APPROXIMATE CONDITIONAL INERENCE
AND THE LINEAR FUNCTIONAL MODEL

A. Dobriyal
Department of Statistics
University of Toronto
Toronto, Canada, M5S 1A1

D.A.S. Fraser
Department of Mathematics
York University
North York, Canada, M3J 1P3

R. Geotys
Department of Psychology
Wilfrid Laurier University
Waterloo, Canada, N2L 3C5

Key Words and Phrases: Linear functional model; conditional inference; von Mises distribution; confidence distribution; structural parameter

ABSTRACT

Approximate conditional inference is developed for the slope parameter of the linear functional model with two variables. It is shown that the model can be transformed so that the slope parameter becomes an angle and nuisance parameters are radial distances. If the nuisance parameters are known an exact confidence interval based on a location-type conditional distribution is available for the angle. More generally, confidence distributions are used to average the conditional distribution over the nuisance parameters yielding an approximate conditional confidence interval that reflects the precision indicated by the data. An example is analyzed.
1. INTRODUCTION

A linear functional model is an extension of the ordinary linear model which has errors in the independent variable \( x \) as well as the dependent variable \( y \). Such models have been used frequently in fields such as econometrics (e.g. Havelmo, 1943; Anderson, 1977) and psychology and sociology (e.g. Jöreskog, 1973).

We consider the linear functional model for the case of two variables; the model has the form

\[
\begin{align*}
    x_i &= X_i + \delta_i \\
    y_i &= Y_i + \varepsilon_i \\
    Y_i &= \theta_1 + \theta_2 X_i
\end{align*}
\]  

(1.1)

where \( X_i, Y_i \) denote mean or true values and \( x_i, y_i \) are corresponding observable quantities with associated errors \( \delta_i, \varepsilon_i \). It is typically assumed that the errors \( \delta_i \) and \( \varepsilon_i \) are uncorrelated and that there is zero correlation from one \( i \)-value to another. The primary parameters are \( \theta_1 \) and \( \theta_2 \). Neyman and Scott (1948) refer to \( \theta_1 \) and \( \theta_2 \) as structural parameters; the mean-value parameters \( X_1, ..., X_n \) are called incidental parameters.

The case of normal errors has been investigated by various authors including Lindley (1947), Tukey (1951), Anderson (1976, 1984), Kendall and Stuart (1979). In the absence of replication the standard maximum likelihood procedure is not available because the model as it stands lacks identifiability (Lindley and El-Sayyad, 1968). A common assumption compensating for this is that the ratio \( \lambda = \sigma_y^2 / \sigma_x^2 \) is known (Kendall and Stuart, 1979).

Confidence interval estimation was investigated by Creasy (1956) who derived intervals for \( \theta = \arctan \theta_2 \) by a pivotal argument. The confidence regions, however, have the disturbing property that they sometimes include the whole parameter space which in turn implies a more disturbing property that the intervals have less than the rated confidence for the remaining cases. Sprott and Viveros (1983) as part of a more general discussion on ratios obtained confidence regions from an approximate linear pivot.

This paper develops an approximate conditional confidence interval for \( \theta_2 \) which uses indicators of precision calculated from the data. The method extends earlier results for nonnested linear models (Fraser and Gebotys, 1984) and the calibration problem (Dobriyal, Fraser and Gebotys, 1987), where it was shown that the overall confidence level is approximate and that the data sensitive nature of the confidence region avoids the difficulties inherent in the pivot used by Creasy (1956).
2. MODEL REDUCTION

Under the assumption of normal errors and known $\lambda$ we can write the model (1.1) as

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N_2 \left( \begin{pmatrix} x_i \\ \theta_1 + \theta_2 x_i \end{pmatrix} ; \sigma^2 I_2 \right) \quad i = 1, ..., n \quad (2.1)$$

The case with $\theta_1 = 0$ is referred to as the homogeneous linear functional model. We will be interested in $\theta_2$ as the primary structural parameter with $X_1, ..., X_n$ and $\theta_1$ treated as nuisance parameters; the analysis is in the context of $\sigma^2$ being known or known approximately.

The general model (2.1) can be transformed to produce effectively the homogeneous case by using an orthogonal transformation (Anderson, 1976). Let $Q = (O_{ij})$ be an $n \times n$ orthogonal matrix with $O_{ij} = n^{-1/2} (j = 1, ..., n)$ and let

$$x_i^* = \Sigma_j Q_{ij} x_j, \quad X_i^* = \Sigma_j Q_{ij} X_j; \quad y_i^* = \Sigma_j Q_{ij} y_j \quad i = 1, ..., n \quad (2.2)$$

The transformed model is

$$\begin{pmatrix} x_1^* \\ y_1^* \end{pmatrix} \sim N_2 \left( \begin{pmatrix} X_1^* \\ \theta_2 X_1^* \end{pmatrix} ; \sigma^2 I_2 \right)$$

$$\begin{pmatrix} x_n^* \\ y_n^* \end{pmatrix} \sim N_2 \left( \begin{pmatrix} X_n^* \\ \theta_1 n^{1/2} + \theta_2 X_n^* \end{pmatrix} ; \sigma^2 I_2 \right) \quad (2.3)$$

Note that the first component is homogeneous with $n$ replaced by $n-1$ and the second component has $\theta_2$ location-aliased with the nuisance parameter $\theta_1$. Accordingly we restrict our attention to the homogeneous case, which can be obtained in general by the preceding analysis.

Now consider the homogeneous model (2.1) with $\theta_1 = 0$. This simpler model exhibits the natural geometrical characteristics discussed in Dobriyal, Fraser and Geotys (1987). Let

$$X_i = \rho_1 \cos \theta$$

$$\theta_2 X_i = \rho_1 \sin \theta \quad (2.4)$$

the special version of (2.1) can then be rewritten as

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \rho_1 \cos \theta \\ \rho_1 \sin \theta \end{pmatrix} ; \sigma^2 I \right) \quad i = 1, ..., n \quad (2.5)$$
The primary parameter is now taken to be the angle $\theta$ rather than the slope $\theta_2 = \tan \theta$, and the nuisance parameters appear as $\rho_1, \ldots, \rho_n$. There is a small technical point in that $\theta \mod(-\pi/2, \pi/2)$ is one-one equivalent to $\theta_2$ and thus $\theta$ itself includes a binary component from the nuisance parameters. Our analysis in the next section focuses on the reduced $\theta$ but $\theta$ itself constitutes the geometrically relevant location parameter.

In the modified form (2.5) it can be shown that $\theta$ is orthogonal (Cox and Reid, 1987) to the redefined nuisance parameters $\rho_1, \ldots, \rho_n$.

3. MODEL ANALYSIS

We now consider the statistical analysis of the reformulated homogeneous linear structural model (2.5) with $\theta$ as the parameter of interest.

First we examine the special case in which the radial parameters $\rho_1, \ldots, \rho_n$ are known. Let
\begin{align}
x_i &= r_i \cos a_i \quad i = 1, \ldots, n \\
y_i &= r_i \sin a_i
\end{align}

and
\begin{align}
f(e \mid r; \rho) &= (2\pi I_0(\kappa))^\frac{1}{2} \exp\{\kappa \cos e\} \\
h(r; \rho) &= \exp\left\{ -\frac{\rho^2}{2} \sum_{j=0}^{\infty} \frac{\rho^2/2}{j!} g(r^2, 2 + 2j) 2r \right\}
\end{align}

where $g(r^2, f)$ is the chi-square(f) density, $I_0(\cdot)$ is the imaginary Bessel function of order 0,
\begin{align}
I_0(\kappa) = \sum_{j=0}^{\infty} (\kappa^2/4)^j / (j!)^2,
\end{align}

and $f(e \mid r; \rho)$ is the von Mises density with precision parameter, $\kappa = r_\rho / \sigma^2$. The density function for the transformed variables $a_1, \ldots, a_n, r_1, \ldots, r_n$ is
\begin{align}
\prod_{i=1}^{n} f(a_i \mid \theta - \theta_i, \rho_i) \prod_{i=1}^{n} h(r_i; \rho_i)
\end{align}

where the $\kappa_i = r_i \rho_i / \sigma^2$.

The model (3.4) can be viewed as a location model similar to that in Fisher (1956) or as a structural model similar to the measurement model in Fraser (1968, 1979). The indicated conditional analysis uses the reduced model
\begin{align}
g(\hat{\theta} - \theta) d\hat{\theta} = k \prod_{i=1}^{n} f(d_i \mid \hat{\theta}^2 - \theta_i, r_i; \rho_i) d\hat{\theta}
\end{align}
for the maximum likelihood estimate \( \hat{\theta} \) on \( (-\pi, \pi) \), where the \( d_0^0 = a_0^0 - \hat{\theta}_0 \) give the deviations of the observed \( a_i \) values from the observed maximum likelihood \( \hat{\theta} \) value. The one-dimensional distribution (3.5) yields a \((1-\alpha)\) confidence interval for \( \theta \mod(\pm \pi/2) \) by simple numerical integration,

\[
\hat{\theta}_0^0 \pm c_\alpha
\]

where

\[
\int_{-c_\alpha}^{c_\alpha} [g(e) + g(e+\pi)]de = 1 - \alpha
\]

for \( e \) examined in the range \((-\pi/2, \pi/2)\).

Now consider the more general problem with \( \rho_1, ..., \rho_n \) as nuisance parameters.

An approach indicated by Cox and Reid (1987) is to condition on the mle (maximum likelihood estimate) of the orthogonalized nuisance parameters. For the present problem this means conditioning on the \( r_1, ..., r_n \) and gives the distribution (3.5); then for calculations a natural choice for the nuisance parameters involves replacing \( \rho_1, ..., \rho_n \) by appropriate maximum likelihood or bias-adjusted maximum likelihood estimates derived from the noncentral chi-square densities \( h(r_i; \rho_i) \).

An alternative approach by Fraser and Reid (1987) involves a differential development of a one-dimensional conditional distribution using likelihood difference arguments. Again this leads to (3.5) and for implementation also requires the use of estimates for the nuisance parameters \( \rho_1, ..., \rho_n \).

We further develop from these two approaches by using the conditional distribution formula (3.5) but replacing each \( f(e_i \mid r_i; \rho) \) by a confidence-distribution average over the possible values for \( \rho_i \):

\[
\bar{f}(\rho_i \mid r_i) = \frac{1}{2\pi} (1 - G(r^2, 2))
\]

\[
+ \sum_{j=0}^{\infty} \frac{(1/2)^j}{j!} \left[ G(r^2, 2+2j) - G(r^2, 4+2j) \right]
\]

\[
\times \frac{1}{4\pi} \int_0^\infty t^2 \rho^{2j} \exp(\rho \cos e_i - \rho^2/2) d\rho^2
\]

where \( G(r^2, f) \) is the chi-square distribution function with \( f \) degrees of freedom; for details see Fraser and Gebotys (1984) and Dobriyal, Fraser and Gebotys (1987); formula (3.7) applies for the case \( \sigma^2 = 1 \). The averaged distribution (3.7) can be calculated on a grid of \( e_i \) values and then compounded as in (3.5) giving

\[
\bar{g}(\theta - \hat{\theta})d\theta = k \Pi_i \bar{f}(d_i^0 + (\hat{\theta} - \theta) \mid r_i) d\delta
\]

A \((1-\alpha)\) confidence region can then be obtained,

\[
\hat{\theta}_0^0 \pm c_\alpha
\]
by the numerical integration

\[
\int_{0}^{\pi} \left[ g(e) + \bar{e}(e+\pi) \right] \, de = 1 - \alpha \tag{3.10}
\]

for \( e \) in \((-\pi/2, \pi/2)\).

To accommodate the simplification \( \sigma^2 = 1 \) we assume in the preceding that the original data has been divided through by the known value for \( \sigma \).

As the distribution (3.8) is typically asymmetrical, an equal-tailed confidence interval

\[
(\bar{\theta} - e_U, \quad \bar{\theta} - e_L) \tag{3.11}
\]

may be preferred where

\[
\int_{-\pi/2}^{\pi/2} \left[ g(e) + \bar{e}(e+\pi) \right] \, de = \alpha/2 \tag{3.12}
\]

\[
\int_{-\pi/2}^{\pi/2} \left[ g(e) + \bar{e}(e+\pi) \right] \, de = \alpha/2 .
\]

4. AN EXAMPLE

We now examine the Cushny-Pebbles data on the effect of two drugs on hours of sleep gained by ten patients. The data are discussed in Student (1908):

<table>
<thead>
<tr>
<th>Drug I</th>
<th>Drug II</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x )</td>
<td>( y )</td>
</tr>
<tr>
<td>1.9</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>-1.6</td>
</tr>
<tr>
<td>1.1</td>
<td>-0.2</td>
</tr>
<tr>
<td>0.1</td>
<td>-1.2</td>
</tr>
<tr>
<td>-0.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>4.4</td>
<td>3.4</td>
</tr>
<tr>
<td>5.5</td>
<td>3.7</td>
</tr>
<tr>
<td>1.6</td>
<td>0.8</td>
</tr>
<tr>
<td>4.6</td>
<td>0.0</td>
</tr>
<tr>
<td>3.4</td>
<td>2.0</td>
</tr>
</tbody>
</table>
Fieller (1954) (also see Creasy, 1956) used the homogeneous linear functional model (2.1) with $\theta_1 = 0$ or (2.5). We use the same model and derive a conditional confidence interval (3.11) using the methods discussed in the preceding section.

An estimate of $\sigma^2$ can be obtained by maximum likelihood together with a degrees-of-freedom adjustment for bias,

$$\hat{\sigma}^2 = \frac{\Sigma_i [y_i - \hat{\theta}_2 x_i]^2}{(n-1)(1+\hat{\theta}_2^2)},$$

giving $\hat{\sigma}^2 = 1.22$. We use this estimate to standardize the data. The observed radial and angular values are

<table>
<thead>
<tr>
<th>$r^0_i$</th>
<th>$a^0_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.83</td>
<td>0.3530</td>
</tr>
<tr>
<td>1.62</td>
<td>-1.1071</td>
</tr>
<tr>
<td>1.01</td>
<td>-0.1798</td>
</tr>
<tr>
<td>1.09</td>
<td>-1.4876</td>
</tr>
<tr>
<td>0.13</td>
<td>-2.3562</td>
</tr>
<tr>
<td>5.03</td>
<td>0.6579</td>
</tr>
<tr>
<td>6.00</td>
<td>0.5922</td>
</tr>
<tr>
<td>1.62</td>
<td>0.4636</td>
</tr>
<tr>
<td>4.16</td>
<td>0.0000</td>
</tr>
<tr>
<td>3.57</td>
<td>0.5317</td>
</tr>
</tbody>
</table>

The conditional location analysis using (3.11) produces the following confidence intervals for $\theta$

95% : (0.2262, 0.7759)
99% : (0.0781, 0.9279)

the corresponding intervals for $\theta_2$ are

95% : (0.2301, 0.9812)
99% : (0.0783, 1.3350).

The maximum likelihood estimate for $\theta_2$ is 0.5432. There is thus evidence that Drug I provides more additional hours of sleep than Drug II.

Creasy (1956), using a t-type pivotal, obtained the $\theta_2$ intervals:

95% : (0.2284, 0.9710)
99% : (0.0766, 1.3100).

Sprott and Viveros (1983) using a linear pivotal approximation obtained the 95% interval (0.2255, 0.9765).
Other model assumptions have been used and have produced different results. Fisher (see Sprott and Viveros, 1984) assumed that there is an $\alpha$ such that $y_1 - \alpha x_1$ is normal$(0, \sigma^2)$. The resulting 95% confidence interval is $(-0.4848, 0.6566)$. Also see Hunter and Lamboy (1979a,b) for a Bayesian analysis.

5. DISCUSSION

We have derived an approximate conditional confidence interval for the structural parameter of the two variable linear structural model. The conditioning makes the interval sensitive to measures of precision available in the data and avoids the sometimes 100%, sometimes undercoverage problems with the usual unconditional intervals.

The example in Section 4 involves moderately high precision with a corresponding low probability for the overcoverage, undercoverage anomaly. The interval can be expected to more carefully reflect the precision implied by the data pattern.

REFERENCES


Received July, 1987.