The Annals of Mathematical Statistics
1972, Vol. 43, No. 3, 898-916

THE DETERMINATION OF LIKELIHOOD AND THE TRANSFORMED REGRESSION MODEL

BY D.A.S FRASER

University of Hawaii and University of Toronto

0. Survey and Summary. The traditional model of statistics is a class of probability measures for a response variable. Under reasonable continuity this can be given as a class C of probability density functions relative to an atom-free measure. With a realized value of the response variable, the model C gives the possible probabilities for that realized value—it gives the likelihood function. The likelihood function can be accepted alone or in conjunction with the distribution of possible likelihood functions.

In a variety of applications, the variation in a response variable can be traced to a well-defined source having a known probability distribution. The model then is not a class of probability measures but is a single probability measure and a class of random variables. Under moderate conditions this can be given as a probability density function and a class C_2 of transformations from the variation space to the response space. And if the distribution for variation is not completely known, the model becomes a class C_1 of probability density functions and a class C_2 of transformations from the variation space to the response space. With an observed response value, the component C_2 identifies a set, the set of possible values for the realized variation. If C_2 is a transformation group, then C_2 identifies a set—*in a partition on the variation space*. Standard probability argument using C_1 then gives the probability of what has been "observed," and the conditional distribution of what has not been "observed": it gives the likelihood function from the identified set, and the conditional density within the identified set. The likelihood function alone or with its distribution gives the information concerning the parameter of C_1; and for any assumed value of that parameter the conditional density gives the information concerning possible values for the realized variation, and accordingly gives the information concerning the parameter of C_2, it being *what stands between the realized variation and the observed response*.

The probability of what is identified as having occurred—the likelihood function—is a fundamental output of a model involving density functions. The determination of this probability can however involve certain complexities as soon as the class C_2 of random variables is no longer effectively a group. Certainly the class C_2 identifies a set on the variation space. But in moderately general cases the range of alternatives can be a partition on the variation space depends on the element of C_2. Thus an 'event' is identified but the range of possible 'events' depends on the parameter of C_2. For two kinds of generalized model (C_1, C_2) this paper explores the determination of *the probability of what*

Received January 6, 1970; revised November 3, 1971.

is identified as having occurred—it explores the determination of the likelihood function.

In Section 1 the notation and results are summarized for the special model (C_1, C_2) with C_2 a transformation group. Two generalizations are examined in Section 2: first, the class C_2 is a group but its application as a transformation group has an additional parameter; second, the class C_2 is a class of *expression* transformations L applied to a group of transformations G, i.e. $C_2 = LG$. These two generalizations are not as distinct as they may at first appear but they are quite distinct in contexts. The transformed regression model is the central example.

Several formulas for volume change in subspaces are recorded in Section 3 and used in Section 4 to make four determinations of likelihood for the generalized model (C_1, C_2). These are applied to the transformed regression model in Section 5 and compared by means of examples in Section 6.

The effects of initial variable on the likelihood functions is examined in Section 7 and two compensating routes for analysis are proposed.

The class L of expression transformations is examined in Section 8 and shown to be a group under mild consistency conditions. A corresponding invariant likelihood is determined in Section 9, and a transit likelihood in Section 10; the power-transformed regression model is examined in Section 11. In Section 12 the transit likelihood is shown to be the natural likelihood when the semi-direct product LG is itself a group.

1. Introduction. For the traditional model of statistics in the continuous case, let: $\mathscr{Y}$ be the space of values for the response Y, an open set in the cartesian product $\mathbb{R}^N$; Φ be the space of values for the quantity ϕ under investigation; and $C = \{f(Y: \phi)\, dY: \phi \in \Phi\}$ be the class of probability distributions for Y where f is continuous in its first argument and dY is Lebesgue measure.

For a realized value Y_0 the model C gives the possible probabilities for that realized value: $f(Y_0: \phi)m$ with m unspecified positive. The model thus gives $L(Y_0: \cdot) = \mathbb{R}^+(Y_0)f(Y_0: \cdot)$ where $\mathbb{R}^+(\cdot): \mathscr{Y} \to \{\mathbb{R}^+\}$ is the constant map from $\mathscr{Y}$ to the single image, the set of positive real numbers $\mathbb{R}^+ = (0, \infty)$.

As notation for the variation-response model in the continuous case, let $\mathscr{U}$ be the space of values for the variation U and $\mathscr{Y}$ be the space of values for the response Y where $\mathscr{Y} = \mathscr{U}$ is an open set in $\mathbb{R}^N$, $C_1 = \{p(U: \rho)\, dU: \rho \in P\}$ be the class of probability distributions for U, and $C_2 = \{Y = \theta U: \theta \in G\}$ be the class of random variables, a transformation group from $\mathscr{U}$ to $\mathscr{Y}$ where the group G is an open set in $\mathbb{R}^Q$. For regularity suppose $\tilde{Y} = hgY$ and $\tilde{g} = hg$ are continuously differentiable re h, g, Y, and suppose there is a continuously differentiable map $[\cdot]: \mathscr{Y} = G$ such that $[gY] = g[Y]$ for all g, Y.

By omission, a variation-response model can generate a traditional model:

$$C = \{p(\theta^{-1}Y: \rho)\, d\theta^{-1}Y = p(\theta^{-1}Y: \rho)J_N(\theta^{-1}Y)J_N^{-1}(Y)\, dY: \theta \in G, \rho \in P\}$$

where $J_N(Y)$ is the Jacobian determinant $J_N([Y]:X) = |\partial[Y]X/\partial X|$ with $X = [Y]^{-1}Y$.

With an observed response Y_0 the component C_2 determines the set $\{g^{-1}Y_0: g \in G\} = \{gY_0: g \in G\} = GY_0$ of possible value for the realized variation U_0; the identified set GY_0 is an element of the partition $\{GU: U \in \mathscr{U}\}$ of $\mathscr{U}$ into orbits GU. The conditional distribution describing the realized variation U_0 can be obtained directly by variable change and normalization,

$$g([U]: D_0, \rho)d[U] = k^{-1}(D_0, \rho)p([U]D_0: \rho)J_N([U]D_0)J_Q^{-1}([U])d[U] ,$$

where $D_0 = [Y_0]^{-1}Y_0$, and $J_Q(g)$ is the Jacobian determinant $|\partial gh/\partial h|$ with h equal to the identity i; for given ρ this distribution gives the information concerning the value of the parameter θ. The probability for the identified set GY_0 can be obtained by division,

$$\frac{p(U:\rho)\,dU}{g([U]: D_0, \rho)d[U]} = k(D_0, \rho)\frac{J_Q([U])}{J_N([U]D_0)} \cdot \frac{dU}{d[U]}$$

using a cross-orbit measure at U, or

$$k(D_0, \rho)\frac{J_Q([Y_0])}{J_N([Y_0]D_0)} \cdot \frac{dY_0}{d[Y_0]}$$

using a cross-orbit measure at Y_0; the likelihood function for the identified set is then

$$L_1(D_0:\rho) = \mathbb{R}^+(D_0)k(D_0, \rho)$$

which can be examined alone or with its distribution to obtain the information concerning ρ. For further details see Fraser (1968).

As an example consider the variation-response model for regression. Let $\mathbf{y}' = (y_1, \cdots, y_n)$ be the vector of response observations, let $\mathbf{u}' = (u_1, \cdots, u_n)$ be the corresponding vector of variables for variation, let V with r row vectors $\mathbf{v}_u'(u = 1, \cdots, r)$ be the basis matrix for the r-dimensional response-level space $\mathscr{L}(V)$, and let $\Pi f(u_i:\rho)\,d\mathbf{u}$ be the distribution describing the variation. For notation amenable to matrix multiplication, let

$$Y = \begin{pmatrix} V \\ \mathbf{y}' \end{pmatrix}, \qquad U = \begin{pmatrix} V \\ \mathbf{u}' \end{pmatrix}, \qquad \theta = \begin{pmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{pmatrix}$$

be matrix labels for the response $\mathbf{y}$, the variation $\mathbf{u}$, and the transformation with regression coefficients $\boldsymbol{\beta}' = (\beta_1, \cdots, \beta_r)$ and variation scaling σ. The model then is (C_1, C_2) where

$$C_1 = \{\textstyle\prod_1^n p(u_i:\rho)\,d\mathbf{u}: \rho \in P\} , \qquad C_2 = \{Y = \theta U: \boldsymbol{\beta} \in \mathbb{R}^r, \sigma \in \mathbb{R}^+\} .$$

2. Two generalizations. As a first generalization let G be a group and suppose its application as a transformation group G_κ from $\mathscr{U}$ to $\mathscr{Y}(=\mathscr{U})$ involves a parameter κ. For an element θ in G let θ_κ designate the corresponding transformation in G_κ. The generalized model is then (C_1, C_2) where

$$C_1 = \{p(U:\rho)\,dU: \rho \in P\} , \qquad C_2 = \{Y = \theta_\kappa U: \theta \in G, \kappa \in K\} .$$

As an example consider the regression model but supose now that the vectors in the response level matrix depend on a parameter κ. This could arise for example if there was doubt as to the natural form of expression for an input variable. For matrix notation let

$$Y = \begin{pmatrix} V_\kappa \\ \mathbf{y}' \end{pmatrix}, \qquad U = \begin{pmatrix} V_\kappa \\ \mathbf{u}' \end{pmatrix}, \qquad \theta = \begin{pmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{pmatrix}$$

but note that Y and U as response and variation do *not* depend on κ whereas θ as a transformation θ_κ does depend on κ as the matrix multiplication shows. The generalized model is then (C_1, C_2) where

$$C_1 = \{\Pi p(u_i : \rho)\, d\mathbf{u} : \rho \in P\}\,, \qquad C_2 = \{Y = \theta_\kappa U : \boldsymbol{\beta} \in \mathbb{R}^r, \sigma \in \mathbb{R}^+, \kappa \in K\}\,.$$

As a second generalization suppose that the *natural* response can be described by a variation-response model with θ in a group G, but that the recorded response is some transformation λ of the natural response. The model is then (C_1, C_2) where

$$C_1 = \{p(U : \rho)\, dU : \rho \in P\}\,, \qquad C_2 = \{Y = \lambda\theta U : \theta \in G, \lambda \in L\}\,.$$

As an example consider the regression model but suppose that the natural response variable, which has the additive form in terms of input variables, is some transform $l(y, \lambda) = \lambda^{-1}y$ of the given response variable. For matrix notation let

$$\lambda^{-1} Y = \begin{pmatrix} V \\ \lambda^{-1}\mathbf{y}' \end{pmatrix}, \qquad U = \begin{pmatrix} V \\ \mathbf{u}' \end{pmatrix}, \qquad \theta = \begin{pmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{pmatrix}.$$

The generalized model is then (C_1, C_2) where

$$C_1 = \{\Pi p(u_i : \rho)\, d\mathbf{u} : \rho \in P\}\,, \qquad C_2 = \{Y = \lambda\theta U : \boldsymbol{\beta} \in \mathbb{R}^r, \sigma \in \mathbb{R}^+, \lambda \in L\}\,.$$

The second generalization can be treated formally as a special case of the first generalization. Let

$$\theta_\lambda = \lambda\theta\lambda^{-1}\,, \qquad U^* = \lambda U$$

and $p(U^* : \rho, \lambda)\, dU^*$ be the distribution of U^*. The second generalization is then (C_1, C_2) where

$$C_1 = \{p(U^* : \rho, \lambda)\, dU^* : \rho \in P, \lambda \in L\}\,, \qquad C_2 = \{Y = \theta_\lambda U^* : \theta \in G, \lambda \in L\}\,;$$

note that a parameter of C_1 is also a parameter of C_2. This alternate form suppresses any explicit recognition of the natural response space, the space on which the group G operates.

The two generalizations can be compounded giving (C_1, C_2) where

$$C_1 = \{p(U : \rho)\, dU : \rho \in P\}\,, \qquad C_2 = \{Y = \lambda\theta_\kappa U : \theta \in G, \kappa \in K, \lambda \in L\}$$

which can be abbreviated

$$C_1 = \{p(U : \lambda)\, dU : \lambda \in \Lambda\}\,, \qquad C_2 = \{Y = \lambda\theta_\lambda U : \theta \in G, \lambda \in \Lambda\}$$

where λ now embraces the three non-group parameters ρ, κ, λ in the original expressions.

3. Volume change in subspaces. Consider a Q-dimensional subspace $\mathscr{L}(M)$ of $\mathbb{R}^N$ when M is a basis matrix of row vectors. A point $\mathbf{x}$ in $\mathscr{L}(M)$ can be represented by $\mathbf{b}$ in terms of the basis $M : \mathbf{x}' = \mathbf{b}'M$. This maps $\mathbf{x}$ in $\mathscr{L}(M)$ into $\mathbf{b}$ in $\mathbb{R}^Q$. Alternatively a point $\mathbf{x}$ in $\mathscr{L}(M)$ can be represented by means of the Q-dimensional linear form $\mathbf{l} = M\mathbf{x}$; this maps $\mathbf{x}$ (here in $\mathscr{L}(M)$) into $\mathbf{l}$ in $\mathbb{R}^Q$. Consider volume change under these two maps.

For the first map the formula $d\mathbf{x} = |M|_+ d\mathbf{b} = |MM'|^{\frac{1}{2}} d\mathbf{b}$ follows trivially if $N = Q$. The last expression uses the inner product matrix MM' for the basis vectors; in terms of the inner product matrix the formula is independent of the embedding Euclidean space $\mathbb{R}^N$; hence

$$d\mathbf{x} = |MM'|^{\frac{1}{2}}\, d\mathbf{b}$$

where $d\mathbf{x}$ expresses Euclidean volume in $\mathscr{L}(M)$.

The second map can be related to the first: $\mathbf{l} = M\mathbf{x} = M(\mathbf{b}'M)' = MM'\mathbf{b}$. Hence $\mathbf{x}' = \mathbf{b}'M = l'(MM')^{-1}M$ and it follows that $\mathbf{l}$ provides coordinates with respect to the basis matrix $(MM')^{-1}M$. The formula for volume change is then

$$\begin{aligned} d\mathbf{x} &= |(MM')^{-1}MM'(MM')^{-1}|^{\frac{1}{2}}\, d\mathbf{l} \\ &= |MM'|^{-\frac{1}{2}}\, d\mathbf{l}\,. \end{aligned}$$

4. Four determinations of likelihood. Consider the generalized variation-response model (C_1, C_2) as defined in Section 2:

$$C_1 = \{p(U : \lambda)\, dU : \lambda \in \Lambda\}\,, \qquad C_2 = \{Y = \lambda\theta_\lambda U : \theta \in G,\ \lambda \in \Lambda\}\,.$$

For given λ this model reduces effectively to the standard variation-response model in Section 1. An observed value Y on the given response space determines the transformed response $Y_\lambda = \lambda^{-1} Y$ on the natural response space; and by the argument in Section 1 this identifies the orbit $G_\lambda U = G_\lambda Y_\lambda$ of possible values for the realized variation U. The conditional distribution describing U is given by

$$\begin{aligned} &g([U] : D_\lambda, \lambda)d[U] \\ &\quad = k^{-1}(D_\lambda, \lambda)p([U]_\lambda D_\lambda : \lambda)J_N([U]_\lambda : D_\lambda)J_Q^{-1}([U]_\lambda)d[U]_\lambda \end{aligned}$$

where $D_\lambda = [Y_\lambda]^{-1} Y_\lambda$. For given λ this distribution describes the realized U in the equation $Y_\lambda = \theta_\lambda U$ and hence gives the information concerning the value of θ.

Now consider the probability of what is identified as having occurred. The observed response Y identifies the orbit $G_\lambda U = G_\lambda Y_\lambda$ on the variation space or equivalently identifies the pre-orbit $\lambda G_\lambda U = \lambda G_\lambda Y_\lambda$ on the given response space. Thus the observed Y is equivalent to the observation $\lambda G_\lambda Y_\lambda$ on the function λG_λ defined on $\mathscr{U}$. The probability for the identified orbit can be obtained by division:

$$\frac{p(U:\lambda)\,dU}{g([U]_\lambda:D_\lambda,\lambda)d[U]_\lambda} = k(D_\lambda,\lambda)\frac{J_Q([U]_\lambda)}{J_N([U]_\lambda:D_\lambda)}\cdot\frac{dU}{d[U]_\lambda}$$

using a cross-orbit measure at U, or

$$k(D_\lambda,\lambda)\frac{J_Q([Y_\lambda]_\lambda)}{J_N([Y_\lambda]_\lambda:D_\lambda)}\cdot\frac{dY_\lambda}{d[Y_\lambda]_\lambda}$$

using a cross-orbit measure at Y_λ. Likelihood can then be obtained by separating the λ-dependence from the cross-orbit measure.

(4.1) The volume differential dY_λ on the natural response space can be expressed in terms of dY on the given response space

$$dY_\lambda = \left|\frac{\partial\lambda^{-1}Y}{\partial Y}\right| dY = |J(\lambda^{-1}:Y)|\,dY\,.$$

The volume differential $d[Y_\lambda]_\lambda$ on the group G can be expressed in terms of volume on the Q dimensional orbit at Y_λ and in terms of volume on the Q dimensional pre-orbit at Y; these last two volumes will be defined if $\mathbb{R}^N$ is given, say, the Euclidean distance. In terms of vector differentials

$$dY = J^{-1}(\lambda^{-1}:Y)\,dY_\lambda = J^{-1}(\lambda^{-1}:Y)W_\lambda(Y_\lambda)d[Y_\lambda]_\lambda$$

where

$$W_\lambda(Y_\lambda) = \frac{\partial[Y_\lambda]_\lambda D_\lambda}{\partial[Y_\lambda]_\lambda}$$

is an $N\times Q$ Jacobian matrix; hence by Section 3

$$d[Y_\lambda]_\lambda = |(--)'(J^{-1}(\lambda^{-1}:Y)W_\lambda(Y_\lambda))|^{-\frac{1}{2}}\,dY\,.$$

The quotient of volume at Y by Euclidean volume on the pre-orbit through Y is Euclidean volume dv_1 in the orthogonal complement to the pre-orbit at Y. Thus the probability for the event identified by Y is

$$k(D_\lambda,\lambda)\frac{J_Q([Y_\lambda]_\lambda)}{J_N([Y_\lambda]_\lambda:D_\lambda)}\cdot\frac{|J(\lambda^{-1}:Y)|}{|(--)'(J^{-1}(\lambda^{-1}:Y)W_\lambda(Y_\lambda))|^{-\frac{1}{2}}}\,dv_1\,;$$

and it has *orthogonal* likelihood

$$L_1 = \frac{\mathbb{R}^+(D_\lambda)k(D_\lambda,\lambda)J_Q([Y_\lambda]_\lambda)\,|J(\lambda^{-1}:Y)|}{J_N([Y_\lambda]_\lambda:D_\lambda\,|(--)'(J^{-1}(\lambda^{-1}:Y)W_\lambda(Y_\lambda))|^{-\frac{1}{2}}}\,.$$

This is the marginal likelihood as used for the regression model (Fraser (1967)) and as obtained in general (Fraser (1968)).

(4.2) In some applications there may be natural sections to orbits on the variation space, sections that are preserved under the group G_λ; a section might join points having a given position in the conditional distributions. For convenience in formulas suppose that the base points $D_\lambda(U)$ form such a section.

Then $[U]_\lambda =$ constant describes a general section on $\mathscr{U}$, $[Y_\lambda]_\lambda =$ constant describes the corresponding section at Y_λ on the natural space, and $[\lambda^{-1} Y]_\lambda =$ constant describes the corresponding section at Y on the given space. This last section can be described alternatively by orthogonality to the row vectors in the $Q \times N$ matrix

$$\frac{\partial [Y_\lambda]_\lambda}{\partial Y} = \frac{\partial [Y_\lambda]_\lambda}{\partial Y_\lambda} \frac{\partial X_\lambda}{\partial Y} = M_\lambda(Y_\lambda) J(\lambda^{-1} : Y) .$$

where

$$M_\lambda(Y_\lambda) = \frac{\partial [Y_\lambda]_\lambda}{\partial Y}$$

is a $Q \times N$ Jacobian matrix. Let v_* describe Euclidean volume orthogonal to the section:

$$dv_* = |(M_\lambda(Y_\lambda)J(\lambda^{-1} : Y))(--)'|^{-\frac{1}{2}} \, d[Y_\lambda]_\lambda .$$

Then the probability for the event identified by Y is

$$k(D_\lambda \lambda) \frac{J_Q([Y_\lambda]_\lambda)}{J_N([Y_\lambda]_\lambda : D_\lambda)} \frac{|J(\lambda^{-1} : Y)|}{|(M_\lambda(Y_\lambda)J(\lambda^{-1} : Y))(--)'|^{\frac{1}{2}}} \, dv_2$$

where $dv_2 = dY/dv_*$ is volume in the section at Y. The corresponding *section* likelihood is

$$L_2 = \frac{\mathbb{R}^+(D_\lambda)k(D_\lambda, \lambda)J_Q([Y_\lambda]_\lambda) \, |J(\lambda^{-1} : Y)|}{J_N([Y_\lambda]_\lambda : D_\lambda) \, |(M_\lambda(Y_\lambda)J(\lambda^{-1} : Y))(--)'|^{\frac{1}{2}}} .$$

This likelihood incorporates effects due to shearing when orbits on the natural space are mapped to the pre-orbits on the given space. For the normal regression model this likelihood coincides with a conditional likelihood given a sufficient statistic as developed by Sprott and Kalbfleisch.

(4.3) As in (4.2) suppose there are natural sections to orbits on the variation space, sections that are preserved under the group G_λ. And in addition suppose the application of G to the variation space does not involve $\lambda : G_\lambda = G$. For notational convenience let the base points $D_\lambda(U)$ form a natural section on $\mathscr{U}$; then other sections are given by $[Y_\lambda] =$ constant. With no information concerning U on its orbit GU in $\mathscr{U}$, replace the conditional distribution on the orbit by a uniform distribution relative to the invariant differential: $cd\mu[U] = cJ_Q^{-1}([U])d[U] = cJ_Q^{-1}([Y_\lambda])d[Y_\lambda]$; this is feasible since the sections interrelate the orbits. The probability differential at Y_λ is then

$$k(D_\lambda, \lambda) \frac{J_Q([Y_\lambda])}{J_N([Y_\lambda] : D_\lambda)} \cdot \frac{dY_\lambda}{d[Y_\lambda]} \cdot cJ_Q^{-1}([Y_\lambda])d[Y_\lambda] ,$$

and at Y is then

$$k(D_\lambda, \lambda) \frac{c \, |J(\lambda^{-1} : Y)|}{J_N([Y_\lambda] : D_\lambda)} \cdot dY .$$

The corresponding *fibre* likelihood is

$$L_3 = \frac{\mathbb{R}^+(D_\lambda)k(D_\lambda, \lambda)\, |J(\lambda^{-1} : Y)|}{J_N([Y_\lambda] : D_\lambda)} .$$

(4.4) A fourth likelihood function can be constructed by customary likelihood methods applied to the traditional model. The traditional model corresponding to (C_1, C_2) is

$$C = \left\{p(\theta_\lambda^{-1} Y_\lambda : \lambda) \frac{J_N(\theta_\lambda^{-1}[Y_\lambda]_\lambda : D_\lambda)}{J_N([Y_\lambda]_\lambda : D_\lambda)} |J(\lambda^{-1} : Y)|\, dY : \theta \in G,\ \lambda \in \Lambda\right\}$$

and the likelihood for Y is

$$\mathbb{R}^+(Y)p(\theta_\lambda^{-1} Y_\lambda : \lambda) \frac{J_N(\theta_\lambda^{-1}[Y_\lambda]_\lambda : D_\lambda)}{J_N([Y_\lambda]_\lambda : D_\lambda)} |J(\lambda^{-1} : Y)| .$$

The *profile* likelihood for λ is obtained by maximizing across λ sections on the likelihood domain:

$$L_4 = \mathbb{R}^+(Y)p(g_\lambda D_\lambda : \lambda) \frac{J_N(g_\lambda : D_\lambda)}{J_N([Y_\lambda]_\lambda : D_\lambda)} |J(\lambda^{-1} : Y)|$$

where g_λ maximizes

$$p(gD_\lambda : \lambda)J_N(g : D_\lambda) .$$

5. The transformed regression model. Consider the regression model as generalized in two ways in Section 2:

$$C_1 = \{\Pi p(u_i : \lambda)\, d\mathbf{u} : \lambda \in \Lambda\} , \qquad C_2 = \{Y = \lambda\theta_\lambda U : \boldsymbol{\beta} \in \mathbb{R}^r, \sigma \in \mathbb{R}^+, \lambda \in \Lambda\}$$

where

$$Y_\lambda = \begin{pmatrix} V_\lambda \\ \mathbf{y}_\lambda' \end{pmatrix}, \qquad U = \begin{pmatrix} V_\lambda \\ \mathbf{u}' \end{pmatrix}, \qquad \theta = \begin{pmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{pmatrix},$$

and $y^\lambda = l(y : \lambda) = \lambda^{-1}y$ for each response coordinate. For notation let $\mathbf{b}_\lambda(\mathbf{u})$, $s_\lambda(\mathbf{u})$, and $\mathbf{d}_\lambda(\mathbf{u})$ be the regression coefficients, residual length, and unit residual for $\mathbf{u}$ on $\mathscr{L}(V_\lambda)$; let

$$[U]_\lambda = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{b}_\lambda'(\mathbf{u}) & s(\mathbf{u}) \end{pmatrix}, \qquad D_\lambda(U) = [E]_\lambda^{-1}E = \begin{pmatrix} V_\lambda \\ \mathbf{d}_\lambda'(\mathbf{u}) \end{pmatrix};$$

and let

$$J(\lambda^{-1} : \mathbf{y}) = \frac{\partial\lambda^{-1}\mathbf{y}}{\partial\mathbf{y}'} = \begin{pmatrix} \dfrac{dy_1^\lambda}{dy_1} & & 0 \\ & \ddots & \\ 0 & & \dfrac{dy_n^\lambda}{dy_n} \end{pmatrix}.$$

For additional details such as the calculation of the Jacobian determinants J_N and J_Q see Fraser (1968). The likelihood functions of Section 4 require matrices

$W_\lambda(Y)$ and $M_\lambda(Y)$:

$$\mathbf{y}_\lambda = D_\lambda'(Y_\lambda)\begin{pmatrix}\mathbf{b}(\mathbf{y}_\lambda)\\ s(\mathbf{y}_\lambda)\end{pmatrix},$$

$$W_\lambda(Y) = \frac{\partial \mathbf{y}_\lambda}{\partial(\mathbf{b}', s)} = D_\lambda'(Y_\lambda);$$

$$\begin{pmatrix}\mathbf{b}_\lambda(\mathbf{y}_\lambda)\\ s(y_\lambda)\end{pmatrix} = (D_\lambda(Y_\lambda)D_\lambda'(Y_\lambda))^{-1}D_\lambda(Y_\lambda)\mathbf{y}_\lambda,$$

$$M_\lambda(Y) = \frac{\partial\begin{pmatrix}\mathbf{b}\\ s\end{pmatrix}}{\partial \mathbf{y}_\lambda'} = (D_\lambda D_\lambda')^{-1}D_\lambda .$$

The four likelihood functions from Section 4 can now be calculated. For the second and third likelihoods, orthogonal (least-squares) sections are used on the natural response space. And for the third likelihood, $V_\lambda = V$. The four likelihood functions are

$$L_1 = \mathbb{R}^+ \frac{k(D_\lambda, \lambda)\,|J(\lambda^{-1}:\mathbf{y})|}{s_\lambda^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda J^{-2}(\lambda^{-1}:\mathbf{y})D_\lambda'|^{-\frac{1}{2}}},$$

$$L_2 = \mathbb{R}^+ \frac{k(D_\lambda, \lambda)\,|J(\lambda^{-1}:\mathbf{y})|}{s_\lambda^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda D_\lambda'|^{-1}\,|D_\lambda J^{2}(\lambda^{-1}:\mathbf{y})D_\lambda'|^{\frac{1}{2}}},$$

$$L_3 = \mathbb{R}^+ \frac{k(D_\lambda, \lambda)\,|J(\lambda^{-1}:\mathbf{y})|}{s^n(\mathbf{y}_\lambda)}$$

$$L_4 = \mathbb{R}^+ \frac{|J(\lambda^{-1}:\mathbf{y})|\,\sup_{ac} c^n \prod_1^n f(\Sigma a_u v_{ui} + c d_i^\lambda : \lambda)}{s_\lambda^n(\mathbf{y}_\lambda)} .$$

For normal variation the normalizing constant $k(D_\lambda, \lambda)$ is given by

$$k^{-1}(D_\lambda, \lambda) = A_{n-r}\,|V_\lambda V_\lambda'|^{\frac{1}{2}} = A_{n-r}\,|D_\lambda D_\lambda'|^{\frac{1}{2}}$$

where $A_f = 2\pi^{f/2}\Gamma(f/2)$ is the area of the unit sphere in $\mathbb{R}^f$. The four likelihood functions are then

$$L_1 = \mathbb{R}^+ \frac{|J(\lambda^{-1}:\mathbf{y})|}{|D_\lambda D_\lambda'|^{\frac{1}{2}} s_\lambda^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda J^{-2}(\lambda^{-1}:\mathbf{y})D_\lambda'|^{-\frac{1}{2}}},$$

$$L_2 = \mathbb{R}^+ \frac{|J(\lambda^{-1}:\mathbf{y})|}{|D_\lambda D_\lambda'|^{-\frac{1}{2}}\, s_\lambda^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda J^{2}(\lambda^{-1}:\mathbf{y})D_\lambda'|^{\frac{1}{2}}},$$

$$L_3 = \mathbb{R}^+ \frac{|J(\lambda^{-1}:\mathbf{y})|}{s^n(\mathbf{y}_\lambda)},$$

$$L_4 = \mathbb{R}^+ \frac{|J(\lambda^{-1}:\mathbf{y})|}{s_\lambda^n(\mathbf{y}_\lambda)} .$$

Consider now the regression model but with known scaling for the variation:

$$C_1 = \{\Pi f(u_i:\lambda)\,d\mathbf{u} : \lambda \in \Lambda\}, \qquad C_2 = \{Y = \lambda\theta_\lambda U : \boldsymbol{\beta} \in \mathbb{R}^r, \lambda \in \Lambda\}$$

where

$$Y_\lambda = \begin{pmatrix} V_\lambda \\ \mathbf{y}_\lambda' \end{pmatrix}, \qquad U = \begin{pmatrix} V_\lambda \\ \mathbf{u}' \end{pmatrix}, \qquad \theta = \begin{pmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & 1 \end{pmatrix},$$

and $y^\lambda = l(y:\lambda) = \lambda^{-1}y$ for each response coordinate. For notation let $\mathbf{b}_\lambda(\mathbf{u})$, $\mathbf{d}_\lambda(\mathbf{u})$ be the regression coefficients and residual vector for $\mathbf{y}$ on $\mathscr{L}(V_\lambda)$; let

$$[U]_\lambda = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{b}_\lambda'(\mathbf{u}) & 1 \end{pmatrix}, \qquad D_\lambda(U) = [U]_\lambda U = \begin{pmatrix} V_\lambda \\ \mathbf{d}_\lambda'(\mathbf{u}) \end{pmatrix},$$

The required matrices are

$$W_\lambda(Y) = J^{-1}(\lambda^{-1}:\mathbf{y})V_\lambda', \qquad M_\lambda(Y) = (V_\lambda V_\lambda')^{-1}V_\lambda J(\lambda^{-1}:\mathbf{y}).$$

Four likelihood functions from Section 4 can now be calculated. For the second and third likelihoods, orthogonal (least-squares) sections are used on the natural response space. And for the third likelihood, $V_\lambda = V$. The four likelihood functions are

$$L_1 = \mathbb{R}^+ \frac{k(D_\lambda, \lambda)\,|J(\lambda^{-1}:\mathbf{y})|}{|V_\lambda J^{-2}(\lambda^{-1}:\mathbf{y})V_\lambda'|^{-\frac{1}{2}}},$$

$$L_2 = \mathbb{R}^+ \frac{k(D_\lambda, \lambda)\,|J(\lambda^{-1}:\mathbf{y})|}{|V_\lambda V_\lambda'|^{-1}\,|V_\lambda J^2(\lambda^{-1}:\mathbf{y})V_\lambda'|^{\frac{1}{2}}},$$

$$L_3 = \mathbb{R}^+ k(D_\lambda, \lambda)\,|J(\lambda^{-1}:\mathbf{y})|,$$

$$L_4 = \mathbb{R}^+ |J(\lambda^{-1}:\mathbf{y})| \sup_{\mathbf{a}} \textstyle\prod_1^n f(\Sigma a_u v_{ui} + d_i^\lambda : \lambda).$$

For normal variation with variance σ_0^2, the normalizing constant is given by

$$k(D_\lambda, \lambda) = (V_\lambda V_\lambda')^{-\frac{1}{2}}(2\pi\sigma_0^2)^{-(n-r)/2} \exp\{-|\mathbf{d}_\lambda|^2/2\sigma_0^2\}.$$

The four likelihood functions are then

$$L_1 = \mathbb{R}^+ \frac{\exp\{-|\mathbf{d}_\lambda|^2/2\sigma_0^2\}\,|J(\lambda^{-1}:\mathbf{y})|}{\sigma_0^{n-r}\,|V_\lambda V_\lambda'|^{\frac{1}{2}}\,|V_\lambda J^{-2}(\lambda^{-1}:\mathbf{y})V_\lambda'|^{-\frac{1}{2}}},$$

$$L_2 = \mathbb{R}^+ \frac{\exp\{-|\mathbf{d}_\lambda|^2/2\sigma_0^2\}\,|J(\lambda^{-1}:\mathbf{y})|}{\sigma_0^{n-r}\,|V_\lambda V'|^{-\frac{1}{2}}\,|V_\lambda J^2(\lambda^{-1}:\mathbf{y})V_\lambda'|^{\frac{1}{2}}},$$

$$L_3 = \mathbb{R}^+ \frac{\exp\{-|\mathbf{d}_\lambda|^2/2\sigma_0^2\}\,|J(\lambda^{-1}:\mathbf{y})|}{\sigma_0^{n-r}},$$

$$L_4 = \mathbb{R}^+ \frac{\exp\{-|\mathbf{d}_\lambda|^2/2\sigma_0^2\}\,|J(\lambda^{-1}:\mathbf{y})|}{\sigma_0^{n}}.$$

6. Comparisons by means of examples. Consider the likelihood functions of Section 4 as they apply to a succession of examples that introduce progressively some of the complexities of the general model in Section 2.

EXAMPLE 1. Consider the location model with normal variation having variance σ_0^2,

$$C_1 = \left\{(2\pi\sigma_0^2)^{-n/2} \exp\left\{-\frac{\Sigma u_i^2}{2\sigma_0^2}\right\} d\mathbf{u} : \sigma_0 \in \mathbb{R}^+\right\}, \qquad C_2 = \left\{\mathbf{u} = \mu\mathbf{1} + \mathbf{u} : \begin{matrix}\mu \in \mathbb{R} \\ \sigma_0 \in \mathbb{R}^+\end{matrix}\right\}.$$

For illustrative purposes here, the identifiable variation underlying the scaling σ_0 is not separated out. The likelihood functions are

$$L_1 = L_2 = L_3 = \mathbb{R}^+ \frac{1}{\sigma_0^{n-1}} \exp\left\{-\frac{\Sigma(y_i - \bar{y})^2}{2\sigma_0^2}\right\},$$

$$L_4 = \mathbb{R}^+ \frac{1}{\sigma_0^{n}} \exp\left\{-\frac{\Sigma(y_i - \bar{y})^2}{2\sigma_0^2}\right\}.$$

Each likelihood is based on the response variable $\Sigma(y_i - \bar{y})^2$ which is $\sigma_0^2\chi^2$ on $n - 1$ df. The first three likelihoods are in fact the likelihood functions from such a variable; the fourth likelihood has an additional factor σ_0^{-1} and is not the likelihood from $\Sigma(y_i - \bar{y})^2$. The maximum-likelihood estimate from L_1, L_2, L_3 is the usual sample standard deviation; the maximum-likelihood estimate from L_4 is the traditional embarrassment $(\Sigma(y_i - \bar{y})^2/n)^{\frac{1}{2}}$.

EXAMPLE 2. The essentials of Example 1 but with nonnormal variation:

$$C_1 = \{f(u_1 : \lambda)f(u_2 : \lambda)\, d\mathbf{u} : \lambda \in \Lambda\}, \qquad C_2 = \begin{Bmatrix} y_1 = u_1 & : \mu \in \mathbb{R} \\ y_2 = \mu + u_2 & \lambda \in \Lambda \end{Bmatrix},$$

$$L_1 = L_2 = L_3 = \mathbb{R}^+ f(y_1 : \lambda)$$

$$L_4 = \mathbb{R}^+ f(y_1 : \lambda) f(a_\lambda : \lambda)$$

where a_λ maximizes $f(a : \lambda)$. The profile likelihood L_4 contains an extraneous factor, the modal height of the density along the orbit.

EXAMPLE 3. Example 2 with the addition of an expression parameter

$$C_1 = \{f(u_1 : \lambda)f(u_2 : \lambda)\, d\mathbf{u} : \lambda \in \Lambda\}; \qquad C_2 = \begin{Bmatrix} y_1^\lambda = u_1 & : \mu \in \mathbb{R} \\ y_2^\lambda = \mu + u_2 & \lambda \in \Lambda \end{Bmatrix}$$

where λ indexes a bijective transformation from y^λ to y.

$$L_1 = L_2 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right|,$$

$$L_3 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right| \cdot \left|\frac{dy_2^\lambda}{dy_2}\right|,$$

$$L_4 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right| \cdot \left|\frac{dy_2^\lambda}{dy_2}\right| f(a_\lambda : \lambda).$$

The likelihoods L_1, L_2 are the likelihood from the response variable y_1. The likelihood L_3 has an extraneous factor measuring dilation along the orbit, and L_4 has a further factor, the modal height of the density along the orbit.

EXAMPLE 4. Example 3 with a more complex expression parameter involving shearing

$$C_1 = \{f(u_1 : \lambda)f(u_2 : \lambda)\, d\mathbf{u} : \lambda \in \Lambda\}, \qquad C_2 = \begin{Bmatrix} y_1^\lambda = u_1 & : \mu \in \mathbb{R} \\ y_2^\lambda + \lambda y_1 = \mu + e_2 & \lambda \in \Lambda \end{Bmatrix}.$$

The expression transformation here is not covered by the diagonal Jacobian matrix in Section 5 but a simple generalization gives the needed formulas:

$$L_1 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right| ,$$

$$L_2 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right| \cdot \frac{\left|\frac{dy_2^\lambda}{dy_2}\right|}{\left|\lambda^2 + \left(\frac{dy_2^\lambda}{dy_2}\right)^2\right|^{\frac{1}{2}}} ,$$

$$L_3 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right| \cdot \left|\frac{dy_2^\lambda}{dy_2}\right| ,$$

$$L_4 = \mathbb{R}^+ f(y_1^\lambda : \lambda) \left|\frac{dy_1^\lambda}{dy_1}\right| \cdot \left|\frac{dy_2^\lambda}{dy_2}\right| f(a_\lambda : \lambda) .$$

The likelihood L_1 is the likelihood from the response variable y_1. The remaining likelihoods all have extraneous factors referring to within-orbit properties.

The four examples suggest that L_1 is the preferred likelihood for inference concerning $\lambda : L_1$ is the likelihood function for the observed orbit; the remaining likelihoods have extraneous factors introducing irrelevant aspects of the observed response.

The succession of examples, however, omits one major complexity that can arise with the generalized model in Section 2—dependence of the pre-orbit partition on the parameter λ. Some effects with this complexity are discussed in the next section.

Without this complexity, L_1 is the correct and only likelihood for the identified orbit. This can be seen analytically by noting that the tangent space $\mathscr{L}(W_\lambda'(Y))$ is independent of λ. Typically the other likelihoods have additional factors referring to within-orbit characteristics.

7. The pre-orbit partition with dependence on λ. Consider the more general models that allow orientation of the pre-orbit to depend on the parameter λ.

The analysis in Section 4 uses the Euclidean distance for the response Y. Now consider a new response $X = h(Y)$ generated by a diffeomorphism h; let $K(Y) = \partial Y/\partial X$ be the Jacobian of old with respect to new variable. The analysis of X can be viewed as an analysis of Y but with the Euclidean distance assigned locally to $K^{-1}(Y)\,dY$. The change in likelihood formulas follows by replacing $J(\lambda^{-1} : Y)$ by $J(\lambda^{-1} : Y)K(Y)$ and the new likelihood functions are

$$L_1(X) = \frac{|(--)'(J^{-1}(\lambda^{-1} : Y)W_\lambda(X_\lambda))|^{-\frac{1}{2}}\,|K(Y)|}{|(--)'(K^{-1}(Y)J^{-1}(\lambda^{-1} : Y)W_\lambda(Y_\lambda))|^{-\frac{1}{2}}} L_1(Y) ,$$

$$L_2(X) = \frac{|(M_\lambda(Y_\lambda)J(\lambda^{-1} : Y))(--)'|^{\frac{1}{2}}\,|K(Y)|}{|(M_\lambda(Y_\lambda)J(\lambda^{-1} : Y)K(Y))(--)'|^{\frac{1}{2}}} L_2(Y) ,$$

$$L_3(X) = L_3(Y) ,$$

$$L_4(X) = L_4(Y) .$$

Thus L_3 and L_4 are independent of the initiating variable for analysis. L_1 and

L_2, however, can depend on the initiating variable, can depend on the metric used on the response space. In fact, L_2 can depend on the initiation variable even when the pre-orbit partition is independent of λ. The implied ranking L_1, L_2, L_3, L_4 in Section 6 must now be qualified with the adverse property associated with L_1 and L_2—the possible dependence on the choice of initiating variable. Two simple routes seem open to accommodating L_1 in these new circumstances.

The transformations λ^{-1} applied to the given response variable Y generate the various possibilities $Y_\lambda = \lambda^{-1} Y$ for the natural response variable. Let τ in place of λ be used to index these possible natural response variables. The model $Y = \lambda\theta U$ for the given response, then becomes $Y_\tau = (\tau^{-1}\lambda)\theta U$ for the possible natural response Y_τ; the transformations $\{\lambda : \lambda \in \Lambda\}$ for Y become $\{\tau^{-1}\lambda : \lambda \in \Lambda\}$ for Y_τ. Let $L_1(\lambda : \tau)$ be the likelihood function L_1 for λ calculated with Y_τ as given variable.

As a first route consider the relative likelihood function $L_1(\lambda : \tau)$ defined on $\Lambda \times \Lambda$. Any τ-section assesses other λ values with respect to $\lambda = \tau$.

As a second route consider a practical modification of the preceding: from an initial variable τ_0 calculate the maximum likelihood value λ_0 from $L_1(\lambda : \tau_0)$; from $\tau_1 = \lambda_0$ calculate the maximum likelihood value λ_1 from $L_1(\lambda : \tau_1)$; iterate; assuming stability use the limiting likelihood form. The need for a relative likelihood can be viewed as a sort of non-linearity and the relative likelihood as a local linearization.

8. The expression transformations. Consider the expression transformation used in the second generalization in Section 2:

$$Y_\lambda = l(Y, \lambda) = \lambda^{-1} Y$$

where Y_λ, for some λ, is the natural response variable that has the standard variation-response model. For the regression model the expression transformation was taken to operate coordinate by coordinate. Now for the general model suppose the expression transformation operates coordinate by coordinate $Y_1, \cdots, Y_N$ on Y.

For the regression model Box and Cox (1964) examined two kinds of expression transformation. The first is the power transformation given by

$$\begin{aligned} l(y, \lambda) &= y^\lambda && \lambda \neq 0, \\ &= \ln y && \lambda = 0, \end{aligned}$$

where y^λ here designates the λth power of y. The transformation maps $(0, \infty)$ onto $(0, \infty)$ for $\lambda \neq 0$ and onto $(-\infty, \infty)$ for $\lambda = 0$. A modified form of the power transformation

$$\begin{aligned} l(y, \lambda) &= \lambda^{-1}(y^\lambda - 1) && \lambda \neq 0 \\ &= \ln y && \lambda = 0 \end{aligned}$$

has continuity at $\lambda = 0$ but the range now depends strongly on λ (the locationscale adjustment can be absorbed by the regression parameter in the typical analysis). The second transformation provides a location adjustment followed by a power transformation:

$$
\begin{aligned}
l(y, \lambda) &= (y + \lambda_2)^{\lambda_1} & \lambda_1 \neq 0 \\
&= \ln (y + \lambda_2) & \lambda_1 = 0 .
\end{aligned}
$$

This transformation maps $(-\lambda_2, \infty)$ onto $(0, \infty)$ for $\lambda_1 \neq 0$ and onto $(-\infty, \infty)$ for $\lambda_1 = 0$.

The context for the first kind of transformation might be: a range of possibilities exists for the natural variable and, whatever it is, the given variable is some transformation of it. This gives primary status to the natural variable.

The context for the second kind of transformation might be: the preceding relative to λ_1; and the given variable is in doubt as to its zero point. The additional argument here gives primary status to the given variable.

Now consider the second generalized model and suppose that it has expression transformations $A = \{\alpha\}$ that are appropriate to the first kind of context. From a given variable Y the class of possible natural response variables is $\{\alpha^{-1} Y : \alpha \in A\}$. But the given variable might equally have been $\alpha_1 \alpha_0^{-1} Y$ where α_0 is the actual transformation. The class of possible natural variables would then be $\{\alpha^{-1}\alpha_1\alpha_0^{-1} Y : \alpha \in A\}$. The equality of these classes gives $A^{-1}\alpha_1\alpha_0^{-1} = A^{-1}$ and hence $\alpha_1 A^{-1}\alpha_1\alpha_0^{-1} = \alpha_1 A^{-1}$ for all α_1. It follows that the transformations $\{\alpha\alpha_0^{-1} : \alpha \in A\}$ form a group. And if the given space is relabelled so the identity is in A, then the transformations A form a group. Thus if the possible natural variables are identified from any given variable then the class A is effectively a group. As examples consider: the power transformations with $\lambda > 0$; the power transformations with $\lambda \neq 0$.

Now suppose that the class A is enlarged to a class Λ as a consequence of certain doubts concerning the given variable. Similar arguments then present Λ as a union of left cosets of the group A. As an example consider: the power-location transformations with $\lambda_1 \neq 0$.

9. Invariant likelihood. Consider the second generalization (Section 2) of the variation-response model. And suppose that the class A of expression transformations A is group, that the group A operates coordinate by coordinate on the natural response, and that A is exactly transitive on any coordinate variable.

The density function for the given response variable is given routinely in terms of the *volume* generated from length for each component variable. A density function for orbit, however, needs a volume measure in subspaces unaligned with the coordinate axes. In Section 4 such a volume measure was generated from the Euclidean inner product—in a sense because "it was there" and not because it related in any natural way to the model. The arbitrariness of this volume measure became apparent with the examination of change of initiating variable in Section 7.

The group A is exactly transitive on each coordinate. For the ith coordinate an invariant length measure can be constructed as follows: let d_i be a reference value; let $\langle Y_i \rangle$ be the transformation (in A) such that $\langle Y_i \rangle d_i = Y_i$; let $J_i(\langle Y_i \rangle)$ be $c_i\, |\partial \langle Y_i \rangle X / \partial X|$ with $X = d_i$ and c_i constant; then $dn_i(Y_i) = dY_i / J_i(\langle Y_i \rangle)$ is invariant. Now consider a transformation λ near the identity and choose c_i or d_i for each coordinate so that the effect of λ on each coordinate is the same; the invariant lengths are standardized. If the group has the same application on each coordinate then the preceding standardization can be obtained by having $c_i \equiv 1$ and $d_i = d$.

Now suppose that at any point in $\mathbb{R}^N$ volume in a subspace is generated from the Euclidean inner product based on the invariant lengths dn_i on each axis. Then in the notation of Section 7, $K(Y) = \text{dia}\, J_i(\langle Y_i \rangle)$ where dia designates the diagonal matrix constructed from the elements that follow it. Thus the change to invariant length is equivalent to replacing $J(\lambda^{-1} : Y)$ by $J(\lambda^{-1} : Y)\, \text{dia}\, J_i(\langle Y_i \rangle)$. The resulting *invariant* likelihood is

$$L_1{}^* = \frac{\mathbb{R}^+(D_\lambda) k(D_\lambda, \lambda) J_Q([Y_\lambda])\, |J(\lambda^{-1} : Y)| \prod_1^n J_i(\langle Y_i \rangle)}{J_N([Y_\lambda] : D_\lambda)\, |(--)'(\text{dia}\, J_i^{-1}(\langle Y_i \rangle) J^{-1}(\lambda^{-1} : Y) W(Y_\lambda))|^{-\frac{1}{2}}} .$$

This invariant likelihood has the properties associated with L_1 in Section 6 and in addition is independent of the choice of initiating variable. The likelihood function is not affected by a common rescaling of the invariant lengths.

Consider briefly the generalization in which A is not necessarily a group. For this suppose that A operates identically on each coordinate axis and is exactly transitive. Let d be a reference point, let $\langle Y_i \rangle$ be the transformation carrying d into Y_i, let $J(\langle Y_i \rangle)$ be $|\partial \langle Y_i \rangle X / \partial X|$ with $X = d$, and let $dn(Y_i) = dY_i / J(\langle Y_i \rangle)$. A change in d will typically change the length measure nonhomogeneously. This length measure may, however, be a more appropriate length measure than the original measure; and if A is a group then $dn(Y_i)$ will be the invariant length. The formula for $L_1{}^*$ can be applied with the present definition of $J(\langle Y_i \rangle)$ and the resulting likelihood is an L_1 likelihood—presumably better than the original L_1 likelihood, certainly better in the case of a group A.

10. Transit likelihood. Consider the second generalization (Section 2) of the variation-response model. And suppose, as in Section 9, that the class A of expression transformations is a group, that the group A operates coordinate by coordinate on the natural response, and that A is exactly transitive on any coordinate variable.

Let $dn_i(Y_i) = dY_i / J_i(\langle Y_i \rangle)$ be the standardized invariant length for the ith coordinate. And suppose that at any point in $\mathbb{R}^N$ volume in a subspace is generated from the Euclidean inner product based on the invariant lengths dn_i. Then the probability for the identified orbit is

$$\frac{k(D_\lambda, \lambda) J_Q([Y_\lambda])\, |J(\lambda^{-1} : Y)| \prod_1^n J_i(\langle Y_i \rangle) \prod_1^n dn_i}{J_N([Y_\lambda] : D_\lambda) d[Y_\lambda]} .$$

The vector differential $d[Y_\lambda]$ on the orbit can be related to the Euclidean differential on the orbit as derived from the invariant lengths dn_i:

$$d\mathbf{n} = \text{dia}^{-1} J_i(\langle Y_i \rangle) J^{-1}(\lambda^{-1} : Y) W(Y_\lambda) d[Y_\lambda] .$$

If the corresponding Euclidean volume element is used on the orbit, then the probability element for orbit is based on Euclidean volume orthogonal to the orbit and the likelihood L_1^* of Section 9 is obtained.

Now suppose the Euclidean differential on the orbit is projected onto the orthogonal complement to the λ-orbit. Let $d\mathbf{n}$ now refer to the projected differential; it can be calculated directly.

The tangent vector to the λ-orbit is $\mathbf{1} = (1, \cdots, 1)'$ in terms of the standardized invariant lengths dn_i. The projection into the orthogonal complement of $\mathbf{1}$ is obtained by $P = I - n^{-1}\mathbf{1}\mathbf{1}'$; the matrix P replaces a column vector by its deviation vector (deviations from the mean). The new $d\mathbf{n}$ in the orthogonal complement to the λ-orbit is then

$$d\mathbf{n} = P\,\text{dia}^{-1} J_i(\langle Y_i \rangle) J^{-1}(\lambda^{-1} : Y) W(Y_\lambda) d[Y_\lambda] ,$$

and the corresponding volume element is

$$|d\mathbf{n}| = |(--)'(P\,\text{dia}^{-1} J_i(\langle Y_i \rangle) J^{-1}(\lambda^{-1} : Y) W(Y_\lambda))|\; |d[Y_\lambda]| .$$

The probability for the identified orbit is then

$$\frac{k(D_\lambda, \lambda) J_Q([Y_\lambda])\; |J^{-1}(\lambda^{-1} : Y)| \prod_1^n J_i(\langle Y_i \rangle)\, dv_t}{J_N([Y_\lambda] : D_\lambda)\; |(--)'(P\,\text{dia}^{-1} J_i(\langle Y_i \rangle) J^{-1}(\lambda^{-1} : Y) W(Y_\lambda))|^{-\frac{1}{2}}}$$

where dv_t measures Euclidean invariant volume along (transit) the λ-orbit and orthogonal to the orbit and λ-orbit. The resulting transit likelihood is

$$L_1{}^t = \frac{\mathbb{R}^+(D_\lambda) k(D_\lambda, \lambda) J_Q([Y_\lambda])\; |J^{-1}(\lambda^{-1} : Y)| \prod J_i(\langle Y_i \rangle)}{J_N([Y_\lambda] : D_\lambda)\; |(--)'(P\,\text{dia}^{-1} J_i(\langle Y_i \rangle) J^{-1}(\lambda^{-1} : Y) W(Y_\lambda))|^{-\frac{1}{2}}} .$$

11. Examples; the power-transformed regression model. Consider the regression model of Section 1 as generalized with the power transformations of Section 8:

$$C_1 = \{\textstyle\prod_1^n f(u_i : \lambda)\, d\mathbf{u} : \lambda \in \Lambda\} , \qquad C_2 = \{Y = \lambda\theta U : \boldsymbol{\beta} \in \mathbb{R}^r, \sigma \in \mathbb{R}^+, \lambda \in \Lambda\}$$

where

$$Y_\lambda = \begin{pmatrix} V \\ \mathbf{y}_\lambda' \end{pmatrix}, \qquad U = \begin{pmatrix} V \\ \mathbf{u}' \end{pmatrix}, \qquad \theta = \begin{pmatrix} I & \mathbf{0} \\ \boldsymbol{\beta}' & \sigma \end{pmatrix},$$

and $l(y : \lambda) = \lambda^{-1} y = y^\lambda (\lambda \neq 0)$ is the power transformation.[1]

The orthogonal likelihood L_1 for λ is

$$L_1 = \mathbb{R}^+ \frac{k(D_\lambda, \lambda)\lambda^{n-r-1} \prod y_i^{\lambda-1}}{s^{n-r-1}(\mathbf{y}_\lambda)\; |D_\lambda\, \text{dia}\,(y_i^{2-2\lambda}) D_\lambda'|^{-\frac{1}{2}}} ,$$

[1] As a reasonable approximation assume that the regression variation-response model is applicable on the positive axis.

which for the normal case becomes

$$L_1 = \mathbb{R}^+ \frac{\lambda^{n-r-1} \prod y_i^{\lambda-1}}{s^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda \operatorname{dia}(y_i^{2-2\lambda})D_\lambda'|^{-\frac{1}{2}}}\,.$$

Now consider the application of the power transformation group on $\mathbb{R}^+$:

$$(\lambda)y = y^{-\lambda} = e^{-\lambda \ln y}\,, \qquad (-1)y = y\,.$$

The transformation $\langle y \rangle = (-\ln y)$ carries e into y,

$$(-\ln y)e = e^{\ln y} = y\,;$$

and the change in length under a transformation is

$$\frac{d(\lambda)y}{dy} = -\lambda y^{-\lambda-1}\,.$$

Hence

$$J((-\ln y)) = |\ln y| \cdot e^{\ln y - 1} = |\ln y| \cdot y/e\,,$$

$$dn(y) = \frac{e\,dy}{y\,|\ln y|}\,;$$

the constant e corresponds to a change in reference point and can accordingly be omitted.

The invariant likelihood for λ is then

$$L_1^* = \mathbb{R}^+ \frac{k(D_\lambda, \lambda) \prod_1^n y_i^\lambda\,|\ln y_i^\lambda|}{s^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda \operatorname{dia}^{-2}(y_i^\lambda\,|\ln y_i^\lambda|)D_\lambda'|^{-\frac{1}{2}}}\,,$$

which for the normal case becomes

$$L_1^* = \mathbb{R}^+ \frac{\prod_1^n y_i^\lambda\,|\ln y_i^\lambda|}{s^{n-r-1}(\mathbf{y}_\lambda)\,|D_\lambda \operatorname{dia}^{-2}(y_i^\lambda\,|\ln y_i^\lambda|)D_\lambda'|^{-\frac{1}{2}}}\,.$$

The transit likelihood for λ is

$$L_1^t = \mathbb{R}^+ \frac{k(D_\lambda, \lambda) \prod_1^n y_i^\lambda\,|\ln y_i^\lambda|}{s^{n-r-1}(\mathbf{y}_\lambda)\,|(--)'(P \operatorname{dia}^{-1} y_i^\lambda\,|\ln y_i^\lambda|\,D_\lambda')|^{-\frac{1}{2}}}$$

which for the normal case becomes

$$L_1^t = \mathbb{R}^+ \frac{\prod_1^n y_i^\lambda\,|\ln y_i^\lambda|}{s^{n-r-1}(\mathbf{y}_\lambda)\,|(--)'(P \operatorname{dia}^{-1} y_i^\lambda\,|\ln y_i^\lambda|\,D_\lambda')|^{-\frac{1}{2}}}$$

As a final example consider the "other side" of the Example 1 in Section 6. Consider the model with normal variation and the expression transformations (λ) which relocate,

$$C_1 = \{(2\pi)^{-n/2} \exp\{-\tfrac{1}{2}\,\Sigma u_i^2\}\,d\mathbf{u}\}\,, \qquad C_2 = \{\mathbf{y} = (\lambda)\sigma\mathbf{u} : \sigma \in \mathbb{R}^+, \lambda \in \mathbb{R}\}\,,$$

where $(\lambda)\mathbf{y} = \mathbf{y} + \lambda\mathbf{1}$; thus λ is the location parameter for the given response. For notation, see Section 5 with $r = 0$.

The orthogonal likelihood is the invariant likelihood:

$$L_1 = L_1^* = \mathbb{R}^+ \frac{A_n}{s^{n-1}(\mathbf{y}_\lambda)\,|\mathbf{d}_\lambda'\mathbf{d}_\lambda|^{-\frac{1}{2}}}$$

$$= \mathbb{R}^+ \frac{1}{s^{n-1}(\mathbf{y}_\lambda)} = \mathbb{R}^+ \frac{1}{(\Sigma(y_i - \lambda)^2)^{(n-1)/2}}$$

$$= \mathbb{R}^+ \frac{1}{\left(1 + \frac{n(\bar{y} - \lambda)^2}{(n-1)s_y^2}\right)^{(n-1)/2}};$$

this is not the likelihood function associated with the ordinary t-test for location.

The transit likelihood is

$$L_1^t = \mathbb{R}^+ \frac{A_n}{s^{n-1}(\mathbf{y}_\lambda)\,|d_\lambda' P\mathbf{d}_\lambda|^{-\frac{1}{2}}}$$

$$= \mathbb{R}^+ \frac{1}{s^{n-1}(\mathbf{y}_\lambda)\,|(\mathbf{y} - \bar{y}\mathbf{1})/s(\mathbf{y}_\lambda)|^{-1}}$$

$$= \mathbb{R}^+ \frac{\Gamma(n/2)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(n-1))} \left(1 + \frac{n(\bar{y} - \lambda)^2}{(n-1)s_x^2}\right)^{\frac{1}{2}(n-1)+\frac{1}{2}};$$

and this is the likelihood function associated with the ordinary t-test.

12. Transit likelihood within a variation-response model. The determination of likelihood has been examined under progressive increases in the complexity of the statistical model. The most developed determination is the transit likelihood of Section 10. Now consider this transit likelihood *within* a variation-response model—for the case where the variation-response model plus expression transformations is in fact a larger variation-response model.

For this consider

$$C_1 = \{p(u)\,dU\}\,, \qquad C_2 = \{Y = \lambda\theta U : \theta \in G, \lambda \in H\}$$

where G, H are transformation groups and where semi direct product $K = HG$ is a transformation group. And in particular consider the information available concerning λ within this variation-response model and how it compares with the transit likelihood treating H as expression transformations applied to the variation-response model involving G.

Let Y be the observed response value. For simplicity of notation take the reference point at Y and consider coordinates for the concealed variation in terms of group elements from K, H, G : $U = kY = ghY$ where k in K is factored as gh with g in G and h in H. The equation $Y = \lambda\theta U$ then becomes $i = \lambda\theta gh$ which factors as $h = \lambda^{-1}$ and $g = \theta^{-1}$.

Now consider the distribution $p(U)\,dU$ on and near the orbit KY. Let dO be volume orthogonal to KY at Y and let p^* be the corresponding density for orbit KU. Let $d\mu_2(h)$ be the left invariant differential on the group H and $p(h)$

be the conditional density for orbit Gh within K. Let $d\mu_1(g)$ be the left invariant differential on the group G and $p(g:h)$ be the conditional density along the orbit Gh. The distribution of U is then

$$p^* dO p(h)\, d\mu_2(h) p(g:h)\, d\mu_1(g) .$$

The variable h contains the information concerning λ; in fact, $h = \lambda^{-1}$. The identified distribution for h is $p(h) d\mu_2(h)$. The corresponding distribution describing λ is

$$p(\lambda^{-1})\, d\mu_2(\lambda^{-1}) .$$

The likelihood function for λ from the distribution $p(h)\, d\mu_2(h)$ is obtained from the transformation λ. Let $w = \lambda h$; then the distribution of W is $p(\lambda^{-1} W)\, d\mu_2(\omega)$ which at $\omega = i$ gives the direct likelihood

$$\mathbb{R}^+ p(\lambda^{-1}) .$$

Now consider the transit likelihood function. For this suppose that H operates exactly transitively on individual coordinates of U, or on pairs, or on triplets, $\cdots$. Let Euclidean coordinates be given at the identity in H. And let dU be the standardized invariant differential vector induced by H applied componentwise to the Euclidean differential at the identity. Length and angle based on dU are then invariant under H. But points on an orbit HU (or KU) remain on that orbit under H. It follows that orthogonality to KU is preserved under H; that dO measures volume orthogonal to KY not just at Y but at all points along HY; that $d\mu_2(h)$ which is H-invariant is just a multiple c of the corresponding Euclidean differential. It follows that the probability element underlying transit likelihood is

$$p^* dO\, p(\lambda^{-1} i)\, d\mu_2(\lambda^{-1} i)$$

and the corresponding transit likelihood is

$$\mathbb{R}^+ p(\lambda^{-1}) .$$

Thus the reflected density function for λ (include an arbitrary multiplying constant), the direct likelihood function for λ, and the transit likelihood function for λ are identical.

REFERENCES

BOX, G.E.P. and COX, D.R. (1964). An analysis of transformations. *J. Roy. Statist. Soc. Ser.* **B 26** 211–252.

FRASER, D.A.S. (1967). Data transformations and the linear model. *Ann. Math. Statist.* **38** 1456–1465.

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