Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

BAYESIAN HYPOTHESIS TESTING OF EQUALITY OF NORMAL COVARIANCE MATRICES

BY H. J. KIM AND S. J. PRESS

Dongguk University and University of California, Riverside

This paper discusses hypothesis testing in multivariate analysis from a Bayesian point of view using the highest posterior density (HPD) region methodology. This approach is applied to the old, but still difficult problem of testing for the equality of normal covariance matrices, and a new Bayesian criterion is developed to carry out the test. Bartlett's classical test results as an approximate special case. It is shown that under the simple case of vague prior distributions for the covariance matrices a Bartlett-like test (Bartlett (1937)) results; but the degrees of freedom are lower, so the classical test weights the evidence against the null hypothesis of equality more heavily than is warranted by the posterior probability distribution, a result analogous to that of Berger and Selke, 1987. Moreover, more general (non-vague) prior distributions will generate a richer class of tests than were previously available.

1. Introduction. This paper concerns hypothesis testing in multivariate analysis from a Bayesian point of view. Generally, estimation and prediction are of much greater interest to Bayesian statisticians than is hypothesis testing, but there are those situations in which hypothesis testing is desirable and appropriate. Those situations are the ones with which we will be concerned in this paper.

We begin in Section2 with a brief summary of the method of Box-Tiao HPD region Bayesian hypothesis testing. In Section 3 and 4 we take up the problem of Bayesian testing for the equality of normal covariance matrices. In Section 3, we develop the joint posterior density for normal precision matrices. In Section 4 we develop the joint posterior density for the "ratios" of normal precision matrices, and we apply the HPD region method of hypothesis testing to develop new Bayesian tests for the equality of normal covariance matrices. In Section 5 we derive an asymptotic distribution appropriate for the required test.

AMS 1980 Subject Classifications: 62H15, 62F15

Key words and phrases: Hypotheses testing, HPD regions, normal covariance matrices.

2. Bayesian Highest Posterior Density (HPD) Region Testing. Box and Tiao, 1973 (p.122) introduced the notion of HPD regions. In the case of a one-dimensional quantity of interest, θ, they defined an "HPD interval" as an interval (a, b) such that for a posterior cdf F for θ, and a preassigned small $\alpha > 0$:

1. $F(b) - F(a) = 1 - \alpha$; and if $p(\cdot)$ denotes the posterior density for θ,
2. $p(\theta \mid X_1, \cdots, X_n)$ is greater than that for any other interval for which (1) holds.

REMARK 1. If (a, b) is an HPD interval, for any $\theta_1 \in (a, b)$ and any $\theta_2 \notin (a, b)$,

$$p(\theta_1 \mid \text{data}) \geq p(\theta_2 \mid \text{data}),$$

and conversely, subject to condition (1).

REMARK 2. For higher dimensional $\boldsymbol{\theta}$ Box and Tiao extended the idea to regions instead of intervals. We merely need to change condition (1) to

$$P\{\boldsymbol{\theta} \in \text{Region } R \mid \text{data}\} = 1 - \alpha.$$

REMARK 3. If $\boldsymbol{\phi} = f(\boldsymbol{\theta})$ defines a one-to-one transformation from $\boldsymbol{\theta}$ to $\boldsymbol{\phi}$, any region of content $(1 - \alpha)$ in the space of $\boldsymbol{\theta}$ transforms into a region of the same content in the space of $\boldsymbol{\phi}$, but the HPD region for $\boldsymbol{\theta}$ will not transform into an HPD region for $\boldsymbol{\phi}$, unless the transformation is linear.

REMARK 4. HPD regions are often ideally suited for testing hypotheses of interest in Bayesian multivariate analysis. This is because in higher dimensions we are generally interested in the event that some vector or matrix belongs to a particular region, and this event can generally be specified either directly or in terms of some monotonic function (for more details, see Box and Tiao, op. cit.).

REMARK 5. It might be noted that while some credibility regions might also be HPD regions, credibility regions need not be HPD regions. Moreover, while credibility regions might work well in one-dimensional problems, HPD regions generally work well in both one-dimensional and higher-dimensional problems.

The probability statements defining HPD regions can be derived directly from the posterior distribution; any kind of prior information may be used (vague or not) so that non-vague prior distributions will lead to a rich family of tests; and no multidimensional integrations are involved once we have the posterior distribution for $\boldsymbol{\theta}$ (we must of course be able to evaluate the integral for the cdf to evaluate the probability content of the distribution). In the sequel, we will adopt this HPD approach to develop a Bayesian test for the equality of normal covariance matrices.

3. Joint Posterior Density for Normal Precision Matrices.

Notation. We have independent p-variate observations from K normal populations, $X_1(i), \cdots, X_{N_i}(i) \sim N(\theta_i, \Sigma_i)$, $i = 1, \cdots, K$ with sample mean vectors $\overline{X}_i = N_i^{-1}\Sigma_1^{N_i} X_\alpha(i)$, and sample sums-of-squares matrices

$$V_i \equiv \Sigma_{\alpha=1}^{N_i}\left[X_\alpha(i) - \overline{X}(i)\right]\left[X_\alpha(i) - \overline{X}(i)\right]'.$$

Prior. In order to express the notion of "knowing little", and to provide a "reference-type prior" that often produces frequentist-types of results, we adopt a vague, Jeffreys-type of prior density (see Geisser and Cornfield (1963) and Jeffreys (1961)) for the mean vectors and the precision matrices, Σ_i^{-1},

$$g(\theta_1, \cdots, \theta_k, \Sigma_1^{-1}, \cdots, \Sigma_K^{-1}) \propto \prod_{i=1}^{K} |\Sigma_i|^{(p+1)/2}.$$

Results are easily extendable to natural conjugate families of prior distributions with little change in results (except for changes in the numbers of degrees of freedom).

Posterior. It is well-known (see e.g., Press, 1982) that the joint posterior distribution of the mean vectors and precision matrices is normal-Wishart. Integrating the posterior density with respect to the θ_i's yields the marginal posterior density for the precision matrices

$$\prod_{i=1}^{K} |\Sigma_i|^{-(n_i-p-1)/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{K} \operatorname{tr}(\Sigma_i^{-1}V_i)\right\},$$
$$i = 1, \cdots, K, \quad n_i \equiv N_i - 1. \tag{1}$$

4. Posterior Distribution of the "Ratios" of Normal Precision Matrices.

Testing for the equality of normal covariance matrices is a problem that was studied from a frequency point of view by Bartlett, 1937. The test (for $p > 1$) is an asymptotic result, and involves what later came to be called the multivariate beta distribution (see Olkin and Rubin (1964), Theorem 3.1).

We begin by making the transformations,

$$\Phi_i = \Sigma_K^{\frac{1}{2}}\Sigma_i^{-1}\Sigma_K^{\frac{1}{2}}, \qquad i = 1, \cdots, K-1,$$
$$Z_i = \Big(I_p + \sum_{i=1}^{K-1} T_i\Big)^{-\frac{1}{2}} T_i \Big(I_p + \sum_{i=1}^{K-1} T_i\Big)^{-\frac{1}{2}}, \qquad i = 1, \cdots, K-1, \tag{2}$$

where $T_i = V_K^{-\frac{1}{2}} V_i^{\frac{1}{2}} \Phi_i V_i^{\frac{1}{2}} V_K^{-\frac{1}{2}}$. The joint posterior distribution of $(T_1, T_2, \cdots, T_{K-1} \mid X)$ is generalized multivariate Beta II (see Kim and Press, 1992,and Tan, 1969). This now leads to the following result.

THEOREM. *The joint posterior distribution of* $\boldsymbol{Z} \equiv (Z_1, \cdots Z_{K-1})$ *is multivariate Beta I, with density given by*

$$f(Z \mid X) = C_1 \prod_{i=1}^{K-1} |Z_i|^{(n_i-p-1)/2} \left| I_p - \sum_{i=1}^{K-1} Z_i \right|^{(n_K-p-1)/2}, \tag{3}$$

where $Z_i > 0$, $I_p - \sum_{i=1}^{K-1} Z_i > 0$, *and where*

$$C_1 = \frac{\Gamma_p(n/2)}{\prod_{i=1}^{K} \Gamma_p(n_i/2)}.$$

PROOF. The result follows by successively making the transformations: $Q = (I_p + \sum_{i=1}^{K-1} T_i)$, $Z_i = Q^{-\frac{1}{2}} T_i Q^{-\frac{1}{2}}$, $i = 1, 2, \cdots, K-2$, from $(T_1, T_2, \cdots, T_{K-1})$ to $(Z_1, Z_2, \cdots, Z_{K-2}, Q)$ with Jacobian $J(T_1, T_2, \cdots, T_{K-1} \rightarrow Z_1, Z_2, \cdots, Z_{K-2}, Q) = \prod_{i=1}^{K-2} |Q|^{\frac{p+1}{2}}$. Then from Q to Z_{K-1} by $Z_{K-1} = Q^{-\frac{1}{2}} T_{K-1} Q^{-\frac{1}{2}} = I_p - Q^{-1} - \sum_{i=1}^{K-2} Z_i$ with $J(Q \rightarrow Z_{K-1}) = |Q|^{p+1}$. The transformations and the relation, $(I_p + \sum_{i=1}^{K-1} T_i)^{-1} = I_p - \sum_{i=1}^{K-1} Z_i > 0$, yield the result. ∎

COROLLARY. *In the special case when* $\Phi_i = I_p$, $i = 1, 2, \cdots, K-1$, *the joint density of the* Z_i*'s takes on the specific value (depending only upon the data)*

$$\begin{aligned} f(Z \mid X) &\propto \frac{\prod_{i=1}^{K-1} |T_i^*|^{(n_i-p-1)/2}}{|I_p + \sum_{i=1}^{K-1} T_i^*|^{\sum_{i=1}^{K} (n_i-p-1)/2}} \\ &\equiv \frac{\prod_{i=1}^{K} |V_i|^{(n_i-p-1)/2}}{|\sum_{i=1}^{K} V_i|^{\sum_{i=1}^{K} (n_i-p-1)/2}}, \end{aligned} \tag{4}$$

where $T_i^* = V_K^{-\frac{1}{2}} V_i V_K^{-\frac{1}{2}}$.

PROOF. When $\Phi_i = I_p, Z_i = [I_p + V_K^{-\frac{1}{2}} (\Sigma_{i=1}^{K-1} V_i) V_K^{-\frac{1}{2}}]^{-\frac{1}{2}} V_K^{-\frac{1}{2}} V_i V_K^{-\frac{1}{2}} [I_p + V_K^{-\frac{1}{2}} (\Sigma_{i=1}^{K-1} V_i) V_K^{-\frac{1}{2}}]^{-\frac{1}{2}}$, for $i = 1, 2, \cdots, K-1$, and $|I_p - \sum_{i=1}^{K-1} Z_i|^{(n_K-p-1)/2} = |I_p + V_K^{-\frac{1}{2}} (\Sigma_{i=1}^{K-1} V_i) V_K^{-\frac{1}{2}}]^{-(n_K-p-1)/2}$. The result follows by noticing these expressions. ∎

Thus, under the hypothesis of equality of covariance matrices, $H : \Sigma_1 = \cdots = \Sigma_K$, $\Phi_i = I_p$ for $i = 1, \cdots, K-1$, and the transformation of the T_i's to the Z_i's is not sensitive to which covariance matrix is used as a reference matrix in Φ_i. We will be testing H against the alternative hypothesis A: the Σ_i's are unequal.

H.P.D. Region. The density function $f(Z \mid X)$ is a monotonic decreasing function of

$$M = -2\log W,$$

where, from (3),

$$W = \left(\frac{\delta_K}{\delta}\right)^{-(p\delta)/2} \prod_{i=1}^{K-1} \left|\frac{\delta_K}{\delta_i} Z_i\right|^{\delta_i/2} \left|I_p - \sum_{i=1}^{K-1} Z_i\right|^{\delta_K/2}, \tag{5}$$

$\delta_i = n_i - p - 1,\ i = 1, \cdots, K$, and $\delta = \sum_{i=1}^K \delta_i$.

Thus, the event $f(Z \mid X) > f(Z_0 \mid X)$ is equivalent to the event $M < -2\log W_0$, where Z_0 and W_0 are obtained by substituting a particular matrix Φ_0 for Φ in the expressions of Z and W, respectively. In particular, our interest is in the point $\Phi_0 = \{\Phi_0; \Phi_{0i} = I_p, \text{ for all } i = 1, \cdots, K-1\}$ which corresponds to the situation $\Sigma_1 = \cdots = \Sigma_K$. By the Corollary, we see that

$$-2\log W_0 = -\left\{C + \sum_{i=1}^{K} \delta_i \log|V_i| - \delta \log\left|\sum_{i=1}^{K} V_i\right|\right\}, \tag{6}$$

where $C = p\delta \log \delta - p\sum_{i=1}^K \delta_i \log \delta_i$. The right side of this expression is similar to Bartlett's criterion, except that n_i in Bartlett's criterion is here replaced by $\delta_i = n_i - p - 1,\ i = 1, \cdots, K$. So the classical test weights the evidence against H more heavily than is warranted by the posterior probability distribution, a result analogous to that of Berger and Selke (1987). The actual test becomes, reject A if

$$P\{M < -2\log W_0\} \leq 1 - \alpha,$$

for some preassigned α, where W_0 is defined in (6). But to carry out the test we need an asymptotic result for M. Such a result is given in the next section.

5. Asymptotic Distribution of M. The asymptotic distribution of M is obtained from the asymptotic Box approximation (Box (1949)). For our context the result is given in the lemma below.

LEMMA. *The h-th moment of W is*

$$E(W^h) = \Psi\left(\frac{\prod_{j=1}^p (\delta/2)^{\delta/2}}{\prod_{j=1}^p \prod_{i=1}^K (\delta_i/2)^{\delta_i/2}}\right)^h \prod_{i=1}^K \Gamma_p\left\{\frac{\delta_i(1+h)+p+1}{2}\right\}$$
$$\cdot \left[\Gamma_p\left\{\sum_{i=1}^K \frac{\delta_i(1+h)+p+1}{2}\right\}\right]^{-1}$$

and is independent of $V_1, \cdots, V_K$, *where* $\Psi = \frac{\Gamma_p(n/2)}{\prod_{i=1}^K \Gamma_p(n_i/2)}$.

PROOF. Using the generalized multivariate Beta I density integral (cf. Tan (1969)):

$$\int \prod_{i=1}^{K-1} |Z_i|^{a_i - \frac{p+1}{2}} \left| I_p - \sum_{i=1}^{K-1} Z_i \right|^{a_K - \frac{p+1}{2}} = \frac{\prod_{i=1}^K \Gamma_p(a_i)}{\Gamma_p(\sum_{i=1}^K a_i)},$$

where $a_i > \frac{1}{2}(p-1)$, $i = 1, \cdots, K$, we obtain, from the distribution of Z_i's, the moment generating function of M:

$$\begin{aligned} Ee^{tM} &= E(W)^{-2t} \\ &= \prod_{i=1}^K \left(\frac{\delta_i}{\delta}\right)^{\delta_i pt} \Psi \int \prod_{i=1}^{K-1} |Z_i|^{\frac{\delta_i}{2}(1-2t)} \left| I_p - \sum_{i=1}^{K-1} Z_i \right|^{\frac{\delta_K}{2}(1-2t)} \prod_{i=1}^{K-1} dZ_i \\ &= C_0 \Psi \prod_{i=1}^K \left(\frac{\delta_i}{\delta}\right)^{\delta_i pt}, \end{aligned}$$

where

$$C_0 = \frac{\prod_{i=1}^K \Gamma_p\{(\delta_i(1-2t) + p + 1)/2\}}{\Gamma_p\{(\sum_{i=1}^K \delta_i(1-2t) + p + 1)/2\}}.$$

Letting $h = -2t$ gives the result. ∎

Set

$$\begin{aligned} &b = p, \quad y_j = \delta/2, \quad \eta_j = \{1 - j + K(p+1)\}/2, \\ &\varepsilon_j = \delta(1-p)/2, \quad j = 1, \cdots, p. \\ &a = pK, \quad x_l = \delta_i/2, \quad \delta_l^* = \{1 - j + (p+1)\}/2, \\ &l = j, p+j, \cdots, (K-1)p + j, \\ &j = 1, \cdots, p. \\ &\beta_l = \delta_i(1-\rho)/2, \quad l = (i-1)p + 1, \cdots, ip, \ i = 1, \cdots, K. \end{aligned}$$

The equation in the lemma can be expressed as

$$E(W)^h = K \left(\frac{\prod_{j=1}^b y_j^{y_j}}{\prod_{l=1}^a x_l^{x_l}} \right)^h \frac{\prod_{l=1}^a \Gamma\{x_l(1+h) + \delta_l^*\}}{\prod_{j=1}^b \Gamma\{y_j(1+h) + \eta_j\}}, \qquad h = 0, 1, \cdots,$$

where $K = \prod_{j=1}^b \Gamma\{y_j + \eta_j\} / \prod_{l=1}^a \Gamma\{x_l + \delta_l^*\}$. Since $\sum_{l=1}^a x_l = \sum_{j=1}^b y_j = p\delta/2$ and $E(W)^0 = 1$, the random variable W, whose moments are certain functions of gamma functions, satisfies the conditions for Box's (1949) theorem of a general asymptotic expansion of the random variable (cf. Anderson (1984, p.311)): such that if we take a second order approximation to the distribution of $M = -2\log W$, the asymptotic H.P.D. region of probability content of the event

$$f(Z \mid X) > f(Z_0 \mid X) \equiv M \leq -2\log W_0,$$

is given by

$$\begin{aligned} P(\rho M \leq -2\rho\log W_0) =& P(\chi_f^2 \leq -2\rho\log W_0) + w_2(P(\chi_{f+4}^2 \leq -2\rho\log W_0)) \\ &- P(\chi_f^2 \leq -2\rho\log W_0)] + O\{(\delta)^{-3}\}, \end{aligned}$$

where

$$\begin{aligned} f &= p(p+1)(K-1)/2, \\ \rho &= \frac{\delta}{\sum_{i=1}^{K} n_i} - \left(\sum_{i=1}^{K}\frac{1}{\delta_i} - \frac{1}{\delta}\right)\frac{2p^2+3p-1}{6(p+1)(K-1)}, \end{aligned}$$

and

$$w_2 = \frac{1}{6}\left\{\sum_{l=1}^{a} B_3(\beta_l + \delta_l^*)/(\rho^2 x_l^2) - \sum_{j=1}^{b} B_3(\varepsilon_j + \eta_j)/(\rho^2 y_j^2)\right\}.$$

Here $B_3(h)$ denotes the Bernoulli polynomial of degree 3, so that $B_3(h) = h^3 - (3/2)h^2 + (1/2)h$.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analyses*, Second Edition, John Wiley and Sons, New York.

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests, *Proc. Royal Soc. of London* (A), **160**, 268–282.

BERGER, J. O. and SELLKE, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of p-values and Evidence *J. Amer. Stat. Assn.* **82**, 112–122.

BOX, G. E. P. (1949). A general distribution theory for a class of likelihood criteria, *Biometrika* **36**, 317–346.

BOX, G. E. P. and TIAO, G.C. (1973). *Bayesian Inference in Statistical Analysis*, Addison Wesley Pub. Co.: Reading, MA.

GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters, *Jour. Roy. Statist. Soc.* (B) **25**, 368–376.

JEFFREYS, H. (1961). Theory of Probability, Third Edition, Clarendon Press: Oxford.

KIM, H. J. and PRESS, S.J. (1992). A Bayesian criterion for testing equality of multivariate normal covariance matrices. Tech Report No.193R, June, 1992, Department of Statistics, University of California, Riverside.

OLKIN, I. and RUBIN, H. (1964). Multivariate beta distributions and independence properties of the wishart distribution, *Ann. Math. Statist.* **35**, 261–269.

PRESS, S. JAMES. (1982). Applied Multivariate Analyses: Using Bayesian and Frequentist Methods of Inference, Second Edition, Malabar, Florida: Robert E. Krieger Publishing Company.

TAN, W.Y. (1969). Note on the multivariate and the generalized multivariate Beta Distributions, *J. Amer. Stat. Assn.* **64**, 230–241.

DEPARTMENT OF STATISTICS
DONGGUK UNIVERSITY
SEOUL 100, KOREA

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
RIVERSIDE, CA 92521, U.S.A.