The Annals of Mathematical Statistics
1970, Vol. 41, No. 4, 1349–1353

NECESSARY AND SUFFICIENT CONDITION FOR CONVERGENCE IN PROBABILITY TO INVARIANT POSTERIOR DISTRIBUTIONS

BY M. STONE

University College London

0. Summary. For statistical problems based on groups, it is shown that there is convergence in probability to an invariant posterior distribution if and only if the posterior corresponds to right-invariant Haar prior and the group can support an asymptotically right-invariant sequence of proper prior distributions.

1. Introduction, definitions and result. Many statistical problems can be reduced to the following model: (i) a space, S, of data points s (ii) a space, Θ, of parameter points θ (iii) a locally compact group, G, of 1–1 transformations acting on both S and Θ, sending each space onto itself (iv) isomorphism of G, S and Θ, with the identity element, e, of G having corresponding points s_0 in S and θ_0 in Θ, while the points corresponding to g in G are gs_0 in S and $g\theta_0$ in Θ (v) given θ, s has a density $f(\theta^{-1}s)$ with respect to the left-invariant Haar measure, μ, transferred from G to S by the isomorphism.

The simplest example of this model is the location problem in which s is a univariate random variable, θ its location parameter, G is the group of addition of real numbers, μ is length on the real line and $f(\theta^{-1}s)$ becomes $f(s-\theta)$, an ordinary density function.

The above model, and the result below, can easily be extended to incorporate ancillary statistics. Many examples of the extended model are to be found in the literature; Fraser (1961), Stone (1965).

The Bayesian analysis of the model often uses a prior representing "ignorance". When G is not compact, the ignorance prior may be improper, that is, non-integrable. Bayes theorem is then used purely formally. We will be concerned only with such cases. For a prior m with element $dm(\theta)$ the posterior probability element is $f(\theta^{-1}s)\,dm(\theta)/\int f(\theta^{-1}s)\,dm(\theta)$ provided the denominator is neither zero nor infinite. Let $P(A \mid s)$ denote the corresponding posterior probability of $A \subset \Theta$. The posterior probability distribution is said to be *invariant* if $P(gA \mid gs) \equiv P(A \mid s)$ for all g. For an invariant posterior obtained by a formal use of Bayes theorem with an improper prior, we may attempt to justify its use as an approximation to the posterior that results from some proper prior. Let $\{f_n\}$ be a sequence of proper prior densities with respect to ν, the right-invariant Haar measure. Let $\{P_n(\cdot \mid \cdot)\}$ denote the corresponding posterior probability measures. Define

$$d_n(s) = \sup_A |P_n(A \mid s) - P(A \mid s)|$$

which measures the closeness of $P_n(\cdot \mid s)$ to the invariant posterior $P(\cdot \mid s)$. Letting $\tilde{s}_n$ denote the random variable $\tilde{s}$ generated *marginally* from $f_n(\theta)$ and $f(\theta^{-1}s)$, we

Received February 12, 1969; revised February 2, 1970.

say that $\{f_n\}$ *induces convergence in probability* to $P(\cdot \mid \cdot)$ if $\text{plim}_{n\to\infty} d_n(\tilde{s}_n) = 0$. (This definition is motivated in Stone (1965).)

We need just one more definition. The sequence $\{f_n\}$ of proper prior densities is said to be *asymptotically right-invariant* if

$$\int |f_n(\theta) - f_n(\theta g)|\, dv(\theta) \to 0$$

as $n \to \infty$, uniformly on every compact set in G. [Not all groups can support such a sequence; Dieudonné (1960), Stone and von Randow (1968), Reiter (1968).]

THEOREM. *If* (i) $\frac{1}{2}\int |f(u) - f(g^{-1}u)|\, d\mu(u)$ *is bounded away from unity on every compact set in* G (ii) *there is an open set* U *on which* f *is positive such that* $V = U^{-1}U$ *generates* G (*that is,* $\lim_{n\to\infty} V^n = G$) *then there exists an invariant posterior distribution to which there is induced convergence in probability by some sequence* $\{f_n\}$ *of proper priors if and only if* (a) *the invariant posterior corresponds to right-invariant Haar prior* (b) $\{f_n\}$ *is asymptotically right-invariant.*

The proof is given in the Appendix.

2. Discussion and applications. Conditions (i) and (ii) of the theorem are satisfied if, for example, f is everywhere positive.

When the trivial extensions to include ancillary statistics have been made, the theorem constitutes an appreciable generalization of the results of Stone (1965). In his theory of structural inference, Fraser (1961, 1968) has implicitly recommended the use of right-invariant Haar prior, since Fraser's structural probability distribution is just the corresponding invariant posterior. Confidence procedures may also be based on this distribution; Hora & Buehler (1966).

Two examples in which G cannot support an asymptotically right-invariant sequence may be easily stated:

(a) *Two-dimensional walk with a random final stage.* G is the set of all distinct walks of all finite lengths on a two-dimensional lattice, where each of the walks is defined by the sequence of consecutive steps, each of unit length either vertically or horizontally. The backward retracing of a portion of walk cancels the portion. The concatenation of walk g_1 followed by walk g_2 is denoted $g_1 g_2$. The observed walk s consists of an initial deterministic portion θ followed by a random portion u with discrete probability distribution $f(u)$. Fraser's structural probability distribution for θ is $f(\theta^{-1}s)$, definable without constraint. However, G is isomorphic to the discrete group with two free generators which is known not to support an asymptotically right-invariant sequence. Thus when f is everywhere positive we cannot justify $f(\theta^{-1}s)$ by our theorem.

(b) *Randomly perturbed* 2×2 *matrices.* G is the set of all non-singular 2×2 matrices. The observed matrix s is a product θu of a parameter matrix θ of unknown elements with an error matrix u which has probability density function $f(u)$ with respect to the left (and right) invariant element of measure $du_{11}\, du_{12}\, du_{21}\, du_{22}/|u|^2$ The structural probability distribution has element

$$f(\theta^{-1}s)\, d\theta_{11}\, d\theta_{12}\, d\theta_{21}\, d\theta_{22}/|\theta|^2.$$

When f is everywhere positive, the latter is not justifiable by our theorem.

We have made no analysis of what may be wrong with the invariant posterior or structural probability distribution in these two cases. We can only at this stage assert our belief, induced by the theorem, that they must be defective in some respect.

APPENDIX

LEMMA 1. *Under condition* (ii) *of the theorem, invariance of* $P(\cdot \mid \cdot)$ *based on prior* m *implies that* m *is a relatively invariant measure with*

$$dm(\theta)/dv(\theta) \equiv q(\theta) \tag{1}$$

$$q(\theta_1 \theta_2) \equiv q(\theta_1)q(\theta_2)$$

for some q.

PROOF. Invariance of $P(\cdot \mid \cdot)$ implies

$$\int_A f(\theta^{-1}s)\, dm(\theta)/h(s) \equiv \int_{gA} f(\theta^{-1}gs)\, dm(\theta)/h(gs)$$

$$\equiv \int_A f(\theta^{-1}s)\, dm(g\theta)/h(gs)$$

where $h(s) = \int f(\theta^{-1}s)\, dm(\theta)$, identically in A, g and s. So, for $\theta^{-1}s \in U$, "$dm(\theta)/h(s) = dm(g\theta)/h(gs)$" identically in g, that is, for $A \subset sU^{-1}$,

$$m(A)/h(s) \equiv m(gA)/h(gs). \tag{2}$$

(Note $h(s) > 0$ for all s for definition of P.)
There must be a g_0 such that $m(E) > 0$ for every open E with $g_0 \in E$; if not, for every g there would be open E_g with $g \in E_g$ and $m(E_g) = 0$, which would imply $m(G) = 0$. Then if E is *any* open set, we may choose g such that $g_0 \in gE$, open, whence, by (2), $m(E) > 0$.

Write $r_g(s) = h(gs)/h(s)$. Then (2) says that $r_g(s') = r_g(s'')$ if there exists open A such that $A \subset s'U^{-1}$ and $A \subset s''U^{-1}$. Since U^{-1} is open, such A will exist if there is a with $a \in s'U^{-1} \cap s''U^{-1}$, that is, if $s' \in s''U^{-1}U = s''V$. So $r_g(s)$ is constant on s_0V, hence on s_0V^2, $s_0V^3, \cdots$. But $V^n \to G$ as $n \to \infty$. So $r_g(s) = r_g$, say, independent of s and, by (2), $m(gA) \equiv r_g m(A)$. Hence m is relatively invariant and by Halmos (1950, page 265) the lemma follows.

LEMMA 2. *Under condition* (i) *of the theorem, convergence in probability to the invariant posterior from prior* (1) *induced by some sequence* $\{f_n\}$ *of proper priors implies* (a) $q \equiv 1$ (b) $\{f_n\}$ *is asymptotically right-invariant.*

PROOF. The invariant posterior element is $h(s)^{-1}q(\theta)f(\theta^{-1}s)\, dv(\theta)$ where $h(s) = \int q(\theta)f(\theta^{-1}s)\, dv(\theta)$. Now, since we have densities, we may write

$$d_n(s) = \tfrac{1}{2}\int \left| \frac{f_n(\theta)f(\theta^{-1}s)}{\int f_n(\theta)f(\theta^{-1}s)\, dv(\theta)} - h(s)^{-1}q(\theta)f(\theta^{-1}s) \right| dv(\theta).$$

Whence, with $h_n(s) = \int f_n(\theta) f(\theta^{-1}s)\, d\nu(\theta)$,

$$(3)\quad Ed_n(\tilde{s}_n) = \tfrac{1}{2}\iint f(\theta^{-1}s)\,|f_n(\theta) - h(s)^{-1}q(\theta)h_n(s)|\, d\nu(\theta)\, d\mu(s)$$

$$= \tfrac{1}{2}\iint f(g^{-1}\theta^{-1}s)\,|f_n(\theta g) - h(s)^{-1}q(\theta g)h_n(s)|\, d\nu(\theta)\, d\mu(s)$$

$$= \tfrac{1}{2}q(g)\iint f(g^{-1}\theta^{-1}s)\,|q(g)^{-1}f_n(\theta g) - h(s)^{-1}q(\theta)h_n(s)|\, d\nu(\theta)\, d\mu(s)$$

$$= \tfrac{1}{2}[1+q(g)^{-1}]^{-1}\iint \{f(\theta^{-1}s)\,|f_n(\theta) - h(s)^{-1}q(\theta)h_n(s)| + f(g^{-1}\theta^{-1}s)\,|q(g)^{-1}f_n(\theta g) - h(s)^{-1}q(\theta)h_n(s)|\}\, d\nu(\theta)\, d\mu(s).$$

But $\alpha\,|a-c| + \beta\,|b-c| \geqq \min\{\alpha, \beta\}|a-b|$. So

$$Ed_n(\tilde{s}_n) \geqq \tfrac{1}{2}[1+q(g)^{-1}]^{-1}\iint \min\{f(\theta^{-1}s), f(g^{-1}\theta^{-1}s)\}\,|f_n(\theta) - q(g)^{-1}f_n(\theta g)| \cdot d\nu(\theta)\, d\mu(s)$$

$$= \tfrac{1}{2}[1+q(g)^{-1}]^{-1}\int |f_n(\theta) - q(g)^{-1}f_n(\theta g)|\, d\nu(\theta) \int \{\tfrac{1}{2}[f(\theta^{-1}s) + f(g^{-1}\theta^{-1}s)] - \tfrac{1}{2}|f(\theta^{-1}s) - f(g^{-1}\theta^{-1}s)|\}\, d\mu(s)$$

$$(4)\qquad = \tfrac{1}{2}[1+q(g)^{-1}]^{-1}\int |f_n(\theta) - q(g)^{-1}f_n(\theta g)|\, d\nu(\theta)\{1 - \tfrac{1}{2}\int |f(u) - f(g^{-1}u)| \cdot d\mu(u)\}$$

$$\geqq \tfrac{1}{2}[1+q(g)^{-1}]^{-1}\,|1-q(g)^{-1}|\,\{1 - \tfrac{1}{2}\int |f(u) - f(g^{-1}u)|\, d\mu(u)\}.$$

Under condition (i), $1 - \frac{1}{2}\int |f(u) - f(g^{-1}u|\, d\mu(u) > 0$ for all g. So the induced convergence in probability implies $Ed_n(\tilde{s}_n) \to 0$ and hence $q(g) \equiv 1$. Then putting $q(g) \equiv 1$ in (4) we obtain, with condition (i), that $\int |f_n(\theta) - f_n(\theta g)|\, d\nu(\theta) \to 0$ uniformly on all compact sets in G, and the lemma is established.

LEMMA 3. *If $\{f_n\}$ is asymptotically right-invariant then $\{f_n\}$ induces convergence in probability to the invariant posterior corresponding to $q \equiv 1$.*

PROOF. For $q \equiv 1$, $h(s)^{-1} = \Delta(s) = d\mu(s)/d\nu(s)$, the modular element. Moreover, $1 = \int \Delta(s) f(u^{-1}s)\, d\nu(u)$. Substituting in (3)

$$Ed_n(\tilde{s}_n) = \tfrac{1}{2}\iint f(\theta^{-1}s)\,|f_n(\theta)\int \Delta(s) f(u^{-1}s)\, d\nu(u) - \Delta(s)\int f_n(u) f(u^{-1}s)\, d\nu(u)| \cdot d\nu(\theta)\, d\mu(s)$$

$$\leqq \tfrac{1}{2}\iiint |f_n(\theta) - f_n(u)| f(\theta^{-1}s) f(u^{-1}s)\Delta(s)\, d\nu(\theta)\, d\nu(u)\, d\mu(s).$$

Changing variables to θ, $g = \theta^{-1}u$, $t = \theta^{-1}s$, we have Jacobian $\Delta(\theta)$ and so

$$Ed_n(\tilde{s}_n) \leqq \tfrac{1}{2}\iint \{\int |f_n(\theta) - f_n(\theta g)|\, d\nu(\theta)\} f(t) f(g^{-1}t)\Delta(t)\, d\nu(g)\, d\mu(t)$$

$$\leqq \tfrac{1}{2}\sup_{g\in C}\int |f_n(\theta) - f_n(\theta g)|\, d\nu(\theta) + \int [1 - \int_C f(g^{-1}t)\Delta(t)\, d\nu(g)] f(t)\, d\mu(t)$$

for arbitrary C. Given $\varepsilon > 0$, choose compact A so that $\int_A f(t)\, d\mu(t) > 1-\varepsilon$ and $\int_{A^{-1}} f(t)\, d\mu(t) > 1-\varepsilon$. Then

$$Ed_n(\tilde{s}_n) \leqq \tfrac{1}{2}\sup_{g\in C}\int |f_n(\theta) - f_n(\theta g)|\, d\nu(\theta) + \int_A [1 - \int_C f(g^{-1}t)\Delta(t)\, d\nu(g)] f(t)\, d\mu(t) + \varepsilon.$$

Take $C = A^2$. Then, for $t \in A$,

$$\int_C f(g^{-1}t)\Delta(t)\, d\nu(g) = \int_{C^{-1}t} f(u)\, d\mu(u) \geqq \int_{A^{-1}} f(u)\, d\mu(u) > 1-\varepsilon.$$

So

$$Ed_n(\tilde{s}_n) \leqq \tfrac{1}{2} \sup_{g \in C} \int |f_n(\theta) - f_n(\theta g)|\, d\nu(\theta) + 2\varepsilon < 3\varepsilon$$

for n large enough since C is compact. The lemma is thereby established.

Lemmas 1, 2, 3 combine to establish the theorem.

REFERENCES

[1] DIEUDONNÉ, J. (1960). Sur le produit de composition II. *J. Math. Pures. Appl.* **39** 275–292.
[2] FRASER, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.
[3] FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.
[4] HALMOS, P. (1950). *Measure Theory*. Van Nostrand, Princeton.
[5] HORA, R. B. and BUEHLER, R. J. (1966). Fiducial theory and invariant estimation. *Ann. Math. Statist.* **37** 643–656.
[6] REITER, H. (1968). *Classical Harmonic Analysis and Locally Compact Groups*. Oxford Univ.
[7] STONE, M. (1965). Right Haar measure for convergence in probability to quasiposterior distributions. *Ann. Math. Statist.* **36** 440–453.
[8] STONE, M. and VON RANDOW, R. (1968). Statistically inspired conditions on the group structure of invariant experiments and their relationship with other conditions on locally compact topological groups. *Z. Wahrscheinlichkeitstheorie und Verw. Gebeite* **10** 70–78.