The Annals of Statistics
1975, Vol. 3, No. 3, 559-572

PARAMETER FACTORIZATION AND INFERENCE BASED ON SIGNIFICANCE, LIKELIHOOD, AND OBJECTIVE POSTERIOR

BY D. A. S. FRASER AND JOCK MACKAY

University of Toronto and University of Manitoba

The concepts of significance, likelihood, and objective posterior have wide ranges of application in statistics. For certain very simple applications—single location variable—there has been recognition that the three concepts produce equivalent numerical results, specifically the equality of observed level of significance, integrated likelihood extreme values, and integrated objective posterior extreme values. The most general model permitting the use of the three concepts for the full parameter, and indeed for component parameters, is a structural model (or *probability-space model*). This paper examines the three concepts for a structural model and shows for both the full parameter and for component parameters the essential equivalence of observed level of significance, integrated likelihood extreme values, and integrated objective posterior extreme values.

1. Introduction and summary. There has been some awareness in the statistical profession that for very simple models a numerical equivalence can be found among significance, likelihood, and objective posterior. For this let y be a real variable with model $f(y-\theta)$ having a real location parameter θ; the variable y can be the resultant of a sufficiency reduction. The significance test of a value θ_0 is based on the observed value $y_0-\theta_0$ in relation to the distribution $f(z)\,dz$ for the departure $z=y-\theta_0$ from hypothesis; the observed level of significance for a departure $y_0-\theta_0$ or larger is usually the tail area probability $\alpha(\theta_0)=\int_{y_0-\theta_0}^{\infty}f(z)\,dz$. The likelihood function from the observed y_0 is given essentially by $f(y_0-\theta)$; the normed tail area likelihood for θ as small as or smaller than θ_0 is given by $\alpha(\theta_0)=\int_{-\infty}^{\theta_0}f(y_0-\theta)\,d\theta$. The objective posterior for θ from the observed y_0 has density $f(y_0-\theta)$ as obtained from the appropriate structural model; the tail area probability for θ as small as or smaller than θ_0 is $\alpha(\theta_0)=\int_{-\infty}^{\theta_0}f(y_0-\theta)\,d\theta$. These evaluations are numerically equal and in fact they correspond by rather obvious 1–1–1 transformations. A similar equivalence is obtained for the two-sided situation or for any other choice of discrepancy measure on the "usual" values of z.

The first of the concepts, the test of significance, has the widest range of applications. The test of significance requires a model but only under the hypothesis; the model may describe the whole response or merely some reduction of the response. For a test of significance the test statistic is chosen pragmatically

Received June 1972; revised July, 1974.

AMS 1970 *subject classifications.* Primary 62A05, 62A10, 62A15, 62A99.

Key words and phrases. Inference procedures, comparisons of inference procedures, significance tests likelihood interval, objective posterior intervals, structural intervals and regions.

to be sensitive to departures from what is expected under the hypothesis. If a full model is available covering the hypotheses and alternatives then the measure of departure is sometimes obtained from power considerations. If a structural model is available, then for certain hypotheses the measure of departure is a necessary consequence of direct probability analysis.

The likelihood function has the next widest range of applications. The likelihood function requires a full model under the various possible hypotheses; the model may describe the whole response or merely some reduction of the response. The likelihood function can be examined in a variety of ways including integrating, profiling, and sectioning; belief in the Bayesian approach sanctifies the first of these ways.

The objective posterior has the most limited range of applications. The objective posterior needs a structural model which has a probability space to describe the underlying variation affecting the system under investigation. The structural model and necessary analysis are summarized in Sections 2 and 3.

The common range of applications for the three concepts is that covered by the structural models. In the context of this common range of applications Section 6 presents the essential equivalence of the three concepts, equivalence as discussed for the simple example.

The equivalence of the three concepts also occurs for component parameters. The factorization of the parameter in a way that permits the use of the various concepts is examined in Sections 4 and 5; the equivalence of the three concepts as presented in Section 6 includes this additional generality of component parameters.

Then in Sections 7 and 8 we have an example that illustrates inference for component prarameters in a familiar context and an addendum that presents the factorization of Haar measure as needed for the analysis.

2. The model. Consider a model for some physical or social system. Let Z be a variable that describes the variation affecting the response of the system; and suppose that Z takes values in a connected open set $\mathscr{S} \subset \mathbb{R}^N$ and has a probability measure P with density f with respect to Lebesgue measure on $(\mathscr{S}, \mathscr{B}^N)$. And let $Y = \theta Z$ be a variable that describes the response of the system where θ is the transformation that presents the response Y from the variation Z; and suppose that θ is an element of a connected open set $G \subset \mathbb{R}^L$ and that G is an exact Lie group on $\mathscr{S}$ (G is exact on $\mathscr{S}$ if $gZ = hZ$ implies $g = h$). The variation in the system is described by the probability space $(\mathscr{S}, \mathscr{B}^N, P)$; and the possible presentations for the response are described by G. The combination $(\mathscr{S}, \mathscr{B}^N, P; G)$ is a structural model or a *probability space model* (Fraser (1968)).

The model can be extended to cover the case where the distribution for the variation is known only up to a parameter λ with values in a parameter space Λ. The extended model is then given by $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\}; G)$; for any specified λ, this is a probability space model.

As an illustration consider the linear model $\mathbf{y} = X\boldsymbol{\beta} + \sigma\mathbf{z}$ where $\mathbf{z}$ is a random sample of n from a distribution P with density f and X is a full rank $n \times r$ design matrix. The group multiplication can be expressed by matrix multiplication $Y = \theta Z$ with

$$Y = \begin{bmatrix} X' \\ \mathbf{y}' \end{bmatrix}, \qquad \theta = \begin{bmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{bmatrix}, \qquad Z = \begin{bmatrix} X' \\ \mathbf{z}' \end{bmatrix}; \tag{2.1}$$

the design matrix is placed as a fixed print in the representations Y for $\mathbf{y}$ and Z for $\mathbf{z}$. The probability space is $\{\mathbb{R}^N - \mathscr{L}(X), \mathscr{B}^N, P^N\}$ where $\mathscr{L}(X)$, the linear space generated by the columns of X, is deleted to satisfy the exactness condition, and the group is $G = \{\theta : \boldsymbol{\beta} \in \mathbb{R}^r, \sigma \in \mathbb{R}^+\}$ using matrix multiplication. The density can be indexed by a parameter λ to cover applications where there is a range of possibilities for the distribution describing the variation; for example λ could be the degrees-of-freedom of a Student distribution thus allowing more probability in the tails of the distribution than with the usual normal distribution ($\lambda = \infty$).

3. The analysis of the model. In this section we analyze an observed response Y_0 in relation to the extended model $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\}; G)$.

First we introduce some notation for some mathematical objects that arise naturally in the course of the analysis. The group G applied to a point Z generates the set $GZ = \{gZ : g \in G\}$ of image points under the action of the group. The various sets GZ with Z in $\mathscr{S}$ are either identical or disjoint. It follows that the class $\{GZ : Z \in \mathscr{S}\}$ of distinct sets is a partition of $\mathscr{S}$ and the mapping $Z \mapsto GZ$ carries Z into the set that contains Z.

The assumptions in Section 2 provide for the following definitions needed in the analysis. Let $[\cdot]$ be a continuously differentiable function from $\mathscr{S}$ into G such that $[hZ] = h[Z]$ for all h in G and Z in $\mathscr{S}$. And then let $D(Z) = [Z]^{-1}Z$ and $Q = \{D(Z) : Z \in \mathscr{S}\} = \{Z : [Z] = e\}$ where e is the identity element of the group. It follows then that $Z = [Z]D(Z) = gD$ with $g = [Z]$ unique in G and $D = D(Z)$ unique in Q. Thus $Z \leftrightarrow (g, D)$ gives a continuous 1–1 correspondence between $\mathscr{S}$ and $G \times Q$. Note that the points in Q are in 1–1 correspondence with the elements of the partition $\{GZ : Z \in \mathscr{S}\}$ of the space $\mathscr{S}$.

Now consider the observed repsonse Y_0 in relation to the extended model $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\}; G)$. According to the model the observed Y_0 is some transformation θ of a realized value Z from the statistical space $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\})$. The information concerning Z can be separated into two parts:

(i) *the value Z is a realization from one of the distributions in* $\{P_\lambda : \lambda \in \Lambda\}$;

(ii) *the value Z must satisfy $Y_0 = \theta Z$ for some θ in G.*

The information (ii) concerning the value Z can be expressed alternatively as:

(ii)′ *the value Z is some point in the set* $\{h^{-1}Y_0 : h \in G\} = GY_0$;

note that this is a set in the partition described in the preceding paragraph. The information can then be expressed as (ii)″ *the function* $D(Z)$ *has the observed value* $D(Y_0) = D_0$ say.

We thus have the information: there is a realization Z from the statistical space $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\})$ and the only observational information is that the function $D = D(Z)$ has the value D_0. Classical probability theory describes this in terms of:

(iii) *the probability for what has been observed,* $D(Z) = D_0$; and
(iv) *the conditional distribution of the unobserved* g *given the observed* $D(Z) = D_0$.

The change of variable $Z \leftrightarrow ([Z], D(Z)) = (g, D)$ provides a coordinate D appropriate to the observational condition and a complementing coordinate g which indexes points given the condition. The Jacobian modulus for the transformation is

$$J(Z) = \left| \frac{\partial Z}{\partial g, D} \right|$$

which involves Euclidean volume for Z in $\mathbb{R}^N$ as compared with Euclidean volume for g in $\mathbb{R}^L$ and for D tangent to Q; typically this is difficult to calculate. The transformation can be reexpressed with two additional steps using a transformation h on G and the inverse transformation h^{-1} on $\mathscr{S}$: $Z \leftrightarrow h^{-1}Z \leftrightarrow (h^{-1}g, D) \leftrightarrow (g, D)$. For this let

$$J_N(h : D) = \left| \frac{\partial hZ}{\partial Z} \right|, \qquad J_L(h : g) = \left| \frac{\partial hg}{\partial g} \right| \tag{3.1}$$

be the easily calculated Jacobian moduli for the transformation h applied on $\mathscr{S}$ and on G. The reexpressed transformation holds for arbitrary h and hence in particular for $h = g$; thus

$$dZ = J_N(g : D) J(D) J_L^{-1}(g : e)\, dg\, dD\,. \tag{3.2}$$

The adjusted differential $J_L^{-1}(g : e)\, dg = d\mu(g)$ is the left Haar measure on G standardized to Euclidean volume at the identity.

The probability differential on the space $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\})$ can now be expressed in terms of the new variables

$$f_\lambda(Z)\, dZ = f_\lambda(g : D)\, d\mu(g) \cdot k_\lambda(D) J(D)\, dD \tag{3.3}$$

where

$$k_\lambda(D) = \int_G f_\lambda(gD) J_N(g : D)\, d\mu(g) \tag{3.4}$$

is the marginal density for D relative $J(D)\, dD$ on Q and

$$f_\lambda(g : D) = k_\lambda^{-1}(D) f_\lambda(gD) J_N(g : D) \tag{3.5}$$

is the conditional density for g relative to the standardized left Haar $d\mu(g)$ on G.

Now consider further the observed response Y_0 and the information concerning the realization Z on the space $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\})$. The classical probability description of Z is in terms of:

(iii) *the probability for what has been observed*, $D(Z) = D_0$; and
(iv) *the conditional distribution of the unobserved g given the observed* $D(Z) = D_0$.

The probability differential $f_\lambda(Z)\,dZ$ can be factored accordingly as in (3.3) and we obtain

(iii) *the likelihood* $k_\lambda(D_0)J(D_0)\,dD$ *for the observed* $D = D_0$.
(iv) *the conditional distribution* $f_\lambda(g : D)\,d\mu(g)$ *describing the unobserved* $g = [Z]$ *given the observed* $D(Z) = D_0$.

This provides at least the likelihood function for the parameter λ; and for given λ it provides probability analysis for the unknown $g = [Z]$ and thus for the unknown θ.

The analysis of the unknown θ is based on the distribution describing g. We have $Y_0 = \theta g D_0$ with a one–one correspondence between possible values for g and possible values for θ; and thus tests, confidence intervals, or probability intervals for θ are obtained directly from the distribution describing g. The conditional distribution given the set GY_0 is perhaps most conveniently described by having the coordinates $g = [Z]$ chosen relative to a surface Q passing through the observed response. If this notational accommodation is made then the preceding equation becomes $Y_0 = \theta g Y_0$ and the one–one correspondence is given simply by the equation $\theta = g^{-1}$ or $g = \theta^{-1}$. The distribution for g then provides in a very simple way, tests, confidence intervals, and probability intervals for θ. The details of this are examined in Section 6 as a special case of inference for parameter components.

4. Parameter factorization. The analysis of a real parameter is a relatively straightforward and accessible part of statistical inference. With a multidimensional parameter however a whole range of complexities and ambiguities arises. In classical statistics there are difficulties connected with for example nuisance parameters, lack of completeness for minimal sufficient statistics, nonexistence of uniformly most powerful tests, nonexistence of uniformly minimum variance unbiased estimates, and others. And in likelihood analysis there are ambiguities and uncertainties connected with integrating or profiling or sectioning the likelihood function. And in Bayesian analysis the meaning for a prior becomes even more questionable and ambiguous.

The common approach to a multidimensional parameter is to separate it into a sequence of real parameters and to analyze the components separately and usually sequentially. The analysis of variance for linear models provides a familiar example; the regression parameters are analyzed in sequence from the bottom towards the top of the analysis-of-variance table.

With the model examined in this paper a multidimensional parameter means a multidimensional group G. And a separation of the parameter into components means a set of new coordinates for an element of the group. If we are then to be concerned with the three basic inference concepts we must restrict our attention to parameter separations that produce components each with the structure that allows the three inference concepts.

As a simple first case suppose that θ can be represented uniquely as $\theta = \theta_2\theta_1$ where θ_1 is an element of a subgroup H_1 of G, is and θ_2 an element of another subgroup H_2 of G; then $G = H_2H_1$ is called a semidirect product of the component groups H_1 and H_2. For the analysis connected with such component parameters we will need a factorization of $g = [Z]$ in the reverse order: $g = h_1h_2$ where h_1 is in H_1 and h_2 is in H_2. The existence and uniqueness of the reverse factorization $G = H_2H_1$ follows from the 1–1 mapping $g \leftrightarrow g^{-1}$.

Now consider the analysis of the model $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\}; G)$ in relation to the observed response Y_0. The conditional distribution given the observed set GY_0 is perhaps most conveniently described in terms of component parameters by having the coordinates $g = [Z]$ chosen relative to a surface Q passing through the observed response. With this choice of notation we obtain

$$Y_0 = \theta g Y_0 = \theta_2\theta_1 h_1 h_2 Y_0$$

and hence $\theta_2\theta_1 = (h_1h_2)^{-1} = h_2^{-1}h_1^{-1}$ or equivalently $\theta_2 = h_2^{-1}$, $\theta_1 = h_1^{-1}$. Thus the possible values for θ_2 are in one-one correspondence with possible values for h_2 by the equation $\theta_2 = h_2^{-1}$ and possible values for θ_1 are in one-one correspondence with possible values for h_1 by the equation $\theta_1 = h_1^{-1}$. For any specified λ, the order of the inference procedure is first the analysis of θ_2 based on the distribution of h_2 given the observed set GY_0, and then the analysis of θ_1 based on the distribution of h_1 given h_2 and the set GY_0. In the second stage the component h_2 or correspondingly θ_2^{-1} is given assumed values based in whole or part on inferences from the first stage of the analysis.

Some justification for this order for the inference procedure may be found in [2]; further discussion will be found in a paper now in preparation. In an application there is often a natural order for examining parameter components. With a structural model this may agree with that just described; the three inference concepts can then be compared. In other cases the three methods are not available for component parameters. Thus for our discussion of the three concepts as applied to component parameters we will assume the factorization form just described.

The generalization covering factorization into more than two components is straightforward: a component on the left is assessed first, the remainder is then examined in the way the original parameter is examined—by assessing a component on the left.

For this generalization we assume a sequence of subgroups of G:

$$G = G_r > G_{r-1} > \cdots > G_0 = \{e\} .$$

And we suppose that $H_r, \cdots, H_1$ is a generating set of subgroups such that $G_s = H_sG_{s-1}$ is a semidirect product. We then have $G = H_r H_{r-1} \cdots H_1$ and θ can be factored as $\theta = \theta_r \theta_{r-1} \cdots \theta_1$ with θ_s unique in H_s.

For an illustration consider the linear model as described in Section 2. The usual analysis of variance procedure suggests the factorization

$$\theta = \begin{bmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ 0 \cdots 0\beta_r & 1 \end{bmatrix} \cdots \begin{bmatrix} I & \mathbf{0} \\ \beta_1 0 \cdots 0 & 1 \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0}' & \sigma \end{bmatrix} = \theta_r \cdots \theta_1 \theta_0 , \tag{4.1}$$

where the component group for each of $\theta_1, \cdots, \theta_r$ is the additive group on the reals and for θ_0 is the multiplicative group on the positive reals. In the usual context the parameter β_1 would give location relative to the one-vector, $\cdots$, up to β_r relative to a vector representing a high power of an input variable, or a high-order interaction, or a covariance difference.

Now in general consider the analysis of the model $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\}; G)$ in relation to the observed response Y_0. Again it is convenient to use coordinates relative to a surface Q passing through the observed response. This leads to a factorization of $g = [Z]$ in the reverse order: $g = h_1 h_2 \cdots h_r$ where h_s is unique in H_s. With this choice of notation we obtain

$$Y_0 = \theta g Y_0 = \theta_r \cdots \theta_1 h_1 \cdots h_r Y_0$$

and hence $\theta_r \cdots \theta_1 = (h_1 \cdots h_r)^{-1} = h_r^{-1} \cdots h_1^{-1}$ or equivalently $\theta_s = h_s^{-1}$ for $s = 1, \cdots, r$. Thus the possible values for θ_s are in one-one correspondence with possible values for h_s by the equation $\theta_s = h_s^{-1}$.

The inference procedure would first involve the assessment of λ on the basis of what has been observed, the set GY_0. Usually this would involve at least a likelihood analysis on the basis of (iii) in Section 3.

For any specified λ, the inference procedure would then involve the analysis of θ_r based on the conditional distribution of h_r given the observed set GY_0. The inference procedure would then involve the analysis of θ_{r-1} based on the conditional distribution of h_{r-1} given h_r and the set GY_0; in this stage the component h_r or correspondingly θ_r^{-1} would be given assumed values based in whole or part on inferences from the preceding stage. The inference procedure would then examine sequentially in the same manner the parameters $\theta_{r-2}, \cdots, \theta_2, \theta_1$.

The conditional distribution can be factored in accord with the preceding inference procedure. This is examined in the next section.

5. Density factorization. The analysis of an observed response Y_0 in Section 3 leads to the factorization $Z = gD$. The corresponding factorization of the probability differential is given by (3.3).

The analysis in Section 4 relative to component parameters leads to the factorization $Z = h_1 h_2 \cdots h_r D$ where $G = H_1 \cdots H_r$ is a semidirect product based on subgroups $G_1 = H_1$, $G_2 = H_1 H_2$, $G_3 = H_1 H_2 H_3, \cdots$. The factorization is used with $D = Y_0$ in accord with the special notation. We now consider the corresponding factorization of the probability differential.

In Section 3 we discussed the change of variable $Z = gD$. We now examine the change of variable $g = h_1 h_2 \cdots h_r$. The Haar measure $d\mu(g)$ can be factored as described in the addendum Section 8:

$$d\mu(g) = d\mu_1(h_1) \cdot \frac{\Delta_{(2)}(h_2)}{\Delta_2(h_2)} d\mu_2(h_2) \cdot \cdots \cdot \frac{\Delta_{(r)}(h_r)}{\Delta_r(h_r)} d\mu_r(h_r)$$

where μ_s, Δ_s are the left Haar measure and modular function for H_s and $\Delta_{(s)}$ is the modular function for $H_1 \cdots H_s$. This factorization can be collected into two parts corresponding to $(s) = 1 \cdots s$ and $[s+1] = s+1 \cdots r$. Let $h_{(s)} = h_1 \cdots h_s$ and $h_{[s+1]} = h_{s+1} \cdots h_r$. Then

$$d\mu(g) = d\mu_{(s)}(h_{(s)}) \cdot d\mu_{[s+1]}(h_{[s+1]})$$

where

$$d\mu_{(s)}(h_{(s)}) = d\mu_1(h_1) \cdot \frac{\Delta_{(2)}(h_2)}{\Delta_2(h_2)} d\mu_2(h_2) \cdot \cdots \cdot \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)} d\mu(h_s)$$

is the left Haar measure on $G_s = H_1 \cdots H_s$ and where

$$d\mu_{[s+1]}(h_{[s+1]}) = \frac{\Delta_{(s+1)}(h_{s+1})}{\Delta_{s+1}(h_{s+1})} d\mu_{s+1}(h_{s+1}) \cdot \cdots \cdot \frac{\Delta_{(r)}(h_r)}{\Delta_r(h_r)} d\mu_r(h_r)$$

is the quotient of left Haar on $G = H_1 \cdots H_r$ by left Haar on the subgroup $G_s = H_1 \cdots H_s$. Note that the quotient of adjacent support measures gives the appropriate component in the full factorization, that is,

$$\frac{d\mu_{[s]}(h_{[s]})}{d\mu_{[s+1]}(h_{[s+1]})} = \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)} d\mu_s(h_s) .$$

The marginal probability differential for $h_{s+1} \cdots h_r D = h_{[s+1]} D$ can be obtained by integrating the left or right side of (3.3):

$$\begin{aligned} &k_\lambda^{[s+1]}(h_{[s+1]} D)\, d\mu_{[s+1]}(h_{[s+1]}) J(D)\, dD \\ &\qquad = f_\lambda^{[s+1]}(h_{[s+1]} : D)\, d\mu_{[s+1]}(h_{[s+1]}) \cdot k_\lambda(D) J(D)\, dD \end{aligned}$$

where

$$\begin{aligned} k_\lambda^{[s+1]}(h_{[s+1]} D) &= \int_{G_s} f_\lambda(h_{[1]} D) J_N(h_{[1]} : D)\, d\mu_{(s)}(h_{(s)}) \\ &= \int_{H_s} k_\lambda^{[s]}(h_{[s]} D) \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)} d\mu_s(h_s) \end{aligned}$$

is the marginal density for $h_{[s+1]} D$ with respect to $d\mu_{[s+1]}(h_{[s+1]}) J(D)\, dD$ and can be calculated iteratively ($s = 1, 2, \cdots, r$) and where

$$\begin{aligned} f_\lambda^{[s+1]}(h_{[s+1]} : D) &= \int_{G_s} f_\lambda(h_{[1]} : D)\, d\mu_{(s)}(h_{(s)}) \\ &= \int_{H_s} f_\lambda^{[s]}(h_{[s]} : D) \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)} d\mu_s(h_s) \end{aligned}$$

is the marginal density for $h_{[s+1]}$ given D with respect to $d\mu_{[s+1]}(h_{[s+1]})$ and can be calculated iteratively ($s = 1, 2, \cdots, r$).

Conditional densities can be obtained by taking the ratio of adjacent marginal densities in the iterations just mentioned. Thus the conditional density of h_s

given $h_{[s+1]} D$ with respect to the quotient measure

$$\frac{d\mu_{[s]}(h_{[s]})}{d\mu_{[s+1]}(h_{[s+1]})} = \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)} d\mu_s(h_s)$$

is

$$f_\lambda^s(h_s : h_{[s+1]}, D) = \frac{k_\lambda^{[s]}(h_{[s]} D)}{k_\lambda^{[s+1]}(h_{[s+1]} D)} = \frac{f_\lambda^{[s]}(h_{[s]} : D)}{f_\lambda^{[s+1]}(h_{[s+1]} : D)} .$$

The probability differential for Z can now be factored in accord with the representation $Z = h_1 \cdots h_r D = h_{(s)} h_{[s+1]} D$:

$$\begin{aligned}
f_\lambda(Z)\, dZ &= f_\lambda^1(h_1 : h_{[2]} D)\, d\mu_1(h_1) \cdots \\
&\qquad f_\lambda^r(h_r : D) \frac{\Delta_{(r)}(h_r)}{\Delta_r(h_r)} d\mu_r(h_r) \cdot k_\lambda(D) J(D)\, dD \\
&= f_\lambda^{(s)}(h_{(s)} : h_{[s+1]}, D)\, d\mu_{(s)}(h_{(s)}) \\
&\qquad \times k_\lambda^{[s+1]}(h_{[s+1]} : D)\, d\mu_{[s+1]}(h_{[s+1]}) J(D)\, dD \qquad (5.1)\\
&= f_\lambda^{(s)}(h_{(s)} : h_{[s+1]} D)\, d\mu_{(s)}(h_{(s)}) \\
&\qquad \times f_\lambda^{[s+1]}(h_{[s+1]} : D)\, d\mu_{[s+1]}(h_{[s+1]}) \cdot k_\lambda(D) J(D)\, dD .
\end{aligned}$$

Note that the middle factorization is in terms of the marginal for $h_{[s+1]} D$ preceded by the conditional of $h_{(s)}$, and that the last factorization is in terms of the marginal for D preceded by the conditional for $h_{[s+1]}$ preceded by the conditional for $h_{(s)}$. Also note that the conditional density

$$f_\lambda^{(s)}(h_{(s)} : h_{[s+1]} D) = \frac{k_\lambda^{[1]}(h_{[1]} D)}{f_\lambda^{[s+1]}(h_{[s+1]} : D)} = \frac{f_\lambda^{[1]}(h_{[1]} : D)}{f_\lambda^{[s+1]}(h_{[s+1]} : D)}$$

is available as a quotient of the marginal densities discussed earlier.

6. The factored analysis. Consider an observed response Y_0 in relation to the extended model $(\mathscr{S}, \mathscr{B}^N, \{P_\lambda : \lambda \in \Lambda\}; G)$. And suppose that the parameter θ in G is factored as $\theta = \theta_r \cdots \theta_1$ where $G = H_r \cdots H_1$ is a semidirect product of subgroups $H_1, \cdots, H_r$ as described in Section 4.

For the analysis we choose the convenient coordinates with Q passing through the observed response $Y_0 = D_0$ and we factor g in the reverse order $g = h_1 \cdots h_r$ with h_s unique in H_s. The inference procedure discussed in Section 4 can be used with the probability factorization (5.1). We obtain

(i) *The likelihood $k_\lambda(D_0)J(D_0)\, dD$ for the observed GY_0.* This can be examined as a likelihood function alone or in relation to the distribution of possible likelihood functions.

(ii) *The conditional distribution $f_\lambda^r(h_r : D)\Delta_{(r)}(h_r)\Delta_r^{-1}(h_r)\, d\mu_r(h_r)$ describing the unobserved h_r given D_0.* For given λ this provides the probability analysis for h_r and thus tests, confidence intervals, and probability intervals for θ_r.

(iii) *The conditional distribution $f_\lambda^{r-1}(h_{r-1} : h_r D)\Delta_{(r-1)}(h_{r-1})\Delta_{r-1}^{-1}(h_{r-1})\, d\mu_{r-1}(h_{r-1})$ describing the unobserved h_{r-1} given $h_r D_0$.* For given λ and $\theta_r = h_r^{-1}$ this provides the probability analysis for h_{r-1} and thus tests, confidence intervals, and probability intervals for θ_{r-1}. And so on, to

(iv) *The conditional distribution* $f_\lambda^1(h_1 : h_{[2]} D)\, d\mu_1(h_1)$ *describing the unobserved* h_1 *given* $h_{[2]} D_0$. For given λ and $\theta_s = h_s^{-1}$ $(s = 2, \cdots, r)$ this provides the probability analysis for h_1 and thus tests, confidence intervals, and probability intervals for θ_1.

6.1. *Observed significance level.* Consider given values for $\theta_{s+1}, \cdots, \theta_r$, λ and a hypothesis that specifies the value of θ_s. Specifying a value for θ_s determines an observed value for $h_s = \theta_s^{-1}$. The hypothesis is tested by comparing the value of h_s observed under the hypothesis with the distribution

$$f_\lambda^s(h_s : h_{[s+1]} D) \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)}\, d\mu_s(h_s)\,. \tag{6.1}$$

The common comparison is by means of the observed level of significance, an integrated tail area based on departure criteria applied to the distribution (6.1) on the space H_s.

6.2. *Integrated likelihood.* For given values of $\theta_{s+1}, \cdots, \theta_r$, λ the model takes the form of a structural model with group G_{s-1} and an additional parameter θ_s with values in a group H_s. The determination of likelihood in some general contexts has been examined in [3] and the most highly structured determination is transit likelihood.

To keep the notation relatively simple we assume that the given values of $\theta_{s+1}, \cdots, \theta_r$ are all equal to the identity e; this can of course be accomplished in effect by using the adjusted response $\theta_{s+1}^{-1} \cdots \theta_r^{-1} Y$.

Now consider the parameter θ_s in relation to the structural model involving $\theta_1 \cdots \theta_{s-1}$ in G_{s-1}. The adjusted response $\theta_s^{-1} Y$ has a probability space model with group G_{s-1} and equation

$$\theta_s^{-1} Y_0 = \theta_{s-1} \cdots \theta_1 h_1 \cdots h_{s-1} \cdot h_s Y_0\,.$$

The probability for the observed set $G_{s-1}\theta_s^{-1} Y_0$ is

$$\begin{aligned} k_\lambda^{[s]}(h_{[s]} D_0)\, d\mu_{[s]}(h_{[s]}) J(D_0)\, dD &= f_\lambda^s(h_s : h_{[s+1]} D_0) \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)}\, d\mu_s(h_s) \\ &\quad \times k_\lambda^{[s+1]}(h_{[s+1]} D_0)\, d\mu_{[s+1]}(h_{[s+1]}) J(D_0)\, dD \end{aligned}$$

where $h_{s+1}, \cdots, h_r$ have their observed values (here equal to e) and h_s^{-1}. The preceding probability is expressed in terms of an invariant measure $d\mu_s(h_s)$ along the H_s orbit through Y_0. It follows that the preceding expression gives transit likelihood and omitting multiplicative constants we obtain

$$f_\lambda^s(\theta_s^{-1} : h_{[s+1]} D_0) \frac{\Delta_{(s)}(\theta_s^{-1})}{\Delta_s(\theta_s^{-1})}\, d\mu_s(h_s)\,.$$

A hypothesized value for θ_s can be assessed by calculating an appropriate tail area using the natural measure $d\mu_s(\theta_s^{-1})$. The resulting expression is equal to the observed level of significance as obtained in Section 6.1.

6.3. *Integrated objective posterior.* With $\theta_{s+1}, \cdots, \theta_r, \lambda$ given, the equation $Y_0 = \theta_r \cdots \theta_1 h_1 \cdots h_r Y_0$ simplifies to

$$\theta_s = h_s^{-1}, \cdots, \theta_1 = h_1^{-1}.$$

The analysis in Section 5 then gives the distribution

$$f_\lambda^{(s)}(h_{(s)} : h_{[s+1]} D_0) \, d\mu_{(s)}(h_{(s)})$$

describing $h_{(s)} = h_1 \cdots h_s$; and the marginal distribution

$$f_\lambda^s(h_s : h_{[s+1]} D_0) \frac{\Delta_{(s)}(h_s)}{\Delta_s(h_s)} \, d\mu_s(h_s)$$

describing h_s alone. The equation $\theta_s = h_s^{-1}$ gives the parameter value that corresponds to a realized value h_s. The distribution describing these possible parameter values is obtained from the correspondence $h_s = \theta_s^{-1}$:

$$f_\lambda^s(\theta^{-1} : h_{[s+1]} D_0) \frac{\Delta_{(s)}(\theta_s^{-1})}{\Delta_s(\theta_s^{-1})} \, d\mu_s(\theta_s^{-1}).$$

A hypothesized value for θ_s can be assessed by calculating an appropriate tail area. The resulting expression is equal to the observed level of significance in Section 6.1 and the integrated likelihood in Section 6.2.

6.4. *Summary.* A value for a component parameter in a structured model can be assessed by an observed level of significance, an integrated likelihood tail area and an integrated objective posterior tail area. These are numerically equal as based on departure criteria applied to the function being integrated on the space which is a group. Thus these basic methods of inference are equivalent when they are all available. Likelihood has a larger range of applications and significance a still larger range.

7. The linear model. Consider the linear model as introduced in Sections 2 and 4. For any distribution f describing variation the various conditional distributions mentioned in Sections 5, 6 can be calculated by computer integration. For the case of normal error the distributions are available analytically and agree *as marginal distributions* with those used in the ordinary analysis of this model. As this classical normal case is familiar and simple it is presented here to exemplify and illustrate the logical form of the theoretical results presented in this paper. The normal model has so many symmetries that almost all approaches are equivalent and distinctions necessary in other contexts are lost. The examination of this familiar model thus emphasizes the inference pattern that is needed in more general contexts.

For simplicity we assume that the design matrix is orthonormal. And for notation we let $[Z] = e$ be the unit sphere $Q = \mathscr{L}^{\perp}(X)$. Then

$$Z = \begin{bmatrix} X' \\ \mathbf{z}' \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{b}'(\mathbf{z}) & s(\mathbf{z}) \end{bmatrix} \begin{bmatrix} X' \\ \mathbf{d}'(\mathbf{z}) \end{bmatrix} = [Z]D(Z)$$

where $\mathbf{b}(\mathbf{z})$ is the vector of regression coefficients, $s(\mathbf{z})$ the residual length, and $\mathbf{d}(\mathbf{z})$ the standardized residual. The probability differential follows immediately from (3.3):

$$\frac{1}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\Sigma b_j^2\}\, d\mathbf{b} \cdot \frac{A_{n-r}}{(2\pi)^{(n-r)/2}} \exp\left\{-\frac{s^2}{2}\right\} s^{n-r-1}\, d\mathbf{b}\, ds \cdot \frac{d\mathbf{d}}{A_{n-r}}$$

where $A_f = 2\pi^{f/2}/\Gamma(f/2)$ is the area of a unit sphere in $\mathbb{R}^f$.

Consider the parameter factorization (4.1).

The group element g can be factored in the reverse order as discussed generally in Section 4:

$$g = \begin{bmatrix} I & \mathbf{0} \\ b_1 \cdots b_r & s \end{bmatrix} = \begin{bmatrix} I & \mathbf{0} \\ 0 \cdots 0 & s \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \\ t_1 0 \cdots 0 & 1 \end{bmatrix} \cdots \begin{bmatrix} I & \mathbf{0} \\ 0 \cdots 0\, t_r & 1 \end{bmatrix}$$

where $\mathbf{t} = \mathbf{b}/s$. The left Haar factorization from Section 5 is simple

$$\frac{d\mathbf{b}\, ds}{s^{r+1}} = \frac{ds}{s}\, dt_1 \cdots dt_r\,.$$

The probability differential can then be reexpressed in terms of the new coordinates: $s, t_1, \cdots, t_r$:

$$\begin{aligned} &f^{[0]}(s, \mathbf{t} : \mathbf{d}) \frac{ds}{s}\, d\mathbf{t} \cdot \frac{d\mathbf{d}}{A_{n-r}} \\ &\quad = \frac{A_{n-r}}{(2\pi)^{n/2}} \exp\left\{-\frac{s^2}{2}(1 + \Sigma t_j^2)\right\} s^n \frac{ds}{s}\, dt_1 \cdots dt_r \cdot \frac{d\mathbf{d}}{A_{n-r}}\,. \end{aligned}$$

The special coordinates relative to the observed $\mathbf{y}$ are not needed as the transformation from $s, t_1, \cdots, t_r$ to $\sigma, \beta_1, \cdots, \beta_r$ is already a coordinate by coordinate transformation:

$$s(\mathbf{y}) = \sigma s\,, \qquad b_j(\mathbf{y}) = \beta_j + s(\mathbf{y})t_j \qquad\qquad j = 1, \cdots, r\,.$$

The marginal differentials from Section 5 can be obtained by simple integration; note the serial numbering $0, 1, \cdots, r$ of the $r + 1$ coordinates:

$$\begin{aligned} f^{[1]}(\mathbf{t} : \mathbf{d})\, d\mathbf{t} \cdot \frac{d\mathbf{d}}{A_{n-r}} &= \frac{A_{n-r}}{A_n} (1 + \textstyle\sum_1^r t_j^2)^{-n/2}\, d\mathbf{t} \cdot \frac{d\mathbf{d}}{A_{n-r}} \\ f^{[2]}(\mathbf{t}_{[2]} : \mathbf{d})\, d\mathbf{t}_{[2]} \cdot \frac{d\mathbf{d}}{A_{n-r}} &= \frac{A_{n-r}}{A_{n-1}} (1 + \textstyle\sum_2^r t_j^2)^{-(n-1)/2}\, d\mathbf{t}_{[2]} \cdot \frac{d\mathbf{d}}{A_{n-r}} \\ &\vdots \\ f^{[r]}(t_r : \mathbf{d})\, dt_r \cdot \frac{d\mathbf{d}}{A_{n-r}} &= \frac{A_{n-r}}{A_{n-r+1}} (1 + t_r^2)^{-(n-r+1)/2}\, dt_r \cdot \frac{d\mathbf{d}}{A_{n-r}} \\ k(\mathbf{d})\, d\mathbf{d} &= \frac{d\mathbf{d}}{A_{n-r}}\,. \end{aligned}$$

The conditional differentials are then obtained by division. The distribution of s given $\mathbf{t}$ and $\mathbf{d}$ is

$$f^0(s:\mathbf{t},\mathbf{d})\frac{ds}{s}=\frac{A_n}{(2\pi)^{n/2}}\exp\{-\tfrac{1}{2}(1+\textstyle\sum_1^r t_j^2)s^2\}((1+\sum_1^r t_j^2)^{\frac{1}{2}}s)^{n-1}(1+\sum_1^r t_j^2)^{\frac{1}{2}}\,ds\,;$$

conditionally, the variable $(1+\sum_1^r t_j^2)^{\frac{1}{2}}s$ is chi with n degrees of freedom. The distribution of t_1 given $t_{[2]}$ and $\mathbf{d}$ is

$$f^1(t_1:t_{[2]},\mathbf{d})\,dt_1=\frac{A_{n-1}}{A_n}\left(1+\frac{t_1^2}{1+\sum_2^r t_j^2}\right)^{-(n-1)/2}\cdot\frac{dt_1}{(1+\sum_2^r t_j^2)^{\frac{1}{2}}}\,;$$

conditionally, the variable $t_1/(1+\sum_2^r t_j^2)^{\frac{1}{2}}$ is canonical[1] t with $n-1$ degrees of freedom. The distribution of intermediate t's follows in the same way. Finally the distribution of t_r given $\mathbf{d}$ is

$$f^r(t_r:\mathbf{d})\,dt_r=\frac{A_{n-r}}{A_{n-r+1}}(1+t_r^2)^{-(n-r+1)/2}\,dt_r\,;$$

conditionally, the variable t_r is canonical t with $n-r$ degrees of freedom.

The significance of a value for β_r, the likelihood for β_r, on the objective posterior β_r are then obtained from

$$t_r=\frac{b_r(\mathbf{y})-\beta_r}{s(\mathbf{y})}$$

with t_r treated as a canonical t variable with $n-r$ degrees of freedom.

Then given β_r the significance of a value for β_{r-1}, the likelihood for β_{r-1}, or the objective posterior for β_{r-1} are obtained from

$$t_{r-1}=\frac{b_{r-1}(\mathbf{y})-\beta_{r-1}}{s(\mathbf{y})}$$

with $t_{r-1}/(1+t_r^2)^{\frac{1}{2}}$ treated as a canonical t-variable with $n-r+1$ degrees of freedom; note that the scaling factor $(1+t_r^2)^{\frac{1}{2}}$ involves the given value for β_r (i.e. a pooled residual).

Then given $\beta_r,\cdots,\beta_2$, the assessment of β_1 is obtained from

$$t_1=\frac{b_1(\mathbf{y})-\beta_1}{s(\mathbf{y})}$$

with $t_1/(1+\sum_2^r t_j^2)^{\frac{1}{2}}$ treated as a canonical t-variable with $n-1$ degrees of freedom; note that the scaling factor $(1+\sum_2^r t_j^2)^{\frac{1}{2}}$ involves the given values for $\beta_r,\cdots,\beta_2$.

Finally given $\beta_r,\cdots,\beta_1$ the assessment of σ is obtained from

$$s=\frac{s(\mathbf{y})}{\sigma}$$

with $s(1+\sum_1^r t_j^2)^{\frac{1}{2}}$ treated as a chi variable with n degrees of freedom.

[1] A normal variable divided by a chi-variable; this avoids the square root of the degrees of freedom in the usual expressions.

8. Addendum: Haar factorization. Consider a semidirect product $G = H_1 H_2$ where H_1 and H_2 are continuous subgroups of G. Let μ_1 and Δ_1 be the standardized left Haar and modular functions for H_1, μ_{12} and Δ_{12} be the standardized left Haar and modular function for $H_1 H_2$, and μ_2 and Δ_2 be the standardized (re-Euclidean volume orthogonal to H_1 at e in $H_1 H_2$) left Haar and modular function for H_2. Then

$$\frac{d\mu_{12}(h_1 h_2)}{\Delta_{12}(h_2)} = d\mu_1(h_1) \frac{d\mu_2(h_2)}{\Delta_2(h_2)}$$

is the standardized left (H_1) — right (H_2) invariant Haar for G. Thus the left Haar factors

$$\begin{aligned} d\mu_{12}(h_1 h_2) &= d\mu_1(h_1) \cdot \frac{\Delta_{12}(h_2)}{\Delta_2(h_2)} d\mu_2(h_2) \\ &= d\mu_1(h_1) \cdot d\mu_{[2]}(h_2), \end{aligned}$$

where the support measure $\mu_{[2]}$ is the quotient of left Haar on $H_1 H_2$ by left Haar on H_1

$$d\mu_{[2]}(h_2) = \frac{d\mu(h_1 h_2)}{d\mu_1(h_1)} = \frac{\Delta_{12}(h_2)}{\Delta_2(h_2)} d\mu_2(h_2).$$

Now consider the semidirect product $G = H_1 \cdots H_r$ as discussed in Section 4. Let $g = h_{(r)} = h_1 \cdots h_r$, $h_{(s)} = h_1 \cdots h_s$, $\mu_{(s)} = \mu_{1\ldots s}$, $\Delta_{(s)} = \Delta_{1\ldots s}$. Then applying the preceding factorization recursively (splitting factors from the right), we have

$$d\mu(g) = d\mu_1(h_1) \cdot \frac{\Delta_{(2)}(h_2)}{\Delta_2(h_2)} d\mu_2(h_2) \cdot \cdots \cdot \frac{\Delta_{(r)}(h_r)}{\Delta_r(h_r)} d\mu_r(h_r)$$

where the cumulative products from the left are the left Haar measures for the corresponding cumulative product groups and the component Haar measures are standardized in terms of Euclidean volume orthogonal to preceding coordinates at the identity.

Acknowledgment. The authors wish to express their appreciation to a referee who provided very helpful incisive and fruitful comments on an earlier version of this paper. The paper was prepared under the partial support of National Research Council of Canada grant A3011.

REFERENCES

[1] FRASER, D. A. S. (1966). Structural probability and a generalization. *Biometrika* **53** 1–9.

[2] FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.

[3] FRASER, D. A. S. (1972). The determination of likelihood and the transformed regression model. *Ann. Math. Statist.* **43** 898–916.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
TORONTO, 181, CANADA

DEPARTMENT OF STATISTICS
UNIVERSITY OF MANITOBA
WINNIPEG, CANADA