RIGHT HAAR MEASURE FOR CONVERGENCE IN PROBABILITY TO QUASI POSTERIOR DISTRIBUTIONS

By M. Stone

University College of Wales

1. Introduction and summary. Many statistical problems contain an infinite parameter space or are analyzed as if they contained one. In a Bayesian analysis of such a problem, there is often, for one or more of several reasons, an attraction in employing an infinite measure on parameter space in the role of 'prior distribution' of the parameter. The employment of such a *quasi* prior distribution consists in its formal substitution as the prior density in Bayes's theorem to produce a *quasi posterior distribution*. (We will qualify the "posterior distribution," obtained in this way, as *quasi*, even if it is a probability distribution with integral one, as will be assumed for the rest of this paper). The attractions referred to are that the quasi prior distribution

(i) may be thought to represent "ignorance" about the parameter;

(ii) may give (quasi) posterior distributions satisfying some "natural" invariance requirement;

(iii) may itself satisfy some "natural" invariance requirement (the Jeffreys invariants);

(iv) may give (quasi) posterior distributions on the basis of which statistical statements may be constructed which closely resemble those of classical statistics.

[A separate argument for the quasi prior distribution is that, for an infinite parameter space, the class of Bayes decision functions may be complete only if the class includes those derived from quasi prior distributions (e.g. Sacks (1963)); but in this paper we will go no further than consideration of posterior distributions.]

In the *foundations* of Bayesian statistics, associated with the names of Ramsey, de Finetti and Savage (but not Jeffreys), quasi prior distributions do not appear. When finally arrived at, subjective prior distributions are finite measures. Moreover they are, for any given person, uniquely determined, so that there should be no question of choice. Hence, matters such as the representation of ignorance, invariance and the degree of resemblance to classical statistics are not relevant. But it is possible to accept this standpoint and then to argue that the consequences of using quasi prior distributions are worth investigating if only as convenient approximations in some sense. As Welch [(1958), p. 778] reveals, such an attitude must have been implicitly adopted by those nineteenth century followers of Bayes and Laplace who ascribed a probability content to the interval between probable error limits of some astronomical or geodetic observation. (With a normal distribution of known variance taken for the observations, the implicit quasi prior distribution was, of course, uniform on the real line.)

Received 16 June 1964; revised 19 October 1964.

One sense of the "approximation" is straightforward. Imagine fixed data. A quasi prior distribution may be a satisfactory approximation to the actual prior distribution *for this set of data* if it locally resembles or simulates the actual distribution over some important compact set of parameter points *determined by the data*; the fact that the quasi prior distribution is not integrable on the complement of this compact set may not be important. For example, if, in sampling from a normal distribution, the quasi prior resembles the actual prior in the region where the likelihood function is not close to zero, then the approximation may be satisfactory. Thus, for given data, actual prior distribution and criterion of satisfactory approximation, the question of deciding whether a certain quasi prior distribution is employable is theoretically straightforward, the answer being based on a direct comparison of the actual posterior and quasi posterior distributions. Furthermore, even without knowledge of the data, a calculation of the prior probability (evaluated by the actual prior distribution) of obtaining data for which the quasi prior distribution is employable will provide the necessary *prospective* analysis. However, when no actual prior distribution is given, two courses of justification of a given quasi prior distribution are available.

One, suggested by a referee, would consist of the demonstration that, for each member of a wide class of proper (and possibly actual) prior distributions, the corresponding posterior distributions are (with high probability) satisfactorily close to the quasi posterior distribution.

The other, which formally avoids any decision as to when two distributions are satisfactorily close to each other, is asymptotic and would ask the question "Does there exist a sequence of proper prior distributions such that, as we proceed down the sequence, the posterior distributions converge in some sense to the quasi posterior distribution?"

Jeffreys [(1957), p. 68] and Wallace [(1959), p. 873] have adopted the latter course. Wallace shows without difficulty that, roughly speaking, given a quasi prior distribution, there exists a sequence of proper prior densities whose corresponding posterior densities tend to the quasi posterior density for each fixed set of data. Reintroducing the concept of satisfactory approximation, the existence of this type of convergence, which may be called Jeffreys-Wallace convergence, assures us that, for all reasonable criteria of approximation, *given the data* there will be a member of the constructed sequence of prior distributions whose corresponding posterior distribution will be satisfactorily approximated by the quasi posterior distribution. Since this particular prior distribution could be the actual prior distribution of some experimenter, a certain justification of the quasi posterior distribution is thereby provided. This justification can be made separately for each set of data thereby yielding an apparently prospective justification of the quasi prior distribution itself (that is, a justification of its use for all sets of data). However, it is clear that the justification is essentially *retrospective*, since the prior selected may depend on the data or, in other words, the convergence (of posterior distributions to quasi posterior distribution) may not be uniform with respect to different data.

In this paper, we will, like Jeffreys and Wallace, adopt the asymptotic justification course (fearing that the more statistically relevant alternative is exceedingly complex) but our justification will be genuinely prospective. To obtain this prospective asymptotic justification of a given quasi prior distribution, we may, perhaps, impose the condition that the convergence be uniform. However, even where it is possible, such a requirement is stronger than is statistically necessary. All we need is *convergence in probability* defined as follows. (The precise definition, applied to a special context, is given later.) *There is* convergence in probability *to the quasi posterior distributions corresponding to a given quasi prior distribution if there exists a sequence of proper prior distributions such that, at each parameter point, the corresponding sequence of posterior densities converges in probability to the quasi posterior density.* The latter use of "converges in probability" is the customary one with the Bayesian slant that the sequence of probability distributions with respect to which it is defined are the marginal distributions of data corresponding to the sequence of prior distributions. From this definition it is clear that, for all reasonable criteria of approximation, for each $\epsilon > 0$ there will be a member of the postulated sequence of prior distributions such that the prior probability of obtaining data for which the quasi posterior distribution is a satisfactory approximation to the posterior distribution corresponding to the member will exceed $1 - \epsilon$. Since this member could be the actual prior distribution of some experimenter, a genuinely prospective justification of the given quasi prior distribution is thereby provided.

Wallace's theorem shows that there is Jeffreys-Wallace convergence to *all* quasi prior distributions, so that all quasi prior distributions are asymptotically justified in the Jeffreys-Wallace sense. However, no equivalent theorem is available for convergence in probability, which requires demonstration for each quasi prior distribution considered.

The above distinction has been drawn previously by Stone (1963), (1964) for data from the normal distribution—univariate and multivariate. The present work provides a generalisation of this case.

Two analytical restrictions are made. The first, essential for the results obtained, is that the experiment generating the data should have the property of invariance under a group of transformations. (Following Fraser (1961), the supposed invariance will generally be conditional on an ancillary statistic). Many statistical problems have this group invariance structure.

The second restriction (which may well be inessential) is that, given any quasi prior distribution, only sequences of prior distributions obtained by *truncations* of the quasi distribution to compact parameter sets will be considered.

In Section 2, the *group invariant structure* of the experiment is outlined. In Section 3, we follow Hartigan (1964) by introducing *relatively invariant prior distributions*. In Section 4, *convergence in probability* in this context is defined and in Theorems 4.1 and 4.2 it is shown that right Haar measure (as quasi prior distribution) is, under certain conditions, sufficient and necessary (among relatively invariant prior distributions) for convergence in probability. In Section 5, general statistical applications are considered.

2. Group invariant structure of the experiment.

SUPPOSITION 1. *The data x has a probability distribution dependent on a parameter θ.*

We write $x = (a, s)$, where a is an ancillary statistic.

SUPPOSITION 2. *The spaces S of points s and Θ of points θ are isomorphic to a group G of transformations g.*

In S and Θ, s_0 and θ_0 (resp.) are the points corresponding to the identity element e of G. Any point g of G is a one-one transformation of S onto S and Θ onto Θ, the isomorphisms being established by (a) $s \leftrightarrow g \Leftrightarrow s = gs_0$ and (b) $\theta \leftrightarrow g \Leftrightarrow \theta = g\theta_0$. Following Fraser (1961), it will often be notationally convenient to index G by the corresponding points of either S or Θ. Thus in the expression "$\theta^{-1}s$", θ^{-1} stands for the point g for which $\theta = (\theta^{-1})^{-1}\theta_0$, while the θ, θ_0 just written have their ordinary meaning. Which meaning is to be attached to a symbol will be clear from the context.

SUPPOSITION 3. *G is a locally compact topological group* (Halmos, 1950, Section 0).

(To avoid triviality, G must not be compact). Supposition 3 implies (ibid. Ch. XI) the existence of two measures μ, ν on the Borel sets of G together with a real valued function $\Delta(g)$, with the properties (to be used extensively later):

$$\Delta(g) > 0, \qquad \Delta(g_1g_2) = \Delta(g_1)\Delta(g_2)$$

$$\mu(gA) = \mu(A), \qquad \nu(Ag) = \nu(A),$$

$$\mu(Ag) = \Delta(g)\mu(A), \qquad \nu(gA) = \Delta(g^{-1})\nu(A),$$

$$\mu(A) = \nu(A^{-1}), \qquad \int h(g)\, d\nu(g) = \int h(g)\Delta(g^{-1})\, d\mu(g).$$

(Integrals without indicated regions of integration will be over the whole space.)

$\mu(\nu)$ is the left (right) invariant Haar measure of G and $\Delta(\cdot)$ is the modular function.

The supposed isomorphism will be used to transfer μ to S and ν to Θ, that is, for $B \subset S$, $C \subset \Theta$, $\mu(B) = \mu(\{g \mid gx_0 \varepsilon B\})$, $\nu(C) = \nu(\{g \mid g\theta_0 \varepsilon C\})$.

SUPPOSITION 4. *Given $\theta \varepsilon \Theta$, $x = (a, s)$ has a probability distribution in which the conditional distribution of $u = \theta^{-1}s$ given a is independent of θ. In addition, the probability density function of a and u with respect to product measure of λ (for the space of a) and μ (for u) exists and may be written with probability element*

$$f(a, u)\, d\lambda(a)\, d\mu(u).$$

There is no proper subgroup G_s of G such that, for any a, $\int_A f(a, u)\, d\mu(u) = 0$ for all Borel sets A disjoint from G_s.

3. Relatively invariant quasi prior distributions. Distributions and quasi prior distributions of θ will be defined by their density with respect to right invariant ν measure.

DEFINITION 3.1. $q_0(\theta)$ is the density of a *quasi prior distribution* for θ if

$\int q_0(\theta)\, d\nu(\theta) = \infty$. (In requiring that a quasi prior density cannot be normed to unity, we differ slightly from Wallace (1959)).

DEFINITION 3.2. The quasi prior density $q_0(\theta)$ is *relatively invariant* if $q_0(\theta)$ is continuous and

$$q_0(\theta_1\theta_2) = q_0(\theta_1)q_0(\theta_2)$$

for $\theta_1 \varepsilon \Theta$, $\theta_2 \varepsilon \Theta$. ($\theta_1\theta_2$ means $\theta_1\theta_2\theta_0$ where $\theta_1 \varepsilon G$, $\theta_2 \varepsilon G$, $\theta_0 \varepsilon \Theta$). Definition 3.2 is consistent with existing terminology in that the measure corresponding to a relatively invariant quasi prior density is relatively invariant (Halmos, 1950, p. 265). Note that $q_0(\theta) \equiv 1$ (corresponding to ν) and $q_0(\theta) \equiv \Delta(\theta)$ (corresponding to μ) are special cases of relatively invariant quasi prior densities when, as will be assumed, G is not compact. Let $\mathcal{R}_Q$ denote the class of relatively invariant quasi priori densities.

DEFINITION 3.3. The *quasi posterior density* of θ given x, say $q(\theta \mid x)$, corresponding to $q_0(\theta)$, is that "posterior density" obtained by formal use of $q_0(\theta)$ as the "prior density" in Bayes's theorem.

We have the joint "probability" element $q_0(\theta)f(a, u)\, d\lambda(a)\, d\mu(u)\, d\nu(\theta)$. The conditional "probability" element for θ is

$$\begin{aligned} q(\theta \mid x)\, d\nu(\theta) & \\ &= q_0(\theta)f(a, u)\, d\lambda(a)\, d\mu(u)\, d\nu(\theta)/\textstyle\int_\theta q_0(\theta)f(a, u)\, d\lambda(a)\, d\mu(u)\, d\nu(\theta) \\ &= q_0(\theta)f(a, \theta^{-1}s)\, d\mu(\theta^{-1}s)\, d\nu(\theta)/\textstyle\int_\theta q_0(\theta)f(a, \theta^{-1}s)\, d\mu(\theta^{-1}s)\, d\nu(\theta) \\ &= q_0(\theta)f(a, \theta^{-1}s)\, d\nu(\theta)/\textstyle\int q_0(\theta)f(a, \theta^{-1}s)\, d\nu(\theta) \end{aligned} \tag{3.1}$$

since $d\mu(\theta^{-1}s) = d\mu(s)$.

From this point on, it will be convenient to work with the quantity

$$v = s^{-1}\theta. \tag{3.2}$$

THEOREM 3.1. *If $q_0(\theta)\ \varepsilon\ \mathcal{R}_Q$, the quasi prior distribution of v is independent of s.*

PROOF. By (3.1),

$$\begin{aligned} q(v \mid x)\, d\nu(v) &= q(\theta \mid x)\, d\nu(\theta) \\ &= q_0(sv)f(a, v^{-1}s_0)\, d\nu(sv)/\textstyle\int_v q_0(sv)f(a, v^{-1}s_0)\, d\nu(sv) \\ &= q_0(v)f(a, v^{-1}s_0)\, d\nu(v)/\textstyle\int q_0(v)f(a, v^{-1}s_0)\, d\nu(v)\,, \end{aligned} \tag{3.3}$$

which is independent of s. This independence allows us to write $q(v \mid x) \equiv q(v \mid a)$. Theorem 3.1 repeats, in this context, the conclusion of Hartigan (1964) that $\mathcal{R}_Q$ is a statistically "natural" class of quasi prior distributions to consider. Their employment leads to statistical procedures which are invariant under transformations which leave the problem invariant.

4. Convergence in probability. Given $q_0(\theta)\ \varepsilon\ \mathcal{R}_Q$ and a sequence of (compact) θ sets $\Theta_1, \Theta_2, \cdots$, define the corresponding *truncation sequence* of prior densities $p_1(\theta), p_2(\theta), \cdots$ by

$$p_i(\theta)\, d\nu(\theta) = q_0(\theta)\, d\nu(\theta)/\textstyle\int_{\Theta_i} q_0(\theta)\, d\nu(\theta), \qquad \theta \in \Theta_i$$
$$= 0, \qquad \text{otherwise.}$$

Then the corresponding sequence of posterior densities $p_1(\theta \mid x_1)$, $p_2(\theta \mid x_2)$, $\cdots$ is given by

$$p_i(\theta \mid x_i)\, d\nu(\theta) = \frac{q_0(\theta) f(a_i, \theta^{-1} s_i)\, d\nu(\theta)}{\int_{\Theta_i} q_0(\theta) f(a_i, \theta^{-1} s_i)\, d\nu(\theta)}, \qquad \theta \in \Theta_i$$
$$= 0, \qquad \text{otherwise.}$$

In terms of $v = s_i^{-1}\theta$,

$$p_i(v \mid x_i)\, d\nu(v) = \frac{q_0(v) f(a_i, v^{-1} s_0)\, d\nu(v)}{\int_{s_i^{-1}\Theta_i} q_0(v) f(a_i, v^{-1} s_0)\, d\nu(v)}, \qquad v \in s_i^{-1}\Theta_i \tag{4.1}$$
$$= 0, \qquad \text{otherwise,}$$

where we have used that fact that $q_0(\theta) \in \mathcal{R}_Q$.

(Subscripts have been attached to x, a, s to indicate that the data is not considered fixed. Each prior will be separately evaluated with respect to the data it generates.) Write $p_i(v \mid x_i) = p_i(v \mid a_i, s_i)$.

DEFINITION 4.1. The sequence $p_1(v \mid a_1, s_1)$, $p_2(v \mid a_2, s_2)$, $\cdots$ *converges in probability* to the quasi posterior density $q(v \mid a)$ if, for each fixed value a of the ancillary statistic, $\operatorname{plim}_{i\to\infty} p_i(v \mid a, s_i)$ exists and

$$\operatorname{plim}_{i\to\infty} p_i(v \mid a, s_i) \equiv q(v \mid a). \tag{4.2}$$

By (4.2), we mean that, for every $\epsilon > 0$ and each v,

$$\lim_{i\to\infty} \textstyle\int_{R(s_i, \epsilon)} p_i(s_i \mid a)\, d\mu(s_i) = 1,$$

where $R(s_i, \epsilon) = \{s_i \mid |p_i(v \mid a, s_i) - q(v \mid a)| < \epsilon\}$ and $p_i(s_i \mid a)$ is the conditional probability density of s_i in the joint marginal distribution of s_i, a_i obtained by integrating the probability element $p_i(\theta) f(a_i, \theta^{-1} s_i)\, d\lambda(a_i) \cdot d\mu(s_i)\, d\nu(\theta)$.

DEFINITION 4.2. If there is convergence in probability to a quasi posterior density, the quasi posterior density is a *probability limit.*

Keeping a_i fixed in Definition 4.1 is justified by the fact that, as an ancillary statistic, the distribution of a_i is independent of θ and therefore of $p_i(\theta)$ and therefore does not change with i. This is in contrast to s_i, whose distribution does depend on i. The restriction to fixed a is simply one of convenience. Since a will be held constant throughout the rest of the paper, explicit reference to it will be omitted. However, it should be remembered that all distributions involving s_i will be conditional on $a_i = a$. The effect of this on notation is that, for example,

$$f(a, v^{-1} s_0) \rightarrow f(v^{-1} s_0)$$
$$p_i(v \mid a, s_i) \rightarrow p_i(v \mid s_i)$$

$$q(v \mid a) \to q(v)$$

$$p_i(s_i \mid a) \to p_i(s_i).$$

We now investigate conditions for convergence in probability.

LEMMA 4.1. *For v such that $q(v) > 0$,* $\text{plim}_{i\to\infty} p_i(v \mid s_i) = q(v)$ iff

(i) $\lim_{i\to\infty} P(s_i \,\varepsilon\, \Theta_i v^{-1}) = 1$ and

(ii) $\lim_{i\to\infty} Er_i(s_i) = 1$

where

$$r_i(s_i) = \int_{s_i^{-1}\Theta_i} q_0(v) f(v^{-1}s_0)\, d\nu(v) \Big/ \int q_0(v) f(v^{-1}s_0)\, d\nu(v).$$

[The probability P and expectation E are evaluated with respect to $p_i(s_i)$.]

PROOF. From (3.3) [in which $q(v \mid x) = q(v)$] and (4.1),

$$\begin{aligned} q(v)/p_i(v \mid s_i) &= r_i(s_i), && s_i \,\varepsilon\, \Theta_i v^{-1} \\ &= \infty, && \text{otherwise.} \end{aligned}$$

So "$\text{plim}_{i\to\infty} p_i(v \mid s_i) = q(v)$" $\Leftrightarrow$ "$\text{plim}_{i\to\infty} q(v)/p_i(v \mid s_i) = 1$" $\Leftrightarrow$ "$\lim_{i\to\infty} P(s_i \,\varepsilon\, \Theta_i v^{-1}) = 1$ and $\text{plim}_{i\to\infty} r_i(s_i) = 1$" $\Leftrightarrow$ "(i) and (ii)," since $0 \leqq r_i(s_i) \leqq 1$.

THEOREM 4.1. *If $q_0(\theta) \,\varepsilon\, \mathcal{R}_Q$ and $q_0(\theta) \not\equiv 1$ then* $\text{plim}_{i\to\infty} p_i(v \mid s_i) \neq q(v)$ *for any v such that $q(v) > 0$.*

PROOF.

$$\begin{aligned} p_i(s_i) d\mu(s_i) &= \int_\theta p_i(\theta) f(\theta^{-1}s_i) d\mu(\theta^{-1}s_i)\, d\nu(\theta) \\ &= \left[\int_{\Theta_i} q_0(\theta) f(\theta^{-1}s_i) d\nu(\theta) \Big/ \int_{\Theta_i} q_0(\theta) d\nu(\theta)\right] d\mu(s_i) \\ &= \left[\int_{s_i^{-1}\Theta_i} q_0(v) f(v^{-1}s_0) d\nu(v) \Big/ \int_{s^{-1}\Theta_i} q_0(v) d\nu(v)\right] d\mu(s_i). \end{aligned}$$

Hence

$$Er_i(s_i) = \frac{\displaystyle\int \frac{\left[\int_{s^{-1}\Theta_i} q_0(v) f(v^{-1}s_0)\, d\nu(v)\right]^2}{\int_{s^{-1}\Theta_i} q_0(v)\, d\nu(v)}\, d\mu(s)}{\displaystyle\int q_0(v) f(v^{-1}s_0)\, d\nu(v)}. \tag{4.3}$$

With $\chi_A(\cdot)$ denoting the characteristic function of a set A, the numerator of (4.3) may be written

$$\begin{aligned} &\int \frac{\int q(v_1) f(v_1^{-1}s_0) \chi_{s^{-1}\Theta_i}(v_1)\, d\nu(v_1) \int q_0(v_2) f(v_2^{-1}s_0) \chi_{s^{-1}\Theta_i}(v_2)\, d\nu(v_2)}{\int_{s^{-1}\Theta_i} q_0(v)\, d\nu(v)}\, d\mu(s) \\ &\quad = \iint q_0(v_1) q_0(v_2) f(v_1^{-1}s_0) f(v_2^{-1}s_0) H(v_1, v_2)\, d\nu(v_1)\, d\nu(v_2), \end{aligned} \tag{4.4}$$

where

$$H(v_1, v_2) = \int \frac{\chi_{s^{-1}\Theta_i}(v_1) \chi_{s^{-1}\Theta_i}(v_2)}{\int_{s^{-1}\Theta_i} q_0(v)\, d\nu(v)}\, d\mu(s).$$

But

$$\int_{s^{-1}\Theta_i} q_0(v) d\nu(v) = \Delta(s) \int_{\Theta_i} q_0(\theta) d\nu(\theta)/q_0(s).$$

So

$$\begin{aligned} H(v_1, v_2) &= [\int q_0(s)\chi_{s^{-1}\Theta_i}(v_1)\chi_{s^{-1}\Theta_i}(v_2)d\nu(s)]/\int_{\Theta_i} q_0(\theta)d\nu(\theta) \\ &\leqq \min_{j=1,2}\{\int q_0(s)\chi_{s^{-1}\Theta_i}(v_j)d\nu(s)\}/\int_{\Theta_i} q_0(\theta)d\nu(\theta) \\ &= \min_{j=1,2}\{q_0(v_j)^{-1}\}. \end{aligned}$$

Substituting in (4.4) and (4.3) gives

$$(4.5)\quad Er_i(s_i) \leqq \frac{\int\int q_0(v_1)q_0(v_2)f(v_1^{-1}s_0)f(v_2^{-1}s_0)\min_{j=1,2}\{q_0(v_j)^{-1}\}\,d\nu(v_1)\,d\nu(v_2)}{\int q_0(v)f(v^{-1}s_0)\,d\nu(v)}.$$

But $\min_{j=1,2}\{q_0(v_j)^{-1}\} = \frac{1}{2}[q_0(v_1)^{-1} + q_0(v_2)^{-1}] - \frac{1}{2}|q_0(v_1)^{-1} - q_0(v_2)^{-1}|$. Substituting in (4.5), one obtains $Er_i(s_i) \leqq 1 - \frac{1}{2}D/\int q_0(v)f(v^{-1}s_0)d\nu(v)$, where

$$D = \int\int |q_0(v_2) - q_0(v_1)|\, f(v_1^{-1}s_0)f(v_2^{-1}s_0)d\nu(v_1)d\nu(v_2),$$

the algebraic reduction involving the identity $\int f(v^{-1}s_0)d\nu(v) = \int f(u)d\mu(u) = 1$. Using $q_0(\cdot) \varepsilon \mathcal{R}_Q$, D can be written

$$\int\int |q_0(u_2)^{-1} - q_0(u_1)^{-1}|\, f(u_1)f(u_2)d\mu(u_1)d\mu(u_2).$$

Hence, unless $q_0(u) =$ constant, a.e. P_f (where P_f is the probability measure whose density element is $f(u)d\mu(u)$), we have $Er_i(s_i) \leqq K < 1$ for some K and $i = 1, 2, \cdots$. Choose an open Borel set B on which P_f and μ (and therefore P_f and ν) are equivalent (mutually absolutely continuous). [For example, for $c > 0$ small enough, $\{u \mid f(u) > c\}$ is a non-empty set whose interior has the required properties.] Since the choice of s_0 is arbitrary, we may suppose that $e \varepsilon B$.

If $q_0(u) =$ constant (C), a.e. P_f, then $q_0(u) = C$, a.e. ν, for $u \varepsilon B$, which, by continuity of $q_0(u)$, implies that $q_0(u) = C$ for $u \varepsilon B$. But $e \varepsilon B$ and $q_0(e) = 1$ (by $q_0(\cdot) \varepsilon \mathcal{R}_Q$) so that $C = 1$ and $q_0(u) = 1$ for $u \varepsilon B$. Then, by $q_0(\cdot) \varepsilon \mathcal{R}_Q$, $q_0(u) = 1$ for $u \varepsilon B^{-1}$; hence $q_0(u) = 1$ for $u \varepsilon B \cup B^{-1}$, hence for $u \varepsilon (B \cup B^{-1})^r$ (for arbitrary positive integer r) and hence for $u \varepsilon E(B) = \bigcup_1^\infty (B \cup B^{-1})^r$. By Hewitt and Ross (1963), p. 34, $E(B)$ is a subgroup of G. Moreover, there exists an increasing sequence of such Borel sets B, $B_1 \subset B_2 \subset \cdots$, such that all such B's obey $B \subset \lim B_i$. Then $E(B_1) \subset E(B_2) \subset \cdots$ and it may be verified that $G_s = \lim E(B_i)$ is a subgroup of G. Suppose $G_s \neq G$. Then for any Borel set A disjoint from G_s, we must have $\int_A f(u)d\mu(u) = 0$ for, by construction, G_s contains all the P_f probability. This contradicts Supposition 4. Hence $G_s = G$. But $q_0(u) = 1$, $u \varepsilon E(B_i)$. Hence $q_0(u) = 1$, $u \varepsilon G_s = G$; or $q_0(u) \equiv 1$. That is, $q_0(u) =$ constant, a.e. P_f, implies $q_0(u) \equiv 1$.

So, unless $q_0(u) \equiv 1$, we have $Er_i(s_i) \leqq K < 1$ for some K and $i = 1, 2, \cdots$. So, unless $q_0(u) \equiv 1$, $\lim Er_i(s_i) \neq 1$ and hence, by Lemma 4.1, plim $p_i(v \mid s_i) \neq q(v)$ for any v such that $q(v) > 0$.

Theorem 4.1 implies that, given an experiment with the group invariant structure, a *necessary* condition that there is convergence in probability to the quasi posterior distribution derived from a particular relatively invariant quasi prior density, using a truncation sequence of prior densities (obtained by truncating the quasi prior density to a sequence of compact θ-sets), is that $q_0(\theta) \equiv 1$, that is, that the quasi prior measure be right invariant Haar measure.

We examine now whether this condition is also sufficient.

DEFINITION 4.3. G is *Haar controllable* if, for each compact (measurable) set C, there exists a sequence $G_1, G_2, \cdots$ such that $\lim_{i\to\infty} \nu(G_i[C])/\nu(G_i) = 1$ where $G_i[C] = \{g \mid gC \subset G_i\}$. (If a conjecture in Section 5(g) is correct, not all locally compact topological groups are Haar controllable.) (The definition is equivalent when stated with left Haar measure and right translations of C.)

THEOREM 4.2. *If G is Haar controllable then the quasi posterior distribution corresponding to right invariant Haar measure is a probability limit.*

PROOF. Choose a sequence $\epsilon_1, \epsilon_2, \cdots$ with $\epsilon_i > 0$ and $\epsilon_i \to 0$ and an increasing sequence of compact sets (in G) $V_1 \subset V_2 \subset \cdots$ with $e \varepsilon V_i$, $i = 1, 2, \cdots$, and $\lim_{i\to\infty} V_i = G$. For each ϵ_i, find a compact set C_i such that $\int_{C^{-1}_i} f(u)d\mu(u) > 1 - \epsilon_i$ and $\int_{C_i v^{-1}} f(u)d\mu(u) > 1 - \epsilon_i$ for $v \varepsilon V_i$. (Such a construction is possible, since we could take $C_i = A_i V_i$, where A_i is a compact set with the property that both $\int_{A_i} f(u)d\mu(u)$ and $\int_{A_i^{-1}} f(u)d\mu(u)$ exceed $1 - \epsilon_i$, and then note that $C_i^{-1} \supset A_i^{-1}$ and $C_i v^{-1} \supset A_i$ for all $v \varepsilon V_i$).

Since G is Haar controllable, we may, for each i, construct a sequence $G_1(i)$, $G_2(i)$, $\cdots$ such that $\lim_{j\to\infty} [\nu(G_j(i)[C_i])/\nu(G_j(i))] = 1$. Let $\Theta_i = G_{j(i)}(i)$ where $j(i)$ is chosen so that

$$\nu(G_{j(i)}(i)[C_i])/\nu(G_{j(i)}(i)) > 1 - \epsilon_i \tag{4.6}$$

and let $p_i(\theta)$ be the truncation of $q_0(\theta)$ to Θ_i. For $q_0(\theta) \equiv 1$,

$$\begin{aligned} P(s_i \varepsilon \Theta_i v^{-1}) &= \int_{\Theta_i} P(s_i \varepsilon \Theta_i v^{-1} \mid \theta) d\nu(\theta) / \int_{\Theta_i} d\nu(\theta) \\ &= \int_{\Theta_i} [\int_{\theta^{-1}\Theta_i v^{-1}} f(u)d\mu(u)] d\nu(\theta) / \int_{\Theta_i} d\nu(\theta). \end{aligned} \tag{4.7}$$

But "$\theta \varepsilon G_{j(i)}(i)[C_i]$" $\Rightarrow$ "$\theta C_i \subset G_{j(i)}(i) = \Theta_i$" $\Rightarrow$ "$\theta^{-1}\Theta_i v^{-1} \supset C_i v^{-1}$." But, since V_i increases to the whole space with i, there exists $i(v)$ such that $v \varepsilon V_i$ for $i > i(v)$; so that

$$\int_{C_i v^{-1}} f(u)d\mu(u) > 1 - \epsilon_i$$

and hence,

$$\int_{\theta^{-1}\Theta_i v^{-1}} f(u)d\mu(u) > 1 - \epsilon_i \tag{4.8}$$

for $i > i(v)$ and $\theta \varepsilon G_{j(i)}(i)[C_i]$. Substituting (4.8) in (4.7), we get, for $i > i(v)$,

$$\begin{aligned} P(s_i \varepsilon \Theta_i v^{-1}) &> (1 - \epsilon_i)[\int_{G_{j(i)}(i)[C_i]} d\nu(\theta) / \int_{G_{j(i)}(i)} d\nu(\theta)] \\ &> (1 - \epsilon_i)^2 \end{aligned}$$

using (4.6). Hence, for any fixed v,

$$\lim_{i\to\infty} P(s_i \varepsilon \Theta_i v^{-1}) = 1. \tag{4.9}$$

Also, when $q_0(\theta) \equiv 1$, we find from (4.3)

$$Er_i(s_i) = \int \frac{[\int_{\Theta_i^{-1}\theta} f(u)\, d\mu(u)]^2}{\int_{\Theta_i} d\nu(\theta)}\, d\nu(\theta).$$

But "$\theta \,\varepsilon\, G_{j(i)}(i)[C_i]$" $\Rightarrow$ "$C_i^{-1}\theta^{-1} \subset \Theta_i^{-1}$" $\Rightarrow$ "$\Theta_i^{-1}\theta \supset C_i^{-1}$." So, for $\theta \,\varepsilon\, G_{j(i)}\,(i)[C_i]$, $\int_{\Theta_i^{-1}\theta} f(u)d\mu(u) > 1 - \epsilon_i$ by construction of C_i. Hence,

$$\begin{aligned} Er_i(s_i) &> (1 - \epsilon_i)^2[\textstyle\int_{G_{j(i)}(i)[C_i]} d\nu(\theta)/\int_{\Theta_i} d\nu(\theta)] \\ &> (1 - \epsilon_i)^3. \end{aligned}$$

So

(4.10) $$\lim_{i\to\infty} Er_i(s_i) = 1.$$

But (4.9) and (4.10) are the two conditions of Lemma 4.1. So, for v such that $q(v) > 0$, $\operatorname{plim}_{i\to\infty} p_i(v \mid s_i) = q(v)$.
But $q(v) = 0 \Rightarrow p_i(v \mid s_i) = 0$. So $\operatorname{plim}_{i\to\infty} p_i(v \mid s_i) \equiv q(v)$ and the theorem is established.

5. Statistical applications.

(a) *Location parameters.* Suppose $x = (x_{11}, \cdots, x_{1n_1}; \cdots; x_{k1}, \cdots, x_{kn_k}) \,\varepsilon\, R^n$ where $n = n_1 + \cdots + n_k$, $\theta = (\theta_1, \cdots, \theta_k) \,\varepsilon\, R^k$, and the probability density function of x with respect to Lebesgue measure in R^n is $f(x_{11} - \theta_1, \cdots, x_{1n_1} - \theta_1; \cdots; x_{k1} - \theta_k, \cdots, x_{kn_k} - \theta_k)$. Here, $a = (x_{11} - \bar{x}_1, \cdots, x_{1n_1} - \bar{x}_1; \cdots; x_{k1} - \bar{x}_k, \cdots, x_{kn_k} - \bar{x}_k)$ and $s = (\bar{x}_1, \cdots, \bar{x}_k)$, where $\bar{x}_i = \sum_j x_{ij}/n_i$, are convenient representations of a and s. G is the group consisting of simple translations with representation $\theta \to \theta + g$, $s \to s + g$, $g \,\varepsilon\, R^k$. Moreover, G is clearly Haar controllable. Hence the only relatively invariant quasi prior distribution, which is a probability limit, is that corresponding to a "uniform prior distribution for θ."

(b) *Scale parameters.* Suppose x and θ as in (a) but $x_{ij} > 0$, $\theta_i > 0$, and the probability density function with respect Lebesgue measure in the positive "quadrant" is

$$(\textstyle\prod_i \theta_i^{-n_i}) f(x_{11}/\theta_1, \cdots, x_{1n_1}/\theta_1; \cdots; x_{k1}/\theta_k, \cdots, x_{kn_k}/\theta_k).$$

Hence, $a = (x_{11}/x_1^*, \cdots, x_{1n_1}/x_1^*; \cdots; x_{k1}/x_k^*, \cdots, x_{kn_k}/x_k^*)$ and $s = (x_1^*, \cdots, x_k^*)$, where $x_i^* = \prod_j x_{ij}^{1/n_i}$, are convenient representations of a and s. G is the group of scale changes with representation $\theta_i \to g_i\theta_i$, $x_i^* \to g_i x_i^*$ where $g_i > 0$, $i = 1, \cdots, k$. For G, $\mu = \nu$ has Radon-Nikodym derivative $\prod_{i=1}^k g_i^{-1}$ with respect to Lebesgue measure in the positive quadrant. That G is Haar controllable is seen by making the transformations $g_i \to e^{g_i}$, $i = 1, 2, \cdots$, noting that in the new representation G is Haar controllable and that Haar controllability is invariant under such changes of representation. Indeed, if we also make the transformations $\theta_i \to e^{\theta_i}$, $x_i^* \to e^{\bar{x}_i}$, we obtain a representation entirely coincident with (a).

(c) *One location and one scale parameter.* Suppose $x = (x_1, \cdots, x_n) \,\varepsilon\, R^n$, $\theta = (\eta, \sigma^2)$, $-\infty < \eta < \infty$, $\sigma > 0$, and the probability density function of x

with respect to Lebesgue measure is $\sigma^{-n}f((x_1-\eta)/\sigma, \cdots, (x_n-\eta)/\sigma)$. Here, $a = ((x_1-\bar{x})/S^{\frac{1}{2}}, \cdots, (x_n-\bar{x})/S^{\frac{1}{2}})$ and $s = (\bar{x}, S)$, where $S = \sum_1^n (x_i-\bar{x})^2$, are convenient representations of a and s. G is the group of transformations $g\theta = (\alpha+\beta\eta, \beta^2\sigma^2)$, $gs = (\alpha+\beta\bar{x}, \beta^2 S)$ $-\infty<\alpha<\infty$, $\beta>0$. As is well known (Halmos, 1950, p. 256), this is a case where $\mu \neq \nu$ but $d\mu(\theta) = d\eta d\sigma/\sigma^2$ and $d\nu(\theta) = d\eta d\sigma/\sigma$. To see that G is Haar controllable, we need a construction. Let $G_i = \{\alpha, \beta \mid -1<\alpha<1, i^{-1}<\beta<1\}$. Let C be any compact set of G. Then C can be contained in a rectangular set of the type $C' = \{\alpha, \beta \mid -A < \alpha < A, A^{-1} < \beta < A\}$. It is easily verified that

$$G_i[C'] = \{\alpha, \beta \mid Ai^{-1} < \beta < A^{-1}, -1+A\beta < \alpha < 1-A\beta\}$$

for $i > A^2$, whence

$$\nu(G_i[C']) = \int_{G_i[C']} d\alpha(d\beta/\beta) = \int_{Ai^{-1}}^{A^{-1}} (2-2A\beta)(d\beta/\beta) \sim 2\log i,$$

while $\nu(G_i) = 2\log i$. Hence $\lim_{i\to\infty} \nu(G_i[C'])/\nu(G_i) = 1$. But $G_i[C'] \subset G_i[C]$ so $\nu(G_i[C'])/\nu(G_i) \leqq \nu(G_i[C])/\nu(G_i) < 1$. Hence $\lim_{i\to\infty}[\nu(G_i[C])/\nu(G_i)] = 1$ and we have established Haar controllability. With greater effort, it can be shown that necessary and sufficient conditions for the more general sequence $G_1, G_2, \cdots$ given by $G_i = \{\alpha, \beta \mid \alpha_{1i} < \alpha < \alpha_{2i}, \beta_{1i} < \beta < \beta_{2i}\}$ to suffice for the demonstration of Haar controllability are

(a) $\rho_{2i}/\rho_{1i} \to \infty$,

(b) $\rho_{2i} \to \infty$,

(c) $\lim\inf_{i\to\infty} [\log \rho_{1\alpha}/\log \rho_{2\alpha}] \geqq 0$

where $\rho_{1i} = \beta_{2i}^{-\frac{1}{2}}(\alpha_{2i}-\alpha_{1i})$, $\rho_{2i} = \beta_{1i}^{-\frac{1}{2}}(\alpha_{2i}-\alpha_{1i})$. From the proof of Theorem 4.2, we know that the θ-sets, used as the basis of the truncation of the quasi prior (right Haar) density, are drawn from the sequences that give Haar controllability. This explains why the conditions (a), (b), (c) are formally identical with the necessary and sufficient conditions for convergence in probability, given for the normal case by Stone (1963); for we have $\alpha \leftrightarrow \eta$[or "$\mu$" in ibid.] and $\beta \leftrightarrow \sigma$[or $\omega^{-\frac{1}{2}}$ in ibid.].

(d) *Several location parameters and one scale parameter.* Suppose x as in (a) but $\theta = (\boldsymbol{\eta}, \sigma^2)$, $\boldsymbol{\eta} \in R^k$, $\sigma > 0$, and the probability density function of x with respect to Lebesgue measure is

$$\sigma^{-n}f((x_{11}-\eta_1)/\sigma, \cdots, (x_{1n_1}-\eta_1)/\sigma; \cdots; (x_{k1}-\eta_k)/\sigma, \cdots, (x_{kn_k}-\eta_k)/\sigma). \tag{5.1}$$

Here, a is $((x_{11}-\bar{x}_1)/S^{\frac{1}{2}}, \cdots, (x_{1n_1}-\bar{x}_1)/S^{\frac{1}{2}}; \cdots; (x_{k1}-\bar{x}_k)/S^{\frac{1}{2}}, \cdots, (x_{kn_k}-\bar{x}_k)/S^{\frac{1}{2}})$ and $s = (\bar{x}_1, \cdots, \bar{x}_k, S)$ where $S = \sum_i \sum_j (x_{ij}-\bar{x}_i)^2$. G is the group of transformations

$$g\theta = (\boldsymbol{\alpha}+\beta\boldsymbol{\eta}, \beta^2\sigma^2),$$

$$gx = (\boldsymbol{\alpha}+\beta\bar{\boldsymbol{x}}, \beta^2 S)$$

with $\boldsymbol{\alpha} \in R^k$ and $\beta > 0$. For this group, $\mu \neq \nu$ but $d\nu(g) = d\boldsymbol{\alpha} d\beta/\beta$. The same kind of argument, as was employed in (c), can be used to show that G is also Haar controllable. A special case of (5.1) is the normal model used in least squares theory and, for this case, $d\nu(g)$ was recommended by Jeffreys (1961, p. 150).

(e) *Several location and scale parameters.* Suppose x as in (a) but $\theta = (\eta_1, \cdots, \eta_k, \sigma_1^2, \cdots, \sigma_k^2)$, $\mathbf{n} \varepsilon R^k$, $\sigma_i > 0$, $i = 1, \cdots, k$, and the probability density function of x is

$$(5.2) \quad (\textstyle\prod_i \sigma_i^{-n_i}) f((x_{11} - \eta_1)\sigma_1, \cdots, (x_{1n_1} - \eta_1)/\sigma_1; \cdots; (x_{k1} - \eta_k)/\sigma_k, \cdots, (x_{kn_k} - \eta_k)/\sigma_k).$$

Here, $a = ((x_{11} - \bar{x}_1)/S_1^{\frac{1}{2}}, \cdots, (x_{1n_1} - \bar{x}_1)/S_1^{\frac{1}{2}}; \cdots; (x_{k1} - \bar{x}_k)/S_k^{\frac{1}{2}}, \cdots, (x_{kn_k} - \bar{x}_k)/S_k^{\frac{1}{2}})$ and $s = (\bar{x}_1, \cdots, \bar{x}_k, S_1, \cdots, S_k)$, where $S_i = \sum_j (x_{ij} - \bar{x}_i)^2$. G is the group of transformations $g\theta = (\alpha_1 + \beta_1\eta_1, \cdots, \alpha_k + \beta_k\eta_k, \beta_1^2\sigma_1^2, \cdots, \beta_k^2\sigma_k^2)$, $\boldsymbol{\alpha} \varepsilon R^k$, $\beta_i > 0$, $i = 1, \cdots, k$. Here,

$$d\nu(g) = d\boldsymbol{\alpha}(d\beta_1/\beta_1)(d\beta_2/\beta_2) \cdots (d\beta_k/\beta_k)$$

is the product measure of the k right Haar measures $d\alpha_i\, d\beta_i/\beta_i$. Since any $g \varepsilon G$ acts separately on each of the k component spaces and each of the separate components of G is Haar controllable, it follows that G is also Haar controllable. A special case of (5.2) is the Behrens-Fisher problem.

(f) *Scale parameters* (*generalization*). In (b), the restrictions that $f = 0$ outside the positive quadrant can be lifted if x_i^*, $i = 1, \cdots, k$, in the description of a, s and G is replaced by $(\sum_j x_{ij}^2)^{\frac{1}{2}}$, $i = 1, \cdots, k$. The isomorphism with (a) is lost but a specialization of (c) is thereby revealed. The (right and left) Haar measure element has Radon-Nikodym derivative $\prod_{i=1}^k g_i^{-1}$ with respect to Lebesgue measure in R^k.

(g) *Multivariate scale.* This is an instructive non-application. Suppose $x = \mathbf{S}$ and $\theta = \mathbf{\Sigma}$ where $\mathbf{S}$ and $\mathbf{\Sigma}$ are positive definite $k \times k$ matrices ($k > 2$). Suppose that, given $\mathbf{\Sigma}$, $\mathbf{S}$ has the probability density function in $R^{\frac{1}{2}k(k+1)}$

$$(5.3) \qquad |\mathbf{\Sigma}|^{-\frac{1}{2}(k+1)} f(\boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$ are the latent roots of $|\mathbf{S} - \lambda\mathbf{\Sigma}| = 0$. (For example, $\mathbf{S}$ might have a Wishart distribution with covariance matrix $\mathbf{\Sigma}$). Observing that (5.3) is invariant under the simultaneous transformations $\mathbf{S} \rightarrow \mathbf{ASA}'$, $\mathbf{\Sigma} \rightarrow \mathbf{A\Sigma A}'$ where $\mathbf{A}$ is any non-singular $k \times k$ matrix, it is clear that, to obtain our group invariant structure, we need a subgroup of the general linear group of nonsingular $k \times k$ matrices that is isomorphic to Θ, the space of all positive definite matrices. Such a subgroup is

$$G = \{\text{Non-singular, upper-triangular } k \times k \text{ matrices}\}$$

(although we could also take any one of the $k!$ other subgroups obtained by permuting the rows (say) of all the matrices in G). If $\mathbf{G}$ is a typical point in G then the isomorphism of G and Θ is established by $\mathbf{G} \leftrightarrow \mathbf{\Sigma} \Leftrightarrow \mathbf{GG}' = \mathbf{\Sigma}$. From Hewitt and Ross (1963, p. 209), we find that $\mu \neq \nu$ and

(5.4) $$d\nu(\theta) = dT_{11}\,dT_{12} \cdots dT_{kk}/|T_{11}T_{22}^2 \cdots T_{kk}^k|,$$

where $\mathbf{T}$ is the upper-triangular $k \times k$ matrix defined by $\mathbf{TT}' = \mathbf{\Sigma}$. From a statistical point of view, the lack of symmetry of (5.4) under permutation of $(1, \cdots, k)$ is objectionable and, in this respect, (5.4) differs from the quasi prior density element $d\mathbf{\Sigma}/|\mathbf{\Sigma}|^{\frac{1}{2}(k+1)} \propto d\mathbf{T}/|T_{11}T_{12} \cdots T_{kk}|^{\frac{1}{2}(k+1)}$, which has been proposed and used by Geisser and Cornfield (1963). It is perhaps fortunate then that the conjecture that G is Haar uncontrollable has strong support. In support of the conjecture, set $k = 2$ and, for simplicity, write $G_{11} = a$, $G_{12} = b$, $G_{22} = c$. For a particular C in Definition 4.3, take the compact set $\{\frac{1}{2} < a < \frac{3}{2}, -\frac{1}{2} < b < \frac{1}{2}, \frac{1}{2} < c < \frac{3}{2}\}$. Consider the left translate g_0C of C by the point $g_0 = (a_0, b_0, c_0)$ where $a_0 > 0$, $c_0 > 0$. Since

$$\begin{pmatrix} a_0 & b_0 \\ 0 & c_0 \end{pmatrix}\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a_0 a & a_0 b + b_0 c \\ 0 & c_0 c \end{pmatrix},$$

it is clear that g_0C is a section of the cylinder $\{\frac{1}{2}a_0 < a < 3a_0/2, \frac{1}{2}c_0 < c < 3c_0/2\}$. The ends of the section are on the planes $b = \pm\frac{1}{2}a_0 + b_0c/c_0$. Now in terms of a, b, c, the right Haar measure element is $da\,db\,dc/|a|\,|c|^2$. The fact that $a_0 > 0$, $c_0 > 0$ means that $a > 0$ and $c > 0$ for $(a, b, c) \in g_0C$ and, defining $L = \log a$, $I = 1/c$, we see that the right Haar measure element for points in g_0C is $dL\,db\,dI$, which is the element of volume in Euclidean space of L, b, I. But, in this space, g_0C is a section of the cylinder

(5.5) $$\{L_0 - \log 2 < L < L_0 + \log \tfrac{3}{2},\quad \tfrac{2}{3}I_0 < I < 2I_0\},$$

where $L_0 = \log a_0$, $I_0 = 1/c_0$. The ends of the sections are on the surfaces $b = \pm\frac{1}{2}\exp L_0 + b_0I_0/I$. (If either $a_0 < 0$ or $c_0 < 0$ or both, it is necessary to redefine L but a similar picture is obtained for the left translate of C.) Now to preserve the possibility of Haar controllability of G, we would have to find a sequence of compact sets $G_1, G_2, \cdots$ such that the vol. $(G_i[C])/\text{vol.}\,(G_i) \to 1$ as $i \to \infty$, where vol. ( ) means the Euclidean volume in the L, b, I space. In view of the inequalities on I in (5.5), the existence of such a sequence seems unlikely, and the conjecture that G is Haar uncontrollable is therefore supported.

(h) *Multivariate scale and location.* The remarks under (g) are equally relevant to the extension where $x = (\mathbf{y}, \mathbf{S})$, $\theta = (\boldsymbol{\eta}, \mathbf{\Sigma})$, $\boldsymbol{\eta}, \mathbf{y} \in R^k$, and the probability density function in $R^{\frac{1}{2}k(k+3)}$ is $|\mathbf{\Sigma}|^{-\frac{1}{2}(k+2)}f[(\mathbf{y} - \boldsymbol{\eta})'\mathbf{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\eta}), \lambda]$. (For example, $\mathbf{y}$ and $\mathbf{S}$ might be the sample mean vector and sample covariance matrix in a random sample from a multivariate normal distribution).

Acknowledgments. Conversations with Professor Lindley and Richard Morton have been valuable.

REFERENCES

Fraser, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.

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

HALMOS, P. (1950). *Measure Theory*. Van Nostrand, New York.

HARTIGAN, J. (1964). Invariant prior distributions. *Ann. Math. Statist.* **35** 836–845.

HEWITT, EDWIN and ROSS, KENNETH A. (1963). *Abstract Harmonic Analysis*. Springer-Verlag, Berlin.

JEFFREYS, SIR HAROLD. (1957). *Scientific Inference*. Cambridge Univ. Press.

JEFFREYS, SIR HAROLD. (1961). *Theory of Probability*. Oxford Univ. Press.

SACKS, J. (1964). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.* **34** 751–768.

STONE, M. (1963). The posterior *t* distribution. *Ann. Math. Statist.* **34** 568–573.

STONE, M. (1964). Comments on a posterior distribution of Geisser & Cornfield, *J. Roy. Statist. Soc. Ser. B.* **26** 274–276.

WALLACE, D. (1959). Conditional confidence level properties. *Ann. Math. Statist.* **30** 864–876.

WELCH, B. L. (1958). 'Student' and small sample theory. *J. Amer. Statist. Assoc.* **53** 777–788.