Non-nested linear models: A conditional confidence approach

D.A.S. Fraser and R.J. Gebotys

York University, University of Toronto, University of Waterloo; Wilfrid Laurier University, University of Toronto

Key words and phrases: Non-nested linear models, confidence interval, location parameter, angular separation of hypotheses, Creasy-Fieller problem, conditional inference, precision-based confidence interval, approximate inference.

AMS 1980 subject classifications: Primary, 62; secondary, J05, A30.

ABSTRACT

The comparison of nested linear models with normal error is well standardized in the common procedures of the analysis of variance. This article considers the comparison of two non-nested linear models that have the same parameter dimension; the comparison is made on the assumption that the true mean lies somewhere in the linear span of the two models. The analysis leads to a precision-based conditional confidence interval for the unsigned angular direction of the true mean, and this in turn provides a confidence assessment of the two directions that correspond to the two models being compared. The confidence interval is an approximate conditional interval (given the distance of the estimate from the intersection of the hypotheses), and its length as a fraction of \( \pi \) indicates the precision of the confidence procedure. The method provides a conditional-inference alternative to a confidence interval available by Creasy-Fieller analysis.

RÉSUMÉ

Il est d'usage courant de comparer des modèles linéaires emboités avec termes d'erreur normaux dans les procédures habituelles d'analyse de la variance. Dans cet article, nous comparons deux modèles linéaires non emboités dont les paramètres sont de même dimension; la comparaison s'effectue en supposant que la moyenne théorique se situe quelque part dans l'espace linéaire engendré par les deux modèles. L'analyse donne lieu à un intervalle de confiance conditionnel basé sur la précision pour la direction angulaire non signée de la moyenne théorique et, à son tour, ceci procure un jugement de confiance portant sur les deux directions qui correspondent aux deux modèles comparés. L'intervalles de confiance est un intervalle approximatif conditionnel (étant donné la distance qui sépare l'estimation de l'intersection des deux hypothèses) et sa longueur, exprimée en fraction de \( \pi \), indique la précision de la procédure d'estimation par intervalle. La méthode, basée sur une inférence conditionnelle, est une solution de rechange à la procédure de Creasy-Fieller.

1. INTRODUCTION

The comparison and testing of nested linear models with normal error is well standardized in the common procedures of the analysis of variance. More generally the problem of selecting independent variables for a linear model has led to a variety of methods, including forward and backward stepwise methods and the all-subsets methods. In this paper we consider the comparison of two non-nested linear models.
that have the same parameter dimension; the comparison is made on the assumption that the true mean lies in the linear span of the two models and the error is normally distributed with known variance. A conditional confidence interval is developed for the unsigned direction of the true mean relative to the intersection of the two models; two specific unsigned directions are shown to correspond to the two models being compared. The method provides an alternative to a confidence interval available from Creasy-Fieeller analysis, an interval that can produce an anomalous full-parameter-space interval with obvious 100% confidence.

Consider the normal model \( y \sim N_p(\eta, \sigma^2 I) \) with unknown mean \( \eta \) and known variance \( \sigma^2 \). We consider confidence-theory inference for the two linear models, \( \eta \in \mathcal{L}_A \) and \( \eta \in \mathcal{L}_B \), where \( \mathcal{L}_A \) and \( \mathcal{L}_B \) are linear subspaces of dimension \( k \), and the mean \( \eta \) is assumed to lie in the span \( \mathcal{L}_{A+B} \) of \( \mathcal{L}_A + \mathcal{L}_B \). Let \( k - d \) be the dimension of the intersection model \( \mathcal{L}_A \cap \mathcal{L}_B \). Then \( d \) is the number of independent vectors to be appended to the intersection model to form \( \mathcal{L}_A \) or \( \mathcal{L}_B \). We first examine the case \( d = 1 \), and then in Section 5 the case \( d > 1 \).

Consider the case \( d = 1 \), in which the models \( \mathcal{L}_A \) and \( \mathcal{L}_B \) differ only with respect to the choice for the "final" regression vector, which can of course be expressed as a vector orthogonal to the intersection model \( \mathcal{L}_{AB} = \mathcal{L}_A \cap \mathcal{L}_B \); the two possibilities by themselves generate one-dimension spaces \( \mathcal{L}_{A:AB} \) and \( \mathcal{L}_{B:AB} \) which are the orthogonal residuals of \( \mathcal{L}_A \) and \( \mathcal{L}_B \) with respect to \( \mathcal{L}_{AB} \). The standard factorization of the normal linear model shows that inference concerning \( \eta \) with respect to \( \mathcal{L}_A \) and \( \mathcal{L}_B \) can be examined entirely in the two-dimensional span \( \mathcal{L}_{A+B:AB} \) of \( \mathcal{L}_{A:AB} \) and \( \mathcal{L}_{B:AB} \).

Let \( \eta_{0+1}^a, B \) and \( \eta_{0+1}^b, B \) be the projections of \( \eta \) and \( y \) to \( \mathcal{L}_{A+B:AB} \); see Figure 1. Also let \( (p, \theta) \) and \( (r, a) \) be polar coordinates for \( \eta_{0+1}^a, B \) and \( \eta_{0+1}^b, B \) relative to a direction, say, bisecting one of the acute angles between \( \mathcal{L}_{A:AB} \) and \( \mathcal{L}_{B:AB} \), and let \( 2c \) be the measure of the acute angle. Note then that the angular direction \( c \) or \( -c + \pi \) designates the model \( \mathcal{L}_A \), and \( -c \) or \( -c + \pi \) designates the model \( \mathcal{L}_B \), as examined orthogonally to \( \mathcal{L}_{AB} \). This paper determines a conditional confidence interval for the direction \( \theta \) of the true mean \( \eta \) with precision based appropriately on the length \( r \) from the intersection model. Tables 1 and 2 in Section 3 record confidence limits for \( z = a - \theta \) conditional on \( r \), for the case \( \sigma = 1 \).

The present development of a confidence interval in the case \( d = 1 \) is in recognition that the two hypotheses correspond to two lines in \( \mathcal{L}_{A+B:AB} \) and that the true mean has a direction from the intersection of these two lines. If the two hypotheses are in standard terminology orthogonal, that is, correspond to orthogonal contrasts such as the row effect and column effect in a \( 2 \times 2 \) factorial analysis, then the angle \( 2c \) used here is \( \pi/2 \). More generally the mean vector for a hypotheses-A model has the form \( \eta_{0+1}^a + \hat{i}_A \), and for a hypothesis-B model has the form \( \eta_{0+1}^b + \hat{i}_B \), where \( \hat{i}_A \) and \( \hat{i}_B \) are unit orthogonal explanatory vectors for the extensions to \( A \) and to \( B \) from \( AB \); the angle between \( \hat{i}_A \) and \( \hat{i}_B \) is \( 2c \). On the assumption that the mean vector lies in the span \( \mathcal{L}_{A+B} \), then the mean vector has the form \( \eta_{AB} + \hat{i} \), where the unit vector \( \hat{i} \) lies in the plane of \( \hat{i}_A \) and \( \hat{i}_B \) and has angular direction given by \( \theta \) with respect to the bisector of \( \hat{i}_A \) and \( \hat{i}_B \).

We will see that the radial distance \( r \) is the clear determinant of the precision of the confidence interval for the direction \( \theta \) of the true mean. The distance \( p \) of the true mean enters as a nuisance parameter, and an averaged confidence interval is determined by using posteriors for that parameter.

Efron (1984) suggests using a parameter \( \Delta \), the difference in the mean squared error using model \( B \) as opposed to model \( A \). The \( C_p \) statistics (Mallows 1973) provide
unbiased estimates of the mean squared error for each model, and the difference between these provides an estimate \( \hat{\Delta} \) of \( \Delta \). A confidence interval for \( \Delta \) is based on bootstrap simulations of the distribution of \( \hat{\Delta} \) or on related Edgeworth and Cornish-Fisher calculations. This approach focuses on prediction properties rather than directly on the intrinsic model-indicator parameter used in this paper.

Hotelling (1940) provides a test for a “null” hypothesis that the mean vector \( \eta \) is equidistant from \( \mathcal{L}_A \) and \( \mathcal{L}_B \). The test does not address confidence in one or other model, as a positive departure can have opposite implications depending on the location of the mean vector.

The Creasy-Fieller method (Fieller 1954) can be applied to the problem and would use the pivotal variable \( v = r \sin(a - \theta) \) with a standard normal distribution (\( \sigma^2 = 1 \) case). For small values of \( r \) this leads to acceptance of all \( \theta \)-values, an anomaly with 100% confidence levels that one usually chooses to avoid.

For the more general case \( (d > 1) \) in which the models differ in more than one dimension, Euler angles can describe the separation between the two possibilities for the subspace of additional vectors. The Euler angles (cosines of these can be
represented as canonical partial correlations) arise in \( d \) orthogonal two-dimensional subspaces, and the methods for the \( d = 1 \) case provide confidence intervals for each of \( d \) angular directions \( \theta_i \) of the true \( \eta \) projected into the \( i \)th subspace. This is surveyed in Section 5 together with an example.

2. FORMULATION

Consider the statistical problem outlined in the introduction of assessing two competing linear models \( \mathcal{L}_A \) and \( \mathcal{L}_B \) of dimension \( k \) in the context that the mean vector is assumed to lie in the span \( \mathcal{L}_{A+B} \) of dimension \( k + d \). We examine the case \( d = 1 \) in this section.

First we note that an orthogonal transformation applied to \( y \) can segregate \( k - 1 \) variables that have the same mean in the two models, \( n - k - 1 \) variables that estimate error, and two variables that correspond to the vectors that distinguish the two models. Inference principles reduce the problem to these latter two variables, which can be presented as coordinates for the projection of \( y \) into \( \mathcal{L}_{A+B} \), which is the orthogonal residual of \( \mathcal{L}_{A+B} \) with respect to \( \mathcal{L}_B \). We use the specialized notation introduced in Section 1 and illustrated in Figure 1.

Consider a reparametrization as it affects the two-dimensional subspace \( \mathcal{L}_{A+B} \).

The two linear models of original dimension \( k \) are represented as the lines \( \mathcal{L}_{A,AB} \) and \( \mathcal{L}_{B,AB} \) through the origin; see Figure 1. A \( k \)-dimensional model consistent with the assumption that \( \eta \in \mathcal{L}_{A+B} \) is also a line through the origin. The intrinsic indexing for such models is given by the direction of such lines, that is, by the direction (with arbitrary sign) of the mean from the origin. Also note that the angle \( 2c \) between \( \mathcal{L}_{A,AB} \) and \( \mathcal{L}_{B,AB} \) indicates how similar the models are and thus relates to how easily they are distinguished one from the other. As a reference direction we choose a bisector of one of the acute angles between \( \mathcal{L}_{A,AB} \) and \( \mathcal{L}_{B,AB} \) and then choose a direction of rotation from that bisector. In accord with this let \((\rho, \theta)\) be the polar coordinates of the mean \( \eta_{A+B}^0 \). We note then that the model \( A \) is given by \( \theta = c, c + \pi \) and the model \( B \) is given by \( \theta = -c, -c + \pi \). Indeed, any \( k \)-dimensional original model consistent with the assumptions is given by a direction, which can be specified as \( \theta = d, d + \pi \).

This paper derives a confidence interval for the model-indexing parameter \( \theta \), which in turn provides a confidence assessment of the models \( A \) and \( B \) specified respectively by \( \theta = c, c + \pi \) and \( \theta = -c, -c + \pi \); as nuisance parameter we have \( \rho \), which is the distance from the origin and is invariant under the rotational symmetries of the problem. In the special situation where only model \( A \) or \( B \) is true, the confidence interval remains a confidence “interval” for the corresponding angle (this is a general property of confidence regions when the parameter space is restricted).

For more detailed notation let \((v_1, v_2)\) be an orthogonal basis for \( \mathcal{L}_{A+B,AB} \). This can be chosen in a symmetric manner by letting \( v_1 \) be the bisector of one of the acute angles between \( \mathcal{L}_{A,AB} \) and \( \mathcal{L}_{B,AB} \), and \( v_2 \) be orthogonal to \( v_1 \) in the plane of \( \mathcal{L}_{A+B,AB} \) with direction chosen so that \( \mathcal{L}_A \) is in the positive quadrant of the new axes; see Figure 1.

As an example, suppose model \( A \) is given by \( E(y) = \alpha x_1 + \beta_1 x_2 + \beta_2 x_3' \), and model \( B \) by \( E(y) = \alpha x_1 + \beta_1 x_2 + \beta_3 x_3' \). Then \( v_1 \) can be taken to be

\[
v_1 = \frac{|x_{3,12}'| x_{3,12}' \pm |x_{3,12}'| x_{3,12}'}{|x_{3,12}'| x_{3,12}' \pm |x_{3,12}'| x_{3,12}'}
\]
with sign chosen so that the inner product \((x'_{3:12}, \pm x''_{3:12}) = |x'_{3:12}| |x''_{3:12}| \cos 2\phi\) is positive; and \(v_2\) would then be

\[v_2 = \frac{x'_{3:12} - (x'_{3:12}, v_1)v_1}{|x'_{3:12} - (x'_{3:12}, v_1)v_1|},\]

for notation we use \(x_{3:12}\) to designate the residual of \(x_3\) orthogonalized to \(x_1, x_2\).

Let \((\eta_1, \eta_2)\) and \((y_1, y_2)\) be the coordinates of \(\eta^0_{A+B}\) and \(y^0_{A+B}\) with respect to \(v_1, v_2\). The polar coordinates for \(\eta^0_{A+B}\) are then obtained from \(\eta_1 = \rho \cos \theta, \eta_2 = \rho \sin \theta\), and for \(y^0_{A+B}\) are obtained from \(y_1 = r \cos a, y_2 = r \sin a\).

3. ANALYSIS

In the simplified coordinates from Section 2, the general model has the form

\[(y_1, y_2) = (r \cos a, r \sin a) = N_2(\rho \cos \theta, \rho \sin \theta, \sigma^2 I_2),\]

where \(\sigma^2\) is known. The specialized model \(A\) has \(\theta = c, c + \pi\) and \(B\) has \(\theta = -c, -c + \pi\). We develop a confidence interval for \(\theta\) with the central purpose of obtaining a confidence evaluation of the special \(\theta\)-values for model \(A\) and model \(B\).

The joint density for \((r, a)\) is

\[(2\pi\sigma^2)^{-1} \exp \left(\frac{-r^2 + \rho^2 - 2r\rho \cos(a - \theta)}{2\sigma^2}\right) \cdot r\]

on \((0, \infty) \times (0, 2\pi)\). It follows that the marginal distribution of \(r^2\) is \(\sigma^2\) chi-squared \(g(2, \rho^2/\sigma^2)\) with noncentrality \(\rho^2/\sigma^2\), and the conditional distribution of \(e = a - \theta\) is Von Mises with density

\[f(e | r) = (2\pi I_0(\kappa))^{-1} \exp(\kappa \cos e)\] \hspace{1cm} (3.1)

where \(I_0(\cdot)\) is the imaginary Bessel function of order zero and \(\kappa = \rho \sigma^2\) measures precision (like reciprocal variance) for this "normal" distribution on the circle. For simplicity we now standardize so that \(\sigma^2 = 1\).

We first investigate confidence-theory inference for \(\theta\) in the context of a given value for the nuisance parameter \(\rho\). The statistical model then is a location model for the variable \(a\) on the circle \((0, 2\pi)\). This follows from the fixed \(\theta\)-free distribution for the "error" \(e = a - \theta\). The standard analysis for transformation-parameter or structural models requires conditioning on the orbits under the rotation group that generates the model. The orbits are indexed by \(r\), and conditional inference then uses the simple location model \(f(a - \theta | r)\) given by (3.1). Confidence limits are readily available by numerical integration. Note that the conditioning is the same whatever the value for the parameter \(\rho\) and is thus viewed as an intrinsic aspect of assessing the angle \(\theta\).

We note that the precision of the location distribution of \(a\) is given by \(\kappa = \rho \sigma^2\) and varies from low when \(r\) is near \(0\) to high as \(r\) becomes large. This precision of course manifests itself in the size of the conditional confidence interval for \(\theta\).

Now consider inference with \(\rho\) as a free parameter. Clearly the precision as described in the \(\rho\)-known case depends on \(\rho\). A natural first approach is to use the conditional inference as just described but with \(\rho\) replaced by its MLE estimate \(\hat{\rho}\). This gives an approximate conditional confidence region for \(\theta\) consistent with various current approaches to conditional inference.

For the preceding method the quantile limits for \(e\) have been calculated for \(r = 0\) to
\[ r = 10 \text{ in steps of 0.2 and 0.5 for confidence levels } \beta = 90\%, 95\%, 99\% \text{ by computer integration using} \]
\[
\int_{-e_\beta}^{e_\beta} f(e \mid r) \, dr + \int_{-\pi-e_\beta}^{\pi-e_\beta} f(e \mid r) \, dr = \beta
\]  
(3.2)

with range reduced modulo \([0, 2\pi]\); the two-piece integral results from the fact that \( \theta \) and \( \theta + \pi \) both represent the same one-dimensional line through the origin. Values are recorded in Table 2 below for this \( \rho \)-estimated density.

The \( \beta \) confidence interval for \( \theta \) as obtained from the location relation \( e = a - \theta \) is given by
\[
(a - e_\beta, a + e_\beta)
\]  
(3.3)

where \( e_\beta \) comes from the tables using the observed \( r \) (recall the standardization to \( \sigma^2 = 1 \)).

As a preferred approach we examine the conditional model \( f(e \mid r) \) in (3.1) and average in some natural pragmatic fashion over possible \( \rho \)-values. One option in this direction is to use the marginal likelihood from the chi-squared distribution for \( r^2 \).

An alternative to this that we prefer is to use the confidence or fiducial distribution for \( \rho \) given \( r \). For this, let \( F(r^2 \mid \rho^2) \) be the noncentral \( \chi^2 \) (2, \( \rho^2 \)) distribution function
\[
F(r^2 \mid \rho^2) = \exp\left(-\frac{\rho^2}{2}\right) \sum_{k=0}^{\infty} \frac{(\rho^2/2)^k}{k!} G(r^2 | 2 + 2k),
\]  
(3.4)

where \( G(\cdot \mid k) \) designates the central \( \chi^2 \) (\( k \)) distribution function. The confidence interval percentage points \( \rho^2_e \) are obtained from \( F(r^2 \mid \rho^2_e) = 1 - \alpha \), and this corresponds to a \( \rho^2 \)-density
\[
\frac{1}{2}F(r^2 \mid \rho^2) - \frac{1}{2} \exp\left(-\frac{\rho^2}{2}\right) \sum_{k=0}^{\infty} \frac{(\rho^2/2)^k}{k!} G(r^2 | 4 + 2k)
\]  
(3.5)

plus discrete probability \( 1 - G(r^2 \mid 2) \) at \( \rho = 0 \). From the fiducial viewpoint there are two reasonable possibilities (Fisher 1935; 1956, p. 133): one is a renormalized version of the density given by (3.5), and the other is the preceding confidence distribution including the discrete component.

In accord with the preceding, we have examined two methods of averaging the conditional density \( f(e \mid r) \) in (3.1): method I uses the fiducial density (3.5) without the discrete component; method II uses the confidence fiducial density given by (3.5) with the discrete component. In method I the average conditional density is \( f^I(e \mid r) \) given by
\[
f^I(e \mid r) = G^{-1}(r^2 \mid 2)(4\pi)^{-1} \sum_{k=0}^{\infty} \frac{(\frac{1}{2})^k}{k!} (G(r^2 | 2 + k) - G(r^2 | 4 + 2k)).
\]  
(3.6)

\[
\int_{0}^{\infty} I_0(rp) \rho^2 \exp\left(rp \cos e - \frac{\rho^2}{2}\right) \, dp^2,
\]  
(3.6)

and in method II is \( f^{II}(e \mid r) \) given by
\[
f^{II}(e \mid r) = \frac{1}{2\pi}(1 - G(r^2 \mid 2)) + G(r^2 \mid 2)f^I(e \mid r).
\]

The quantile limits have been calculated in accord with (3.2) and are recorded in Table 1. The confidence interval is then given by (3.3). For larger \( r \) methods I and II
Table 1: Angular limit $e_\beta(r)$ at three confidence levels.

<table>
<thead>
<tr>
<th>$r$</th>
<th>Method-I averaging</th>
<th></th>
<th>Method-II averaging</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
<td>90%</td>
</tr>
<tr>
<td>0.01</td>
<td>1.418</td>
<td>1.497</td>
<td>1.558</td>
<td>1.418</td>
</tr>
<tr>
<td>0.20</td>
<td>1.414</td>
<td>1.492</td>
<td>1.558</td>
<td>1.414</td>
</tr>
<tr>
<td>0.40</td>
<td>1.405</td>
<td>1.488</td>
<td>1.558</td>
<td>1.414</td>
</tr>
<tr>
<td>0.60</td>
<td>1.387</td>
<td>1.479</td>
<td>1.558</td>
<td>1.414</td>
</tr>
<tr>
<td>0.80</td>
<td>1.366</td>
<td>1.470</td>
<td>1.553</td>
<td>1.405</td>
</tr>
<tr>
<td>1.00</td>
<td>1.335</td>
<td>1.453</td>
<td>1.549</td>
<td>1.392</td>
</tr>
<tr>
<td>1.20</td>
<td>1.291</td>
<td>1.431</td>
<td>1.545</td>
<td>1.370</td>
</tr>
<tr>
<td>1.40</td>
<td>1.239</td>
<td>1.401</td>
<td>1.536</td>
<td>1.335</td>
</tr>
<tr>
<td>1.60</td>
<td>1.174</td>
<td>1.357</td>
<td>1.531</td>
<td>1.287</td>
</tr>
<tr>
<td>1.80</td>
<td>1.095</td>
<td>1.305</td>
<td>1.518</td>
<td>1.217</td>
</tr>
<tr>
<td>2.00</td>
<td>1.008</td>
<td>1.226</td>
<td>1.497</td>
<td>1.121</td>
</tr>
<tr>
<td>2.20</td>
<td>0.916</td>
<td>1.139</td>
<td>1.466</td>
<td>1.012</td>
</tr>
<tr>
<td>2.40</td>
<td>0.833</td>
<td>1.043</td>
<td>1.422</td>
<td>0.899</td>
</tr>
<tr>
<td>2.60</td>
<td>0.759</td>
<td>0.951</td>
<td>1.361</td>
<td>0.798</td>
</tr>
<tr>
<td>2.80</td>
<td>0.689</td>
<td>0.864</td>
<td>1.274</td>
<td>0.716</td>
</tr>
<tr>
<td>3.00</td>
<td>0.633</td>
<td>0.785</td>
<td>1.156</td>
<td>0.646</td>
</tr>
<tr>
<td>3.20</td>
<td>0.585</td>
<td>0.720</td>
<td>1.047</td>
<td>0.589</td>
</tr>
<tr>
<td>3.40</td>
<td>0.541</td>
<td>0.663</td>
<td>0.936</td>
<td>0.545</td>
</tr>
<tr>
<td>3.60</td>
<td>0.502</td>
<td>0.615</td>
<td>0.873</td>
<td>0.506</td>
</tr>
<tr>
<td>3.80</td>
<td>0.471</td>
<td>0.576</td>
<td>0.807</td>
<td>0.471</td>
</tr>
<tr>
<td>4.00</td>
<td>0.445</td>
<td>0.537</td>
<td>0.746</td>
<td>0.445</td>
</tr>
<tr>
<td>4.50</td>
<td>0.388</td>
<td>0.467</td>
<td>0.637</td>
<td>0.388</td>
</tr>
<tr>
<td>5.00</td>
<td>0.345</td>
<td>0.414</td>
<td>0.558</td>
<td>0.345</td>
</tr>
<tr>
<td>5.50</td>
<td>0.310</td>
<td>0.375</td>
<td>0.502</td>
<td>0.310</td>
</tr>
<tr>
<td>6.00</td>
<td>0.284</td>
<td>0.340</td>
<td>0.454</td>
<td>0.284</td>
</tr>
<tr>
<td>6.50</td>
<td>0.262</td>
<td>0.314</td>
<td>0.414</td>
<td>0.262</td>
</tr>
<tr>
<td>7.00</td>
<td>0.240</td>
<td>0.288</td>
<td>0.384</td>
<td>0.240</td>
</tr>
<tr>
<td>7.50</td>
<td>0.227</td>
<td>0.270</td>
<td>0.358</td>
<td>0.227</td>
</tr>
<tr>
<td>8.00</td>
<td>0.209</td>
<td>0.253</td>
<td>0.332</td>
<td>0.209</td>
</tr>
<tr>
<td>8.50</td>
<td>0.196</td>
<td>0.236</td>
<td>0.314</td>
<td>0.196</td>
</tr>
<tr>
<td>9.00</td>
<td>0.188</td>
<td>0.222</td>
<td>0.297</td>
<td>0.188</td>
</tr>
<tr>
<td>9.50</td>
<td>0.179</td>
<td>0.209</td>
<td>0.279</td>
<td>0.179</td>
</tr>
<tr>
<td>10.00</td>
<td>0.170</td>
<td>0.205</td>
<td>0.275</td>
<td>0.170</td>
</tr>
</tbody>
</table>

converge (vanishing effect of $\rho = 0$), and in the limit $e_\beta = \sin^{-1}(z_{1-\beta}/r)$, or $e_\beta = z_{1-\beta}/r$, where $z_\alpha$ is the right tail $\alpha$-point for the standard normal; this follows by noting that for large $r$, $r \sin \alpha$, $r \cos \alpha$ is approximated by $(r\alpha, r)$ with high probability. For small $r$ we have that $e_\beta$ approximates $\pi/2\beta$, giving an interval that is a fraction $\beta$ of the possible range $\pi$; this can be noted at the value $r = 0.01$ in Table 1.

All three methods involve conditioning on $r$ and thus conform to the theoretically indicated procedure for the $\rho$-known case. Some theory now developing indicates that this is an optimal conditioning procedure in the general case (Cox and Reid 1987; Fraser and Reid 1988). Methods I and II involve averaging to compensate for inadequacies in the initial method that uses an estimated $\rho$.

For a comparison we note that the Fieller (1954) solution for the Creasy-Fieller problem gives the $\beta$ confidence interval $(a \pm \sin^{-1}(\min(1, z_{1-\beta}/r)))$ for the line at angle $\theta$ or $\theta + \pi$. Note that for $r \leq z_{1-\beta}/2$, the confidence interval covers the full parameter range $(0, 2\pi)$, an interval with clear 100% conditional confidence.
Table 2: Angular limit \( e_\beta(r) \) at three confidence levels.

<table>
<thead>
<tr>
<th>( r )</th>
<th>( \beta )-estimation</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.418</td>
<td>1.497</td>
<td>1.558</td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>1.414</td>
<td>1.492</td>
<td>1.558</td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>1.414</td>
<td>1.492</td>
<td>1.558</td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>1.409</td>
<td>1.492</td>
<td>1.558</td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>1.401</td>
<td>1.488</td>
<td>1.558</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.374</td>
<td>1.475</td>
<td>1.553</td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>1.326</td>
<td>1.449</td>
<td>1.549</td>
<td></td>
</tr>
<tr>
<td>1.40</td>
<td>1.248</td>
<td>1.401</td>
<td>1.540</td>
<td></td>
</tr>
<tr>
<td>1.60</td>
<td>1.130</td>
<td>1.322</td>
<td>1.518</td>
<td></td>
</tr>
<tr>
<td>1.80</td>
<td>0.999</td>
<td>1.200</td>
<td>1.479</td>
<td></td>
</tr>
<tr>
<td>2.00</td>
<td>0.886</td>
<td>1.073</td>
<td>1.409</td>
<td></td>
</tr>
<tr>
<td>2.20</td>
<td>0.794</td>
<td>0.960</td>
<td>1.291</td>
<td></td>
</tr>
<tr>
<td>2.40</td>
<td>0.720</td>
<td>0.868</td>
<td>1.169</td>
<td></td>
</tr>
<tr>
<td>2.60</td>
<td>0.659</td>
<td>0.794</td>
<td>1.065</td>
<td></td>
</tr>
<tr>
<td>2.80</td>
<td>0.611</td>
<td>0.733</td>
<td>0.977</td>
<td></td>
</tr>
<tr>
<td>3.00</td>
<td>0.567</td>
<td>0.681</td>
<td>0.908</td>
<td></td>
</tr>
<tr>
<td>3.20</td>
<td>0.528</td>
<td>0.633</td>
<td>0.842</td>
<td></td>
</tr>
<tr>
<td>3.40</td>
<td>0.497</td>
<td>0.593</td>
<td>0.790</td>
<td></td>
</tr>
<tr>
<td>3.60</td>
<td>0.467</td>
<td>0.559</td>
<td>0.742</td>
<td></td>
</tr>
<tr>
<td>3.80</td>
<td>0.441</td>
<td>0.528</td>
<td>0.702</td>
<td></td>
</tr>
<tr>
<td>4.00</td>
<td>0.419</td>
<td>0.502</td>
<td>0.663</td>
<td></td>
</tr>
<tr>
<td>4.50</td>
<td>0.371</td>
<td>0.445</td>
<td>0.587</td>
<td></td>
</tr>
<tr>
<td>5.00</td>
<td>0.336</td>
<td>0.397</td>
<td>0.528</td>
<td></td>
</tr>
<tr>
<td>5.50</td>
<td>0.303</td>
<td>0.362</td>
<td>0.476</td>
<td></td>
</tr>
<tr>
<td>6.00</td>
<td>0.279</td>
<td>0.331</td>
<td>0.436</td>
<td></td>
</tr>
<tr>
<td>6.50</td>
<td>0.257</td>
<td>0.305</td>
<td>0.404</td>
<td></td>
</tr>
<tr>
<td>7.00</td>
<td>0.240</td>
<td>0.284</td>
<td>0.375</td>
<td></td>
</tr>
<tr>
<td>7.50</td>
<td>0.223</td>
<td>0.266</td>
<td>0.349</td>
<td></td>
</tr>
<tr>
<td>8.00</td>
<td>0.209</td>
<td>0.249</td>
<td>0.327</td>
<td></td>
</tr>
<tr>
<td>8.50</td>
<td>0.196</td>
<td>0.233</td>
<td>0.308</td>
<td></td>
</tr>
<tr>
<td>9.00</td>
<td>0.188</td>
<td>0.223</td>
<td>0.288</td>
<td></td>
</tr>
<tr>
<td>9.50</td>
<td>0.177</td>
<td>0.207</td>
<td>0.275</td>
<td></td>
</tr>
<tr>
<td>10.00</td>
<td>0.166</td>
<td>0.201</td>
<td>0.262</td>
<td></td>
</tr>
</tbody>
</table>

The present approach targets on a conditional \( \beta \) confidence level given \( r \). For small \( r \) the conditional approach gives a proportion \( \beta \) of the available angular range (based on the limiting uniform distribution), rather than the 100% with the Fieller interval; for medium \( r \) the conditional interval is larger (it is determined by the conditional distribution); for large \( r \) the intervals become the same. The length of the confidence interval in comparison with the model separation angle \( 2\pi \) provides background information for the discrimination between the two model possibilities.

In Section 6 we discuss the Monte Carlo assessment of the approximate confidence limits just described. Reasonably close approximation is obtained over the critical range of \( \rho^2 \) values, with some moderate preference for method II over method I.

A survey of methods concerning the Fieller problem may be found in Wallace (1980); also see Sprott and Viveros (1985).
4. AN EXAMPLE

Efron (1984) illustrates a bootstrap-based prediction method of comparing non-nested linear models using data found in Williams (1959). Williams considers tests to determine the relation between maximum compressive strength parallel to the grain (y) and density (x_{11}) for pine lumber. Some specimens of pine contain a large amount of resin, which contributes to the density but little to the strength; the resin content of the specimens was determined and the adjusted density figure (x_{21}) calculated. The question of concern is whether density or adjusted density is a better determinant of wood strength.

Model A has \( \eta = \mu \mathbf{1} + \beta_{11} x_{11} \), whereas model B has \( \eta = \mu \mathbf{1} + \beta_{21} x_{21} \). For a sample of \( n = 42 \), the full residual mean square on 30 degrees of freedom was used as an estimate and replacement for \( \sigma^2 (\tilde{\sigma} = 276.387) \). Rescaling into standard units for which \( \sigma \) is taken equal to 1 required dividing all values \( y, x_{11}, \) and \( x_{21} \) by 276.387. The observed \( r \) was calculated to be 19.762 with relevant angles \( \alpha = 0.0907 \) and \( c = 0.1448 \).

The confidence intervals for \( \theta \) and in particular for the angles \( \theta(A) = -0.1448 \) for model A and \( \theta(B) = +0.1448 \) for model B are \( 0.0907 \pm 0.0873 = (0.003, 0.178), 0.0907 \pm 0.1004 = (-0.0097, 0.1911), 0.0907 \pm 0.1353 = (-0.0446, 0.226) \) at levels 90\%, 95\%, 99\%. These intervals were calculated by direct computer integration. The confidence intervals all bracket the model-B angle and provide support for the adjusted density model in preference to the unadjusted density model.

Efron (1984) lets \( \Delta \) be the difference in mean squared error of prediction (MSEP) between model B and model A (\( \Delta = \text{MSEP}_B - \text{MSEP}_A \)). The point estimate for \( \Delta \) using the Williams data is -20.144. The null hypothesis \( \Delta = 0 \) is rejected at \( \alpha = 0.1 \). Hotelling (1940) suggests a test that depends on the fact that the sum of squares of the regression of \( y \) on \( x_i \), where \( i = 1, 2 \), is the square of a linear function of \( y \). He tests whether these have the same expectation. The corresponding \( t \)-value for the Williams data is 1.79. The null hypothesis \( \eta_A = \eta_B \) is rejected at \( \alpha = 0.1 \). Neither method produces directly a confidence region for the indicator for the true hypothesis A or B.

Now consider the Fieller (1954) method. The 95\% confidence interval is \( (0.0907 \pm \sin^{-1}(1.96/19.762)) = (0.0907 \pm 0.0993) = (-0.0086, 0.1900) \), which is, perhaps misleadingly, shorter than the interval calculated by the conditional method.

5. MULTIPARAMETER SEPARATION

Now consider the more general case in which models \( \mathcal{L}_A \) and \( \mathcal{L}_B \) differ in \( d \) dimensions, with \( d > 1 \). The model structure at issue is \( \mathcal{L}_{A:AB} \) versus \( \mathcal{L}_{B:AB} \), both of dimension \( d \); recall that the span of these spaces is \( \mathcal{L}_{A:AB} \) of dimension \( 2d \). Then as discussed in Efron (1984) the \( 2d \)-dimensional space \( \mathcal{L}_{A:AB} \) decomposes into \( d \) orthogonal 2-dimensional spaces, each of the type discussed in Section 2. More specifically, there are \( d \) pairs of unit vectors \( a_i, b_i \), each pair being orthogonal to each other pair and with inner product \( (a_i, b_i) = \cos 2c_i \), the angles \( 2c_1, \ldots, 2c_d \) being Euler angles between the two spaces. We then obtain

\[
\mathcal{L}_{A:AB} = \mathcal{L}(a_1, \ldots, a_d),
\]
\[
\mathcal{L}_{B:AB} = \mathcal{L}(b_1, \ldots, b_d).
\]
For preliminary discussion it is convenient to suppose that simplified coordinates $y_1, \ldots, y_{2d}$ are available in the pattern discussed in Section 3, but without the requirement of symmetry in relation to the models $\mathcal{L}_{A:AB}$ and $\mathcal{L}_{B:AB}$; the axes for such coordinates can arise from a Gram-Schmidt orthogonalization of $X_A$ and then $X_B$, where for example $X_A$ designates the design matrix for $\mathcal{L}_A$ and where the first columns of $X_A$ are assumed to represent $\mathcal{L}_{A:AB}$. With respect to the simplified axes, let $\bar{\mathcal{P}}_A$ and $\bar{\mathcal{P}}_B$ be the projection matrices to $\mathcal{L}_{A:AB}$ and $\mathcal{L}_{B:AB}$ respectively (both are $2d \times 2d$). Then by the singular-value decomposition (SVD) theorem applied to $\bar{\mathcal{P}}_A \bar{\mathcal{P}}_B$ we have

$$
\bar{\mathcal{P}}_A \bar{\mathcal{P}}_B = \sum_{i=1}^{d} \cos^2 c_i \bar{a}_i \bar{b}_i
$$

(5.2)

where $\mathcal{L}_{A:AB} = \mathcal{L}(\bar{a}_1, \ldots, \bar{a}_d)$, $\mathcal{L}_{B:AB} = \mathcal{L}(\bar{b}_1, \ldots, \bar{b}_d)$, in the present simplified coordinates.

More directly, at the expense of a decomposition of an $n \times n$ matrix (rather than $2d \times 2d$), let $P_A$ and $P_B$ be the projection matrices for $\mathcal{L}_A$ and $\mathcal{L}_B$ in the original coordinates. For the SVD let $\cos 2c_1, \ldots, \cos 2c_d$ be the eigenvalues not equal to 0 or 1, and $\bar{a}_1, \ldots, \bar{a}_d$ and $\bar{b}_1, \ldots, \bar{b}_d$ be the corresponding eigenvectors. Then the angles $2c_i$ are those discussed in the preceding paragraph, and $\mathcal{L}_{A:AB} = (\bar{a}_1, \ldots, \bar{a}_d)$, $\mathcal{L}_{B:AB} = (\bar{b}_1, \ldots, \bar{b}_d)$ in the original coordinates. For related detail see Hotelling (1940, Section 7), Anderson (1958, Ch. 12), Dempster (1968, Ch. 5), Eaton (1983, Ch. 1).

For the assessment of the model $\mathcal{L}_A$ with respect to $\mathcal{L}_B$ now let $(r_1, \cos a_1, r_1 \sin a_1)$ and $(\rho, \cos \theta_1, \rho \sin \theta_1)$ be the coordinates of $y$ and $\eta$ with respect to the symmetric axes developed in Section 3 and used now with respect to $\mathcal{L}(\bar{a}_i)$ and $\mathcal{L}(\bar{b}_i)$ as model-A and model B ingredients. Then a $\beta$-level confidence interval for the angle $\theta_1$ is given by

$$(a_1 \pm e_\beta(r_1)),
$$

(5.3)

where $e_\beta(r_1)$ comes from Table 1 using the observed $r_1$ (after standardization to $\sigma^2 = 1$). Model $A$ corresponds to $\theta_1 = c_1$, or $c_1 + \pi$, and model $B$ to $\theta_1 = -c_1$ or $-c_1 + \pi$. The confidence intervals for the various angles $\theta_1, \ldots, \theta_d$ are statistically independent by virtue of the eigenvector orthogonality; the precision of an interval with respect to separating the hypothesis is available from the size of the interval as indicated by $2e_\beta(r_1)$ in relation to the separation of the hypothesis as given by $2c_1$.

As an example we consider the data from Draper and Smith (Ch. 7). The data concern a number of components that can affect a rocket engine's performance: temperature $x_{11}$, static fire $x_{12}$, vibration $x_{21}$, and shock $x_{22}$. As the engine is tested, the preceding variables are recorded together with chamber pressure $y$. We investigate whether the agitation variables (vibration and shock) or the heat variables (temperature and static fire) provide a better model for chamber pressure, and at the same time obtain a confidence assessment for blends of these variables.

The formal model $A$ has $\eta = \mu + \beta_{11} x_{11} + \beta_{12} x_{12}$, and model $B$ has $\eta = \mu + \beta_{21} x_{21} + \beta_{22} x_{22}$. For the sample of $n = 24$, the full residual mean square on 19 degrees of freedom was used as an estimate and replacement for $\sigma^2$. Let $X_A$ and $X_B$ be the corresponding $n \times 3$ design matrices. With $P_A = X_A(X_A'X_A)^{-1}X_A'$ and $P_B$ defined similarly, the SVD of $P_A P_B$ gives eigenvalues 1.06687, 0.0562. For the eigenvalue $\cos 2c_1 = 0.06687$ we have

\[ 2c_1 = 0.8382, \quad r_1 = 11.882, \quad a_1 = 0.5311, \]
and for the eigenvalue \( \cos 2c_2 = 0.0562 \) we have

\[
2c_2 = 1.514, \quad r_2 = 6.837, \quad a_2 = 0.6566.
\]

The confidence intervals for \( \theta_1 \) and in particular for \( \theta_1(A) = 0.414 \) and \( \theta_1(B) = -0.414 \) are (0.387, 0.675), (0.361, 0.701), (0.308, 0.753) at levels 90%, 95%, 99%. The confidence intervals for \( \theta_2 \) and in particular for \( \theta_2(A) = 0.757 \) and \( \theta_2(B) = -0.757 \) are (0.408, 0.905), (0.359, 0.953), (0.266, 1.05) at levels 90%, 95%, 99%. The confidence intervals all bracket the model-A angles and provide strong support for the heat-variable model in preference to the agitation-variable model.

6. MONTE CARLO SIMULATION

A Monte Carlo simulation was used to estimate the marginal confidence level for the intervals \( (a \pm e_\rho(r)) \) recorded in (3.3) and thus to assess the nominal confidence level \( \beta \). First generate \( (z_1, z_2) \) from the \( \text{N}(0, 1) \). Calculate \( (y_1, y_2) = (e_1 + \rho, e_2) \) where the parameter \( \rho \) is given. Let \( r \) be the distance to the origin, \( a \) the smaller angle between \( (y_1, y_2) \) and the first axis. Advance the counter corresponding to \( \beta \) by unity if \( a \) is less than or equal to \( e_\rho(r) \) in magnitude. The method of simulation gives the \( \chi^2(2, r^2) \) distribution for \( r^2 \) and the conditional Von Mises \( (\kappa = \rho r) \) distribution for \( a \). The one-dimensional space for the true \( \eta \) is the first axis: \( \theta = 0 \). Because of the rotational symmetry the validation applies equally for all \( \theta \), not just the chosen \( \theta = 0 \).

Table 3 displays the results for a sample size \( N = 10,000 \) and \( \beta = 95\% \) using the (Table 1) intervals based on the method-I and method-II averaging, and using Table 2 intervals based on the estimated \( \rho \).

ACKNOWLEDGEMENT

The authors express their appreciation for many helpful suggestions, particularly those of one referee, who also indicated key directions for further work that space has presented us from addressing here. One of these, particularly deserving of attention, involves a 75%-25% blending of the unconditional and conditional intervals to stabilize the confidence level. We offer our special thanks.

REFERENCES


Fraser, D.A.S. (1979). Inference and Linear Models. DAI, Univ. of Toronto Textbook Store, Toronto.

Received 2 July 1985
Revised 22 July 1987
Accepted 23 July 1987

Department of Mathematics
York University
Toronto, Ontario M3J 1P3

Department of Psychology
Wilfrid Laurier University
Waterloo, Ontario N2L 3C5