac{\partial^2 \log L}{\partial(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})'\partial(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})}\right) \\ &= diag\left(n_1 \boldsymbol{\Psi}'\boldsymbol{\Psi}, n_2 \boldsymbol{\Psi}'\boldsymbol{\Psi}, (n_1+n_2)\boldsymbol{\Lambda}_{\psi 1}, \ldots, (n_1+n_2)\boldsymbol{\Lambda}_{\psi p}\right), \end{aligned} \tag{47}$$

where, $\boldsymbol{\Lambda}_{\psi h}$ is defined by (44). Thus Fact 2.2 follows. □

Combine the Fisher information matrix of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$ and the Jacobian from $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$ into $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})$, we have the Fisher information matrix of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})$.

Fact 2.3 *The Fisher information matrix of* $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})$ *is*

$$\boldsymbol{I}_\Delta = diag\left(n_1 \boldsymbol{\Delta}' \Xi^2 \boldsymbol{\Delta}, n_2 \boldsymbol{\Delta}' \Xi^2 \boldsymbol{\Delta}, (n_1+n_2)\boldsymbol{\Lambda}_{\tau 1}, \ldots, (n_1+n_2)\boldsymbol{\Lambda}_{\tau p}\right), \tag{48}$$

where for $h = 1, \ldots, p$,

$$\begin{aligned} \boldsymbol{\Lambda}_{\tau h} &= \begin{pmatrix} \psi_{hh}^2 \left(\boldsymbol{\Delta}_{h-1}' \Xi_{h-1}^2 \boldsymbol{\Delta}_{h-1}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2}{\psi_{hh}^2} \end{pmatrix}, & (49) \\ \Xi_h &= \textit{the upper left } h \times h \textit{ matrix of } \Xi, & (50) \\ \boldsymbol{\Delta}_h &= \textit{the upper left } h \times h \textit{ matrix of } \boldsymbol{\Delta}. & (51) \end{aligned}$$

Proof. From (34), the Jacobian matrix of the transformation from $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})$ into $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})$ is

$$J_\psi = \frac{\partial(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\psi})}{\partial(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})} = diag(\boldsymbol{I}_p, \boldsymbol{I}_p, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_p), \tag{52}$$

where, for $h = 1, \ldots, p$, $\boldsymbol{\Lambda}_h$ is as defined in (42). Thus, the Fisher information matrix of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})$ is

$$\begin{aligned} \boldsymbol{I}_\Delta &= J_\psi' \boldsymbol{I}_\psi J_\psi \\ &= diag\left(n_1 \boldsymbol{\Delta}' \Xi^2 \boldsymbol{\Delta}, n_2 \boldsymbol{\Delta}' \Xi^2 \boldsymbol{\Delta}, (n_1+n_2)\boldsymbol{\Lambda}_{\tau 1}, \ldots, (n_1+n_2)\boldsymbol{\Lambda}_{\tau p}\right), & (53) \end{aligned}$$

where $\boldsymbol{\Lambda}_{\tau h}$ is as defined by (49). Fact 2.3 is then proved. □

2.8.2 The Objective Priors

One of the commonly used objective prior is the Jeffreys' prior, introduced by Jeffreys (1961). The incentive of the Jeffreys prior is that a prior should be invariant with respect to re-parametrization. It has big advantage over the constant priors introduced by Laplace (1812), such that same prior is used for various parameterizations. The Jeffreys' prior is derived by taking priors proportional to the square root of the determinant of the Fisher information matrix. In our case, the Jeffreys' prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ is given by the following proposition.

Proposition 2.1 *The Jeffreys' prior for* $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ *from the two population multivariate normal with common covariance is*

$$\pi_J(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) \quad \propto \quad \prod_{h=1}^{p} \psi_{hh}^{2h-p}. \tag{54}$$

Proof. Take the square root of the determinant of the Fisher information matrix as following

$$\begin{aligned} |\boldsymbol{I}_\Delta|^{\frac{1}{2}} &= \left(2\prod_{h=1}^{p}\psi_{hh}^4 \prod_{h=1}^{p}\prod_{l=1}^{h}\psi_{ll}^{-2}\prod_{h=1}^{p}\psi_{hh}^{2(h-1)}\right)^{\frac{1}{2}} \\ &= \left(2\prod_{h=1}^{p}\psi_{hh}^4 \prod_{h=1}^{p}\psi_{hh}^{-2(p-h+1)}\prod_{h=1}^{p}\psi_{hh}^{2(h-1)}\right)^{\frac{1}{2}} \\ &= 2^{\frac{1}{2}}\prod_{h=1}^{p}\psi_{hh}^{2h-p}. \end{aligned} \tag{55}$$

□

The Jeffreys' prior has been proved to work well for one-parameter cases, but may not work adequately for inference on multiple parameters within a model (see Berger & Bernardo (1992a)). Reference priors (see Bernardo (1979), Berger & Bernardo (1989), and Berger & Bernardo (1992a)) were then introduced to address multi-parameter cases by grouping the parameters and dividing them into the interest parameters and nuisance parameters with specific order of interest. For every grouping and ordering, a reference prior is derived. In most cases, treating each parameter as an individual group is recommended, and the corresponding reference prior derived is called a **one-at-a-time reference prior**. In some specific cases, different orderings of the parameters may end up with the same one-at-a-time reference prior. The model considered here is one of these specific cases. This is one of the motivations of re-parametrization from $\boldsymbol{\Psi}$ to $(\Xi, \boldsymbol{\Delta})$.

Fact 2.3 shows that the Fisher information matrix of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\tau})$ is a block diagonal matrix with the blocks corresponding to $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\delta}_{h,h-1}$ are functions free of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\delta}_{h,h-1}$ correspondingly. Based on the algorithms of reference priors given in Berger & Bernardo

(1992a), the one-at-a-time reference priors for such case depends on the elements of the Fisher information matrix corresponding to ψ_{hh} only and is given in the following proposition.

Lemma 2.2 *Under the parameterization* $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$*:*

(a) *The one-at-a-time reference prior for* $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ *with any ordering is of the form*

$$\pi_R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi,\ \boldsymbol{\Delta}) \propto \prod_{h=1}^{p} \frac{1}{\psi_{hh}}. \tag{56}$$

(b) *The right-Haar prior for* $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ *is same as its one-at-a-time reference prior given in part (a).*

Proof. For part (a), from the property of reference prior algorithm given by Berger & Bernardo (1992a), when the Fisher information matrix is block diagonal as in this case, the one-at-a-time reference prior is given by

$$\pi_R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) \quad \propto \quad \prod_{h=1}^{p} \frac{1}{|\psi_{hh}^2|^{\frac{1}{2}}} = \prod_{h=1}^{p} \frac{1}{\psi_{hh}}. \tag{57}$$

For part (b), from Example 1.3, we know that the left-Haar measure of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Psi})$ is

$$v_l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Psi}) \quad \propto \quad \frac{d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_1 d\boldsymbol{\Psi}}{\prod_{h=1}^{p} \psi_{hh}^{h}} \tag{58}$$

and the Jacobian from $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Psi})$ to $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ is $J_\psi = \prod_{h=1}^{p} \psi_{hh}^{h-1}$, thus the left-Haar prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ is given by

$$\begin{aligned} v_l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) &= v_l(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Psi}) \times J_\psi^{-1} \\ &= \frac{d\boldsymbol{\mu}_1 d\boldsymbol{\mu}_1 d\Xi d\boldsymbol{\Delta}}{\prod_{h=1}^{p} \psi_{hh}}. \end{aligned} \tag{59}$$

A special note is that the $\boldsymbol{\Psi}$ is a Cholesky decomposition for $\boldsymbol{\Omega}$ such that $\boldsymbol{\Omega} = \boldsymbol{\Psi}'\boldsymbol{\Psi}$, thus the left-Haar measure for $\boldsymbol{\Psi}$ is more true to be the right-Haar measure from the model point of view. Thus from now on, we rename this left-Haar measure as right-Haar measure. □

(54) and (56) clearly show that both the Jeffreys' prior and the one-at-a-time reference prior are special cases of the following class of priors,

$$\pi(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) \propto \prod_{h=1}^{p} \frac{1}{\psi_{hh}^{a_h}}. \tag{60}$$

The a_h are real values. For the rest of this chapter, we confine our discussion within this class of priors.

2.9 Posterior Distribution of KL Divergence

One of the important merits of objective Bayesian analysis is that posteriors of parameters offer great flexibility for all kinds of inference for interesting parameters or functions of parameters. This section is dedicated to the study of joint posterior distributions, marginal posterior distributions, and conditional posterior distributions of model parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$, the KL divergence, and related parts of the KL divergence, under the class of objective priors given by (60).

Let s_{hl} be the element of $\boldsymbol{S}$ located at the h^{th} row and l^{th} column. For $h = 1, \ldots, p$, we define some new notation:

$$\begin{aligned}
\boldsymbol{s}_{h,h-1} &= (s_{h1}, \ldots, s_{h(h-1)})', && (61)\\
\boldsymbol{S}_h &= \text{the upper left } h \times h \text{ matrix of } \boldsymbol{S}, && (62)\\
t_h &= \begin{cases} s_{11}, & \text{if } h = 1, \\ s_{hh} - \boldsymbol{s}'_{h,h-1}\boldsymbol{S}_{h-1}\boldsymbol{s}_{h,h-1} & \text{if } h = 2, \ldots, p, \end{cases} && (63)\\
\boldsymbol{g}_h &= \boldsymbol{S}_{h-1}^{-1}\boldsymbol{s}_{h,h-1}, && (64)\\
\boldsymbol{\delta}_{h,h-1} &= (\delta_{h1}, \ldots, \delta_{h(h-1)})'. && (65)
\end{aligned}$$

In order to derive the posteriors, we need the following proposition to expand $tr\{\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\}$.

Proposition 2.2 *We have the following formula,*

$$tr\{\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\} = \sum_{h=1}^{p} \psi_{hh}^2 t_h + \sum_{h=2}^{p} \left[\psi_{hh}^2(\boldsymbol{\delta}_{h,h-1} + \boldsymbol{g}_h)'\boldsymbol{S}_{h-1}(\boldsymbol{\delta}_{h,h-1} + \boldsymbol{g}_h)\right]. \tag{66}$$

Proof. Let $\Xi_h = diag(\psi_{11}, \ldots, \psi_{hh})$, then

$$\begin{aligned}
\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S} &= \begin{pmatrix} \boldsymbol{\Delta}_{p-1} & \mathbf{0} \\ \boldsymbol{\delta}'_{p,p-1} & 1 \end{pmatrix}' \begin{pmatrix} \Xi_{p-1} & \mathbf{0} \\ \mathbf{0} & \psi_{pp}^2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Delta}_{p-1} & \mathbf{0} \\ \boldsymbol{\delta}'_{p,p-1} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{S}_{p-1} & \boldsymbol{s}_{p,p-1} \\ \boldsymbol{s}'_{p,p-1} & s_{pp} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Delta}'_{p-1}\Xi_{p-1}^2\boldsymbol{\Delta}_{p-1}\boldsymbol{S}_{p-1} + \psi_{pp}^2\boldsymbol{\delta}_{p,p-1}(\boldsymbol{\delta}'_{p,p-1}\boldsymbol{S}_{p-1} + \boldsymbol{s}'_{p,p-1}) & * \\ * & \psi_{pp}^2(\boldsymbol{\delta}'_{p,p-1}\boldsymbol{s}_{p,p-1} + s_{pp}) \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Delta}'_{p-1}\Xi_{p-1}^2\boldsymbol{\Delta}_{p-1}\boldsymbol{S}_{p-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{pmatrix} \\
&\quad + \begin{pmatrix} \psi_{pp}^2\boldsymbol{\delta}_{p,p-1}(\boldsymbol{\delta}'_{p,p-1}\boldsymbol{S}_{p-1} + \boldsymbol{s}'_{p,p-1}) & * \\ * & \psi_{pp}^2(\boldsymbol{\delta}'_{p,p-1}\boldsymbol{s}_{p,p-1} + s_{pp}) \end{pmatrix}, \qquad (67)
\end{aligned}$$

where $*$ parts are unrelated to the calculations of $tr(\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})$. Clearly,

$$\begin{aligned}
tr\{\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\} &= tr\{\boldsymbol{\Delta}'_{p-1}\Xi_{p-1}^2\boldsymbol{\Delta}_{p-1}\boldsymbol{S}_{p-1}\} \\
&\quad + tr\{\psi_{pp}^2\boldsymbol{\delta}_{p,p-1}(\boldsymbol{\delta}'_{p,p-1}\boldsymbol{S}_{p-1} + \boldsymbol{s}'_{p,p-1})\} + \psi_{pp}^2(\boldsymbol{\delta}'_{p,p-1}\boldsymbol{s}_{p,p-1} + s_{pp}) \\
&= tr\{\boldsymbol{\Delta}'_{p-1}\Xi_{p-1}^2\boldsymbol{\Delta}_{p-1}\boldsymbol{S}_{p-1}\} \\
&\quad + \psi_{pp}^2 t_p + \psi_{pp}^2(\boldsymbol{\delta}_{p,p-1} + \boldsymbol{g}_p)'\boldsymbol{S}_{p-1}(\boldsymbol{\delta}_{p,p-1} + \boldsymbol{g}_p). \qquad (68)
\end{aligned}$$

By iteration, Proposition 2.2 follows. □

Using the above expansion, we get the posteriors for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ and collect them into the following lemma.

Lemma 2.3 *Assume that $n_1+n_2-h-a_h > 0$ for any $h = 1, \ldots, p$ and the class of objective priors in (60), the following results are valid:*

(a). *The joint posterior of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ is proper.*

(b). *The conditional posteriors of $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ given $(\Xi, \boldsymbol{\Delta})$ are independently normal distributions with the following distributions,*

$$\begin{aligned}
(\boldsymbol{\mu}_1 \mid \Xi, \boldsymbol{\Delta}; \boldsymbol{X}, \boldsymbol{Y}) &\sim N_p\left(\overline{\boldsymbol{x}}, \frac{1}{n_1}(\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1}\right), \qquad (69) \\
(\boldsymbol{\mu}_2 \mid \Xi, \boldsymbol{\Delta}; \boldsymbol{X}, \boldsymbol{Y}) &\sim N_p\left(\overline{\boldsymbol{y}}, \frac{1}{n_2}(\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1}\right). \qquad (70)
\end{aligned}$$

(c). *The posteriors of* $\boldsymbol{\delta}_{h,h-1}$ *given* $(\psi_{11},\ldots,\psi_{hh})$ *are mutually independent. The posterior of* $\boldsymbol{\delta}_{h,h-1}$ *depends only on* ψ_{hh} *and is given by*

$$(\boldsymbol{\delta}_{h,h-1} \mid \psi_{hh}; \boldsymbol{X}, \boldsymbol{Y}) \sim N_{h-1}(-\boldsymbol{g}_h, \psi_{hh}^{-2}\boldsymbol{S}_{h-1}^{-1}),\ h = 2,\ldots,p. \tag{71}$$

(d). *The marginal posterior of* $(\psi_{11},\ldots,\psi_{hh})$ *is*

$$(\psi_{hh}^2 \mid \boldsymbol{X}, \boldsymbol{Y}) \sim Gamma(\frac{1}{2}(n_1+n_2-h-a_h), \frac{1}{2}t_h), \quad h = 1,\ldots,p. \tag{72}$$

Proof. The joint posterior density function of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ under the class of priors (60) is

$$\begin{aligned}
& [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}|\boldsymbol{X}, \boldsymbol{Y}] \\
\propto\ & L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}|\boldsymbol{X}, \boldsymbol{Y}) \times \prod_{h=1}^{p} \frac{1}{\psi_{hh}^{a_h}} \\
\propto\ & (2\pi)^{-\frac{n_1+n_2}{2}p} \prod_{h=1}^{p} \psi_{hh}^{n_1+n_2-a_h} exp\{-\frac{n_1}{2}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}})'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}}) \\
& -\frac{n_2}{2}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}})'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \overline{\boldsymbol{x}})\} etr\{-\frac{1}{2}\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\}.
\end{aligned} \tag{73}$$

Part (a) and (b) hold immediately. For part (c), note that the marginal posterior of $(\Xi, \boldsymbol{\Delta})$ given $(\boldsymbol{X}, \boldsymbol{Y})$ is

$$\begin{aligned}
[\Xi, \boldsymbol{\Delta}|\boldsymbol{X}, \boldsymbol{Y}] &\propto \prod_{h=1}^{p} \psi_{hh}^{n_1+n_2-2-a_h} etr\{-\frac{1}{2}\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\} \\
&\propto \prod_{h=1}^{p} \psi_{hh}^{n_1+n_2-2-a_h} \times \\
&\quad \exp\left\{-\frac{1}{2}\left[\sum_{h=1}^{p} \psi_{hh}^2 t_h + \sum_{h=2}^{p} (\psi_{hh}^2\boldsymbol{\delta}_{h,h-1} + \boldsymbol{g}_h)'\boldsymbol{S}_{h-1}(\psi_{hh}^2\boldsymbol{\delta}_{h,h-1} + \boldsymbol{g}_h)\right]\right\}.
\end{aligned} \tag{74}$$

Similarly we integrate out $\boldsymbol{\delta}_{h,h-1}$ and get the marginal distribution of Ξ as

$$\begin{aligned}
(\psi_{11},\ldots,\psi_{pp}|\boldsymbol{X}, \boldsymbol{Y}) &\propto \prod_{h=1}^{p} \psi_{hh}^{n_1+n_2-h-1-a_h} \exp\left\{-\frac{1}{2}\sum_{h=1}^{p} \psi_{hh}^2 t_h\right\}, \\
(\psi_{11}^2,\ldots,\psi_{pp}^2|\boldsymbol{X}, \boldsymbol{Y}) &\propto \prod_{h=1}^{p} (\psi_{hh}^2)^{\frac{1}{2}(n_1+n_2-h-a_h)-1} \exp\left\{-\frac{1}{2}\sum_{h=1}^{p} (\psi_{hh}^2) t_h\right\}.
\end{aligned} \tag{75}$$

Part(d) of Lemma 2.3 is then proved. □

Lemma 2.3 shows that, under priors (60), if $n_1 + n_2 - h - a_h > 0$ for any $h = 1, \ldots, p$, the marginal posterior distribution of ψ_{hh}, conditional posterior distributions of $\boldsymbol{\delta}_{h,h-1}$ and $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ are all proper. Therefore, by plugging in the respective posterior distributions and conditional posterior distributions, posterior distributions of functions of parameters may be derived.

In this study, the interesting function is the KL divergence $K = \frac{1}{2}\|\boldsymbol{\xi}\|$ as defined in (28). where

$$\begin{aligned}
\boldsymbol{\xi} &= \Xi\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\
&= \begin{pmatrix} \psi_{11}(\mu_{11} - \mu_{21}) \\ \psi_{22}\delta_{21}(\mu_{11} - \mu_{21}) + \psi_{22}(\mu_{12} - \mu_{22}) \\ \vdots \\ \psi_{pp}\delta_{p1}(\mu_{11} - \mu_{21}) + \ldots + \psi_{pp}\delta_{p(p-1)}(\mu_{1(p-1)} - \mu_{2(p-1)}) + \psi_{22}(\mu_{1p} - \mu_{2p}) \end{pmatrix} \\
&= \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_p \end{pmatrix}.
\end{aligned} \tag{76}$$

Thus, ξ_h is a function of ψ_{hh}, $\boldsymbol{\delta}_{h,h-1}$, $(\mu_{11}, \ldots, \mu_{1h})$, and $(\mu_{21}, \ldots, \mu_{2h})$. The simulation of ξ_h becomes possible based on the clear structure of the posteriors of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ given by Lemma 2.3.

In this study, we will write posteriors in the form of **constructive random posteriors**, introduced by Berger & Sun (2008). Such notion of constructive random posteriors can be used for posterior computation. More importantly, this will provide a powerful tool for verifying the frequentist matching properties of the Bayesian procedures, which is the content of Section 3.14. The main feature is to use the notation that * appended to a random variable denotes randomness arising from the constructive posterior (i.e., from the random variables used in simulation from the posterior), while a random variable without a

* refers to randomness arising from the (frequentist) distribution of a statistic. Also, let Z_{ij} denote standard normal random variables and χ^2_m denote chi-squared random variables with specified degree of freedom. Whenever several of these occur in an expression, they are all independent (except that random variables of the same type and with the same index refer to the same random variable). Finally, we reserve quartile notation for posterior quartiles, with respect to the * distributions.

For later use, we decompose $\boldsymbol{S}$ into the product a lower triangle matrix $\boldsymbol{V}$ through Cholesky decomposition, such as

$$\boldsymbol{S} \;=\; \boldsymbol{V}\boldsymbol{V}'. \tag{77}$$

Some useful relationships between $\boldsymbol{S}$ and $\boldsymbol{V}$ are given below.

Fact 2.4 *For $h = 1, \ldots, p$,*

(a). $\boldsymbol{S}_h = \boldsymbol{V}_h\boldsymbol{V}_h'$. *$\boldsymbol{S}_h$ and $\boldsymbol{V}_h$ are the upper and left $h \times h$ matrix of $\boldsymbol{S}$ and $\boldsymbol{V}$ respectively.*

(b). *Define $|\boldsymbol{S}_0| = 1$, then $t_h = v_{hh}^2 = \frac{|\boldsymbol{S}_h|}{|\boldsymbol{S}_{h-1}|}$.*

(c). $\boldsymbol{S}_{h-1}^{-1}\boldsymbol{s}_{h,h-1} = \boldsymbol{V}_{h-1}'^{-1}\boldsymbol{v}_{h,h-1}$.

Proof. For $h = 1, \ldots, p$,

$$\begin{aligned}
\boldsymbol{S}_h^{-1} &= \begin{pmatrix} \boldsymbol{S}_{h-1}^{-1} + \boldsymbol{S}_{h-1}^{-1}\boldsymbol{s}_{h,h-1}'\boldsymbol{S}_{h-1}^{-1}\boldsymbol{s}_{h,h-1}t_h^{-1} & -t_h^{-1}\boldsymbol{S}_{h-1}^{-1}\boldsymbol{s}_{h,h-1} \\ -t_h^{-1}\boldsymbol{s}_{h,h-1}'\boldsymbol{S}_{h-1}^{-1} & t_h^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{V}_{h-1}'^{-1}\boldsymbol{V}_{h-1}^{-1} + \boldsymbol{V}_{h-1}'^{-1}\boldsymbol{v}_{h,h-1}\boldsymbol{v}_{h,h-1}'\boldsymbol{V}_{h-1}v_{hh}^{-2} & -v_{hh}^{-2}\boldsymbol{V}_{h-1}'^{-1}\boldsymbol{v}_{h,h-1} \\ -v_{hh}^{-2}\boldsymbol{v}_{h,h-1}'\boldsymbol{V}_{h-1}^{-1} & v_{hh}^{-2} \end{pmatrix} \\
&= \boldsymbol{V}_h'^{-1}\boldsymbol{V}_h^{-1}.
\end{aligned} \tag{78}$$

Therefore, Fact 2.4 follows. □

Here, we rewrite the posteriors of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Xi}, \boldsymbol{\Delta})$ from Lemma 2.3 into constructive random posteriors and collect them in the following lemma.

Lemma 2.4 *Under priors (60), assume that* $|\boldsymbol{S}_h| \neq 0$ *and* $n_1 + n_2 - h - a_h > 0$*:*

(a). *The constructive random posterior of* ψ_{hh} *given* $\boldsymbol{X}, \boldsymbol{Y}$ *has the expression*

$$\psi^*_{hh} = \frac{\chi^*_{n_1+n_2-h-a_h}}{\sqrt{t_h}}, \quad h = 1, \ldots, p. \tag{79}$$

(b). *The constructive random posterior of* $\boldsymbol{\delta}_{h,h-1}$ *given* $\boldsymbol{X}$, $\boldsymbol{Y}$, *and* ψ^*_{hh} *has the expression*

$$\boldsymbol{\delta}^*_{h,h-1} = -\boldsymbol{S}^{-1}_{h-1}\boldsymbol{s}_{h,h-1} + \frac{1}{\psi^*_{hh}}\boldsymbol{V}'^{-1}_{h-1}\boldsymbol{z}^*_{h,h-1}. \tag{80}$$

(c). *Let* $c_0 = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, *the constructive random posteriors of* $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, *and* $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ *given* $\boldsymbol{X}$, $\boldsymbol{Y}$, Ξ^*, *and* $\boldsymbol{\Delta}^*$ *have the expressions*

$$\boldsymbol{\mu}^*_1 = \overline{\boldsymbol{x}} + \sqrt{\frac{1}{n_1}}(\Xi^*\boldsymbol{\Delta}^*)^{-1}\boldsymbol{z}^*_{2,p}, \tag{81}$$

$$\boldsymbol{\mu}^*_2 = \overline{\boldsymbol{y}} + \sqrt{\frac{1}{n_2}}(\Xi^*\boldsymbol{\Delta}^*)^{-1}\boldsymbol{z}^*_{3,p}, \tag{82}$$

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^* = \overline{\boldsymbol{x}} - \overline{\boldsymbol{y}} + c_0(\Xi^*\boldsymbol{\Delta}^*)^{-1}\boldsymbol{z}^*_{4,p}. \tag{83}$$

In addition, define $\boldsymbol{W}^* = \Xi^*\boldsymbol{\Delta}^*\boldsymbol{V}$, then $\boldsymbol{W}^* = (w^*_{hl})$ is a lower triangle matrix, and we have the following corollary.

Corollary 2.2 *Assume the priors (60), if* $n_1 + n_2 - h - a_h > 0$, *the distribution of* w^*_{hl} *has the following properties,*

(a). *all* w^*_{hl} *are mutually independent.*

(b). $w^{2*}_{hh} \sim \chi^2_{n_1+n_2-h-a_h}$, $\quad h = 1, \ldots, p.$

(c). $w^*_{hl} \sim N(0, 1)$, $\quad h = 2, \ldots, p,\ l = 1, \ldots, h-1.$

Proof. In the previous proof, we derive the joint posterior density function of $(\Xi, \boldsymbol{\Delta})$ as

$$[\Xi, \boldsymbol{\Delta}|\boldsymbol{X}, \boldsymbol{Y}] \propto \prod_{h=1}^{p} \psi_{hh}^{n_1+n_2-2-a_h} etr\{-\frac{1}{2}\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\}. \tag{84}$$

Since $\boldsymbol{S} = \boldsymbol{V}\boldsymbol{V}'$ and $\boldsymbol{W} = \Xi\boldsymbol{\Delta}\boldsymbol{V}$, we have

$$\begin{aligned} [\boldsymbol{W}|\boldsymbol{X},\boldsymbol{Y}] &\propto \prod_{h=1}^{p} w_{hh}^{n_1+n_2-h-1-a_h} etr\{-\frac{1}{2}\boldsymbol{W}'\boldsymbol{W}\} \\ &\propto \prod_{h=1}^{p} w_{hh}^{n_1+n_2-h-1-a_h} exp\{-\frac{1}{2}\prod_{h=1}^{p}\prod_{j=1}^{h} w_{hj}^2\}. \end{aligned} \tag{85}$$

□

From Corollary 2.2, the constructive random posterior of $\boldsymbol{\xi}$ is derived and given as below.

Proposition 2.3 *Assume the priors (60),*

$$\boldsymbol{\xi}^* = \boldsymbol{W}^*\boldsymbol{V}^{-1}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}_{4p}^*, \tag{86}$$

$$K = \frac{1}{2}\boldsymbol{\xi}^{*\prime}\boldsymbol{\xi}^*. \tag{87}$$

Proof. Plug $(\boldsymbol{\mu}_1^*, \boldsymbol{\mu}_2^*, \Xi^*, \boldsymbol{\Delta}^*)$ into expression of $\boldsymbol{\xi}$, then

$$\begin{aligned} \boldsymbol{\xi}^* &= \Xi^*\boldsymbol{\Delta}^*(\boldsymbol{\mu}_1^* - \boldsymbol{\mu}_2^*) \\ &= \Xi^*\boldsymbol{\Delta}^*(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}_{4,p}^* \\ &= \Xi^*\boldsymbol{\Delta}^*\boldsymbol{V}\boldsymbol{V}^{-1}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}_{4,p}^* \\ &= \boldsymbol{W}^*\boldsymbol{V}^{-1}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}_{4,p}^*. \end{aligned} \tag{88}$$

□

Based on Lemma 2.4, the constructive random posteriors of each component of $\boldsymbol{\xi}$ and K are given in the following fact.

Proposition 2.4 *Let $b_h = n_1 + n_2 - h - a_h$. Under prior (60),*

$$\begin{aligned} \xi_1^* &= \chi_{b_1}^* \frac{(\overline{x}_1 - \overline{y}_1)}{v_{11}} + c_0 z_{41}^*, \\ \xi_h^* &= \begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}' \boldsymbol{V}_h^{-1}(\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{y}}_h) + c_0 z_{4h}^*, \quad h = 2, \ldots, p. \end{aligned} \tag{89}$$

$$K^* = \frac{1}{2}\boldsymbol{\xi}^{*\prime}\boldsymbol{\xi}^*. \tag{90}$$

Proof. Under prior (60)

$$
\begin{aligned}
\boldsymbol{\xi}^* &= \Xi^*\boldsymbol{\Delta}^*\left(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}} + c_0(\Xi^*\boldsymbol{\Delta}^*)^{-1}\boldsymbol{z}^*_{4,p}\right) \\
&= \Xi^*\boldsymbol{\Delta}^*(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p} \\
&= \begin{pmatrix}
\psi^*_{11}(\overline{x}_1 - \overline{y}_1) \\
\psi^*_{22}\delta^*_{21}(\overline{x}_1 - \overline{y}_1) + \psi^*_{22}(\overline{x}_2 - \overline{y}_2) \\
\vdots \\
\psi^*_{pp}\delta^*_{p1}(\overline{x}_1 - \overline{y}_1) + \ldots + \psi^*_{pp}\delta^*_{p(p-1)}(\overline{x}_{(p-1)} - \overline{y}_{(p-1)}) + \psi^*_{22}(\overline{x}_p - \overline{y}_p)
\end{pmatrix} \\
&\quad + c_0\boldsymbol{z}^*_{4,p},
\end{aligned} \tag{91}
$$

that is,

$$
\xi^*_1 = \chi^*_{b_1}\frac{(\overline{x}_1 - \overline{y}_1)}{v_{11}} + c_0 z^*_{41}, \tag{92}
$$

$$
\begin{aligned}
\xi^*_h &= \psi^*_{hh}\boldsymbol{\delta}'^*_{h,h-1}(\overline{\boldsymbol{x}}_{h-1} - \overline{\boldsymbol{y}}_{h-1}) + \psi^*_{22}(\overline{x}_h - \overline{y}_h) + c_0 z^*_{4h} \\
&= \left(\chi^*_{b_h}(-\frac{\boldsymbol{V}'^{-1}_{h-1}\boldsymbol{v}_{h,h-1}}{v_{hh}}) + \boldsymbol{V}'^{-1}_{h-1}\boldsymbol{z}^*_{h,h-1}\right)'(\overline{\boldsymbol{x}}_{h-1} - \overline{\boldsymbol{y}}_{h-1}) \\
&\qquad + \chi^*_{b_h}\frac{1}{v_{hh}}(\overline{x}_h - \overline{y}_h) + c_0 z^*_{4h}.
\end{aligned} \tag{93}
$$

Thus Fact 2.4 follows. □

Proposition 2.4 shows that we can plug the observed sample mean and sample covariance matrix $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}, \boldsymbol{V})$ into the K^* and run direct Monte Carlo simulations to simulate the posterior density of K.

2.10 Frequentist Matching Property

2.10.1 Preliminary

Section 2.9 derives the constructive random posteriors of KL divergence under the class of priors (60). In order to put these constructive random posteriors into practice, we need to justify their use. For general discussion, there is no available loss function. The frequentist

Statistics is widely acceptable to be more "objective". Thus, one approach for justification is to compare the results from the constructive random posterior to its counterparts from frequentist method. Which is very similar to the calibrated Bayes proposed by Little (2006). As is customary in such comparisons, the one-sided interval $(\theta_L, \theta_\alpha(x))$ of a function of parameters $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ is studied. Here θ_L is the lower bound of θ, and $\theta_\alpha(x)$ is the posterior quartile of θ, defined by

$$P(\theta < \theta_\alpha(x)|x) = \alpha. \tag{94}$$

The interesting quantity is the frequentist coverage of the corresponding confidence interval

$$P(\theta < \theta_\alpha(x)|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}). \tag{95}$$

The closer this coverage is to the nominal α, the more non-informative of the priors is judged to be. The priors could be treated as a better prior for the Bayesian procedure in the sense of "Objectiveness". The order of frequentist matching is defined as below

$$P(\theta < \theta_\alpha(X_n)|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) = \begin{cases} \alpha, & \text{exact matching;} \\ \alpha + o\left((\frac{1}{\sqrt{n}})^1\right), & \text{1st order matching;} \\ \alpha + o\left((\frac{1}{\sqrt{n}})^2\right), & \text{2nd order matching.} \end{cases}$$

The best we can get from this kind of matching comparison is **exact matching**. If exact matching can not be fulfilled, asymptotic matching properties could be studied. First order or second order matching offer desirable asymptotic matching properties, as the difference decreases proportional to n and n^2 respectively, where n is the sample size. Datta & Mukerjee (2004) also introduced a series of **higher order** frequentist matching methods, such as quartile matching, distribution function matching, and highest posterior density region matching. In this section, the frequentist matching properties of the posteriors derived in Section 2.9 are discussed. We prove that one-at-a-time reference prior is exact frequentist matching prior for each of the ξ_h, certainly they are first order matching prior as well.

From Section 2.7, we know that $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}, \boldsymbol{S})$ are the sufficient statistics for the model parameters, and Section 2.9 shows that constructive random posteriors of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ and

KL divergence are functions of these sufficient statistics and standard distributions. This section starts from the study of the distribution of these sufficient statistics under the parameterization $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$.

The joint distribution of $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2})$ is

$$\begin{aligned}
& f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{n_2} | \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) \\
= \;& (2\pi)^{-\frac{n_1+n_2}{2}p} |\Xi|^{n_1+n_2} exp\{-\frac{1}{2}\sum_{i=1}^{n_1} (\boldsymbol{x}_i - \boldsymbol{\mu}_1)'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{x}_i - \boldsymbol{\mu}_1) \\
& -\frac{1}{2}\sum_{j=1}^{n_2} (\boldsymbol{y}_j - \boldsymbol{\mu}_2)'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\boldsymbol{y}_j - \boldsymbol{\mu}_2)\} \\
= \;& (2\pi)^{-\frac{n_1+n_2}{2}p} \prod_{h=1}^{p} \psi_{hh}^{n_1+n_2} exp\{-\frac{n_1}{2}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_1)'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_1) \\
& -\frac{n_2}{2}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_2)'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_2)\} etr\{-\frac{1}{2}\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\}, \qquad (96)
\end{aligned}$$

where $\overline{\boldsymbol{x}}$, $\overline{\boldsymbol{y}}$, and $\boldsymbol{S}$ are as defined in (25).

This joint distribution (96) implies Fact 2.5.

Fact 2.5

(a). *$(\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}, \boldsymbol{S})$ are sufficient statistics for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$, as we claim in Section 2.7.*

(b). *$\overline{\boldsymbol{x}}$, $\overline{\boldsymbol{y}}$, and $\boldsymbol{S}$ are mutually independent.*

(c). *The distributions of $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ are*

$$\overline{\boldsymbol{x}} \sim N_p(\boldsymbol{\mu}_1, \frac{1}{n_1}(\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1}), \qquad (97)$$

$$\overline{\boldsymbol{y}} \sim N_p(\boldsymbol{\mu}_2, \frac{1}{n_2}(\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1}). \qquad (98)$$

(d). *$\boldsymbol{S}$ follows a Wishart distribution with parameters p, n_1+n_2-2, and $(\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1}$. (Refer to Gupta & Nagar (2000, Page 92-93) plus some simple derivations).*

Let $n = n_1 + n_2 - 2$, then

$$\boldsymbol{S} \sim W_p(n, (\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1}). \tag{99}$$

The frequentist distribution of $\boldsymbol{V}$ is given by Lemma 2.5.

Lemma 2.5 *The elements of $\boldsymbol{V}$ follow the following distributions,*

(a). $\psi_{hh}^2 v_{hh}^2$ *independently* $\sim \chi^2_{n-h+1}$, $h = 1, \ldots, p$.

(b). $\psi_{hh}(\sum_{l=k}^{h} v_{lk}\delta_{hl})$ *independently* $\sim N(0,1)$, *for* $1 \leq k \leq h-1 \leq p-1$.

Proof. Fact 2.5 gives the result that $\boldsymbol{S} \sim W_p(n_1 + n_2 - 2, (\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta})^{-1})$. From Gupta & Nagar (2000, page14), the Jacobian determinant of the transformation from $\boldsymbol{S}$ into the lower triangle matrix $\boldsymbol{V}$ is $2^p \prod\limits_{h=1}^{p} v_{hh}^{p-h+1}$.

Let $n = n_1 + n_2 - 2$, the p.d.f of $\boldsymbol{V}$ is

$$\begin{aligned}
f(\boldsymbol{V}|\Xi, \boldsymbol{\Delta}) &= \left(2^{\frac{1}{2}p}\Gamma_p(\frac{1}{2}n)|\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}|^{-\frac{1}{2}n}\right)^{-1} |\boldsymbol{S}|^{\frac{1}{2}(n-p-1)} etr\left(-\frac{1}{2}\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{S}\right) \cdot 2^p \prod_{h}^{p} v_{hh}^{p-h+1} \\
&= \left(2^{\frac{1}{2}p}\Gamma_p(\frac{1}{2}n)\right)^{-1} |\Xi|^n |\boldsymbol{V}|^{n-p-1} etr\left(-\frac{1}{2}\boldsymbol{V}'\boldsymbol{\Delta}'\Xi^2\boldsymbol{\Delta}\boldsymbol{V}\right) \cdot 2^p \prod_{h}^{p} v_{hh}^{p-h+1} \\
&= 2^{\frac{1}{2}p}\Gamma_p^{-1}(\frac{1}{2}n) \prod_{h=1}^{p} \psi_{hh}^n \prod_{h=1}^{p} v_{hh}^{n-h} etr\left(-\frac{1}{2}(\Xi\boldsymbol{\Delta}\boldsymbol{V})'(\Xi\boldsymbol{\Delta}\boldsymbol{V})\right) \\
&\propto \prod_{h=1}^{p} v_{hh}^{n-h} exp\left(-\frac{1}{2}\sum_{h=1}^{p}\psi_{hh}^2 v_{hh}^2\right) exp\left(-\frac{1}{2}\sum_{h=2}^{p}\sum_{k=1}^{h-1}\left(\psi_{hh}\sum_{l=k}^{h} v_{lk}\delta_{hl}\right)^2\right), \qquad (100)
\end{aligned}$$

where, $\Gamma_p^{-1}(\frac{1}{2}n)$ is the multivariate gamma function defined as,

$$\Gamma_p^{-1}(\frac{1}{2}n) = \int_{\boldsymbol{A}>0} |\ \boldsymbol{A}\ |^{\frac{n-p-1}{2}} etr(\boldsymbol{A})d\boldsymbol{A}. \tag{101}$$

for $n > p - 1$, and integral over the space of $p \times p$ symmetric positive definite matrices. Then, Lemma 2.5 follows. □

Corollary 2.3 *Let $\boldsymbol{W} = \Xi\boldsymbol{\Delta}\boldsymbol{V}$, then $\boldsymbol{W} = (w_{hl})$ is a lower triangle matrix with positive diagonal elements and density function*

$$f(\boldsymbol{W}) = \frac{2^{\frac{1}{2}p}}{\Gamma_p(\frac{1}{2}(n_1+n_2-2))}\prod_{h=1}^{p}\left(w_{hh}^{n-h}etr(-\frac{1}{2}w_{hh}^2)\right)etr(-\frac{1}{2}\sum_{h=1}^{p-1}\sum_{l=h+1}^{p}w_{lh}^2). \quad (102)$$

Thus, for $h = 1,\ldots,p$, let $c_h = n_1+n_2-1-h$,

(a). $w_{hh}^2 \sim \chi^2_{c_h}, \quad h = 1,\ldots,p.$

(b). $w_{hl} \sim N(0,1), \quad h = 2,\ldots,p,\ l = 1,\ldots,h-1.$

In the subsections below, we evaluate the frequentist properties of the posterior credible intervals for $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_p)'$ under the class of priors in (60).

2.10.2 Exact Frequentist Matching using Constructive Random Posterior

In this subsection we prove the exact frequentist matching properties of all ξ_h using constructive random posterior. To prove the exact frequentist matching property of ξ_h^*, we need to reorganize the expression of ξ_h^*. From (89) and Fact 2.4, we have

$$\xi_1^* = \chi_{b_1}^*\frac{1}{\psi_{11}v_{11}}\psi_{11}(\overline{x}_1-\overline{y}_1)+c_0z_{41}^*, \quad (103)$$

$$\begin{aligned}
\xi_h^* &= \begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}'\boldsymbol{V}_h^{-1}(\overline{\boldsymbol{x}}_h-\overline{\boldsymbol{y}}_h)+c_0z_{4h}^* \\
&= \begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}'\boldsymbol{V}_h^{-1}(\Xi_h\boldsymbol{\Delta}_h)^{-1}(\Xi_h\boldsymbol{\Delta}_h)(\overline{\boldsymbol{x}}_h-\overline{\boldsymbol{y}}_h)+c_0z_{4h}^* \\
&= \begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}'(\Xi_h\boldsymbol{\Delta}_h\boldsymbol{V}_h)^{-1}(\Xi_h\boldsymbol{\Delta}_h)(\overline{\boldsymbol{x}}_h-\overline{\boldsymbol{y}}_h)+c_0z_{4h}^* \\
&= \begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}'\boldsymbol{W}_h^{-1}(\Xi_h\boldsymbol{\Delta}_h)(\overline{\boldsymbol{x}}_h-\overline{\boldsymbol{y}}_h)+c_0z_{4h}^*, \quad h=2,\ldots,p. \quad (104)
\end{aligned}$$

Thus, for $h = 2, \ldots, p$, the α^{th} quartile of ξ_1^* and ξ_h^* are

$$\xi_{1,\alpha}^* = \left[\chi_{b_1}^* \frac{1}{\psi_{11} v_{11}} \psi_{11}(\overline{x}_1 - \overline{y}_1) + c_0 z_{41}^*\right]_\alpha, \quad (105)$$

$$\xi_{h,\alpha}^* = \left[\begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^{2*} \end{pmatrix}' \boldsymbol{W}_h^{-1}(\Xi_h \boldsymbol{\Delta}_h)(\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{y}}_h) + c_0 z_{4h}^*\right]_\alpha. \quad (106)$$

In order to prove frequentist matching for complicated parameter functions, the following lemma from Berger & Sun (2008) is utilized frequently.

Lemma 2.6 *Let Y_α denote the α quartile of any random variable Y.*

(a) *If $g(\cdot)$ is a monotonically increasing function, $[g(Y)]_\alpha = g(Y_\alpha)$ for any $\alpha \in (0, 1)$.*

(b) *For any $a > 0$, $b \in \mathbb{R}$, $[aY + b]_\alpha = aY_\alpha + b$.*

(c) *If W is a positive random variable, $[WY]_\alpha \geq 0$ if and only if $Y_\alpha \geq 0$.*

Here, we propose another lemma to be employed in the verification of frequentist matching.

Lemma 2.7 *Let $\boldsymbol{W}$, $\boldsymbol{Y}_1$, and $\boldsymbol{Y}_2$ be mutually independent random vectors. Let $g(\cdot)$ be a real function. If $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are identically distributed, then*

$$P\big(g(\boldsymbol{W}, \boldsymbol{Y}_1) \leq [g(\boldsymbol{W}, \boldsymbol{Y}_2)]_\alpha\big) = \alpha. \quad (107)$$

Proof.

$$\begin{aligned} P\big(g(\boldsymbol{W}, \boldsymbol{Y}_1) \leq [g(\boldsymbol{W}, \boldsymbol{Y}_2)]_\alpha\big) &= E_{\boldsymbol{W}}\Big(E\Big(P\big(g(\boldsymbol{W}, \boldsymbol{Y}_1) \leq [g(\boldsymbol{W}, \boldsymbol{Y}_2)]_\alpha\big) \mid \boldsymbol{W}\big)\big) \\ &= E_{\boldsymbol{W}}\big(\alpha\big) \\ &= \alpha. \end{aligned} \quad (108)$$

$\square$

Proposition 2.5 *The one-at-a-time reference prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ is exact frequentist matching prior for ξ_h, $h = 1, \ldots, p$.*

Proof. Under the one-at-a-time reference prior, $a_h = 1$, $b_h(= n_1 + n_2 - h - a_1) = c_h(= n_1 + n_2 - h - 1)$. Given $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Xi}, \boldsymbol{\Delta})$, use part (c) of Lemma 2.6, the frequentist coverage of $[\xi_1^*]_\alpha$ is

$$\begin{aligned}
& P(\xi_1 < [\xi_1^*]_\alpha | \mu_{11}, \mu_{21}, \psi_{11}) \\
= \;& P\left(\xi_1 < \left[\chi_{b_1}^* \frac{1}{\psi_{11} v_{11}} \psi_{11}(\overline{x}_1 - \overline{y}_1) + c_0 z_{41}^*\right]_\alpha\right) \\
= \;& P\left(0 < \left[(\frac{\chi_{c_1}^*}{\chi_{c_1}} - 1)\xi_1 + c_0 \left(\frac{\chi_{c_1}^*}{\chi_{c_1}} z_{41} - z_{41}^*\right)\right]_\alpha\right) \\
= \;& P\left(-\frac{\xi_1}{\chi_{c_1}} - \frac{z_{41}}{\chi_{c_1}} < \left[-\frac{\xi_1}{\chi_{c_1}^*} - \frac{z_{41}^*}{\chi_{c_1}^*}\right]_\alpha\right) \\
= \;& \alpha.
\end{aligned} \tag{109}$$

For $h = 2, \ldots, p$, let $\boldsymbol{\mu}_{i,h}$ be the first h elements of $\boldsymbol{\mu}_i$ and denote

$$\boldsymbol{\zeta}_{h-1} = \boldsymbol{W}_{h-1}^{-1} \left(\boldsymbol{\Xi}_{h-1} \boldsymbol{\Delta}_{h-1}(\boldsymbol{\mu}_{1,h-1} - \boldsymbol{\mu}_{2,h-1}) + c_0 \boldsymbol{z}_{4,h-1}\right),$$

which is a random vector independent of $\boldsymbol{w}_{h,h-1}$, $\boldsymbol{z}_{h,h-1}^* \chi_{b_h}^{2*}$, and $\chi_{c_h}^2$. We have

$$\begin{aligned}
& P(\xi_h < [\xi_h^*]_\alpha | \boldsymbol{\mu}_{1,h}, \boldsymbol{\mu}_{2,h}, \boldsymbol{\Xi}_h, \boldsymbol{\Delta}_h) \\
= \;& P\left(\xi_h < \left[\begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}' \boldsymbol{W}_h^{-1} (\boldsymbol{\Xi}_h \boldsymbol{\Delta}_h)(\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{y}}_h) + c_0 z_{4h}^*\right]_\alpha\right) \\
= \;& P\left(\xi_h < \left[\begin{pmatrix} \boldsymbol{z}_{h,h-1}^* \\ \chi_{b_h}^* \end{pmatrix}' \boldsymbol{W}_h^{-1} \left(\boldsymbol{\Xi}_h \boldsymbol{\Delta}_h (\boldsymbol{\mu}_{1,h} - \boldsymbol{\mu}_{2,h}) + c_0 \boldsymbol{z}_{4,h}\right) + c_0 z_{4h}^*\right]_\alpha\right) \\
= \;& P\left(\xi_h < \left[(\boldsymbol{z}_{h,h-1}'^* - \frac{\chi_{b_h}^*}{w_{hh}} \boldsymbol{w}_{h,h-1}')\boldsymbol{\zeta}_h + \frac{\chi_{b_h}^*}{w_{hh}} \xi_h + \frac{\chi_{b_h}^*}{w_{hh}} c_0 (z_{4h} - z_{4h}^*)\right]_\alpha\right) \\
= \;& P\left(0 < \left[(\boldsymbol{z}_{h,h-1}'^* - \frac{\chi_{b_h}^*}{\chi_{c_h}} \boldsymbol{w}_{h,h-1}')\boldsymbol{\zeta}_h + (\frac{\chi_{b_h}^*}{\chi_{c_h}} - 1)\xi_h + \frac{\chi_{b_h}^*}{\chi_{c_h}} c_0 (z_{4h} - z_{4h}^*)\right]_\alpha\right).
\end{aligned} \tag{110}$$

Use part (c) of Lemma 2.6, (110) is further reduced as

$$\begin{aligned}
&= P\Bigg(0 < \Bigg[(\frac{1}{\chi^*_{b_h}}\boldsymbol{z}'^*_{h,h-1} - \frac{1}{\chi_{c_h}}\boldsymbol{w}'_{h,h-1})\boldsymbol{\zeta}_{h-1} \\
&\quad +(\frac{1}{\chi_{c_h}} - \frac{1}{\chi^*_{b_h}})\xi_h + c_0(\frac{z_{4h}}{\chi_{c_h}} - \frac{z^*_{4h}}{\chi^*_{b_h}})\Bigg]_\alpha\Bigg) \\
&= P\Bigg(\frac{1}{\chi_{c_h}}\left(\boldsymbol{w}'_{h,h-1}\boldsymbol{\zeta}_{h-1} - \xi_h - c_0 z_{4h}\right) < \left[\frac{1}{\chi^*_{c_h}}\left(\boldsymbol{z}'^*_{h,h-1}\boldsymbol{\zeta}_{h-1} - \xi_h - c_0 z^*_{4h}\right)\right]_\alpha\Bigg) \\
&= \alpha. \qquad (111)
\end{aligned}$$

The Monte Carlo simulation in Session 2.10.3 is consistent with this justification. □

In this study, we are more interested in the KL divergence K. We want to check the frequentist matching properties of the K^*. Under the class of priors (60), follow the procedures introduced above, we are interested in

$$P(K < [K^*]_\alpha | \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}). \qquad (112)$$

Proposition 2.6 *None of the priors in the class of priors given in (60) is an exact frequentist matching prior for K if $K = 0$.*

Proof. Let $\boldsymbol{W}^*\boldsymbol{W}^{-1} = \boldsymbol{W}^\#$. From (88)

$$\begin{aligned}
& P(K < [K^*]_\alpha | \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}) \\
= \;& P\left(\frac{1}{2}\boldsymbol{\xi}'\boldsymbol{\xi} < \left[\frac{1}{2}\boldsymbol{\xi}^{*'}\boldsymbol{\xi}^*\right]_\alpha \mid \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta}\right) \\
= \;& P\left(\boldsymbol{\xi}'\boldsymbol{\xi} < \left[\left(\boldsymbol{W}^*\boldsymbol{V}^{-1}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p}\right)'\left(\boldsymbol{W}^*\boldsymbol{V}^{-1}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p}\right)\right]_\alpha\right) \\
= \;& P\left(\boldsymbol{\xi}'\boldsymbol{\xi} < \left[\left(\boldsymbol{W}^*(\Xi\boldsymbol{\Delta}\boldsymbol{V})^{-1}\Xi\boldsymbol{\Delta}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p}\right)'\left(\boldsymbol{W}^*(\Xi\boldsymbol{\Delta}\boldsymbol{V})^{-1}\Xi\boldsymbol{\Delta}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p}\right)\right]_\alpha\right) \\
= \;& P\left(\boldsymbol{\xi}'\boldsymbol{\xi} < \left[\left(\boldsymbol{W}^*\boldsymbol{W}^{-1}(\boldsymbol{\xi} + c_0\boldsymbol{z}_{4,p}) + c_0\boldsymbol{z}^*_{4,p}\right)'\left(\boldsymbol{W}^*\boldsymbol{W}^{-1}(\boldsymbol{\xi} + c_0\boldsymbol{z}_{4,p}) + c_0\boldsymbol{z}^*_{4,p}\right)\right]_\alpha\right). \qquad (113)
\end{aligned}$$

when $\boldsymbol{\xi} = \boldsymbol{0}$, that is $K = 0$, (113) becomes

$$P\left(0 < \left[c_0^2(\boldsymbol{W}^\#\boldsymbol{z}_{4,p} + \boldsymbol{z}^*_{4,p})'(\boldsymbol{W}^\#\boldsymbol{z}_{4,p} + \boldsymbol{z}^*_{4,p})\right]_\alpha\right) = 1. \qquad (114)$$

Therefore, there is not a prior within the class of priors (60) that offers exact frequentist coverage for $[K^*]_\alpha$ when $K = 0$. $\square$

When $K = 0$, the two population being considered are the same. We are expecting that, there is no a prior within the class of priors given in (60) that is exact matching prior for any value of K. In Subsection 2.10.3, we check the frequentist coverage of some values of K under various priors through simulation.

2.10.3 Numerical Examples through Monte Carlo Simulations

One of the advantages of expressing posteriors as constructive random posteriors is that it offers a powerful tool for verifying the frequentist coverage of the posteriors obtained through Bayesian procedures. Another important advantage is that direct Monte Carlo simulation of the posteriors becomes straightforward. Given Proposition 2.3, we can run computer simulations to investigate the frequentist coverage of the K^* or ξ_h^* for specific values of $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ and given sample sizes $(n_1,\, n_2)$. For simplification of simulation, we need to choose a concise form of $\boldsymbol{\xi}^*$ by reorganizing as below:

$$\begin{aligned}
\boldsymbol{\xi}^* &= \boldsymbol{W}^*\boldsymbol{V}^{-1}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p} \\
&= \boldsymbol{W}^*(\Xi\boldsymbol{\Delta}\boldsymbol{V})^{-1}\Xi\boldsymbol{\Delta}(\overline{\boldsymbol{x}} - \overline{\boldsymbol{y}}) + c_0\boldsymbol{z}^*_{4,p} \\
&= \boldsymbol{W}^*\boldsymbol{W}^{-1}(\Xi\boldsymbol{\Delta}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + c_0\boldsymbol{z}_{4,p}) + c_0\boldsymbol{z}^*_{4,p} \\
&= \boldsymbol{W}^*\boldsymbol{W}^{-1}(\boldsymbol{\xi} + c_0\boldsymbol{z}_{4,p}) + c_0\boldsymbol{z}^*_{4,p}, \qquad (115)\\
K^* &= \frac{1}{2}\|\boldsymbol{\xi}^*\|, \qquad (116)
\end{aligned}$$

which clearly shows that $\boldsymbol{\xi}^*$ depends on $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$ only through $\boldsymbol{\xi}$. So is K^*. The Monte Carlo simulation procedure is as follow:

Step 0. Choose the parameter set $\boldsymbol{\xi}$ of interest and specify the sample size $(n_1,\, n_2)$.

Step 1. Simulate a sample of $\boldsymbol{W}$ and $\boldsymbol{z}_{4,p}$.

Step 2. For each sample in Step 1, simulate m_2 values of $\boldsymbol{W}^*$ and $\boldsymbol{z}^*_{4,p}$ and plug them into K^* (or ξ^*_h). Sort K^* (or ξ^*_h) from the smallest to the largest.

Step 3. Record a 1 if the true value of K (or ξ_h) is less than K^*_α(or $\xi^*_{h\alpha}$).

Step 4. Repeat Step 1-3 for m_1 times. The total number of 1's recorded divided by m_1 is the corresponding frequentist coverage for the α credible interval.

If the prior used is exact frequentist matching prior, the simulated frequentist coverage will be very close to α, the difference is from sampling error with variance $\frac{\alpha(1-\alpha)}{m_1}$. So the difference will decrease when the sample size increases, and eventually goes to zero. One of the drawbacks of investigating frequentist coverage through Monte Carlo simulation is that one simulation can address only a specific set of parameters. With continuous parameters that have an infinite number of parameter sets, it is impossible to numerically verify frequentist coverage for all of them. However some systematically chosen parameters sets might build up a frequentist coverage image clear enough for practical use.

For fast simulation, we program the simulation with FORTRAN. We run a set of pioneering simulations. The pioneering simulations are carried out with two goals: (i) to show that an exact matching prior for parameter A may result in poor frequentist coverage for parameter B; (ii) to further explore the frequentist coverage of credible intervals of $(K, \boldsymbol{\xi})$ under non-exact matching priors with the hope of finding a prior that offers "good" frequentist coverage for K.

The simulations evaluate the frequentist coverage of $(K, \boldsymbol{\xi})$ for:

(1) two different dimensions: $p = 3$ and $p = 5$;

(2) two parameter values: $\xi_h = h$ and $\xi_h = 2h$, $h = 1, \ldots, p$;

(3) two non-informative priors: one-at-a-time reference prior and Jeffreys' prior;

(4) various sample sizes: $n_1 = n_2 = n$, $n = 5, 10, 20, 50, 100, 200$.

The simulation results are reported in Tables 1 - 8. Since K is of our prime interest for this study, we summarize all simulation outcomes for K in Table 9 for easy comparison. Here is a list of findings from the simulations:

(1) One-at-a-time reference prior is the exact matching prior for all elements of $\boldsymbol{\xi}$.

(2) Under Jeffreys' prior, the frequentist coverage of ξ_h decreases as h increases, with ξ_h in the middle position among ξ_hs performing the best (ξ_2 and ξ_3 in the $p = 3$ and $p = 5$ cases, respectively).

(3) Under Jeffreys' prior, frequentist coverage of K is very close to the nominal α even in relatively small sample size, and biases are mostly to the bigger side of α; Thus Jeffreys' prior can be a good non-informative prior candidate for K.

(4) Increasing sample size generally improves frequentist coverage.

Table 1: Frequentist Coverage for Credible of $(K, \boldsymbol{\xi})$ with One-at-a-time Reference Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$; $p = 3$, $\xi_1 = 1, \xi_2 = 3, \xi_3 = 3$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
5, 5; 384s	K	.0403	.0828	.1481	.7435	.9906	.9958	.9982
	ξ_1	.0085	.0219	.0489	.4987	.9440	.9702	.9892
	ξ_2	.0102	.0254	.0503	.5016	.9485	.9735	.9899
	ξ_3	.0088	.0243	.0503	.5004	.9510	.9743	.9905
10, 10; 546s	K	.0258	.0601	.1096	.6816	.9821	.9923	.9970
	ξ_1	.0099	.0257	.0508	.5071	.9508	.9749	.9900
	ξ_2	.0092	.0245	.0508	.4969	.9511	.9758	.9916
	ξ_3	.0092	.0226	.0481	.4996	.9493	.9746	.9912
20, 20; 615s	K	.0204	.0459	.0861	.6242	.9790	.9904	.9964
	ξ_1	.0091	.0237	.0481	.4944	.9501	.9756	.9905
	ξ_2	.0092	.0225	.0468	.4998	.9527	.9750	.9897
	ξ_3	.0102	.0226	.0469	.4989	.9519	.9769	.9909
50, 50; 627s	K	.0179	.0397	.0760	.5779	.9683	.9838	.9940
	ξ_1	.0100	.0250	.0509	.5052	.9498	.9738	.9896
	ξ_2	.0104	.0256	.0513	.5010	.9466	.9722	.9880
	ξ_3	.0106	.0257	.0508	.4941	.9479	.9765	.9894
100, 100; 647s	K	.0140	.0353	.0685	.5656	.9629	.9819	.9940
	ξ_1	.0087	.0251	.0496	.5039	.9493	.9727	.9905
	ξ_2	.0099	.0250	.0510	.4993	.9503	.9742	.9902
	ξ_3	.0098	.0242	.0488	.5094	.9483	.9737	.9888
200, 200; 649s	K	.0117	.0305	.0595	.5431	.9592	.9792	.9927
	ξ_1	.0092	.0224	.0470	.4977	.9488	.9748	.9894
	ξ_2	.0101	.0237	.0482	.5092	.9489	.9745	.9888
	ξ_3	.0088	.0217	.0458	.4985	.9492	.9751	.9894

Table 2: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with Jeffreys' prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$; $p = 3$, $\xi_1 = 1, \xi_2 = 2, \xi_3 = 3$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
5, 5; 388s	K	.0200	.0441	.0838	.6362	.9813	.9915	.9976
	ξ_1	.0283	.0592	.1072	.6035	.9476	.9679	.9839
	ξ_2	.0093	.0249	.0519	.5082	.9520	.9758	.9901
	ξ_3	.0003	.0022	.0062	.2494	.8752	.9303	.9715
10, 10; 544s	K	.0149	.0383	.0754	.6021	.9753	.9894	.9966
	ξ_1	.0207	.0476	.0851	.5744	.9582	.9765	.9892
	ξ_2	.0091	.0221	.0481	.4964	.9509	.9756	.9897
	ξ_3	.0030	.0086	.0209	.3716	.9109	.9527	.9789
20, 20; 611s	K	.0135	.0325	.0638	.5717	.9671	.9848	.9943
	ξ_1	.0156	.0365	.0706	.5571	.9582	.9782	.9912
	ξ_2	.0098	.0268	.0503	.5021	.9497	.9733	.9889
	ξ_3	.0048	.0118	.0256	.4109	.9237	.9615	.9838
50, 50; 627s	K	.0118	.0318	.0613	.5374	.9627	.9814	.9935
	ξ_1	.0129	.0308	.0577	.5377	.9593	.9808	.9908
	ξ_2	.0111	.0256	.0500	.4938	.9489	.9762	.9911
	ξ_3	.0075	.0179	.0351	.4439	.9315	.9669	.9870
100, 100; 650s	K	.0124	.0308	.0597	.5275	.9575	.9798	.9920
	ξ_1	.0109	.0289	.0578	.5241	.9566	.9782	.9904
	ξ_2	.0099	.0251	.0537	.5029	.9530	.9776	.9910
	ξ_3	.0080	.0211	.0421	.4612	.9338	.9654	.9862
200, 200; 647s	K	.0118	.0274	.0541	.5237	.9546	.9785	.9923
	ξ_1	.0106	.0276	.0542	.5210	.9536	.9782	.9915
	ξ_2	.0110	.0264	.0489	.4974	.9498	.9764	.9903
	ξ_3	.0093	.0210	.0439	.4698	.9419	.9712	.9898

Table 3: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with One-at-a-time Reference Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$; $p = 3$, $\xi_1 = 2, \xi_2 = 4, \xi_3 = 6$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
$5, 5$; 392s	K	.0361	.0818	.1458	.7346	.9916	.9962	.9985
	ξ_1	.0114	.0278	.0551	.5049	.9486	.9733	.9897
	ξ_2	.0088	.0257	.0509	.5007	.9505	.9742	.9906
	ξ_3	.0106	.0246	.0497	.5048	.9524	.9776	.9904
$10, 10$; 544s	K	.0275	.0612	.1095	.6673	.9836	.9921	.9972
	ξ_1	.0093	.0237	.0501	.4997	.9514	.9744	.9892
	ξ_2	.0111	.0269	.0532	.4979	.9514	.9778	.9922
	ξ_3	.0107	.0259	.0508	.5052	.9515	.9760	.9916
$20, 20$; 612s	K	.0224	.0494	.0899	.6188	.9753	.9892	.9957
	ξ_1	.0114	.0264	.0502	.5018	.9466	.9714	.9892
	ξ_2	.0103	.0251	.0493	.4893	.9498	.9752	.9906
	ξ_3	.0125	.0261	.0528	.5061	.9531	.9757	.9917
$50, 50$; 621s	K	.0151	.0378	.0733	.5629	.9663	.9827	.9951
	ξ_1	.0106	.0255	.0490	.4968	.9508	.9778	.9914
	ξ_2	.0080	.0210	.0455	.4892	.9502	.9756	.9908
	ξ_3	.0099	.0226	.0495	.5015	.9505	.9739	.9894
$100, 100$; 644s	K	.0130	.0348	.0684	.5551	.9619	.9832	.9946
	ξ_1	.0099	.0238	.0512	.5076	.9446	.9727	.9899
	ξ_2	.0096	.0247	.0490	.5009	.9505	.9754	.9914
	ξ_3	.0104	.0281	.0527	.5076	.9507	.9745	.9906
$200, 200$; 645s	K	.0110	.0293	.0570	.5336	.9582	.9797	.9918
	ξ_1	.0109	.0265	.0486	.5009	.9513	.9761	.9910
	ξ_2	.0101	.0246	.0490	.5008	.9505	.9740	.9912
	ξ_3	.0091	.0236	.0483	.5009	.9505	.9752	.9897

Table 4: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with Jeffreys' Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Xi}, \boldsymbol{\Delta})$; $p = 3$, $\xi_1 = 2, \xi_2 = 4, \xi_3 = 6$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
5, 5; 398s	K	.0184	.0414	.0749	.5941	.9779	.9898	.9971
	ξ_1	.0383	.0725	.1239	.6450	.9700	.9847	.9934
	ξ_2	.0104	.0275	.0504	.4929	.9508	.9765	.9899
	ξ_3	.0007	.0015	.0054	.2444	.8651	.9232	.9660
10, 10; 541s	K	.0125	.0322	.0678	.5760	.9730	.9868	.9946
	ξ_1	.0221	.0479	.0887	.6051	.9663	.9841	.9940
	ξ_2	.0103	.0242	.0477	.5025	.9518	.9752	.9900
	ξ_3	.0035	.0090	.0216	.3574	.9093	.9525	.9787
20, 20; 608s	K	.0101	.0297	.0604	.5500	.9627	.9831	.9940
	ξ_1	.0158	.0384	.0730	.5676	.9643	.9827	.9934
	ξ_2	.0097	.0255	.0546	.4973	.9524	.9758	.9908
	ξ_3	.0037	.0121	.0259	.4022	.9219	.9599	.9827
50, 50; 616s	K	.0115	.0277	.0555	.5194	.9589	.9811	.9935
	ξ_1	.0145	.0313	.0639	.5371	.9604	.9830	.9933
	ξ_2	.0094	.0243	.0483	.4985	.9516	.9770	.9901
	ξ_3	.0064	.0177	.0372	.4351	.9331	.9646	.9849
100, 100; 642s	K	.0107	.0282	.0571	.5177	.9533	.9764	.9894
	ξ_1	.0128	.0295	.0599	.5323	.9601	.9798	.9915
	ξ_2	.0105	.0261	.0508	.5005	.9501	.9771	.9918
	ξ_3	.0071	.0192	.0402	.4619	.9356	.9662	.9853
200, 200; 642s	K	.0107	.0279	.0526	.5187	.9500	.9757	.9911
	ξ_1	.0118	.0279	.0571	.5224	.9564	.9782	.9913
	ξ_2	.0109	.0284	.0551	.5005	.9479	.9745	.9897
	ξ_3	.0077	.0202	.0434	.4731	.9426	.9704	.9873

Table 5: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with One-at-a-time Reference Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Xi}, \boldsymbol{\Delta})$; $p = 5$, $\xi_1 = 1, \xi_2 = 2, \ldots, \xi_5 = 5$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
10, 10; 921s	K	.0503	.1045	.1784	.7798	.9933	.9974	.9991
	ξ_1	.0101	.0245	.0505	.4993	.9506	.9732	.9890
	ξ_2	.0119	.0268	.0518	.5009	.9474	.9727	.9895
	ξ_3	.0099	.0242	.0488	.5056	.9498	.9742	.9898
	ξ_4	.0115	.0241	.0500	.4934	.9493	.9737	.9881
	ξ_5	.0117	.0270	.0551	.5070	.9534	.9779	.9908
20, 20; 1041s	K	.0308	.0725	.1254	.6977	.9866	.9940	.9982
	ξ_1	.0088	.0239	.0520	.4998	.9531	.9773	.9913
	ξ_2	.0098	.0227	.0477	.4998	.9537	.9781	.9909
	ξ_3	.0107	.0233	.0474	.4975	.9520	.9755	.9909
	ξ_4	.0089	.0235	.0500	.5091	.9516	.9756	.9895
	ξ_5	.0097	.0238	.0517	.4973	.9515	.9772	.9898
50, 50; 1063s	K	.0213	.0497	.0917	.6334	.9789	.9911	.9971
	ξ_1	.0105	.0237	.0481	.5041	.9489	.9735	.9887
	ξ_2	.0102	.0257	.0494	.5066	.9503	.9753	.9904
	ξ_3	.0099	.0256	.0512	.4975	.9510	.9768	.9920
	ξ_4	.0095	.0249	.0496	.5006	.9499	.9736	.9900
	ξ_5	.0105	.0262	.0513	.5050	.9500	.9757	.9897
100, 100; 1094s	K	.0174	.0387	.0728	.5891	.9700	.9870	.9948
	ξ_1	.0091	.0228	.0463	.5057	.9527	.9751	.9900
	ξ_2	.0105	.0249	.0487	.5019	.9492	.9733	.9901
	ξ_3	.0102	.0264	.0501	.5075	.9477	.9743	.9890
	ξ_4	.0088	.0236	.0495	.5006	.9484	.9739	.9895
	ξ_5	.0093	.0220	.0470	.4969	.9480	.9749	.9906
200, 200; 1091s	K	.0170	.0350	.0656	.5603	.9645	.9820	.9932
	ξ_1	.0101	.0235	.0478	.5000	.9470	.9747	.9890
	ξ_2	.0108	.0265	.0486	.4953	.9464	.9756	.9913
	ξ_3	.0094	.0238	.0484	.4990	.9492	.9760	.9895
	ξ_4	.0113	.0248	.0540	.4971	.9472	.9751	.9900
	ξ_5	.0109	.0252	.0507	.4921	.9497	.9755	.9882

Table 6: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with Jeffreys' Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$; $p = 5$, $\xi_1 = 1, \xi_2 = 2 \ldots, \xi_5 = 5$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
10, 10; 895s	K	.0100	.0275	.0557	.5570	.9720	.9875	.9961
	ξ_1	.0406	.0768	.1261	.6337	.9600	.9764	.9895
	ξ_2	.0254	.0516	.0952	.6251	.9714	.9861	.9958
	ξ_3	.0092	.0250	.0498	.5021	.9491	.9756	.9898
	ξ_4	.0017	.0075	.0186	.3560	.9047	.9488	.9771
	ξ_5	.0003	.0012	.0041	.1952	.8170	.8929	.9472
20, 20; 1039s	K	.0106	.0312	.0586	.5427	.9651	.9829	.9934
	ξ_1	.0268	.0570	.0975	.6058	.9652	.9796	.9913
	ξ_2	.0162	.0375	.0713	.5703	.9632	.9831	.9939
	ξ_3	.0103	.0252	.0503	.4940	.9512	.9755	.9893
	ξ_4	.0034	.0126	.0264	.4070	.9236	.9586	.9806
	ξ_5	.0025	.0060	.0153	.3149	.8815	.9352	.9680
50, 50; 1054s	K	.0115	.0273	.0549	.5296	.9564	.9798	.9926
	ξ_1	.0192	.0414	.0749	.5652	.9608	.9802	.9910
	ξ_2	.0143	.0355	.0681	.5416	.9608	.9816	.9927
	ξ_3	.0109	.0250	.0489	.4965	.9488	.9742	.9888
	ξ_4	.0081	.0169	.0348	.4457	.9321	.9658	.9850
	ξ_5	.0044	.0121	.0261	.3901	.9126	.9541	.9819
100, 100; 1093s	K	.0110	.0265	.0530	.5264	.9588	.9786	.9918
	ξ_1	.0154	.0344	.0678	.5419	.9611	.9810	.9924
	ξ_2	.0140	.0322	.0624	.5394	.9597	.9798	.9939
	ξ_3	.0089	.0248	.0522	.5022	.9512	.9738	.9898
	ξ_4	.0082	.0183	.0373	.4654	.9428	.9706	.9879
	ξ_5	.0044	.0122	.0299	.4279	.9270	.9609	.9843
200, 200; 1093s	K	.0105	.0255	.0518	.5137	.9559	.9785	.9923
	ξ_1	.0119	.0314	.0595	.5402	.9595	.9816	.9936
	ξ_2	.0112	.0283	.0576	.5182	.9564	.9787	.9908
	ξ_3	.0102	.0247	.0484	.5085	.9527	.9778	.9911
	ξ_4	.0083	.0203	.0437	.4795	.9404	.9696	.9868
	ξ_5	.0078	.0167	.0349	.4470	.9367	.9658	.9861

Table 7: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with One-at-a-time Reference Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Xi, \boldsymbol{\Delta})$; $p = 5$, $\xi_1 = 2, \xi_2 = 4, \ldots, \xi_5 = 10$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
10, 10; 900s	K	.0452	.0928	.1582	.7661	.9909	.9961	.9988
	ξ_1	.0097	.0251	.0505	.4977	.9486	.9745	.9903
	ξ_2	.0092	.0256	.0526	.5060	.9517	.9785	.9915
	ξ_3	.0104	.0253	.0537	.5003	.9504	.9759	.9893
	ξ_4	.0086	.0225	.0457	.5003	.9538	.9772	.9902
	ξ_5	.0102	.0257	.0492	.4986	.9481	.9747	.9893
20, 20; 1028s	K	.0295	.0630	.1136	.6916	.9851	.9936	.9984
	ξ_1	.0098	.0254	.0482	.5041	.9491	.9738	.9896
	ξ_2	.0103	.0253	.0515	.5005	.9501	.9745	.9888
	ξ_3	.0107	.0271	.0522	.4976	.9526	.9758	.9904
	ξ_4	.0101	.0243	.0490	.4934	.9493	.9736	.9888
	ξ_5	.0091	.0230	.0465	.4983	.9504	.9774	.9901
50, 50; 1041s	K	.0224	.0520	.0899	.6199	.9761	.9898	.9963
	ξ_1	.0115	.0266	.0494	.4966	.9543	.9788	.9921
	ξ_2	.0096	.0220	.0471	.5004	.9500	.9741	.9909
	ξ_3	.0097	.0254	.0499	.5004	.9510	.9771	.9919
	ξ_4	.0097	.0268	.0514	.4971	.9504	.9741	.9891
	ξ_5	.0098	.0242	.0501	.5041	.9502	.9737	.9880
100, 100; 1082s	K	.0175	.0414	.0763	.5874	.9659	.9838	.9937
	ξ_1	.0094	.0235	.0496	.5035	.9520	.9765	.9905
	ξ_2	.0095	.0242	.0500	.4997	.9492	.9756	.9910
	ξ_3	.0093	.0261	.0531	.5080	.9507	.9753	.9898
	ξ_4	.0091	.0233	.0478	.4962	.9490	.9774	.9921
	ξ_5	.0116	.0270	.0524	.4927	.9437	.9723	.9887
200, 200; 1084s	K	.0139	.0356	.0675	.5634	.9661	.9855	.9945
	ξ_1	.0106	.0266	.0533	.4982	.9498	.9754	.9900
	ξ_2	.0103	.0259	.0517	.5040	.9543	.9783	.9904
	ξ_3	.0101	.0249	.0487	.5083	.9512	.9752	.9904
	ξ_4	.0100	.0250	.0518	.4917	.9515	.9770	.9902
	ξ_5	.0101	.0234	.0462	.4948	.9518	.9757	.9903

Table 8: Frequentist Coverage for Credible Interval of $(K, \boldsymbol{\xi})$ with Jeffreys' Prior for $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Xi}, \boldsymbol{\Delta})$; $p = 5$, $\xi_1 = 2, \xi_2 = 4, \ldots, \xi_5 = 10$; $m_1 = m_2 = 10,000$.

(n_1,n_2; time)	α	.01	.025	.05	.50	.95	.975	.99
10, 10; 892s	K	.0093	.0226	.0477	.5358	.9718	.9868	.9968
	ξ_1	.0466	.0909	.1496	.7000	.9823	.9921	.9967
	ξ_2	.0260	.0568	.1005	.6226	.9738	.9871	.9955
	ξ_3	.0107	.0246	.0504	.4965	.9492	.9743	.9880
	ξ_4	.0031	.0085	.0177	.3428	.9072	.9493	.9777
	ξ_5	.0004	.0015	.0042	.2007	.8205	.8983	.9502
20, 20; 1030s	K	.0091	.0250	.0513	.5242	.9604	.9797	.9930
	ξ_1	.0284	.0594	.1064	.6436	.9750	.9896	.9957
	ξ_2	.0182	.0391	.0794	.5839	.9656	.9831	.9936
	ξ_3	.0111	.0268	.0518	.4999	.9470	.9732	.9883
	ξ_4	.0041	.0112	.0272	.3999	.9174	.9572	.9814
	ξ_5	.0009	.0047	.0129	.3027	.8766	.9305	.9672
50, 50; 1043s	K	.0091	.0249	.0515	.5175	.9574	.9774	.9920
	ξ_1	.0201	.0443	.0832	.5958	.9676	.9847	.9934
	ξ_2	.0138	.0330	.0679	.5560	.9622	.9827	.9930
	ξ_3	.0112	.0254	.0487	.4995	.9485	.9718	.9894
	ξ_4	.0056	.0154	.0350	.4480	.9329	.9652	.9862
	ξ_5	.0027	.0098	.0245	.3862	.9109	.9529	.9779
100, 100; 1084s	K	.0135	.0283	.0542	.5190	.9582	.9764	.9896
	ξ_1	.0153	.0350	.0682	.5575	.9638	.9818	.9933
	ξ_2	.0123	.0320	.0612	.5452	.9615	.9800	.9930
	ξ_3	.0100	.0250	.0514	.5013	.9486	.9729	.9901
	ξ_4	.0082	.0213	.0409	.4646	.9399	.9672	.9853
	ξ_5	.0058	.0162	.0324	.4298	.9260	.9589	.9820
200, 200; 1085s	K	.0095	.0272	.0505	.5046	.9555	.9783	.9918
	ξ_1	.0122	.0324	.0629	.5503	.9593	.9792	.9925
	ξ_2	.0121	.0290	.0576	.5197	.9557	.9775	.9917
	ξ_3	.0078	.0230	.0485	.4928	.9476	.9717	.9892
	ξ_4	.0081	.0202	.0410	.4671	.9418	.9726	.9889
	ξ_5	.0072	.0203	.0390	.4463	.9339	.9636	.9861

Table 9: Summary of Frequentist Coverage for Credible Interval of K
p-dimensional; (J)ffreys'r prior or (R)ference prior; $n_1 = n_2 = n$; $m_1 = m_2 = 10,000$.

p, ξ_h, prior	$n \setminus \alpha$	.01	.025	.05	.50	.95	.975	.99
3, 1h, R	5	.0403	.0828	.1481	.7435	.9906	.9958	.9982
3, 1h, R	10	.0258	.0601	.1096	.6816	.9821	.9923	.9970
3, 1h, R	20	.0204	.0459	.0861	.6242	.9790	.9904	.9964
3, 1h, R	50	.0179	.0397	.0760	.5779	.9683	.9838	.9940
3, 1h, R	100	.0140	.0353	.0685	.5656	.9629	.9819	.9940
3, 1h, R	200	.0117	.0305	.0595	.5431	.9592	.9792	.9927
3, 1h, J	5	.0200	.0441	.0838	.6362	.9813	.9915	.9976
3, 1h, J	10	.0149	.0383	.0754	.6021	.9753	.9894	.9966
3, 1h, J	20	.0135	.0325	.0638	.5717	.9671	.9848	.9943
3, 1h, J	50	.0118	.0318	.0613	.5374	.9627	.9814	.9935
3, 1h, J	100	.0124	.0308	.0597	.5275	.9575	.9798	.9920
3, 1h, J	200	.0118	.0274	.0541	.5237	.9546	.9785	.9923
3, 2h, R	5	.0361	.0818	.1458	.7346	.9916	.9962	.9985
3, 2h, R	10	.0275	.0612	.1095	.6673	.9836	.9921	.9972
3, 2h, R	20	.0224	.0494	.0899	.6188	.9753	.9892	.9957
3, 2h, R	50	.0151	.0378	.0733	.5629	.9663	.9827	.9951
3, 2h, R	100	.0130	.0348	.0684	.5551	.9619	.9832	.9946
3, 2h, R	200	.0110	.0293	.0570	.5336	.9582	.9797	.9918
3, 2h, J	5	.0184	.0414	.0749	.5941	.9779	.9898	.9971
3, 2h, J	10	.0125	.0322	.0678	.5760	.9730	.9868	.9946
3, 2h, J	20	.0101	.0297	.0604	.5500	.9627	.9831	.9940
3, 2h, J	50	.0115	.0277	.0555	.5194	.9589	.9811	.9935
3, 2h, J	100	.0107	.0282	.0571	.5177	.9533	.9764	.9894
3, 2h, J	200	.0107	.0279	.0526	.5187	.9500	.9757	.9911
5, 1h, R	10	.0503	.1045	.1784	.7798	.9933	.9974	.9991
5, 1h, R	20	.0308	.0725	.1254	.6977	.9866	.9940	.9982
5, 1h, R	50	.0213	.0497	.0917	.6334	.9789	.9911	.9971
5, 1h, R	100	.0174	.0387	.0728	.5891	.9700	.9870	.9948
5, 1h, R	200	.0170	.0350	.0656	.5603	.9645	.9820	.9932
5, 1h, J	10	.0100	.0275	.0557	.5570	.9720	.9875	.9961
5, 1h, J	20	.0106	.0312	.0586	.5427	.9651	.9829	.9934
5, 1h, J	50	.0115	.0273	.0549	.5296	.9564	.9798	.9926
5, 1h, J	100	.0110	.0265	.0530	.5264	.9588	.9786	.9918
5, 1h, J	200	.0105	.0255	.0518	.5137	.9559	.9785	.9923
5, 2h, R	10	.0452	.0928	.1582	.7661	.9909	.9961	.9988
5, 2h, R	20	.0295	.0630	.1136	.6916	.9851	.9936	.9984
5, 2h, R	50	.0224	.0520	.0899	.6199	.9761	.9898	.9963
5, 2h, R	100	.0175	.0414	.0763	.5874	.9659	.9838	.9937
5, 2h, R	200	.0139	.0356	.0675	.5634	.9661	.9855	.9945
5, 2h, J	10	.0093	.0226	.0477	.5358	.9718	.9868	.9968
5, 2h, J	20	.0091	.0250	.0513	.5242	.9604	.9797	.9930
5, 2h, J	50	.0091	.0249	.0515	.5175	.9574	.9774	.9920
5, 2h, J	100	.0135	.0283	.0542	.5190	.9582	.9764	.9896
5, 2h, J	200	.0095	.0272	.0505	.5046	.9555	.9783	.9918

Chapter 3 Objective Bayesian Analysis of Star-Shape Gaussian Graphical Model

3.11 Introduction

3.11.1 Problem Statement and Notation

As stated in Whittaker (1990), exploring the inter-relationships between several variables is central to applied multivariate statistical analysis. An important objective of graphical modeling is to describe and explain these relationships through conditioning on and controlling for some other variables. The multivariate normal distribution plays an important role in multivariate statistical analysis. A large literature exists on the problem of estimating the covariance matrix $\mathbf{\Sigma}$ and precision matrix $\mathbf{\Omega}$ in saturated multivariate normal distribution with a positive definite matrix $\mathbf{\Sigma}$. See, for example, Haff (1980), Sinha & Ghosh (1987), Krishnamoorthy & Gupta (1989), Yang & Berger (1994), and others. The problem is that as the number of variables, p, increases, the number of parameters to be estimated, $p(p+1)/2$, increases faster. If the sample size n is smaller then $p(p+1)/2$, the estimate will lack good statistical properties, and may not exist. Furthermore, it could be hard to interpret models with a large number of parameters as well.

In practice, it is typical that some of the variables share common characteristics and some are loosely dependent. Recently, Sun & Sun (2005) considered the admissibility and minimaxity of the maximum likelihood estimates and Bayesian estimates (under objective priors) of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ for the star-shape model under various loss functions. The key feature of the star-shape model is that variables are grouped into several subgroups of variables, and conditional on one specific group of variables, which serves as the global variables, the other groups are independent of each other. For the multivariate normal distribution case, this is equivalent to having zeros at certain off-diagonal elements in $\mathbf{\Omega}$. In particular, assume that $\boldsymbol{x} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$. Vector $\boldsymbol{x}$ is partitioned into $k+1$ sub-groups, that is, $\boldsymbol{x} = (\boldsymbol{x}_0', \boldsymbol{x}_1', \ldots, \boldsymbol{x}_k')'$, where $\boldsymbol{x}_i'$ is p_i-dimensional and $\sum_{i=0}^k p_i = p$. We assume that, given $\boldsymbol{x}_0$, the other subvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are mutually conditionally independent. From Whittaker (1990) and Lauritzen (1996), the precision matrix $\mathbf{\Omega} = \mathbf{\Sigma}^{-1}$ has the following special structure,

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Omega}_0 & \mathbf{\Omega}_{01} & \mathbf{\Omega}_{02} & \cdots & \mathbf{\Omega}_{0k} \\ \mathbf{\Omega}_{10} & \mathbf{\Omega}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Omega}_{20} & \mathbf{0} & \mathbf{\Omega}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Omega}_{k0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Omega}_k \end{pmatrix}. \tag{117}$$

We give diagrams for a few examples when $k = 2, 3$, and 6 in Figure 3.11.1. Figure 3.11.1 shows three star-shape examples (a), (b) and (c), with 2, 3 and 6 sets of local variables respectively. More examples can be found in Whittaker (1990).

Sun & Sun (2005) considered a Cholesky decomposition $\mathbf{\Psi}$ of $\mathbf{\Omega}$ and showed that the MLE of $\mathbf{\Psi}$ is inadmissible and can be dominated by the Bayesian estimate under the right-Haar prior under various loss functions. Note that the admissibility and minimaxity of estimation depend on the choices of a loss function. We also consider alternative parameterization of $\mathbf{\Psi}$, namely, $(\mathbf{\Xi}, \mathbf{\Delta})$, a standard version considered in Pourahmadi (1999).

One of the purposes of this chapter is to consider the properties of Bayesian credible intervals of $\mathbf{\Psi}$ and $\mathbf{\Omega}$.

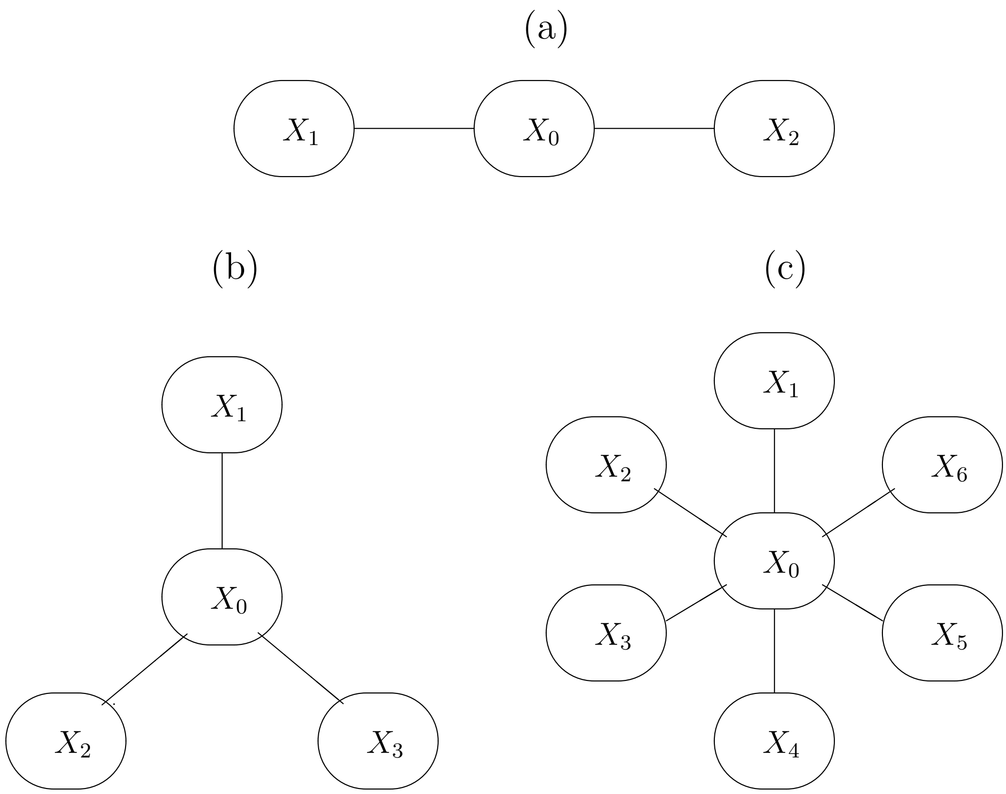

Figure 1: Several graphs of conditional independence: (a) $k = 2$; (b) $k = 3$, and (c) $k = 6$.

In particular, we consider several objective priors of $\mathbf{\Psi}$ including the Jeffreys' prior, Berger and Bernardo's (1992a) reference prior, and the right-Haar prior.

3.11.2 A Cholesky Decomposition of the Precision Matrix

Sun & Sun (2005) employed a kind of Cholesky decomposition of $\mathbf{\Omega}$, in the sense that

$$\mathbf{\Omega} = \mathbf{\Psi}'\mathbf{\Psi}, \tag{118}$$

where $\mathbf{\Psi}$ is given by

$$\mathbf{\Psi} = \begin{pmatrix} \mathbf{\Psi}_0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Psi}_{10} & \mathbf{\Psi}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Psi}_{20} & \mathbf{0} & \mathbf{\Psi}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Psi}_{k0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Psi}_k \end{pmatrix}. \tag{119}$$

Here

$$\mathbf{\Psi}_i = \begin{pmatrix} \psi_{i,11} & 0 & \cdots & 0 \\ \psi_{i,21} & \psi_{i,22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{i,p_i1} & \psi_{i,p_i2} & \cdots & \psi_{i,p_ip_i} \end{pmatrix}_{p_i \times p_i}, \quad i = 0, 1, \ldots, k, \tag{120}$$

$$\mathbf{\Psi}_{i0} = \begin{pmatrix} \psi_{i0,11} & \psi_{i0,12} & \cdots & \psi_{i0,1p_0} \\ \psi_{i0,21} & \psi_{i0,22} & \cdots & \psi_{i0,2p_0} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{i0,p_i1} & \psi_{i0,p_i2} & \cdots & \psi_{i0,p_ip_0} \end{pmatrix}_{p_i \times p_0}, \quad i = 1, \ldots, k. \tag{121}$$

Without loss of generality, we assume that the diagonal elements of $\mathbf{\Psi}$ is positive, i.e., $\psi_{i,jj} > 0$, for $i = 0, 1, \ldots, k,\ j = 1, \ldots, p_i$.

Based on this parameterization, Sun & Sun (2005) derived and compared the maximum likelihood estimator and Bayesian estimators of $\mathbf{\Omega}$ under various loss functions. Some non-informative priors of $\mathbf{\Psi}$ were also derived.

For later use, we also define,

$$
\begin{aligned}
\mathbf{\Psi}_{i,j} &= \begin{pmatrix} \psi_{i,11} & 0 & \cdots & 0 \\ \psi_{i,21} & \psi_{i,22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{i,j1} & \psi_{i,j2} & \cdots & \psi_{i,jj} \end{pmatrix}_{j\times j}, \quad i=0,\ldots,k,\ j=1,\ldots,p_i; \qquad (122)\\
\mathbf{\Psi}_{i0,j} &= \begin{pmatrix} \psi_{i0,11} & \cdots & \psi_{i0,1p_0} \\ \vdots & \ddots & \vdots \\ \psi_{i0,j1} & \cdots & \psi_{i0,jp_0} \end{pmatrix}_{j\times p_0}, \quad i=1,\ldots,k,\ j=1,\ldots,p_i; \qquad (123)\\
\widetilde{\mathbf{\Psi}}_{i,0} &= \mathbf{\Psi}_0, \quad i=1,\ldots,k; \qquad (124)\\
\widetilde{\mathbf{\Psi}}_{i,j} &= \begin{pmatrix} \mathbf{\Psi}_0 & \mathbf{0} \\ \mathbf{\Psi}_{i0,j} & \mathbf{\Psi}_{i,j} \end{pmatrix}, \quad i=1,\ldots,k,\ j=1,\ldots,p_i; \qquad (125)\\
\boldsymbol{\psi}_{0,j} &= (\psi_{0,j1},\ldots,\psi_{0,j(j-1)})', \quad j=2,\ldots,p_0; \qquad (126)\\
\widetilde{\boldsymbol{\psi}}_{i,j} &= (\psi_{i0,j1},\ldots,\psi_{i0,jp_0},\psi_{i,j1},\ldots,\psi_{i,j(j-1)})', \quad i=1,\ldots,k,\ j=1,\ldots,p_i. \qquad (127)
\end{aligned}
$$

3.11.3 Alternative Parameterization to the Cholesky Decomposition

In this chapter, we employ another parameterization over $\mathbf{\Psi}$ by further decomposing $\mathbf{\Psi}$ into a diagonal matrix $\mathbf{\Xi}$ and a unit lower triangle matrix $\mathbf{\Delta}$, that is

$$\mathbf{\Psi} = \mathbf{\Xi}\mathbf{\Delta}, \qquad (128)$$

where

$$\mathbf{\Xi} = diag(\mathbf{\Xi}_0, \mathbf{\Xi}_1, \ldots, \mathbf{\Xi}_k), \tag{129}$$

$$\mathbf{\Xi}_i = diag(\psi_{i,11}, \psi_{i,22}, \ldots, \psi_{i,p_ip_i}), \quad i = 0, 1, \ldots, k;$$

$$\mathbf{\Delta} = \begin{pmatrix} \mathbf{\Delta}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Delta}_{10} & \mathbf{\Delta}_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Delta}_{k0} & \mathbf{0} & \cdots & \mathbf{\Delta}_k \end{pmatrix}, \tag{130}$$

$$\mathbf{\Delta}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \delta_{i,21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i,p_i1} & \delta_{i,p_i2} & \cdots & 1 \end{pmatrix}, \quad i = 0, 1, \ldots, k, \tag{131}$$

$$\mathbf{\Delta}_{i0} = \begin{pmatrix} \delta_{i0,11} & \delta_{i0,12} & \cdots & \delta_{i0,1p_0} \\ \delta_{i0,21} & \delta_{i0,22} & \cdots & \delta_{i0,2p_0} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i0,p_i1} & \delta_{i0,p_i2} & \cdots & \delta_{i0,p_ip_0} \end{pmatrix}, \quad i = 1, \ldots, k. \tag{132}$$

Similar to notations in (122)-(127), we define

$$
\begin{aligned}
\mathbf{\Xi}_{i,j} &= diag(\psi_{i,11}, \psi_{i,22}, \ldots, \psi_{i,jj}), \quad i = 0, 1, \ldots, k, \ j = 1, \ldots, p_i; & (133)\\
\mathbf{\Delta}_{i,j} &= \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \delta_{i,21} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{i,j1} & \delta_{i,j2} & \cdots & 1 \end{pmatrix}_{j\times j}, \quad i = 0, \ldots, k, \ j = 1, \ldots, p_i; & (134)\\
\mathbf{\Delta}_{i0,j} &= \begin{pmatrix} \delta_{i0,11} & \cdots & \delta_{i0,1p_0} \\ \vdots & \ddots & \vdots \\ \delta_{i0,j1} & \cdots & \delta_{i0,jp_0} \end{pmatrix}_{j\times p_0}, \quad i = 1, \ldots, k, \ j = 1, \ldots, p_i; & (135)\\
\widetilde{\mathbf{\Xi}}_{i,0} &= \mathbf{\Xi}_0, \quad i = 1, \ldots, k; & (136)\\
\widetilde{\mathbf{\Xi}}_{i,j} &= diag(\psi_{0,11}, \ldots, \psi_{0,p_0p_0}, \psi_{i,11}, \ldots, \psi_{i,jj}), \quad i = 1, \ldots, k, \ j = 1, \ldots, p_i; & (137)\\
\widetilde{\mathbf{\Delta}}_{i,0} &= \mathbf{\Delta}_0, \quad i = 1, \ldots, k; & (138)\\
\widetilde{\mathbf{\Delta}}_{i,j} &= \begin{pmatrix} \mathbf{\Delta}_0 & \mathbf{0} \\ \mathbf{\Delta}_{i0,j} & \mathbf{\Delta}_{i,j} \end{pmatrix}, \quad i = 1, \ldots, k, \ j = 1, \ldots, p_i; & (139)\\
\boldsymbol{\delta}_{0,j} &= (\delta_{0,j1}, \ldots, \delta_{0,j(j-1)})', \quad j = 2, \ldots, p_0; & (140)\\
\widetilde{\boldsymbol{\delta}}_{i,j} &= (\delta_{i0,j1}, \ldots, \delta_{i0,jp_0}, \delta_{i,j1}, \ldots, \delta_{i,j(j-1)})', \quad i = 1, \ldots, k, \ j = 1, \ldots, p_i. & (141)
\end{aligned}
$$

By Pourahmadi (1999), the $\widetilde{\boldsymbol{\delta}}_{i,j}$ are the negatives of the linear least-squares predictor of x_{ij} based on its predecessors $x_{01}, \ldots, x_{0p_0}$, $x_{i1}, \ldots, x_{i(j-1)}$ and the $\psi^{-2}_{i,jj}$ are the prediction error variances $var(x_{ij} - \hat{x}_{ij})$.

Based on this new parameterization, we will derive non-informative priors such as the Jeffreys' prior and reference prior. We also derive the posteriors of $(\mathbf{\Xi}, \mathbf{\Delta})$ under a class of objective priors. We then give the constructive random posteriors of the parameters. As mentioned in Chapter 1.5, such constructive random posteriors are quite useful in Bayesian computation, and play an important role in proving the frequentist properties of these posteriors.

The rest of this chapter is organized as follows. In Section 3.12, we discuss the relationship between $\boldsymbol{\Psi}$ and $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$, which will allow us to use some results from Sun & Sun (2005) and facilitate our derivations, then we derive various non-informative priors and the Haar measures. In Section 3.13, we derive the posteriors of $\boldsymbol{\Psi}$ and $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ and write them in the form of constructive random posteriors. In Section 3.14, we study the frequentist properties of the constructive random posteriors given in Section 3.13. In Section 3.15, we analyze a real example.

3.12 The Objective Priors

3.12.1 The Relationship Between the Two Parameterizations

This subsection is dedicated to building the elemental level relationships between the parameterizations $\boldsymbol{\Psi}$ and $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ and $\boldsymbol{\Omega}$. It follows from (119) and (128) that

$$\begin{aligned} \boldsymbol{\Psi}_i &= \boldsymbol{\Xi}_i \boldsymbol{\Delta}_i, \quad i = 0, \ldots, k; && (142) \\ \boldsymbol{\Psi}_{i0} &= \boldsymbol{\Xi}_i \boldsymbol{\Delta}_{i0}, \quad i = 1, \ldots, k. && (143) \end{aligned}$$

This is equivalent to componentwise,

$$\begin{aligned} \psi_{i,jl} &= \psi_{i,jj}\delta_{i,jl}, \quad i = 0, \ldots, k,\ j = 2, \ldots, p_i,\ l = 1, \ldots, j-1; && (144) \\ \psi_{i0,jl} &= \psi_{i,jj}\delta_{i0,jl}, \quad i = 1, \ldots, k,\ j = 1, \ldots, p_i,\ l = 1, \ldots, p_0. && (145) \end{aligned}$$

From (117), (119), and (128), we get

$$\begin{aligned} \boldsymbol{\Omega}_0 &= \boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0 + \sum_{i=1}^{k} \boldsymbol{\Psi}_{i0}'\boldsymbol{\Psi}_{i0} = \boldsymbol{\Delta}_0'\boldsymbol{\Xi}_0\boldsymbol{\Delta}_0 + \sum_{i=1}^{k} \boldsymbol{\Delta}_{i0}'\boldsymbol{\Xi}_0\boldsymbol{\Delta}_{i0}, && (146) \\ \boldsymbol{\Omega}_i &= \boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i = \boldsymbol{\Delta}_i'\boldsymbol{\Xi}_i\boldsymbol{\Delta}_i, \quad i = 1, \ldots, k, && (147) \\ \boldsymbol{\Omega}_{i0} &= \boldsymbol{\Psi}_i'\boldsymbol{\Psi}_{i0} = \boldsymbol{\Delta}_i'\boldsymbol{\Xi}_i\boldsymbol{\Delta}_{i0}, \quad i = 1, \ldots, k. && (148) \end{aligned}$$

For later use, we define

$$\begin{aligned} \boldsymbol{\psi} &= (\psi_{0,11}, \boldsymbol{\psi}_{0,2}', \psi_{0,22}, \ldots, \boldsymbol{\psi}_{0,p_0}', \psi_{0,p_0p_0}, \widetilde{\boldsymbol{\psi}}_{1,1}', \psi_{1,11}, \ldots, \widetilde{\boldsymbol{\psi}}_{k,p_k}', \psi_{k,p_kp_k})', && (149) \\ \boldsymbol{\tau} &= (\psi_{0,11}, \boldsymbol{\delta}_{0,2}', \psi_{0,22}, \ldots, \boldsymbol{\delta}_{0,p_0}', \psi_{0,p_0p_0}, \widetilde{\boldsymbol{\delta}}_{1,1}', \psi_{1,11}, \ldots, \widetilde{\boldsymbol{\delta}}_{k,p_k}', \psi_{k,p_kp_k})'. && (150) \end{aligned}$$

Clearly, there are

$$\frac{1}{2}\left[p_0(p_0+1)+\sum_{i=1}^{k}p_i(2p_0+p_i+1)\right]$$

parameters in either $\boldsymbol{\psi}$ or $\boldsymbol{\tau}$. This is the same as the number of parameters in $\boldsymbol{\Psi}$ or $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$.

3.12.2 Objective Priors of the Cholesky Decomposition

Sun & Sun (2005) derived the invariant Haar prior, Jeffreys' prior, and reference prior for $\boldsymbol{\Psi}$. These three types of priors are the most popular non-informative priors. For later use, we include their results as the following proposition.

Proposition 3.7 *Consider the star-shape model with the parameters* $\boldsymbol{\Psi}$.

(a) *The Jeffreys' prior and right-Haar prior for* $\boldsymbol{\psi}$ *are of the same form, given by*

$$\pi_J(\boldsymbol{\psi}) \propto \prod_{j=1}^{p_0}\psi_{0,jj}^{-j}\prod_{i=1}^{k}\prod_{j=1}^{p_i}\psi_{i,jj}^{-(p_0+j)}. \tag{151}$$

(b) *The reference prior of* $\boldsymbol{\psi}$ *with the ordered group* $\{(\psi_{0,11}), (\boldsymbol{\psi}_{0,2}), (\psi_{0,22}), \ldots, (\boldsymbol{\psi}_{0,p_0}), (\psi_{0,p_0p_0}), (\widetilde{\boldsymbol{\psi}}_{1,1}), (\psi_{1,11}), \ldots, (\widetilde{\boldsymbol{\psi}}_{k,p_k}), (\psi_{k,p_kp_k})\}$ *is*

$$\pi_R(\boldsymbol{\psi}) \propto \prod_{i=0}^{k}\prod_{j=1}^{p_i}\psi_{i,jj}^{-1}. \tag{152}$$

(c) *The right-Haar prior for* $\boldsymbol{\Psi}$ *in terms of matrix multiplication is*

$$\boldsymbol{V}_G^r(d\boldsymbol{\Psi}) = \frac{d\boldsymbol{\Psi}}{\prod_{j=0}^{p_0}\psi_{0,jj}^{j}\prod_{i=1}^{k}\prod_{j=1}^{p_i}\psi_{i,jj}^{p_0+j}}. \tag{153}$$

3.12.3 Objective Priors of the Alternative Parameterization

The Fisher information matrix is the launching point for the derivation of non-informative priors, such as Jeffreys' prior and reference priors. Thus, we need to derive the Fisher information matrix for $\boldsymbol{\tau}$ (equivalent to $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$). Instead of directly working on deriving the Fisher information matrix of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$, we derive the Jacobian matrix for the transformation from $\boldsymbol{\Psi}$ into $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$. Then we use Proposition 3.7 to construct the Fisher information matrix of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$.

Proposition 3.8 *For $j = 1, \ldots, p_0$, define*

$$\boldsymbol{\Lambda}_{0j} = \begin{pmatrix} \psi_{0,jj}\boldsymbol{I}_{j-1} & \boldsymbol{\delta}_{0,j} \\ \boldsymbol{0}_{1\times(j-1)} & 1 \end{pmatrix}_{j\times j}, \tag{154}$$

$$\boldsymbol{\Lambda}^{\tau}_{0j} = \begin{pmatrix} \psi^2_{0,jj}(\boldsymbol{\Delta}'_{0,j-1}\boldsymbol{\Xi}^2_{0,j-1}\boldsymbol{\Delta}_{0,j-1})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2}{\psi^2_{0,jj}} \end{pmatrix}; \tag{155}$$

and for $i = 1, \ldots, k$, $j = 1, \ldots, p_i$, define

$$\boldsymbol{\Lambda}_{ij} = \begin{pmatrix} \psi_{i,jj}\boldsymbol{I}_{p_0+j-1} & \widetilde{\boldsymbol{\delta}}_{i,j} \\ \boldsymbol{0}_{1\times(p_0+j-1)} & 1 \end{pmatrix}_{(p_0+j)\times(p_0+j)}, \tag{156}$$

$$\boldsymbol{\Lambda}^{\tau}_{ij} = \begin{pmatrix} \psi^2_{i,jj}(\widetilde{\boldsymbol{\Delta}}'_{i,j-1}\widetilde{\boldsymbol{\Xi}}^2_{i,j-1}\widetilde{\boldsymbol{\Delta}}_{i,j-1})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2}{\psi^2_{i,jj}} \end{pmatrix}. \tag{157}$$

(a) The Jacobian matrix of transformation from $\boldsymbol{\psi}$ into $\boldsymbol{\tau}$ is

$$\boldsymbol{J}(\boldsymbol{\psi}, \boldsymbol{\tau}) = diag(\boldsymbol{\Lambda}_{01}, \ldots, \boldsymbol{\Lambda}_{0p_0}, \boldsymbol{\Lambda}_{11}, \ldots, \boldsymbol{\Lambda}_{1p_1}, \ldots, \boldsymbol{\Lambda}_{k1}, \ldots, \boldsymbol{\Lambda}_{kp_k}). \tag{158}$$

(b) The Fisher information matrix of $\boldsymbol{\tau}$ is

$$I^*(\boldsymbol{\tau}) = diag(\boldsymbol{\Lambda}^{\tau}_{01}, \ldots, \boldsymbol{\Lambda}^{\tau}_{0p_0}, \ldots, \boldsymbol{\Lambda}^{\tau}_{k1}, \ldots, \boldsymbol{\Lambda}^{\tau}_{kp_k}). \tag{159}$$

Proof. Part (a) is easy. We prove only Part (b). For any positive integer j, define $\boldsymbol{e}_j = (0, \cdots, 0, 1)' \in \mathbb{R}^j$. From Sun & Sun (2005) and some matrix algebra, the Fisher information matrix of $\boldsymbol{\psi}$ has the form,

$$I(\boldsymbol{\psi}) = diag(\boldsymbol{\Lambda}^{\psi}_{01}, \ldots, \boldsymbol{\Lambda}^{\psi}_{0p_0}, \boldsymbol{\Lambda}^{\psi}_{11}, \ldots, \boldsymbol{\Lambda}^{\psi}_{1p_1}, \ldots, \boldsymbol{\Lambda}^{\psi}_{k1}, \ldots, \boldsymbol{\Lambda}^{\psi}_{kp_k}), \tag{160}$$

where for $j = 1, \ldots, p_0$, define $\boldsymbol{B}_{0,j-1} = (\Xi_{0,j-1}\boldsymbol{\Delta}_{0,j-1})^{-1}$, then

$$\boldsymbol{\Lambda}_{0j}^{\psi} = \begin{pmatrix} \boldsymbol{B}_{0,j-1} & \boldsymbol{0} \\ -\boldsymbol{\delta}_{0,j}'\boldsymbol{B}_{0,j-1} & \frac{1}{\psi_{0,jj}} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_{0,j-1}' & -\boldsymbol{B}_{0,j-1}'\boldsymbol{\delta}_{0,j} \\ \boldsymbol{0} & \frac{1}{\psi_{0,jj}} \end{pmatrix} + \frac{1}{\psi_{0,jj}^2}\boldsymbol{e}_j\boldsymbol{e}_j'; \quad (161)$$

and for $i = 1, \ldots, k$, $j = 1, \ldots, p_i$, define $\boldsymbol{B}_{i,j-1} = (\widetilde{\Xi}_{i,j-1}\widetilde{\boldsymbol{\Delta}}_{i,j-1})^{-1}$, then

$$\boldsymbol{\Lambda}_{ij}^{\psi} = \begin{pmatrix} \boldsymbol{B}_{i,j-1} & \boldsymbol{0} \\ -\widetilde{\boldsymbol{\delta}}_{i,j}'\boldsymbol{B}_{i,j-1} & \frac{1}{\psi_{i,jj}} \end{pmatrix} \begin{pmatrix} \boldsymbol{B}_{i,j-1}' & -\boldsymbol{B}_{i,j-1}'\widetilde{\boldsymbol{\delta}}_{i,j} \\ \boldsymbol{0} & \frac{1}{\psi_{i,jj}} \end{pmatrix} + \frac{1}{\psi_{i,jj}^2}\boldsymbol{e}_{p_0+j}\boldsymbol{e}_{p_0+j}'. \quad (162)$$

Consider the Jacobian matrix $\boldsymbol{J}(\boldsymbol{\psi}, \boldsymbol{\tau})$ given by (158),

$$\begin{aligned} I^*(\boldsymbol{\tau}) &= \boldsymbol{J}(\boldsymbol{\psi}, \boldsymbol{\tau})' I(\boldsymbol{\psi}) \boldsymbol{J}(\boldsymbol{\psi}, \boldsymbol{\tau}) \\ &= diag(\boldsymbol{\Lambda}_{01}'\boldsymbol{\Lambda}_{01}^{\psi}\boldsymbol{\Lambda}_{01}, \ldots, \boldsymbol{\Lambda}_{0p_0}'\boldsymbol{\Lambda}_{0p_0}^{\psi}\boldsymbol{\Lambda}_{0p_0}, \ldots, \boldsymbol{\Lambda}_{k1}'\boldsymbol{\Lambda}_{k1}^{\psi}\boldsymbol{\Lambda}_{k1}, \ldots, \boldsymbol{\Lambda}_{kp_k}'\boldsymbol{\Lambda}_{kp_k}^{\psi}\boldsymbol{\Lambda}_{kp_k}). \end{aligned}$$

It is easy to versify that $\boldsymbol{\Lambda}_{ij}'\boldsymbol{\Lambda}_{ij}^{\psi}\boldsymbol{\Lambda}_{ij} = \boldsymbol{\Lambda}_{ij}^{\tau}$. Part (b) then follows. □

The Fisher information matrix of $\boldsymbol{\tau}$ given by Proposition 3.8 is a block diagonal matrix, where the blocks corresponding to $\widetilde{\boldsymbol{\delta}}_{i,j}$ ($\boldsymbol{\delta}_{0,j}$) are free of $\widetilde{\boldsymbol{\delta}}_{i,j}$ ($\boldsymbol{\delta}_{0,j}$). Such a structure offers great convenience for deriving various non-informative priors, especially for the reference priors. It is one of the important incentives for re-parameterization from $\boldsymbol{\Psi}$ into $(\Xi, \boldsymbol{\Delta})$.

Proposition 3.9 *Consider the star-shape model with the parameters* $(\Xi, \boldsymbol{\Delta})$.

(a) *The Jeffreys' prior of* $(\Xi, \boldsymbol{\Delta})$ *is*

$$\pi_J(\Xi, \boldsymbol{\Delta}) \propto \prod_{j=1}^{p_0} \psi_{0,jj}^{-(p-2j+2)} \prod_{i=1}^{k}\prod_{j=1}^{p_i} \psi_{i,jj}^{-(p_i-p_0-2j+2)}. \quad (163)$$

(b) *The reference prior of* $(\Xi, \boldsymbol{\Delta})$ *for any ordering of its components is of the form*

$$\pi_R(\Xi, \boldsymbol{\Delta}) \propto \prod_{i=0}^{k}\prod_{j=1}^{p_i} \psi_{i,jj}^{-1}. \quad (164)$$

(c) *The reference prior of* $(\Xi, \boldsymbol{\Delta})$ *in (164) is the same as the right Haar measure for* $\boldsymbol{\Psi}$ *given in (153).*

Proof. For Part (a), the Jeffreys' prior is the square root of the determinant of the Fisher information matrix, which is equivalent to the product of all $|\mathbf{\Lambda}^{\tau}_{ij}|^{\frac{1}{2}}$ given below

$$|\mathbf{\Lambda}^{\tau}_{0j}|^{\frac{1}{2}} = \frac{2\psi_{0,jj}^{j-2}}{\prod_{l=1}^{j}\psi_{0,ll}}, \quad j=1,\dots,p_0; \tag{165}$$

$$|\mathbf{\Lambda}^{\tau}_{ij}|^{\frac{1}{2}} = \frac{2\psi_{i,jj}^{j-2}}{\prod_{l=1}^{p_0}\psi_{0,ll}\prod_{l=1}^{j}\psi_{i,ll}}, \quad i=1,\dots,k,\ j=1,\dots,p_i. \tag{166}$$

Here we propose another intuitive approach for determining out the power of each ψ_{ij} in the square root of the determinant of Fisher information matrix for $(\Xi, \mathbf{\Delta})$.
For, $i=1,\dots,k;\ j=1,\dots,p_i$, define

$$\begin{aligned}
A_{ij} &= \text{number of non-zero elements of } (\Xi,\mathbf{\Delta}) \text{ in the same row and column as } \psi_{i,jj}\equiv 1,\\
B_{ij} &= \text{number of non-zero elements of } (\Xi,\mathbf{\Delta}) \text{ in the same row as } \psi_{i,jj},\\
C_{ij} &= \text{number of non-zero elements of } (\Xi,\mathbf{\Delta}) \text{ in the same column as } \psi_{i,jj},\\
a_{ij} &= \text{the negative power of } \psi_{i,jj} \text{ in the Jeffreys' prior}\\
&= A_{ij}-(B_{ij}-1)+(C_{ij}-1).
\end{aligned}$$

where A_{ij}, B_{ij}, and C_{ij} correspond to the term $\frac{1}{\psi^2_{i,jj}}$, $\psi_{i,jj}$, and $(\widetilde{\mathbf{\Delta}}'_{i,j-1}\widetilde{\Xi}^2_{i,j-1}\widetilde{\mathbf{\Delta}}_{i,j-1})^{-1}$ (or $(\mathbf{\Delta}'_{0,j-1}\Xi^2_{i,j-1}\mathbf{\Delta}_{0,j-1})^{-1}$ for $i=0$) in $\mathbf{\Lambda}^{\tau}_{ij}$ respectively. Thus the Jeffrey's prior is easily derived.

Table 10: Determine the Jeffreys' Prior for $(\Xi, \mathbf{\Delta})$ Via Counting

ψ_{ij}	A_{ij}	-$(B_{ij}-1)$	$C_{ij}-1$	a_{ij}
ψ_{0j}	1	$-(j-1)$	$p-j$	$p-2j+2$
ψ_{ij}	1	$-(p_0+j-1)$	p_i-j	p_i-p_0-2j+2

For Part(b), notice that the Fisher information matrix $\boldsymbol{I}^*(\Xi,\mathbf{\Delta}) = \boldsymbol{I}^*(\boldsymbol{\tau})$ is a block diagonal matrix. Furthermore, for each of the blocks $\mathbf{\Lambda}^{\tau}_{ij}$, the upper left corner part is

$$\psi^2_{i,jj}(\widetilde{\mathbf{\Delta}}'_{i,j-1}\widetilde{\Xi}^2_{i,j-1}\widetilde{\mathbf{\Delta}}_{i,j-1})^{-1},$$

which is not a function of their corresponding parameters $\widetilde{\boldsymbol{\delta}}_{i,j}$. By the algorithm for deriving reference priors in Berger & Bernardo (1992a), we know that the reference prior with any orderings of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ is the same and is determined only by the element in the right and lower corner of $\boldsymbol{\Lambda}_{ij}^{\tau}$, given by $2/\psi_{i,jj}^2$. This is proportional to $\prod_{i=0}^{k}\prod_{j=1}^{p_i}\psi_{i,jj}^{-1}$.

For Part (c), from Jacobian matrix (158), we have that

$$|\boldsymbol{\Lambda}_{0j}| = \psi_{0,jj}^{j-1} \quad and \quad |\boldsymbol{\Lambda}_{ij}| = \psi_{i,jj}^{p_0+j-1}. \tag{167}$$

Thus,

$$\begin{aligned} \pi_R(\boldsymbol{\Xi}, \boldsymbol{\Delta}) \times \prod_{i=0}^{k}\prod_{j=1}^{p_i}\frac{1}{|\boldsymbol{\Lambda}_{ij}|} &= \prod_{i=0}^{k}\prod_{j=1}^{p_i}\frac{1}{\psi_{i,jj}} \times \prod_{j=1}^{p_0}\frac{1}{\psi_{0,jj}^{j-1}}\prod_{i=1}^{k}\prod_{j=1}^{p_i}\frac{1}{\psi_{i,jj}^{p_0+j-1}} \\ &= \prod_{j=1}^{p_0}\frac{1}{\psi_{0,jj}^{j}}\prod_{i=1}^{k}\prod_{j=1}^{p_i}\frac{1}{\psi_{i,jj}^{p_0+j}} \;=\; \boldsymbol{V}_G^{\tau}(d\boldsymbol{\Psi}). \end{aligned} \tag{168}$$

The result holds. □

It is obvious that the priors in Proposition 3.9 are special cases of the class of priors

$$\pi(\boldsymbol{\Xi}, \boldsymbol{\Delta}) \propto \prod_{i=0}^{k}\prod_{j=1}^{p_i}\psi_{i,jj}^{-a_{ij}}, \tag{169}$$

which is equivalent to the class of priors for $\boldsymbol{\Psi}$,

$$\pi(\boldsymbol{\Psi}) \propto \prod_{i=0}^{k}\prod_{j=1}^{p_i}\psi_{i,jj}^{-a_{ij}}, \tag{170}$$

where all a_{ij} are real values. For the rest of this chapter, we will confine our discussion under this class of priors. We consider this general class of priors for the reason that if one is interested in a specific parameter only, then there may be infinite number of different choice of priors in this class that will satisfy exact matching property. This will offer more flexibility for choosing a non-informative prior in practice.

3.13 The Posterior Distributions

Bayesian analysis is based on the posteriors distributions of parameters. In this section, we will derive posteriors of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$. As in Chapter 1.5, we will write them in the form of **constructive random posteriors**.

3.13.1 Sufficient Statistics and Notations

Let $\boldsymbol{x}_h = (\boldsymbol{x}'_{h0}, \boldsymbol{x}'_{h1}, \ldots, \boldsymbol{x}'_{hk})'$, $h = 1, \ldots, n$, be a simple random sample of size n from $N_p(\mathbf{0}, \boldsymbol{\Omega}^{-1})$, the star-shape graphical model. Here $\boldsymbol{x}_{hi} = (x_{hi1}, \ldots, x_{hip_i})', i = 0, \cdots, k$. It is well known that the non-normalized sample covariance matrix,

$$\boldsymbol{S} = \sum_{h=1}^{n} \boldsymbol{x}_h \boldsymbol{x}'_h = (\boldsymbol{S}_{ij})_{i,j=0,\ldots,k} \tag{171}$$

is sufficient for $\boldsymbol{\Omega}$ or $\boldsymbol{\Psi}$ with the Wishart$_p(n, \boldsymbol{\Omega})$ distribution. (See, for example, Gupta & Nagar, 2000, p88). Here $\boldsymbol{S}_{ij}$ is a $p_i \times p_j$ matrix given by

$$\boldsymbol{S}_{ij} = \sum_{h=1}^{n} \boldsymbol{x}_{hi} \boldsymbol{x}'_{hj} \equiv (s_{ij,lm})_{l=1,\cdots,p_i,m=1,\cdots,p_j}, \qquad i, j = 0, \ldots, k. \tag{172}$$

Then the density of $\boldsymbol{S}$ given $\boldsymbol{\Psi}$ is given by

$$f(\boldsymbol{S} \mid \boldsymbol{\Psi}) \propto |\boldsymbol{S}|^{n-p-1} |\boldsymbol{\Psi}|^n \exp\Big\{-\frac{1}{2} tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S})\Big\}. \tag{173}$$

Lemma 3.8 *Define $\boldsymbol{S}_i = \boldsymbol{S}_{ii}$, for $i = 0, 1, \cdots, k$. We have the expression for $tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S})$.*

(a) $tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S}) = tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S}^{\#})$, where

$$\boldsymbol{S}^{\#} = \begin{pmatrix} \boldsymbol{S}_0 & \boldsymbol{S}_{01} & \boldsymbol{S}_{02} & \cdots & \boldsymbol{S}_{0k} \\ \boldsymbol{S}_{10} & \boldsymbol{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{S}_{20} & \mathbf{0} & \boldsymbol{S}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_{k0} & \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{S}_k \end{pmatrix}. \tag{174}$$

(b) We have a decomposition,

$$tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S}) = tr(\boldsymbol{\Psi}_0 \boldsymbol{S}_0 \boldsymbol{\Psi}'_0) + \sum_{i=1}^{k} \eta_i, \tag{175}$$

where

$$\eta_i = tr\left\{ (\boldsymbol{\Psi}_{i0},\ \boldsymbol{\Psi}_i) \begin{pmatrix} \boldsymbol{S}_0 & \boldsymbol{S}_{0i} \\ \boldsymbol{S}_{i0} & \boldsymbol{S}_i \end{pmatrix} \begin{pmatrix} \boldsymbol{\Psi}'_{i0} \\ \boldsymbol{\Psi}'_i \end{pmatrix} \right\}. \tag{176}$$

(c) Define $\boldsymbol{S}_{i\cdot 0} = \boldsymbol{S}_i - \boldsymbol{S}_{i0}\boldsymbol{S}_0^{-1}\boldsymbol{S}_{i0}'$. *Then*

$$\eta_i = tr(\boldsymbol{\Psi}_i \boldsymbol{S}_{i\cdot 0}\boldsymbol{\Psi}_i') + tr\Big\{(\boldsymbol{\Psi}_{i0} + \boldsymbol{\Psi}_i \boldsymbol{S}_{i0}\boldsymbol{S}_0^{-1})\boldsymbol{S}_0(\boldsymbol{\Psi}_{i0} + \boldsymbol{\Psi}_i \boldsymbol{S}_{i0}\boldsymbol{S}_0^{-1})'\Big\}. \tag{177}$$

Proof. For Part (a), note that $tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S}) = tr(\boldsymbol{\Omega}\boldsymbol{S}) = tr(\boldsymbol{\Omega}\boldsymbol{S}^{\#}) = tr(\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{S}^{\#})$. Parts (b) and (c) follow from matrix algebra. □

From Lemma 3.8(a), $\boldsymbol{S}^{\#}$ is also sufficient. Part (b) is useful in finding the posterior distributions of $\boldsymbol{\Psi}_0$, and $(\boldsymbol{\Psi}_{i0},\ \boldsymbol{\Psi}_i)$. Part (c) is useful in finding the conditional posterior distributions of $\boldsymbol{\Psi}_{i0}$ and the marginal posterior distributions of $(\boldsymbol{\Psi}_0, \boldsymbol{\Psi}_1, \cdots, \boldsymbol{\Psi}_k)$.

For convenience, we introduce some notations,

$$\boldsymbol{S}_{i,j} = \begin{pmatrix} s_{i,11} & \cdots & s_{i,1j} \\ \vdots & \ddots & \vdots \\ s_{i,j1} & \cdots & s_{i,jj} \end{pmatrix}_{j\times j}, \quad i = 0,1,\ldots,k,\ j = 1,\ldots,p_i; \tag{178}$$

$$\boldsymbol{s}_{0,j} = (s_{0,j1}, s_{0,j2}, \ldots, s_{0,j(j-1)})', \quad j = 2,\ldots,p_0; \tag{179}$$

$$\boldsymbol{S}_{i0,j} = \begin{pmatrix} s_{i0,11} & \cdots & s_{i0,1p_0} \\ \vdots & \ddots & \vdots \\ s_{i0,j1} & \cdots & s_{i0,jp_0} \end{pmatrix}_{j\times p_0}, \quad i = 1,\ldots,k,\ j = 1,\ldots,p_i; \tag{180}$$

$$\widetilde{\boldsymbol{S}}_{i,0} = \boldsymbol{S}_0, \quad i = 1,\ldots,k; \tag{181}$$

$$\widetilde{\boldsymbol{S}}_{i,j} = \begin{pmatrix} \boldsymbol{S}_0 & \boldsymbol{S}_{i0,j}' \\ \boldsymbol{S}_{i0,j} & \boldsymbol{S}_{i,j} \end{pmatrix}, \quad i = 1,\ldots,k,\ j = 1,\ldots,p_i; \tag{182}$$

$$\widetilde{\boldsymbol{s}}_{i,j} = (s_{i0,j1}, \ldots, s_{i0,jp_0}, s_{i,j1}, \ldots, s_{i,j(j-1)})', \quad i = 1,\ldots,k,\ j = 1,\ldots,p_i; \tag{183}$$

$$u_{0j} = \begin{cases} s_{0,11}, & \text{if } j = 1, \\ s_{0,jj} - \boldsymbol{s}_{0,j}'\boldsymbol{S}_{0,j-1}^{-1}\boldsymbol{s}_{0,j}, & \text{if } j = 2,\ldots,p_0; \end{cases} \tag{184}$$

$$u_{ij} = s_{i,jj} - \widetilde{\boldsymbol{s}}_{i,j}'\widetilde{\boldsymbol{S}}_{i,j-1}^{-1}\widetilde{\boldsymbol{s}}_{i,j}, \quad i = 1,\ldots,k,\ j = 1,\ldots,p_i; \tag{185}$$

$$\boldsymbol{g}_{0j} = \boldsymbol{S}_{0,j-1}^{-1}\boldsymbol{s}_{0,j}, \quad j = 2,\ldots,p_0; \tag{186}$$

$$\boldsymbol{g}_{ij} = \widetilde{\boldsymbol{S}}_{i,j-1}^{-1}\widetilde{\boldsymbol{s}}_{i,j}, \quad i = 1,\ldots,k,\ j = 1,\ldots,p_i. \tag{187}$$

Lemma 3.9 *We have the recursive formulas,*

$$tr(\boldsymbol{\Psi}_0\boldsymbol{S}_0\boldsymbol{\Psi}_0') = \sum_{j=1}^{p_0}\psi_{0,jj}^2 u_{0j} + \sum_{j=2}^{p_0}\psi_{0,jj}^2(\boldsymbol{\delta}_{0,j}+\boldsymbol{g}_{0j})'\boldsymbol{S}_{0,j-1}(\boldsymbol{\delta}_{0,j}+\boldsymbol{g}_{0j}), \quad (188)$$

$$\eta_i = \sum_{j=1}^{p_i}\psi_{i,jj}^2 u_{ij} + \sum_{j=1}^{p_i}\psi_{i,jj}^2(\widetilde{\boldsymbol{\delta}}_{i,j}+\boldsymbol{g}_{ij})'\widetilde{\boldsymbol{S}}_{i,j-1}(\widetilde{\boldsymbol{\delta}}_{i,j}+\boldsymbol{g}_{ij}). \quad (189)$$

3.13.2 Posteriors of the Cholesky Decomposition

Under the class of priors (170), we derive the posterior of $\boldsymbol{\Psi}$ and summarize them into the following theorem.

Theorem 3.2 *For $i = 0, \dots, k$, $j = 1, \dots, p_i$, define $f_{ij} = n - a_{ij} + 1$. Consider the prior (170) and assume that $f_{ij} > 0$.*

(a) *Given $\psi_{i,jj}$, the conditional posteriors of the off-diagonal vectors $\widetilde{\boldsymbol{\psi}}_{i,j}$ ($\boldsymbol{\psi}_{0,j}$) are mutually independent multivariate normal distributions with mean $-\psi_{i,jj}\boldsymbol{g}_{ij}$ ($-\psi_{0,jj}\boldsymbol{g}_{0j}$) and the precision matrix $\widetilde{\boldsymbol{S}}_{i,j-1}^{-1}$ ($\boldsymbol{S}_{0,j-1}^{-1}$), that is,*

$$(\boldsymbol{\psi}_{0,j}|\psi_{0,jj};\boldsymbol{S}_{0,j}) \sim N_{j-1}(-\psi_{0,jj}\boldsymbol{g}_{0j}, \boldsymbol{S}_{0,j-1}^{-1}), \quad j = 1, \dots, p_0; \quad (190)$$

$$(\widetilde{\boldsymbol{\psi}}_{i,j}|\psi_{i,jj};\widetilde{\boldsymbol{S}}_{i,j}) \sim N_{p_0+j-1}(-\psi_{i,jj}\boldsymbol{g}_{ij}, \widetilde{\boldsymbol{S}}_{i,j-1}^{-1}), \quad i = 1, \dots, k,\ j = 1, \dots, p_i. \quad (191)$$

(b) *The marginal posteriors of $\psi_{i,jj}^2$ are independent $Gamma(\frac{f_{ij}}{2}, \frac{u_{ij}}{2})$.*

(c) *Let $\chi_{f_{ij}}^{2*}$ be independent draws from chi-squared distributions with the indicated degree of freedom, and let $\boldsymbol{z}_{i,j}^*$ ($\boldsymbol{z}_{0,j}^*$) be independent draws from $N(\boldsymbol{0}, \boldsymbol{I}_{p_0+j-1})$ ($N(\boldsymbol{0}, \boldsymbol{I}_{j-1})$).*

i) The elements of constructive random posterior of $\boldsymbol{\Psi}$ given $\boldsymbol{X}$(or$\boldsymbol{S}$) can be expressed *as*

$$\begin{aligned}
\psi^*_{i,jj} &= \sqrt{\frac{\chi^{2*}_{f_{ij}}}{u_{ij}}}, \quad i=1,\ldots,k,\ j=1,\ldots,p_i, && (192)\\
\boldsymbol{\psi}^*_{0,j} &= -\psi^*_{0,jj}\boldsymbol{g}_{0j} + \boldsymbol{S}^{-\frac{1}{2}}_{i,j-1}\boldsymbol{z}^*_{0,j}\\
&= -\sqrt{\frac{\chi^{2*}_{f_{0j}}}{u_{0j}}}\boldsymbol{g}_{0j} + \boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}^*_{0,j}, \quad j=2,\ldots,p_0, && (193)\\
\widetilde{\boldsymbol{\psi}}^*_{i,j} &= -\psi^*_{i,jj}\boldsymbol{g}_{ij} + \widetilde{\boldsymbol{S}}^{-\frac{1}{2}}_{i,j-1}\boldsymbol{z}^*_{i,j}\\
&= -\sqrt{\frac{\chi^{2*}_{f_{ij}}}{u_{ij}}}\boldsymbol{g}_{ij} + \widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j}, \quad i=1,\ldots,k,\ j=1,\ldots,p_i, && (194)
\end{aligned}$$

where $\widetilde{\boldsymbol{V}}_{i,j-1}$ *is the Cholesky decomposition of* $\widetilde{\boldsymbol{S}}_{i,j-1}$*, such that* $\widetilde{\boldsymbol{S}}_{i,j-1} = \widetilde{\boldsymbol{V}}_{i,j-1}\widetilde{\boldsymbol{V}}'_{i,j-1}$.
ii) The constructive random posterior of $\boldsymbol{\Sigma}$ *has the expression,*

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Psi}^{*-1}\boldsymbol{\Psi}^{*\prime-1}. \tag{195}$$

3.13.3 Posteriors of the Alternative Parameterization

Under the class of priors (169), we derive the posteriors of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ and summarize them into Theorem 3.3. Define

$$b_{ij} = \begin{cases} n - a_{ij} - j + 2, & \text{if } i=0,\ j=1,\ldots,p_0,\\ n - a_{ij} - p_0 - j + 2, & \text{if } i=1,\ldots,k,\ j=1,\ldots,p_i. \end{cases} \tag{196}$$

Theorem 3.3 *Under the class of priors (169) and assume* $b_{ij} > 0$*:*

(a) *Given* $\psi_{i,jj}$*, the conditional posteriors of the off-diagonal vectors* $\widetilde{\boldsymbol{\delta}}_{i,j}$ $(\boldsymbol{\delta}_{0,j})$ *are mutually independent multivariate normal distributions,*

$$\begin{aligned}
(\boldsymbol{\delta}_{0,j}|\psi_{0,jj};\boldsymbol{S}_{0,j}) &\sim N(-\boldsymbol{g}_{0j}, (\psi^2_{0,jj}\boldsymbol{S}_{0,j-1})^{-1}), \quad j=2,\ldots,p_0; && (197)\\
(\widetilde{\boldsymbol{\delta}}_{i,j}|\psi_{i,jj};\widetilde{\boldsymbol{S}}_{i,j}) &\sim N(-\boldsymbol{g}_{ij}, (\psi^2_{i,jj}\widetilde{\boldsymbol{S}}_{i,j-1})^{-1}), \quad i=1,\ldots,k,\ j=1,\ldots,p_i. && (198)
\end{aligned}$$

(b) *The marginal posteriors of* $\psi^2_{i,jj}$ *are independent Gamma*$(\frac{b_{ij}}{2}, \frac{u_{ij}}{2})$*,*

(c) *Let $\chi^{2*}_{b_{ij}}$ be independent draws from chi-squared distributions with the indicated degree of freedom, and let $\boldsymbol{z}^*_{0,j}$ and $\boldsymbol{z}^*_{i,j}$ be independent draws from $N(\boldsymbol{0}, \boldsymbol{I}_{j-1}))$ and $N(\boldsymbol{0}, \boldsymbol{I}_{p_0+j-1})$, respectively. The constructive random posterior of $(\Xi, \boldsymbol{\Delta})$ given $\boldsymbol{S}$ can be expressed as*

$$\psi^*_{i,jj} = \sqrt{\frac{\chi^{2*}_{b_{ij}}}{u_{ij}}}, \quad i = 0, \ldots, k, \ j = 1, \ldots, p_i, \tag{199}$$

$$\begin{aligned}\boldsymbol{\delta}^*_{0,j} &= -\boldsymbol{g}_{0j} + \frac{1}{\psi^*_{0,jj}} \boldsymbol{S}^{-\frac{1}{2}}_{0,j-1} \boldsymbol{z}^*_{0,j} \\ &= -\boldsymbol{g}_{0j} + \sqrt{\frac{u_{0j}}{\chi^{2*}_{b_{0j}}}} \widetilde{\boldsymbol{V}}'^{-1}_{0,j-1} \boldsymbol{z}^*_{0,j}, \quad j = 2, \ldots, p_0,\end{aligned} \tag{200}$$

$$\begin{aligned}\widetilde{\boldsymbol{\delta}}^*_{i,j} &= -\boldsymbol{g}_{ij} + \frac{1}{\psi^*_{i,jj}} \widetilde{\boldsymbol{S}}^{-\frac{1}{2}}_{i,j-1} \boldsymbol{z}^*_{i,j} \\ &= -\boldsymbol{g}_{ij} + \sqrt{\frac{u_{ij}}{\chi^{2*}_{b_{ij}}}} \widetilde{\boldsymbol{V}}'^{-1}_{i,j-1} \boldsymbol{z}^*_{i,j} \quad i = 1, \ldots, k, \ j = 1, \ldots, p_i.\end{aligned} \tag{201}$$

Furthermore, plug $\psi^*_{i,jj}$, $\boldsymbol{\delta}^*_{0,j}$, $\widetilde{\boldsymbol{\delta}}^*_{i,j}$ into Ξ and $\boldsymbol{\Delta}$ to get Ξ^* and $\boldsymbol{\Delta}^*$. The constructive random posterior of $\boldsymbol{\Sigma}$ has the expression

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Delta}^{*-1} \Xi^{*-2} \boldsymbol{\Delta}^{*\prime-1}. \tag{202}$$

Proof. Under the class of priors (169), from Lemma 3.9 and formulas (144)-(145), the joint posterior density function of $(\Xi, \boldsymbol{\Delta})$ is

$$\begin{aligned}[\Xi, \boldsymbol{\Delta} | \boldsymbol{S}] &\propto [\boldsymbol{S} | \Xi, \boldsymbol{\Delta}] \times \pi(\Xi, \boldsymbol{\Delta}) \\ &\propto \prod_{i=0}^{k} \prod_{j=1}^{p_i} \psi^{n-a_{ij}}_{i,jj} exp\Big\{ -\frac{1}{2} \sum_{i=0}^{k} \sum_{j=1}^{p_i} \psi^2_{i,jj} u_{ij} \Big\} \\ &\quad \times exp\Big\{ -\frac{1}{2} \sum_{j=2}^{p_0} (\boldsymbol{\delta}_{0,j} + \boldsymbol{g}_{0j})' \psi^2_{0,jj} \boldsymbol{S}_{0,j-1} (\boldsymbol{\delta}_{0,j} + \boldsymbol{g}_{0j}) \Big\} \\ &\quad \times exp\Big\{ -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{p_i} (\widetilde{\boldsymbol{\delta}}_{i,j} + \boldsymbol{g}_{ij})' \psi^2_{i,jj} \widetilde{\boldsymbol{S}}_{i,j-1} (\widetilde{\boldsymbol{\delta}}_{i,j} + \boldsymbol{g}_{ij}) \Big\}.\end{aligned} \tag{203}$$

Therefore, Part (a) of Theorem 3.3 directly follows.

To prove Part (b), integrate out the $\boldsymbol{\delta}_{0,j}$ and $\widetilde{\boldsymbol{\delta}}_{i,j}$ from (203)

$$[\Xi \mid \boldsymbol{S}] \propto \prod_{i=0}^{k} \prod_{j=1}^{p_i} \psi^{n-a_{ij}}_{i,jj} exp\Big\{ -\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{p_i} \psi^2_{i,jj} u_{ij} \Big\} \prod_{i=0}^{k} \prod_{j=1}^{p_i} |\psi^2_{i,jj} \widetilde{\boldsymbol{S}}_{i,j-1}|^{-\frac{1}{2}}. \tag{204}$$

Then, use $\psi^2_{i,jj}$ as interesting variables, we obtain

$$\begin{aligned}
[\psi^2_{i,jj}|\boldsymbol{S}] &\propto \prod_{i=0}^{k}\prod_{j=1}^{p_i} {\psi^2_{i,jj}}^{\frac{n-a_{ij}-1}{2}} exp\Big\{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{p_i}\psi^2_{i,jj}u_{ij}\Big\} \times \prod_{i=0}^{k}\prod_{j=1}^{p_i} {\psi^{2}_{i,jj}}^{(-\frac{I(i>0)p_0+j-1}{2})} \\
&\propto \prod_{i=0}^{k}\prod_{j=1}^{p_i} (\psi^2_{i,jj})^{\frac{n-a_{ij}-I(i>0)p_0-j+2}{2}-1} \exp\Big\{-\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{p_i}\psi^2_{i,jj}u_{ij}\Big\}. \qquad (205)
\end{aligned}$$

This implies Part (b). Part (c) gives the constructive random posteriors based on Parts (a) and (b).

□

3.14 Frequentist Matching Properties

In this section, we study the matching property of the constructive random posteriors derived in Section 3.13. We first employ the constructive random posteriors to prove exact matching priors for all $\boldsymbol{\Psi}$ and $(\boldsymbol{\Xi},\boldsymbol{\Delta})$ in one subsection. Then we give the Monte Carlo simulation result in another subsection.

3.14.1 Frequentist Distributions of Sufficient Statistics

The key to proving exact matching properties of a constructive random posterior parameter (function of parameters) is to figure out the distribution the sufficient statistics or function of sufficient statistics used in the expression of the constructive random posterior. In this star-shape case, the sufficient statistics is $\boldsymbol{S}$ or $\boldsymbol{S}^{\#}$. Thus, we start this subsection from the derivation of the distribution of $\boldsymbol{S}$.

Fact 3.6 *From Gupta & Nagar (2000, P88), let* $\boldsymbol{X} \sim N_{p,n}(\mathbf{0},\boldsymbol{\Sigma}\otimes\mathbf{I_n})$ *and define* $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}'$, $n \geq p$. *Then* $\boldsymbol{S} \sim W_p(n,\boldsymbol{\Sigma})$, *a Whishart distribution.*

In this study, we use parameterizations $\boldsymbol{\Psi}$ and $(\boldsymbol{\Xi},\boldsymbol{\Delta})$. Thus we have $\boldsymbol{S} \sim W_p(n,(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1})$ and $\boldsymbol{S} \sim W_p(n,(\boldsymbol{\Delta}'\boldsymbol{\Xi}^2\boldsymbol{\Delta})^{-1})$ respectively.

Let $\boldsymbol{I}_p$ denote a p-dimensional identity matrix. We define

$$\boldsymbol{A}_0 = (\boldsymbol{I}_{p_0} \ \boldsymbol{0}_{p_0\times(p_1+\cdots+p_k)}), \tag{206}$$

$$\boldsymbol{A}_i = \begin{pmatrix} \boldsymbol{I}_{p_0} & \boldsymbol{0}_{p_0\times(p_1+\cdots+p_{i-1})} & \boldsymbol{0}_{p_0\times p_i} & \boldsymbol{0}_{p_0\times(p_{i+1}+\cdots+p_k)} \\ \boldsymbol{0}_{p_i\times p_0} & \boldsymbol{0}_{p_i\times(p_1+\cdots+p_{i-1})} & \boldsymbol{I}_{p_i} & \boldsymbol{0}_{p_i\times(p_{i+1}+\cdots+p_k)} \end{pmatrix}, \quad i = 1,\ldots,k. \tag{207}$$

We now give the frequentist distribution of some related partitions of $\boldsymbol{S}$.

Fact 3.7 *The marginal distributions of $\boldsymbol{S}_0$ and $\widetilde{\boldsymbol{S}}_{i,p_i}$ are as follows*

$$\boldsymbol{S}_0 \mid \boldsymbol{\Psi} \sim W_{p_0}(n, (\boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0)^{-1}), \tag{208}$$

$$\widetilde{\boldsymbol{S}}_{i,p_i} \mid \boldsymbol{\Psi} \sim W_{p_0+p_i}(n, (\widetilde{\boldsymbol{\Psi}}_{i,p_i}'\widetilde{\boldsymbol{\Psi}}_{i,p_i})^{-1}). \tag{209}$$

Proof. From Gupta & Nagar (2000, P96), we have

$$\begin{aligned} \boldsymbol{A}_0\boldsymbol{S}\boldsymbol{A}_0' &\sim W_{p_0}(n, \boldsymbol{A}_0(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{A}_0'), \\ \boldsymbol{A}_i\boldsymbol{S}\boldsymbol{A}_i' &\sim W_{p_0+p_i}(n, \boldsymbol{A}_i(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{A}_i'). \end{aligned}$$

After some matrix algebra, we also have

$$\begin{aligned} \boldsymbol{A}_0\boldsymbol{S}\boldsymbol{A}_0' &= \boldsymbol{S}_0, \\ \boldsymbol{A}_0(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{A}_0' &= (\boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0)^{-1}; \\ \boldsymbol{A}_i\boldsymbol{S}\boldsymbol{A}_i' &= \begin{pmatrix} \boldsymbol{S}_0 & \boldsymbol{S}_{i0}' \\ \boldsymbol{S}_{i0} & \boldsymbol{S}_i \end{pmatrix} = \widetilde{\boldsymbol{S}}_{i,p_i}, \\ \boldsymbol{A}_i(\boldsymbol{\Psi}'\boldsymbol{\Psi})^{-1}\boldsymbol{A}_i' &= \left(\begin{pmatrix} \boldsymbol{\Psi}_0 & \boldsymbol{0} \\ \boldsymbol{\Psi}_{i0} & \boldsymbol{\Psi}_i \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Psi}_0 & \boldsymbol{0} \\ \boldsymbol{\Psi}_{i0} & \boldsymbol{\Psi}_i \end{pmatrix}\right)^{-1} = \left(\widetilde{\boldsymbol{\Psi}}_{i,p_i}'\widetilde{\boldsymbol{\Psi}}_{i,p_i}\right)^{-1}. \end{aligned}$$

Therefore, Fact 3.7 is proved. □

To verify matching properties, we need the frequentist distributions of $\boldsymbol{g}_{ij}$ and u_{ij}. This can be derived from Fact 3.7.

Lemma 3.10 *Define*

$$c_{ij} \;=\; \begin{cases} n-j+1, & \text{if } i=0,\ j=1,\ldots,p_0, \\ n-p_0-j+1, & \text{if } i=1,\ldots,k,\ j=1,\ldots,p_i. \end{cases} \tag{210}$$

(a) If $c_{0j}>0$, we have,

$$\boldsymbol{g}_{0j} \mid \psi_{0,jj}, \boldsymbol{\psi}_{0,j}, \boldsymbol{S}_{0,j-1} \;\sim\; N_{j-1}\big(-\frac{\boldsymbol{\psi}_{0,j}}{\psi_{0,jj}}, \psi_{0,jj}^{-2}\boldsymbol{S}_{0,j-1}^{-1}\big), \quad j=2,\ldots,p_0, \tag{211}$$

$$u_{0j} \mid \psi_{0,jj}, \boldsymbol{S}_{0,j-1} \;\sim\; Gamma\big(\frac{c_{0j}}{2}, \frac{\psi_{0,jj}^2}{2}\big), \quad j=1,\ldots,p_0, \tag{212}$$

or

$$\boldsymbol{g}_{0j} \mid \psi_{0,jj}, \boldsymbol{\delta}_{0,j}, \boldsymbol{S}_{0,j-1} \;\sim\; N_{j-1}\big(\boldsymbol{\delta}_{0,j}, \psi_{0,jj}^{-2}\boldsymbol{S}_{0,j-1}^{-1}\big), \quad j=2,\ldots,p_0, \tag{213}$$

$$u_{0j} \mid \psi_{0,jj}, \boldsymbol{S}_{0,j-1} \;\sim\; Gamma\big(\frac{c_{0j}}{2}, \frac{\psi_{0,jj}^2}{2}\big), \quad j=1,\ldots,p_0. \tag{214}$$

(b) For $i=1,\ldots,k$, $j=1,\ldots,p_i$, if $c_{ij}>0$, we have,

$$\boldsymbol{g}_{ij} \mid \psi_{i,jj}, \widetilde{\boldsymbol{\psi}}_{i,j}, \widetilde{\boldsymbol{S}}_{i,j-1} \;\sim\; N_{p_0+j-1}\big(-\frac{\widetilde{\boldsymbol{\psi}}_{i,j}}{\psi_{i,jj}}, \psi_{i,jj}^{-2}\widetilde{\boldsymbol{S}}_{i,j-1}^{-1}\big), \tag{215}$$

$$u_{ij} \mid \psi_{i,jj}, \widetilde{\boldsymbol{S}}_{i,j-1} \;\sim\; Gamma\big(\frac{c_{ij}}{2}, \frac{\psi_{i,jj}^2}{2}\big); \tag{216}$$

or equivalently,

$$\boldsymbol{g}_{ij} \mid \psi_{i,jj}, \widetilde{\boldsymbol{\delta}}_{i,j}, \widetilde{\boldsymbol{S}}_{i,j-1} \;\sim\; N_{p_0+j-1}\big(-\widetilde{\boldsymbol{\delta}}_{i,j}, \psi_{i,jj}^{-2}\widetilde{\boldsymbol{S}}_{i,j-1}^{-1}\big), \tag{217}$$

$$u_{ij} \mid \psi_{i,jj}, \widetilde{\boldsymbol{S}}_{i,j-1} \;\sim\; Gamma\big(\frac{c_{ij}}{2}, \frac{\psi_{i,jj}^2}{2}\big). \tag{218}$$

Proof. By partition of a symmetric matrix iteratively, we have

$$|\boldsymbol{S}_{0,p_0}| \;=\; \frac{|\boldsymbol{S}_{0,p_0}|}{|\boldsymbol{S}_{0,p_0-1}|}\times\cdots\times\frac{|\boldsymbol{S}_{0,2}|}{|\boldsymbol{S}_{0,1}|}\times|\boldsymbol{S}_{0,1}| \;=\; \prod_{l=1}^{p_0} u_{0l}, \tag{219}$$

$$|\widetilde{\boldsymbol{S}}_{i,p_i}| \;=\; \frac{|\widetilde{\boldsymbol{S}}_{i,p_i}|}{|\widetilde{\boldsymbol{S}}_{i,p_i-1}|}\times\frac{|\widetilde{\boldsymbol{S}}_{i,p_i-1}|}{|\widetilde{\boldsymbol{S}}_{i,p_i-2}|}\times\cdots\times\frac{|\widetilde{\boldsymbol{S}}_{i,1}|}{|\widetilde{\boldsymbol{S}}_{i,0}|}\times|\widetilde{\boldsymbol{S}}_{i,0}| \;=\; |\boldsymbol{S}_0|\prod_{l=1}^{p_i} u_{il}. \tag{220}$$

In general,

$$|\boldsymbol{S}_{0,j}| \;=\; \prod_{l=1}^{j} u_{0l}, \quad j=1,\ldots,p_0; \tag{221}$$

$$|\widetilde{\boldsymbol{S}}_{i,j}| \;=\; |\boldsymbol{S}_0|\prod_{l=1}^{j} u_{il}, \quad i=1,\ldots,k,\ j=1,\ldots,p_i. \tag{222}$$

The Jacobian matrix from $(u_{01}, \boldsymbol{g}_{02}, u_{02}, \ldots, \boldsymbol{g}_{kp_k}, u_{kp_k})$ to $(\boldsymbol{S}_{0,1}, \ldots, \boldsymbol{S}_{0,p0}, \widetilde{\boldsymbol{S}}_{1,1}, \ldots, \widetilde{\boldsymbol{S}}_{k,p_k})$ is

$$J_{gu} = diag(1, \boldsymbol{S}_{0,1}^{-1}, 1, \ldots, \boldsymbol{S}_{0,p_0-1}^{-1}, 1, \ldots, \widetilde{\boldsymbol{S}}_{k,1}^{-1}, 1, \ldots, \widetilde{\boldsymbol{S}}_{k,p_k-1}^{-1}, 1). \tag{223}$$

The marginal density functions of $(u_{ij}, \boldsymbol{g}_{ij}, j = 1, \cdots, p_i)$ can then be derived from Fact 3.7. For $i = 0$,

$$\begin{aligned}
[u_{01}, \boldsymbol{g}_{0j}, u_{0j}, j = 1, \cdots, p_0 \mid \boldsymbol{\Xi}_0, \boldsymbol{\Delta}_0] \;&\propto\; [\boldsymbol{S}_0 \mid \boldsymbol{\Xi}_0, \boldsymbol{\Delta}_0] \times |J(\boldsymbol{S}_0, (\boldsymbol{g}_{0j}, u_{0j}))| \\
\propto\;& |\boldsymbol{S}_0|^{\frac{1}{2}(n-p_0-1)} etr\Big\{ -\frac{1}{2}(\boldsymbol{\Delta}_0'\boldsymbol{\Xi}_0^2\boldsymbol{\Delta}_0)\boldsymbol{S}_0 \Big\} \times \prod_{j=1}^{p_0-1} |\boldsymbol{S}_{0,j}| \\
\propto\;& \prod_{j=1}^{p_0} u_{0j}^{\frac{1}{2}(n-j-1)} \exp\Big\{ -\frac{1}{2}\sum_{j=1}^{p_0} \psi_{0,jj}^2 u_{0j} - \frac{1}{2}\sum_{j=2}^{p_0} (\boldsymbol{\delta}_{0,j} + \boldsymbol{g}_{0j})'\psi_{0,jj}^2 \boldsymbol{S}_{0,j-1}(\boldsymbol{\delta}_{0,j} + \boldsymbol{g}_{0j}) \Big\},
\end{aligned} \tag{224}$$

and for $i > 1$,

$$\begin{aligned}
[\boldsymbol{g}_{ij}, u_{ij}, j = 1, \cdots, p_i \mid \widetilde{\boldsymbol{\Xi}}_{i,p_i}, \widetilde{\boldsymbol{\Delta}}_{i,p_i}, \boldsymbol{S}_0] &\propto [\widetilde{\boldsymbol{S}}_{i,p_i} \mid \widetilde{\boldsymbol{\Xi}}_{i,p_i}, \widetilde{\boldsymbol{\Delta}}_{i,p_i}, \boldsymbol{S}_0] \times |J(\widetilde{\boldsymbol{S}}_{i,p_i}, (\boldsymbol{g}_{ij}, u_{ij}))| \\
\propto\;& |\widetilde{\boldsymbol{S}}_{i,p_i}|^{\frac{1}{2}(n-p_0-p_i-1)} etr\Big\{ -\frac{1}{2}(\widetilde{\boldsymbol{\Delta}}_{i,p_i}'\widetilde{\boldsymbol{\Xi}}_{i,p_i}^2\widetilde{\boldsymbol{\Delta}}_{i,p_i})\widetilde{\boldsymbol{S}}_{i,p_i} \Big\} \times \prod_{j=1}^{p_i-1} |\widetilde{\boldsymbol{S}}_{i,j}| \\
\propto\;& \prod_{j=1}^{p_i} u_{0j}^{\frac{1}{2}(n-p_0-j-1)} \exp\Big\{ -\frac{1}{2}\sum_{j=1}^{p_i} \psi_{i,jj}^2 u_{ij} - \frac{1}{2}\sum_{j=1}^{p_i} (\widetilde{\boldsymbol{\delta}}_{i,j} + g_{ij})'\psi_{i,jj}^2 \widetilde{\boldsymbol{S}}_{i,j-1}(\widetilde{\boldsymbol{\delta}}_{i,j} + \boldsymbol{g}_{ij}) \Big\}.
\end{aligned} \tag{225}$$

Therefore, Lemma 3.10 is proved. □

3.14.2 Exact Frequentist Matching

From the previous results, we conclude the frequentist matching properties in the following theorems and corollaries.

Theorem 3.4 *Under the Cholesky decomposition $\boldsymbol{\Psi}$, for any $i = 0, \ldots, k$ and $j = 1, \ldots, p_i$,*

(a) $P(\psi_{i,jj} < [\psi_{i,jj}^*]_\alpha) = \alpha$,

(b) $P(\psi_{i,jl} < [\psi_{i,jl}^*]_\alpha) = \alpha, \quad l = 1, \ldots, j-1,$ *and*

(c) $P(\psi_{i0,jl} < [\psi^*_{i0,jl}]_\alpha) = \alpha, \quad l = 1, \ldots, p_0,$

if and only if $a_{ij} = I(i > 0)p_0 + j$ *within the class of priors given in (170). Therefore, the matching prior belongs to the class of priors given by (170).*

Proof. Let $\chi_n = \sqrt{\chi^2_n}$. For the class of priors (170), if and only if $a_{ij} = I(i > 0)p_0 + j$, we have f_{ij}=c_{ij} and

$$
\begin{aligned}
P(\psi_{i,jj} < [\psi^*_{i,jj}]_\alpha \mid \boldsymbol{\Psi}) &= P\left(\psi_{ij} < \left[\frac{\chi^*_{f_{ij}}}{\sqrt{u_{ij}}}\right]_\alpha\right) \\
&= P\left(\psi_{i,jj}\sqrt{u_{ij}} < \left[\chi^*_{f_{ij}}\right]_\alpha\right) \\
&= P\left(\chi_{c_{ij}} < \left[\chi^*_{c_{ij}}\right]_\alpha\right) \\
&= \alpha, \qquad (226)
\end{aligned}
$$

Let $\boldsymbol{e}_{0j,l} = (0, \ldots, 0, 1, 0, \ldots, 0)' \in \mathbb{R}^{j-1}$, with the l^{th} element being 1 and all others being 0,

$$
\begin{aligned}
& P(\psi_{0,jl} < [\psi^*_{0,jl}]_\alpha \mid \boldsymbol{\Psi}) \\
= {} & P(\boldsymbol{e}'_{0j,l}\boldsymbol{\psi}_{0,j} < [\boldsymbol{e}'_{0j,l}\boldsymbol{\psi}^*_{0,j}]_\alpha \mid \boldsymbol{\Psi}) \\
= {} & P\left(\boldsymbol{e}'_{0j,l}\boldsymbol{\psi}_{0,j} < \left[-\boldsymbol{e}'_{0j,l}\psi^*_{0,jj}\boldsymbol{g}_{0j} + \boldsymbol{e}'_{0j,l}\boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}^*_{0,j}\right]_\alpha\right) \\
= {} & P\left(\boldsymbol{e}'_{0j,l}\boldsymbol{\psi}_{0,j} < \left[-\boldsymbol{e}'_{0j,l}\frac{\psi^*_{0,jj}}{\psi_{0,jj}}\psi_{0,jj}\boldsymbol{g}_{0j} + \boldsymbol{e}'_{0j,l}\boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}^*_{0,j}\right]_\alpha\right) \\
= {} & P\left(\boldsymbol{e}'_{0j,l}\boldsymbol{\psi}_{0,j} < \left[\boldsymbol{e}'_{0j,l}\frac{\psi^*_{0,jj}}{\psi_{0,jj}}\boldsymbol{\psi}_{0,j} + \boldsymbol{e}'_{0j,l}\boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}_{0,j} + \boldsymbol{e}'_{0j,l}\boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}^*_{0,j}\right]_\alpha\right) \\
= {} & P\left(\boldsymbol{e}'_{0j,l}(\frac{\psi^*_{0,jj}}{\psi_{0,jj}}\boldsymbol{\psi}_{0,jj} + \boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}_{0,j}) < \left[\boldsymbol{e}'_{0j,l}(\frac{\psi^*_{0,jj}}{\psi_{0,jj}}\boldsymbol{\psi}_{0,j} + \boldsymbol{V}'^{-1}_{0,j-1}\boldsymbol{z}^*_{0,j})\right]_\alpha\right) \\
= {} & \alpha. \qquad (227)
\end{aligned}
$$

Let $\boldsymbol{e}_{ij,t} = (0, \ldots, 0, 1, 0, \ldots, 0)' \in \mathbb{R}^{p_0+j-1}$, with the t^{th} elements being 1 and all others being 0, then

$$\begin{aligned}
& P(\psi_{i,jt} < [\psi^*_{i,jt}]_\alpha \mid \boldsymbol{\Psi}) \\
= \; & P(\boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{\psi}}_{i,j} < [\boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{\psi}}^*_{i,j}]_\alpha \mid \boldsymbol{\Psi}) \\
= \; & P\left(\boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{\psi}}_{i,j} < \left[-\boldsymbol{e}'_{ij,t}\widetilde{\psi}^*_{i,jj}\boldsymbol{g}_{ij} + \boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j}\right]_\alpha\right) \\
= \; & P\left(\boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{\psi}}_{i,j} < \left[-\boldsymbol{e}'_{ij,t}\frac{\widetilde{\psi}^*_{i,jj}}{\psi_{i,jj}}\psi_{i,jj}\boldsymbol{g}_{ij} + \boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j}\right]_\alpha\right) \\
= \; & P\left(\boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{\psi}}_{i,j} < \left[\boldsymbol{e}'_{ij,t}\frac{\psi^*_{i,jj}}{\psi_{i,jj}}\widetilde{\boldsymbol{\psi}}_{i,j} + \boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}_{i,j} + \boldsymbol{e}'_{ij,t}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j}\right]_\alpha\right) \\
= \; & P\left(\boldsymbol{e}'_{ij,t}(\frac{\psi^*_{i,jj}}{\psi_{i,jj}}\widetilde{\boldsymbol{\psi}}_{i,jj} + \widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}_{i,j}) < \left[\boldsymbol{e}'_{ij,t}(\frac{\psi^*_{i,jj}}{\psi_{i,jj}}\widetilde{\boldsymbol{\psi}}_{i,j} + \widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j})\right]_\alpha\right) \\
= \; & \alpha, \qquad (228)
\end{aligned}$$

where, for $t = 1, \ldots, p_0$, $\psi_{i,jt} = \psi_{i0,jt}$, and for $t = p_0, \ldots, p_0 - j - 1$, $\psi_{i,jt} = \psi_{i,j(t-p_0)}$. □

From Theorem 3.4, the following corollary automatically follows.

Corollary 3.4 *The right-Haar prior for $\boldsymbol{\Psi}$ within the class of priors given by (170) is exact frequentist matching prior for all $\psi_{i0,jl}$, $\psi_{i,jl}$, and $\psi_{i,jj}$.*

Under the alternative parametrization, we attain a corresponding theorem on frequentist matching properties as well.

Theorem 3.5 *Under the parameterization $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$, for any $i = 0, \ldots, k$ and $j = 1, \ldots, p_i$,*

(a) $P(\psi_{i,jj} < [\psi^*_{i,jj}]_\alpha) = \alpha$,

(b) $P(\delta_{i,jl} < [t^*_{i,jl}]_\alpha) = \alpha$, $\quad l = 1, \ldots, j-1$, *and*

(c) $P(\delta_{i0,jl} < [t^*_{i0,jl}]_\alpha) = \alpha$, $\quad l = 1, \ldots, p_0$,

if and only if $a_{ij} = 1$ within the class of priors given in (169). Therefore, the matching prior belongs to the class of priors given by (169).

Proof. Let $\boldsymbol{e}_{ij,l} = (0, \ldots, 0, 1, 0, \ldots, 0)$, a $(I(i > 0)p_0 + j - 1)$ dimensional vector with the l^{th} elements being 1 and all others being 0. If and only if $a_{ij} = 1$, we have b_{ij}=c_{ij}, such that

$$\begin{aligned} P(\psi_{i,jj} < [\psi^*_{i,jj}]_\alpha|\boldsymbol{\Xi}) &= P\left(\psi_{i,jj} < \left[\frac{\chi^*_{b_{ij}}}{\sqrt{u_{ij}}}\right]_\alpha\right) \\ &= P\left(\psi_{i,jj}\sqrt{u_{ij}} < \left[\chi^*_{b_{ij}}\right]_\alpha\right) \\ &= P\left(\chi_{c_{ij}} < \left[\chi^*_{c_{ij}}\right]_\alpha\right) \\ &= \alpha, \end{aligned} \tag{229}$$

and

$$\begin{aligned} & P(\delta_{i,jl} < [t^*_{i,jl}]_\alpha|\boldsymbol{\Xi}, \boldsymbol{\Delta}) \\ = \;& P(\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{\delta}}_{i,j} < [\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{\delta}}^*_{i,j}]_\alpha|\boldsymbol{\Xi}, \boldsymbol{\Delta}) \\ = \;& P\left(\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{\delta}}_{i,j} < \left[-\boldsymbol{e}'_{ij,l}\boldsymbol{g}_{ij} + \psi^*_{i,jj}\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j}\right]_\alpha\right) \\ = \;& P\left(\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{\delta}}_{i,j} < \left[\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{\delta}}_{i,j} + \psi_{i,jj}\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}_{i,j} + \psi^*_{i,jj}\boldsymbol{e}'_{ij,l}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j}\right]_\alpha\right) \\ = \;& P\left(\boldsymbol{e}'_{ij,l}(\widetilde{\boldsymbol{\delta}}_{i,j} + \psi_{i,jj}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}_{i,j}) < \left[\boldsymbol{e}'_{ij,l}(\widetilde{\boldsymbol{\delta}}_{i,j} + \psi^*_{i,jj}\widetilde{\boldsymbol{V}}'^{-1}_{i,j-1}\boldsymbol{z}^*_{i,j})\right]_\alpha\right) \\ = \;& \alpha. \end{aligned} \tag{230}$$

The results hold. □

Theorem 3.5 implies another corollary below.

Corollary 3.5 *The right-Haar prior for* $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ *within the class of priors given by (169) is exact frequentist matching prior for all* $\psi_{i,jj}$, $\delta_{i,jl}$, *and* $\delta_{i0,jl}$.

3.14.3 Numerical Example

As in Chapter 1.5, another way of verifying frequentist matching properties is through Monte Carlo simulation. The Monte Carlo simulation procedure for the star-shape model is as follows:

Step 0. Choose the parameter set $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ of interest and specify the sample size n.

Step 1. Simulate a sample of $\boldsymbol{X}$ and calculate $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}'$.

Step 2. For each sample in Step 1, simulate m_2 values of $(\boldsymbol{\Xi}^*, \boldsymbol{\Delta}^*)$ using the constructive random posterior expressions.

Step 3. Record a 1 if the true value of $\psi_{i,jj}$ (or $\delta_{ij,kl}$) is less than $(\psi^*_{i,jj})_\alpha$(or $(\delta^*_{ij,kl})_\alpha$).

Step 4. Repeat Step 1-3 for m_1 times. The total number of 1's recorded divided by m_1 is a sample of the corresponding frequentist coverage for the α credible interval, with a mean of α and a variance of $\frac{\alpha(1-\alpha)}{m_1}$, under exact matching.

In this section, we give an example to show the adequacy of the frequentist matching of the $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$. Assume we are interested in a star-shape case with parameters $k = 3$, $p_0 = 3,\ p_1 = 2,\ p_2 = 2,\ p_3 = 3$, and

$$\boldsymbol{\Psi} = \left(\begin{array}{ccc|cc|cc|ccc}
1 & & & & & & & & & \\
.5 & 1 & & & & & & & & \\
0.5 & 0.6 & 1 & & & & & & & \\
\hline
0.3 & 0.5 & 0.8 & 2 & & & & & & \\
0.5 & 0.6 & 0.4 & 1 & 4 & & & & & \\
\hline
0.3 & 0.5 & 0.8 & 0 & 0 & 1 & & & & \\
0.5 & 0.7 & 0.9 & 0 & 0 & 0.3 & 1 & & & \\
\hline
0.3 & 0.5 & 0.8 & 0 & 0 & 0 & 0 & 1 & & \\
0.5 & 0.7 & 0.9 & 0 & 0 & 0 & 0 & 0.3 & 1 & \\
0.5 & 0.7 & 0.9 & 0 & 0 & 0 & 0 & 0.3 & 1 & 3
\end{array}\right). \tag{231}$$

We follow the monte carlo procedures with the sample size $n = 20$. Having proved that the right-Haar prior is an exact matching prior for all the model parameters, we first run the simulation using the right-Haar prior to assure that the program is correctly coded. Then, we rerun the simulation with Jeffrey's prior. We know that Jeffreys' prior is not exactly matching for all parameters, but we do not know how much different the frequentist coverage will be from the nominal α. The simulation offers information for answering this question.

The frequentist coverage for all elements of $(\Xi, \boldsymbol{\Delta})$ with the specified values from the Monte Carlo simulation is reported in Table 11-12. We choose $\alpha = .01$, .025, .05, .50, .95, .975, and .99. Here is a list of findings from these simulations:

(1) Under the right-Haar prior, all coverage probabilities for each element of $(\Xi, \boldsymbol{\Delta})$ are very close to the nominal α. This confirms the correctness of the program coding.

(2) Under the Jeffreys' prior, the program only differs from the program for right-Haar prior simulation in the corresponding degree of freedoms. The report shows that the frequentist coverage for some of the parameters, such as $\psi_{0,11}$ is very different from the corresponding nominal α. This, again, shows the limitation of Jeffreys' prior for multi-parameter case.

It is possible to check how sample size affects the coverage or how the coverage differs from one set of parameter values to another by simple modifications of the above program. It is also possible to check frequentist coverage for other functions of parameters of interest such as $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ by further modification on the program. These actions, requiring considerably computer time, will be taken at a later date.

3.15 Application

In Whittaker (1990), a data set that includes the mathematics marks of five mathematics exams of 88 students are considered. It is shown that the marks of five mathematics exams could be grouped into three sets (algebra), (mechanics, vectors), and (analysis, statistics). It is also shown that conditional on (algebra), (mechanics, vectors) are independent of (analysis, statistics). Figure 2 shows relationship among student marks of these math subjects under the star-shape scheme, which is a specific example of Figure 1(a).

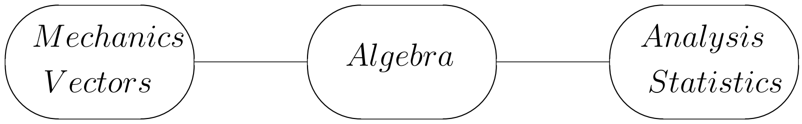

Figure 2: Independence graph of the mathematics marks: $k = 2$.

The data set gives a Gaussian graphical model with $k = 2$, $p_0 = 1$, $p_1 = 2$, $p_2 = 2$, and $n = 88$. Since one degree of freedom is used to estimate $(u_{01}, u_{11}, u_{12}, u_{21}, u_{22})$, the degrees of freedom left are 87. The normalized sample covariance matrix of (Algebra, Mechanics, Vectors, Analysis, Statistics) is

$$\boldsymbol{S} = \left(\begin{array}{c|cc|cc} 111.60 & & & & \\ \hline 100.43 & 302.29 & & & \\ 84.19 & 125.78 & 170.88 & & \\ \hline 110.84 & 105.07 & 93.60 & 217.88 & \\ 120.49 & 116.07 & 97.89 & 153.77 & 294.37 \end{array} \right). \tag{232}$$

Based on this normalized sample covariance $\boldsymbol{S}$, we draw the posterior distribution of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$, calculate the posterior precision matrix $\boldsymbol{\Omega}$ and covariance matrix $\boldsymbol{\Sigma}$. The marginal posterior densities of these parameters are plotted in Figures 3, 4, and 5. It is interesting that all the off-diagonal elements $\delta_{10,21}$, $\delta_{20,11}$, $\delta_{20,21}$, $\delta_{1,21}$, and $\delta_{2,21}$ are in negative ranges.

Table 11: Frequentist Coverage for Credible Intervals of $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$ with Right-Haar Prior for $(\boldsymbol{\Xi}, \boldsymbol{\Delta})$; $n = 20$, $m_1 = m_2 = 10,000$; $k = 3$, $p_0 = 3$, $p_1 = p_2 = 2$, $p_3 = 3$.

α	.01	.025	.05	.50	.95	.975	.99
$\psi_{0,11}$	.0104	.0246	.0522	.5034	.9489	.9744	.9899
$\delta_{0,21}$	.0101	.0241	.0498	.4971	.9498	.9747	.9895
$\psi_{0,22}$	.0109	.0264	.0525	.4941	.9508	.9764	.9898
$\delta_{0,31}$	.0081	.0235	.0482	.4938	.9498	.9742	.9891
$\delta_{0,32}$	.0097	.0246	.0487	.5016	.9486	.9744	.9898
$\psi_{0,33}$	.0094	.0237	.0475	.5077	.9519	.9779	.9910
$\delta_{10,11}$	.0109	.0252	.0515	.5027	.9510	.9740	.9904
$\delta_{10,12}$	.0100	.0282	.0520	.5010	.9504	.9741	.9900
$\delta_{10,13}$	.0129	.0276	.0504	.5029	.9493	.9761	.9896
$\psi_{1,11}$	.0085	.0228	.0480	.5005	.9457	.9726	.9904
$\delta_{10,21}$	.0105	.0238	.0484	.5093	.9549	.9770	.9912
$\delta_{10,22}$	.0100	.0253	.0503	.4995	.9505	.9757	.9909
$\delta_{10,23}$	.0101	.0239	.0504	.5010	.9495	.9745	.9902
$\delta_{1,21}$	.0091	.0262	.0506	.5030	.9533	.9783	.9926
$\psi_{1,22}$	.0110	.0256	.0482	.4986	.9485	.9747	.9900
$\delta_{20,11}$	.0101	.0234	.0490	.5055	.9520	.9749	.9893
$\delta_{20,12}$	.0106	.0264	.0488	.5087	.9546	.9789	.9916
$\delta_{20,13}$	.0083	.0236	.0488	.5010	.9548	.9778	.9909
$\psi_{2,11}$	.0095	.0242	.0490	.4963	.9498	.9753	.9894
$\delta_{20,21}$	.0100	.0262	.0537	.5024	.9530	.9736	.9886
$\delta_{20,22}$	.0105	.0261	.0485	.4982	.9507	.9762	.9897
$\delta_{20,23}$	.0093	.0228	.0493	.5013	.9539	.9769	.9900
$\delta_{2,21}$	.0108	.0251	.0483	.4960	.9501	.9755	.9903
$\psi_{2,22}$	.0107	.0264	.0510	.4977	.9510	.9752	.9908
$\delta_{30,11}$	.0108	.0255	.0516	.5012	.9515	.9757	.9916
$\delta_{30,12}$	.0095	.0233	.0483	.4993	.9495	.9756	.9907
$\delta_{30,13}$	.0090	.0242	.0506	.4927	.9494	.9737	.9888
$\psi_{3,11}$	.0093	.0239	.0498	.5064	.9565	.9784	.9907
$\delta_{30,21}$	.0098	.0227	.0466	.4935	.9506	.9758	.9885
$\delta_{30,22}$	.0101	.0256	.0523	.5041	.9500	.9740	.9906
$\delta_{30,23}$	.0094	.0225	.0497	.5003	.9476	.9718	.9888
$\delta_{3,21}$	.0084	.0246	.0521	.5093	.9480	.9742	.9899
$\psi_{3,22}$	.0104	.0253	.0498	.4999	.9491	.9756	.9904
$\delta_{30,31}$	.0108	.0281	.0545	.5021	.9530	.9759	.9896
$\delta_{30,32}$	.0090	.0242	.0512	.5020	.9516	.9752	.9880
$\delta_{30,33}$	.0105	.0267	.0544	.5037	.9533	.9749	.9898
$\delta_{3,31}$	.0110	.0267	.0511	.5003	.9507	.9750	.9914
$\delta_{3,32}$	.0090	.0221	.0516	.5024	.9527	.9774	.9913
$\psi_{3,33}$	.0092	.0271	.0495	.5029	.9489	.9723	.9872

Table 12: Frequentist Coverage for Credible Intervals of $(\Xi, \boldsymbol{\Delta})$ with Jeffreys' Prior for $(\Xi, \boldsymbol{\Delta})$; $n = 20$, $m_1 = m_2 = 10,000$; $k = 3$, $p_0 = 3$, $p_1 = p_2 = 2$, $p_3 = 3$.

α	.01	.025	.05	.50	.95	.975	.99
$\psi_{0,11}$	.0000	.0000	.0000	.0375	.5221	.6589	.7914
$\delta_{0,21}$	.0020	.0068	.0190	.5059	.9809	.9935	.9989
$\psi_{0,22}$	.0000	.0001	.0004	.0875	.6647	.7734	.8787
$\delta_{0,31}$	.0030	.0102	.0248	.5009	.9722	.9891	.9967
$\delta_{0,32}$	.0032	.0086	.0248	.5012	.9747	.9889	.9977
$\psi_{0,33}$	.0001	.0013	.0038	.1753	.7870	.8695	.9349
$\delta_{10,11}$	.0138	.0328	.0622	.4974	.9361	.9646	.9838
$\delta_{10,12}$	.0131	.0306	.0581	.4967	.9314	.9635	.9819
$\delta_{10,13}$	.0171	.0342	.0623	.5024	.9362	.9683	.9854
$\psi_{1,11}$	.0236	.0542	.0982	.6340	.9731	.9885	.9967
$\delta_{10,21}$	.0184	.0401	.0688	.5074	.9276	.9596	.9798
$\delta_{10,22}$	.0197	.0430	.0758	.5055	.9306	.9610	.9817
$\delta_{10,23}$	.0200	.0439	.0720	.5019	.9317	.9610	.9812
$\delta_{1,21}$	.0187	.0436	.0772	.5093	.9317	.9615	.9810
$\psi_{1,22}$	.0596	.1094	.1778	.7431	.9878	.9947	.9983
$\delta_{20,11}$	.0164	.0349	.0608	.4989	.9395	.9682	.9857
$\delta_{20,12}$	.0143	.0331	.0630	.5056	.9415	.9698	.9863
$\delta_{20,13}$	.0134	.0340	.0627	.4993	.9392	.9690	.9863
$\psi_{2,11}$	.0252	.0575	.1045	.6450	.9760	.9902	.9976
$\delta_{20,21}$	.0177	.0413	.0758	.5014	.9258	.9596	.9817
$\delta_{20,22}$	.0177	.0359	.0673	.4992	.9311	.9616	.9836
$\delta_{20,23}$	.0191	.0435	.0791	.5041	.9313	.9620	.9825
$\delta_{2,21}$	.0200	.0408	.0728	.4996	.9288	.9587	.9780
$\psi_{2,22}$	.0587	.1129	.1840	.7490	.9868	.9938	.9985
$\delta_{30,11}$	.0123	.0316	.0563	.5017	.9449	.9700	.9866
$\delta_{30,12}$	.0125	.0290	.0556	.4969	.9454	.9721	.9878
$\delta_{30,13}$	.0122	.0298	.0557	.5033	.9450	.9743	.9892
$\psi_{3,11}$	.0169	.0380	.0713	.5676	.9659	.9834	.9940
$\delta_{30,21}$	.0178	.0369	.0665	.5051	.9330	.9619	.9819
$\delta_{30,22}$	.0164	.0355	.0661	.5038	.9359	.9647	.9846
$\delta_{30,23}$	.0169	.0371	.0669	.5050	.9375	.9684	.9867
$\delta_{3,21}$	.0180	.0404	.0687	.5034	.9332	.9640	.9830
$\psi_{3,22}$	.0366	.0782	.1402	.6962	.9839	.9922	.9971
$\delta_{30,31}$	.0205	.0443	.0783	.4981	.9205	.9546	.9788
$\delta_{30,32}$	.0210	.0446	.0764	.5075	.9257	.9588	.9790
$\delta_{30,33}$	.0219	.0421	.0756	.5014	.9256	.9550	.9778
$\delta_{3,31}$	.0231	.0455	.0768	.4989	.9220	.9538	.9768
$\delta_{3,32}$	.0204	.0426	.0733	.5019	.9202	.9547	.9786
$\psi_{3,33}$	.0876	.1518	.2286	.7991	.9922	.9963	.9986

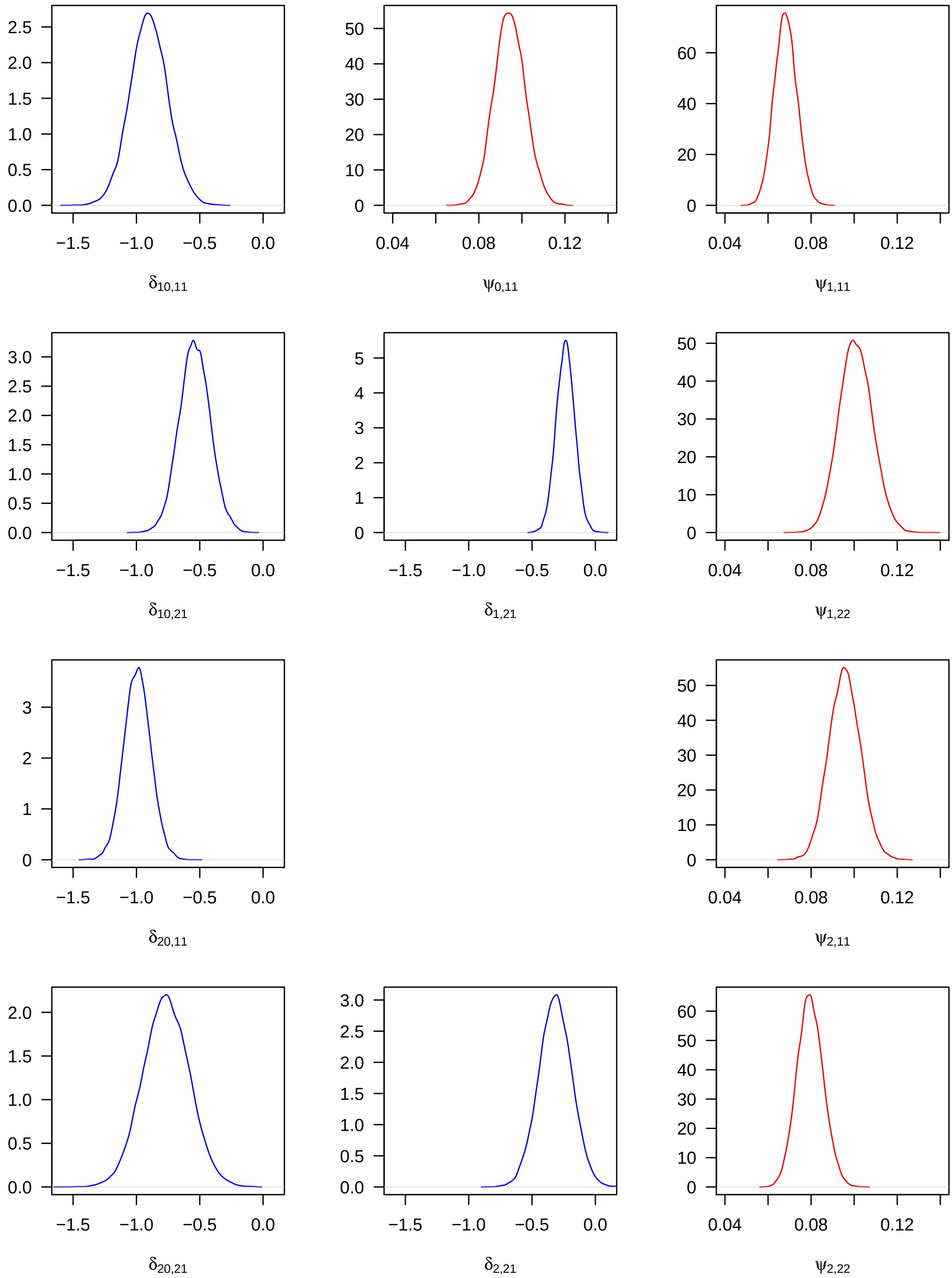

Figure 3: Marginal posterior distribution of $(\boldsymbol{\Xi},\boldsymbol{\Delta})$

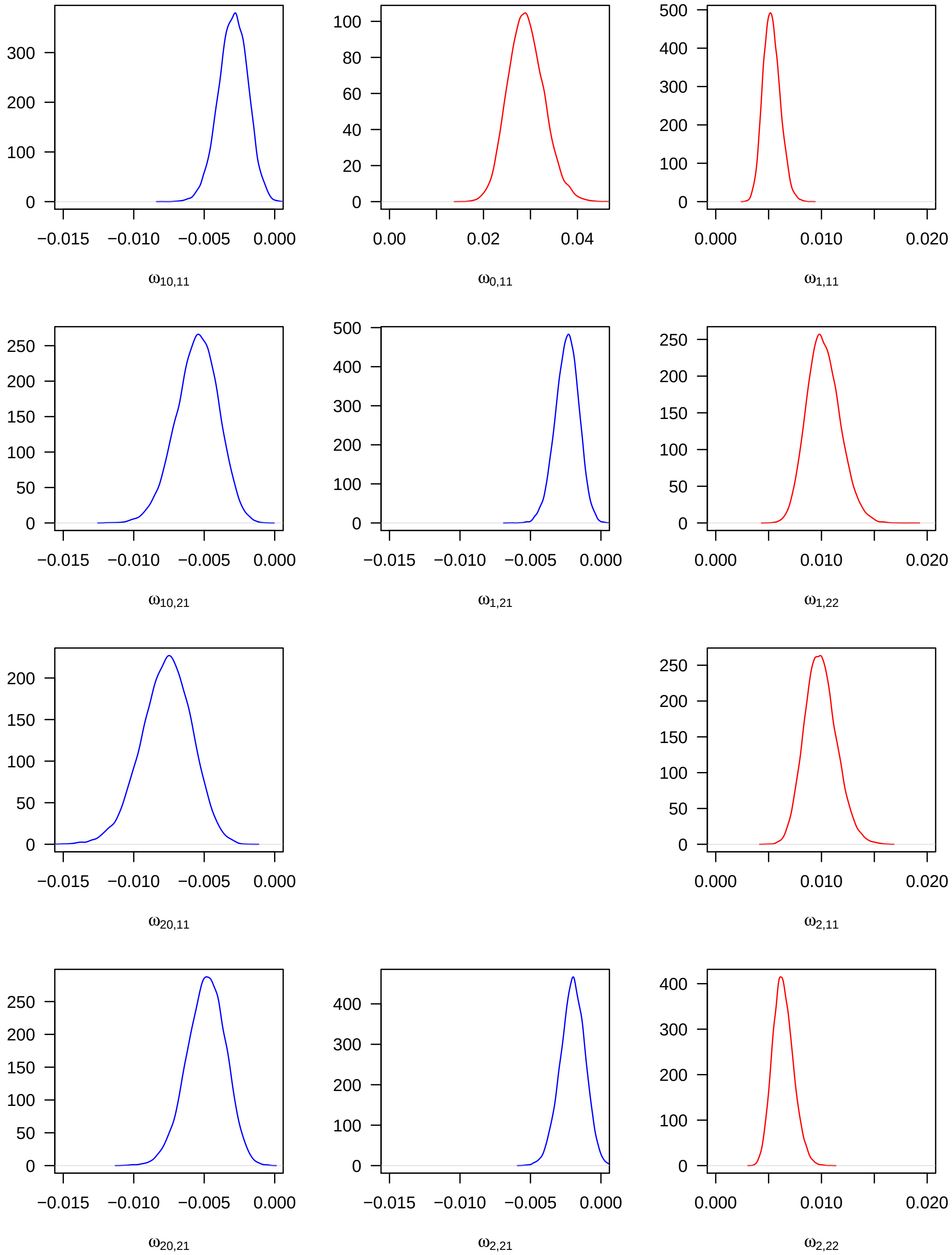

Figure 4: Marginal posterior distribution of $\mathbf{\Omega}$

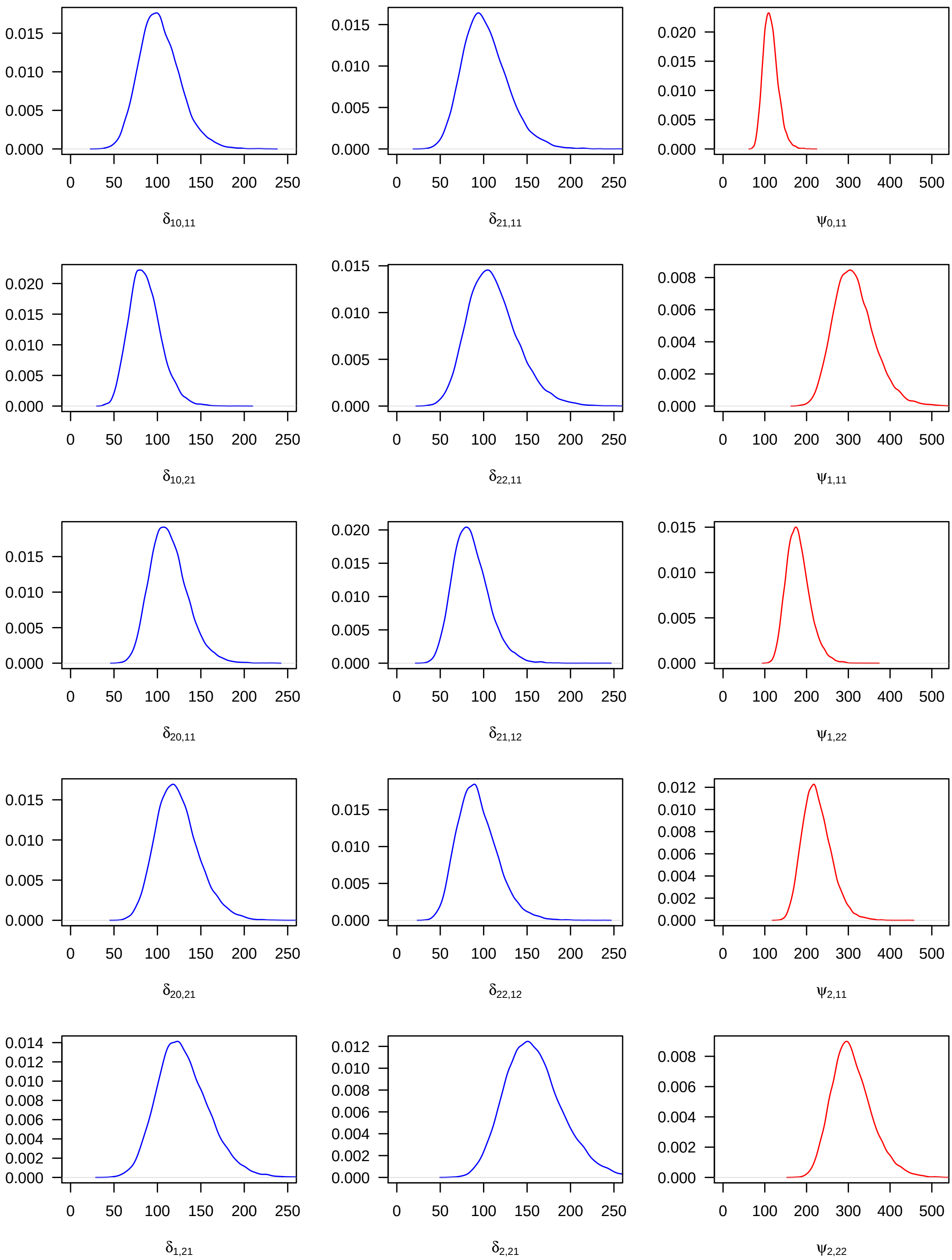

Figure 5: Marginal posterior distribution of $\mathbf{\Sigma}$

Chapter 4 Objective Bayesian Analysis of Partial Correlation Coefficients

4.16 Introduction

The Partial Correlation Coefficient (PCC) is a useful tool for verifying the conditional dependence/independence between two Gaussian variables. An application could be found in de la Fuente et al. (2004), where they used PCC for discovery of meaningful associations in genomic data. However, the frequentist distribution of the estimator of the PCC is not straightforward. In this study, we first derive an exact frequentist matching prior for PCC for Gaussian models similar to Chapter 2.10.3, then we use this prior to derive the posterior distribution of PCC. The posteriors of PCC is written in constructive random posterior form which is a function of sufficient statistics and some standard distributions. Thus the frequentist distribution of the PCC is equivalent to the constructive random posterior of PCC. Finally, we simplify the testing function for testing the null hypothesis that a PCC is zero, which implies conditional independence.

4.16.1 Problem Statement and Notation

In Chapter 2.10.3, the objective Bayesian analysis of the star-shape model was studied. The model was original introduced by Sun & Sun (2005). In this study, we continue on the track of Chapter 2.10.3 but focus on the study of PCC. We start from the PCC of x_{i1} and x_{i2} conditional on $\boldsymbol{x}_0$. From Chapter 2.10.3, we know that the joint distribution of $\boldsymbol{x}_0$ and $\boldsymbol{x}_i$ is

$$(\boldsymbol{x}_0, \boldsymbol{x}_i) \sim N\left(\mathbf{0}, \left[\begin{pmatrix} \boldsymbol{\Psi}_0 & \mathbf{0} \\ \boldsymbol{\Psi}_{i0} & \boldsymbol{\Psi}_i \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Psi}_0 & \mathbf{0} \\ \boldsymbol{\Psi}_{i0} & \boldsymbol{\Psi}_i \end{pmatrix}\right]^{\mathbf{-1}}\right). \tag{233}$$

By the theoretical properties of the multivariate normal distribution, we obtain the conditional distribution in the following fact.

Fact 4.8 *The conditional distribution of $\boldsymbol{x}_i$ given $\boldsymbol{x}_0$ is*

$$(\boldsymbol{x}_i \mid \boldsymbol{x}_0) \sim N\left(\mathbf{0}, (\boldsymbol{\Psi}_\mathbf{i}'\boldsymbol{\Psi}_\mathbf{i})^{\mathbf{-1}}\right). \tag{234}$$

Proof. The covariance of $(\boldsymbol{x}_0, \boldsymbol{x}_i)$ is

$$\begin{aligned}
\boldsymbol{\Sigma} &= \left[\begin{pmatrix} \boldsymbol{\Psi}_0 & \mathbf{0} \\ \boldsymbol{\Psi}_{i0} & \boldsymbol{\Psi}_i \end{pmatrix}' \begin{pmatrix} \boldsymbol{\Psi}_0 & \mathbf{0} \\ \boldsymbol{\Psi}_{i0} & \boldsymbol{\Psi}_i \end{pmatrix}\right]^{-1} \\
&= \begin{pmatrix} (\boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0)^{-1} & -(\boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_{i0}'\boldsymbol{\Psi}_i'^{-1} \\ -\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Psi}_{i0}(\boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0)^{-1} & (\boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i)^{-1} + \boldsymbol{\Psi}_i^{-1}\boldsymbol{\Psi}_{i0}(\boldsymbol{\Psi}_0'\boldsymbol{\Psi}_0)^{-1}\boldsymbol{\Psi}_{i0}'\boldsymbol{\Psi}_i'^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_{0i} \\ \boldsymbol{\Sigma}_{0i} & \boldsymbol{\Sigma}_{ii} \end{pmatrix}.
\end{aligned} \tag{235}$$

Thus, the conditional covariance of $\boldsymbol{x}_i|\boldsymbol{x}_0$ is

$$\boldsymbol{\Sigma}_{ii} - \boldsymbol{\Sigma}_{i0}\boldsymbol{\Sigma}_{00}^{-1}\boldsymbol{\Sigma}_{0i} = (\boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i)^{-1}. \tag{236}$$

Fact 4.8 is proved. □

After some algebra, the PCC of x_{i1} and x_{i2} conditional on $\boldsymbol{x}_0$ is derived as

$$\begin{aligned} \rho_{i,12\cdot\mathbf{0}} &= \frac{-\psi_{i,21}}{\sqrt{\psi_{i,11}^2 + \psi_{i,21}^2}} \\ &= \frac{\frac{-\psi_{i,21}}{\psi_{i,11}}}{\sqrt{1 + \left(\frac{-\psi_{i,21}}{\psi_{i,11}}\right)^2}}, \end{aligned} \tag{237}$$

and,

$$\begin{aligned} \frac{\rho_{i,12\cdot\mathbf{0}}}{\sqrt{1-\rho_{i,12\cdot\mathbf{0}}^2}} &= -\frac{\psi_{i,21}}{\psi_{i,11}} \\ &= -\delta_{i,21}. \end{aligned} \tag{238}$$

By Pourahmadi (1999), $-\delta_{i,21}$ is the regression coefficient of x_{i1} in the linear least-squares of x_{i2} based on its predecessors $x_{01}, \ldots, x_{0p_0}, x_{i1}$. This is an interesting result. It makes clear sense in that if $-\delta_{i,21} = 0$, which means that x_{i1} does not offer further information in predicting x_{i2} if $\boldsymbol{x}_0$ is already in the regression model, that is, the PCC of x_{i2} and x_{i1} conditional on $\boldsymbol{x}_0$ is 0. It is also well studied and known that under the null hypothesis $H_0: \ -\delta_{i,21} = 0$, the least square predictor of $-\delta_{i,21}$ divided by its predicted standard error follows a t-distribution with degree of freedom $n-k$, where n is sample size and k is the number of coefficients in the regression model to be predicted, including the intercept.

4.16.2 Sufficient Statistics and Notations

Let $\boldsymbol{x}_h = (\boldsymbol{x}'_{h0}, \boldsymbol{x}'_{h1}, \ldots, \boldsymbol{x}'_{hk})'$, $h = 1, \ldots, n$, be a simple random sample of size n from $N_p(\mathbf{0}, \boldsymbol{\Omega}^{-1})$, the star-shape graphical model. Here $\boldsymbol{x}_{hi} = (x_{hi1}, \ldots, x_{hip_i})', i = 0, \cdots, k$. It is well known that the non-normalized sample covariance matrix,

$$\boldsymbol{S} = \sum_{h=1}^{n} \boldsymbol{x}_h \boldsymbol{x}'_h = (\boldsymbol{S}_{ij})_{i,j=0,\ldots,k} \tag{239}$$

is a sufficient statistic for $\boldsymbol{\Omega}$ or $\boldsymbol{\Psi}$ and follows the Wishart$_p(n, \boldsymbol{\Omega})$ distribution. (See, for example, Gupta & Nagar, 2000, p88). Here $\boldsymbol{S}_{ij}$ is a $p_i \times p_j$ matrix given by

$$\boldsymbol{S}_{ij} = \sum_{h=1}^{n} \boldsymbol{x}_{hi}\boldsymbol{x}'_{hj} \equiv (s_{ij,lm})_{l=1,\cdots,p_i,m=1,\cdots,p_j}, \quad i,j = 0,\ldots,k. \tag{240}$$

Let $\boldsymbol{X}_0 = (\boldsymbol{x}_{10}, \ldots, \boldsymbol{x}_{n0})'$, define

$$\boldsymbol{H}_0 = \boldsymbol{X}_0(\boldsymbol{X}'_0\boldsymbol{X}_0)^{-1}\boldsymbol{X}'_0, \tag{241}$$

$$\boldsymbol{x}_{ij} = (x_{1ij}, x_{2ij}, \ldots, x_{nij})', \tag{242}$$

$$\boldsymbol{s}_{i10} = (s_{i0,11}, \ldots, s_{i0,1p_0})', \tag{243}$$

$$\boldsymbol{s}_{i20} = (s_{i0,21}, \ldots, s_{i0,2p_0})', \tag{244}$$

$$\widetilde{\boldsymbol{s}}_{i20} = (s_{i0,21}, \ldots, s_{i0,2p_0}, s_{i,21})', \tag{245}$$

$$\widetilde{\boldsymbol{S}}_{i,1} = \begin{pmatrix} \boldsymbol{S}_0 & \boldsymbol{s}_{i10} \\ \boldsymbol{s}'_{i10} & s_{i,11} \end{pmatrix}, \tag{246}$$

$$u_{i1} = s_{i,11} - \boldsymbol{s}'_{i10}\boldsymbol{S}_0^{-1}\boldsymbol{s}_{i10}, \tag{247}$$

$$u_{i2} = s_{i,22} - \boldsymbol{s}'_{i20}\boldsymbol{S}_0^{-1}\boldsymbol{s}_{i20}, \tag{248}$$

$$u_{i12} = s_{i,21} - \boldsymbol{s}'_{i10}\boldsymbol{S}_0^{-1}\boldsymbol{s}_{i20}, \tag{249}$$

$$\widetilde{u}_{i2} = s_{i,22} - \widetilde{\boldsymbol{s}}'_{i20}\widetilde{\boldsymbol{S}}_{0,i1}^{-1}\widetilde{\boldsymbol{s}}_{i20}. \tag{250}$$

Based on these notations, we get the sample PCC of (x_{i1} *and* $x_{i2} \mid \boldsymbol{x}_0$) in the following proposition.

Proposition 4.10

(a) *From least square linear regression,*

$$r_{i,12\cdot\mathbf{0}} = \frac{\boldsymbol{x}'_{i1}(\boldsymbol{I} - \boldsymbol{H}_0)\boldsymbol{x}_{i1}}{\sqrt{\boldsymbol{x}'_{i1}(\boldsymbol{I} - \boldsymbol{H}_0)\boldsymbol{x}_{i1}\boldsymbol{x}'_{i2}(\boldsymbol{I} - \boldsymbol{H}_0)\boldsymbol{x}_{i2}}}. \tag{251}$$

(b) *From conditional sample covariance matrix,*

$$r_{i,12\cdot\mathbf{0}} = \frac{u_{i12}}{\sqrt{u_{i1}u_{i2}}}. \tag{252}$$

Proof. For part (a), we know that the residual of least square linear regression of x_{ij} on $\boldsymbol{x}_0$ is $(\boldsymbol{I}-\boldsymbol{H}_0)\boldsymbol{x}_{ij}$ and $(\boldsymbol{I}-\boldsymbol{H}_0)$ is idempotent. The result is straightforward from the definition of PCC.

For part(b), the sample covariance matrix of $(\boldsymbol{x}_0, x_{i1}, x_{i2})$ is

$$\left(\begin{array}{c|cc} \boldsymbol{S}_{00} & \boldsymbol{s}_{i10} & \boldsymbol{s}_{i20} \\ \hline \boldsymbol{s}'_{i10} & s_{i11} & s_{i12} \\ \boldsymbol{s}'_{i20} & s_{i21} & s_{i22} \end{array}\right). \tag{253}$$

Thus, the conditional sample covariance of $(x_{i1}, x_{i2}|\boldsymbol{x}_0)$ is

$$\begin{pmatrix} s_{i11} & s_{i12} \\ s_{i21} & s_{i22} \end{pmatrix} - \begin{pmatrix} \boldsymbol{s}'_{i10} \\ \boldsymbol{s}'_{i20} \end{pmatrix} \boldsymbol{S}_{00}^{-1} \begin{pmatrix} \boldsymbol{s}_{i10} & \boldsymbol{s}_{i20} \end{pmatrix} = \begin{pmatrix} u_{i1} & u_{i12} \\ u_{i12} & u_{i2} \end{pmatrix}. \tag{254}$$

Therefore, the sample PCC in the part (b) follows. □

Fact 4.9 *The relationship of $\widetilde{u}_{i2}$, u_{i2}, and $r_{i,12\cdot\mathbf{0}}$ is given by*

$$\widetilde{u}_{i2} = u_{i2}(1 - r^2_{i,12\cdot\mathbf{0}}). \tag{255}$$

Proof.

$$\begin{aligned}
\widetilde{u}_{i2} &= s_{i,22} - \begin{pmatrix} \boldsymbol{s}'_{i20} & s_{i21} \end{pmatrix} \widetilde{\boldsymbol{S}}_{i,1}^{-1} \begin{pmatrix} \boldsymbol{s}_{i20} \\ s_{i21} \end{pmatrix} \\
&= s_{i,22} - \begin{pmatrix} \boldsymbol{s}'_{i20} & s_{i21} \end{pmatrix} \begin{pmatrix} \boldsymbol{S}_0 & \boldsymbol{s}_{i10} \\ \boldsymbol{s}'_{i10} & s_{i,11} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{s}_{i20} \\ s_{i21} \end{pmatrix} \\
&= u_{i2}\left(1 - \frac{u_{i12}^2}{u_{i1}u_{i2}}\right) \\
&= u_{i2}(1 - r^2_{i,12\cdot\mathbf{0}}).
\end{aligned} \tag{256}$$

□

Consider the same class of priors as given in (170)

$$\pi(\boldsymbol{\Psi}) \propto \prod_{i=0}^{k}\prod_{j=1}^{p_i} \psi_{i,jj}^{-a_{ij}}, \tag{257}$$

From Theorem 3.2 (c) of Chapter 2.10.3, we get the constructive random posterior of $\psi_{i,21}$ as

$$\psi^*_{i,11} = \sqrt{\frac{\chi^{2*}_{f_{i1}}}{u_{i1}}}, \tag{258}$$

$$\psi^*_{i,21} = -\sqrt{\frac{\chi^{2*}_{f_{i2}}}{u_{i1}}} \frac{r_{i,12\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,12\cdot\mathbf{0}}}} + \frac{z^*_{i,1}}{\sqrt{u_{i1}}}, \tag{259}$$

$$\begin{aligned} Y^* &= \frac{-\psi^*_{i,21}}{\psi^*_{i,11}} \\ &= \sqrt{\frac{\chi^{2*}_{f_{i2}}}{\chi^{2*}_{f_{i1}}}} \frac{r_{i,12\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,12\cdot\mathbf{0}}}} + \frac{z^*_{i,1}}{\sqrt{\chi^{2*}_{f_{i1}}}}, \end{aligned} \tag{260}$$

where, $f_{ij} = n - a_{ij} + 1$. Plug $\psi^*_{i,11}$ and $\psi^*_{i,21}$ into (261), we get the constructive random posterior

$$\begin{aligned} \rho^*_{i,12\cdot\mathbf{0}} &= \frac{\frac{-\psi^*_{i,21}}{\psi^*_{i,11}}}{\sqrt{1+\left(\frac{-\psi^*_{i,21}}{\psi^*_{i,11}}\right)^2}} \\ &= \frac{Y^*}{\sqrt{1+Y^{*2}}} \\ &= \phi(Y^*). \end{aligned} \tag{261}$$

4.17 Exact Frequentist Matching Property

To verify the frequentist matching property of the random posterior $\rho^*_{i,12\cdot\mathbf{0}}$, we derive the frequentist distribution for the statistics involved in the expression of $\rho^*_{i,12\cdot\mathbf{0}}$, which is $\frac{r_{i,12\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,12\cdot\mathbf{0}}}}$.

Define

$$c_{ij} = \begin{cases} n-j+1, & \text{if } i=0,\ j=1,\ldots,p_0, \\ n-p_0-j+1, & \text{if } i=1,\ldots,k,\ j=1,\ldots,p_i. \end{cases} \tag{262}$$

Lemma 4.11 *We have the distribution*

$$\frac{r_{i,12\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,12\cdot\mathbf{0}}}} = \sqrt{\frac{\chi^2_{c_{i1}}}{\chi^2_{c_{i2}}}}\frac{\rho_{i,12\cdot\mathbf{0}}}{\sqrt{1-\rho^2_{i,12\cdot\mathbf{0}}}} + \frac{z_{i,1}}{\sqrt{\chi^2_{c_{i2}}}}. \tag{263}$$

Proof. After some matrix algebra from Lemma 3.10 (b), we have

$$\left(\frac{u_{i12}}{u_{i1}}|\psi_{i,11},\psi_{i,21},\psi_{i,22},u_{i1}\right) \sim N\left(-\frac{\psi_{i,21}}{\psi_{i,22}}, \frac{1}{\psi^2_{i,22}u_{i1}}\right), \tag{264}$$

and u_{i1} is independent of $\tilde{u}_{i2}$

From (251) - (255),

$$\begin{aligned}
\frac{r_{i,12\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,12\cdot\mathbf{0}}}} &= \frac{u_{i12}}{\sqrt{u_{i1}}}\frac{1}{\sqrt{\widetilde{u}_{i2}}} \\
&= \frac{u_{i12}}{u_{i1}}\sqrt{\frac{u_{i1}}{\widetilde{u}_{i2}}} \\
&= \left(-\frac{\psi_{i,21}}{\psi_{i,22}} + \frac{z_{i,1}}{\psi_{i,22}\sqrt{u_{i1}}}\right)\frac{\psi_{i,22}\sqrt{u_{i1}}}{\sqrt{\chi^2_{c_{i2}}}} \\
&= \sqrt{\frac{\chi^2_{c_{i1}}}{\chi^2_{c_{i2}}}}\left(\frac{-\psi_{i,21}}{\psi_{i,11}}\right) + \frac{z_{i,1}}{\sqrt{\chi^2_{c_{i2}}}}. \qquad (265)
\end{aligned}$$

Thus Lemma 4.11 is proved. □

Theorem 4.6 *The exact matching property of posterior PCC:*

(a) *For any* $\alpha \in [0,1]$*,*

$$P(\rho_{i,12\cdot\mathbf{0}} < [\rho^*_{i,12\cdot\mathbf{0}}]_\alpha) = \alpha, \tag{266}$$

if and only if $a_{i1} = p_0 + 1$ *and* $a_{i2} = p_0 + 2$ *in the class of priors given in (257).*

(b) *The right-Haar prior given in (151) is a exact Frequentist matching prior for* $\rho_{i,12\cdot\mathbf{0}}$*.*

Proof. For part (a) by definition of f_{ij} and c_{ij}, we know that $f_{i1} = c_{i1}$ and $f_{i2} = c_{i2}$ if and only if $a_{i1} = p_0 + 1$ and $a_{i2} = p_0 + 2$. The following equations will hold, if and only if

$f_{i1} = c_{i1}$ and $f_{i2} = c_{i2}$,

$$\begin{aligned}
& P\left(\rho_{i,12\cdot\mathbf{0}} < \left[\rho^*_{i,12\cdot\mathbf{0}}\right]_\alpha \mid \mathbf{\Psi}\right) \\
= \;& P\left(\rho_{i,12\cdot\mathbf{0}} < [\phi(Y^*)]_\alpha\right) \\
= \;& P\left(\phi^{-1}(\rho_{i,12\cdot\mathbf{0}}) < [Y^*]_\alpha\right) \\
= \;& P\left(\frac{\rho_{i,12\cdot\mathbf{0}}}{\sqrt{1-\rho^2_{i,12\cdot\mathbf{0}}}} < \left[\sqrt{\frac{\chi^{2*}_{f_{i2}}}{\chi^{2*}_{f_{i1}}}}\frac{r_{i,12\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,12\cdot\mathbf{0}}}} + \frac{z^*_{i,1}}{\sqrt{\chi^{2*}_{f_{i1}}}}\right]_\alpha\right) \\
= \;& P\left(-\sqrt{\frac{\chi^2_{c_{i1}}}{\chi^2_{c_{i2}}}}\frac{\rho_{i,12\cdot\mathbf{0}}}{\sqrt{1-\rho^2_{i,12\cdot\mathbf{0}}}} + \frac{z_{i,1}}{\sqrt{\chi^2_{c_{i2}}}} < \left[-\sqrt{\frac{\chi^{2*}_{c_{i1}}}{\chi^{2*}_{c_{i2}}}}\frac{\rho_{i,12\cdot\mathbf{0}}}{\sqrt{1-\rho^2_{i,12\cdot\mathbf{0}}}} + \frac{z^*_{i,1}}{\sqrt{\chi^{2*}_{c_{i2}}}}\right]_\alpha\right) \\
= \;& \alpha. \qquad (267)
\end{aligned}$$

For part (b), the right-Haar prior for $\mathbf{\Psi}$ given in (151) has $a_{i1} = p_0 + 1$ and $a_{i2} = p_0 + 2$. □

Corollary 4.6 *By switching the pair of interesting variables x_{ij} and x_{il} to the position of x_{i1} and x_{i2} conditional on $\boldsymbol{x}_0$, the corresponding right-Haar prior will be the exact matching priors for the PCC of this new pair of variables. Equivalently, we can find the frequentist distributions of PCC for any pair of x_{ij} and x_{il}, WLOG, assume $j < l$, by using the constructive random posteriors formula*

$$\rho^*_{i,jl\cdot\mathbf{0}} = \frac{Y^*}{\sqrt{1+Y^{*2}}}, \qquad (268)$$

$$Y^* = \sqrt{\frac{\chi^{2*}_{c_{i2}}}{\chi^{2*}_{c_{i1}}}}\frac{r_{i,jl\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,jl\cdot\mathbf{0}}}} + \frac{z^*_{i,1}}{\sqrt{\chi^{2*}_{c_{i1}}}}. \qquad (269)$$

4.18 Use PCC as a Tool for Testing Conditional Independence

Based on the above posterior distribution of PCC, as it is exactly frequentist matching under the right-Haar prior, we can find cut-offs for testing that a PCC is zero in frequentist way.

Proposition 4.11 *For any pair of x_{ij} and x_{il}, the one-sided $1-\alpha$ level test for H_0 : $\rho_{i,jl\cdot\mathbf{0}} < 0$ is to reject null hypotheses when* $\left[\sqrt{\chi^{*2}_{ci2}}\frac{r_{i,jl\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,jl\cdot\mathbf{0}}}} + z^*_{i,1}\right]_\alpha > 0$.

Proof.

$$\begin{aligned}
& \left[\rho^*_{i,jl\cdot\mathbf{0}}\right]_\alpha > 0 \\
\Leftrightarrow\ & \left[\phi(Y^*)\right]_\alpha > 0 \\
\Leftrightarrow\ & \left[Y^*\right]_\alpha > 0 \\
\Leftrightarrow\ & \left[\sqrt{\frac{\chi^{*2}_{ci2}}{\chi^{*2}_{ci1}}}\frac{r_{i,jl\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,jl\cdot\mathbf{0}}}} + \frac{z^*_{i,1}}{\sqrt{\chi^{*2}_{ci1}}}\right]_\alpha > 0 \\
\Leftrightarrow\ & \left[\sqrt{\chi^{*2}_{ci2}}\frac{r_{i,jl\cdot\mathbf{0}}}{\sqrt{1-r^2_{i,jl\cdot\mathbf{0}}}} + z^*_{i,1}\right]_\alpha > 0. \qquad (270)
\end{aligned}$$

□

From de la Fuente et al. (2004), the order of PCC is defined by the number of variables that are conditioned on. Thus $\rho_{i,jl\cdot\mathbf{0}}$ is the $p_0^{th}-order$ PCC, as it is conditional on p_0 variables. Using the constructive random posterior of PCC given above and (252), the posterior PCC can be simulated to any arbitrary order quickly. (271)-(273) allow the calculation of sample PCC of orders $0-2$ as similar equations exist to calculate higher-order sample PCCs.

$$\text{sample correlation: } r_{xy} = \frac{cov(xy)}{var(x)var(y)} \qquad (271)$$

$$\text{first-order correlation: } r_{xy\cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{(1-r^2_{xz})(1-r^2_{yz})} \qquad (272)$$

$$\text{second-order correlation: } r_{xy\cdot zq} = \frac{r_{xy\cdot z} - r_{xq\cdot z}r_{xq\cdot z}}{(1-r^2_{xq\cdot z})(1-r^2_{yq\cdot z})} \qquad (273)$$

4.19 Application

In the previous Chapter, we use the mathematics marks example, where we start with the assumption that conditional on (Algebra), (Mechanics, Vectors) are independent of

(Analysis, Statistics). Define x_0=Algebra, x_{11}=Mechanics, x_{12}=Vectors, x_{21}=Analysis, and x_{22}=Statistics. This assumption implies that

$$H_{01} : \rho_{1,12\cdot\mathbf{0}} \neq 0; \tag{274}$$

$$H_{02} : \rho_{2,12\cdot\mathbf{0}} \neq 0; \tag{275}$$

$$H_{03} : \rho_{(1,j)(2,l)\cdot\mathbf{0}} = 0, \quad j, l = 1, 2. \tag{276}$$

The sample size of this mathematics marks data is $n = 88$. One degree of freedom is used to estimate $(u_{01}, u_{11}, u_{12}, u_{21}, u_{22})$, thus the degree of freedom left for estimation of covariance matrix is 87. The sample covariance matrix of (Algebra, Mechanics, Vectors, Analysis, Statistics) is

$$\boldsymbol{S} = \left(\begin{array}{c|cc|cc} 111.60 & & & & \\ \hline 100.43 & 302.29 & & & \\ 84.19 & 125.78 & 170.88 & & \\ \hline 110.84 & 105.07 & 93.60 & 217.88 & \\ 120.49 & 116.07 & 97.89 & 153.77 & 294.37 \end{array}\right) \tag{277}$$

The sample correlation coefficients are

$$\boldsymbol{r} = \left(\begin{array}{c|cc|cc} 1.0000000 & & & & \\ \hline 0.5467886 & 1.0000000 & & & \\ 0.6096531 & 0.5534190 & 1.0000000 & & \\ \hline 0.7108136 & 0.4094097 & 0.4850891 & 1.0000000 & \\ 0.6647711 & 0.3891001 & 0.4364615 & 0.6071782 & 1.0000000 \end{array}\right) \tag{278}$$

The sample PCC ($r_{i,jl\cdot\mathbf{0}}$) conditional on Algebra is

$$\left(\begin{array}{c|cc|cc} \textbf{Algebra} & Mechanics & Vectors & Analysis & Statistics \\ \hline Mechanics & 1.0000000 & & & \\ Vectors & 0.3315880 & 1.0000000 & & \\ \hline Analysis & 0.0352254 & 0.0927982 & 1.0000000 & \\ Statistics & 0.0409458 & 0.0526575 & 0.2562522 & 1.0000000 \end{array}\right) \tag{279}$$

See Figure 6 for a graphical representation of sample PCC conditional on Algebra using balloon plots. The plot shows that we are not able to reject $H_{03} : \rho_{(1,j)(2,l)\cdot\mathbf{0}} = 0, j, l = 1, 2$. For further information we also simulate the posterior distribution of the PCC from (268), see Figure 7, which offers more evidence on the correctness of our assumption $H_{01} - H_{03}$.

The ratio $(\frac{r_{i,jl\cdot\mathbf{0}}}{1-r^2_{i,jl\cdot\mathbf{0}}})$ of sample PCC conditional on Algebra is

$$\left(\begin{array}{c|cc|cc} \textbf{Algebra} & Mechanics & Vectors & Analysis & Statistics \\ \hline Mechanics & 1.0000000 & & & \\ Vectors & 0.3514728 & 1.0000000 & & \\ \hline Analysis & 0.0352473 & 0.0932004 & 1.0000000 & \\ Statistics & 0.0409802 & 0.0527307 & 0.2651040 & 1.0000000 \end{array}\right). \tag{280}$$

These are the statistics that are directly involved in the constructive random posterior of PCCs. We use them in our simulation of the density of posterior PCCs. See Figure 7. The figure shows clearly the dependence/independence structure among mathematical marks conditioned on Algebra, which confirm the graphical assumption used in previous chapter.

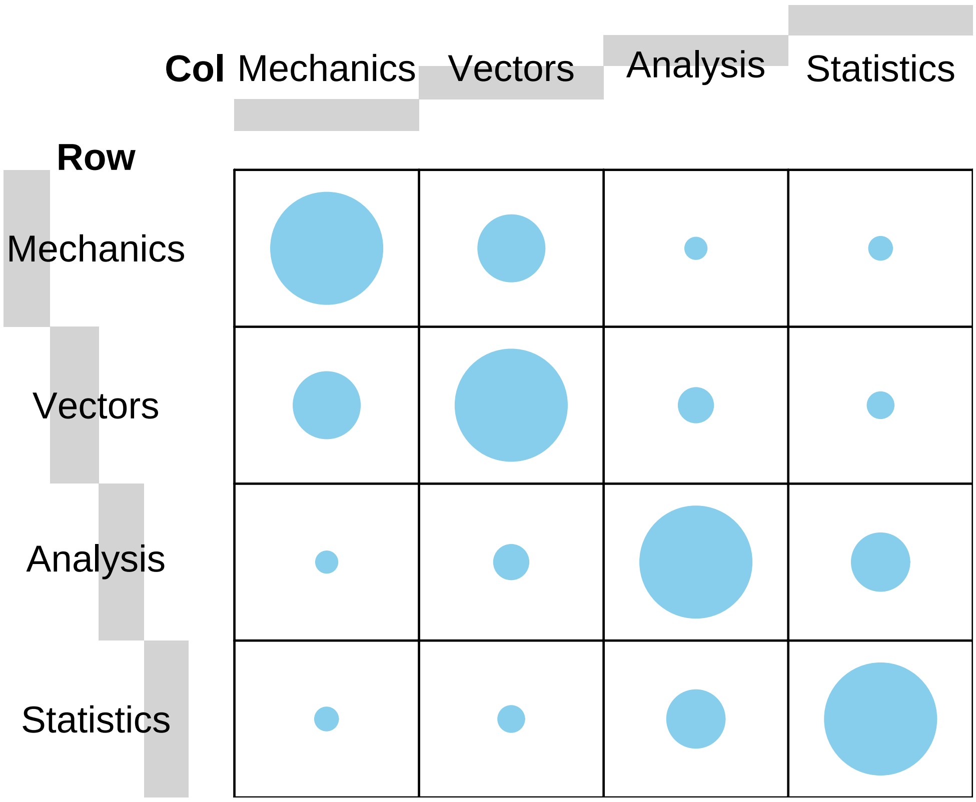

Figure 6: Balloon plot of sample PCC for mathematical marks conditional on algebra

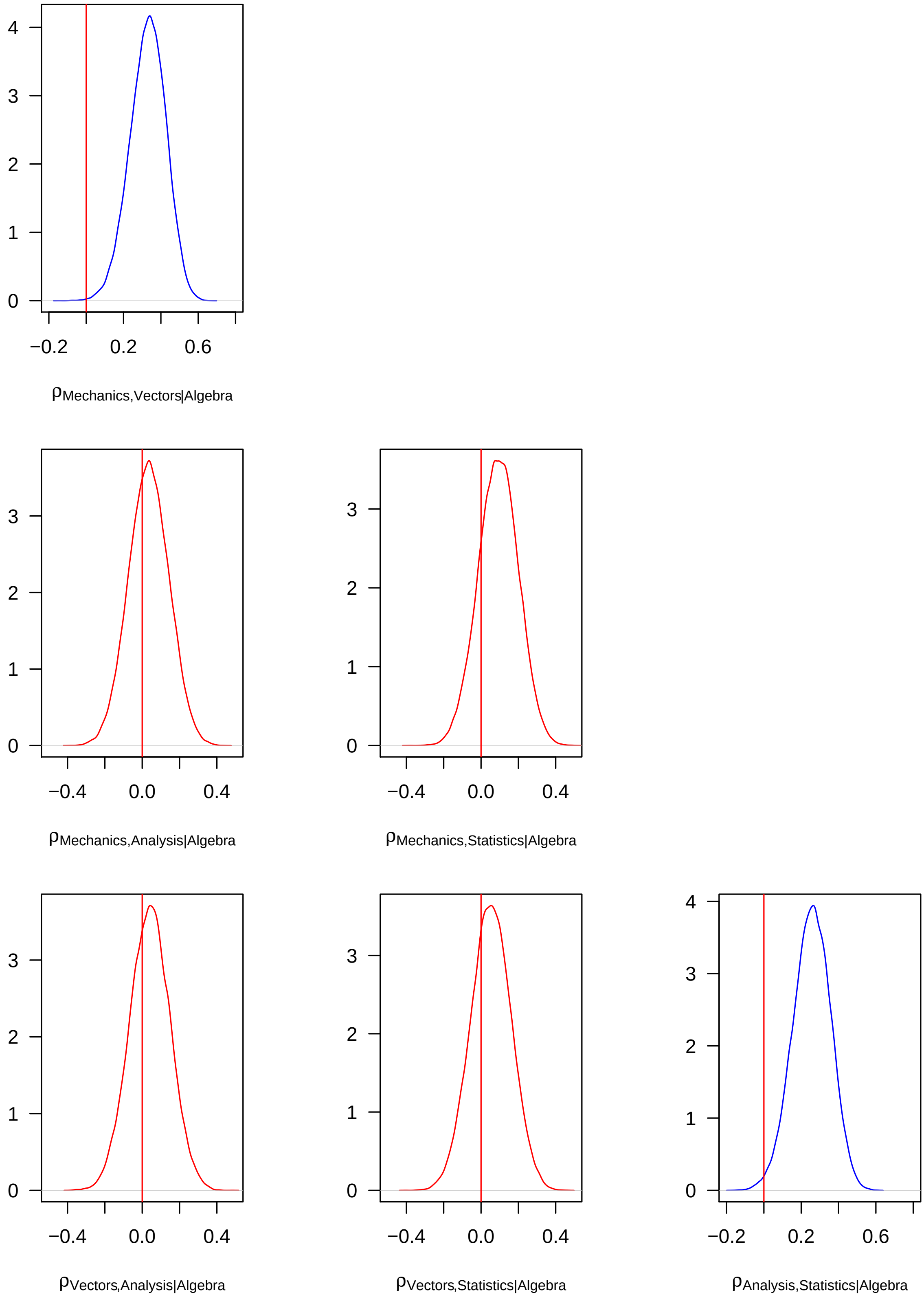

Figure 7: Posterior PCC for mathematical marks

Bibliography

Bayes, T. (2001), 'An essay towards solving a problem in the doctrine of chances', *Rev. R. Acad. Cienc. Exactas Fís. Nat. (Esp.)* **95**(1-2), 61–80. Translated from the English original [R. Soc. Lond. Philos. Trans. **53** (1763), 370–418] by Miguel Ángel Gómez Villegas, Francisco Javier Girón González-Torre, María Lina Martínez García and David Ríos Insua.

Berger, J. O. & Bernardo, J.-M. (1989), 'Estimating a product of means: Bayesian analysis with reference priors', *J. Amer. Statist. Assoc.* **84**(405), 200–207.

Berger, J. O. & Bernardo, J. M. (1992a), 'Ordered group reference priors with application to the multinomial problem', *Biometrika* **79**(1), 25–37.

Berger, J. O. & Bernardo, J. M. (1992b), Reference priors in a variance components problem, *in* 'Bayesian analysis in statistics and econometrics (Bangalore, 1988)', Vol. 75 of *Lecture Notes in Statist.*, Springer, New York, pp. 177–194.

Berger, J. O. & Bernardo, J. M. (1992c), *On the development of reference priors (with discussion)*, Bayesian Statistics, 4, Oxford University Press.

Berger, J. O., Boukai, B. & Wang, Y. (1997), 'Unified frequentist and Bayesian testing of a precise hypothesis', *Statist. Sci.* **12**(3), 133–160. With comments by Dennis V. Lindley, Thomas A. Louis and David Hinkley and a rejoinder by the authors.

Berger, J. O., Boukai, B. & Wang, Y. (1999), 'Simultaneous Bayesian-frequentist sequential testing of nested hypotheses', *Biometrika* **86**(1), 79–92.

Berger, J. O. & Sun, D. C. (2008), 'Objective priors for the bivariate normal model', *Ann. Statist.* **36**(2), 963–982.

Bernardo, J.-M. (1979), 'Reference posterior distributions for Bayesian inference', *J. Roy. Statist. Soc. Ser. B* **41**(2), 113–147. With discussion.

Casella, G. (1985), 'An introduction to empirical Bayes data analysis', *Amer. Statist.* **39**(2), 83–87.

Casella, G. & Berger, R. L. (1987), 'Reconciling Bayesian and frequentist evidence in the one-sided testing problem', *J. Amer. Statist. Assoc.* **82**(397), 106–111, 123–139. With comments and a rejoinder by the authors.

Cox, D. R. (2005), 'Frequentist and bayesian statistics: a critique', *Proceeding of PHYSTAT05* .

URL: *http://www.physics.ox.ac.uk/phystat05/proceedings/files/papbayesrev.pdf*

Dass, S. C. & Berger, J. O. (2003), 'Unified conditional frequentist and Bayesian testing of composite hypotheses', *Scand. J. Statist.* **30**(1), 193–210.

Datta, G. S. & Ghosh, M. (1996), 'On the invariance of noninformative priors', *Ann. Statist.* **24**(1), 141–159.

Datta, G. S. & Mukerjee, R. (2004), *Probability matching priors: higher order asymptotics*, Vol. 178 of *Lecture Notes in Statistics*, Springer-Verlag, New York.

de la Fuente, A., Bing, N., Hoeschele, I. & Mendes, P. (2004), 'Discovery of meaningful associations in genomic data using partial correlation coefficients', *Bioinformatics* **20**(18), 3565–3574.

URL: *http://bioinformatics.oxfordjournals.org/cgi/content/abstract/20/18/3565*

Eaton, M. L. (1989), *Group Invariance Applications in Statistics*, Institute of Mathematical Statistics and American Statistical Association.

Efron, B. (2005), 'Bayesians, frequentists, and scientists', *J. Amer. Statist. Assoc.* **100**(469), 1–5.

Ghosh, J. K. & Mukerjee, R. (1993), 'On priors that match posterior and frequentist distribution functions', *Canad. J. Statist.* **21**(1), 89–96.

Ghosh, M. & Mukerjee, R. (1998), Recent developments on probability matching priors, *in* 'Applied statistical science, III', Nova Sci. Publ., Commack, NY, pp. 227–252.

Good, I. J. (1992), 'The Bayes/non-Bayes compromise: a brief review', *J. Amer. Statist. Assoc.* **87**(419), 597–606.

Gupta, A. K. & Nagar, D. K. (2000), *Matrix variate distributions*, Vol. 104 of *Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics*, Chapman & Hall/CRC, Boca Raton, FL.

Haff, L. R. (1980), 'Empirical Bayes estimation of the multivariate normal covariance matrix', *Ann. Statist.* **8**(3), 586–597.

Jeffreys, H. (1961), *Theory of probability*, Third edition, Clarendon Press, Oxford.

Kass, R. E. & Wasserman, L. (1996), 'The selection of prior distributions by formal rules', *J. Amer. Statist. Assoc.* **91**(435), 1343–1370.
URL: *http://www.jstor.org/stable/2291752*

Krishnamoorthy, K. & Gupta, A. K. (1989), 'Improved minimax estimation of a normal precision matrix', *Canad. J. Statist.* **17**(1), 91–102.

Laplace, P. S. (1995), *A philosophical essay on probabilities*, english edn, Dover Publications Inc., New York. Translated from the sixth French edition by Frederick William Truscott and Frederick Lincoln Emory, With an introductory note by E. T. Bell.

Lauritzen, S. L. (1996), *Graphical models*, Vol. 17 of *Oxford Statistical Science Series*, The Clarendon Press Oxford University Press, New York. , Oxford Science Publications.

Little, R. J. (2006), 'Calibrated Bayes: a Bayes/frequentist roadmap', *Amer. Statist.* **60**(3), 213–223.

Pourahmadi, M. (1999), 'Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation', *Biometrika* **86**(3), 677–690.

Samaniego, F. J. & Reneau, D. M. (1994), 'Toward a reconciliation of the Bayesian and frequentist approaches to point estimation', *J. Amer. Statist. Assoc.* **89**(427), 947–957.

Sinha, B. K. & Ghosh, M. (1987), 'Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss', *Statist. Decisions* **5**(3-4), 201–227.

Sun, D. & Berger, J. O. (2007), Objective priors for the multivariate normal model, *in* 'Bayesian statistics, 8 (Peñíscola, 2007)', Oxford Univ. Press, New York, pp. 525–566.

Sun, D. & Sun, X. (2005), 'Estimation of the multivariate normal precision and covariance matrices in a star-shape model', *Ann. Inst. Statist. Math.* **57**(3), 455–484.

Whittaker, J. (1990), *Graphical models in applied multivariate statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester.

Yang, R.-y. & Berger, J. O. (1994), 'Estimation of a covariance matrix using the reference prior', *Ann. Statist.* **22**(3), 1195–1211.