Ann. Inst. Statist. Math.
Vol. 40, No. 2, 353-366 (1988)

ON GENERALIZATION OF THE ANALYSIS OF VARIANCE

D. A. S. FRASER[1*] AND P. MCDUNNOUGH[2]

[1] *Department of Mathematics, York University, Toronto, Ontario, Canada M3T 1P3*
[2] *Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 1A1*

(Received February 27, 1986; revised August 12, 1987)

Abstract. The standard analysis of variance procedures were developed and organized primarily in the context of the normal linear model; central to this organization is the orthogonality of components and the use of orthogonal projections. This paper examines two model-type generalizations of the normal linear model: the regression model with nonnormal error and the exponential linear model. Principles of conditioning and measurement are used to develop corresponding analysis-of-variance procedures. In each case a linear fibre or foliation structure replaces orthogonality; however, for the intersection of the two model-types, which is the normal linear model, the two quite-different fibre-foliation structures reduce to a product space structure, which with the appropriate inner product, is the usual orthogonality. For implementation, conditional-marginal densities are involved, the marginalization aspect being the restricting aspect: the marginalization degree is the number of nuisance parameters for the regression model-type and is the complement of the number of free parameters for the exponential model-type. Approximations are available and will be discussed subsequently.

Key words and phrases: Analysis-of-variance, exponential model, transformation model, sequential parameter structure.

1. Introduction

Analysis of variance exists as perhaps the most wide-spread technique of statistical analysis. The technique both in name and broad substance is due to Fisher (1925) with one of his first works on the subject providing most of the major directions for the topic.

The present form of the analysis of variance is organized largely in a pattern based on successive hypotheses and derived primarily in the context of the *normal* linear model. In this paper, we direct our attention to two model-

*University of Toronto and University of Waterloo.

types that substantially generalize the normal linear model; both generalizations retain the linearity property but generalize in quite different directions with only the normal linear model in the intersection. For the two model-types we examine the statistical procedures appropriate to testing successive hypotheses.

The first model-type is the transformation or structural model as given by $y=X\beta+\sigma z$ where z has a specified distribution on R^n and σ is known or is unknown (Sections 3 and 4). The second model-type is the continuous exponential model $f(y|\theta)=\exp\{t'(y)\theta-\phi(\theta)\}h(y)$ on R^n (Section 5). If the parameters and variables are equivalenced by $\theta'=(-1/2\sigma^2, \beta'/\sigma^2)$ and $t=y'(y, X)$, then these two model-types intersect in the normal linear model with IID normal error (Heichelheim (1966)).

The standard analysis of variance is concerned with testing a succession of linear hypotheses. In this paper, we examine such a succession in the *fully partitioned case* involving one real parameter at each step. The ordinary approach in such cases involves the introduction of one parameter at a time and the corresponding *freeing-up* of the model to obtain a closer fit to the data. The sequential testing then follows by examining the fitted components in the reverse order.

Consider the successive introduction of real parameters, one at a time. *Initially*, in some complete sense, there is a fully specified distribution say $f(y)$ for the response. A first parameter, say θ_1 in Ω_1, is introduced freeing the distribution up as say $f(y|\theta_1)$ with a hypothesis $\theta_1=\theta_1^0$ giving the initial $f(y|\theta_1^0)=f(y)$. A second parameter say θ_2 in Ω_2 is then introduced freeing the distribution further to say $f(y|\theta_2, \theta_1)$ with a hypothesis $\theta_2=\theta_2^0$ giving the preceding $f(y|\theta_2^0, \theta_1)=f(y|\theta_1)$. And finally, in this manner a p-th parameter say θ_p in Ω_p is introduced freeing the model as $f(y|\theta_p, \theta_{p-1},\ldots, \theta_1)$ with a hypothesis $\theta_p=\theta_p^0$ giving $f(y|\theta_p^0, \theta_{p-1},\ldots, \theta_1)=f(y|\theta_{p-1},\ldots, \theta_1)$. For reasons to be clarified later we write the full parameter as $\theta_p/\cdots/\theta_2/\theta_1$. In this notation, we are concerned with fitting the succession of models indicated by $\theta_p^0/\cdots/\theta_2^0/\theta_1^0$, $\theta_p^0/\cdots/\theta_2^0/\theta_1,\ldots, \theta_p/\cdots/\theta_2/\theta_1$. The testing is conducted in the reverse order: $\theta_p=\theta_p^0$, $\theta_{p-1}=\theta_{p-1}^0,\ldots, \theta_2=\theta_2^0$, $\theta_1=\theta_1^0$.

For the ordinary regression analysis of variance the θ_j^0 values are zero; this specialization is obtained from the more general case by removing from the initial data, the fixed effects indicated by the θ_j^0.

For the transformation model-type (σ known case) the succession of models takes the form $y=e$, $y=\theta_1 x_1+e,\ldots, y=\theta_1 x_1+\cdots+\theta_p x_p+e$ involving a sequence of explanatory vectors $x_1,\ldots, x_p$. In this succession, $\mathbf{0}$, $\{\theta_1 x_1\}$, $\{\theta_1 x_1+\theta_2 x_2\},\ldots$ enter as a sequence of increasing sets describing possible *locations* for the distribution. For the exponential model-type $f(y|\theta)=\exp\{y'X\theta-\phi(\theta)\}h(y)$, the same succession provides a sequence of increasing sets indexing *linear forms* in y which in turn determine the density.

For the parameter structure, consider for simplicity the case $p=3$ and, for familiar interpretation, the regression case with the succession of possible locations $\mathbf{0}$, $\theta_1 \boldsymbol{x}_1$, $\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2$, $\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 + \theta_3 \boldsymbol{x}_3$. With θ_1, θ_2 free, the introduction of θ_3 is in terms of the vector $\boldsymbol{x}_3$; $\theta_3 = \theta_3^0\ (=\mathbf{0})$ gives a 2-dimensional region in R^n, $\{\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 + \theta_3^0 \boldsymbol{x}_3 \colon \theta_1, \theta_2 \in R\}$; an alternative value for θ_3, say θ_3', also gives a 2-dimensional region $\{\theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 + \theta_3' \boldsymbol{x}_3 \colon \theta_1, \theta_2 \in R\}$. Thus, the additional parameter corresponds not directly to the vector $\boldsymbol{x}_3$ but to the 2-dimensional affine set $L(\boldsymbol{x}_1, \boldsymbol{x}_2) + \theta_3 \boldsymbol{x}_3$; that is, the vector $\boldsymbol{x}_3 + \alpha_1' \boldsymbol{x}_1 + \alpha_2' \boldsymbol{x}_2$ would equally have accomplished the generalization. In a similar manner, one step earlier the additional parameter θ_2 corresponds not directly to a vector $\boldsymbol{x}_2$ but to the 1-dimensional affine set $L(\boldsymbol{x}_1) + \theta_2 \boldsymbol{x}_2$; that is, the vector $\boldsymbol{x}_2 + \alpha_1'' \boldsymbol{x}_1$ would equally have accomplished the generalization.

The notation $\theta_p / \cdots / \theta_2 / \theta_1$ is thus used to indicate the implicit *file-system* tree structure with θ_p indexing at the top level, θ_{p-1} indexing within each branch from the top level, and finally θ_1 indexing within each tree branch from the second bottom level. The particular branch. $\theta_p^0 / \cdots / \theta_2^0 / \theta_1^0$ records the succession of hypotheses and the initial model $f(y)$ corresponds to a *bottom* vertex on the tree.

A notational issue is how to relate θ_1 in the $\theta_2 = \theta_2^0$ branch with θ_1 in some other branch say $\theta_2 = \theta_2'$; and similarly θ_2 in the $\theta_3 = \theta_3^0$ branch with θ_2 in some other branch say $\theta_3 = \theta_3'$. This can be viewed as a nuisance parameter difficulty and one pragmatic resolution is to have the natural estimate of θ_1 when $\theta_2 = \theta_2^0$ be the estimate of the appropriate θ_1 for each of the other θ_2 values, and so on up the tree.

The succession of hypothesis generates a *tree structure* and the usual product space notation for the parameter space provides a misleading view that does not address the nuisance parameter issues or indeed more general issues.

In this paper, we restrict our attention to successive tests of hypotheses and to the related confidence interval at a terminal stage; the pattern of the successive hypotheses is assumed given. Theoretically indicated analyses for the two generalized model-types are developed; some background details may be found in Fraser and MacKay (1975, 1976). We do not address, however, the larger questions concerning the choice of the sequence of hypothesis, or concerning exploratory analyses that can follow from the techniques developed, or concerning issues arising if the sequence is not fully split to individual real parameters. Some development of tests for this last problem may be found in Fraser and Massam (1985).

For the two general model-types, the methods available in the literature for testing successive hypotheses and forming confidence intervals are pragmatic or asymptotic: for the regression model with nonnormal error the usual methods address the marginal distribution for elements in the usual analysis of variance; for the continuous exponential the usual method, GLIM for example, examines likelihood drop or deviance in relation to the asymptotic chi-square approximation.

In this paper, we use conditioning and measurement principles to determine appropriate variables for the two model-types and thus obtain corresponding tests and confidence regions. This leads to generalized analysis of variance procedures appropriate to each type, procedures that are fundamentally different, with opposite sample space structures; however, when specialized to the intersection normal linear model, they become equivalent; in a sense quite different techniques reduce to the same technique, a consequence of the rotational symmetry of the IID normal.

In conclusion, we note that the ordinary analysis of variance is organized largely in terms of the criteria of orthogonality and statistical independence for the components of the usual analysis of variance table. In the two more general contexts, these criteria are not directly appropriate or relevant, and are replaced by a sample space tree structure and the use of conditional-marginal distributions. We discuss the principles and conditioning briefly in Section 2 and then examine the transformation model with known (unknown) scaling in Sections 3 and 4 and the continuous exponential model in Section 5. Section 6 contains some concluding remarks.

2. The principles and the tests and confidence intervals

We investigate transformation and exponential linear models from the viewpoint of a fully partitioned succession of hypotheses. Testing is performed in a sequence opposite to that for fitting and for an 'i-th' row involves testing $\theta_p^0/\cdots/\theta_{i+1}^0/\theta_i/\theta_{i-1}/\cdots/\theta_1=\theta_p^0/\cdots/\theta_{i+1}^0/\theta_i^0/\theta_{i-1}/\cdots/\theta_1$ and constructing correspondingly a confidence interval for θ_i. Theoretical principles discussed in this section and elaborated on in succeeding sections will lead for each model type to a *real-valued pivotal quantity* t_i and its corresponding *pivotal density* $h_i(t_i)$.

Given the pivotal variable and pivotal density the details for testing the hypothesis and forming a confidence interval are relatively straightforward. For testing the hypothesis $\theta_i=\theta_i^0$, the observed value of the corresponding pivotal t_i can be calculated and an *observed level of significance* (OLS) obtained; for this OLS some freedom of choice is available: equi-probability tails; equi-density tails; conical test (Fraser and Massam (1985)) relative to a modal or central pivotal value.

For a confidence region for θ_i, a $(1-\alpha)$ acceptance region for the pivotal variable t_i can be based on one of the just mentioned OLS tail criteria and the standard inversion produces the confidence interval or region for θ_i. A minor additional complication arises in the exponential case and will be discussed in Section 5.

Both the tests and confidence regions are computationally straightforward provided the key variable and its distribution is available at each stage. Marginalization aspects can restrict the availability of the needed densities; approximations will be discussed subsequently.

Conditionality in one of two forms is the primary theoretical principle used. Before discussing this we do mention a central issue concerning conditional versus unconditional procedures. As conditional procedures quite generally lead to a shortfall with respect to all the standard optimality criteria (thus larger variance, less power and greater confidence-interval mean length), the choice of a conditional procedure cannot be supported by these standard macro properties.

Our conditionality approach involves the search for a *conditioning* variable that describes a known characteristic of the physical context (Fraser (1968)) whose occurrence is not related to the parameter at issue, or the search for a *conditioned variable* that uniquely measures the particular parameter at issue. For some current directions for conditionality see Amari (1982) and Barndorff-Nielsen (1980).

For the transformation model, we work with the error based or structural model version. In the context of observed data and the structural model, a hypothesized parameter value allows the direct calculation of an error value which by basic probability theory, then predicates a conditional distribution (Fraser (1979)). The same result can be obtained in terms of conditioning on an ancillary. We specifically avoid this alternative basis: conditioning on ancillaries (Evans *et al.* (1986)) implies the strong likelihood principle and this is viewed as *too strong* a principle; also, conditioning in the context of multiple ancillaries effectively leads to contradictions (Fraser (1973, 1979)).

For the exponential model-type, we seek a conditioned variable that uniquely measures the parameter at issue; this is obtained by likelihood map arguments (Fraser (1979), Subsection 4.3). As part of this, the conditioning variable will have a distribution that depends on the free nuisance parameter and perhaps indeed on the parameter at issue; while the latter possibility may be less than one might wish for, we view it as a simple byproduct of obtaining a unique conditioned variable that involves *only* the parameter at issue.

The use of different approaches for the two model directions might be viewed as an expediency. In part, this can be acknowledged but we do note that the two model-types have different structure which in turn provides different access for measuring or assessing the particular parameter. The two model-types are examined in detail in the next three sections.

3. Nonnormal regression, known scaling

Consider the regression model $\boldsymbol{y}=\beta_1\boldsymbol{x}_1+\cdots+\beta_r\boldsymbol{x}_r+\boldsymbol{e}=X_r\beta_r+\boldsymbol{e}$ where $\boldsymbol{e}$ has a known distribution and X_r has full column rank. Let $\theta=\beta_r/\cdots/\beta_1$ reflect the model fitting and reverse testing pattern and suppose the null effects have been removed so that $\beta_j^0=\mathbf{0}$ for each j.

For notational convenience later, suppose there is a saturated set of vectors $\boldsymbol{x}_1,\ldots,\ \boldsymbol{x}_n$ so that the model can be written $\boldsymbol{y}=X_n\beta_n+\boldsymbol{e}$, with $\beta_{r+1}=\cdots=\beta_n=0$ for the model as initially given. Also let $\boldsymbol{v}_1\in L(X_1)$, $\boldsymbol{v}_2\in L(X_2),\ldots,\ \boldsymbol{v}_n\in L(X_n)$ be any linearly independent set of basis vectors and let $V_r=(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_r)$ and $X_n\beta_n=V_n\alpha_n$; then $L(X_r)=L(V_r)$ and $X_n=V_nT$, $V_n=X_nT^{-1}$, $\alpha_n=$

$T\beta_n$, $\beta_n = T^{-1}\alpha_n$ where T is an $n\times n$ upper triangular matrix. Let $\boldsymbol{b}_n(\boldsymbol{y})=\boldsymbol{b}_n$ and $\boldsymbol{a}_n(\boldsymbol{y})=\boldsymbol{a}_n$ be the coordinates with respect to X_n and V_n; then $a_n=Tb_n$ and $b_n=T^{-1}\boldsymbol{a}_n$. This alternative notation will be used in Section 4.

Now consider testing the i-th parameter $\beta_i=\beta_i^0$ $(=\mathbf{0})$, that is, testing $\beta_r^0/\cdots/\beta_i^0/\beta_{i-1}/\cdots/\beta_1$ given $\beta_r^0/\cdots/\beta_{i+1}^0/\beta_i/\cdots/\beta_1$. The given model then has the form

$$y = \beta_1\boldsymbol{x}_1 + \cdots + \beta_i\boldsymbol{x}_i + \boldsymbol{e} \; . \tag{3.1}$$

For a response vector $\boldsymbol{y}^0$ we can solve (3.1) for error characteristics:

$$\boldsymbol{y}^0 + L(X_i) = \boldsymbol{e} + L(X_i) \; , \quad \boldsymbol{e} \in \boldsymbol{y}^0 + L(X_i) \; , \tag{3.2}$$

and thus

$$(b_n(\boldsymbol{e}),\ldots, b_{i+1}(\boldsymbol{e})) = (b_n(\boldsymbol{y}^0),\ldots, b_{i+1}(\boldsymbol{y}^0)) \; , \tag{3.3}$$

The more restricted model with $\beta_i=\beta_i^0$ (for confidence use, allow a non zero value) gives in the same manner

$$(b_n(\boldsymbol{e}),\ldots, b_i(\boldsymbol{e})) = (b_n(\boldsymbol{y}^0),\ldots, b_i(\boldsymbol{y}^0) - \beta_i^0) \; . \tag{3.4}$$

Thus the single step tightening of the parameter hypotheses gives us *precisely one additional datum*,

$$b_i(\boldsymbol{e}) = b_i(\boldsymbol{y}^0) - \beta_i^0 = b_i^0 \; . \tag{3.5}$$

Note that the parameter β_i in model (3.1) occurs uniquely in the pivotal variable

$$b_i(\boldsymbol{e}) = b_i(\boldsymbol{y}^0) - \beta_i \; , \tag{3.6}$$

as examined conditionally given $b_{i+1}^0,\ldots, b_n^0$ (for simplicity we write $b_j(\boldsymbol{y}^0)=b_j^0$).

Let $f(\boldsymbol{e})$ be the density for $\boldsymbol{e}$. Then $f(\boldsymbol{y}-X_n\beta_n)$ is the density for $\boldsymbol{y}$ and

$$f(X_n(\boldsymbol{b}_n - \beta_n))\,|X_n| \; , \quad g(\boldsymbol{b}_n) = f(X_n\boldsymbol{b}_n)\,|X_n| \tag{3.7}$$

are the densities for $\boldsymbol{b}_n(\boldsymbol{y})=X_n^{-1}\boldsymbol{y}$, $\boldsymbol{b}_n(\boldsymbol{e})=b_n(\boldsymbol{y})-\beta_n$. Now let $g_{[i]}(b_i,\ldots, b_n)$ be the marginal density of $(b_i,\ldots, b_n)$ obtained from (3.7). Then the conditional density of the pivotal $b_i=b_i(\boldsymbol{y})-\beta_i$ given $b_{i+1}^0,\ldots, b_n^0$ is

$$g_i(b_i) = \frac{g_{[i]}(b_i,\, b_{i+1}^0,\ldots,\, b_n^0)}{g_{[i+1]}(b_{i+1}^0,\ldots,\, b_n^0)} \; . \tag{3.8}$$

Tests and a confidence interval for β_i, then follow as discussed in Section 2. The hypothesis $\beta_i=0$ gives the observed value $b_i=b_i(y^0)$ for the pivotal variable $b_i=b_i(y)-\beta_i$ and this can be compared with the 1-dimensional distribution $g_{[i]}(b_i)$ and a corresponding OLS calculated. The confidence interval for β_i given $\beta_{i+1}=\cdots=\beta_n=0$ is obtained from a $(1-\alpha)$ interval (b_i^L, b_i^U),

$$\int_{b_i^L}^{b_i^U} g_i(b_i)\,db_i\ , \tag{3.9}$$

for the pivotal $b_i=b_i(y)-\beta_i$, and is given by

$$(b_i(y^0) - b_i^U,\ b_i(y^0) - b_i^L)\ . \tag{3.10}$$

The essentials of the sequential testing can be summarized in a generalized analysis of variance table with i-th row as follows:

Parameter:	$\beta_r^0/\cdots/\beta_{i+1}^0/\beta_i/\cdots/\beta_1$,
Effect:	$\beta_i = \beta_i^0$,
Variable:	$b_n^0/\cdots/b_{i+1}^0/b_i/\cdots/b_1$,
Observed:	$b_i = b_i(y^0) = b_i^0$,
L-drop:	Δ_i .

For this we note that the successive hypotheses generate a sample space structure that is in direct correspondence with the parameter space structure $\beta_r/\cdots/\beta_1$. Accordingly, the sample point can be designated as $b_n(y)/\cdots/b_1(y)$: b_n indexes an $(n-1)$-dimensional affine subspace, b_{n-1} an $(n-2)$-dimensional space *within* a preceding space, and so on in the pattern of introduced information in the succession of hypotheses. Note carefully that the sample space partitions and tree structure are *independent* of the coordinates used and thus are not affected by a change from the (b_i) coordinates to some alternative (a_i) coordinates as described at the beginning of this section.

The likelihood drop analysis for such a sequence of hypotheses gives Δ_i for the i-th row which is the reduction in maximum log-likelihood when β_i is restricted to β_i^0. The probability density for the data can be written

$$g_{[i+1]}(b_{i+1}^0,\dots,b_n^0)g_i(b_i^0-\beta_i)f(b_1^0-\beta_1,\dots,b_{i-1}^0-\beta_{i-1}\,|\,b_i^0-\beta_i)\ , \tag{3.11}$$

where f designates the conditional density for $b_1,\dots,b_{i-1}$. Then

$$\Delta_i = \sup_{\beta_i}(l_1(\beta_i) + l_2(\beta_i))\ , \tag{3.12}$$

where $l_1(\beta_i)=\ln g_i(b_i^0-\beta_i)-\ln g_i(b_i^0-\beta_i^0)$ is likelihood drop from β_i to β_i^0 for the theoretically indicated inference variable $b_i(y)$ given $b_{i+1}^0,\dots,b_r^0$, and

$$l_2(\beta_i) = \sup_{b_1,\ldots,b_{i-1}} \ln f(b_1,\ldots, b_{i-1} \mid b_i^0 - \beta_i)$$
$$- \sup_{b_1,\ldots,b_{i-1}} \ln f(b_1,\ldots, b_{i-1} \mid b_i^0 - \beta_i^0)$$
$$= \ln m(b_i^0 - \beta_i) - \ln m(b_i^0 - \beta_i^0) ,$$

where $m(b_i^0-\beta_i)$ is the maximum attainable density for $b_1,\ldots, b_{i-1}$ when the i-th parameter is β_i. The first term in (3.12) is central to what likelihood drop should record; the second term, however, is indicating some *estimation effectiveness* for rows higher in the table and thus can be viewed as *inappropriate* to the assessment of β_i.

4. Nonnormal regression, unknown scaling

Consider now the ordinary regression model $\boldsymbol{y}=\beta_1\boldsymbol{x}_1+\cdots+\beta_r\boldsymbol{x}_r+\sigma\boldsymbol{e}=X_r\beta_r+\sigma\boldsymbol{e}$ where $\boldsymbol{e}$ has density $f(\boldsymbol{e})$ and X_r has full column rank. We assume that the parameters have been ordered so that $\theta=\beta_r/\cdots/\beta_1/\sigma$ reflects the model fitting and reverse testing pattern and that relocation gives $\beta_i^0=0,\ldots, \beta_1^0=0$. The analysis in this section corresponds closely to that in the preceding section but some details need special attention.

For notational convenience *only*, we allow the more general model

$$\boldsymbol{y} = X_{n-1}\beta_{n-1} + \sigma\boldsymbol{e} \tag{4.1}$$

using details from the preceding section. For this the parameter tree is $\beta_{n-1}/\cdots/\beta_1/\sigma$ with $\sigma>0$. This is not a fully saturated model with $\sigma>0$, but is easily handled and casts light on more general situations. We also add a complementing vector $\boldsymbol{x}_n$ but for some notational simplicity take it to be orthogonal to X_{n-1} and of unit length. Then $b_{n-1}(\boldsymbol{y})=(X'_{n-1}X_{n-1})^{-1}X_{n-1}\boldsymbol{y}$ gives the first $(n-1)$ coordinates of $\boldsymbol{b}_n(\boldsymbol{y})$ and $b_n(\boldsymbol{y})$ is the signed length of the residual. We write $c(\boldsymbol{y})=\operatorname{sgn} b_n(\boldsymbol{y})$ and $s(\boldsymbol{y})=|b_n(\boldsymbol{y})|$; let c^0 be the observed sign and s^0 the observed length for the residual.

Now consider the sample space structure generated by the successive hypotheses on the full model. For the general model $\beta_{n-1}/\cdots/\sigma$, we have $\boldsymbol{y}=X_{n-1}\beta_{n-1}+\sigma\boldsymbol{e}$ from which $c(\boldsymbol{e})=c(\boldsymbol{y})$ is the unique error function available with observed value $c(\boldsymbol{e})=c(\boldsymbol{y}^0)=c^0$.

The unique function of β_{n-1} that is available is

$$\frac{b_{n-1}(\boldsymbol{y}) - \beta_{n-1}}{s(\boldsymbol{y})} = \frac{b_{n-1}(\boldsymbol{e})}{s(\boldsymbol{e})} = t_{n-1}(\boldsymbol{e}) , \tag{4.2}$$

and it has observed value $t_{n-1}^0=b_{n-1}(\boldsymbol{y}^0)/s(\boldsymbol{y}^0)$ under the hypothesis $\beta_{n-1}=0$; the relevant distribution is that of t_{n-1} given c^0.

With hypothesized values $\beta_{i+1}=0,\ldots, \beta_{n-1}=0$ we obtain the unique function

$$\frac{b_i(y) - \beta_i}{s(y)} = \frac{b_i(e)}{s(e)} = t_i(e) , \tag{4.3}$$

of β_i, and it has observed value $t_i^0 = b_i(y^0)/s(y^0)$ under the additional $\beta_i = 0$. Thus we see that the single step tightening of the parameter hypothesis gives one additional error characteristic

$$t_i = (b_i(y) - \beta_i)/s(y) \tag{4.4}$$

from a response y and parameter valued β_i. The distribution of t_i for tests and confidence intervals is conditional on values calculated from preceding hypotheses, that is given c^0, $t_{n-1}^0, \ldots, t_{i+1}^0$.

Finally, given all β's$=0$ we obtain the unique function $s(y)/\sigma = s(e)$ of σ, and it has observed value s^0 under the further hypothesis $\sigma = \sigma^0$; the relevant distribution is that of s given c^0, $t_{n-1}^0, \ldots, t_1^0$.

Tests and confidence intervals are constructed as indicated in Section 2 and in the pattern in Section 3. For example, the $(1-\alpha)$ confidence interval for β_i given $\beta_{i+1} = \cdots = \beta_n = 0$ is $(b_i(y) - s(y)t_i^U, b_i(y) - s(y)t_i^L)$ where (t_i^L, t_i^U) is a $(1-\alpha)$ interval for the distribution, say $h_i(t_i)$ for t_i.

We now record the relevant density functions for $s, t_1, \ldots, t_{n-1}$. The joint density for $s, t_1, \ldots, t_{n-1}, c$ is available from (3.7),

$$g(st_1, \ldots, st_{n-1}, sc)s^{n-1} , \tag{4.5}$$

from which the marginal for $s, t_i, \ldots, t_{n-1}, c$ is

$$g_{[i]}(st_i, \ldots, st_{n-1}, sc)s^{n-i} , \tag{4.6}$$

and for $t_i, \ldots, t_{n-1}, c$ is

$$h_{[i]}(t_i, \ldots, t_{n-1}, c) = \int_0^\infty g_{[i]}(st_i, \ldots, st_{n-1}, sc)s^{n-i}ds . \tag{4.7}$$

The conditional density for t_i given $t_{i+1}, \ldots, t_{n-1}, c$ is

$$h_i(t_i) = \frac{h_{[i]}(t_i, \ldots, t_{n-1}, c)}{h_{[i+1]}(t_{i+1}, \ldots, t_{n-1}, c)} , \tag{4.8}$$

and this provides the test of significance for β_i and the related confidence interval.

The essentials of the sequential testing can be recorded in an analysis of variance table with $(i+1)$-st row

Parameter:	$\beta_{n-1}^0/\cdots/\beta_{i+1}^0/\beta_i/\cdots/\beta_1/\sigma$,
Effect:	$\beta_i = \beta_i^0$,
Variable:	$c^0/t_{n-1}^0/\cdots/t_{i+1}^0/t_i/\cdots/t_1/s$,
Observed:	$t_i = t_i^0$,
L-drop:	Δ_i ,

where $t_i^0=(b_i(y^0)-\beta_i^0)/s(y^0)$. The initial row of the table corresponds to examining the observed s^0 conditional on $c^0/t_{n-1}^0/\cdots/t_1^0$; the relevant density is given by $g_{[n]}(st_1^0,\ldots, st_{n-1}^0, sc^0)s^{n-1}/h_{[n]}(t_1^0,\ldots, t_{n-1}^0, c^0)$.

As in the preceding section we note that the succession of hypotheses indicated by $\beta_{n-1}/\cdots/\beta_1/\sigma$ generates a corresponding sample space tree partition which can be designated as $c(y)/b_{n-1}(y)/\cdots/b_1(y)/s(y)$. Of course this sample space tree partition is independent of the coordinates used and for example would be unaffected by a shift to coordinates $\boldsymbol{a}(y)$ relative to V_n as defined in Section 2.

The likelihood drop analysis for the sequence of hypotheses gives Δ_i for the $(i+1)$-st row, which can be written

$$\Delta_i = \sup_{\beta_1} (l_1(\beta_i) + l_2(\beta_i)) , \tag{4.9}$$

where $l_1(\beta_i)$ is likelihood drop from β_i to β_i^0 for the appropriate conditional inference variable $t_i=(b_i^0-\beta_i)/s^0$ and $l_2(\beta_i)$ is the corresponding drop in the maximum attainable log conditional density for the lower order variables; this latter term is measuring potential estimation effectiveness for the remaining nuisance parameters and is inappropriate as a contribution to the assessment of β_i.

5. Exponential linear model

Consider a continuous exponential linear model, with parameter $\theta_p/\cdots/\theta_1$ as arranged for a succession of hypotheses:

$$f(y) = \exp\left\{\sum_1^p \theta_i a_i(y) - \phi(\theta)\right\} H(y) . \tag{5.1}$$

If the hypothesized values $\theta_p^0,\ldots, \theta_1^0$ are non zero, then a parameter relocation and corresponding adjustment gives an expression (5.1) with the new $\theta_i^0=0$.

We note that $(a_1(y),\ldots, a_p(y))$ is a sufficient statistic for the full parameter; in more fundamental terms we note that the probability at a sample point as given by the likelihood depends only on that vector, other differences being notational and parameter-free (Fraser (1979)). Accordingly, we simplify and rewrite the model

$$f(y\,|\,\theta) = \exp\left\{\sum_1^p \theta_i y_i - \phi(\theta)\right\} h(y)\,, \tag{5.2}$$

with y now designating the sufficient statistic. If we also relocate the y_i at their mean value when $(\theta_1,\ldots,\theta_p)=(0,\ldots,0)$, then y is the score function at that full-hypothesis parameter value.

Again we emphasize the tree structure $\theta_p/\cdots/\theta_1$ of the parameter and note that the notation change $\tilde{\theta}=T\theta$ and $\tilde{y}'=y'T^{-1}$ where T is upper triangular with one's on the diagonal leaves the model form and parameter-testing succession unchanged. Thus for example at the second stage there is no non-arbitrary association of parameter values for θ_1 when $\theta_2=\theta_2'$ with say values for θ_1 when $\theta_2=\theta_2''$ and is a manifestation of the tree structure.

Now consider inference for θ_p with the initial parameters $\theta_1,\ldots,\theta_{p-1}$ taken as nuisance parameters. For measuring, assessing, or evaluating θ_p without influence from $\theta_1,\ldots,\theta_{p-1}$ we have by the arguments in (Fraser (1979), p. 81, (4-10)) the unique variable y_p examined conditionally given $y_1,\ldots,y_{p-1}$. This is not the full sufficiency-ancillarity (Fraser (1979), p. 80). Rather it is a structured version of weak sufficiency-ancillarity (Fraser (1979), p. 84), and has uniqueness from its construction procedure. We take y_p given $y_{p-1}^0,\ldots,y_1^0$ as the fundamental variable for assessing θ_p with $\theta_{p-1},\ldots,\theta_1$ as nuisance parameters. Note that we do not appeal to properties of similar tests or of completeness of a null distribution. This would provide an alternative approach to the present testing issue, but by preference our present focus is on structure, on available variables, and on how the information on these variables relates to the parameters at issue.

Consider the conditional density for this variable y_p given $y_1^0,\ldots,y_{p-1}^0$: let $f(y\,|\,\theta)=\exp\{\theta y-\phi(\theta)\}h(y)$ where we have $\theta=\theta_p$, $y=y_p$, $h(y)=h(y_1^0,\ldots,y_{p-1}^0,y_p)$, and

$$\exp\{\phi(\theta)\} = \int \exp\{\theta y\}h(y)dy\,, \tag{5.3}$$

note that the sample space for y may be less than R and even dependent on $y_1^0,\ldots,y_{p-1}^0$. Let $G(t,\theta)$ be the distribution function

$$G(t,\theta) = \int^t \exp\{\theta s - \phi(\theta)\}h(s)ds\,. \tag{5.4}$$

For testing $\theta=0$ the observed level of significance is

$$\text{OLS} = 2\min\{G(t^0,0),\ 1-G(t^0,0)\}\,, \tag{5.5}$$

using equi-probability tails or is

$$\text{OLS} = \begin{cases} G(t^0,0)/G(t^m,0), & t^0 < t^m \\ (1-G(t^0,0))/(1-G(t^m,0)), & t^0 > t^m\,, \end{cases} \tag{5.6}$$

using the conical test (Fraser and Massam (1985)) where t^m is say the median or the modal value for the distribution $f(y|0)$.

The median estimate is $\hat{\theta}$ defined by $G(t^0, \hat{\theta})=1/2$; it is also a central 0% confidence 'interval', in the pattern next described.

The $(1-\alpha)$ central confidence interval for θ is $(\hat{\theta}_1, \hat{\theta}_2)$ where

$$G(t^0, \hat{\theta}_2) = \alpha/2, \quad G(t^0, \hat{\theta}_1) = 1 - \alpha/2 . \tag{5.7}$$

All these characteristics are available from 1-dimensional integrations, a single integration for OLS, and iterations on this for the median estimate and confidence interval.

Now consider testing $\theta_{p-1}=\theta_{p-1}^0$ (=0) given $\theta_p^0/\theta_{p-1}/\cdots/\theta_1$. With θ_p=0 the marginal density $\boldsymbol{y}_{p-1}=(y_1,\dots, y_{p-1})$ is

$$\exp\{\theta_{p-1}y_{p-1} + \cdots + \theta_1 y_1 - \phi(\theta_1,\dots, \theta_{p-1}, 0)\}h_{p-1}(\boldsymbol{y}_{p-1}), \tag{5.8}$$

where $h_{p-1}(\boldsymbol{y}_{p-1})=\int h(\boldsymbol{y})dy_p$. This is of the preceding general form but with p replaced by $p-1$. Tests of significance, median estimates, confidence intervals can be found as above where we take $\theta=\theta_{p-1}, y=y_{p-1}, h(y)=h_{p-1}(y_1^0,\dots, y_{p-2}^0, y)$.

On the assumption of no effect for θ_p and θ_{p-1}, the process can be continued in the pattern just described to test $\theta_{p-2},\dots, \theta_1$. For implementation the single complication centers on deriving the successive marginal densities.

The sample structure derived from the successive hypotheses is represented by $y_1/y_2/\cdots/y_p$. For example, the test for the step from $\theta_p^0/\theta_{p-1}/\cdots/\theta_1$ to $\theta_p^0/\theta_{p-1}^0/\theta_{p-2}/\cdots/\theta_1$ uses the statistic $y_1^0/\cdots/y_{p-2}^0/y_{p-1}/y_p$ with observed value $y_1^0/\cdots/y_{p-1}^0/y_p$. Note that this tree-partition structure is in some sense directly opposite to that for the regression models.

The essentials of the sequential testing can be summarized in a general analysis of variance table. For this we can consider a fully saturated exponential model with $p=n$ and the sufficiency reduction corresponds to the preceding analysis with higher order θ values equal to zero. The i-th row of the table would have the form

Parameter:	$\theta_n^0/\cdots/\theta_{i+1}^0/\theta_i/\cdots/\theta_1$,
Effect:	$\theta_i = \theta_i^0$,
Variable:	$y_1^0/\cdots/y_{i-1}^0/y_i/\cdots/y_n$,
Observed:	$y_i = y_i^0$,
L-drop:	Δ_i .

The standard likelihood-drop analysis for the succession of hypotheses would have

$$\Delta_i = \sup_{\theta_i} \{l_1(\theta_i) + l_2(\theta_i)\} ,$$

where $l_1(\theta_i)$ is the appropriate drop for θ_i to θ_i^0 using the theoretically indicated variable y_i given $y_1^0, \ldots, y_p^0$ and

$$
\begin{aligned}
l_2(\theta_i) = & \sup_{\theta_1, \ldots, \theta_{i-1}} \ln f(y_1^0, \ldots, y_{i-1}^0; \theta_1, \ldots, \theta_{i-1}, \theta_i) \\
& - \sup_{\theta_1, \ldots, \theta_{i-1}} \ln f(y_1^0, \ldots, y_{i-1}^0; \theta_1, \ldots, \theta_{i-1}, \theta_i^0) ,
\end{aligned}
$$

where f is the marginal density for $y_1^0, \ldots, y_{i-1}^0$ under $\theta_n^0 / \cdots / \theta_{i+1}^0 / \theta_i / \cdots / \theta_1$. The second term $l_2(\theta_i)$ records a logarithmic maximum for the conditioning variables; this profile value arises in part because the conditioning variables are not fully ancillary for the conditional use of y_i and it enters in an uncalibrated way that distorts the information from the theoretically indicated i-th variable.

6. Concluding remarks

The succession of hypotheses indicated by the notation $\theta_p / \theta_{p-1} / \cdots / \theta_1$ has led to a succession of tests indicated by $b_n / b_{n-1} / \cdots / b_1$ and $c / t_{n-1} / \cdots / t_1 / s$ for the general regression models and by $y_1 / y_2 / \cdots / y_n$ for the exponential linear model. These two general tree-partition patterns have opposite sample space structure: conditioning on values lower in the table for the regression case and on values higher in the table for the exponential case. Correspondingly, the marginalization is over the number of nuisance parameters in the regression case and over the number of free variables in the exponential case.

The *normal* regression model belongs to both general model types. How then do the opposite types of conditional analyses relate one to the other?

For this the conditional procedures in Sections 3 and 4, with some moderate analysis and choice of convenient coordinates, can be shown to be effectively independent of the conditioning and thus equivalent to the ordinary marginal-type analysis. In a similar manner, the conditional procedure involving the tree partition in Section 5 can be examined for the normal case, and the usual orthogonal coordinates show that the conditional procedure is again the same as the ordinary marginal procedure. Thus two quite distinct procedures devolve to the same procedure in the context of the *rotationally-symmetric* normal.

For implementation of the methods discussed in this paper conditional distributions are used and are of course readily accessible. These conditional distributions are, however, in the context of some preceding marginalization: over the nuisance parameters in the regression context and over the free variables in the exponential case. This marginalization is the restricting factor for implementation; some approximation procedures will be discussed subsequently with background details in Fraser (1980, 1987).

The widely used likelihood-drop procedures are examined against the theoretically indicated procedures and are found to contain potential for bias and distortion.

REFERENCES

Amari, S.-I. (1982). Geometrical theory of asymptotic ancillarity and conditional inference, *Biometrika*, **69**, 1–17.

Barndorff-Nielsen, O. (1980). Conditionality resolutions, *Biometrika*, **67**, 293–310.

Evans, M., Fraser, D. A. S. and Monette, G. (1986). On arguments from sufficiency and conditionality to likelihood, *Canad. J. Statist.*, **14**, 181–200.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh.

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

Fraser, D. A. S. (1973). The elusive ancillarity, *Multivariate Statistical Inference*, (eds. D. G. Kabe and R. P. Gupta), North-Holland, Amsterdam.

Fraser, D. A. S. (1979). *Inference and Linear Models*, McGraw-Hill.

Fraser, D. A. S. (1980). Inference and the structural model for ANOVA and MANOVA, *Handbook of Statistics*, (ed. P.R. Krishnaiah), North-Holland, Amsterdam.

Fraser, D. A. S. (1987). Fibre analysis and tangent models, *Statist. Hefte*, **28**, 163–181.

Fraser, D. A. S. and MacKay, J. (1975). Parameter factorization and inference based on significance, likelihood and objective posterior, *Ann. Statist.*, **3**, 559–572.

Fraser, D. A. S. and MacKay, J. (1976). On the equivalence of standard inference procedures, *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, (eds. W. L. Harper and C. A. Hooker), Vol. II, 47–62, D. Reidel, Dordrecht, Holland.

Fraser, D. A. S. and Massam, H. (1985). Conical tests: observed levels of significance and confidence regions, *Statistical Papers*, **26**, 1–18.

Heichelheim, P. C. A. (1966). Comparison between sufficiency and structural methods, Ph. D. Thesis, University of Toronto.