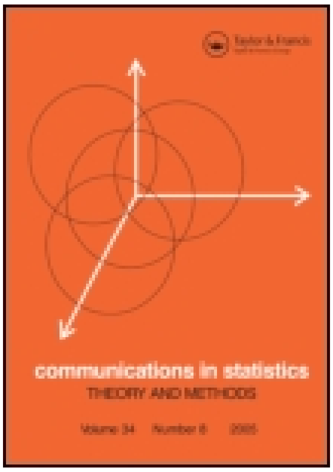

Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

A study of linear model alysis by an error contour emulation

Gordon H. Fick [a]

[a] University of Calgary , Calgary , Canada

Published online: 27 Jun 2007.

COMMUN. STATIST.-THEOR. METH., 13(16), 2029-2047 (1984)

A STUDY OF LINEAR MODEL ANALYSIS BY AN ERROR CONTOUR EMULATION

Gordon H. Fick

University of Calgary
Calgary, Canada

Key Words and Phrases: conditionality; likelihood function; linear model; marginal likelihood function, necessary analysis; non-normal analysis.

ABSTRACT

This paper considers the conditional approach to linear models in which the exact theoretical results are unavailable except in terms of multiple integrals. A class of multidimensional error distributions that emulate elongated error distributions is discussed. The appropriate conditional distributions are derived along with several properties of these distributions.

1. INTRODUCTION

Consider a system in which a linear model of the following form is appropriate:

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \sigma \underset{\sim}{z}$$

where $\underset{\sim}{y}$ displays a response vector of n observations obtained at input levels recorded in the k columns of X. The vector $\underset{\sim}{z}$ displays the errors which follow some distribution with density function $f_\lambda(\underset{\sim}{z})$.

0361-0926/84/1316-2029$3.50/0

Our interest centres around the location of the response: $X\underset{\sim}{\beta}$, and the scaling of the response: σ. In addition, we wish to study λ which indexes the distributions for the error form.

A traditional linear model analysis assumes that the error density f_λ is known to be standard normal and leads to the familiar distribution theory for the regression coefficients and the standard deviation. Confidence regions and tests of significance can be obtained directly from this distribution theory.

Similar conditional distributions can be derived for any error density f_λ, but such distributions are usually available only in terms of multiple integrals. Typically, the properties of inference procedures derived from these distributions can be assessed only through simulation and approximation.

The main references for this topic are section 3.4 in Box and Tiao (1973), section 11.3 in Fraser (1976), section 6.4 in Fraser (1979) and Sprott (1980).

A common form of nonnormality is extended tail length or elongation (as described in Tukey (1977)). There are many parametric classes of densities which reflect varying degrees of elongation. The student class is given by

$$f_\lambda(\underset{\sim}{z}) \propto \prod_{i=1}^{n} \left\{1 + z_i^2/\omega^2(\lambda)\right\}^{-\frac{(\lambda+1)}{2}}, \quad \lambda \in (0, +\infty).$$

The exponential power class is given by

$$f_\lambda(\underset{\sim}{z}) \propto \prod_{i=1}^{n} exp\left\{-\frac{1}{2}\,|z_i/\omega(\lambda)|^\lambda\right\}, \quad \lambda \in (0, 2].$$

(The functions$\omega(\lambda)$ are normally chosen to give $\underset{\sim}{\beta}$ and σ fixed interpretations that do not depend on λ).

The shape of the n dimensional contours of f_λ determines the form of the derived conditional distributions. There is increasing empirical evidence derived from simulation studies to suggest that the form of the n dimensional contours near the coordinate axes in $\mathcal{R}^n$ has the most dominant effect on the derived

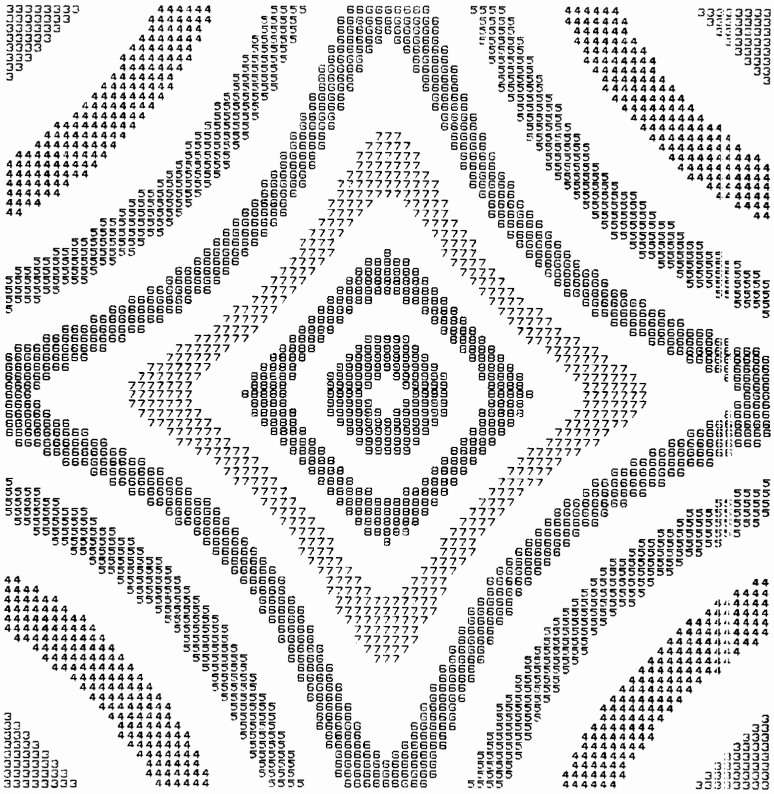

Figure 1

Contours of the bivariate student class with $\lambda = 3$

distributions. It is these derived distributions that determine the inferences obtained from the model and the data. (Fick (1978)).

In this article, a general class of distributions is exhibited that can be used to emulate the error contours of long tailed distributions. The main advantage of this emulating class of distributions is that the desired conditional distributions can be derived analytically. (All of the required integrations can be made without the need for numerical integrations).

Figure 1 and Figure 2 illustrate the effect of elongation

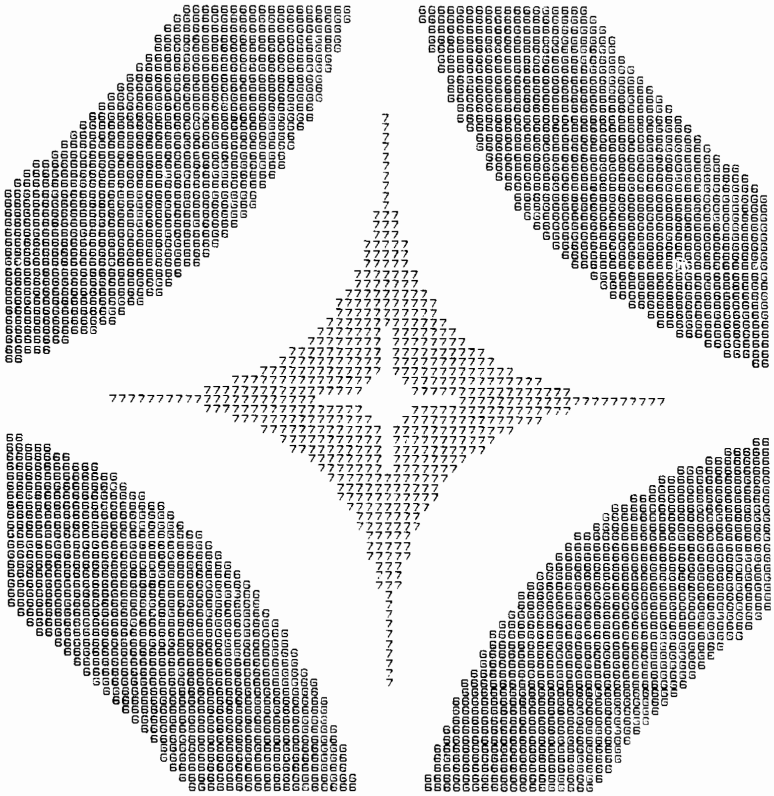

Figure 2

Contours of the bivariate exponential power class with $\lambda = 0.5$

on 2 dimensional contours of distributions f_λ (Figure 1 shows the student density with $\lambda = 3$, Figure 2 shows the exponential power density with $\lambda = 0.5$). The dominant characteristic of these contours is the manner in which they protrude in the directions of the coordinate axes. A typical contour will be closest to the origin at a 45 degree angle to any coordinate axis.

2. THE EMULATING DISTRIBUTIONS

The student class of distributions and the exponential power class of distributions discussed in the introduction are examples of long tailed (or elongated) densities that form the basis of a considerable amount of research into the robustness of conditional inference methods. In each class, there is a need for a multi-dimensional integration (A discussion of the integration required to obtain the conditional distributions is presented in section 3). There seem to be very few examples available in the literature, in which the integration can be completely carried out analytically. If $\underset{\sim}{z} \sim N\,(\underset{\sim}{0},\,\Sigma)$ $|\Sigma| > 0$, all of the required integrations can in fact, be made. The derivations are reviewed in section 4. A related result can be found in Fraser, Guttman and Styan (1976).

We can try to copy the protruding contours illustrated in Figures 1 and 2 with appropriately chosen normal distributions. For example, the lobe-like contour along the z_1 axis will be similar to the elliptical contour found in

$$(2\pi)^{-n/2}\,\tau_1^{-\frac{1}{2}}\prod_{i=2}^{n} exp\left\{-\,z_i^2/2\right\}\,exp\left\{-\,z_1^2/2\tau_1\right\}$$

$$\left\{\text{call this density } f_{\Sigma_1},\ \Sigma_1 = diag\,(\tau_1,\,1,\,1,\,\ldots,\,1)\right\}.$$

To gain an approximation to the lobe-like contours along all n coordinate axes, we can consider the density:

$$\sum_{i=1}^{n} c_i\,f_{\Sigma_i}\,(\underset{\sim}{z}),\quad \sum_{i=1}^{n} c_i = 1,\ c_i \geq 0$$

where:

$$f_{\Sigma_k}\,(\underset{\sim}{z}) = (2\pi)^{-n/2}\,\tau_k^{-\frac{1}{2}}\prod_{j\neq k} exp\left\{-\,z_j^2/2\right\}\,exp\left\{-\,z_k^2/2\tau_k\right\}$$

$$= (2\pi)^{-n/2}\,|\,\Sigma_k\,|^{-\frac{1}{2}}\,exp\left\{-\underset{\sim}{z}'\,\Sigma_k^{-1}\,\underset{\sim}{z}/2\right\}$$

with $\Sigma_k = diag\,(1,\,1,\,\ldots,\,\tau_k,\,1,\,\ldots,\,1)$

(and τ_k is in the k^{th} position of the diagonal).

This density has nearly ellipsoidal contours near each coordinate axis and has a shape that is similar to those in Figures 1 and 2. Figures 3 and 4 show 2 dimensional contours of

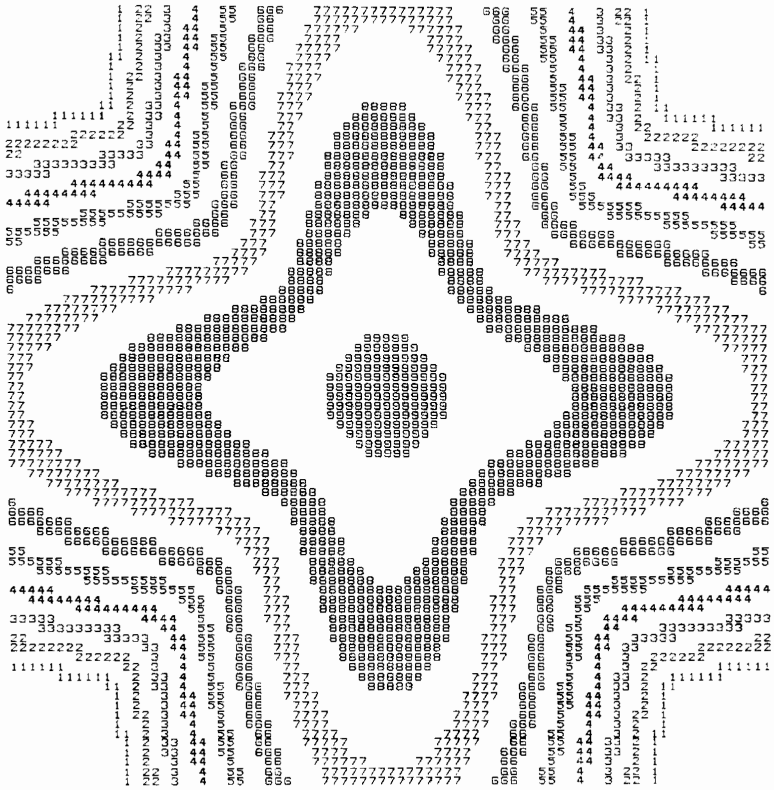

Figure 3

Contours of a sum of 2 rescaled normal distributions
$$\left(\tau_1 = \tau_2 = 9;\ c_1 = c_2 = \frac{1}{2}\right)$$

typical densities $\sum_{i=1}^{n} c_i f_{\Sigma_i}$. It is the shape of the contours that we wish to compare. Note that symmetry in f_λ appears to require that each $c_i = n^{-1}$ and all τ_k be equal. The formulae that will be derived in later sections are no more complicated by considering general c_i and τ_k and there are some indications that, for purposes of simulation and Monte Carlo studies, it may be more efficient to consider varying c_i and τ_k. Some brief

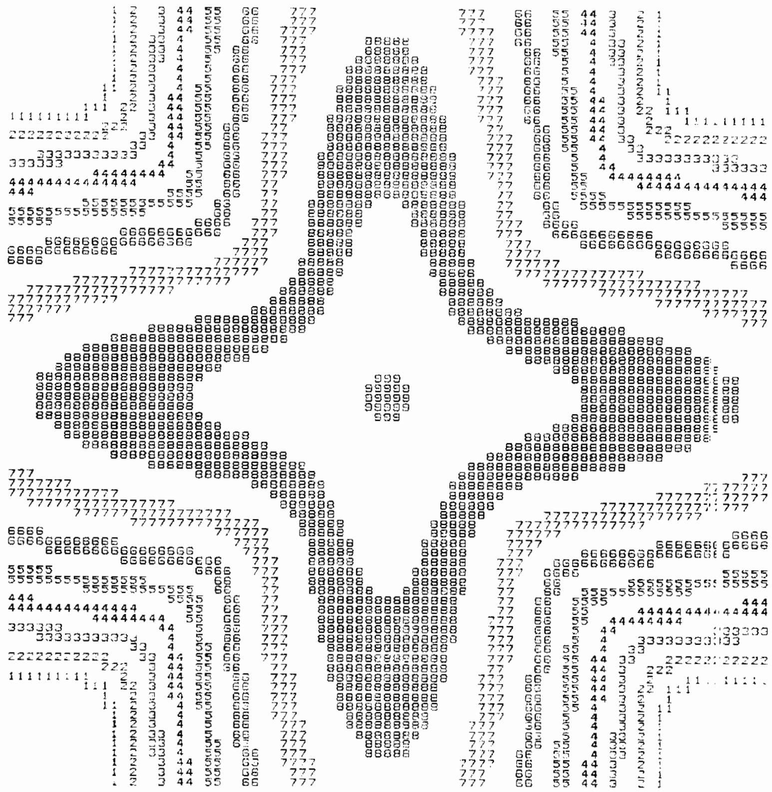

Figure 4

Contours of a sum of 2 rescaled normal distributions

$$\left\{\tau_1 = \tau_2 = 10;\ c_1 = c_2 = \frac{1}{2}\right\}$$

comments on the selection of c_i and τ_k will be made at the end of this article.

3. A MOTIVATION AND SUMMARY OF THE CONDITIONAL DISTRIBUTIONS

For purposes of exposition, it is convenient to place the linear model in a canonical form. The $n \times k$ matrix X can be written as $X = VU$ where V is an $n \times r$ (column) orthonormal matrix ($V'V = I$) and U is an $r \times k$ upper triangular matrix with $r \leq k$. Such a decomposition can be obtained by one of a variety of methods including Gram-Schmidt, Householder, Givens and others.

The model can then be presented as

$$\underset{\sim}{y} = V\,\underset{\sim}{\alpha} + \sigma\,\underset{\sim}{z}$$

where $\underset{\sim}{\alpha} = U\,\underset{\sim}{\beta}$. We of course have that $L(V) = L(X)$ [where $L(V)$ is the linear space spanned by the columns of V]. It is assumed that $n > r + 1$. If this model is appropriate in a design setting, then X might already contain a set of orthogonal contrasts, in which case U is diagonal. Other reasons for presenting the model in an orthonormal form will appear in the next section.

For the moment, consider the analysis of a model with standard normal errors:

$$f_\lambda(\underset{\sim}{z}) = f(\underset{\sim}{z}) = exp\left\{-\frac{1}{2}\,\underset{\sim}{z}'\underset{\sim}{z}\right\}/(2\pi)^{n/2} .$$

We could present the response vector $\underset{\sim}{y}$ as

$\underset{\sim}{y} = V\,\underset{\sim}{a}(\underset{\sim}{y}) + s(\underset{\sim}{y})\,\underset{\sim}{d}(\underset{\sim}{y})$, where:

$\underset{\sim}{a}(\underset{\sim}{y}) = V'\underset{\sim}{y}$ (the regression coefficients)

$s^2(\underset{\sim}{y}) = \|\,\underset{\sim}{y} - V\,\underset{\sim}{a}(\underset{\sim}{y})\,\|^2$ (the squared length of the residual)

and $\underset{\sim}{d}(\underset{\sim}{y}) = \dfrac{1}{s(\underset{\sim}{y})}\,(\underset{\sim}{y} - V\,\underset{\sim}{a}(\underset{\sim}{y}))$ (the unit residual vector).

We normally consider inference for $\underset{\sim}{\alpha}$ based on

$$(\underset{\sim}{a}(\underset{\sim}{y}) - \underset{\sim}{\alpha})/s(\underset{\sim}{y})$$

which has an r-variate Student (n-r) distribution with density function:

$$\frac{A_{n-r}}{A_n}\,(1 + \underset{\sim}{T}'\underset{\sim}{T})^{-n/2} \qquad \underset{\sim}{T} \in \mathbb{R}^r$$

$$\left\{A_f = 2\pi^{f/2}/\Gamma(f/2)\right\}.$$

Inference concerning σ would be based on

$$s(\underset{\sim}{y})/\sigma$$

which has a chi (n-r) distribution with density function:

$$\Gamma^{-1}\left(\frac{n-r}{2}\right) exp\left\{-\frac{1}{2}s^2\right\} s^{n-r-1}\left(\frac{1}{2}\right)^{\frac{n-r}{2}-1} \qquad s \in R^+$$

These distributions can also be derived directly from the error variable $\underset{\sim}{z}$. Indeed, the linear model

$$\underset{\sim}{y} = V\underset{\sim}{\alpha} + \sigma\underset{\sim}{z}$$

displays the response variable $\underset{\sim}{y}$ as a relocated and rescaled version of $\underset{\sim}{z}$. As we did with the response vector $\underset{\sim}{y}$, we can also present the variation $\underset{\sim}{z}$ as

$$\underset{\sim}{z} = V\underset{\sim}{a}(\underset{\sim}{z}) + s(\underset{\sim}{z})\,\underset{\sim}{d}(\underset{\sim}{z}) \,..$$

We now notice that

$\underset{\sim}{d}(\underset{\sim}{z}) = \underset{\sim}{d}(\underset{\sim}{y})$ (we will write just $\underset{\sim}{d}$ from now on).

When $\underset{\sim}{y}$ is observed, we can record the information concerning the unknown $\underset{\sim}{z}$ in the form $\underset{\sim}{d}(\underset{\sim}{z}) = \underset{\sim}{d}$.

Now notice that

$$\begin{aligned}\underset{\sim}{y} &= V\underset{\sim}{a}(\underset{\sim}{y}) + s(\underset{\sim}{y})\underset{\sim}{d} = V\underset{\sim}{\alpha} + \sigma\underset{\sim}{z}\\ &= V\underset{\sim}{\alpha} + \sigma(V\underset{\sim}{a}(\underset{\sim}{z}) + s(\underset{\sim}{z})\underset{\sim}{d})\\ &= V(\underset{\sim}{\alpha} + \sigma\underset{\sim}{a}(\underset{\sim}{z})) + \sigma\, s(\underset{\sim}{z})\underset{\sim}{d}\end{aligned}$$

so that

$$\underset{\sim}{a}(\underset{\sim}{y}) = \underset{\sim}{\alpha} + \sigma\underset{\sim}{a}(\underset{\sim}{z})$$

$$s(\underset{\sim}{y}) = \sigma\, s(\underset{\sim}{z})$$

and

$$\underset{\sim}{a}(\underset{\sim}{z})/s(\underset{\sim}{z}) = (\underset{\sim}{a}(\underset{\sim}{y}) - \underset{\sim}{\alpha})/s(\underset{\sim}{y})$$

$$s(\underset{\sim}{z}) = s(\underset{\sim}{y})/\sigma\,.$$

When using regression analysis based on standard normal errors, the distributions used with our inferences for $\underset{\sim}{\alpha}$ and σ can be interpreted as being derived from the distribution for $\underset{\sim}{z}$.

It appears that we have not used the information supplied by $\underset{\sim}{y}$ that informs us that

$$\underset{\sim}{d}(\underset{\sim}{z}) = \underset{\sim}{d}\ .$$

But when $\underset{\sim}{z}$ is standard normal, $\underset{\sim}{a}(\underset{\sim}{z})$, $s(\underset{\sim}{z})$ and $\underset{\sim}{d}(\underset{\sim}{z})$ are actually statistically independent and the distribution for $\underset{\sim}{d}$ is uniform on the unit sphere in $L^{\perp}(V)$ given by

$$\frac{1}{A_{n-r}} \qquad \underset{\sim}{d} \in \text{unit sphere in } L^{\perp}(V).$$

($L^{\perp}(V)$ is the orthogonal complement to $L(V)$).

Now suppose the density for $\underset{\sim}{z}$ is not necessarily normal but is in fact

$$f_{\lambda}(\underset{\sim}{z}).$$

We now have $r + 2$ parameters to study $(\underset{\sim}{\alpha}, \sigma, \lambda)$.

The necessary distributions for inference are recorded below for reference in later sections. They are derived in, for example, Fraser (1976) Section 11.3. The unit residual vector $\underset{\sim}{d}$ records the observable value of $\underset{\sim}{d}(\underset{\sim}{z})$ and $(\underset{\sim}{a} = \underset{\sim}{a}(\underset{\sim}{z}),\ s = s(\underset{\sim}{z}))$ records the unobservable part of $\underset{\sim}{z}$. It is argued from many perspectives that all inferences should be based on the marginal distribution for $\underset{\sim}{d}$ and the conditional distribution for $(\underset{\sim}{a}, s)$ given $\underset{\sim}{d}$.

The joint distribution for $(\underset{\sim}{a}, s, \underset{\sim}{d})$ is:

$$f_{\lambda}(V\underset{\sim}{a} + s\underset{\sim}{d})\ s^{n-r-1} \quad .$$

The marginal for $\underset{\sim}{d}$ is:

$$h_{\lambda}(\underset{\sim}{d}) = \int_{\mathbb{R}^r}\int_{\mathbb{R}^+} f_{\lambda}(V\underset{\sim}{a} + s\underset{\sim}{d})\ s^{n-r-1}\, d\underset{\sim}{a}\, d s \ .$$

To assess the parameter λ we have a likelihood function, L, for λ:

$$L(\underset{\sim}{d}|\lambda) \propto h_{\lambda}(\underset{\sim}{d}) \ .$$

For inference concerning $\underset{\sim}{\alpha}$, we have

$$(\underset{\sim}{a}(\underset{\sim}{y}) - \underset{\sim}{\alpha})/s(\underset{\sim}{y}) = \underset{\sim}{a}(\underset{\sim}{z})/s(\underset{\sim}{z}) = \underset{\sim}{T}$$

together with the conditional distribution for $\underset{\sim}{T}$ given $\underset{\sim}{d}$ that has density:

$$h_{\lambda}^{-1}(\underset{\sim}{d}) \int_{\mathbb{R}^+} f_{\lambda}(s(V\underset{\sim}{T} + \underset{\sim}{d}))\ s^{n-1}\ ds$$

$$\text{(call it } g_\lambda^L(\underset{\sim}{T}|\underset{\sim}{d})\text{)}.$$

For inference concerning σ we have that

$$s(\underset{\sim}{y})/\sigma = s(\underset{\sim}{z}) = s$$

together with the conditional distribution for s given $\underset{\sim}{d}$ that has density:

$$h_\lambda^{-1}(\underset{\sim}{d}) \int_{\mathbb{R}^r} f_\lambda(s(V\underset{\sim}{T} + \underset{\sim}{d}))\, s^{n-1}\, d\underset{\sim}{T}$$

$$\text{(call it } g_\lambda^S(s|\underset{\sim}{d})\text{)}.$$

These two distributions are considered primary by many authors in statistical inference although authors do disagree on the motivation and principles involved in the derivation and interpretation of them.

Several authors have evaluated these distributions numerically (Fraser (1976), Box and Tiao (1973), Efron and Hinkley (1978), Barnard (1974), with the extensive use of numerical quadrature and Monte Carlo). Empirical studies into the properties of such distributions have been considered by Fraser (1979), and Box and Tiao (1973).

The direct approximation of these distributions has been considered by Lund (1967) and Sprott (1980). When the rank of X is equal to 1, the distributions can be obtained with high numerical accuracy and efficiency. It is when the rank is greater than one that there are many obstacles related to quadrature that remain to be solved in order to obtain these distributions numerically.

4. THE DERIVATION OF THE CONDITIONAL DISTRIBUTIONS USING THE EMULATING ERROR FORMS

This section begins with the derivations of the appropriate distributions when the error distribution is $N(\underset{\sim}{0},\Sigma)$. We will then use these results to derive the desired distributions from the emulating error distributions described in section 2.

Consider the model

$$\underset{\sim}{y} = V\underset{\sim}{\alpha} + \sigma\underset{\sim}{z}$$

where $\underset{\sim}{z}$ is distributed with density:

$$f_{\Sigma}(\underset{\sim}{z}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\underset{\sim}{z}'\,\Sigma^{-1}\,\underset{\sim}{z}\right\} .$$

(Here, Σ can be an arbitrary positive definite convariance matrix.)

The marginal for $\underset{\sim}{d}$, $h_{\Sigma}(\underset{\sim}{d})$ has the form

$$\int_{\underset{\sim}{a}}\int_{s} \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(V\underset{\sim}{a}+s\underset{\sim}{d})'\Sigma^{-1}(V\underset{\sim}{a}+s\underset{\sim}{d})\right\} s^{n-r-1}\, d\underset{\sim}{a}\, ds$$

$$= \int_{\underset{\sim}{a}}\int_{s} \frac{\exp\left\{-\frac{1}{2}(\underset{\sim}{a}+s(V'\Sigma^{-1}V)^{-1}V'\Sigma^{-1}\underset{\sim}{d})'(V'\Sigma^{-1}V)(\underset{\sim}{a}+s(V'\Sigma^{-1}V)^{-1}V'\Sigma^{-1}\underset{\sim}{d})\right\}}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}$$

$$\cdot \exp\left\{-\frac{1}{2}s^2\left[\underset{\sim}{d}'[\Sigma^{-1}-\Sigma^{-1}\, V(V'\Sigma^{-1}V)^{-1}\, V'\Sigma^{-1}]\underset{\sim}{d}\right]\right\} s^{n-r-1}\, d\underset{\sim}{a}\, ds$$

$$= \int_{s} \frac{|V'\Sigma^{-1}V|^{-\frac{1}{2}}}{(2\pi)^{\frac{n-r}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\underset{\sim}{d}'R\underset{\sim}{d})s^2\right\} s^{n-r-1}\, ds$$

where $R = \Sigma^{-1} - \Sigma^{-1}\, V(V'\Sigma^{-1}V)^{-1}\, V'\Sigma^{-1}$.

Now if $u = \frac{1}{2}\underset{\sim}{d}'R\underset{\sim}{d}\; s^2$ then we obtain

$$= \int_{u} \frac{|V\Sigma^{-1}V|^{-\frac{1}{2}}}{(2\pi)^{\frac{n-r}{2}}|\Sigma|^{\frac{1}{2}}} \cdot \frac{2^{\frac{n-r}{2}-1}}{(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}} \exp(-u)\; u^{\frac{n-r}{2}-1}\, du$$

$$= \frac{1}{A_{n-r}} \cdot \frac{|V'\Sigma^{-1}V|^{-\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} \cdot \frac{1}{(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}} .$$

The conditional density for $(\underset{\sim}{a},s)$ given $\underset{\sim}{d}$ is now easily seen to be:

$$\frac{A_{n-r}|V'\Sigma^{-1}V|^{\frac{1}{2}}\,(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(V\underset{\sim}{a}+s\underset{\sim}{d})'\Sigma^{-1}(V\underset{\sim}{a}+s\underset{\sim}{d})\right\} s^{n-r-1} .$$

The conditional density for s given $\underset{\sim}{d}$ is

$$\frac{(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}}{\Gamma\left(\frac{n-r}{2}\right)} \, exp\left\{-\frac{1}{2}(\underset{\sim}{d}'R\underset{\sim}{d})s^2\right\} s^{n-r-1} \left(\frac{1}{2}\right)^{\frac{n-r}{2}-1}$$

The conditional for $\underset{\sim}{T}$ given $\underset{\sim}{d}$ is given by

$$\int_s A_{n-r} \frac{|V'\Sigma^{-1}V|^{\frac{1}{2}}(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}}{(2\pi)^{n/2}} \, exp\left\{-\frac{1}{2}s^2(V\underset{\sim}{T}+\underset{\sim}{d})'\,\Sigma^{-1}\,(V\underset{\sim}{T}+\underset{\sim}{d})\right\} s^{n-1}\, ds$$

$$= \frac{A_{n-r}}{A_n} |V'\Sigma^{-1}V|^{\frac{1}{2}} \, (\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}} \cdot \frac{1}{\left((V\underset{\sim}{T}+\underset{\sim}{d})'\Sigma^{-1}(V\underset{\sim}{T}+\underset{\sim}{d})\right)^{n/2}} \, .$$

This distribution is, in fact, a relocated and rescaled Student $(n-r)$ on $\mathcal{R}^r$ given by

$$\frac{A_{n-r}}{A_n} |W|^{-\frac{1}{2}} \, (1 + (\underset{\sim}{T}-\underset{\sim}{\mu})' W^{-1} \, (\underset{\sim}{T}-\underset{\sim}{\mu}))^{-n/2}.$$

We will write

$$\underset{\sim}{T} \sim \text{Student}_{n-r}(\underset{\sim}{\mu}, W) \text{ on } \mathcal{R}^r \, .$$

To find $\underset{\sim}{\mu}$ and W, we expand the above quadratic form:

$$(V\underset{\sim}{T}+\underset{\sim}{d})'\,\Sigma^{-1}\,(V\underset{\sim}{T}+\underset{\sim}{d})$$

$$= (\underset{\sim}{T} + (V'\Sigma^{-1}V)^{-1}\,V'\Sigma^{-1}\underset{\sim}{d})'\,(V'\Sigma^{-1}V)(\underset{\sim}{T} + (V'\Sigma^{-1}V)^{-1}\,V'\Sigma^{-1}\underset{\sim}{d})$$

$$+ \underset{\sim}{d}'(\Sigma^{-1} - \Sigma^{-1}\,V(V'\Sigma^{-1}V)^{-1}\,V'\Sigma^{-1})\underset{\sim}{d}$$

so that

$$\underset{\sim}{\mu} = -(V'\Sigma^{-1}V)^{-1}\,V'\Sigma^{-1}\underset{\sim}{d}$$

and $W = (\underset{\sim}{d}'R\underset{\sim}{d})(V'\Sigma^{-1}V)^{-1}$.

We now describe how the above results can be applied to the case $\Sigma = \Sigma_k$ as defined in section 2. It is convenient to display the matrix V as a column of rows:

$$V = \begin{pmatrix} \underset{\sim}{v}_1' \\ \cdot \\ \cdot \\ \cdot \\ \underset{\sim}{v}_n' \end{pmatrix} \qquad \underset{\sim}{v}_i' = \left(v_{i1}, v_{i2}, \ldots, v_{ir}\right)$$

When Σ is set equal to Σ_k, it will be seen that the conditional densities for $\underset{\sim}{T}$ and s given $\underset{\sim}{d}$ depend on the k^{th} component of $\underset{\sim}{d}$ only and are functionally independent of all but certain functions of V.

Theorem:

Consider the linear model:

$\underset{\sim}{y} = V\underset{\sim}{\alpha} + \sigma \underset{\sim}{z}$

where $\underset{\sim}{z}$ is distributed with density:

$$\sum_{i=1}^{n} c_i f_{\Sigma_i} \qquad (c_i > 0,\ \Sigma c_i = 1)$$

where $f_{\Sigma_i}(\underset{\sim}{z}) = (2\pi)^{-n/2}|\Sigma_i|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\underset{\sim}{z}'\,\Sigma_i^{-1}\,\underset{\sim}{z}\right\}$

and $\Sigma_i = diag\ (1, \ldots 1, \tau_i, 1, \ldots 1)$

with τ_i in the i^{th} position.

Then

a) The conditional density for $\underset{\sim}{T}$ given $\underset{\sim}{d}$ is

$$\sum_{i=1}^{n} c_i\, h_{\Sigma_i}(\underset{\sim}{d})\, g^{L}_{\Sigma_i}(\underset{\sim}{T}|\underset{\sim}{d}) \Big/ \sum_{i=1}^{n} c_i\, h_{\Sigma_i}(\underset{\sim}{d})$$

b) The conditional density for s given $\underset{\sim}{d}$ is

$$\sum_{i=1}^{n} c_i\, h_{\Sigma_i}(\underset{\sim}{d})\, g^{S}_{\Sigma_i}(s|\underset{\sim}{d}) \Big/ \sum_{i=1}^{n} c_i\, h_{\Sigma_i}(\underset{\sim}{d})$$

with

i) $h_{\Sigma_i}(\underset{\sim}{d}) = A^{-1}_{n-r}\,(1 + \gamma_k\, \underset{\sim}{v}_k'\, \underset{\sim}{v}_k)^{-\frac{1}{2}}\,(\tau_k)^{-\frac{1}{2}}\,(1 + \rho_k\, d_k^2)^{-(n-r)/2}$

ii) $g^{L}_{\Sigma_i}$ - the density for Student $(\underset{\sim}{\mu}_k,\ W_k)$ on $\mathbb{R}^r$

$\underset{\sim}{\mu}_k = -\rho_k\, d_k\, \underset{\sim}{v}_k$

$W_k = (1 + \rho_k\, d_k^2)(I - \rho_k\, \underset{\sim}{v}_k\, \underset{\sim}{v}_k')$

iii) $g^{S}_{\Sigma_k}$ - the density for $(1 + \rho_k\, d_k^2)^{-\frac{1}{2}}\, chi(n-r)$

using $\gamma_k = \tau_k^{-1} - 1$ and $\rho_k = \gamma_k/(1 + \gamma_k\, \underset{\sim}{v}_k'\, \underset{\sim}{v}_k)$

proof:

We begin by evaluating the various matrix expressions when $\Sigma = \Sigma_k$.

1) $V'\Sigma_k^{-1}V = V'(I + \Sigma_k^{-1} - I)V$

$= I + V'\ diag(0, \ldots \gamma_k, \ldots 0)V$

$= I + (\underset{\sim}{0}, \ldots \gamma_k \underset{\sim}{v}_k, \ldots \underset{\sim}{0})V$

$= I + \gamma_k \underset{\sim}{v}_k \underset{\sim}{v}_k'$

2) Using Theorem 1.4.1 (iv) in Srivastava and Khatri (1979),

$(V'\Sigma_k^{-1}V)^{-1} = I - \gamma_k \underset{\sim}{v}_k \underset{\sim}{v}_k' (1 + \gamma_k \underset{\sim}{v}_k' \underset{\sim}{v}_k)^{-1}$

$= I - \rho_k \underset{\sim}{v}_k \underset{\sim}{v}_k'$

3) Using $V'\underset{\sim}{d} = 0$,

$V'\Sigma_k^{-1}\underset{\sim}{d} = \gamma_k d_k \underset{\sim}{v}_k$

4) By a direct calculation,

$(V'\Sigma_k^{-1}V)^{-1}\ V'\Sigma_k^{-1}\underset{\sim}{d} = \rho_k d_k \underset{\sim}{v}_k$

5) Using $\underset{\sim}{d}'\underset{\sim}{d} = 1$

$d'\Sigma_k^{-1}d = 1 + \gamma_k d_k^2$

6) By a direct calculation,

$\underset{\sim}{d}'R_k\underset{\sim}{d} = 1 + \rho_k d_k^2$

7) $(\underset{\sim}{d}'R_k\underset{\sim}{d})\ (V'\Sigma_k^{-1}V)^{-1} = (1 + \rho_k d_k^2)\ (I - \rho_k \underset{\sim}{v}_k \underset{\sim}{v}_k')$

Now, to obtain expression (a) above, we observe that $h_{\Sigma_k}(\underset{\sim}{d})\ g^L_{\Sigma_k}(\underset{\sim}{T}|\underset{\sim}{d})$ is the joint density for $(\underset{\sim}{T},\underset{\sim}{d})$ if the error density is f_{Σ_k}. If the error density is $\Sigma c_i f_{\Sigma_i}$, then $\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d})\ g^L_{\Sigma_i}(T|d)$ records the functional form of the conditional density for $\underset{\sim}{T}$ given $\underset{\sim}{d}$. $\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d})$ is the divisor to ensure that expression (a) integrates to 1.

Expression (b) is obtained in the same way.

5. OBSERVATIONS

Some observations about the results are in order.

a) By presenting the linear model in an orthonormal form, (with $V'V = I$) one can see how the conditional distributions are dependent on $L(V)$. These results are easily presented in terms of $X = VU$ if the original model presented is of primary interest.

b) The marginal h_{Σ_k} is actually a function of d_k^2 only. The dependence on $L(V)$ is entirely in terms of $\underset{\sim}{v}_k' \underset{\sim}{v}_k$.

3) The density $g_{\Sigma_k}^L$ is conditional upon d_k^2 in a form that depends on only $\underset{\sim}{v}_k \underset{\sim}{v}_k'$ and $\underset{\sim}{v}_k' \underset{\sim}{v}_k$.

d) The density $g_{\Sigma_k}^S$ is a rescaled chi distribution that has a scaling factor dependent on d_k^2 and $\underset{\sim}{v}_k' \underset{\sim}{v}_k$.

e) If all τ_k are set equal to 1, we of course obtain the usual distributions as described in section 3 by noting that $\gamma_k = \rho_k = 0$.

f) If τ_k is greater than 1, then $\underset{\sim}{\mu}$ will be in the direction $d_k \underset{\sim}{v}_k$. Clearly the larger the unit residual component d_k, the further $\underset{\sim}{\mu}$ will be from $\underset{\sim}{0}$ (this compares favorably with the empirical results found in Fraser (1979)).

g) Notice that the conditional distributions are weighted sums of the component g densities. The weights are modulated by the marginals h_{Σ_i}.

h) The conditional distributions will typically be asymmetric, unless there is only one $c_i > 0$.

6. CONCLUSION

The form of the conditional distributions obtained from the emulating error form is substantially influenced by the functions h_{Σ_i}. If the unit residual $\underset{\sim}{d}$ is close to the k^{th} coordinate axis, then h_{Σ_k} will be large relative to the other h_{Σ_i}, $i \neq k$. In applications, it seems to be unnecessary to compute the components of conditional distributions with low weights, $c_i\, h_{\Sigma_i}(\underset{\sim}{d})$. These components will not affect the conditional distributions appreciably.

To choose a range of plausible values for $\underset{\sim}{\tau} = (\tau_1, \ldots, \tau_n)'$ one could consult the marginal likelihood function for $\underset{\sim}{\tau}$:

$$L(\underset{\sim}{d}|\underset{\sim}{\tau}) \propto \sum_{i=1}^{n} c_i\, h_{\Sigma_i}(\underset{\sim}{d}) \ .$$

In many applications, it would be meaningful to formally take all $\tau_k = \tau$ equal. A plot of the marginal likelihood in τ would then be possible. A collection of relative likelihood intervals could be computed to aid in the selection of plausible values of τ (see Kalbfleisch (1981) for an extensive discussion of relative likelihood intervals).

A typical analysis would normally begin with all $c_i = \frac{1}{n}$. This choice will often closely emulate the actual error density $\prod_{i=1}^{n} f_\lambda$. This choice may not be the best for emulating the conditional densities, g_λ^L and g_λ^S, however. If $h_{\Sigma_k}(\underset{\sim}{d})$ is considerably larger than the remaining $h_{\Sigma_i}(\underset{\sim}{d})$, an improvement in the emulation of the conditionals is often made by choosing $c_k = 1$. For contour emulation to aid in the understanding of the behaviour of conditional analyses with elongated errors, an analysis should include a range of plausible σ_i values.

ACKNOWLEDGEMENTS

The author would like to thank D.A.S. Fraser, R.J. Mackay and a referee for their suggestions and comments.

This work was supported by the Natural Sciences and Engineering Research Council of Canada through Grant No. A3457.

BIBLIOGRAPHY

Barnard, G.A. (1974), Conditionality, Pivotals and Robust Estimation. Aarhus Univ. Dept. of Theoretical Statistics 61-80.

Box, G.E.P and Tiao, G.C. (1973), *Bayesian Inference in Statistical Analysis*. Reading Mass: Addison-Wesley Co.

Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton University Press.

Efron, B. and Hinkley, D.V. (1978), Assessing the accuracy of the maximum likelihood estimator : Observed versus expected Fisher information. *Biometrika* 45, 457-488.

Fick, G.H. (1978), The analysis of the linear model with general error distributions, Toronto: Ph.D. Thesis, University of Toronto.

Fraser, D.A.S. (1976), *Probability and Statistics : Theory and Applications*. Toronto: DAI Press.

Fraser, D.A.S. (1979), *Inference and Linear Models*, New York: McGraw Hill Inc.

Fraser, D.A.S., Guttman, I., and Styan, G.P.H. (1976), Serial correlation and distributions on the sphere, *Commun. Statis. - Theory and Method* A5, 97-118.

Kalbfleisch, J.G. (1981), *Probability and Statistical Inference* Vol. II, New York: Springer-Verlag Inc.

Lund, D.R. (1967). Parameter estimation in a class of power distributions, Madison Wisc. Ph.D. Thesis, University of Wisconsin.

Sprott, D.A. (1980), Maximum likelihood in small samples:

Estimation in the presence of nuisance parameters. *Biometrika* 67, 515-524.

Srivastava, M.S. and Khatri, C.G. (1979), *An Introduction to Multivariate Statistics*, New York: Elsevier North Holland Inc.

Tukey, J. (1977), *Exploratory Data Analysis*, Reading Mass: Addison-Wisley Co.

Received by Editorial Board member March, 1982; Revised April, 1984.

Recommended by George P. H. Styan, McGill University, Montreal, Canada.

Refereed by Ioannis A. Koutrouvelis, Virginia Commonwealth University, Richmond, VA and Anonymously.