The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 326–339

STRUCTURAL DISTRIBUTIONS WITHOUT EXACT TRANSITIVITY[1]

BY JAMES V. BONDAR

University of Wisconsin and University of British Columbia

This is an extension of D.A.S. Fraser's *structural inference* to statistical problems invariant under a group which is not necessarily exactly transitive on either the sample or parameter spaces. Explicit formulas are given for the extended structural distribution of the parameter given observations, in a class of cases for which a relatively invariant measure exists. The connection with Bayesian inference with invariant priors is discussed.

0. Introduction. The concept of structural inference has been extensively treated in [3], notably for invariant statistical models in which the group is exactly transitive on the parameter space and exact on the sample space. Here we discuss an extension of structural inference to cases in which the group is not exact, and generalize some results of [1]. This material is of potential interest, not only to adherents of structural inference, but also to those who calculate Bayes posteriors from uniform priors, since the structural distribution is often equal to the Bayes posterior (see Section 3). In Section 1, we show the existence of extended structural distributions for a class of invariant models, and discuss some properties. Explicit formulae for these distributions are derived in Section 2 for two classes of special cases when relatively invariant measures exist. In Section 3 and Section 4 it is shown for these two classes of special cases, that the extended structural inference, a type of pistimetric inference, and Bayes inference with right invariant priors give the same distribution for the parameter given the observations (modulo the fact that they may not be defined on the same σ-field of the parameter space). Of course, these apparently similar distributions are given different interpretations by the adherents of these three approaches to inference.

1. The extended structural distribution. General information on invariant models is found in [5], Chapter 6, and a brief summary of the required measure theory is found in the appendix to the present paper. We assume that we are dealing with an invariant statistical model, i.e. with a sample space $\mathscr{X}$ and group G of transformations of $\mathscr{X}$, with generic elements x and g respectively; θ in the parameter space Ω indexes probability measures $P(\cdot;\theta)$ such that if

Received January 5, 1970; revised June 14, 1971.

[1] This research was started with the aid of the Office of Naval Research at the Department of Statistics of the University of Wisconsin, and finished at the Department of Mathematics of the University of British Columbia. A previous version was prepared with the aid of Grant No. A3011 of the National Research Council of Canada.

the random variable x has a distribution indexed by θ in Ω, then for any g in G, the random variable gx has a distribution indexed by θ' in Ω (we define $g\theta = \theta'$). We further assume:

1. $\theta_1 \neq \theta_2$ implies $P(\cdot\,; \theta_1) \neq P(\cdot\,; \theta_2)$.
2. $\mathscr{X}$ is a submanifold of some E^q embedded (in the sense of differentiable manifolds) in E^q and given the induced topology; the set $\mathscr{B}$ of measurable subsets of $\mathscr{X}$ is assumed to be the Borel field of $\mathscr{X}$.
3. All of the probabilities $P(\cdot\,; \theta)$ are dominated by the same σ-finite measure, called the *ground measure*.
4. $f(x; \theta)$ is a continuous function on $\mathscr{X} \times \Omega$ except for some smooth surface embedded in $\mathscr{X} \times \Omega$.
5. G is transitive on Ω.
6. G is a separable Lie group, and $(g, x) \to gx$ is continuous on $G \times \mathscr{X}$.
7. $\mathscr{X}$ is the Cartesian product of some orbit Gx with some other space $\mathscr{A}$: $\mathscr{A}$ is a surface in a Euclidean space.

For any x in $\mathscr{X}$, $Gx = \{gx : g \in G)$ is the *orbit* of x under G. If $Gx = \mathscr{X}$, we say that G is *transitive* on $\mathscr{X}$. If for every x_1, x_2 in $\mathscr{X}$ there is at most one g in G such that $gx_1 = x_2$, then we say that G is *exact* on $\mathscr{X}$. If G is exact and transitive it is also called *exactly transitive*. A model obeying assumptions 1 to 5 for which G is exact on Ω is called a *transformation parameter model* (TPM).

We choose an arbitrary point θ_0 in Ω; we associate each θ in Ω with the left coset $\sigma(\theta) = [\theta]H(\theta_0)$ in $G/H(\theta_0)$, where $[\theta]$ is any group element such that $[\theta]\theta_0 = \theta$, and $H(\theta) = \{g \mid g\theta = \theta\}$. We equip $G/H(\theta_0)$ with the finest topology such that σ is continuous. Since σ is both a homeomorphism and an isomorphism (see [7], Chapter 3), we can "by abuse of language" identify θ and the coset $\sigma(\theta)$. Similarly, we choose an arbitrary point $D(x)$ in each Gx such that $D(y) = D(x)$ if $y \in Gx$, and $x \to D(x)$ is continuous (this is possible by assumption 7); we associate any x with the left coset $\pi(x) = [x]H(Dx)$ in $G/H(Dx)$, where $[x]$ is any element of G such that $[x]Dx = x$, and $H(x) = \{g \mid gx = x\}$. We equip $G/H(Dx)$ with the finest topology such that π is continuous. π is both a homeomorphism and isomorphism of the orbit Gx onto $G/H(Dx)$, so whenever we are considering the conditional statistical model given Gx, we may identify x and the coset $\pi(x)$. We can then speak, for example, of $x^{-1} = H(Dx)[x]^{-1}$, $\theta_0 = H(\theta_0)$ and $Dx = H(Dx)$; and if G is exact on $\mathscr{X}$, then $H(x) = e$ (the identity of G).

The map $\rho : g \to gH(Dx)$, is a map from G onto $G/H(D)$. Note that different Dx may have different $H(Dx)$, hence ρ depends on the orbit. To emphasize this, we may write ρ_{Dx} instead of ρ. If S is a subset of G such that $S \cdot H(D) = S$, it will often be merely pedantic to distinguish between S and $\rho(S)$.

In addition to assumptions 1 to 7, we shall from time to time invoke the following assumption:

8. The function D can be (and has been) chosen such that $H(Dx) = H(Dy)$ for all $x, y \in \mathscr{X}$.

If $\mathscr{X}$ consists of a single orbit, then 7 and 8 are automatically satisfied. The cross-product required by assumption 7 can always be replaced by a local cross-product, thus generalizing our results somewhat.

Now, if x has a θ–distribution, then $[\theta]^{-1}x$ has a θ_0–distribution. We write $E = H(\theta_0)[\theta]^{-1}x$ and the mapping $x \to E$ then induces a probability from the conditional probability on Gx, onto the set $H(\theta_0)\backslash Gx = \{H(\theta_0)y \mid y \in Gx\}$. Thus E is a random variable defined on almost all Gx, taking values in $H(\theta_0)\backslash Gx$ and has a distribution (the *error distribution* given Gx) which does not depend on the parameter value θ.

If we use the above-mentioned identification which associates θ with $\sigma(\theta)$, x with $\pi(x)$, then on a specific orbit

$$E = H(\theta_0)[\theta]^{-1}x$$

becomes

$$E = \theta^{-1}x \,. \tag{1}$$

From (1),

$$x^{-1}\theta = E^{-1}$$

or

$$H(x)\theta = xE^{-1} \tag{2a}$$

or

$$H(Dx)[x]^{-1}\theta = H(Dx)E^{-1} \,. \tag{2b}$$

Given a value of x, equation (2a) induces a probability on the set of cosets of form $H(x)\theta$. This probability is, by (2b), the distribution of the inverse of the error variable multiplied on the left by x. To be precise, if S is measurable in Ω such that $H(x)S = S$ (or equivalently, if S is measurable in G such that $H(x)SH(\theta_0) = S$), we define

$$P(S \mid x) = P(S^{-1}x \mid Gx; \theta_0) \,. \tag{3}$$

$P(\cdot \mid x)$ is thus a measure on the σ–field of such S's in Ω, which we shall call the *extended structural distribution* of θ given x. Because all the spaces in sight ($\mathscr{A}$, $\mathscr{X}$ and Gx) are Euclidean and $x \to Gx$ is a measurable map from $\mathscr{X}$ onto $\mathscr{A}$ (assumption 7), the conditional probability in (3) exists (Lehmann, page 44) and is unique up to a null set.

There are two reasons for being interested in the probability in (3). Firstly,

$P(S \mid x)$ is the fiducial distribution obtained from the pivotal quantity $\theta^{-1}x$ given the ancillary statistic Gx. A second reason, more persuasive to many people is

THEOREM 1. *If assumptions* 1 *to* 7 *hold for an invariant statistical model, and* $P(S \mid x) = \alpha$, *then* S *is a confidence set of size* α *for* θ.

PROOF.

$$\begin{aligned} \alpha &= P(S \mid x) = P(S^{-1}x \mid Gx; \theta_0) \\ &= P(E \subset S^{-1}x \mid Gx) \\ &= P(Ex^{-1} \subset S^{-1} \mid Gx) \\ &= P(xE^{-1} \subset S \mid Gx) \\ &= P(\theta \in S \mid x) . \end{aligned}$$

The latter expression, as is usual with confidence sets, is interpreted as the probability that the *fixed* value θ is covered by the *random* set S. □

If in (3), S is the entire parameter space, we get $P(\Omega \mid x) = P(G^{-1}x \mid Gx; \theta_0) = 1$, hence:

THEOREM 2. *The structural probability is a probability.* (a.e.)

A *quantity* is any function of $\mathscr{X} \times \Omega$. An *invariant quantity* $(x, \theta) \to q(x, \theta)$ is a quantity such that $q(x, \theta) = q(gx, g\theta)$ for all x, g and θ. The following theorem will be useful in finding maximal (finest) invariant quantities:

THEOREM 3. *In an invariant model satisfying assumptions* 1 *to* 4, $(x, \theta) \to (Gx, G\theta, H(\theta_i)[\theta]^{-1}[x]H(Dx))$ *is a maximal invariant quantity, and if* G *is transitive on* Ω, *it is also a pivotal quantity.* ($i = G\theta$ *and* θ_i *is an arbitrary point in* i.)

PROOF. The quantity is invariant since its value at $(gx, g\theta)$ is $(Ggx, Gg\theta, H(\theta_i)[g\theta]^{-1}[gx]H(Dx))$. Since $[g\theta] = g[\theta]h$ where $h \in H(\theta_i)$, and similarly for $[gx]$, the quantity becomes $(Gx, G\theta, H(\theta_i)[\theta]^{-1}[x]H(Dx)) = q(x, \theta)$.

On the other hand, if $q(x, \theta)$ is an invariant quantity, we have $q(x, \theta) = q([x]H(Dx), [\theta]H(\theta_i)\theta_i) = q(H(\theta_i)[\theta]^{-1}[x]H(Dx), \theta_i)$, which is a function of $D(x)$, of i, and of $H(\theta_i)[\theta]^{-1}[x]H(Dx)$.

If G is transitive on Ω then $G\theta$ is constant; Gx has a constant distribution (see [1], Section 2) and equation (2b) tells us that $H(\theta_i)[\theta]^{-1}[x]H(Dx)$ equals $\theta^{-1}[x]H(Dx) = (H(Dx)E^{-1})^{-1} = E$, which has a distribution independent of θ. □

COROLLARY. *The extended structural probability is* homogeneous *in the sense that* $P(gS \mid gx) = P(S \mid x)$ *for almost all* x. *Any distribution for* θ *given* x (*fiducial, structural, Bayes or whatever*) *which is homogeneous in this sense, or any confidence interval for* θ *which is homogeneous, will be a function of the maximal invariant quantity.*

DEFINITION. If HS_α (α in some index set I) is a partition of the H-orbits of

G, a *complete set of invariants* for the class of sets HS_α, $\alpha \in I$, is a mapping which takes each HS_α into an element (member of G) in that HS_α. In other words, it is a system of distinct representatives for the class of sets HS_α, $\alpha \in I$.

Rule to find the maximal invariant quantity: A maximal invariant quantity is neither more nor less than a complete set of invariants for the sets $\theta^{-1}x = H(\theta_0)[\theta]^{-1}[x]H(Dx)$ in G (all $x \in \mathscr{X}$). Thus, the maximal invariant quantity is equivalent to a complete set of invariants for the $H(\theta_0)$–orbits $H(\theta_0)[\theta]^{-1}x$. For example, $[\theta]^{-1}x$ and $[\theta]^{-1}[x]$ are such invariants. Examples of this rule may be found in Section 2. A consequence of the rule is that the maximal invariant quantity is a function of cosets $H(\theta_0)[x]$, and hence the dimension of the range of the maximal invariant quantity is no greater than the dimension of the coset space $H(\theta_0)\backslash G$, which in turn equals the dimension of $\Omega = G/H(\theta_0)$, namely $\dim G - \dim H(\theta_0)$ ([6], Section 6.2.1). This may be interesting in view of a dictum sometimes proposed, that the pivotal quantity used for fiducial inference should have a dimension no greater than that of Ω.

2. Two special cases. In the remainder of this paper, we will wish to get explicit formulae for generalized structural distributions in important cases, and for this purpose it would be useful to express probability density functions with respect to an invariant measure on $\mathscr{X}$. However, if G is not exact on $\mathscr{X}$, then we do not in general have an invariant measure, but in many statistical models we have a *left relatively invariant measure* λ on $\mathscr{X}$, i.e. there exists $\delta : G \to R$ such that $d\lambda(gx) = \delta(g) \cdot d\lambda(x)$. E.g., for many of the models of multivariate analysis, G is a group of matrices, and Lebesgue measure on E^n is left relatively invariant with $\delta(g) = \det(g)$. (See Appendix or [7] Chapter 3 for the properties of relatively invariant measures.) δ is called the *modulus* of λ, and is not the group modulus Δ.

LEMMA 1. *If H is a compact subgroup of the locally compact group G, then there exists a left invariant measure on G/H. If λ is a left relatively invariant measure on G/H with modulus δ, any integral with respect to λ may be written*

$$\int_S f(x) \cdot d\lambda(x) = k \int_{\rho^{-1}(S)} f(gD)\delta(g) \cdot d\mu(g)$$

for some $k > 0$ and with $D = H$ in G/H; S is any measurable subset of G/H, and μ is left invariant measure on G.

PROOF. The existence of a left invariant measure is a corollary of Weil's theorem (for a statement of Weil's theorem, see Appendix, or [7], page 140, Corollary 2). Now,

$$\int_H f(ghx)\, d\mu_H(h) = f(gx) \cdot \mu_H(H)$$

($\mu_H(H) < \infty$ by compactness of H). Therefore we may set $f^*(g) = f(gD)/\mu_H(H)$ in Weil's theorem, and

$$\int_S f(x)d\lambda(x) = k_1 \int_{\rho^{-1}(S)} f^*(g)\delta(g)\,d\mu(g)$$
$$= k_1\mu_H(H)^{-1} \int_{\rho^{-1}(S)} f(gD)\delta(g)\,d\mu(g)\,. \quad \square$$

A. *A general case*. Making use of assumptions 7 and 8, we may construct a measure ξ on $\mathscr{X}$ which is the product of some convenient measure λ on some orbit Gx, with a measure l on the set of orbits. If there exists a left relatively invariant measure on G/H (see Weil's theorem in Appendix for a necessary and sufficient condition for existence), then we shall use it as our λ:

$$d\xi(x) = d\lambda(\pi(x))\,dl(Gx)\,.$$

If the $P(\cdot\,; \theta)$'s are continuous, then they will be dominated by ξ if l is properly chosen (setting l equal to the marginal distribution of Gx will suffice). If $dm(x)$ is the element of a measure on $\mathscr{X}$ which dominates each $P(\cdot\,; \theta)$ and is dominated by ξ, then

$$dP(x; \theta) = C(x)f(x; \theta)\,d\xi(x) \tag{4}$$
$$= C(x)f([\theta]^{-1}x; \theta_0)\delta^{-1}([\theta])\,d\lambda(\pi(x))\,dl(Gx)$$

where $f(\cdot\,; \theta) = dP(\cdot\,; \theta)/dm$, $C(\cdot) = dm/d\xi$, and δ is the left modulus of λ. Now,

$$C(gy) = dm(gy)/d\xi(gy)$$
$$= \frac{dm(gy)}{d\lambda(g\pi(y))\cdot dl(Dy)}$$
$$= \frac{J(g, y)\,dm(y)}{\delta(g)\,d\lambda(\pi(y))\cdot dl(Dy)} = J(g, y)\delta^{-1}(g)C(y)\,,$$

where $J(g, y)$ is the Jacobian of $y \to gy$ with respect to $dm(y)$, evaluated at y. If we set $y = D(x)$ and $g = [x]$, then $C(x) = J([x], Dx)\delta^{-1}([x])$. From (4),

$$dP(x \mid Gx; \theta) = K(Dx)J([x], Dx)\delta^{-1}([x])f([\theta]^{-1}x; \theta_0)\delta^{-1}([\theta])d\lambda(\pi(x))\,. \tag{5}$$

If S is a measurable set in $\Omega = G/H(\theta_0)$ such that $H(x)S = S$, the structural distribution is given by (3).

$$P(S \mid x) = K(D) \int J([y], Dx)\delta^{-1}([y])f(y; \theta_0)\,d\lambda(\pi(y)) \tag{6a}$$

where the integral is taken over y in Gx such that $\pi(y) \in S^{-1}[x]$. If $H(Dx)$ is compact, then by Lemma 1 this becomes (we absorb constants into k):

$$\begin{aligned} &k(D) \int_{\rho^{-1}(S^{-1}[x])} J(g, Dx)\delta^{-1}(g)f(gDx; \theta_0)\delta(g)\,d\mu(g) \\ &\qquad = k(D) \int_{\rho^{-1}(S^{-1})} J(g, Dx)f(gx; \theta_0)\,d\mu(g) \\ &\qquad = k(D) \int_{\rho^{-1}(S)} J(g^{-1}, Dx)f(g^{-1}x; \theta_0)\,d\nu(g) \end{aligned} \tag{6b}$$

where ν is the right invariant measure $\nu(E) = \mu(E^{-1})$ on G. Hence we have proved:

THEOREM 4. *If assumptions* 1 *to* 8 *hold, and there exists a left relatively invariant measure λ on $G/H(Dx)$; if λ dominates the conditional distributions of $\pi(x)$ given Gx, then the conditional distribution of x is given by* (5); *the structural distribution is given by* (6a). *If $H(x)$ is compact, then the structural distribution is given by* (6b).

REMARKS. If G is compact, then $H(x)$ is automatically compact, for $H(x)$ is the inverse image of the closed set $\{x\}$ under the (continuous) map $g \to gx$.

If $dm(x)$ is chosen to be $d\xi(x)$, then $J(g, Dx) = d\xi(gD)/d\xi(D) = \delta(g)$.

If G is a semi-direct product of $H(\theta_0)$ and some other subgroup, and if G is exact on $\mathscr{X}$, then one may use the ingenious "marginal analysis" of D.A.S. Fraser to derive expression (6b), ([3], Chapter 5).

B. *Special Case* (transformation parameter models). If assumptions 1 to 8 hold, and G is exactly transitive on Ω, then equations (1) and (2a) for the conditional model given the orbit become:

$$x = \theta E\,, \tag{7}$$

$$H(x)\theta = xE^{-1} \tag{8}$$

where each side of (7) is a left coset $gH(Dx)$ and the two sides of (8) are right cosets $H(x)g$. By the rule of Section 1, $\theta^{-1}x$ is maximal invariant. If $H(x)$ is itself compact, then Theorem 4 holds. $\Omega = G$, so (6) becomes

$$P(S \mid x) = k(D) \int_S J(\theta^{-1}, Dx) f(\theta^{-1}x; \theta_0)\, d\nu(\theta)\,, \tag{9}$$

defined for all measurable $S = H(x)S$ in G.

This is a generalization of Fraser (1968) page 64, where several formulas for the density g^* of θ given x are stated. Except for the last one, which contains an error, these formulas agree with (9).

EXAMPLE 1. $\mathscr{X}$ is the real line R; G is the group generated by translations and reflections. Let $(t, +1)$ translate the points of $\mathscr{X}$ to the right by t; let $(t, -1)$ reflect $\mathscr{X}$ in the origin and then translate by t. If zero is chosen as D, then we can use $(x, 1)$ as $[x]$ ($(x, -1)$ would also suffice as a choice of $[x]$). $(t_1, a_1)\cdot(t_2, a_2) = (a_1t_2 + t_1, a_1a_2)$; $(t, a)^{-1} = (-at, a)$. $H(D) = \{(0, 1), (0, -1)\}$ is compact and $d\mu(t, a) = dt$ is left invariant on G; dx is left invariant on $\mathscr{X}$, so we shall use it as our $d\lambda(x)$. $\delta \equiv \Delta \equiv J \equiv 1$.

Let $dP(x; \theta_0) = f(x; \theta_0)\, dx$, where no translate of f is an even function. By (5),

$$dP(x; (t, 1)) = f(x - t)\, dx$$
$$dP(x; (t, -1)) = f(-x + t)\, dx\,.$$

The pivotal quantity is $\theta^{-1}x = a_\theta(x - t_\theta)$ where $\theta = (t_\theta, a_\theta)$.

Now, $H(x)\cdot(t, a) = \{(t, a), (2x - t, -a)\}$. Hence, the structural probability $P(S \mid x)$ is defined iff S is of the form $S = \{(t_\theta, 1) : t_\theta \in T\} \cup \{(t_\theta, -1) : t_\theta \in 2x - t\}$

where T is measurable. By (9)

$$\begin{aligned} P(S \mid x) &= k \int_{(t,a) \in S} f(\theta^{-1}x; \theta_0)\, dt \\ &= k \int_T f(x-t)\, dt + k \int_{2x-T} f(-x+t)\, dt \\ &= 2k \int_T f(x-t)\, dt \end{aligned}$$

where k must be $\frac{1}{2}$ to make $P(\Omega \mid x) = 1$.

General Remark. It is of interest to note that this extended structural distribution could not have a frequency interpretation in terms of confidence sets if we tried to define it on all Borel sets of Ω. To see this, suppose that a probability for θ given x was defined on every measurable subset S of Ω, such that S is a confidence set for θ of size $P(S \mid x)$. If we consider an experiment in which a large number of samples x_i, $i = 1, 2, \cdots$ are taken, and an extended structural distribution for θ given x_i is obtained for each i, and if Nature chooses $\theta = (0, +1)$ every time (unbeknown to the statistician), then the condition of frequency interpretability implies that the marginal probability for $a_\theta = 1$ is equal to unity for almost all x. Similarly, if we consider the same experiment with $\theta = (0, -1)$ every time, we see that the marginal structural probability of $a_\theta = -1$ must be unity for almost all x. These two requirements are contradictory.

EXAMPLE 2. (Sample of size one from normal (μ, σ^2)). $\mathscr{X}$ is the real line R^1; G is the location-scale group $\{(t, a); a > 0\}$ for which $(t, a)x = ax + t$ and $\Delta(t, a) = 1/a$([3], page 63). Multiplication and inverse are as in Example 1. If we choose θ_0 to be the $N(0, 1)$ distribution and Dx to be 0, then $H(D) = \{(0, a) \mid a > 0\}$ and $H(\theta_0) = \{e\}$. There is no invariant measure on $\mathscr{X}$, but the Lebesgue measure element dx is relatively invariant with modulus $\delta(t, a) = a$.

By (5),

$$\begin{aligned} dP(x \mid (\mu, \sigma)) &= (2\pi)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\theta^{-1}x)^2\right] \cdot \delta(\mu, \sigma) \cdot dx \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \cdot dx\,. \end{aligned}$$

The sets on which structural probability is defined are generated by those of form

$$S = \left\{(\mu, \sigma) \,\middle|\, m_1 < \frac{x-\mu}{\sigma} < m_2\right\}$$

since

$$H(x)(\mu, \sigma) = \{(x - ax + a\mu, a\sigma) \mid a > 0\}\,.$$

By (3),

$$P(S \mid x) = P(S^{-1}x; \theta_0) = P(m_1 < z < m_2)$$

where z is $N(0, 1)$. Since $H(x)$ is not compact, (9) cannot be used.

It is of interest to note that parameter sets of the form $\{\theta \mid 0 < \sigma < k\}$ have no structural probability, although there are confidence intervals for σ of this form ([8]).

REMARK. In both Examples 1 and 2, our sample size is so low that the dimension of the sample space is less than or equal to the dimension of the parameter space. If we take a larger sample, our sample space will have dimension greater than that of Ω, and then G becomes exact, but no longer transitive on $\mathscr{X}$. This often occurs in practice: a transformation parameter model with G not exact on $\mathscr{X}$ is typically found in cases in which the sample size is "too small" or there are "too many parameters," a fact which is made more plausible by the following (somewhat heuristic) argument due to Andrew Kalotay. If a sample $\mathbf{x} = (x_1, \cdots, x_n)$ is taken from the model, and G acts on $\mathscr{X}$ so that $g(x_1, \cdots, x_n) = (gx_1, \cdots, gx_n)$, then the stability subgroup of $(x_1, \cdots, x_n)$ is $\bigcap_{i=1}^n H(x_i) = \bigcap_{i=1}^n [x_i]H(D)[x_i]^{-1}$. For almost all $\mathbf{x}$, increasing n by 1 will decrease the dimension of $\bigcap H(x_i)$ unless $\bigcap H(x_i)$ is already equal to some normal subgroup of G which is contained in $H(D)$. If this normal subgroup is called N, we may then consider the model to be invariant over the group G/N, and this reduced group G/N will be exact on almost all of $\mathscr{X}$ for n large enough that $\bigcap H(x_i) = N$.

EXAMPLE 3. (Multilinear model of D.A.S. Fraser and L. Steinberg [4]; a special case of this model is discussed in [3] pages 225–242).

$$G = \left\{ g = \begin{bmatrix} I_{(r)} & 0_{(rxp)} \\ B_{(pxr)} & C_{(pxp)} \end{bmatrix} : \det C > 0 \right\}$$

is a subgroup of the $(r + p) \times (r + p)$ matrices;

$$X = \begin{pmatrix} \mathbf{v}_1, \mathbf{v}_2 \cdots \mathbf{v}_n \\ \mathbf{x}_1, \mathbf{x}_2 \cdots \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & & \vdots \\ v_{r1} & \cdots & v_{rn} \\ x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pn} \end{pmatrix}$$

is a generic element of $\mathscr{X}$ (the v's are fixed, the x_{ij}'s range through the reals). When X has the θ_0 distribution, the $\mathbf{x}_i$ are i.i.d. $N(\mathbf{0}, I_{(pxp)})$. If $n > p$, then G is exact on almost all orbits of $\mathscr{X}$, however G is not exact on Ω, since

$$gX = \begin{pmatrix} \mathbf{v}_1, \cdots, \mathbf{v}_n \\ \mathbf{x}_1^*, \cdots, \mathbf{x}_n^* \end{pmatrix}$$

where the $\mathbf{x}_i^* = B\mathbf{v}_i + C\mathbf{x}_i$ are independent column vectors with the multivariate distribution $N(B\mathbf{v}_i, CC')$. If $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ is of rank r and $n > p > r$, then all $B\mathbf{v}_i = 0$ implies $B = 0$, hence

$$H(\theta_0) = \{g \mid C \in SO(p),\ B = 0\}$$

where $SO(p)$ is the set of $p \times p$ orthogonal matrices with positive determinant.

As is shown in [4], the volume element $dX = dx_{11} \cdots dx_{1n} \cdot dx_{21} \cdots dx_{pn}$ is relatively invariant with modulus $|g|^n$; left and right invariant measure elements on G are $d\mu(g) = dg/|g|^{p+r}$ and $d\nu(g) = dg/|g|^p$ respectively, where $dg = dB\, dC$ and $|g| = |C(g)|$. We now write $[\theta]$ as

$$[\theta] = \begin{bmatrix} I & 0 \\ B & \Gamma \end{bmatrix}.$$

If $f_n(X; \theta_0)\, dX = \prod_{i=1}^n f(X_i)\, dX$ is the probability element for X when θ_0 is the parameter, then by (5):

$$\begin{aligned} dP(X; \theta) &= k'(D) J([X], D) f_n([\theta]^{-1}X; \theta_0) \cdot d\mu(X) \\ &= k'(D)|C(X)|^n \prod f(\Gamma^{-1}(\mathbf{x}_i - \mathscr{B}\mathbf{v}_i)) \cdot |\Gamma|^{-p-r}\, d\mathscr{B}\, d\Gamma\,, \end{aligned}$$

and the structural probability of a set S in $G/H(\theta_0)$ is

$$\begin{aligned} P(S \mid X) &= k(D) \int_{\rho^{-1}(S)} J^{-1}(g, DX) f_n(g^{-1}X; \theta_0) \cdot d\nu(g) \\ &= k(D) \int_{\rho^{-1}(S)} \prod f(\Gamma^{-1}(\mathbf{x}_i - \mathscr{B}\mathbf{v}_i)) |\Gamma|^{-p-n}\, d\mathscr{B}\, d\Gamma\,. \end{aligned} \tag{9a}$$

It would be nice to find a probability element of the form $g^*(\theta \mid X)\, d\theta$, but this may be difficult, since θ is a coset, and $d\theta$ is thus not the simple $p(p + r)$-order infinitesimal $d\mathscr{B}\, d\Gamma$, but is an infinitesimal of order $\dim \Omega = \dim G - \dim H(\theta) = (2r + p + 1)p/2$, whose expression may be difficult to find in terms of the coefficients b_{ij} and γ_{ij} of $\mathscr{B}$ and Γ.

3. A relation of structural with Bayesian inference. In this section we invoke assumptions 1 to 7. If b is nonnegative and ξ is a σ-additive, σ-finite positive measure on Ω, then $dB(\theta) = b(\theta) \cdot d\xi(\theta)$ is said to be the element of the *prior quasi-probability* B, and b is a *prior quasi-density*. If $B(\Omega) = 1$, then B is a *prior probability*. Let $f(\cdot\,; \theta)$ be the density of x with respect to a σ-finite measure λ. A quasi-density is said to be *admissible* with respect to ξ and the model if

$$h(x) = \int_\Omega b(\theta) f(x; \theta)\, d\xi(\theta) < \infty$$

for all x a.e. (λ). The Bayes posterior probability element for θ given x is

$$dP_b(\theta \mid x) = \frac{b(\theta) \cdot f(x; \theta)}{h(x)}\, d\xi(\theta)\,.$$

If the prior distribution is defined on a sub-σ-field of the Borel field of Ω, then the posterior $P_b(S \mid x)$ is defined only on this sub-σ-field.

THEOREM 5. *If the model satisfies assumptions* 1 *to* 8, *and* $H(x)$ *and* $H(\theta_0)$ *are compact, then there is a right invariant measure* λ_Ω *on* Ω. *This* λ_Ω *is also a left* *relatively invariant measure with left modulus $\Delta(g)^{-1}$; it is an admissible prior quasi-density; the Bayes posterior given this prior and the observation x will agree with the structural distribution $P(S \mid x)$ for all S for which the structural distribution is defined.*

PROOF. Existence of λ_Ω follows from Weil's theorem (note that $\Omega = G/H$, hence right invariance is defined); by Nachbin, Proposition 27, its left modulus is $\Delta(g)^{-1}$. Now,

$$\begin{aligned} \int_S f(x;\theta)\, d\lambda_\Omega(\theta) &= \int_S f([\theta]^{-1}x; \theta_0)\delta^{-1}([\theta])\, d\lambda_\Omega(\theta) \\ &= \int_{\rho^{-1}(S)} f(g^{-1}x;\theta_0)\, d\mu(g) \qquad \text{(by Lemma 1)} \\ &= \int_{\rho^{-1}(S)} f(g^{-1}x;\theta_0)\Delta(g)\, d\nu(g)\,. \end{aligned} \tag{10}$$

Comparing with (6b) and bearing in mind that $J(g, Dx) = \delta(g) = \Delta(g)^{-1}$, we see that (10) equals $k(D)^{-1}P(S \mid x)$. Now, $h(x) = \int_\Omega f(x;\theta)\, d\lambda_\Omega(\theta) = k(D)^{-1}P(\Omega \mid x)$ by the above; this equals $k(D)^{-1}$ by Theorem 2, which proves admissibility. The Bayes posterior is

$$\begin{aligned} P_b(S \mid x) &= h(x)^{-1} \int_S f(x;\theta)\, d\lambda_\Omega(\theta) \\ &= P(S \mid x)\,. \quad \square \end{aligned}$$

COROLLARY. *Under the assumptions of Theorem 5, if λ is any left relatively invariant measure on Ω with left modulus δ, then $d\lambda_\Omega = \delta([\theta])^{-1}\Delta([\theta])^{-1}\, d\lambda(\theta)$ is the element of an admissible quasi-prior which yields Bayes posterior probabilities equal to the structural ones (for $H(x)\cdot S = S$).*

Note that the structural probability is defined on measurable sets of right $H(x)$-cosets, while the Bayes solution is defined on all measurable subsets of $\Omega = G$, whether they are sets of cosets or not. However, we see that the two solutions are equal if both exist for a given set S.

EXAMPLE 1. If $x_1, \cdots, x_n$ are i.i.d. observations on a sphere, $G = SO(3)$ and $G\theta = \Omega$, then Theorem 5 holds even if $f(\cdot\,; \theta_0)$ has symmetries.

EXAMPLE 2. In the multivariate model of Example 3, Section 2, the density of the structural distribution will be the same as that of the Bayes posterior obtained from the prior

$$P(\theta \in S) = \int_{\rho^{-1}(S)} \frac{d\mathscr{B}\cdot d\Gamma}{|\Gamma|^p}$$

where

$$g = \begin{pmatrix} I & 0 \\ \mathscr{B} & \Gamma \end{pmatrix}$$

is a generic group element.

4. A "consistency" criterion. Here we prove that extended structural inference satisfies a certain "consistency" criterion proposed (e.g. [2] and [10]) by

Sprott. Let x be a sample from a statistical model $(\mathscr{X}, \Omega, \{P_\theta\})$ which is invariant under a group G; let y be a sample independent of x from another invariant statistical model $(Y, \Omega, \{P_\theta'\})$ with the same parameter space and the same group G; in both models, we want G to act in the same way on Ω. We say that the models satisfy *criterion L* if the structural distribution for x, used as a prior distribution for a Bayes analysis of y, would yield a posterior distribution whose density function is the same as that for the structural distribution from the combined sample (x, y). This definition may remind readers of A. D. Roy's pistimetric inference (Roy, (1960)), in which a fiducial distribution from part of the sample is used as a Bayes prior for the rest of the sample. In those cases in which fiducial and structural answers coincide, criterion L says that pistimetric and pure structural inference give the same distributions for the parameter. By Section 3, these distributions are also Bayes. Condition L implies another consistency property: if the structural distribution from x is used as a prior for a Bayes analysis given the observation y, then the posterior distribution is unchanged if the order of x and y is reversed. Fraser has shown ([1] page 275) that L holds if the models are such that G is exactly transitive on $\mathscr{X}$, $\mathscr{Y}$ and Ω. The results of Sections 1 and 2 permit more general conclusions, as we now explain.

THEOREM. *If assumptions* 1 *to* 8 *hold for both models, and* $H(x)$, $H(y)$ *and* $H(\theta)$ *are compact, then the models satisfy criterion L.*

PROOF. We shall put the right invariant measure λ_Ω from Theorem 5 on Ω; $d\xi_1 = d\lambda_1 \times dl_1$ on the first model and $d\xi_2 = d\lambda_2 \times dl_2$ on the second model, where $d\lambda_1$ and $d\lambda_2$ are left invariant. Let the density functions of the models be f_1 and f_2 respectively. By (6b) the structural distribution given x is (since $J = \delta = 1$):

$$dP(S \mid x) = k(Dx) \int_{\rho^{-1}(S)} f_1(g^{-1}x; \theta_0)\, d\nu(g)\,.$$

Using this as a prior on Ω for a Bayes analysis of y, the joint Bayes probability of θ and y is

$$\begin{aligned} dP(\theta, y) &= dP(y; \theta)\cdot dP(\theta \mid x) \\ &= kf_2([\theta]^{-1}y; \theta_0)\, d\xi(y)\cdot \int_{\rho^{-1}(\theta)} f_1(g^{-1}x; \theta_0)\, d\nu(g)\,. \end{aligned}$$

By Lemma 1, the last term of this expression is $f_1([\theta]^{-1}x; \theta_0)\mu_H(H(\theta_0))$, hence

$$\begin{aligned} dP_b(\theta \in S \mid y) &= k \int_S f_2([\theta]^{-1}y; \theta_0) f_1([\theta]^{-1}x; \theta_0)\cdot d\lambda_\Omega(\theta) \\ &= k \int_{\rho^{-1}(S)} f_2(g^{-1}y; \theta_0) f_1(g^{-1}x; \theta_0)\, d\nu(g)\,. \end{aligned} \tag{11}$$

Now let us look at the combined model: it has the sample space $\mathscr{X} \times \mathscr{Y}$; G acts on the sample space like so: $g(x, y) = (gx, gy)$. We can use $[(x, y)] = [x]$, in which case $D(x, y) = (Dx, [x]^{-1}y)$. $H(x, y) = H(x) \cap H(y)$, which is compact. We can put the measure λ_Ω on Ω, and the measure $\lambda_1 \times l_1 \times \lambda_2 \times l_2$ on $\mathscr{X} \times \mathscr{Y}$.

The joint density of the observation (x, y) is

$$dP((x, y); \theta) = f_1(x; \theta)f_2(y; \theta)\, d\lambda_1(\pi_1(x))\, dl_1(Gx)\, d\lambda_2(\pi_2(y))\, dl_2(Gy) .$$

By (6b), the structural distribution given the combined sample is

$$P(S \mid (x, y)) = k \int_{\rho^{-1}(S)} f_1(g^{-1}x; \theta_0)f_2(g^{-1}y; \theta_0)\, d\nu(g) ,$$

the same as (11). □

APPENDIX

Measure theory. It is known that if the group G has a locally compact topology, and the group operations are continuous with respect to this topology, then there is a non-trivial measure μ which is left-*invariant* (i.e. $\mu(F) > 0$ if F is open; $\mu(E) = \mu(gE)$ for all g in G and measurable $E \subset G$). If ν is defined by $\nu(E) = \mu(E^{-1})$, then ν is right-invariant. There is a homomorphism Δ called the *group modulus*, $\Delta : G \to R^+$ such that $d\mu(g) = \Delta(g)\, d\nu(g)$; $\Delta = 1$ (and $\mu = \nu$) if G is commutative or compact, $\mu(G) < \infty$ if G is compact. (See e.g. Nachbin, Chapter 2.)

λ is a left relatively invariant measure on $\mathscr{X}$ with modulus δ if $d\lambda(gx) = \delta(g)\, d\lambda(x)$ (g fixed in G; x a generic point of $\mathscr{X}$), where $\delta(g)$ is independent of x. Then δ is a continuous homomorphism on $G((g_1 g_2) = \delta(g_1) \cdot \delta(g_2))$, and not to be confused with the *group* modulus Δ.

In what follows, H is a closed subgroup of the locally compact group $G \cdot \Delta^G$ and Δ^H are the moduli of G and H respectively. (See [7] page 138.)

Weil's theorem. There exists a unique non-trivial left relatively invariant measure λ on G/H with modulus δ iff δ is a continuous homomorphism from G to the positive reals such that $\Delta^H(h)/\Delta^G(h) = \delta(h)$ for all h in H. For any real integrable function f on G/H, and measurable subset S of G/H, then the integral of f with respect to λ may be written as

$$\int_S f(x)\, d\lambda(x) = \int_{\rho^{-1}(S)} f^*(g)\delta(g)\, d\mu(g) \tag{12}$$

where μ is any non-trivial left invariant measure on G, and f^* is any function on G such that

$$f(x) = \int_H f^*(xh)\, d\mu_H(h) ,$$

(μ_H being any left invariant measure on H). Conversely, the right-hand side of (12) defines a left relatively invariant integral with modulus δ if f^* is as defined, and the above conditions on δ hold. Such an f^* must exist for any integrable f.

Acknowledgments. The writer wishes to thank D. A. S. Fraser for his help at all stages in the preparation of this paper. Conversations with others, notably R. P. Bhargava, A. Kalotay and L. Steinberg have also been useful.

REFERENCES

[1] FRASER, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.
[2] FRASER, D. A. S. (1962). On the consistency of the fiducial method. *J. Roy. Statist. Soc. Ser. B* **24** 425–434.
[3] FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.
[4] FRASER, D. A. S. and STEINBERG, L. (1968). Structural inference and marginal likelihood for transforms of the multilinear model. Circulated at I.M.S. meeting, Madison.
[5] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.
[6] MONTGOMERY, D. and ZIPPIN, L. (1955). *Topological Transformation Groups*. Interscience, New York.
[7] NACHBIN, L. (1965). *The Haar Integral*. Van Nostrand, Princeton.
[8] ROSENBLATT, M. (1966). Query No. 16. *Technometrics* **8** 367–368.
[9] ROY, A. D. (1960). A note on prediction from an autoregressive process using pistimetric probability. *J. Roy. Statist. Soc. Sèr. B* **22** 97–103.
[10] SPROTT, D. A. (1960). Necessary restrictions for distributions a posteriori. *J. Roy. Statist. Soc. Ser. B* **22** 312–318.