A BAYESIAN ANALOGUE OF PAULSON'S LEMMA AND ITS USE IN TOLERANCE REGION CONSTRUCTION WHEN SAMPLING FROM THE MULTI-VARIATE NORMAL

IRWIN GUTTMAN*

(Received Feb. 16, 1970)

1. Introduction

Suppose we are sampling on a k dimensional random variable X, defined over R^k, Euclidean space of k dimensions, and let $\{A\}=\mathfrak{A}$ denote a σ-algebra of subsets A of R^k. We assume that $\mathfrak{A}\subset B=$Borel subsets of R^k.

Suppose further that X has the absolutely continuous distribution

$$(1.1)\qquad F(\boldsymbol{x}\mid\boldsymbol{\theta})=\int_{-\infty}^{x_k}\cdots\int_{-\infty}^{x_1} f(y_1,\cdots,y_k\mid\boldsymbol{\theta})dy_1\cdots dy_k$$

and $\boldsymbol{\theta}\in\Omega$, with Ω an indexing set which we will refer to as the parameter space. We denote a random sample of n independent observations on X by $(X_1,\cdots,X_n)$ or $\{X_i\}$. We have the following definitions.

DEFINITION 1.1. A statistical tolerance region $S(X_1,\cdots,X_n)$ is a statistic defined over $R^k\times\cdots\times R^k=R^{kn}$, which takes values in the σ-algebra $\mathfrak{A}$.

This definition, then, implies that a statistical tolerance region is a statistic which is a set function, and maps "the point" $(X_1,\cdots,X_n)\in R^{kn}$ into the region $S(X_1,\cdots,X_n)\in\mathfrak{A}$, that is, $S(\{X_i\})\subset R^k$. When constructing such statistical tolerance regions, various criteria may be used. One that is often borne in mind is contained in the following definition.

DEFINITION 1.2. $S(X_1,\cdots,X_n)$ is a β-expectation statistical tolerance region if

$$(1.2)\qquad E_{\{X_i\}}\{F[S(X_1,\cdots,X_n)\mid\boldsymbol{\theta}]\}=\beta$$

for all $\boldsymbol{\theta}\in\Omega$, where

$$(1.2a)\qquad F[S\mid\boldsymbol{\theta}]=\int\cdots\int_S f(\boldsymbol{y}\mid\boldsymbol{\theta})d\boldsymbol{y}\ .$$

* This research was supported in part by the Wisconsin Alumni Research Foundation.
Present address is: Centre de Recherches Mathématiques, Université de Montréal.

The quantity $F[S\,|\,\boldsymbol{\theta}]=F[S(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n)\,|\,\boldsymbol{\theta}]$ is called the coverage of the region S, and will be denoted by $C[S]$. We note that it can be viewed as the probability of an observation, say, $\boldsymbol{Y}$, falling in S, where $\boldsymbol{Y}$ is independent of $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$ and, of course, has distribution $F(\boldsymbol{y}\,|\,\boldsymbol{\theta})$. Now because $S(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n)$ is a random set function, $F[S(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n)\,|\,\boldsymbol{\theta}]$ is a random variable and has the distribution of its own. Hence, constructing an S to satisfy (1.2), simply implies that we are imposing the condition that S be such that the distribution of its coverage $F[S\,|\,\boldsymbol{\theta}]$ has expectation (mean value) β. Paulson [6] has given a very interesting connection between statistical tolerance regions and prediction regions.

PAULSON'S LEMMA. *If on the basis of a given sample on a k-dimensional random variable* $\boldsymbol{X}$, *a k-dimensional "confidence" region* $S(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n)$ *of level* β *is found for statistics* $T_1,\cdots,T_k$, *where* $T_i=T_i(\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_q)$, *where the* $(k\times 1)$ *vector observations* $\boldsymbol{Y}_j$, $j=1,\cdots,q$ *are independent observations on* $\boldsymbol{X}$, *and independent of* $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$, *and if* C *is defined to be such that*

$$C=\int_S dG(\boldsymbol{t}) \tag{1.3}$$

where $G(\boldsymbol{t})$ *is the distribution function of* $\boldsymbol{T}$ $(\boldsymbol{T}'=(T_1,\cdots,T_k)$, *and* $T_i=T_i\,(\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_q)$, *then*

$$E[C]=\beta\ . \tag{1.4}$$

Before we prove this lemma, we remark that our interest will be for the case $q=1$, (that is, we will have one future observation $\boldsymbol{Y}_1=\boldsymbol{Y}$), and $T_i=T_i(\boldsymbol{Y})=Y_i$ so that $\boldsymbol{T}=\boldsymbol{Y}$ and $G(\boldsymbol{t})=F(\boldsymbol{y}\,|\,\boldsymbol{\theta})$. Hence, C given by (1.3) is simply the coverage of $S(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n)$. Note that C depends here on S and $\boldsymbol{\theta}$ and indeed we may write $C=C_{\theta}(S)$.

In these circumstances, then, Paulson's Lemma then gives us an operational method for constructing a statistical tolerance region of β-expectation, namely:

Find S, a prediction region of level β for a future observation $\boldsymbol{Y}$. If this is done, then S is a tolerance region of β-expectation.

PROOF OF PAULSON'S LEMMA. The joint distribution function of $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$, is

$$\prod_{i=1}^{n} F(\boldsymbol{x}_i\,|\,\boldsymbol{\theta})\ .$$

Now the left-hand side of (1.4) may be written as

$$E[C]=\int_{R^{kn}}\int_S dG(\boldsymbol{t})d\prod_{i=1}^{n} F(\boldsymbol{x}_i\,|\,\boldsymbol{\theta})\ . \tag{1.5}$$

But the right-hand side of (1.5) is the probability that $\boldsymbol{T}$ lies in S, and we are given that $S=S(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n)$ is β-level confidence region for $\boldsymbol{T}$ (or prediction region for $\boldsymbol{T}$). Hence, the right-hand side of (1.5) has value β, and we have that $E[C]=\beta$.

As an illustration of the above, we take the case $q=1$, and suppose that sampling is on the k-dimensional normal variable $N(\boldsymbol{\mu},\Sigma)$. It is well known that a $100\beta\%$ prediction region for $\boldsymbol{Y}$, where $\boldsymbol{Y}=N(\boldsymbol{\mu},\Sigma)$, constructed on the basis of the random sample of n independent observations $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$ [where $\boldsymbol{Y}$, $\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n$ are all independent] is

$$S(\{\boldsymbol{X}_i\})=\{\boldsymbol{Y}\mid(\boldsymbol{Y}-\bar{\boldsymbol{X}})'V^{-1}(\boldsymbol{Y}-\bar{\boldsymbol{X}})\leqq C_\beta\} \tag{1.6}$$

where the mean vector $\bar{\boldsymbol{X}}$ and the sample variance-covariance matrix $\boldsymbol{V}$ are defined by

$$\begin{aligned}\bar{\boldsymbol{X}}&=(\bar{X}_1,\cdots,\bar{X}_k)=n^{-1}\left(\sum_{i=1}^n X_{i1},\cdots,\sum_{i=1}^n X_{ik}\right)'=n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\\ \boldsymbol{V}&=(n-1)^{-1}\sum_{i=1}^n(\boldsymbol{X}_i-\bar{\boldsymbol{X}})(\boldsymbol{X}_i-\bar{\boldsymbol{X}})'\ ,\end{aligned} \tag{1.6a}$$

with C_β given by

$$C_\beta=[(n-1)k/(n-k)][1+n^{-1}]F_{k,n-k;1-\beta} \tag{1.6b}$$

and $F_{k,n-k;1-\beta}$ is the point exceeded with probability $1-\beta$ when using the Snedecor-F distribution with $(k,n-k)$ degrees of freedom. Hence, by Paulson's Lemma, S given by (1.6) is a β-expectation tolerance region. This region is known to have certain optimum properties—see, for example, Fraser and Guttman [1].

We now approach the problem of constructing "β-expectation" tolerance regions from the Bayesian point of view. We will see that there is a direct analogue of Paulson's Lemma, which arises in a natural and interesting way.

2. The Bayesian approach

In the Bayesian framework, references about the parameters $\boldsymbol{\theta}$ in a statistical model are summarized by the posterior distribution of the parameters which are obtained by the use of a theorem due to the Reverend Thomas Bayes. This theorem, a simple statement of conditional probability, states that the distribution of the parameters $\boldsymbol{\theta}$, given that $\boldsymbol{X}_i$ is observed to be $\boldsymbol{x}_i$, $i=1,\cdots,n$, is

$$p[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]=cp(\boldsymbol{\theta})p[\{\boldsymbol{x}_i\}\mid\boldsymbol{\theta}]\ , \tag{2.1}$$

where the normalizing constant c is such that

$$c^{-1}=\int_{\Omega} p(\boldsymbol{\theta})p[\{\boldsymbol{x}_i\}\mid\boldsymbol{\theta}]d\boldsymbol{\theta}\ , \tag{2.1a}$$

$p(\boldsymbol{\theta})$ is the (marginal) distribution of the vector of parameters $\boldsymbol{\theta}$, and $p[\{\boldsymbol{x}_i\}\mid\boldsymbol{\theta}]$ is the distribution of the observations $\{\boldsymbol{X}_i\}$, given $\boldsymbol{\theta}$. When we are indeed given that $\{\boldsymbol{X}_i\}=\{\boldsymbol{x}_i\}$, then we often call $p[\{\boldsymbol{x}_i\}\mid\boldsymbol{\theta}]$ the likelihood function of $\boldsymbol{\theta}$ and denote it be $l[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]$.

The ingredients of (2.1) may be interpreted as follows. The distribution $p(\boldsymbol{\theta})$ represents our knowledge about the parameters $\boldsymbol{\theta}$ before the data are drawn, while $l[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]$ represents information given to us about $\boldsymbol{\theta}$ from the data $\{\boldsymbol{x}_i\}$, and finally, $p[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]$ represents our knowledge of $\boldsymbol{\theta}$ after we observe the data. For these reasons, $p(\boldsymbol{\theta})$ is commonly called the *a-priori* or *prior* distribution of $\boldsymbol{\theta}$, and $p[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]$, the *a-posteriori* or *posterior* distribution of $\boldsymbol{\theta}$. Using this interpretation, then, Bayes' theorem provides a formal mechanism by which our a-priori information is combined with sample information to give us the *posterior distribution* of $\boldsymbol{\theta}$, which effectively summarizes all the information we have about $\boldsymbol{\theta}$.

Now the reader will recall that S is a tolerance region of β-expectation if its coverage $C[S]$ has *expectation* β. Now, from a Bayesian point of view, once having seen the data, that is, having observed $\{\boldsymbol{X}_i\}=\{\boldsymbol{x}_i\}$, then $C[S]=\int_S f(\boldsymbol{y}\mid\boldsymbol{\theta})d\boldsymbol{y}$ is a function only of the parameters $\boldsymbol{\theta}$, and the expectation referred to is the expectation with respect to the *posterior distribution* of the parameters $\boldsymbol{\theta}$, that is, on the basis of the given data $\{\boldsymbol{x}_i\}$, we wish to construct S such that

$$E[C[S]\mid\{\boldsymbol{x}_i\}]=\int_{\Omega}\int_S f(\boldsymbol{y}\mid\boldsymbol{\theta})p[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]d\boldsymbol{y}d\boldsymbol{\theta}=\beta \tag{2.2}$$

where $\boldsymbol{Y}$ has the same distribution as the $\boldsymbol{X}_i$, namely $f(\cdot\mid\boldsymbol{\theta})$. Now it is interesting to note that, assuming the conditions of Fubini's Theorem hold, so that we may invert the order of integration, we have that

$$\begin{aligned} E[C[S]\mid\{\boldsymbol{x}_i\}]&=\int_S\int_{\Omega} f(\boldsymbol{y}\mid\boldsymbol{\theta})p[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]d\boldsymbol{\theta}d\boldsymbol{y} \\ &=\int_S h[\boldsymbol{y}\mid\{\boldsymbol{x}_i\}]d\boldsymbol{y}\ . \end{aligned} \tag{2.3}$$

Now the density $h[\boldsymbol{y}\mid\{\boldsymbol{x}_i\}]$, where

$$h[\boldsymbol{y}\mid\{\boldsymbol{x}_i\}]=\int_{\Omega} f(\boldsymbol{y}\mid\boldsymbol{\theta})p[\boldsymbol{\theta}\mid\{\boldsymbol{x}_i\}]d\boldsymbol{\theta} \tag{2.4}$$

is (examining the right-hand side of (2.4)) simply the conditional distribution of $\boldsymbol{Y}$, given the data $\{\boldsymbol{x}_i\}$, where $\boldsymbol{Y}$ may be regarded as an additional observation from $f(\boldsymbol{x}\mid\boldsymbol{\theta})$, additional to and independent of $\boldsymbol{X}_1$,

$\cdots, X_n$. This density $h[\boldsymbol{y} \mid \{\boldsymbol{x}_i\}]$ is the Bayesian estimate of the distribution of $\boldsymbol{Y}$ and has been called the predictive or future distribution of $\boldsymbol{Y}$. (For further discussion, see, for example, Guttman [4] and the references cited therein.)

Hence, (2.2) and (2.3) have the very interesting implication that S is a β-expectation tolerance region if it is a (predictive) β-confidence region for $\boldsymbol{Y}$, where $\boldsymbol{Y}$ has the predictive distribution $h[\boldsymbol{y} \mid \{\boldsymbol{x}_i\}]$ defined by (2.4). This is the Bayesian analogue of Paulson's Lemma given in Section 1 with $q=1$. To repeat in another way, for a particular f, we need only find $h[\boldsymbol{y} \mid \{\boldsymbol{x}_i\}]$ and a region S that is such that

$$\Pr(\boldsymbol{Y} \in S) = \int_S h[\boldsymbol{y} \mid \{\boldsymbol{x}_i\}] d\boldsymbol{y} = \beta \,. \tag{2.5}$$

We summarize the above results in the following lemma.

Lemma 2.1. *If on the basis of observed data* $\{\boldsymbol{x}_i\}$, a predictive β-*confidence region* $S=S(\{\boldsymbol{x}_i\})$ *is constructed such that*

$$\int_S h[\boldsymbol{y} \mid \{\boldsymbol{x}_i\}] d\boldsymbol{y} = \beta \tag{2.6}$$

where the predictive distribution is given by (2.4), *and if* $C[S]$ *is the coverage of* S, *that is*

$$C[S] = C[S(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)] = \int_S f(\boldsymbol{y} \mid \boldsymbol{\theta}) d\boldsymbol{y} \tag{2.7}$$

where f *is the common distribution of the independent random variables,* $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$, $\boldsymbol{Y}$, *then the posterior expectation of* $C[S]$ *is* β, *that is* S *is of* β-*expectation.*

(The proof is simple and utilizes relations (2.3) and (2.6).)

3. Sampling from the k-variate normal

We suppose in this section that sampling is from the k-variate normal $N(\boldsymbol{\mu}, \Sigma)$, whose distribution is given by

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma) = (2\pi)^{-k/2} |\Sigma^{-1}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} , \tag{3.1}$$

where $\boldsymbol{\mu}$ is $(k \times 1)$ and Σ is a $(k \times k)$ symmetric positive definite matrix. It is convenient to work with the set of parameters $(\boldsymbol{\mu}, \Sigma^{-1})$ here, and accordingly, suppose that the prior for this situation is the conjugate prior or Raiffa and Schlaifer given by

$$(3.2) \quad p(\boldsymbol{\mu}, \Sigma^{-1})d\boldsymbol{\mu}d\Sigma^{-1} \propto |\Sigma^{-1}|^{(n_0-k-1)/2} \cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}[(n_0-1)V_0+n_0(\boldsymbol{\mu}-\bar{\boldsymbol{x}}_0)(\boldsymbol{\mu}-\bar{\boldsymbol{x}}_0)']\right\}d\boldsymbol{\mu}d\Sigma^{-1}$$

where $\bar{\boldsymbol{x}}_0$ is a $(k\times 1)$ vector of known constants and V_0 is a $(k\times k)$ symmetric positive definite matrix of known constants*. It is to be noted that if n_0 tends to 0 and $(n_0-1)V_0$ tends to the zero matrix, then (3.2) tends to the "in-ignorance" prior advocated by Geisser [2] and Geisser and Cornfield [3], viz

$$(3.3) \quad p(\boldsymbol{\mu}, \Sigma^{-1})d\boldsymbol{\mu}d\Sigma^{-1} \propto |\Sigma^{-1}|^{-(k+1)/2}d\boldsymbol{\mu}d\Sigma^{-1}\ .$$

Now it is easy to see that if n independent observations $\boldsymbol{X}_i$ are taken from (3.1), and we observe $\{\boldsymbol{X}_i\}$ to be $\{\boldsymbol{x}_i\}$, that the likelihood function is given by

$$(3.4) \quad l[\boldsymbol{\mu}, \Sigma^{-1}|\{\boldsymbol{x}_i\}]=(2\pi)^{-nk/2}|\Sigma^{-1}|^{n/2} \cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}[(n-1)V+n(\bar{\boldsymbol{x}}-\boldsymbol{\mu})(\bar{\boldsymbol{x}}-\boldsymbol{\mu})']\right\}$$

where

$$\bar{\boldsymbol{x}}=n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i$$

and

$$(n-1)V=\sum_{i=1}^{n}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})'\ .$$

(the abbreviation "tr A" stands for the trace of the matrix A.) Now combining (3.2) and (3.4) using Bayes' Theorem gives us that the posterior of $(\boldsymbol{\mu}, \Sigma^{-1})$ is such that

$$(3.5) \quad p[\boldsymbol{\mu}, \Sigma^{-1}|\{\boldsymbol{x}_i\}] \propto |\Sigma^{-1}|^{(n+n_0-k-1)/2} \cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}[(n_0-1)V_0+(n-1)V + n_0(\boldsymbol{\mu}-\bar{\boldsymbol{x}}_0)(\boldsymbol{\mu}-\bar{\boldsymbol{x}}_0)'+n(\boldsymbol{\mu}-\bar{\boldsymbol{x}})(\boldsymbol{\mu}-\bar{\boldsymbol{x}})']\right\}\ .$$

Now the exponent of (3.5) may be written in simplified form on "completing the square" in $\boldsymbol{\mu}$, that is, the term in square brackets in the exponent of (3.5) may be written, after some algebra, as

$$(3.6) \quad (n+n_0)(\boldsymbol{\mu}-\bar{\bar{\boldsymbol{x}}})(\boldsymbol{\mu}-\bar{\bar{\boldsymbol{x}}})'+(n_0-1)V_0+(n-1)V+R$$

* We are in effect saying that our prior information on $\boldsymbol{\mu}$ and Σ^{-1} is such that we expect $\boldsymbol{\mu}$ to be $\bar{\boldsymbol{x}}_0$, with dispersion, that is, variance co-variance matrix of $\boldsymbol{\mu}$, to be $(n_0-1)V_0/[n_0(n_0-k-2)]$, and that we expect Σ^{-1} to be V_0^{-1} etc.

where

(3.7) $$\bar{\bar{x}}=(n+n_0)^{-1}(n_0\bar{x}_0+n\bar{x})$$

and

(3.7a) $$R=n\bar{x}\bar{x}'+n_0\bar{x}_0\bar{x}_0'-(n+n_0)(\bar{\bar{x}}\bar{\bar{x}}')=\frac{nn_0}{n+n_0}(\bar{x}-\bar{x}_0)(\bar{x}-\bar{x}_0)' .$$

Hence, we may now write (3.5) as follows:

(3.8) $$p(\boldsymbol{\mu}, \Sigma^{-1} \mid \{x_i\})=c\,|\Sigma^{-1}|^{(n+n_0-k-1)/2} \cdot \exp\left\{-\frac{1}{2}[\operatorname{tr}\Sigma^{-1}Q+(n_0+n)(\boldsymbol{\mu}-\bar{\bar{x}})'\Sigma^{-1}(\boldsymbol{\mu}-\bar{\bar{x}})]\right\}$$

where

(3.8a) $$Q=(n_0-1)V_0+(n-1)V+\frac{n_0n}{n_0+n}(\bar{x}-\bar{x}_0)(\bar{x}-\bar{x}_0)' ,$$

and c is the normalizing constant necessary to make (3.8) a density, that is, integrate to 1. Now to determine c, we first integrate with respect to $\boldsymbol{\mu}$ and then Σ^{-1}, and in so doing, we make use of the identities derived from the k-variate normal and k-order Wishart distributions, viz

(3.9a) $$\int\cdots\int \exp\left\{-\frac{1}{2}[(\boldsymbol{\eta}-\boldsymbol{a})'N^{-1}(\boldsymbol{\eta}-\boldsymbol{a})]\right\}d\boldsymbol{\eta}=(2\pi)^{k/2}\,|N^{-1}|^{-1/2}$$

(3.9b) $$\int\cdots\int |H|^{(m-k-1)/2}\exp\left(-\frac{1}{2}\operatorname{tr} M^{-1}H\right)dH =2^{mk/2}\pi^{k(k-1)/4}\,|M|^{m/2}\prod_{i=1}^{k}\Gamma[(m+1-i)/2] .$$

As is easily verified, performing the integration yields

(3.10) $$c=(n+n_0)^{k/2}\,|Q|^{(n+n_0-1)/2}\Big/2^{k(n+n_0)/2}\pi^{k(k+1)/4}\prod_{i=1}^{k}\Gamma[(n+n_0-i)/2] .$$

We are now in the position of being able to find the predictive density of a future observation $\boldsymbol{Y}$, conditional on $\boldsymbol{x}$. The first factor of the integrand of (2.4) has functional form (3.1), and the second factor is given by (3.8). Hence, we find that the predictive distribution of $\boldsymbol{Y}$ is given by

(3.11) $$h[\boldsymbol{y} \mid \{x_i\}]=\int\cdots\int\int\cdots\int c(2\pi)^{-k/2}\,|\Sigma^{-1}|^{(n+n_0-k)/2} \cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}[Q+W]d\boldsymbol{\mu}d\Sigma^{-1}\right\}$$

where c is given by (3.10), Q by (3.8a), and where W is such that

$$W=(n+n_0)(\mu-\bar{\bar{x}})(\mu-\bar{\bar{x}})'+(\mu-y)(\mu-y)' . \tag{3.12}$$

Again "completing the square" in μ in (3.12), we easily, but tediously, find that

$$W=(n+n_0+1)(\mu-\bar{\bar{y}})(\mu-\bar{\bar{y}})'+\frac{n+n_0}{n+n_0+1}(y-\bar{\bar{x}})(y-\bar{\bar{x}})' \tag{3.13}$$

with

$$\bar{\bar{y}}=(n+n_0+1)^{-1}[(n+n_0)\bar{\bar{x}}+y] .$$

The integration with respect to μ in (3.11), using (3.9a), gives us

$$h[y\,|\,\{x_i\}]=\int\cdots\int c(n+n_0+1)^{-k/2}\,|\Sigma^{-1}|^{(n+n_0-k-1)/2} \cdot \exp\left\{-\frac{1}{2}\operatorname{tr}\Sigma^{-1}\left[Q+\frac{n+n_0}{n+n_0+1}(y-\bar{\bar{x}})(y-\bar{\bar{x}})'\right]\right\}d\Sigma^{-1} . \tag{3.14}$$

Integrating (3.14) with the help of (3.9b), and substituting for the value of c given by (3.10), we find that

$$h[y\,|\,\{x_i\}]=\left(\frac{n+n_0}{n+n_0+1}\right)^{k/2}\frac{\Gamma[(n+n_0)/2]\,|Q^{-1}|^{1/2}}{\Pi^{k/2}\Gamma[(n+n_0-k)/2]} \cdot \left|I+\frac{n+n_0}{n+n_0+1}Q^{-1}(y-\bar{\bar{x}})(y-\bar{\bar{x}})'\right|^{-(n+n_0)/2} . \tag{3.15}$$

Now using the identity (proved in the appendix)

$$|I_{n_1}-AB|=|I_{n_2}-BA| \tag{3.16}$$

where A is $(n_1\times n_2)$ and B is $(n_2\times n_1)$, we have the result that the predictive density of Y is given by

$$h[y\,|\,\{x_i\}]=\left(\frac{n+n_0}{n+n_0+1}\right)^{k/2}\frac{\Gamma[(n+n_0)/2]\,|Q^{-1}|^{1/2}}{\Pi^{k/2}\Gamma[(n+n_0-k)/2]} \cdot \left(1+\frac{n+n_0}{n+n_0+1}(y-\bar{\bar{x}})'Q^{-1}(y-\bar{\bar{x}})\right)^{-(n+n_0)/2} \tag{3.17}$$

that is, we have the interesting result that the predictive distribution of Y, given $\{x_i\}$, is related to the k-variate t-distribution, degrees of freedom $(n+n_0-k)$. As may be seen from properties of the multivariate-t (see, for example, Tiao and Guttman [8]), we have that

$$\frac{n_0+n}{n+n_0+1}(Y-\bar{\bar{x}})'Q^{-1}(Y-\bar{\bar{x}})=\frac{k}{n+n_0-k}F_{k,n+n_0-k} . \tag{3.18}$$

Suppose now that we are interested in the "central" 100β % of the normal distribution (3.1), that is, in the set

$$A^k=\{\boldsymbol{y}\,|\,(\boldsymbol{y}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\leqq\chi^2_{k,1-\beta}\} \tag{3.19}$$

where $\chi^2_{k,1-\beta}$ is the point exceeded with probability $(1-\beta)$ when using the chi-square distribution with k degrees of freedom. Given μ and Σ or (μ, Σ^{-1}), we have that

$$P(\boldsymbol{Y}\in A^k\,|\,\boldsymbol{\mu}, \Sigma)=\beta \tag{3.20}$$

that is, if we knew $(\boldsymbol{\mu}, \Sigma)$, A^k would be a $100\beta\,\%$ predictive region for $\boldsymbol{Y}$. Now since we don't know $(\boldsymbol{\mu}, \Sigma)$, we use as an estimator of the density (3.1), the density (3.17), after observing the data $\{\boldsymbol{x}_i\}$. Hence, a sensible predictive region for $\boldsymbol{Y}$ is the central ellipsoidal region of (3.17), namely the region

$$S(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n)=\left\{\boldsymbol{y}\,\middle|\,\frac{n_0+n}{n_0+n+1}(\boldsymbol{y}-\bar{\bar{\boldsymbol{x}}})'[Q/(n_0+n-k)]^{-1}(\boldsymbol{y}-\bar{\bar{\boldsymbol{x}}})\leqq kF_{k,n+n_0-k;1-\beta}\right\} \tag{3.21}$$

and it is easy to see, from (3.18), that

$$P(\boldsymbol{Y}\in S\,|\,\{\boldsymbol{x}_i\})=\beta\,. \tag{3.22}$$

Thus, by Lemma 2.1, we have that S defined by (3.21) is a tolerance region of (posterior) β-expectation.

We note that if $n_0=0$ and $(n_0-1)V_0$ is the zero matrix, that is, if the so-called "in-ignorance" prior given by (3.3) is the appropriate prior, then the above results imply that the β-expectation region is of the form

$$S(\{\boldsymbol{x}_i\})=\left\{\boldsymbol{y}\,\middle|\,\frac{n}{n+1}(\boldsymbol{y}-\bar{\boldsymbol{x}})'[(n-1)V/(n-k)]^{-1}(\boldsymbol{y}-\bar{\boldsymbol{x}})\leqq kF_{k,n-k;1-\beta}\right\} \tag{3.23}$$

which is interesting, since this latter result is in agreement with the sampling theory result (1.6), as may be easily verified.

It is to be finally remarked, that the lemma of Section 2 is quite general and may be used when sampling is from any population. In fact, the case of the single exponential is discussed in Guttman [5].

Appendix

We give a proof, due to George Tiao, of (3.16). Consider the matrix equations (A is $(n_1\times n_2)$ and B is $(n_2\times n_1)$)

$$\begin{bmatrix} I_{n_1} & A \\ B & I_{n_2}\end{bmatrix}\begin{bmatrix} I_{n_1} & -A \\ 0 & I_{n_2}\end{bmatrix}=\begin{bmatrix} I_{n_1} & 0 \\ B & I_{n_2}-BA\end{bmatrix} \tag{A.1}$$

and

$$\left[\begin{array}{c|c} I_{n_1} & A \\ \hline B & I_{n_2} \end{array}\right]\left[\begin{array}{c|c} I_{n_1} & 0 \\ \hline -B & I_{n_2} \end{array}\right]=\left[\begin{array}{c|c} I_{n_1}-AB & A \\ \hline 0 & I_{n_2} \end{array}\right]. \tag{A.2}$$

Now taking determinants of both sides of (A.1) and (A.2) yields

$$\begin{gathered} |M|\cdot|I_{n_1}|\cdot|I_{n_2}|=|I_{n_1}|\cdot|I_{n_2}-BA| \\ \text{and} \\ |M|\cdot|I_{n_1}|\cdot|I_{n_2}|=|I_{n_1}-AB|\cdot|I_{n_2}| \end{gathered} \tag{A.3}$$

where

$$M=\left[\begin{array}{c|c} I_{n_1} & A \\ \hline B & I_{n_2} \end{array}\right].$$

Using (A.3) we have the result

$$|I_{n_1}-AB|=|I_{n_2}-BA|=|M|\ .$$

UNIVERSITY OF WISCONSIN AND UNIVERSITY OF MASSACHUSETTS

REFERENCES

[1] Fraser, D. A. S. and Guttman, Irwin (1956). Tolerance regions, *Ann. Math. Statist.*, **27**, 162–179.

[2] Geisser, S. (1965). Bayesian estimation in multivariate analysis, *Ann. Math. Statist.*, **36**, 150–159.

[3] Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters, *Jour. Roy. Statist. Soc.*, Ser. B, **25**, 368–376.

[4] Guttman, Irwin (1967). The use of the concept of a future observation in goodness-of-fit problems, *Jour. Roy. Statist. Soc., Ser. B*, **29**, 83–100.

[5] Guttman, Irwin (1968). Tolerance regions: A survey of its literature. VI. The Bayesian approach, *Tech. Rep.* No. 126, *Department of Statistics, University of Wisconsin.*

[6] Paulson, E. (1943). A note on tolerance limits, *Ann. Math. Statist.*, **14**, 90–93.

[7] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, Harvard University Press.

[8] Tiao, G. C. and Guttman, Irwin (1965). The multivariate inverted beta distribution with applications, *Jour. Amer. Statist. Assoc.*, **60**, 793–805.