*The Annals of Statistics*
1981, Vol. 9, No. 4, 768–776

# INNER STATISTICAL INFERENCE II

BY C. VILLEGAS

*Simon Fraser University*

According to an invariance principle, for some models having a certain group structure, there is a uniquely defined prior representing ignorance, which is called the inner prior. It is shown that the corresponding posterior probability of a likelihood region has a simple frequency interpretation as a mean conditional confidence level. The central multivariate normal model is considered as an example.

**1. Introduction.** The central problem of statistical inference, as seen from a logical Bayesian viewpoint, is to find a logical posterior distribution of the unknown parameter that represents all the information contained in the data and the model, or, equivalently, to find a logical prior that represents ignorance concerning the true value of the parameter.

It has to be strongly emphasized that logical priors are relative to a given group structure. In the present paper it will be shown that, if only one group is considered as an integral part of the model, then under suitable conditions there is a uniquely defined logical prior, called the inner prior, which, according to an invariance argument, represents ignorance concerning the true value of the unknown parameter. It will be shown also that inner posterior probabilities of likelihood sets have a simple frequency interpretation as mean conditional confidence levels.

The concept of an inner prior is due to Peisakoff (1950, page 40). The frequency interpretation for inner inferences was developed by Villegas (1977a) for the univariate normal model. Applications of this approach to multivariate models can be found in Villegas (1971, 1972, 1977b).

For some background summary information on groups and invariant measures, the reader is referred to Chapter 2 of Giri (1977); and to Section 2.1 of Dawid, Stone and Zidek (1973). For more complete information, the reader is referred to Nachbin (1965) and to Chapter 1 of Bourbaki (1974).

**2. The inner prior.** In the present paper "random variable" will be synonymous with "random quantity" in the sense of de Finetti (1974), but the concept of probability will not be restricted to have only a subjective interpretation.

In many important cases the sample space $\mathscr{X}$ has a geometric structure that can be characterized by a group of transformations $G$. For example, if $\mathscr{X}$ is a $p$-dimensional affine space, then $G$ may be the group of $p$-dimensional affine transformations or if $\mathscr{X}$ is a Cartesian product of $n$ copies of a $p$-dimensional affine space, then $G$ may be the direct product of $n$ copies of the group of $p$-dimensional affine transformations.

Let $\mathscr{P}$ be a family of probability distributions on the sample space $\mathscr{X}$, indexed by a parameter space $\Omega$:

$$\mathscr{P} = \{\mathscr{P}_\theta : \theta \in \Omega\}.$$

We shall say that the family $\mathscr{P}$ is invariant under the action of a group $G$ operating to the left on $\mathscr{X}$ if there is a group of transformations of $\Omega$, called $\bar{G}$, and a homomorphism

Received June, 1978; revised June, 1980.

Research supported by the Natural Sciences and Engineering Research Council Canada Grant A8097.

*AMS* 1970 *subject classifications.* Primary 62A05, 62A15; secondary 62H10, 62H99.

*Key words and phrases.* Logical Bayesian inference; inner inference, conditional confidence; Bayesian multivariate analysis; logical probability; multivariate normal distribution.

$g \to \bar{g} : G \to \bar{G}$ such that, if $g \in G$ and a random variable $X \in \mathscr{X}$ has a $\mathscr{P}$-distribution with parameter $\theta$, then the transformed random variable $gX$ has a $\mathscr{P}$-distribution with parameter $\bar{g}\theta$.

A statistical $G$-model is a system $(\mathscr{X}, G, \mathscr{P}, \Omega)$ where $\mathscr{X}$ is a measurable space, called the sample space, $G$ is a group operating to the left on $\mathscr{X}$, and $\mathscr{P} = \{P_\theta : \theta \in \Omega\}$ is a family of probability distributions on $\mathscr{X}$ which is invariant under $G$ and is indexed by a measurable parameter space $\Omega$. It is usually assumed that there is a unique point in $\Omega$ that corresponds to any given probability in $\mathscr{P}$. If $\mathscr{X} = G$ and the action of $G$ on $\mathscr{X}$ is left multiplication on $G$, then the model can be briefly described as a system $(G, \mathscr{P}, \Omega)$, and may be called a statistical group.

An invariance principle says that, when the model has a certain group structure, statistical inferences should be equivariant under the action of the group. This requirement stems from the realization that, if $y = gx$ for some fixed $g \in G$, then the model for the transformed response $y$ is identical with that for $x$. According to Bayesian philosophy it is also possible to make statements concerning the unknown $\theta$ in the absence of data, and these statements can be summarized in a prior distribution. By the same invariance principle such statements, as well as the corresponding prior distribution, should be invariant under the action of the group $G$. But since the group $G$ acts on $\Omega$ only through the action of the group $\bar{G}$, the invariance principle requires that a logical prior representing ignorance concerning $\theta$ should be invariant under the action of $\bar{G}$, or, in other words, that it should be, in Kleinian sense, a geometric prior.

In most statistical applications $\Omega$ is a locally compact Hausdorff space which is a homogeneous space of a locally compact group with a countable base, and the isotropy group of a reference point $a \in \Omega$ is compact. Then, by Proposition 2 of the Appendix A, the above mentioned logical prior exists and is uniquely defined up to an arbitrary positive factor.

It will be convenient to call the above mentioned logical prior the *inner prior*, to distinguish it from the outer prior which will be described in the next section. Inner and outer priors usually assign to the parameter space an infinite measure. They require therefore a special axiomatization which is given in Villegas (1967).

**3. The outer prior.** Let $\phi : \bar{G} \times \Omega \to \Omega$ be the action of the group $\bar{G}$ on $\Omega$, defined by $\phi(\bar{g}, \theta) = \bar{g}\theta$. Choose an origin or reference point $a \in \Omega$. Let $\phi_a : \bar{G} \to \Omega$ be the transformation $\bar{g} \to \bar{g}a$. This transformation can be used to induce a prior on $\Omega$, if a measure on $\bar{G}$ is given.

We shall prove that, if $\bar{G}$ is locally compact, and acts transitively on $\Omega$, then this measure on $\bar{G}$ is uniquely determined, up to an arbitrary scale factor, by the condition that the induced prior should be invariant under a change of the origin or reference point $a \in \Omega$. The corresponding prior is then uniquely determined, and will be called the outer prior.

In effect, let $a_* \in \Omega$ be a new origin. Then there is an element $\bar{h} \in \bar{G}$ such that $a = \bar{h}a_*$. It follows that

$$\phi(\bar{g}, a_*) = \phi(\bar{g}\bar{h}^{-1}, a).$$

Therefore $\phi_{a_*}$ is the composition of the right translation $\bar{g} \to \bar{g}\bar{h}^{-1}$ followed by $\phi_a$. Let $\lambda$ be a measure on $\bar{G}$. The transformation $\phi_a$ induces a measure $\mu$ on $\Omega$ defined, for any measurable $B \subset \Omega$, by $\mu(B) = \lambda(A)$, where $A \subset \bar{G}$ is the preimage of $B$ under $\phi_a$. Let $A_*$ be the preimage of $B$ under $\phi_{a_*}$. Clearly we have $A_* = A\bar{h}$. Since the induced measure $\mu$ should be independent of the origin or reference point $a \in \Omega$, we must have $\lambda(A) = \lambda(A\bar{h})$, which means that $\lambda$ is a right invariant measure in $\bar{G}$. It is well known that, if $\bar{G}$ is locally compact, then a right invariant measure in $\bar{G}$, called the right Haar measure, exists and is uniquely defined up to an arbitrary positive factor.

In order to guarantee that the induced measure on $\Omega$ does not assign an infinite value to any compact set, and that it does not assign a zero value to any open set, we need some additional assumptions. It will be sufficient to assume that $\Omega$ is a locally compact homogeneous space and that the preimage under $\phi_a$ of any compact set is a compact set.

It follows that the outer prior is the prior induced by the right Haar measure on $\bar{G}$. This prior has been considered by Stein (1965, page 223). Since the transformation $\phi_a : \bar{G} \to \Omega$ is equivariant under left multiplication, the inner prior is the prior induced by the left Haar measure on $\bar{G}$. Therefore, the inner prior is, in general, different from the outer prior. This apparent duality in logical theories of statistical inference will be considered in more detail in a future paper.

**4. Transformation parameter models.** We specialize now the model by assuming that the parameter space $\Omega$ is the group $G$, and that the group $\bar{G}$ is the group of left translations in $G$. It follows that if $x$ is a random variable whose distribution corresponds to $\theta \in \Omega$, then the random variable

$$t = \theta^{-1}x \tag{4.1}$$

has a distribution corresponding to the identity element $e \in G$. The distribution of $t$ is therefore independent of $\theta$ and will be denoted by $P$. This model is called a transformation parameter model (Fraser 1966).

We assume that there is an invariant measure $\mu$ on $\mathscr{X}$, that is, a measure $\mu$ such that $\mu(gA) = \mu(A)$, for any measurable subset $A \subset \mathscr{X}$ and any $g \in G$. We further assume that the probability measure $P$ has a known probability density $f$ with respect to $\mu$. Then it is easily seen that, for any $\theta$, the density of $x$ with respect to $\mu$ is given by $f(\theta^{-1}x)$. The likelihood function is the function of $\theta$ which is obtained when the observed $x$ is substituted in $f(\theta^{-1}x)$.

According to the discussion in Section 2, the inner prior for $\theta$ should be a left invariant prior on $G$. We shall assume that $G$ is a locally compact group. It is well known that under this condition, a left invariant prior on $G$, called the left Haar measure, exists and is uniquely defined up to an arbitrary scale factor. We shall denote this inner prior by $\pi$, and the corresponding differential by $\pi(dx)$. It follows that the inner posterior differential is

$$f(\theta^{-1}x)\pi(d\theta) \tag{4.2}$$

It will be assumed that, for any $x \in \mathscr{X}$, the corresponding integral over $G$ is finite,

$$\int f(\theta^{-1}x)\pi\,(d\theta) < \infty, \tag{4.3}$$

so that the posterior measure can be normalized to obtain an inner probability distribution for $\theta$.

The outer prior introduced in the previous section is the right Haar prior, and it is well known that, according to a fiducial argument, it is the prior that represents ignorance with respect to the parameter $\theta$ (Fraser 1961, page 668; Fraser 1972, page 778; see also Peisakoff 1950, page 38).

It should be noted that, if the prior for $\theta$ is the left Haar prior, then the prior for $\theta^{-1}$ is the right Haar prior. It is in order to avoid the confusion between left and right Haar priors that may arise when such reparameterizations are made that we use the name inner prior to denote the Haar measure obtained by the invariance argument of Section 2.

**5. Conditional confidence.** We specialize our model still further by assuming that the sample space $\mathscr{X}$ is the group $G$ and that the action of $G$ on $\mathscr{X}$ is simply the left multiplication on $G$. This simplification is usually obtained after a sufficiency reduction. The model can now be breifly described as a pair $(G, \mathscr{P})$ and may be called a statistical principal group, by analogy with the concept of homogeneous principal $G$-set (Bourbaki, 1974 page 60). It should be noted that the invariant measure $\mu$ in $\mathscr{X}$ is now the left invariant Haar measure in $G$. We also assume that the function $f$ assumes its maximum value at a unique point $a \in G$. Therefore the maximum likelihood estimate of $\theta$ is simply $\hat{\theta} = xa^{-1}$.

A likelihood set for $\theta$ is a set $\mathscr{R}(x)$ defined by

$$\mathscr{R}(x) = \{\theta : f(\theta^{-1}x) > c\}$$

for some $c > 0$. Clearly,

$$\theta \in \mathscr{R}(x) \Leftrightarrow x^{-1}\theta \in \mathscr{R}(e). \tag{5.1}$$

$\mathscr{R}(x)$ is a confidence set with confidence coefficient $1 - \alpha$, which depends on the chosen $c$. According to the frequency interpretation of probability, the confidence coefficient $1 - \alpha$ is the long run frequency of coverage of the true $\theta$ in a sequence of future repetitions of the experiment. More precisely, the confidence coefficient $1 - \alpha$ is the probability of coverage of the true $\theta$ by a future hypothetical likelihood set. In symbols,

$$P_\theta(\theta \in \mathscr{R}(X)) = 1 - \alpha \tag{5.2}$$

where $X$ is the corresponding future hypothetical observation. As is well known, the probability on the left side will lose its meaning if a known $x$ is substituted for the unknown $X$.

Future repetitions used in frequency interpretations of statistical inferences are not real but hypothetical or simulated, and they should be considered only as a means of learning from the data. In Villegas (1977a) it is suggested that better inferences may be obtained if only future hypothetical samples similar to the data are considered, because in this way the noise may be reduced and we may get a better picture of what the actual sample has to say about the population.

If $X$ being similar to $x$ is defined to mean $\hat{\theta} \in \mathscr{R}(X)$, then the conditional confidence level of the actual likelihood set $\mathscr{R}(x)$ is defined to be the conditional probability, given $x$, that a future likelihood region $\mathscr{R}(X)$ will cover the true $\theta$ assuming that it covers $\hat{\theta}$. In symbols, the conditional confidence level of $\mathscr{R}(x)$ is defined to be

$$P_\theta\{\theta \in \mathscr{R}(X) \mid x, \hat{\theta} \in \mathscr{R}(X)\} \tag{5.3}$$

(Villegas, 1977a). This concept is based only on sampling theory and does not assume that any prior for $\theta$ is given. Equivalently, the conditional confidence level is the quotient of

$$P_\theta\{\theta \in \mathscr{R}(X), \hat{\theta} \in \mathscr{R}(X) \mid x\} \tag{5.4}$$

with

$$P_\theta\{\hat{\theta} \in \mathscr{R}(X) \mid x\}. \tag{5.5}$$

Obviously, from (5.1),

$$\theta \in \mathscr{R}(X) \Leftrightarrow T \in \mathscr{R}(e)^{-1},$$

where $e$ is the identity element, and $T = \theta^{-1}X$. We also have

$$\begin{aligned} \hat{\theta} \in \mathscr{R}(X) &\Leftrightarrow ax^{-1}X \in \mathscr{R}(e)^{-1} \\ &\Leftrightarrow T \in ta^{-1}\mathscr{R}(e)^{-1}. \end{aligned} \tag{5.6}$$

Therefore, the denominator (5.5) is equal to

$$P\{T \in ta^{-1}\mathscr{R}(e)^{-1} \mid t\} \tag{5.7}$$

and the numerator (5.4) is equal to

$$P\{T \in \mathscr{R}(e)^{-1}, T \in ta^{-1}\mathscr{R}(e)^{-1} \mid t\}. \tag{5.8}$$

Since the distribution of $T$ is independent of $\theta$, it follows that the conditional confidence level (5.3), which is the quotient of (5.8) divided by (5.7), is a function of the unknown $t = \theta^{-1}x$, and to emphasize this fact we shall denote it by $C(\theta^{-1}x)$.

**6. Frequency interpretation.** In Villegas (1977a) it was shown that, for the univariate normal model with unknown mean $\mu$ and unknown standard deviation $\sigma$, the inner posterior probability of a likelihood set for $\mu$ is equal to a mean conditional confidence level.

In the present section a similar result will be obtained for the model of Section 5. We shall prove that the inner posterior probability of the likelihood set $\mathscr{R}(x)$ is equal to a posterior mean of the conditional confidence level. More precisely, we shall prove that the inner posterior probability of $\mathscr{R}(x)$ is equal to the posterior mean of $C(\theta^{-1}x)$ for a person whose only information about $\theta$ comes from the knowledge that the event $\hat{\theta} \in \mathscr{R}(X)$ actually did occur (where $\hat{\theta}$ is regarded as given). This posterior mean, given $\hat{\theta} \in \mathscr{R}(X)$, will be simply called *the mean conditional confidence level* and will be denoted by $\mathscr{E}C$:

$$\mathscr{E}C = \mathscr{E}_x[C(\theta^{-1}x) \mid \hat{\theta} \in \mathscr{R}(X)].$$

We use a special notation, using $x$ as an indexing symbol, to indicate that $x$ is given, but this fact should not be allowed to have any influence on the posterior distribution of $\theta$. It will be shown later that $\mathscr{E}C$ does not depend on the actual values of $x$ and $\hat{\theta}$.

Assuming that, for the above mentioned person, the prior for $\theta$ is the left Haar prior with differential $\pi(d\theta)$, it follows that the prior for the new variable $t = \theta^{-1}x$, where $x$ is given, is the right Haar prior with differential $\nu(dt)$. Since in this experiment the only information concerning $\theta$ is the realization of the event $\hat{\theta} \in \mathscr{R}(X)$ for a given $x$, it follows that, by virtue of (5.6) the likelihood function for this experiment is given by (5.7). Therefore the corresponding posterior differential of $t$ is simply proportional to

$$\mathscr{P}\{T \in ta^{-1}\mathscr{R}(e)^{-1} \mid t\}\nu(dt). \tag{6.1}$$

The mean conditional confidence level $\mathscr{E}C$ is then equal to the weighted mean of $C(t)$ with the weighting differential (6.1). Equivalently, $\mathscr{E}C = C_1/C_2$, where

$$C_1 = \int P\{T \in \mathscr{R}(e)^{-1},\, T \in ta^{-1}\mathscr{R}(e)^{-1} \mid t\}\nu\ (dt) \tag{6.2}$$

and

$$C_2 = \int P\{T \in ta^{-1}\mathscr{R}(e)^{-1} \mid t\}\nu\ (dt). \tag{6.3}$$

Since the sampling differential of the random quantity $T$ is $f(T)\mu(dT)$, we have

$$C_1 = \int \nu\ (dt) \int \chi_{\mathscr{R}(e)^{-1}}(T)\chi_{ta^{-1}\mathscr{R}(e)^{-1}}(T)f(T)\mu\ (dT),$$

where $\chi_A(\cdot)$ denotes the indicator function for a set $A$. The substitution $T^{-1} = y$ gives

$$C_1 = \int \nu\ (dt) \int \chi_{\mathscr{R}(e)^{-1}}(y^{-1})f(y^{-1})\chi_{\mathscr{R}(e)a}(yt)\nu\ (dy).$$

This is the $L_1$-norm (with right Haar measures) of the convolution of the positive functions $\chi_{\mathscr{R}(e)^{-1}}(\cdot)f(\cdot)$ and $\chi_{\mathscr{R}(e)a}(\cdot)$. It is well known that the $L_1$-norm of the convolution of two positive functions is the product of the $L_1$-norms of the convoluted functions. This follows for example from Proposition 2 page 125 of Bourbaki (1963). Therefore,

$$C_1 = \nu(\mathscr{R}(e)) \int_{\mathscr{R}(e)^{-1}} f(t)\nu\ (dt).$$

Similarly,

$$C_2 = \nu(\mathscr{R}(e)) \int f(t)\nu\ (dt).$$

Therefore, the mean conditional confidence level is

$$\mathscr{E}C = c^{-1} \int_{\mathscr{R}(e)^{-1}} f(t)\nu\,(dt),$$

where $c = \int f(t)\nu\,(dt)$ is assumed to be finite. The substitution $t = \theta^{-1}x$ gives

$$\mathscr{E}C = c^{-1} \int_{\mathscr{R}(x)} f(\theta^{-1}x)\pi\,(d\theta).$$

Therefore, the mean conditional confidence level is equal to the inner posterior probability of the likelihood set $\mathscr{R}(x)$, and is independent of the actual value of $x$.

**7. The central multivariate normal model.** The sample space is the space $R^{p\times n}$ of all $p \times n$ matrices $Y = [y_1 \cdots y_n]$. We assume that the columns of the observed $Y$ are a random sample of size $n \geq p$ from a central normal distribution with unknown positive definite $p \times p$ covariance matrix $\Sigma$. Consider the uniquely defined triangular decomposition

(7.1) $$\Sigma = \Lambda\Lambda',$$

where $\Lambda$ is a $p \times p$ positive lower triangular matrix (that is, a lower triangular matrix with positive diagonal elements). We shall reparameterize the model, taking $\Lambda$ as the new parameter, which may be called the population triangular deviation. Notice that interest in $\Lambda$ as a parameter implicitly assumes an interest in a particular ordered basis (the canonical basis) on $p$-dimensional space. To go along with this, we shall choose as the group $G$ the group of $p \times p$ positive lower triangular matrices, and the action of $G$ on $R^{p\times n}$ will be matrix multiplication. The parameter space $\Omega$ is the same $G$ considered as a set. The associated group $\bar{G}$ is again $G$ and the action of $\bar{G}$ on $\Omega$ is matrix multiplication.

A sufficient statistic is the sample covariance matrix

$$\frac{1}{n} YY',$$

which, by neglecting a null event, will be assumed to be positive definite. The positive lower triangular $p \times p$ matrix $L$, uniquely defined by the triangular decomposition

(7.2) $$\frac{1}{n} YY' = LL'$$

will be called the sample triangular deviation, and is another sufficient statistic.

The sampling distribution of $Y$ is characterized by the fact that

(7.3) $$\Lambda^{-1}Y = U$$

has the standard normal distribution in $R^{p\times n}$. Let the $p \times p$ positive lower triangular matrix $V$ be defined by

(7.4) $$\sqrt{n}\,\Lambda^{-1}L = V.$$

Then $V$ is also uniquely defined by

$$UU' = VV'$$

and by the appendix B the sampling distribution of $V$ is the left chi distribution with $n$ degrees of freedom, whose sampling differential is proportional to

(7.5) $$f(V)\pi(dV),$$

where $f$ is given by

(7.6) $$f(V) = |V|^n \exp - \tfrac{1}{2}\|V\|^2.$$

Here $|V|$ denotes the absolute value of the determinant of $V$, $\|V\|$ denotes the Euclidean norm of $V$, defined by

$$\|V\|^2 = \sum \|v_i\|^2,$$

where $\|v_i\|$ denotes the norm of $v_i$, the $i$th column of $V$, and $\pi(dV)$ is the left Haar differential

$$\pi(dV) = \prod_{i=1}^{p} v_{ii}^{-i}\, dV, \tag{7.7}$$

where $v_{ii}$ is the $i$th entry in the principal diagonal of $V$.

Substitution in (7.5) of the variable $L$ defined by (7.4) gives

$$f(\sqrt{n}\, \Lambda^{-1}L)\pi(dL). \tag{7.8}$$

This is the sampling differential of $L$. The likelihood function is therefore proportional to $f(\sqrt{n}\, \Lambda^{-1}L)$, and it can be checked that $L$ is the maximum likelihood estimate of $\Lambda$, and is therefore a minimal sufficient statistic. Let $t(\cdot)$ be defined by $L = t(Y)$. Now for any $g \in G$, $t(gY) = gt(Y)$. Therefore, as has been proved in general by Fraser (1966), the minimal sufficient statistic $L = t(Y)$ is equivariant. The sufficiency reduction produces therefore a reduced model that has the simple form of Section 5. The inner prior is therefore the left Haar prior (7.7) and the inner posterior differential is proportional to

$$f(\sqrt{n}\, \Lambda^{-1}L)\pi(d\Lambda). \tag{7.9}$$

The inner posterior differential of the matrix $V$ defined by (7.4) is then $f(v)\pi^+(dv)$. Here $\pi^+$ is the right Haar measure defined by

$$\pi^+(dV) = \prod_{i=1}^{p} v_{ii}^{-p+i-1}\, dV. \tag{7.10}$$

This distribution may be called the right (triangular) chi distribution with $n - p + 1$ degrees of freedom.

Since this reduced model has the form considered in Section 5, the inner inferences have the conditional frequency interpretation described in Section 6.

An alternative approach was considered by Villegas (1971). In this alternative approach no sufficiency reduction is made and the group $G$ is the group of nonsingular $p \times p$ matrices. The parameter $\theta$ is a nonsingular $p \times p$ matrix and has a uniquely defined factorization

$$\theta = \Lambda O$$

where $\Lambda$ is positive lower triangular and $O$ is orthogonal. The factor $O$ is a nuisance parameter which is unidentifiable in a normal model, and the inner posterior distribution for $\Lambda$ turns out to be the same as in the reduced model with observation $L$ and parameter $\Lambda$.

Tiao and Zellner (1964) and Geisser (1965) developed a Bayesian theory for the multivariate normal model using

$$|\Sigma|^{\frac{p+1}{2}}\, d\Sigma^{-1} \tag{7.11}$$

as the prior for the inverse of the covariance matrix. This prior was derived by Jeffreys (1961) in the case $p = 2$ and it was considered, for arbitrary values of $p$ by Geisser and Cornfield (1963). In Villegas (1971) it is shown that (7.11) is the inner prior for $\Sigma^{-1}$. It follows that the inner posterior differential of $\Sigma^{-1}$ is proportional to

$$|\Sigma|^{-\frac{n-p-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr} YY'\Sigma^{-1}\right) d\Sigma^{-1} \tag{7.12}$$

if $\Sigma$ is positive definite and is zero otherwise. In other words, the inner posterior distribution of $\Sigma^{-1}$ is the Wishart distribution with scale matrix $(YY')^{-1}$ and $n$ degrees of freedom.

Bayesian inferences based on the prior (7.11) are therefore inner inferences and have the same conditional frequency interpretation, and this explains the so-called strong inconsistency in the bivariate normal example of Stone (1976).

It should be noted that the inner posterior distribution of $\Sigma^{-1}$ (or $\Lambda$) is equivariant under a change of orthogonal basis. This is due to the fact that, by Proposition A3.7 of the Appendix to Villegas (1972), the inner prior for $\Lambda$ is invariant under orthogonal transformations of $R^p$.

The affine multivariate normal model may be better modelled with two groups, the group of translations and the group of positive lower triangular transformations, given as integral parts of the model, and will be considered in a future paper.

## APPENDIX A

1. LEMMA. *Let $H$ be a compact subgroup of a locally compact group $G$, and let $\Delta_r$ be the right-hand modulus of $G$. Then*

$$\Delta_r(h) = 1$$

*for any $h \in H$.*

PROOF. The proof is identical to the second proof of Proposition 13, page 81 of Nachbin (1965).

2. PROPOSITION. *Let $E$ be a homogeneous space under the locally compact group $G$. If the isotropy (or stability) subgroup $H$ of a reference point $a \in E$ is compact, then there is one invariant positive measure on $E$ which is uniquely defined up to an arbitrary positive factor.*

PROOF. The conclusion follows immediately from the previous Lemma and Theorem 1, page 138 of Nachbin (1965).

## APPENDIX B

Assume the $p \times 1$ vectors $u_1, \cdots u_n$ are a random sample of size $n \geq p$ from a standard normal distribution. Let $U$ be the $p \times n$ matrix whose $i$th column is $u_i$. With probability 1 the matrix $U$ will be a full rank matrix and therefore a positive lower triangular matrix $V$, will be uniquely defined by the triangular decomposition

$$UU' = VV'. \tag{1}$$

Let $R^{p\times n}$ be the space of all $p \times n$ matrices and let $G$ be the group of $p \times p$ positive lower triangular matrices. Let $\phi : R^{p\times n} \to G$ be the map defined by (1). Assume that $G$ operates on both spaces $R^{p\times n}$ and $G$ by left matrix multiplication. It can be seen that $\phi$ is equivariant under $G$, that is,

$$\phi(gU) = g\phi(U),$$

for any $g \in G$ and $U \in R^{p\times n}$. The measure with differential

$$\mu(dU) = \frac{dU}{|\phi(U)|^n} \tag{2}$$

is invariant under $G$. Since $\phi$ is equivariant under $G$, the induced measure on $G$ is also invariant under $G$ and is therefore the uniquely defined left Haar measure in $G$, whose differential is proportional to (7.7).

The sampling differential of $U$ is proportional to

$$(\exp - \tfrac{1}{2} \| U \|^2)\, dU. \tag{3}$$

But

$$\| Y \|^2 = \text{tr } YY' = n \text{ tr } LL' = n \| L \|^2.$$

Therefore, the sampling differential of $Y$, by (2) and (3), is proportional to

$$f \circ \phi(U) \mu(dU), \tag{4}$$

where $f$ is the function defined on $G$ by

$$f(V) = | V |^n \exp -\tfrac{1}{2} \| V \|^2. \tag{5}$$

By a standard theorem on change of variables it follows that the sampling differential for $V$ is proportional to

$$f(V) \pi(dV). \tag{6}$$

This distribution will be called the left ($p \times p$ triangular) chi distribution, with $n$ degrees of freedom, to distinguish it from the right chi distribution with $n - p + 1$ degrees of freedom, whose differential is

$$f(V) \pi^+(dV), \tag{7}$$

where $\pi^+(dV)$, given by (7.10) is the right Haar measure in $G$. (Compare this derivation with the first derivation of the Wishart distribution in Giri (1977) page 113.)

**Acknowledgment.** The author wishes to thank J. V. Zidek for a reference to Stein's 1965 paper, and a referee for suggesting the use of convolutions in Section 6.

## REFERENCES

Bourbaki, N. (1963). *Integration* (Chapter 7, 8). Hermann, Paris.
Bourbaki, N. (1974). *Algebra I.* Addison-Wesley, Reading, Mass.
Dawid, A.P., Stone, M. and Zidek, J. (1973). Marginalization paradoxes in Bayesian and structural inferences (with discussion). *J. Roy. Statist. Soc. B* **35** 189–233.
de Finetti, B. (1974). *Theory of Probability. I.* Wiley, New York.
Fraser, D.A.S. (1961). On fiducial inference. *Ann. Math. Statist.* **32** 661–676.
Fraser, D.A.S. (1966). On sufficiency and conditional sufficiency. *Sankhyā Ser. A.* 145–150.
Fraser, D.A.S. (1972). Bayes, likelihood or structural. *Ann. Math. Statist.* **43** 898–916.
Geisser, S. (1965). Bayesian estimation in multivariate analysis. *Ann. Math. Statist.* **36** 150–159.
Geisser, S. and Cornfield, J. (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. B* **25** 368–376.
Giri, N.C. (1977). *Multivariate Statistical Inference.* Academic, New York.
Nachbin, L. (1965). *The Haar Integral.* Van Nostrand, Princeton.
Peisakoff, M.P. (1950). Transformation parameters. Ph.D. Thesis. Princeton Univ.
Stein, C. (1965). Approximation of improper prior measures by prior probability measures. In *Bernoulli Bayes Laplace* (J. Neyman and L.M. LeCam, eds.) Springer, Berlin.
Stone, M. (1976). Strong inconsistency from uniform priors (with discussion). *J. Amer. Statist. Assoc.* **71** 114–125.
Tiao, G. C. and Zellner, A. (1964). On the Bayesian estimation of multivariate regression. *J. Roy. Statist. Soc. B* **26** 277–285.
Villegas, C. (1967). On qualitative probability. *Amer. Math. Monthly* **74** 661–669.
Villegas, C. (1971). On Haar priors. In *Foundations of Statistical Inference* (V.P. Godambe and D.A. Sprott, eds.) Holt, Rinehart and Winston, Toronto.
Villegas, C. (1972). Bayesian inference in linear relations. *Ann. Math. Statist.* **43** 1767–1791.
Villegas, C. (1977a). Inner statistical inference. *J. Amer. Statist. Assoc.* **72** 453–458.
Villegas, C. (1977b). On the representation of ignorance. *J. Amer. Statist. Assoc.* **72** 651–654.

Department of Mathematics
Simon Fraser University
Burnaby, B.C.
Canada V5A 1S6