DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

ON ESTIMATING THE MEAN IN BAYESIAN FACTOR ANALYSIS

Daniel B. Rowe

SOCIAL SCIENCE WORKING PAPER 1096

July 2000

On Estimating the Mean in Bayesian Factor Analysis

Daniel B. Rowe[1]
Division of Humanities and Social Sciences,
California Institute of Technology, USA

Abstract

In the Bayesian factor analysis model (Press & Shigemasu, 1989), the sample size was assumed to be large enough to estimate the overall population mean by the sample mean. In this paper, the procedure of estimating the population mean by the sample mean is compared to estimating it along with the other parameters both by Gibbs sampling and Iterated Conditional Modes. Results show that even in small samples, the Gibbs sampling and iterated conditional modes estimates of the mean are for practical purposes identical to the sample mean. Thus, the population mean is adequately estimated by its sample value.

1 Introduction

In the Bayesian factor anlysis model first proposed by Press and Shigemasu, 1989 (henceforth PS89) the classical normal sampling model was assumed, but the disturbance covariance matrix was assumed to be a full positive definite matrix. One of the prior assumptions, however, was that

[1]Requests for reprints should be addressed to Daniel B. Rowe, Division of Humanities and Social Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA (e-mail: drowe@hss.caltech.edu).

the expected value of the disturbance covariance matrix was diagonal in order to represent traditional views of the factor model containing "common" and "specific" factors. Natural conjugate prior distributions were specified for the unknown matrices.

In PS89, the model parameters were estimated using a large sample approximation for one of the terms in the joint posterior distribution with the result that the marginal posterior distribution of the factor scores was found to be approximately matrix T. The factor loading matrix was estimated conditional on the factor scores, and the disturbance covariance matrix was estimated conditional on estimates of the factor scores and the factor loadings. The need for this large sample approximation was alleviated in Rowe & Press, 1998 (henceforth RP98) by estimating the model parameters by both Gibbs sampling (Geman & Geman, 1984 and Gelfand & Smith, 1990) and iterated conditional modes (Lindley & Smith, 1972).

In both PS89 and RP98 the sample size was assumed to be large enough to estimate the overall population mean by the sample mean well enough for it to be ignored after subtracting it out. The subject addressed here is, how good is this approximation for various sample sizes. In evaluating the estimation of the population mean by the sample mean, Gibbs sampling and ICM estimates are computed.

In this paper, the same model as in PS89 and RP98 is adopted, and the unknown quantities including the population mean are estimated by Gibbs sampling and by Iterated Conditional Modes (ICM). For both approaches conditional posterior distributions for each of the parameters given the other parameters and the data. All four can be found explicitly. Gibbs marginal and ICM conditional estimators may then readily be found from the conditional posterior distributions.

The plan of the paper is to review the model and to adopt prior distributions in Section 2. Present the conditional mean and modal estimaters found in PS89, obtain conditional posterior distributions along with the Gibbs sampling and ICM algorithms in Section 3. In Section 4 an example is presented, and the results for the population mean, from both the Gibbs sampling and the ICM estimation methods are compared to those of PS89 in which the population mean is estimated by the sample mean.

2 Model

2.1 Likelihood Function

The Bayesian factor analysis model (assuming a nonzero overall population mean) is:

$$\underset{(p \times 1)}{(x_j|\mu, \Lambda, f_j)} = \underset{(p \times 1)}{\mu} + \underset{(p \times m)}{\Lambda} \underset{(m \times 1)}{f_j} + \underset{(p \times 1)}{\epsilon_j}, \quad m < p, \tag{2.1}$$

for $j = 1, \ldots, n$, where x_j is the j^{th} observation, μ is the overall population mean, Λ is a matrix of constants called the factor loading matrix; f_j is the factor score vector for subject j; and the ϵ_j's are assumed to be mutually uncorrelated and normally distributed $N(0, \Psi)$ variables.

In the traditional model, Ψ is taken to be a diagonal matrix so that-common and specific factors can readily distinguished. In the the current model, Ψ is taken to be a general symmetric, positive definite covariance matrix with the property of being diagonal on the average, i.e., $E(\Psi) =$ a diagonal matrix.

It is assumed that (μ, Λ, F, Ψ) are unobservable but fixed quantities, and that we can write the distribution of each x_j as

$$p(x_j|\mu, \Lambda, f_j, \Psi) = (2\pi)^{-\frac{p}{2}}|\Psi|^{-\frac{1}{2}}e^{-\frac{1}{2}(x_j-\mu-\Lambda f_j)'\Psi^{-1}(x_j-\mu-\Lambda f_j)}. \tag{2.2}$$

If proportionality is denoted by "$\propto$" and the Kroneker product by $\otimes$ then, the likelihood for (μ, Λ, F, Ψ) is

$$p(X|\mu, \Lambda, F, \Psi) \propto |\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')} \tag{2.3}$$

where the p-variate observation vectors on n subjects are, $X' = (x_1, \ldots, x_n)$, the factor scores are $F' = (f_1, \ldots, f_n)$, and the errors of observation are $E' = (\epsilon_1, \ldots, \epsilon_n)$. The notation $p(\cdot)$ will denote "density"; densities will be distinguished by their arguments. The proportionality constant in (2.3) depends only on (p, n) and not on (μ, Λ, F, Ψ).

2.2 Priors

The same prior distributions are adopted as in PS89 and again in RP98 so that the joint prior density is:

$$p(\mu, \Lambda, F, \Psi) \propto p(\mu)p(\Lambda|\Psi)p(\Psi)p(F), \tag{2.4}$$

where

$$p(\mu) \quad \propto \quad \text{a constant} \tag{2.5}$$

$$p(\Lambda|\Psi) \propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2} tr \Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'}, \tag{2.6}$$

$$p(\Psi) \propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2} tr \Psi^{-1} B}, \tag{2.7}$$

$$p(F) \propto e^{-\frac{1}{2} tr F'F} \tag{2.8}$$

with $H, B, \Psi > 0$ and B a diagonal matrix. In PS89 and again in RP98, a noninformative prior was implicitly specified for the population mean. The matrix Λ conditional on Ψ has elements which are jointly normally distributed, and hyperparameters (Λ_0, H) are to be assesssed. The matrix Ψ^{-1} follows a Wishart distribution, with hyperparameters (ν, B) which are to be assessed. It is assumed that $E(\Psi|B)$ and $E(\Psi)$ are diagonal, in order to represent traditional views of the factor model containing "common" and "specific" factors.

The joint normal distribution for $(\Lambda|\Psi)$ comes from writing $\Lambda' \equiv (\lambda_1, \ldots, \lambda_p)$, as $\lambda \equiv vec(\Lambda') = (\lambda_1', \ldots, \lambda_p')'$; then $\mathrm{var}(\lambda|\Psi) = \Psi \otimes H^{-1}$, which can be written as a matrix normal distribution (Kotz and Johnson, 1985, p. 326–333). Also, as in PS89 and RP98, $H = h_0 I$, for some preassigned scalar h_0. Assessment of the hyperparameters is simplified by these interpretations.

2.3 Joint Posterior

Using Bayes rule, combine (2.3)–(2.8), to get the joint posterior density of the parameters

$$p(\mu, F, \Lambda, \Psi|X) \propto e^{-\frac{1}{2} tr F'F} |\Psi|^{-\frac{(n+m+\nu)}{2}} e^{-\frac{1}{2} tr \Psi^{-1} U} \tag{2.9}$$

where $U = (X - e_n \otimes \mu' - F\Lambda')'(X - e_n \otimes \mu' - F\Lambda') + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

3 Estimation

3.1 Conditional Modes

In PS89, the population mean is estimated by the sample mean as $\hat{\mu} = \bar{x}$. It is noted that marginal estimates may not be found for Λ and Ψ. Alternatively, marginalization and conditional estimation is used. The marginal mean and mode $E(F|X) = \hat{F}$ is computed, the conditional mean and mode $E(\Lambda|\hat{F}, X) = \hat{\Lambda}$, then the mean value $E(\Psi|\Lambda, \hat{F}, X) = \hat{\Psi}$ can be found as well as the mode.

The joint posterior distribution is integrated with respect to the disturbance covariance matrix Ψ and the factor loadings Λ to obtain

$$p(F|X) \propto \frac{e^{-\frac{1}{2}trF'F}|H + F'F|^{\frac{n+\nu-2p-1}{2}}}{|A + (F - \hat{F})'P(F - \hat{F})|^{\frac{(n+m+\nu-p-1)}{2}}}, \tag{3.1}$$

where

$$\begin{aligned} \hat{F} &= P^{-1}(X - e_n \otimes \bar{x}')W^{-1}\Lambda_0 H \qquad (3.2) \\ P &= I_n - (X - e_n \otimes \bar{x}')W^{-1}(X - e_n \otimes \bar{x}')' \\ W &= (X - e_n \otimes \bar{x}')'(X - e_n \otimes \bar{x}') + B + \Lambda_0 H \Lambda_0' \\ A &= H - (\Lambda_0 H)'W^{-1}\Lambda_0 H \\ &\quad -[(X - e_n \otimes \bar{x}')W^{-1}\Lambda_0 H]'P^{-1}[(X - e_n \otimes \bar{x}')W^{-1}\Lambda_0 H]. \end{aligned}$$

When the sample size n is large, $F'F \approx nI_m$, by the weak law of large numbers. The two terms in the numerator can now be incorporated into the proportionality constant and the marginal posterior density of F becomes

$$p(F|X) \propto \frac{1}{|A + (F - \hat{F})'P(F - \hat{F})|^{\frac{(n+\nu-p-1)}{2}}}. \tag{3.3}$$

This is the kernel of a matrix T-distribution. The large sample posterior mean (and modal) estimator of F is $E(F|X) = \hat{F}$. Now estimate Λ for given $F = \hat{F}$.

The conditional distribution of Λ for given F is

$$p(\Lambda|F, X) \propto \frac{1}{|R + (\Lambda - \hat{\Lambda})(H + F'F)(\Lambda - \hat{\Lambda})'|^{-\frac{(n+m+\nu-p-1)}{2}}}, \tag{3.4}$$

where

$$\begin{aligned} R &= B + (X - e_n \otimes \bar{x}')'(X - e_n \otimes \bar{x}') + \Lambda_0 H \Lambda_0' - \\ &\quad [(X - e_n \otimes \bar{x}')'F + \Lambda_0 H](H + F'F)^{-1}[(X - e_n \otimes \bar{x}')'F + \Lambda_0 H]' \\ \hat{\Lambda} &= [(X - e_n \otimes \bar{x}')'F + \Lambda_0 H](H + F'F)^{-1}. \end{aligned}$$

That is $(\Lambda|F, X)$ follows a matrix T-distribution. Their posterior conditional mean (and modal) estimator of Λ is $E(\Lambda|\hat{F}, X)$, or

$$\hat{\Lambda} = [(X - e_n \otimes \bar{x}')'\hat{F} + \Lambda_0 H](H + \hat{F}'\hat{F})^{-1}. \tag{3.5}$$

The covariance matrix Ψ is estimated conditional upon $(\Lambda, F) = (\hat{\Lambda}, \hat{F})$. The conditional density of $(\Psi|\hat{\Lambda}, \hat{F}, X)$ is

$$p(\Psi|\hat{\Lambda}, \hat{F}, X) \propto \frac{e^{-\frac{1}{2} tr \Psi^{-1} \hat{G}}}{|\Psi|^{\frac{(n+m+\nu)}{2}}}, \quad \Psi > 0 \tag{3.6}$$

where

$$\begin{aligned} \hat{G} &= [(X - e_n \otimes \bar{x}') - \hat{F}\hat{\Lambda}']'[(X - e_n \otimes \bar{x}') - \hat{F}\hat{\Lambda}'] + \\ &\quad (\hat{\Lambda} - \Lambda_0)H(\hat{\Lambda} - \Lambda_0)' + B. \end{aligned} \tag{3.7}$$

The conditional posterior mean $E(\Psi|\hat{\Lambda}, \hat{F}, X)$ of $p(\Psi|\hat{\Lambda}, \hat{F}, X)$ is

$$\hat{\Psi} = \frac{\hat{G}}{n + m + \nu - 2p - 2}. \tag{3.8}$$

It should be noted that the conditional mode of $p(\Psi|\hat{\Lambda}, \hat{F}, X)$ is not the same as the conditional mean. The conditional mode is

$$\hat{\Psi}_{mode} = \frac{\hat{G}}{n+m+\nu}. \tag{3.9}$$

The estimators $(\hat{F}, \hat{\Lambda}, \hat{\Psi}_{mode})$ are conditional posterior modal estimators.

3.2 Conditional Posterior Densities

The four posterior conditional distributions are as follows.

$$\begin{aligned} p(\mu|\Lambda, F, X, X) &\propto p(\mu)p(X|\mu, F, \Lambda, \Psi) \\ &\propto |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')} \\ &\propto e^{-\frac{1}{2}(\mu-\tilde{\mu})'\left(\frac{\Psi}{n}\right)^{-1}(\mu-\tilde{\mu})} \end{aligned} \tag{3.10}$$

where $\tilde{\mu} = \bar{x} - \Lambda\bar{f}$. Note that the mean of the distribution is not $\bar{x}$.

$$\begin{aligned} p(\Lambda|\mu, F, \Psi, X) &\propto p(\Lambda|\Psi)p(X|\mu, F, \Lambda, \Psi) \\ &\propto |\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'} \\ &\cdot\ |\Psi|^{-\frac{N}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')} \\ &\propto e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\tilde{\Lambda})(H+F'F)(\Lambda-\tilde{\Lambda})'} \end{aligned} \tag{3.11}$$

where $\tilde{\Lambda} = [(X - e_n \otimes \mu')'F + \Lambda_0 H](H + F'F)^{-1}$.

$$\begin{aligned} p(\Psi|\mu, F, \Lambda, X) &\propto p(\Psi)p(\Lambda|\Psi)p(X|\mu, F, \Lambda, \Psi) \\ &\propto |\Psi|^{-\frac{\nu}{2}} e^{-\frac{1}{2}tr\Psi^{-1}B}|\Psi|^{-\frac{m}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(\Lambda-\Lambda_0)H(\Lambda-\Lambda_0)'} \\ &\cdot\ |\Psi|^{-\frac{n}{2}} e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')} \\ &\propto |\Psi|^{-\frac{(n+m+\nu)}{2}} e^{-\frac{1}{2}tr\Psi^{-1}U} \end{aligned} \tag{3.12}$$

where $U = (X - e_n \otimes \mu' - F\Lambda')(X - e_n \otimes \mu' - F\Lambda')' + (\Lambda - \Lambda_0)H(\Lambda - \Lambda_0)' + B$.

$$\begin{aligned} p(F|\mu, \Lambda, \Psi, X) &\propto p(F)p(X|\mu, F, \Lambda, \Psi) \\ &\propto e^{-\frac{1}{2}trF'F}|\Psi|^{-\frac{n}{2}}e^{-\frac{1}{2}tr\Psi^{-1}(X-e_n\otimes\mu'-F\Lambda')'(X-e_n\otimes\mu'-F\Lambda')} \\ &\propto e^{-\frac{1}{2}tr(F-\tilde{F})(I_m+\Lambda'\Psi^{-1}\Lambda)(F-\tilde{F})'} \end{aligned} \tag{3.13}$$

where $\tilde{F} = (X - e_n \otimes \mu')\Psi^{-1}\Lambda(I_m + \Lambda'\Psi^{-1}\Lambda)^{-1}$.

The modes of these conditional distributions are $\tilde{F}$, $\tilde{\Lambda}$ (as defined above), and

$$\tilde{\Psi} = \frac{U}{n+m+\nu}, \tag{3.14}$$

respectively.

3.3 The Gibbs Sampling Algorithm

For Gibbs estimation of the posterior, we start with initial values for μ, F, and Ψ say $\bar{\mu}_{(0)}$, $\bar{F}_{(0)}$, and $\bar{\Psi}_{(0)}$. Then cycle through

$$\begin{aligned} \bar{\Lambda}_{(i+1)} &= \text{a random sample from } p(\Lambda|\bar{\mu}_{(i)}, \bar{F}_{(i)}, \bar{\Psi}_{(i)}, X) \\ \bar{\Psi}_{(i+1)} &= \text{a random sample from } p(\Psi|\bar{\mu}_{(i)}, \bar{F}_{(i)}, \bar{\Lambda}_{(i+1)}, X) \\ \bar{F}_{(i+1)} &= \text{a random sample from } p(F|\bar{\mu}_{(i)}, \bar{\Lambda}_{(i+1)}, \bar{\Psi}_{(i+1)}, X) \\ \bar{\mu}_{(i+1)} &= \text{a random sample from } p(\mu|\bar{F}_{(i+1)}, \bar{\Lambda}_{(i+1)}, \bar{\Psi}_{(i+1)}, X). \end{aligned}$$

The first s random samples are discarded and the remaining t samples are kept. The means of the remaining random samples

$$\bar{\mu} = \frac{1}{t}\sum_{l=1}^{t}\bar{\mu}_{(k)} \quad \bar{F} = \frac{1}{t}\sum_{k=1}^{t}\bar{F}_{(s+k)} \quad \bar{\Lambda} = \frac{1}{t}\sum_{k=1}^{t}\bar{\Lambda}_{(s+k)} \quad \bar{\Psi} = \frac{1}{t}\sum_{k=1}^{t}\bar{\Psi}_{(s+k)}.$$

are the posterior estimates of the parameters.

3.4 The ICM Algorithm

For Iterated Conditional Modes estimatiom of the posterior, we start with an initial value for $\tilde{\mu}$ and $\tilde{F}$, say $\mu_{(0)}$ and $\tilde{F}_{(0)}$ and then cycle through

$$
\begin{aligned}
\tilde{\Lambda}_{(i+1)} &= [(X - e_n \otimes \tilde{\mu}'_{(i)})'\tilde{F}_{(i)} + \Lambda_0 H)(H + \tilde{F}'_{(i)}\tilde{F}_{(i)})^{-1} \\
\tilde{\Psi}_{(i+1)} &= \{[(X - e_n \otimes \tilde{\mu}'_{(i)}) - \tilde{F}_{(i)}\tilde{\Lambda}'_{(i+1)}]'[(X - e_n \otimes \tilde{\mu}'_{(i)}) - \tilde{F}_{(i)}\tilde{\Lambda}'_{(i+1)}] + \\
&\quad (\tilde{\Lambda}_{(i+1)} - \Lambda_0)H(\tilde{\Lambda}_{(i+1)} - \Lambda_0)' + B\}/(n + m + \nu) \\
\tilde{F}_{(i+1)} &= (X - e_n \otimes \tilde{\mu}'_{(i)})\tilde{\Psi}^{-1}_{(i+1)}\tilde{\Lambda}_{(i+1)}(I_m + \tilde{\Lambda}'_{(i+1)}\tilde{\Psi}^{-1}_{(i+1)}\tilde{\Lambda}_{(i+1)})^{-1} \\
\tilde{\mu}_{(i+1)} &= \bar{x} - \tilde{\Lambda}_{(i+1)}\bar{\tilde{f}}_{(i+1)}.
\end{aligned}
$$

until convergence is reached with the joint posterior modal estimator $(\tilde{\mu}, \tilde{F}, \tilde{\Lambda}, \tilde{\Psi})$. The mean of the factor score vectors has been denoted by $\bar{\tilde{f}}$.

4 Example

In this section the ICM and the Gibbs Sampler procedures for estimating the parameters of the Bayesian factor analysis model are used and the resulting estimators are compared with those obtained by estimating the population mean by the sample mean as in PS89 for various sample sizes. The data is extracted from an example in Kendall 1980, p.53. The problem as stated in PS89 and again in RP98 is the following.

> There are 48 applicants for a certain job, and they have been scored on 15 variables regarding their acceptability. They are:

(1)	Form of letter application	(9)	Experience
(2)	Appearance	(10)	Drive
(3)	Academic ability	(11)	Ambition
(4)	Likeabiliy	(12)	Grasp
(5)	Self-confidence	(13)	Potential
(6)	Lucidity	(14)	Keenness to join
(7)	Honesty	(15)	Suitability
(8)	Salesmanship		

The raw scores of the applicants on these 15 variables, measured on the same scale, are presented in Table 1. The question is, Is there an underlying subset of factors that explain the variation observed in the scores? If so, then the applicants could be compared more easily.

Table 1: Raw scores of 48 applicants scaled on 15 variables.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	6	7	2	5	8	7	8	8	3	8	9	7	5	7	10
2	9	10	5	8	10	9	9	10	5	9	9	8	8	8	10
3	7	8	3	6	9	8	9	7	4	9	9	8	6	8	10
4	5	6	8	5	6	5	9	2	8	4	5	8	7	6	5
5	6	8	8	8	4	5	9	2	8	5	5	8	8	7	7
6	7	7	7	6	8	7	10	5	9	6	5	8	6	6	6
7	9	9	8	8	8	8	8	8	10	8	10	8	9	8	10
8	9	9	9	8	9	9	8	8	10	9	10	9	9	9	10
9	9	9	7	8	8	8	8	5	9	8	9	8	8	8	10
10	4	7	10	2	10	10	7	10	3	10	10	10	9	3	10
11	4	7	10	0	10	8	3	9	5	9	10	8	10	2	5
12	4	7	10	4	10	10	7	8	2	8	8	10	10	3	7
13	6	9	8	10	5	4	9	4	4	4	5	4	7	6	8
14	8	9	8	9	6	3	8	2	5	2	6	6	7	5	6
15	4	8	8	7	5	4	10	2	7	5	3	6	6	4	6
16	6	9	6	7	8	9	8	9	8	8	7	6	8	6	10
17	8	7	7	7	9	5	8	6	6	7	8	6	6	7	8
18	6	8	8	4	8	8	6	4	3	3	6	7	2	6	4
19	6	7	8	4	7	8	5	4	4	2	6	8	3	5	4
20	4	8	7	8	8	9	10	5	2	6	7	9	8	8	9
21	3	8	6	8	8	8	10	5	3	6	7	8	8	5	8
22	9	8	7	8	9	10	10	10	3	10	8	10	8	10	8
23	7	10	7	9	9	9	10	10	3	9	9	10	9	10	8
24	9	8	7	10	8	10	10	10	2	9	7	9	9	10	8
25	6	9	7	7	4	5	9	3	2	4	4	4	4	5	4
26	7	8	7	8	5	4	8	2	3	4	5	6	5	5	6
27	2	10	7	9	8	9	10	5	3	5	6	7	6	4	5
28	6	3	5	3	5	3	5	0	0	3	3	0	0	5	0
29	4	3	4	3	3	0	0	0	0	4	4	0	0	5	0
30	4	6	5	6	9	4	10	3	1	3	3	2	2	7	3
31	5	5	4	7	8	4	10	3	2	5	5	3	4	8	3
32	3	3	5	7	7	9	10	3	2	5	3	7	5	5	2
33	2	3	5	7	7	9	10	3	2	2	3	6	4	5	2
34	3	4	6	4	3	3	8	1	1	3	3	3	2	5	2
35	6	7	4	3	3	0	9	0	1	0	2	3	1	5	3
36	9	8	5	5	6	6	8	2	2	2	4	5	6	6	3
37	4	9	6	4	10	8	8	9	1	3	9	7	5	3	2
38	4	9	6	6	9	9	7	9	1	2	10	8	5	5	2
39	10	6	9	10	9	10	10	10	10	10	8	10	10	10	10
40	10	6	9	10	9	10	10	10	10	10	10	10	10	10	10
41	10	7	8	0	2	1	2	0	10	2	0	3	0	0	10
42	10	3	8	0	1	1	0	0	10	0	0	0	0	0	10
43	3	4	9	8	2	4	5	3	6	2	1	3	3	3	8
44	7	7	7	6	9	8	8	6	8	8	10	8	8	6	5
45	9	6	10	9	7	7	10	2	1	5	5	7	8	4	5
46	9	8	10	10	7	9	10	3	1	5	7	9	9	4	4
47	0	7	10	3	5	0	10	0	0	2	2	0	0	0	0
48	0	6	10	1	5	0	10	0	0	2	2	0	0	0	0

Note that the initial values for the ICM and Gibbs sampling estimation procedures have little effect on the final result, because for ICM we have unimodal posterior conditional distributions so we are sure to converge to the mode quickly, and for Gibbs sampling, we have a burn-in period. We choose the initial value for $\tilde{\mu}$ and $\tilde{F}$ to be $\tilde{\mu}_{(0)} = \bar{x}$ and $\tilde{F}_{(0)} = \hat{F}$, the estimator of PS89. This choice of the initial value hastens convergence.

The same underlying structure is postulated as in as PS89, a model with 4 factors. This choice is based upon PS89 having carried out a principal components analysis and having found that 4 factors accounted for 81.5% of the variance. Based upon underlying theory they constructed the prior factor loading matrix

$$\Lambda_0' = \begin{pmatrix} 0 & 0 & 0 & 0 & .7 & .7 & 0 & .7 & 0 & .7 & .7 & .7 & .7 & 0 & 0 \\ 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & .7 \\ 0 & 0 & 0 & .7 & 0 & 0 & .7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In PS89, the hyperparameter H was assessed as $H = 10I_4$, B was assessed as $B = 0.2I_{15}$, and ν was assessed as $\nu = 33$. The factor scores, factor loadings, and disturbance variances and covariances may now be estimated. It was found that a burn in period of $5,000$ samples worked well, so then the next $25,000$ samples were taken for the Gibbs estimates.

Table 2 displays the PS89, Gibbs sampling, and ICM estimates of the population mean respectively. Note that for all sample sizes considered, the exact Gibbs sampling and ICM estimates are nearly identical to the PS89 estimates which are the sample means.

Table 2: PS89, Gibbs Sampling, and ICM estimates of the mean.

p	n=48	n=40	n=30	n=20	n=10
1	6.0000	6.0000	6.1333	6.3500	7.1000
2	7.0833	7.3000	7.7333	7.9500	8.0000
3	7.0833	6.7000	6.9667	7.3500	6.7000
4	6.1458	6.4500	6.5000	6.2000	6.4000
5	6.9375	7.3750	7.4667	7.8000	8.0000
6	6.3333	6.8500	6.8667	7.2000	7.6000
7	8.0417	8.2750	8.0333	7.9500	8.5000
8	4.7917	5.4000	5.5333	5.9000	6.5000
9	4.2292	4.1750	4.5000	5.7500	6.9000
10	5.3125	5.7250	6.2333	6.5000	7.6000
11	5.9792	6.5000	6.7667	7.3500	8.1000
12	6.2500	6.7500	6.9333	7.6000	8.2000
13	5.6875	6.1250	6.4333	7.1000	7.5000
14	5.5625	6.2500	6.2667	6.1000	7.0000
15	5.9583	6.1000	6.8333	7.7500	8.8000
1	6.0006	5.9998	6.1338	6.3503	7.1005
2	7.0832	7.2989	7.7332	7.9511	8.0007
3	7.0835	6.6992	6.9663	7.3498	6.7008
4	6.1461	6.4499	6.5002	6.2011	6.3995
5	6.9376	7.3750	7.4668	7.7996	7.9998
6	6.3333	6.8504	6.8667	7.1991	7.6003
7	8.0418	8.2752	8.0329	7.9501	8.4988
8	4.7913	5.4003	5.5327	5.8998	6.4999
9	4.2286	4.1751	4.4994	5.7495	6.9002
10	5.3124	5.7250	6.2331	6.5000	7.6001
11	5.9793	6.4995	6.7672	7.3502	8.0998
12	6.2498	6.7497	6.9342	7.5991	8.1996
13	5.6878	6.1244	6.4331	7.1015	7.4991
14	5.5623	6.2503	6.2670	6.1008	7.0003
15	5.9584	6.0997	6.8333	7.7502	8.8006
1	6.0000	6.0000	6.1333	6.3500	7.1000
2	7.0833	7.3000	7.7333	7.9500	8.0000
3	7.0833	6.7000	6.9667	7.3500	6.7000
4	6.1458	6.4500	6.5000	6.2000	6.4000
5	6.9375	7.3750	7.4667	7.8000	8.0000
6	6.3333	6.8500	6.8667	7.2000	7.6000
7	8.0417	8.2750	8.0333	7.9500	8.5000
8	4.7917	5.4000	5.5333	5.9000	6.5000
9	4.2292	4.1750	4.5000	5.7500	6.9000
10	5.3125	5.7250	6.2333	6.5000	7.6000
11	5.9792	6.5000	6.7667	7.3500	8.1000
12	6.2500	6.7500	6.9333	7.6000	8.2000
13	5.6875	6.1250	6.4333	7.1000	7.5000
14	5.5625	6.2500	6.2667	6.1000	7.0000
15	5.9583	6.1000	6.8333	7.7500	8.8000

Table 3: Mean Absolute Deviation with Population Mean.

	n=48	n=40	n=30	n=20	n=10
PS89	0.0000	0.3636	0.5369	0.9092	1.4847
Gibbs	0.0002	0.3636	0.5370	0.9093	1.4846
ICM	0.0000	0.3636	0.5369	0.9092	1.4847

Since the true population mean us unknown, the sample value when $n = 48$ is taken to be the true population mean value. Tables 3 and 4

Table 4: Mean Square Error with Population Mean.

	n=48	n=40	n=30	n=20	n=10
PS89	0.0000	0.1699	0.3680	1.1066	2.8144
Gibbs	0.0000	0.1699	0.3680	1.1067	2.8145
ICM	0.0000	0.1699	0.3680	1.1066	2.8144

display the mean absolute deviations and mean square erors between this assumed true value and the estimated values for each sample size indicated for the PS89, Gibbs sampling, and ICM estimates of the population mean. Note that for all sample sizes considered, the exact Gibbs sampling and ICM estimates are nearly identical to the PS89 estimates which are the sample means.

5 Conclusion

The procedure of estimating the population mean by its sample value needed to be investigated. It has been shown that regardless of the sample size or the estimation procedure, estimating the population mean by the sample mean in Bayesian factor analysis is sufficient.

References

[1] A. E. Gelfand and A. F. M Smith. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, 6:721–741, 1984.

[3] M. Kendall. *Multivariate Analysis*. Charles Griffin & Company LTD, London, second edition, 1980.

[4] D. V. Lindley and A. F. M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B*, 34(1), 1972.

[5] A. O'Hagen. *Kendalls' Advanced Theory of Statistics, Volume 2B Bayesian Inference*. John Wiley and Sons Inc., New York, 1994.

[6] S. J. Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference.* Robert E. Krieger Publishing Company, Malabar, Florida, 1982.

[7] S. J. Press. *Bayesian Statistics: Principles, Models, and Applications.* John Wiley and Sons, New York, 1989.

[8] S. J. Press and K. Shigemasu. Bayesian inference in factor analysis. In *Contributions to Probability and Statistics*, chapter 15. Springer-Verlag, 1989.

[9] S. J. Press and K. Shigemasu. Bayesian inference in factor analysis-Revised. Technical Report No. 243, Department of Statistics, University of California, Riverside, CA 92521, May 1997.

[10] D. B. Rowe. *Correlated Bayesian Factor Analysis.* PhD thesis, Department of Statistics, University of California, Riverside, CA 92521, 1998.

[11] D. B. Rowe and S. J. Press. Gibbs sampling and hill climbing in Bayesian factor analysis. Technical Report No. 255, Department of Statistics, University of California, Riverside, CA 92521, May 1998.

[12] D. M. Shera. *Bayesian Factor Analysis.* PhD thesis, Harvard University, School of Public Health, Boston, MA 02115, 2000.