

1 Introduction

1.1 Motivation

We consider the maximum likelihood estimation of the correlation matrix of a multivariate normal distribution under order restrictions among component correlations. For example we want to estimate the correlation matrix of the four variables X_1, X_2, X_3 and X_4 where X_1 denotes the test scores in an entrance examination acquired by students and X_2, X_3 and X_4 denote respectively the test scores acquired by the same student in the pre-tests executed in one month, two months and three months ago respectively. Putting ρ_{ij} the correlation of the variable X_i and X_j it seems to be natural to assume the following two types of monotonicity among component correlations,

$$H_c : \rho_{12} > \rho_{13} > \rho_{14} \quad \text{and} \quad \rho_{23} > \rho_{24} \quad (1)$$

and

$$H_r : \rho_{13} < \rho_{23} \quad \text{and} \quad \rho_{14} < \rho_{24} < \rho_{34} \quad (2)$$

Thus the situation where we now consider seems to be considerably realistic. Here in general by H_c we denotes the column decreasing hypothesis, that is, $\rho_{i,j} > \rho_{i,j+1}$ for all $1 \leq i \leq p$ and $1 \leq j \leq p-1$ and by H_r the row increasing hypothesis, that is, $\rho_{i,j} < \rho_{i+1,j}$ for all $1 \leq i \leq p-1$ and $1 \leq j \leq p$. Further we put $H_1 = H_r \cap H_c$. Further we denote by H_0 "no hypothesis", that is, the space is the set of all correlation matrices and by H_2 we denote $H_1 \cap \{\rho_{ij} > 0 \text{ for all } i, j\}$. Thus we have the following decreasing sequence of the hypotheses (or the sets represented by the hypotheses)

$$H_0 \rightarrow \{H_c, H_r\} \rightarrow H_1 = H_r \cap H_c \rightarrow H_2. \quad (3)$$

The problem is to obtain the mle for the correlation matrix under each hypothesis. Though for simplicity we treat the hypothesis H_0 and H_2 in this paper the other hypotheses are treated in a similar way.

1.2 Likelihood function

In this paper we assume the vector $X = (X_1, X_2, \dots, X_p)$ has the p dimensional normal distribution $\mathcal{N}_p(0, \Sigma)$ and that we have a sample of size N from this distribution. Then the sufficient static for Σ is the sample covariance matrices $S = (s_{ij})$ where $s_{i,j} = \sum_{k=1}^N X_{ik} X_{jk} / N$ and X_{ik} is the k -th sample of the variable X_i . Let $W = NS$. Then W has the Wishart distribution, $W(N, p, \Sigma)$. The density of $W(N, p, \Sigma)$ is given by

$$\frac{1}{2^{Np/2} \cdot p(\frac{N}{2}) |\Sigma|^{\frac{N}{2}}} \text{etr}(-\frac{1}{2} \Sigma^{-1} W) |W|^{(N-p-1)/2}. \quad (4)$$

Now let R denote the sample correlation matrix, that is,

$$R = (r_{ij}), \quad r_{ij} = \frac{w_{ij}}{\sqrt{w_{ii}w_{jj}}} \quad (5)$$

and let $v_i = w_{ii}$ for $i = 1, 2, \dots, p$ and $V = (v_1, v_2, \dots, v_p)$. Then the transformation of the variables

$$W = (w_{ij}) \rightarrow (R, V) \quad (6)$$

has the Jacobian

$$dW = \bigwedge_{i \leq j}^p dw_{ij} = \prod_{i=1}^p v_i^{(p-1)/2} \bigwedge_{i < j}^p dr_{ij} \bigwedge_{i=1}^p dv_i. \quad (7)$$

We have the decomposition

$$W = \text{diag}(\sqrt{v_1}, \sqrt{v_2}, \dots, \sqrt{v_p}) R \text{diag}(\sqrt{v_1}, \sqrt{v_2}, \dots, \sqrt{v_p}) \quad (8)$$

$$= V^{1/2} R V^{1/2}, \quad (9)$$

where $V = \text{diag}(v_1, v_2, \dots, v_p)$. Note that for simplicity we use the same symbol V for the diagonal matrix $\text{diag}(v_1, v_2, \dots, v_p)$ and the vector (v_1, v_2, \dots, v_p) and that this usage will appear in other parts again without mentioning it.

Then the joint density of (R, V) is given by

$$\frac{1}{2^{Np/2} \cdot {}_p(N/2) |\Sigma|^{N/2}} \text{etr} \left(-\frac{1}{2} \Sigma^{-1} V^{1/2} R V^{1/2} \right) |W|^{(N-p-1)/2} \prod_{i=1}^p v_i^{(p-1)/2}.$$

Here we consider the corresponding decomposition of the population covariance matrix Σ into the population correlation matrix, and the population variances $\Delta = (\sigma_{11}, \sigma_{22}, \dots, \sigma_{pp})$, that is,

$$\Sigma = \text{diag}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \dots, \sqrt{\sigma_{pp}}), \quad \text{diag}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \dots, \sqrt{\sigma_{pp}}) \quad (10)$$

$$= \Delta^{1/2}, \quad \Delta^{1/2}. \quad (11)$$

Then we have the following log likelihood function for, and Δ

$$-\frac{N}{2} \log |, | + N \log \prod_{i=1}^p t_i - N \log \prod_{i=1}^p \sqrt{v_i} - \frac{1}{2} \text{trace}(\Sigma^{-1} T R T), \quad (12)$$

where $t_i = \sqrt{\frac{v_i}{\sigma_{ii}}} = \sqrt{\frac{N s_{ii}}{\sigma_{ii}}}$ and $T = \Delta^{-1/2} D^{1/2} = \text{diag}(t_1, \dots, t_p)$.

As far as the maximization it suffices to consider the following equivalent function with the likelihood function

$$\ell(\Sigma, \Delta, t) = -\frac{N}{2} \log |, | + N \log \prod_{i=1}^p t_i - \frac{1}{2} \text{trace}(\Sigma^{-1} T R T). \quad (13)$$

Note that in this paper we call the function $\ell(\Sigma, \Delta, t)$ of (13) as the likelihood function.

1.3 Maximization of the function $\ell(\Gamma, t)$

We need to maximize $\ell(\cdot, \cdot, t)$, that is, to take

$$\sup_{\Gamma \in H_*} \sup_{t > 0} \ell(\cdot, \cdot, t) \quad (14)$$

where H_* denotes any hypothesis in Subsection 1.1 and $t > 0$ means that $t_i > 0$ for all i . In order to take the above maximization we execute the following two maximization processes iteratively.

(A) Correlation maximization process

For a given t^i , the i -th step value of t , we take the maximum with respect to correlations, that is,

$$\sup_{\Gamma \in H_*} \ell(\cdot, \cdot, t^{(i)}) = \ell(\cdot, \cdot^{(i+1)}, t^{(i)}). \quad (15)$$

(B) Variance maximization process

For a given $\cdot, \cdot = \cdot^{(i+1)}$ we take the maximum with respect to variances, that is,

$$\sup_{t > 0} \ell(\cdot, \cdot^{(i+1)}, t) = \ell(\cdot, \cdot^{(i+1)}, t^{(i+1)}). \quad (16)$$

The maximization process (A) is done by a random search. That is, we generate a certain number, say k , of random correlation matrices on the hypothesis H_* and find a \cdot , which gives the maximum value of $\ell(\cdot, \cdot, t_i)$ among them. Note that we used $k = 1$ in this paper for the limitation of computing time. This means that maximization about \cdot , is done only through the iteration of (A) and (B). The maximization process is done by solving a certain non-linear equation numerically.

We iterate both processes and stop at a predetermined number of times. Finally we adopt a pair of (\cdot, \cdot, t) as the maximum likelihood estimator which gives the maximal value of the likelihood among all pairs that appear in the iteration. In section 2 we show in the case of $H_* = H_0$, that is, no hypothesis case, two maximization (A) and (B) are solved exactly and need not any iteration and they are given by the usual mles, that is, sample correlation matrix and sample variances respectively. In Section 3 we treat the process (A). There, we argue how to generate the uniform random elements over the hypothesis considered by the Gibbs sampling in detail. In Section 4 we argue about the maximization process (B) in detail. We will show the maximum is obtained as the fixed point of the iteration of some vector valued non-linear functional. In Section 5 we will study the property of the mle by simulation.

2 Estimation of R under Hypothesis H_0

The unusual feature of our inference method can be characterized as the separate inference for the correlation matrix and the variance vectors. This is introduced in order to make an inference for the correlation matrix under the order restriction among the correlations. Under the hypothesis H_0 the mle for the correlation matrix is the sample correlation matrix usually derived by the normalization of the sample covariance matrix which is the mle of the population covariance matrix. Here we show that our separate inference for the correlation matrix and the variances gives the usual mle of population variances and a population correlation matrix.

Though this should be so, it seems that there has not appeared such proof based on this separate inference in literature. Further this also supports partially the validity of our separate inference method under the other hypotheses. So we give the proof below.

Theorem 2.1 *Under the hypothesis H_0 the mles for Σ , μ , and Δ obtained by the maximization steps (A) and (B) give the usual mles, that is, the sample correlation matrix and sample variance vectors respectively.*

Proof. First we prepare some notations. For $p \times p$ square matrix A we define $\overline{\overline{A}}$ as a matrix which has the same non-diagonal elements with A and all 0 diagonals. Also like an usual convention $diag(c_1, c_2, \dots, c_p)$ denotes a diagonal matrix with the diagonal elements (c_1, \dots, c_p) . Further we define the product $C \otimes D$ of two matrices C and D by $C \otimes D = (c_{ij}d_{ij})$. Taking the partial derivative of the likelihood function $\ell(\Sigma, \mu, t)$ with respect to Σ , μ , and t respectively, we have

$$\frac{\partial \ell}{\partial \Sigma} = \frac{1}{N} \overline{\overline{\Sigma^{-1} T R T^{-1} \Sigma^{-1}}} \quad (17)$$

$$\frac{\partial \ell}{\partial \mu} = (\Sigma^{-1} \otimes R) t \quad (18)$$

$$(19)$$

where the partial derivative with respect to Σ is defined by

$$\frac{\partial \ell}{\partial \Sigma} = \begin{pmatrix} 0 & \frac{\partial \ell}{\partial \rho_{12}} & \dots & \frac{\partial \ell}{\partial \rho_{1p}} \\ \frac{\partial \ell}{\partial \rho_{12}} & 0 & \dots & \frac{\partial \ell}{\partial \rho_{2p}} \\ \dots & \dots & 0 & \dots \\ \frac{\partial \ell}{\partial \rho_{p1}} & \dots & \frac{\partial \ell}{\partial \rho_{p,p-1}} & 0 \end{pmatrix}. \quad (20)$$

Let $diag(\mu_1, \mu_2, \dots, \mu_p)$ and $diag(\nu_1, \nu_2, \dots, \nu_p)$ be the diagonal matrix with diagonal elements of Σ^{-1} and $\frac{1}{N} \Sigma^{-1} T R T^{-1} \Sigma^{-1}$ respectively and further let $diag(m_1, m_2, \dots, m_p)$ be the difference between them. Adding $diag(m_1, m_2, \dots, m_p)$ to both side of the

equation (17) we have

$$,^{-1} - \frac{1}{N},^{-1}TRT,^{-1} = \text{diag}(m_1, m_2, \dots, m_p). \quad (21)$$

Multiplying , from right both side of the equation (21) we have

$$I - \frac{1}{N},^{-1}TRT = \text{diag}(m_1, m_2, \dots, m_p), . \quad (22)$$

The (i, i) element of the matrix of the left hand side of the equation (22) is

$$1 - \frac{1}{N} \sum_{k=1}^p \rho^{ik} t_i r_{ki} t_k = 1 - \frac{1}{N} t_i \sum_{k=1}^p \rho^{ik} r_{ik} t_k. \quad (23)$$

From (18) the right hand side of the above (23) is 0 and thus all the diagonal elements m_i are 0. Then from (21)

$$, = \frac{1}{N}TRT. \quad (24)$$

Noting the diagonal elements of (24) we have

$$1 = \frac{t_i^2}{N}, \quad \forall i. \quad (25)$$

This means $v_i = w_{ii}/N = s_{ii}$ and , = R , which complete the proof of Theorem 2.1.

3 Gibbs sampling and uniform random correlations

Here we consider to generate the uniformly distributed random elements over the set of the correlation matrices. We use the Gibbs sampling method. The Gibbs sampling is one of the methods to generate random vectors obeying the target distribution in the limit through a certain Markov chain. In general the one step of the Markov chain is composed of the iterative generations of one dimensional random variable from the conditional distribution given all other coordinate variables. For the definition of Gibbs sampling see Tanner [7] and for its applications to restricted inference see Gelfand et. al[1]. In our Gibbs sampling we cyclically generate correlation coefficients with other correlation coefficient fixed. The following Lemma 3.1 and Theorem 3.1 states how the method works.

Lemma 3.1 *Let $\rho = (\rho_{ij}) > 0$ be a correlation matrix, and ρ_{-ij} be the set of all upper half components of ρ , other than ρ_{ij} . For fixed ρ_{-ij} , in order that $\rho = (\rho_{ij})$ is a positive definite correlation matrix, ρ_{ij} must be in an interval.*

Proof. Without loss of generality we assume $i = 1$ and $j = 2$. The condition that ρ is a positive definite correlation matrix is all the q ($q = 1, 2, \dots, p$)-th main diagonal square matrix from the right corner have a positive determinant. Since the q -th ($q \leq p - 1$) main diagonal matrix is unrelated to the element ρ_{12} , the condition is equivalent to that the determinant of ρ is positive. We now show this condition implies that ρ_{12} is in some interval. Let us expand the determinant $|\rho|$ with respect to the second column. We have

$$\begin{aligned} (-1)\rho_{12} & \begin{vmatrix} \rho_{12} & \rho_{23} & ** & \rho_{2p} \\ \rho_{13} & 1 & ** & \rho_{3p} \\ * & * & ** & * \\ \rho_{1p} & \rho_{3p} & ** & 1 \end{vmatrix} + \begin{vmatrix} 1 & \rho_{13} & ** & \rho_{1p} \\ \rho_{13} & 1 & ** & \rho_{3p} \\ * & * & ** & * \\ \rho_{1p} & \rho_{3p} & ** & 1 \end{vmatrix} - \rho_{32} \begin{vmatrix} 1 & \rho_{13} & ** & \rho_{1p} \\ \rho_{12} & * & ** & \rho_{2p} \\ * & * & ** & * \\ \rho_{1p} & \rho_{3p} & ** & 1 \end{vmatrix} \\ & + (-1)^p \begin{vmatrix} 1 & \rho_{13} & ** & \rho_{1p} \\ \rho_{12} & * & ** & \rho_{2p} \\ * & * & ** & * \\ \rho_{1,p-1} & \rho_{3,p-1} & ** & \rho_{p-1,p} \end{vmatrix} > 0 \end{aligned}$$

The left hand side of the above equation is a polynomial of degree two in $x = \rho_{12}$ and the coefficient of x^2 is the main diagonal matrix obtained from ρ , by deleting first and second columns and rows from ρ , and so is positive. Hence the condition is the inequality

$$-ax^2 + bx + c > 0, \quad a > 0.$$

This proves the ρ_{12} is in the interval $[\alpha, \beta]$, where α and β are two real roots of the equation

$$-ax^2 + bx + c = 0.$$

Theorem 3.1 *The iterative generations of the random numbers on 1-dimensional intervals appeared in Lemma 3.1 generate uniform random correlations $\rho \in H_0$. Further if we add the restriction such as $\rho_{i,j-1} > \rho_{ij} > \rho_{i,j+1}$ to the interval for generating ρ_{ij} it generate the uniform random correlation matrices on the hypothesis H_r and if we add the restriction such as $\rho_{i-1,j} > \rho_{ij} > \rho_{i+1,j}$ to the interval for generating ρ_{ij} it generate the uniform correlation matrices on the hypothesis H_r . For H_1 it suffices to add the both types of restrictions. Last for H_2 we only need to add the conditions of $\rho_{i,j} > 0$ to that for H_1 .*

Proof. From the general definition of Gibbs sampling and Lemma 3.1 this is obvious.

Table 1: **Relative volume of Hypothesis H_0**

p	3	4	5	6	7
Relative Volume	61.7%	18.3%	2.2%	0.095%	0.0013%

In the maximization process (A) we generate uniformly distributed random correlation matrices over the hypothesis, say, H_0 by the Gibbs sampling to seek the correlation matrix maximizing $\ell(\cdot, t)$ for a given t . A seemingly more easy way to generate random correlations on H_0 is to generate uniform random elements in the box $[-1, 1]^{\frac{p(p-1)}{2}}$ and take out the element when it is in H_0 and a positive definite correlation matrix. However the relative size of H_0 to the box is very small even for $p \geq 4$ as is shown in Table 1. This is a main reason why we propose the Gibbs sampling.

Next we show by simulation that the Gibbs sampling really generates the uniformly distributed random elements. For simplicity we consider the case of the hypothesis H_0 and H_2 . We check the uniformity of Gibbs sampling by comparing the marginal distribution of r_{12} with the exact marginal distribution obtained by the lattice point counting method in which the probability of a set is given by the number of points in the set divided by the total number of lattice points. Note that though for dimension $p = 4$ we can use this lattice point counting method, it gets rapidly difficult for larger p . That is another reason why we propose the Gibbs random search. Of course the convergence of the marginal distribution does not necessarily implies the convergence of the joint distribution, it supports partially the convergence.

Table 2
Marginal distribution of r_{12} by the uniform distribution
in $H_0(p = 4)$

/	$-1 \sim -0.8$	$-0.8 \sim -0.6$	$-0.6 \sim -0.4$	$-0.4 \sim -0.2$	$-0.2 \sim 0.0$
Gibbs(%)	2.77	7.63	11.24	13.64	14.82
Exact(%)	2.81	7.59	11.21	13.61	14.77

/	$0 \sim 0.2$	$0.2 \sim 0.4$	$0.4 \sim 0.6$	$0.6 \sim 0.8$	$0.8 \sim 1.0$
Gibbs(%)	14.79	13.57	11.69	7.57	2.80
Exact(%)	14.77	13.61	11.21	7.59	2.81

Table 3
Marginal distribution of r_{12} by the uniform distribution
in $H_2(p = 4)$

/	0 ~ 0.2	0.2 ~ 0.4	0.4 ~ 0.6	0.6 ~ 0.8	0.8 ~ 1.0
Gibbs(%)	2.45	13.83	27.67	34.44	21.61
Exact(%)	1.14	11.97	27.72	36.35	22.82

We see from Table 2 and 3 the Gibbs random correlation matrices give the same marginal distribution with that of the uniform distribution on H_0 . Moreover from our simulation experience this holds for the other hypothesis.

4 Maximization process (B)

In maximization process (B) for given $\rho \in H$ we seek $t = (t_1, t_2, \dots, t_p)$ which maximize the likelihood $\ell(\rho, t)$.

Theorem 4.1 *For a given $\rho \in H$ $t = (t_1, t_2, \dots, t_p)$ which maximizes the likelihood is the solution of the following non-linear equation,*

$$N/t = Bt,$$

$$\text{where } B = (b_{ij}) = (\rho^{ij} r_{ij}) \text{ and } 1/t = (1/t_1, \dots, 1/t_p).$$

Proof. The equation is obtained by taking the partial derivative of the likelihood with respect to t .

Our second main result is the following two theorems.

Theorem 4.2 *The above non-linear equation has the unique solution in the positive orthant.*

Proof. The condition given by the equation with the variable t_i is

$$t_i = \frac{c_i \pm \sqrt{c_i^2 + 4nb_{ii}}}{2b_{ii}}, \tag{26}$$

where

$$B = (b_{ij}) \text{ and } c_i = \sum_{j=1}^p b_{ij} t_j. \tag{27}$$

This implies that the set of the points satisfying i -th component equation makes two hyper surfaces with $t_i > 0$ and $t_i < 0$ respectively. Thus the solution set is the intersection of these $2p$ hyper surfaces and it consists of $2p$ points. Clearly the only one solution is in the positive orthant.

Theorem 4.3 *If the iterative method described below converges, it converges the unique positive orthant solution.*

Description: *Take a starting point $t^{(0)} = (t_1^{(0)}, t_2^{(0)}, \dots, t_p^{(0)})$ in the positive orthant. Let $t^{(0)}(-i)$ be the set of $(p-1)$ components of $t^{(0)}$ with the i -th component deleted. First for the given $t^{(0)}(-1)$ we solve the equation with respect to the first component t_1 , which is a degree two polynomial equation in t_1 . The equation has the real solution and we adopt the positive solution and we denote it as t_1^1 . Then we put $t^{(0)}$ with the first component replaced by t_1^1 as $t^{(0),1}$. For given $t^{(0),1}(-2)$ we solve the second equation which is a polynomial equation of degree 2 with respect to the t_2 . It also has the positive solution. We denote this positive solution as t_2^1 and denote $t^{(0),1}$ with the second component replaced by t_2^1 as the $t^{(0),2}$. Like this we have $t^{(0),3}, \dots, t^{(0),p}$ and we denote the set $t^{(0),p}$ as $t^{(1)}$. For this $t^{(1)}$ we repeat the same procedure above and we get $t^{(2)}$. And in the same way we have the sequence $t^{(0)}, t^{(1)}, \dots$. This sequence converges the only positive orthant solution.*

Proof. First we solve a equation of degree two in t_1 and put the positive solution as $t_1 = f_1(t_2, t_3, \dots, t_p)$. Next we solve a second equation of degree two in t_2 with t'_1 in place of t_1 and put the positive solution as $t'_2 = f_2(t'_1, t_3, \dots, t_p)$. Likewise we have $t'_3 = f_3(t'_1, t'_2, \dots, t_p), \dots, t'_p = f_p(t'_1, \dots, t'_p)$. Now we define the map f from \mathfrak{R}^p to itself by $(t_1, t_2, \dots, t_p) \rightarrow (t'_1, t'_2, \dots, t'_p)$. Clearly this map describes just the one-step of our above iteration. Our iteration method thus can be described as the iterative applications of the same mapping to the starting point. That is, the sequence of the points is given by

$$t^{(n)} = f^{(n)}(t^{(0)}) = f(f(\dots f(t^{(0)}) \dots)). \quad (28)$$

If these sequence of points converge to some point $t^{(\infty)}$ it is necessarily the fixed point of the mapping. Since $t^{(\infty)}$ is the fixed point we have

$$t^{(\infty)} = f(t^{(\infty)}).$$

Then we have $t_1^{(\infty)} = f_1(t_2^{(\infty)}, t_3^{(\infty)}, \dots, t_p^{(\infty)})$, $t_2^{(\infty)} = f_2(t_1^{(\infty)}, t_3^{(\infty)}, \dots, t_p^{(\infty)})$, \dots , $t_p^{(\infty)} = f_p(t_1^{(\infty)}, t_2^{(\infty)}, \dots, t_{p-1}^{(\infty)})$. This show the fixed point $t^{(\infty)}$ is the solution of the non-linear equation in Theorem 4.1. Moreover the process shows the solution is in the positive orthant.

From our experience the non-linear equation can be solved by the process(B) with at most 20 times iterations.

Remark 4.1 Theorem above does not make sure the convergence of our iteration method. However it converges in our all simulation experiences. Partial result about convergence will appear in Sakata et al.[6].

5 Comparisons of relative errors of mle and others

In this section we compare the relative error, the mean absolute error divided by a true value, among three estimators of the population correlation matrix $\rho \in H_2$, R_{gibbs} , R_{pava} and R . R_{gibbs} is the pseudo mle obtained by Gibbs sampling, R_{pava} is the estimator obtained matrix order PAVA algorithm(see the page 27 of Robertson et al[4] for the definition of matrix order PAVA algorithm). We briefly describe our simulation scheme now.

First we choose $\rho \in H_2$ which we call a true parameter and generate a sample of size n , $Y^{(1)}, Y^{(2)}, \dots, Y^{(n)}$ from p -dimensional normal distribution $N_p(0, \rho)$. Then we calculate the sample covariance $C = (c_{ij})$ by

$$c_{ij} = \frac{1}{n} \sum_{k=1}^n Y_i^{(k)} Y_j^{(k)} \quad (29)$$

and the sample correlation matrix $R = (r_{ij})$ where

$$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}} \sqrt{c_{jj}}} \quad (30)$$

Note that R_{pava} calculated from R by PAVA algorithm is also a correlation matrix by the definition of PAVA. Though it has no direct relation to maximization of the likelihood function it may be a competitive one to R_{gibbs} . In order to run a Gibbs sampling on H_2 , first we choose an initial correlation matrix ρ_0 in H_2 and an initial variance vector V_0 . For each Gibbs sample of ρ , we perform the maximization process (B) for variances and after then we calculate the likelihood for the set (ρ, V) and register it. After repeating this step 50000 times we take out R_{gibbs} which gives the most highest likelihood values among the 50000 Gibbs samples. Repeating this simulation 100 times we obtained the relative error of R_{gibbs} . For other two statics we also calculated their relative errors by 100 simulated values.

We should note that in the above simulation ρ_0 and $\rho \in H_2$ are obtained as

$$\rho_{ij} = \sqrt{\frac{\kappa_1 + \kappa_2 + \dots + \kappa_i}{\kappa_1 + \kappa_2 + \dots + \kappa_j}} \quad (31)$$

with $\kappa_1 = 0.8, \kappa_2 = 0.6$ and $\kappa_3 = 0.4$ for ρ and $\kappa_1 = 0.5, \kappa_2 = 0.7$ and $\kappa_3 = 0.8$ for ρ_0 . Note that this type of correlation matrix is that of (Y_1, Y_2, \dots, Y_p) where

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \\ &\dots \\ Y_p &= X_1 + X_2 + \dots + X_p \end{aligned}$$

and where $EX_i = 0$, $Var(X_i) = \kappa_i$ and X_i are mutually independent.

In the below we present the simulation results about the relative errors.

Table 4
True $R_* \in H_1$

\	X_1	X_2	X_3	X_4
X_1	1.0000	0.9258	0.8452	0.7559
X_2	*	1.0000	0.9129	0.8165
X_3	*	*	1.0000	0.8944

Table 5($n = 10$)
Relative errors $\times 100$ of R (upper)
 R_{pava} (middle) and R_{gibbs} (lower)

\	X_1	X_2	X_3	X_4
X_1	0.00	8.93	20.44	26.05
	0.00	9.1	18.6	24.7
	0.00	8.90	19.74	25.56
X_2	*	0.00	9.62	15.50
	*	0.00	9.35	15.10
	*	0.00	9.16	15.26
X_3	*	*	0.00	6.68
	*	*	0.00	5.77
	*	*	0.00	6.60

Table 6($n = 20$)
Relative errors $\times 100$ of R (upper)
 R_{pava} (middle) and R_{gibbs} (lower)

\	X_1	X_2	X_3	X_4
X_1	0.00	6.59	14.28	17.45
	0.00	6.38	13.3	19.4
	0.00	6.64	14.56	17.48
X_2	*	0.00	6.63	10.80
	*	0.00	6.31	10.92
	*	0.00	7.12	10.75
X_3	*	*	0.00	4.26
	*	*	0.00	4.44
	*	*	0.00	4.33

Remark 5.1 . Simulation study for mean square errors will appear in Sakata et al.[5].

Remark 5.2 It might happen that in place of R we had better use the Olkin-Platt unbiased correction R^* where

$$r_{ij}^* = r_{ij} + \frac{r_{ij}(1 - r_{ij}^2)}{n - 1}. \quad (32)$$

See Olkin-Platt[4] for the definition and the properties of the estimator. However it is not sure whether R^* is always a positive definite. And so we did not used it here.

Remark 5.3 If we search \hat{R} in an appropriate neighbourhood of R in stead of searching the whole space of a hypothesis we could search more efficiently. However in that method it seems a little bit difficult to determine an initial starting point \hat{R}_0 . So the method was not used in this paper. However that direction may be of value to be developed in the future.

In conclusion we proposed a new method to estimate the correlation matrix of normal distribution under the order restriction among correlations. Our method seems to work well, though we could not show that R_{gibbs} is specially superior R or R_{pava} in the relative error. It might happen that the relative error of R_{gibbs} become more less if we take a more longer step of Gibbs sample, though we could not do so due to the ability of our computer. In any way R_{gibbs} is more natural and so more recommendable than R in the context that we can assume some kind of restrictions among correlations. Finally we would like to note the following. The sample variance vector has no information about the population correlation matrix, however the sample correlation has some information about the population variances. So we cannot separate the likelihood function into the part of variances and the part of correlations. Looking at our iterative methods closely, our method should be considered as an estimation method of both variances and correlations under order restrictions among correlations. So we had better take into account the estimation error in the variance estimation. In this respect further study will be needed.

Acknowledgements

The first author deeply owes to the Center of Mathematics and its Application of Australia National University for the use of facilities and for the research environment.

References

- [1] Gelfand,A.E.,Smith,A.F.M. and Lee,T.M(1995). Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, *J.A.S.A.*, vol.87,523-532.

- [2] Muirhead, R.J.(1982). Aspects of Multivariate Statistical Theory, *John Wiley and Sons Inc.* **10**,Chapter 5.
- [3] Olkin,I.L. and Pratt,J.W(1958). Unbiased estimation of certain correlations,*Ann. Math. Statist.*,*29*,201-211.
- [4] Robertson, Wright and Dykstra(1988). Order restricted statistical inference, *Wiley Sons*
- [5] Sakata,T. and Sawae,R.(1997). A study of the relative error and mean square error of the mle of correlation matrix under inequality constraints among correlations. Preprint.
- [6] Sakata,T. and Sawae,R.(1997). Numerical solution of a certain nonlinear equation arising in maximal likelihood estimation of correlation matrix under inequality constraints. Preprint.
- [7] Tanner,M.A.(1991). Tool for Statistical Inference:Observed Data and Data Augumentation Methods. *Lecture Notes in Statistics, Vol.67 Springer Verlag.*

Toshio SAKATA
Department of Computer Science
Faculty of Engineering
Kumamoto University
Kurokami 39-1 Kumamoto 860, Japan

Ryuichi SAWAE
Department of Applied Mathematics
Faculty of Science
Okayama University of Science
Ridai-cyo Okayama 700, Japan.