THE ANALYSIS OF THE LINEAR MODEL WITH GENERAL ERROR DISTRIBUTIONS

GORDON HILTON FICK

1978

THE ANALYSIS OF THE LINEAR MODEL WITH GENERAL ERROR DISTRIBUTIONS

by

GORDON HILTON FICK

Department of Statistics

University of Toronto

A Thesis submitted in conformity with
the requirements for
the Degree of Doctor of Philosophy
in the University of Toronto

To Murphy

Acknowledgements

I wish to express my deep appreciation to my supervisor, Professor D.A.S. Fraser, for his constant encouragement, inspiration and ideas during both my graduate and undergraduate days at the University of Toronto.

I would also like to thank Professor D.F. Andrews and Professor I. Guttman for their guidance and suggestions.

I am grateful to many people for their encouragement and support during the development of this thesis. I would like to thank them here:
H.R. Audet, D. Brenner, M.B. Cairns, G. Cybenko, D.B. De Lury, M. Evans, G.L. Fick, R.A. Fisher, J.S. Gillies, J.L. Kelk, L. Lapzcak, H.D. McGowan, A. McIntosh, K-W. Ng, D. Pregibon and L.A. Scott.

Special thanks to Mrs. Fran Mitchell for her outstanding typing.

I also acknowledge with gratitude the financial support I received from the National Research Council of Canada and the Department of Mathematics of the University of Toronto.

Gordon Fick

May 1978.

CONTENTS

CHAPTER 1

INTRODUCTION

A. Motivation

A substantial collection of theoretical results has been developed since the early 1960's based on the presentation of a statistical problem as a structural model. Research is vigorous and exciting. One reason for this is that the assumptions needed to apply the methods that are derived from the model are fewer than more traditional analyses. The major difference is the assumption of normality of the errors. This assumption is no longer needed to obtain exact tests of significance or confidence intervals when a structural model can be used.

The implementation of the analyses has been delayed for various reasons. The most substantial is most of the distribution theory that arises from the analysis involves multiple integration that (except in special cases) cannot be handled analytically.

For reasons that become clear later, it is essentially impossible to develop tables to display percentage points for these distributions along lines similar to the t and F distributions. But analysis with non-normal error forms can be performed. The analysis of each data set requires the computation of distributions that are essentially special to that data. As the complexity of the model increases, the computer time needed to perform the analysis increases. As computers get faster, relative costs will go down. At present, the cost of a statistical analysis usually forms a very small part of large applied research projects. The practical researcher should not be discouraged from spending more than a dollar for the statistical computing needed for a $10,000 research project.

We begin with a summary of the ideas needed to apply the methods. Applying these methods to familiar regression models is straightforward.

In Chapter 5, several of the ideas illustrated in Chapters 3 and 4 are displayed in the general setting of this chapter. Presentation in this form offers a unifying perspective for the tools illustrated in the more specialized situations of Chapters 3 and 4.

The applied reader can proceed quite comfortably to Chapters 2, 3 and 4.

B. The Model

Many commonly used models can be presented in the general framework of a structural model. Consider a response variable Y (with space S). In many applications, we have grounds for a statistical model that displays Y in terms of another variable Z (with space S).

In fact, a realization of Y is obtained through an unknown transformation of a concealed realization of Z.

We symbolize this as

$$Y = \theta Z$$

$$\theta \in G .$$

Where the transformations θ form a group G, the presentation of Z is made explicitly in the model. The statistical problem is to then make inferential statements about the unknown transformation θ.

The variable Z is described by means of a probability distribution. In a general setting, it may be known to come from one of a class of distributions, the characteristics of these distributions describing properties of Y not already covered by θ. We will use λ to index such classes of distributions (with space Λ).

The model can now be presented as

$$Y = \theta Z \qquad \theta \in G$$

$$f_\lambda(Z) \qquad \lambda \in \Lambda \; .$$

It is an error- or variation - based model. It is not equivalent to a traditional response-based model.

C. The Analysis

Now suppose that Y is realized. The model, together with the observed Y is called the inference base. It summarizes all the information available for statistical analysis.

We now summarize the methods available through the use of the inference base alone, methods that *necessarily* follow from the data and the model only. It is the object of this thesis to describe how the *use* of these methods leads to substantial insight in real situations and to also display simulation studies investigating properties of such methods.

We begin by asking, 'What information is available about the unknowns described in the inference base?' Certainly statistical statements about the unknown Z should lead to knowledge about θ since $Y = \theta Z$ and Y is observed. But we can write $Z = \theta^{-1}Y$ for some θ in G. This identifies Z as a point on the orbit GY of the observed Y where

$$GZ = \{gZ : g \in G\} .$$

In fact we can write

$$GZ = GY$$

and thus we obtain the observed value of the function GZ of the variation Z.

There is no differential information concerning where the unknown but realized Z lies on the orbit GY .

Let us summarize our information formally.

$$\begin{array}{ll} Y & \\ Y = \theta Z & \theta \in G \\ f_\lambda(Z) & \lambda \in \Lambda \\ GZ = GY & . \end{array} \qquad (1.1)$$

This summary contains arbitrary ingredients that are unconsciously presented but have no relevant bearing on its use. Specifically, the distribution $f_\lambda(Z)$ is no longer the appropriate distribution to describe Z in light of the information $GZ = GY$.

The correct distribution to describe Z is in fact the *conditional* distribution given the information $GZ = GY$.

We can now formally display these observations as a factorization of the inference base.

We have

a) the marginal model for GZ with its observed value,

b) the conditional distribution given $GZ = GY$ describing Z together with the presentation $Y = \theta Z$ for some θ in G .

The preceding suggests that we examine Z in terms of the orbits GZ and the position of Z on an orbit.

For this it is convenient to choose a reference point on each orbit; let $D(Z)$ be the reference point on the orbit GZ. We can then record the position of Z on the orbit by finding a transformation $[Z]$ in G which generates Z from $D(Z)$

$$Z = [Z]D(Z) \ .$$

For notational simplicity it is convenient to write $Z = gD$ when it is clear that $g = [Z]$ and $D = D(Z)$.[1]

We now describe the marginal distribution for D and the conditional distribution for $[Z]$ given D using the powerful tool of invariant measures.

On the space S in $\mathbb{R}^N$ the familiar Euclidean volume measure is

$$V_N(A) = \int_A dZ \qquad A \subset S \ .$$

Let h denote a transformation in G. We have

[1] We assume the exactness of G and the differentiability of all functions involved (see Fraser (1978)).

$$J_N(h : Z) = \left|\frac{\partial hZ}{\partial Z}\right|_+ \, [2]$$

$$J_N(Z) = J_N\big([Z] : D(Z)\big)$$

where $dZ/J_N(Z)$ is an invariant measure since

$$\int_{hA} \frac{dZ}{J_N(Z)} = \int_A \frac{dZ}{J_N(Z)} \, .$$

On the group G consider the action of G acting on itself (on the left)

$$J_L(h : g) = \left|\frac{\partial hg}{\partial g}\right|$$

$$J_L(g) = J_L(g : i) \qquad (1.2)$$

where i denotes the identity element of G so that

$$\int_{hB} \frac{dg}{J_L(g)} = \int_B \frac{dg}{J_L(g)} \qquad h \in G \, .$$

On the group G, consider the action of G acting on itself (on the right)

[2] From now on, all Jacobians are positive.

$$J_L^*(h : g) = \left|\frac{\partial gh}{\partial g}\right|$$

$$J_L^*(g) = J_L^*(g : i) \tag{1.3}$$

so that

$$\int_{Bh} \frac{dg}{J_L^*(g)} = \int_B \frac{dg}{J_L^*(g)} \qquad h \in G \ .$$

To display the marginal and conditional distributions derived from the inference base in terms of g and D, we require

$$f_\lambda(gD)\left|\frac{\partial Z}{\partial(g\ ,\ D)}\right| dgdD \ ,$$

i.e., the distribution for Z described in terms of the new coordinates.

The Jacobian $J(Z) = \left|\frac{\partial Z}{\partial(S\ ,\ D)}\right|$ is the new key ingredient.

It can be written as

$$J(Z) = J_N(h : h^{-1}Z)J(h^{-1}Z)J_L(h^{-1} : g)$$

using any h on G and corresponding h^{-1} on S .

If we choose $h = g$ we find that

$$J(Z) = J_N(g : D) J(D) J_L(g^{-1} : g)$$

$$= J_N(Z) J(D) J_L^{-1}(g) \ .$$

Thus the change of variable can be expressed as

$$dZ = J_N(gD) J(D) J_L^{-1}(g) dg dD \ .$$

These observations would serve little purpose except for the fact that J_N and J_L are usually easily calculated and the factor $J(D)$ is not needed to display the conditional distribution for $[Z]$ given D since it enters in as a constant which can be incorporated into the derivation of the norming constant.

The joint distribution for g and D can now be displayed as

$$f_\lambda(gD) J_N(gD) J(D) J_L^{-1}(g) dg dD \ .$$

The marginal for D is

$$h_\lambda(D) dD = \int_G f_\lambda(gD) J_N(gD) J(D) J_L^{-1}(g) dg \cdot dD \ . \qquad (1.4)$$

The conditional for $g = [Z]$ given D is

$$g_\lambda(g : D) dg = h_\lambda^{-1}(D) f_\lambda(gD) J_N(gD) J(D) J_L^{-1}(g) dg \ . \qquad (1.5)$$

Recall that D is observed and displays all the information about the unknown Z. We have the observed D together with the probability of what has been observed; $h_\lambda(D)$. This probability depends only on λ. The assessment of λ is based on the likelihood function for λ

$$L(D \mid \lambda) = c h_\lambda(D)$$

where c is any arbitrary constant. It is often called the marginal likelihood function for λ being derived from the marginal probability of what has been observed about Z.

Plausible λ values might be chosen based on the observed likelihood function. We then could consult the conditional distributions $g_\lambda(g : d)$ for these λ values for our study of the unknown θ with the observation that

$$[Y] = \theta[Z]$$

$$g_\lambda\big([Z] ; D(Z)\big) .$$

D. Comments

Familiar inferential statements can be derived from the above such as tests of significance, estimates and confidence regions and numerous illustrations follow in Chapters 2, 3 and 4. The theoretical formulation of such tools will not be displayed here (see Fraser (1978)). Specific instances will be described later.

The techniques described here are very firmly based. The essential factorization of the inference base is unique. The tools used to display the coordinates $[Z]$ and D are convenient but clearly not essential to display inferential statements about the unknowns θ and λ. The choices of $[Z]$ and $D(Z)$ are completely arbitrary and terminal statements of significance and confidence concerning θ depend only on the choice of λ values and that is determined by a uniquely displayed marginal probability and its associated observed marginal likelihood.

However, there is a catch. If the family f_λ has enough symmetries then the integration over the group can possibly be performed analytically. For many interesting families of error distributions, the integration can be performed only numerically. The only qualification for this is that with some situations, certain expressions can be displayed as infinite series and the like. But on practical grounds such formulae are likely not useful. Some encouraging work has been done with approximate expressions however, {Lund (1967), Sprott (1977)}.

Direct numerical quadrature and monte carlo appear to be the strongest tools available. With the advent of high speed computers, considerable research has been devoted to the development of efficient integration techniques. As computers get faster, computer time will become less and less important. This thesis takes full advantages of these tools in displaying the conditional distributions and marginal likelihood functions.

In Chapter 2, the choice of error distribution is described and a large group of families of error distributions are discussed.

In Chapter 3, the location-scale model is analyzed with real data. The ideas of robustness and resistance are considered and encouraging results are included.

In Chapter 4, the regression model is analyzed and similar issues are addressed along with indications for future work.

In Chapter 5, many of the ideas introduced in Chapters 2, 3 and 4 are placed in the general setting described in this chapter. They have the advantage of clean display and allow things to be seen from a more unifying perspective.

CHAPTER 2

THE LINEAR MODEL AND THE ERROR DISTRIBUTION

A. The Inference Base and Necessary Analysis

Consider a sequence of repetitions of a system under various settings of input variables. Let $\underset{\sim}{y} = (y_1 , \ldots , y_n)'$ designate the sequence for the response and suppose that the response has a linear location model

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

where $X = (\underset{\sim}{x}_1 , \ldots , \underset{\sim}{x}_r)$ records r linearly independent vectors based on the input variables.

For the variation $\underset{\sim}{z}$, we have the densities $f_\lambda(\underset{\sim}{z})$ $\lambda \in \Lambda$ and for the response we have the variation based model

$$\{\underset{\sim}{y} = X\underset{\sim}{\beta} + \sigma\underset{\sim}{z} , (\underset{\sim}{\beta} , \sigma) \in \mathbb{R}^r \times \mathbb{R}^+\}$$

giving the set of possible functions for the response $\underset{\sim}{y}$ in terms of $\underset{\sim}{z}$.

If the repetitions on the system are independent then the error distribution can be displayed more firmly as

$$\prod_{i=1}^{n} f_\lambda(z_i) \ .$$

It will be clear in which context the notation f_λ is intended.

The transformation group here is the regression-scale group (see Fraser (1968)). The realized orbit for $\underset{\sim}{z}$ is

$$L^+(X : \underset{\sim}{y}) = \{a_1 \underset{\sim}{x}_1 + \cdots + a_r \underset{\sim}{x}_r + c\underset{\sim}{y} : a_j \in \mathbb{R} , c \in \mathbb{R}^+\} \ .$$

For some of the discussion later on it will be convenient to display properties of the analysis from a model displayed in a slightly more canonical form. It is trivial to show that the matrix X can be written as

$$X = VT$$

where $V = (\underset{\sim}{v}_1 , \ldots , \underset{\sim}{v}_r)$ is orthonormal and T is upper triangular. The Gram Schmidt orthogonalization process is one method that could be used to determine V and T. Then the model can be written as

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

where $\underset{\sim}{\alpha} = T\underset{\sim}{\beta}$.

Clearly $L^+(V ; \underset{\sim}{y}) = L^+(X ; \underset{\sim}{y})$.

Convenient coordinates on the orbit are given by $\underset{\sim}{a}(\underset{\sim}{z}) = V'\underset{\sim}{z}$ so that the vector $\underset{\sim}{z}$ has projection $V\underset{\sim}{a}(\underset{\sim}{z})$ on the space $L(V)$.[3] The vector $\underset{\sim}{z}$ has projection $\underset{\sim}{z} - V\underset{\sim}{a}(\underset{\sim}{z})$ on the orthogonal complement $L^{\perp}(V)$. Thus

$$\underset{\sim}{z} = V\underset{\sim}{a}(z) + \left(\underset{\sim}{z} - V\underset{\sim}{a}(\underset{\sim}{z})\right) = V\underset{\sim}{a}(\underset{\sim}{z}) + s(\underset{\sim}{z})\underset{\sim}{d}(\underset{\sim}{z})$$

where $s(\underset{\sim}{z})$ is the length of the residual vector and $\underset{\sim}{d}(\underset{\sim}{z})$ is the unit residual vector. This suggests using the vectors $\underset{\sim}{v}_1 \cdots \underset{\sim}{v}_r , \underset{\sim}{d}$ as the $r+1$ basis vectors for $L^{+}(V , \underset{\sim}{z})$ with coordinates $a_1(\underset{\sim}{z})$, $a_2(\underset{\sim}{z}) \cdots a_r(\underset{\sim}{z})$, $s(\underset{\sim}{z})$. $\left(\underset{\sim}{a}(\underset{\sim}{z}) , s(\underset{\sim}{z})\right)$ identifies $\underset{\sim}{z}$ on $L^{+}(V , \underset{\sim}{z})$ and $\underset{\sim}{d}(\underset{\sim}{z})$ indexes the different subspaces (i.e., $\underset{\sim}{d}$ is the reference point).

We have

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

Accordingly

$$\begin{aligned}
\underset{\sim}{a}(\underset{\sim}{y}) &= \underset{\sim}{\alpha} + \sigma\underset{\sim}{a}(\underset{\sim}{z}) \\
s(\underset{\sim}{y}) &= \sigma s(\underset{\sim}{z}) \ .
\end{aligned} \qquad (2.1)$$

[3] $L(V)$ — the space spanned by the columns of V .

And we can note that

$$\underset{\sim}{b}(\underset{\sim}{y}) = T^{-1}\underset{\sim}{a}(\underset{\sim}{y})$$

if we wish to study the parameter $\underset{\sim}{\beta}$ relative to the original basis vectors.

Actually this method has been found to be one of the most numerically stable methods to carry out standard regression calculations.

Indeed, very often in experimental situations, it is the coefficients $a_1 \cdots a_r$ that are of primary interest; perhaps for the purpose of testing component hypotheses such as polynomial fitting or the partitioning of degrees of freedom in factorial experiments.

To obtain the marginal and conditional distributions we first need to make the change of variable

$$\underset{\sim}{z} \longleftrightarrow \{\underset{\sim}{a}(\underset{\sim}{z}) , s(\underset{\sim}{z}) , \underset{\sim}{d}\} .$$

For $\underset{\sim}{a} = \underset{\sim}{a}(\underset{\sim}{z})$ we have Euclidean volume $d\underset{\sim}{a}$; for $s = s(\underset{\sim}{z})$ we have Euclidean length ds and for $\underset{\sim}{d}(\underset{\sim}{z})$ we have $s^{n-r-1}\, da$ where da is used for surface volume on the unit sphere in $\mathcal{L}^{\perp}(V)$. This gives

$$d\underset{\sim}{z} = d\underset{\sim}{a}\, ds\, s^{n-r-1}\, da .$$

By substitution, we obtain

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

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

$$h_\lambda(\underset{\sim}{d})da = \int_{\mathbb{R}^+}\int_{\mathbb{R}^r} f_\lambda(V\underset{\sim}{a} + s\underset{\sim}{d})\,s^{n-r-1}\,d\underset{\sim}{a}ds \cdot da \ . \qquad (2.2)$$

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

$$h_\lambda^{-1}(\underset{\sim}{d})\,f_\lambda(V\underset{\sim}{a} + s\underset{\sim}{d})\,s^{n-r-1}\,d\underset{\sim}{a}ds \ . \qquad (2.3)$$

Here we have argued towards the distributions directly. Using the tools from Chapter 1 we find that

$$J_n\big((\underset{\sim}{b}\ , s) : \underset{\sim}{d}\big) = s^n$$

$$J_{r+1}\big((\underset{\sim}{b}\ , s)\big) = s^{r+1}$$

$$J(\underset{\sim}{d}) = 1 \ .$$

Now consider separately the two parameter components $\underset{\sim}{\alpha}\ , \sigma$; Equation (2.1) can be rearranged so that $\underset{\sim}{\alpha}$ and σ are separated.

$$s^{-1}(\underset{\sim}{y})\big(\underset{\sim}{a}(\underset{\sim}{y}) - \underset{\sim}{\alpha}\big) = s^{-1}(\underset{\sim}{z})\underset{\sim}{a}(\underset{\sim}{z}) = \underset{\sim}{T}(\underset{\sim}{z})$$

$$\sigma^{-1}s(\underset{\sim}{y}) = s(\underset{\sim}{z}) \ .$$

This separation is essentially unique, up to re-expression of the individual components. The important observation here is that there is a one-to-one correspondence between the unknown $\underset{\sim}{\alpha}$ and the unknown $\underset{\sim}{T}$. Probability statements about $\underset{\sim}{T}$ can be directly interpreted as confidence statements about $\underset{\sim}{\alpha}$. Similarly with σ and s .

Specifically, the specification of $\underset{\sim}{\alpha}$ with σ unknown determines $\underset{\sim}{z}$ as a ray in $L^+(V\,;\,\underset{\sim}{z})$ and $\underset{\sim}{T}$ represents as a vector the coordinates on $L^+(V\,;\,\underset{\sim}{z})$ needed to index such rays. The specification of σ with $\underset{\sim}{\alpha}$ unknown determines an hyperplane parallel to the hyperplane $L(V)$ and s represents the coordinate necessary to display the distance between the two hyperplanes.

The distribution for $\underset{\sim}{T} = \underset{\sim}{T}(\underset{\sim}{z})$ is

$$g_\lambda^L(\underset{\sim}{T} : \underset{\sim}{d})\,d\underset{\sim}{T} = h_\lambda^{-1}(\underset{\sim}{d}) \int_0^\infty f_\lambda\big(s(V\underset{\sim}{T} + \underset{\sim}{d})\big)s^{n-1}\,ds \cdot d\underset{\sim}{T} \ . \tag{2.4}$$

The distribution for $s = s(\underset{\sim}{z})$ is

$$g_\lambda^S(s : \underset{\sim}{d})\,ds = h_\lambda^{-1}(\underset{\sim}{d}) \int_{\mathbb{R}^r} f_\lambda\big(s(V\underset{\sim}{T} + \underset{\sim}{d})\big)s^{n-1}d\underset{\sim}{T} \cdot ds \ . \tag{2.5}$$

We will have occasion to display somewhat more familiar coordinates from time to time involving the $\underset{\sim}{t}_{\underset{\sim}{z}}$ statistic and $s_{\underset{\sim}{z}}$ statistic where

$$s_{\underset{\sim}{z}} = s/\sqrt{n-r} \quad \text{and} \quad \underset{\sim}{t}_{\underset{\sim}{z}} = \underset{\sim}{a}(\underset{\sim}{z})/s_{\underset{\sim}{z}} \; .$$

From a numerical point of view this has advantages for the integration removing part of the distribution's dependency on sample size. This will be clearer later. The distributions for $\underset{\sim}{t}_{\underset{\sim}{z}}$ and $s_{\underset{\sim}{z}}$ are found by the obvious transformations. The additional constants tend to complicate the expressions and we will usually display the less cluttered ones.

B. Standardization

There are a large number of classes of distributions that display interesting properties that we may wish to identify from the data. All we require is the functional form of such families as norming constants and other quantities can be found easily with the use of numerical integration. It is important that such families (always indexed by λ) do not reflect the characteristics displayed in the model by $\underset{\sim}{\alpha}(\underset{\sim}{\beta})$ or σ . In other words the characteristics to be handled by $\underset{\sim}{\alpha}$ and σ should be fixed within the family $\{f_\lambda : \lambda \in \Lambda\}$. For example, if all the members of the family f_λ have standard deviation 1 then σ can be interpreted as the response standard deviation no matter which member of the error family is used.

We may wish to contemplate families which in their canonical form have changing location and scaling depending on λ . It is the characteristics other than location and scaling that we are interested in having the parameter λ handle.

One possibility is to standardize the distributions with respect to mean and standard deviation. That is, to require that

$$\int_{-\infty}^{\infty} z f_\lambda(z)dz = 0 \quad , \quad \int_{-\infty}^{\infty} z^2 f_\lambda(z)dz = 1 \ .$$

This of course would not work for many long tailed distributions like the Cauchy.

Another possibility that seems reasonable for symmetric distributions is to standardize with respect to median and standard error. That is,

$$\int_{-\infty}^{0} f_\lambda(z)dz = 0.5 \quad , \quad \int_{-1}^{1} f_\lambda(z)dz = 0.6827 \ .$$

For the normal distribution this agrees with the first standardization.

Yet another possibility that seems appealing for asymmetric distributions and agrees with the preceding for symmetric distributions is to standardize so that $(-1 \ , +1)$ is a central 68.27% interval; that is,

$$\int_{-\infty}^{-1} f_\lambda(z)dz = 0.15865 = \int_{1}^{\infty} f_\lambda(z)dz \ .$$

Clearly for a model including possibly asymmetric families the particular application may dictate the interpretation that is desired for the unknowns $\underset{\sim}{\alpha}$ and σ . The computer program which handles the determination of such standardized densities can be easily modified to handle any form of standardization.

One other type deserves a brief mention. If the error family is made up of members with positive density on

$(0\ ,+\infty)$ only, we might wish to standardize so that the interval $(0\ ,1)$ has a fixed probability content, perhaps to agree with a folded normal distribution (0.6827), or perhaps an exponential distribution (0.6321) .

It is likely that in such cases that the original distributions would not be relocated and that $V\underset{\sim}{\alpha}$ would be interpreted as the starting corner for the response. It should be mentioned here that such variable carrier models can be easily implemented using the methods described here.

The computer program is set up to determine the norming constant and appropriate values to determine the standardized densities. Values of the natural logarithm of the standardized densities are then stored in an array and the expression

$$f_\lambda(V\underset{\sim}{a} + s\underset{\sim}{d})$$

is evaluated as the exponential of the sum of n logarithms. The values of the n numbers are determined by a simple linear interpolation determined by each of the component coordinates of $V\underset{\sim}{a} + s\underset{\sim}{d}$. Through experience this has been shown to reduce computer time dramatically as a large amount of time would have to be spent with function evaluation. Numerical accuracy with this method has been very encouraging.

The remaining sections of this chapter consider a number of families that have been suggested as offering insight into the study of linear models; some for their direct application to data and others for their contribution to the understanding of how the distributions that are used for inference respond to the data.

In some instances, complete or partial reduction of expressions can be made by direct analytical integration. These are included where possible.

C. The Normal Distribution

The normal distribution is certainly the most important functional form for displaying an error system. Many random systems have error patterns that follow this distribution. Arguments based on randomization or the central limit theorem or numerous other physical theories offer support for the use of it. Most of the families studied were constructed so as to include the normal distribution so that direct comparisons (via the marginal likelihood and then with the conditional distributions) can be made. We have

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

The conditional distribution for $a(\underset{\sim}{z})$, $s(\underset{\sim}{z})$ has the form

$$\frac{1}{(2\pi)^{r/2}} \exp\{-\tfrac{1}{2}\underset{\sim}{a}'\underset{\sim}{a}\}d\underset{\sim}{a} \cdot \frac{A_{n-r}}{(2\pi)^{\frac{n-r}{2}}} \exp\{-\frac{s^2}{2}\}s^{n-r-1}\ ds \qquad (2.6)$$

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

Note that $a_1 \cdots a_r$, s are independent and indeed independent of $\underset{\sim}{d}$.

The marginal distribution for $\underset{\sim}{d}$ is that of a uniform distribution on the unit sphere in $L^{\perp}(V)$. This is directly observable from the fact that $f(\underset{\sim}{z})$ is spherical and is therefore constant on spheres centred at the origin. Other distributions derived from the normal error form appear later.

D. The Student Distributions

The Student t-distributions arise naturally from classical normal analysis as the marginal distribution for the t-statistic. As the degrees of freedom is lowered, there is more probability in the tails. There is a great deal of empirical evidence to suggest that certain data sets come from error distributions with more tail probability than the normal. The Student family offers rational tails rather than exponential tails. The functional form is

$$\frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\sqrt{\pi\lambda}\;\Gamma\left(\frac{\lambda}{2}\right)}\,(1+z^2/\lambda)^{-\frac{(\lambda+1)}{2}} \qquad \lambda \in (0\,,\infty)\;.$$

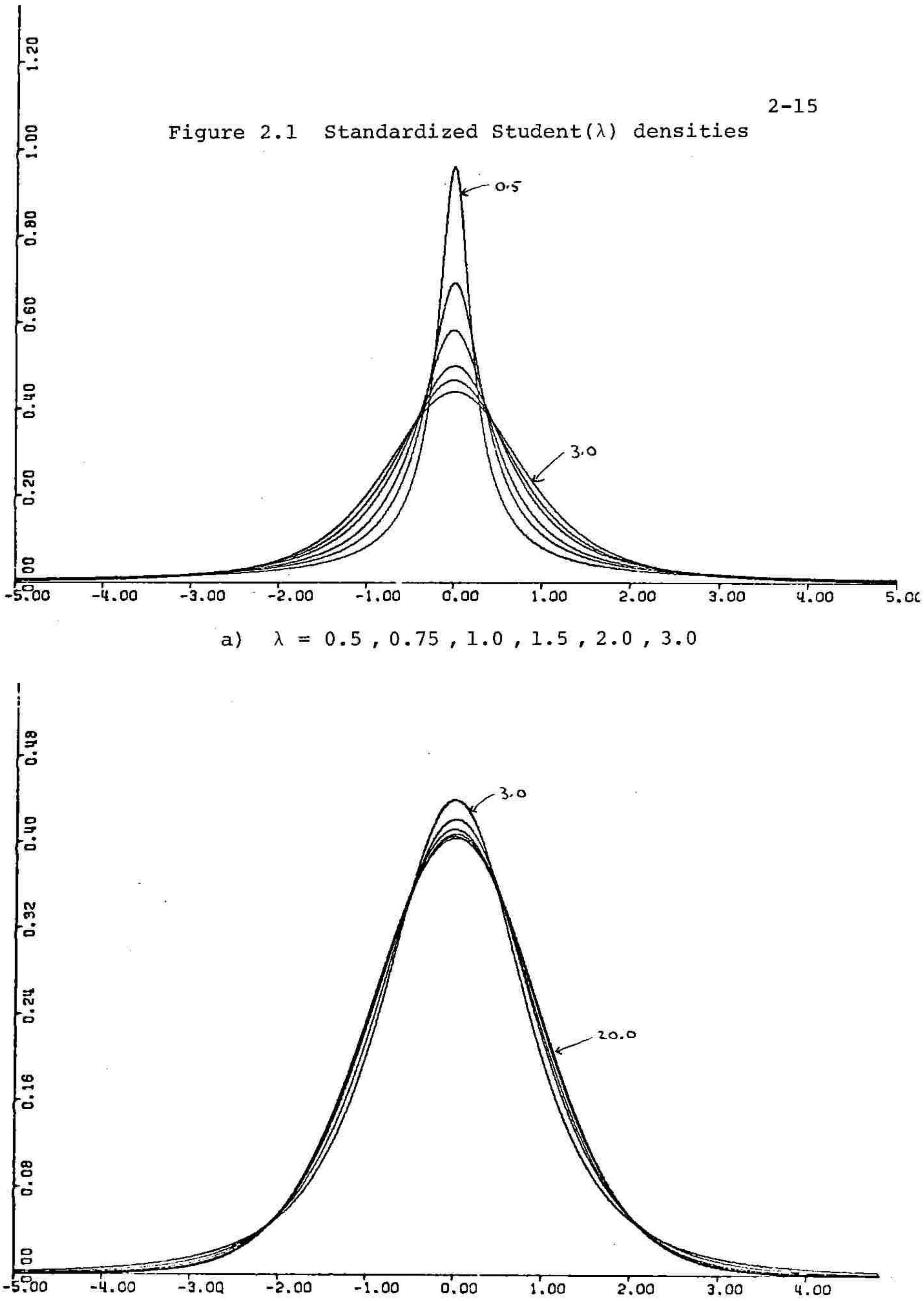
They are symmetric and so have median 0 . They have 68.27% probability in $(-\ell_\lambda\,,\ell_\lambda)$; here are some values of ℓ_λ .

λ	1	3	6	10	15	25	∞
ℓ_λ	1.8367	1.1966	1.0903	1.0524	1.0343	1.0202	1.000

Some representative standardized Student densities are plotted in Figure 2.1.

Many of the illustrations included in this thesis are based on this error form. There are arguments made in Chapter 3 that give strong support for its usage in practical situations.

Figure 2.1 Standardized Student(λ) densities



a) λ = 0.5 , 0.75 , 1.0 , 1.5 , 2.0 , 3.0

b) λ = 3.0 , 5.0 , 8.0 , 12.0 , 20.0

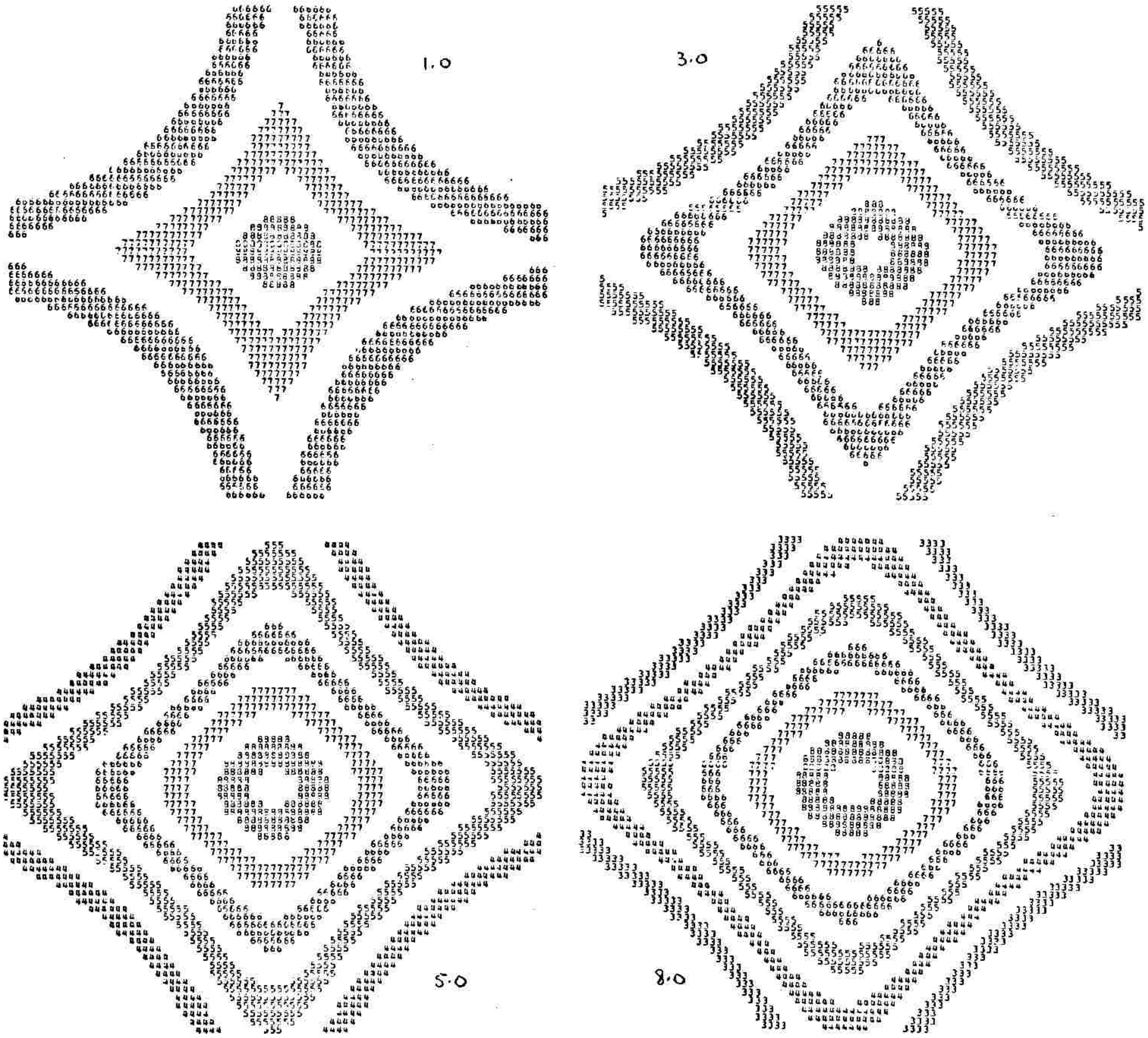


Figure 2.2 Student(λ) contours

$\lambda = 1.0 , 3.0 , 5.0 , 8.0$

$n = 2$

$f_{\lambda}(z_1)f_{\lambda}(z_2)$

Unfortunately, there appears to be no way to display the conditional distributions derived from its use apart from expressions involving integrals. Even samples of size 2 and 3 from the Cauchy distribution lead to horrendous expressions. Numerical integration based on simple Gaussian quadrature rules gives very accurate results with the location-scale model discussed in Chapter 3. The theory behind Gaussian quadrature rules is summarized in the Scientific Subroutine Package manual available through IBM. Additional comments on numerical integration are made in Chapter 3.

It was mentioned that the joint error distribution $\prod_{i=1}^{n} f_\lambda(z_i)$ was spherical if f_λ was the normal. With longer tailed distributions like the Student family the contours of the joint distribution develop lobes along the coordinate axes and the joint pattern can be best described as resembling a children's jack from the game of jacks. The 2 dimensional contours are displayed in Figure 2.2 and an understanding of this type of phenomena appears to be very useful with the monte carlo studies of Chapters 3 and 4.

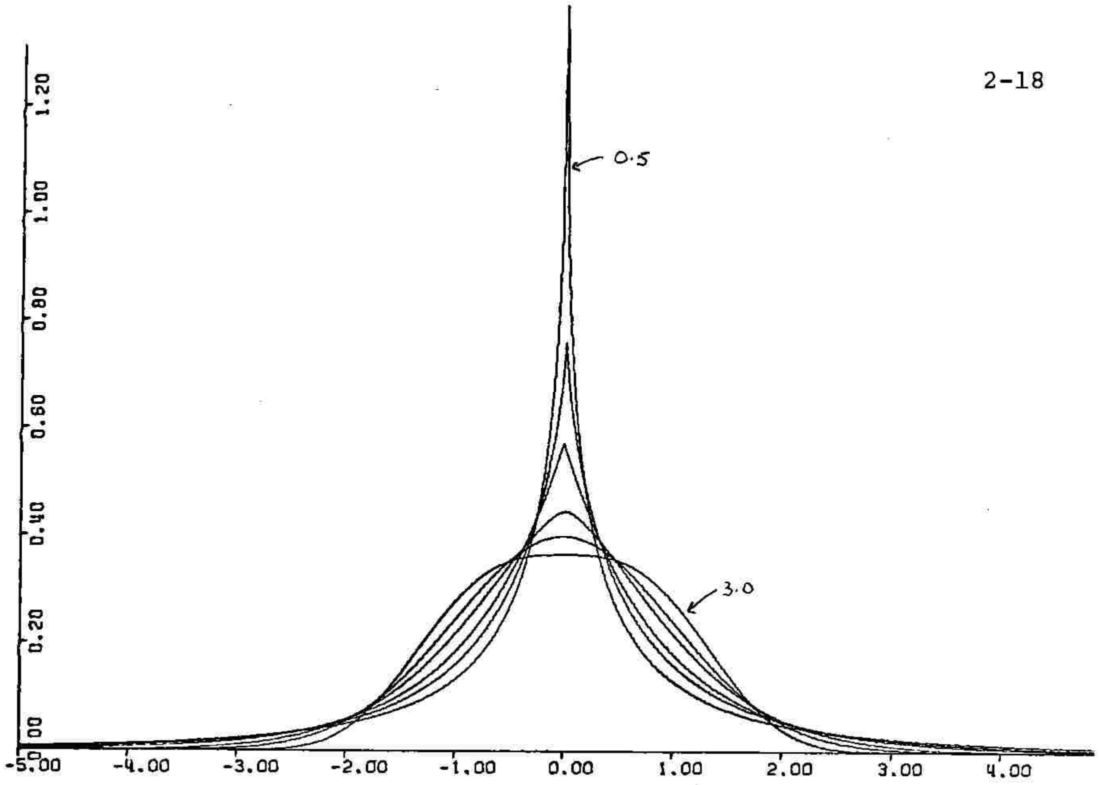
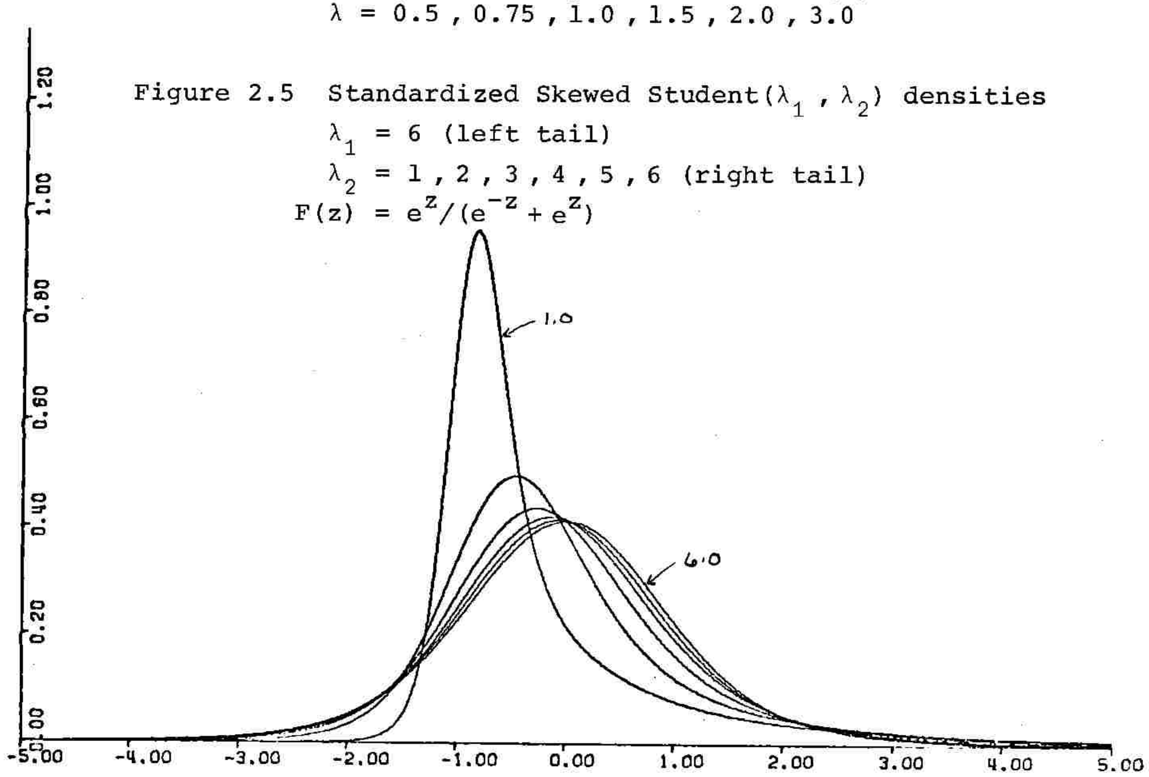


Figure 2.3 Standardized Exponential Power (λ) densities
λ = 0.5 , 0.75 , 1.0 , 1.5 , 2.0 , 3.0

Figure 2.5 Standardized Skewed Student(λ_1 , λ_2) densities
λ_1 = 6 (left tail)
λ_2 = 1 , 2 , 3 , 4 , 5 , 6 (right tail)
$F(z) = e^z/(e^{-z} + e^z)$



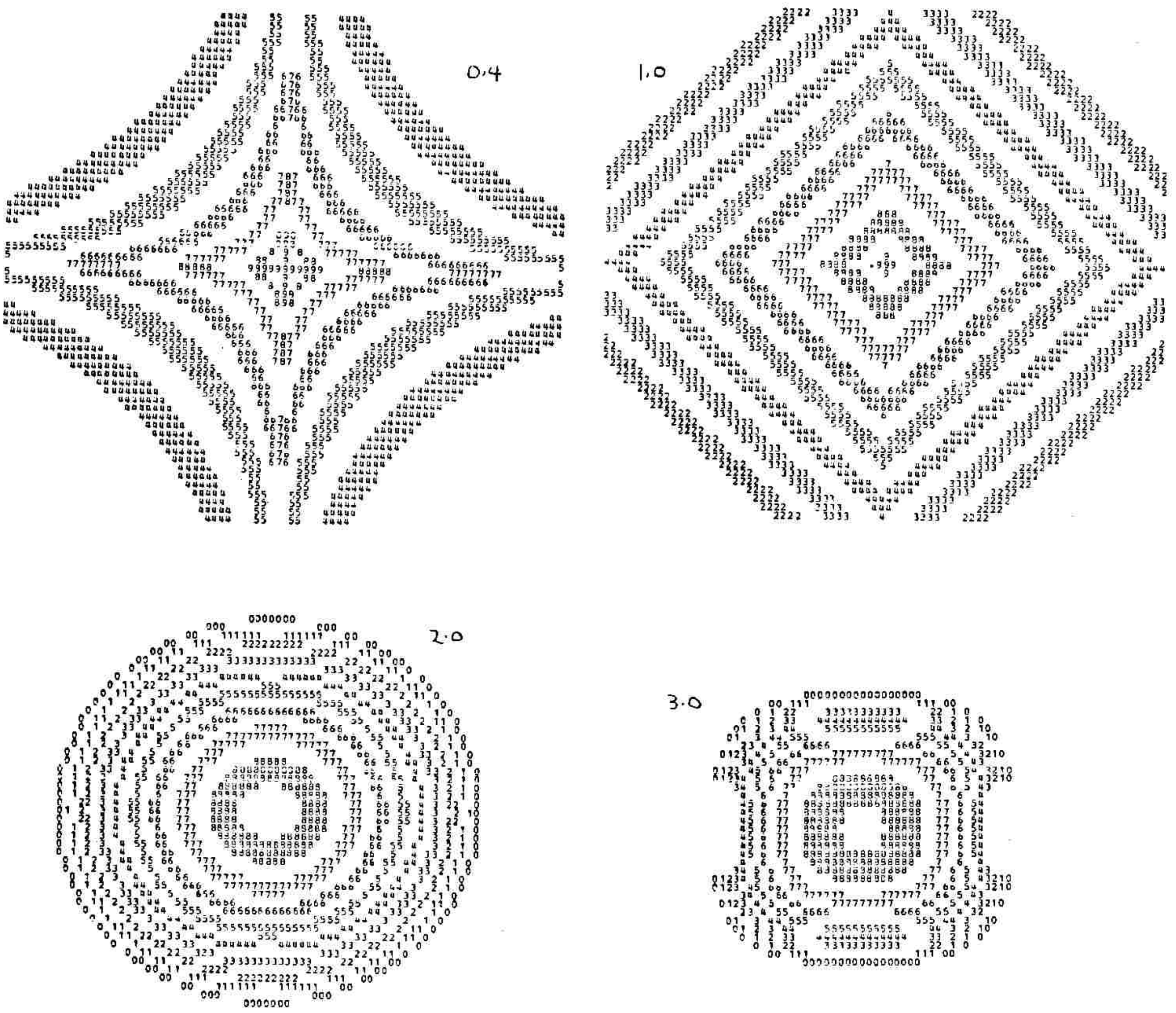


Figure 2.4 Exponential Power (λ) contours

$\lambda = 0.4 , 1.0 , 2.0 , 3.0$

$n = 2$

$f_\lambda(z_1) f_\lambda(z_2)$

E. The Exponential Power Distributions

Another symmetric family for exhibiting tails longer than the normal is the exponential power family. The density function is given by

$$f_\lambda(z) = \frac{1}{2^{1+\frac{1}{\lambda}}\,\Gamma\left(1+\frac{1}{\lambda}\right)} \exp\{-\tfrac{1}{2}|z|^\lambda\} \qquad \lambda \in (0\,,\infty)\,. \qquad (2.7)$$

Representative standardized densities are plotted in Figure 2.3. Notice the rather unnatural cusp at the origin when $\lambda \leq 1$. At best, this family can only be viewed as an approximation to the actual error family. It does have the advantage of including error forms with tails shorter than the normal for $\lambda > 2$. This family was considered extensively by Box and Tiao (1973) and also Barnard (1974). As $\lambda \to \infty$, the standardized distribution tends to a uniform distribution on $(-1.46477\,, 1.46477)$

$$f_\infty(z) = 0.34135 \qquad -1.46477 < z < 1.46477\,.$$

Two dimensional contours of the error distribution are displayed in Figure 2.4.

It turns out that partial integration can be made with this family.

We can compute

$$\int_0^\infty \pi f_\lambda\left(s(\Sigma T_u v_{ui} + d_i)\right)s^{n-1}\,ds$$

which is the joint distribution of $\underset{\sim}{T}$ and $\underset{\sim}{d}$. First let $c(\lambda) = \left(2^{1+1/\lambda}\,\Gamma(1+\frac{1}{\lambda})\right)^{-1}$ and $r(\underset{\sim}{T}\,,\,V) = \sum_{i=1}^{n}\left|\sum_{u=1}^{r} T_u v_{ui} + d_i\right|^{\lambda}$ then we obtain

$$\int_0^\infty c^n(\lambda)\exp\{-\tfrac{1}{2}\,r(\underset{\sim}{T}\,,\,V)s^\lambda\}s^{n-1}\,ds\ .$$

Let $u = s^\lambda$ to obtain

$$\frac{c^n(\lambda)}{\lambda}\int_0^\infty \exp\left\{-\frac{r(\underset{\sim}{T}\,,\,V)}{2}\,u\right\}u^{\frac{n}{\lambda}-1}\,du\ ,$$

now let $w = \dfrac{r(\underset{\sim}{T}\,,\,V)}{2}\,u$

to obtain

$$\frac{c^n(\lambda)}{\lambda}\left(\frac{2}{r(\underset{\sim}{T}\,,\,V)}\right)^{\frac{n}{\lambda}}\int_0^\infty \exp\{-w\}w^{\frac{n}{\lambda}-1}\,dw$$

$$= \frac{c^n(\lambda)}{\lambda}\,\Gamma\left(\frac{n}{\lambda}\right)2^{\frac{n}{\lambda}}\left(r(\underset{\sim}{T}\,,\,V)\right)^{-\frac{n}{\lambda}}$$

$$= \frac{\Gamma\left(\frac{n}{\lambda}\right)}{\lambda 2^n\Gamma^n\left(1+\frac{1}{\lambda}\right)}\left(\sum_{i=1}^{n}\left|\sum_{u=1}^{r} T_u v_{ui} + d_i\right|^{\lambda}\right)^{-\frac{n}{\lambda}}\ . \qquad (2.8)$$

This is the joint distribution for $\underset{\sim}{T}$ and $\underset{\sim}{d}$. To determine the conditional distribution for $\underset{\sim}{T}$ given $\underset{\sim}{d}$ now only involves an r dimensional integral $\left(h_\lambda(\underset{\sim}{d})\right)$. The actual distribution will differ by constants depending on the standardization. These symbols were suppressed to increase the clarity of presentation.

As an illustration, consider $\lambda = 2$ corresponding to the normal distribution: the conditional distribution is then

$$\frac{h_\lambda^{-1}(\underset{\sim}{d})\,\Gamma\left(\frac{n}{2}\right)}{2\cdot 2^n\Gamma^n\left(1+\frac{1}{\lambda}\right)}\left(\sum_{i=1}^{n}\left|\sum_{u=1}^{n} T_u v_{ui} + d_i\right|^2\right)^{-\frac{n}{2}}$$

$$= \frac{A_{n-r}\Gamma\left(\frac{n}{2}\right)}{2\pi^{n/2}}\left(1+\underset{\sim}{T}'\underset{\sim}{T}\right)^{-\frac{n}{2}}$$

$$= \frac{A_{n-r}}{A_n}\left(1+\underset{\sim}{T}'\underset{\sim}{T}\right)^{-\frac{n}{2}} \qquad (2.9)$$

which is the canonical Student(n-r) on $\mathbb{R}^r$ (see Fraser (1976), p. 79).

F. Asymmetric Distributions

Traditionally, when there appears to be evidence of an asymmetry or skewness in the error system, some nonlinear transformation on the response is considered. Quite often such asymmetries are linked with a hetrogeneity of the variance making even stronger evidence for data transformation.

Of course the analysis described earlier in this section can handle skewness head on by including an allowance for asymmetry in the parameter λ . If tail length is also a consideration then it makes sense to consider λ as two dimensional. There is no change in the analysis except that now it appears reasonable to consult a contour plot of the marginal likelihood function for $\underset{\sim}{\lambda} = (\lambda_1 , \lambda_2)$.

One interesting family that was devised to display the phenomena is the following class of distributions which have functional form

$$\left(1 + F(-z)z^2/\lambda_1\right)^{-\frac{\lambda_1 + 1}{2}} \left(1 + F(z)z^2/\lambda_2\right)^{-\frac{\lambda_2 + 1}{2}} \qquad (2.10)$$

where F(z) is some symmetric distribution function. For z large and negative F(-z) will be close to 1 and F(z) will be close to zero, making the left tail of this distribution similar to a Student(λ_1) density. For z large and positive the density will be very similar to a Student(λ_2) density. If $\lambda_1 = \lambda_2$ we obtain yet another symmetric class of densities. If $\lambda_1 < \lambda_2$ the densities are negatively skewed and if $\lambda_1 > \lambda_2$ the densities are positively skewed. As λ_1 and $\lambda_2 \to \infty$ these distributions tend to $N(0, 1)$. This has been called the skewed Student(λ_1, λ_2) family. Several examples of the standardized skewed Student densities are illustrated in Figure 2.5.

Often in life-testing problems a simple scale model with Weibull errors is used as a basis for analysis

$$\underset{\sim}{\omega} = \sigma \underset{\sim}{z}$$

$$f_\beta(\underset{\sim}{z}) = \beta x^{\beta-1} \exp\{-z^\beta\} \qquad z > 0 .$$

If in fact one considers

$$y = \log \omega$$

then one obtains a location-scale model

$$\underset{\sim}{y} = \mu \underset{\sim}{1} + \sigma \underset{\sim}{z}$$

$$f(\underset{\sim}{z}) = \exp(+z)\exp\{-\exp\{+z\}\}$$

where $\mu = \ell n\ \sigma$ and $\sigma = \beta^{-1}$.

This error distribution is another example of an asymmetric error distribution. In fact $-z$ has what is called the standard extreme value distribution. This model has been investigated by A. Dobriyal and A. McIntosh at the

University of Toronto (see Fraser (1978), Section 2.4).

There are many other asymmetric distributions that would be interesting to consider: the noncentral t and F distributions, other variations on the addition or multiplication of symmetric densities, adding modulating factors, etc.

G. Scaled Normal Distributions

We now consider the situation in which the errors have a multivariate normal distribution $N(\underset{\sim}{0}\ ,\ \Sigma)$. We are admitting here the possibility of dependency of the errors and several special cases will be considered here and in later sections. The distributions derived in this section will be used later as an aid to understanding other distributions and also for some monte carlo studies.

We have

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

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

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

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

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

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

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

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

$$= \frac{1}{A_{n-r}} \cdot \frac{|V'\Sigma^{-1}V|^{-\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} \cdot \frac{1}{(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}} \qquad (2.11)$$

so that the conditional distribution for $(\underset{\sim}{a}\,,\,s)$ given $\underset{\sim}{d}$ is

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

(2.12)

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

$$\frac{(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}}}{\Gamma\left(\frac{n-r}{2}\right)} \exp\{-\tfrac{1}{2}(\underset{\sim}{d}'R\underset{\sim}{d})s^2\}\,s^{n-r-1}\,ds \ . \qquad (2.13)$$

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

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

$$= \frac{A_{n-r}}{A_n}\,|V\Sigma^{-1}V|^{\frac{1}{2}}(\underset{\sim}{d}'R\underset{\sim}{d})^{\frac{n-r}{2}} \cdot \frac{1}{\left((V\underset{\sim}{T}+\underset{\sim}{d})'\Sigma^{-1}(V\underset{\sim}{T}+\underset{\sim}{d})\right)^{n/2}} \; . \qquad (2.14)$$

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

$$\frac{A_{n-r}}{A_n}\,|W|^{-\frac{1}{2}}\left(1+(\underset{\sim}{T}-\underset{\sim}{\mu})W^{-1}(T-\underset{\sim}{\mu})\right)^{n/2} \; . \qquad (2.15)$$

We will write

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

To find $\underset{\sim}{\mu}$ and W we display the quadratic form in (2.14) as

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

so that

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

and

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

For monte carlo work in Chapters 3 and 4, we will be interested in the distribution of components of $\underset{\sim}{T} = \begin{pmatrix} \underset{\sim}{T}_1 \\ \underset{\sim}{T}_2 \end{pmatrix}$ on $\mathbb{R}^{r_1} \times \mathbb{R}^{r_2} = \mathbb{R}^r$.

Our method of derivation begins with a consideration of the canonical Student distribution, $\text{Student}_{n-r}(\underset{\sim}{0} \ , I)$ on $\mathbb{R}^r$ which has density

$$\frac{A_{n-r}}{A_n}(1 + \underset{\sim}{x}'\underset{\sim}{x})^{-n/2} \ .$$

If we write $\underset{\sim}{x} = (\underset{\sim}{x}_1 \ , \underset{\sim}{x}_2)'$,

then the distribution can be written as

$$\frac{A_{n-r}}{A_{n-r_1}} \cdot \frac{1}{(1 + \underset{\sim}{x}_2'\underset{\sim}{x}_2')^{\frac{n-r_1}{2}}} \cdot \frac{A_{n-r_1}}{A_n} \cdot \frac{1}{(1 + \underset{\sim}{x}_2'\underset{\sim}{x}_2)^{\frac{r_1}{2}}} \cdot \frac{1}{\left[1 + \frac{\underset{\sim}{x}_1'\underset{\sim}{x}_1}{(1 + \underset{\sim}{x}_2'\underset{\sim}{x}_2)}\right]^{\frac{n}{2}}} \tag{2.16}$$

in other words, the marginal for $\underset{\sim}{x}_2$ is $\text{Student}_{n-r}(0 \ , I)$ on $\mathbb{R}^{r_2}$ and the conditional for $\dfrac{\underset{\sim}{x}_1}{(1 + \underset{\sim}{x}_2'\underset{\sim}{x}_2)^{\frac{1}{2}}}$ given $\underset{\sim}{x}_2$ is $\text{Student}_{n-r_1}(0 \ , I)$ on $\mathbb{R}^{r_1}$.

The distribution for $\underset{\sim}{T}$ can be described as

$$\underset{\sim}{T} = \underset{\sim}{\mu} + \Gamma \underset{\sim}{x} \quad \text{where} \quad \Gamma\Gamma' = W .$$

Now if

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \quad \text{and} \quad \underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}$$

then

the marginal for $\underset{\sim}{T}_2$ is $\text{Student}_{n-r}(\underset{\sim}{\mu}_2, W_{22})$ on $\mathbb{R}^{r_2}$

and the

conditional for $\underset{\sim}{T}_1$ given $\underset{\sim}{T}_2$ is $\text{Student}_{n-r_1}(\underset{\sim}{\nu}, z)$ on $\mathbb{R}^{r_1}$

where

$$\underset{\sim}{\nu} = \underset{\sim}{\mu}_1 - W_{12}W_{22}^{-1}(\underset{\sim}{T}_2 - \underset{\sim}{\mu}_2)$$

$$z = \left(1 + (\underset{\sim}{T}_2 - \underset{\sim}{\mu}_2)'W_{22}^{-1}(\underset{\sim}{T}_2 - \underset{\sim}{\mu}_2)\right)^{-\frac{1}{2}} \cdot [W_{11} - W_{12}W_{22}^{-1}W_{21}] . \tag{2.17}$$

One justification for the above result proceeds in the following way.

Let

$$\underset{\sim}{e} = \begin{pmatrix} I & -W_{12}W_{22}^{-1} \\ O & I \end{pmatrix} \underset{\sim}{T} .$$

This splits the quadratic form in expression (2.15) into two pieces as in expression (2.16). Splitting the joint density into conditional and marginal resembles the corresponding expressions in the canonical case. You won't be disappointed if I spare you this expression.

The use of these results will appear in Section 4D with monte carlo integration. The important idea here is that all of the formulae can be worked out analytically (as is practically characteristic of normal theory results). Later on we will be using particular sums of scaled normal distributions to attempt to emulate the contours of nonnormal distributions for which the analytical results are not practible.

H. Spherical Distributions

We noted earlier that if the error distribution is made up of independent standard normal variables that the joint distribution for $\underset{\sim}{z}$ is spherical.

In fact, if the errors are independent and identically distributed then the uniform distribution for $\underset{\sim}{d}$ characterizes the normal distribution as the parent.

Theorem

Let $\underset{\sim}{z}$ be a random sample from a symmetric continuous density function f and let $\underset{\sim}{d}$ denote the standardized residual vector. Then the distribution $h(\underset{\sim}{d})$ characterizes the parent distribution f . If the distribution for $\underset{\sim}{d}$ is uniform, this characterizes the normal.

Results of this nature have been proven by Prokhorov (1965), Zinger (1956) and Zinger and Linnik (1964).

The presentation of these results and other related interesting characterizations can be found in Kagan, Linnik, Rao (1973, Section 13.5).

It is interesting however to consider the case where $\underset{\sim}{z}$ has a general spherical distribution

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

We know that

$$h_\lambda(\underset{\sim}{d}) = \frac{1}{A_{n-r}}$$

so that the conditional distribution for $\underset{\sim}{a}$ and s given $\underset{\sim}{d}$ is of course just the marginal

$$A_{n-r}\, g_\lambda\big((V\underset{\sim}{a} + s\underset{\sim}{d})'(V\underset{\sim}{a} + s\underset{\sim}{d})\big)s^{n-r-1}\, d\underset{\sim}{a}\, ds$$

$$= A_{n-r}\, g_\lambda(\underset{\sim}{a}'\underset{\sim}{a} + s^2)s^{n-r-1}\, d\underset{\sim}{a}\, ds \ . \qquad (2.18)$$

The marginal distribution for $\underset{\sim}{T}$ is

$$\int_s A_{n-r}\, g_\lambda\big(s^2(1 + \underset{\sim}{T}'\underset{\sim}{T})\big)s^{n-1}\, ds \cdot d\underset{\sim}{T}$$

and letting $u = s(1 + \underset{\sim}{T}'\underset{\sim}{T})^{\frac{1}{2}}$

$$= \int_u A_{n-r}\, g_\lambda(u^2)u^{n-1}\, du \frac{1}{(1 + \underset{\sim}{T}'\underset{\sim}{T})^{n/2}}\, d\underset{\sim}{T}$$

which is just

$$\frac{A_{n-r}}{A_n} \cdot \frac{1}{(1 + \underset{\sim}{T}'\underset{\sim}{T})^{n/2}}$$

which is the $\text{Student}_{n-r}(0\ ,\ I)$ on $\mathbb{R}^r$. Notice that this shows that

$$\int_u g_\lambda(u^2)u^{n-1}du = \frac{1}{A_n} \quad \text{when the integral is convergent.}$$

One interesting nonnormal spherical distribution is $\text{Student}_\lambda(0\,,\,I)$ on $\mathbb{R}^n$ with density

$$\frac{A_\lambda}{A_{\lambda+n}} \cdot \frac{1}{(1+\underset{\sim}{z}'\underset{\sim}{z})^{\frac{\lambda+n}{2}}} \;.$$

CHAPTER 3

ANALYSIS OF THE LOCATION-SCALE MODEL

A. A Classical Example

We now begin an examination of the simplest linear model, the location-scale model. We have a data set $\underset{\sim}{y}$ and the following model

$$\underset{\sim}{y}$$

$$\underset{\sim}{y} = \mu\underset{\sim}{1} + \sigma\underset{\sim}{z}$$

$$f_\lambda(\underset{\sim}{z}) \quad \text{or} \quad \Pi f_\lambda(z_i) \ .$$

A great deal of the results seen in this chapter have direct applicability to the understanding of more complicated linear models with general error distributions.

We now examine the Darwin data as recorded in Fisher (1971).

The data came from an experiment to compare the heights of cross-and self-fertilized plants. The design involved 15 pairs of plants, each pair consisting of a cross-and a self-fertilized plant grown under the same conditions in the same pot. The data available are the fifteen differences in height cross- minus self-fertilized

49	23	24	-67	28
75	8	41	60	16
14	-48	6	56	29

It is of interest later to note the two extreme values on the left tail of the sample.

For illustration we consider 2 error families; the standardized Student(λ) distributions (Section 2D) and the standardized exponential power (λ) distributions (Section 2E).

It perhaps seems reasonable to consider the Student distributions to allow for longer tails than the normal and the exponential power distributions to allow for the possibility of shorter tails than the normal (here $\lambda > 2$) .

(i) Likelihood Analysis

The observed value of $\underset{\sim}{d}$ is

0.1987	0.0146	0.0217	-0.6224	0.0500
0.3828	-0.0916	0.1421	0.2766	-0.0399
-0.0491	-0.4881	-0.1057	0.2483	0.0571

Let the likelihood function derived from the Student analysis be

$$L_{ST}(\underset{\sim}{d} \mid \lambda) = A_{n-1}h_{\lambda}(\underset{\sim}{d}) \ .$$

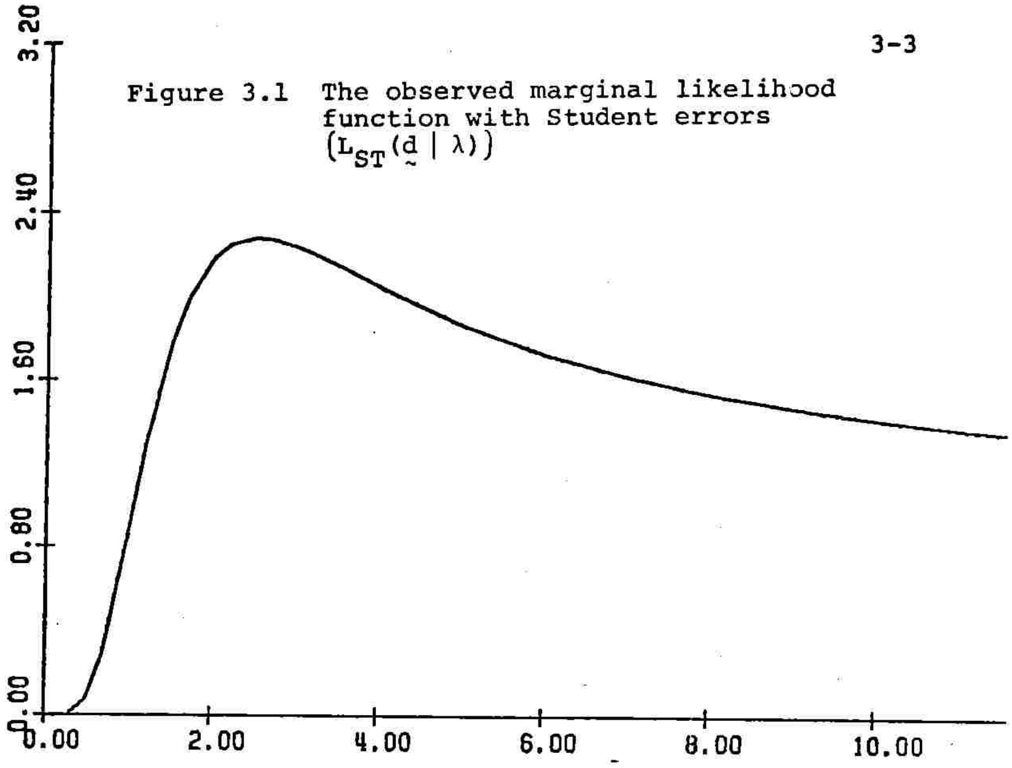


Figure 3.1 The observed marginal likelihood function with Student errors $\left(L_{ST}(\underset{\sim}{d} \mid \lambda)\right)$

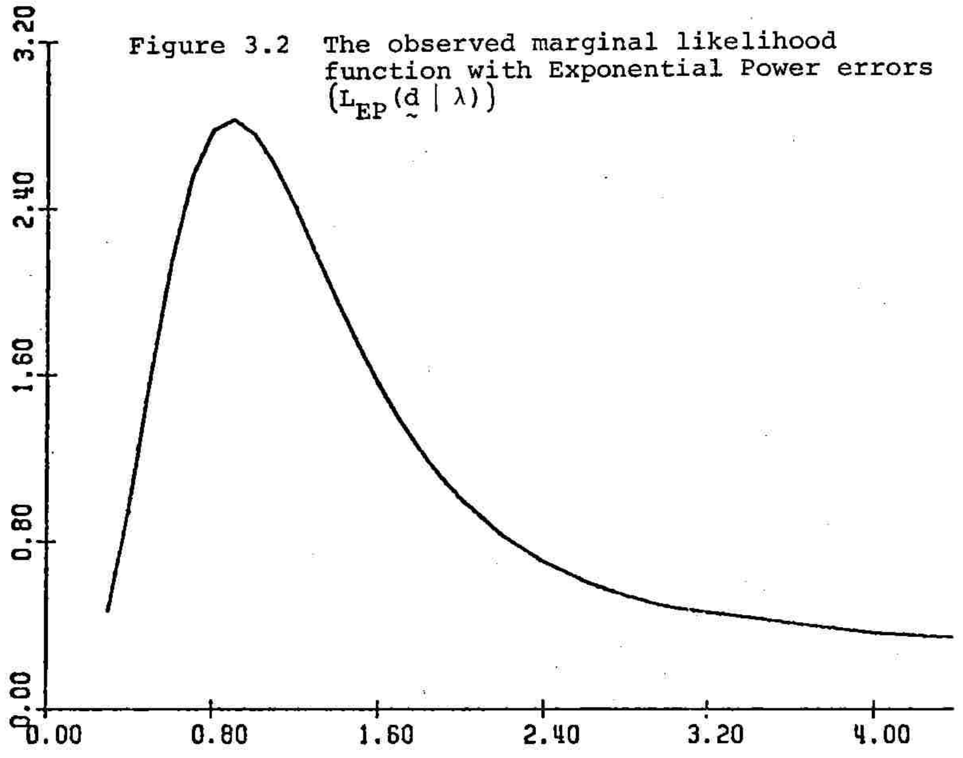


Figure 3.2 The observed marginal likelihood function with Exponential Power errors $\left(L_{EP}(\underset{\sim}{d} \mid \lambda)\right)$

We have chosen the representative curve that has $L_{ST}(\underset{\sim}{d} \mid \infty) = 1$. This allows for direct likelihood ratio comparison with the classical normal analysis.

From the exponential power analysis we will have

$$L_{EP}(\underset{\sim}{d} \mid \lambda) = A_{n-1} h_\lambda(\underset{\sim}{d})$$

so that $L_{EP}(\underset{\sim}{d} \mid 2) = 1$ corresponding to a normal analysis. We remark again that for $\lambda < 1$ there is a rather unnatural cusp at the origin that becomes more and more annoying as λ gets small.

The computer program produces the likelihoods in both tabulated and graphic form. They are plotted in Figures 3.1 and 3.2 . L_{ST} suggests that λ values in the range from 1 to 9 are reasonable for the remainder of the analysis. L_{EP} offers evidence against short tailed distributions. For illustration we consider λ values from 0.5 to 2 (further comments about this situation will be made in Section 3H.

(ii) Inference for μ and σ

Figures 3.3 and 3.4 display the t-statistic distributions for selected values of λ from the Student and exponential power analyses respectively. Recall that we are using the more familiar form for the t statistic here

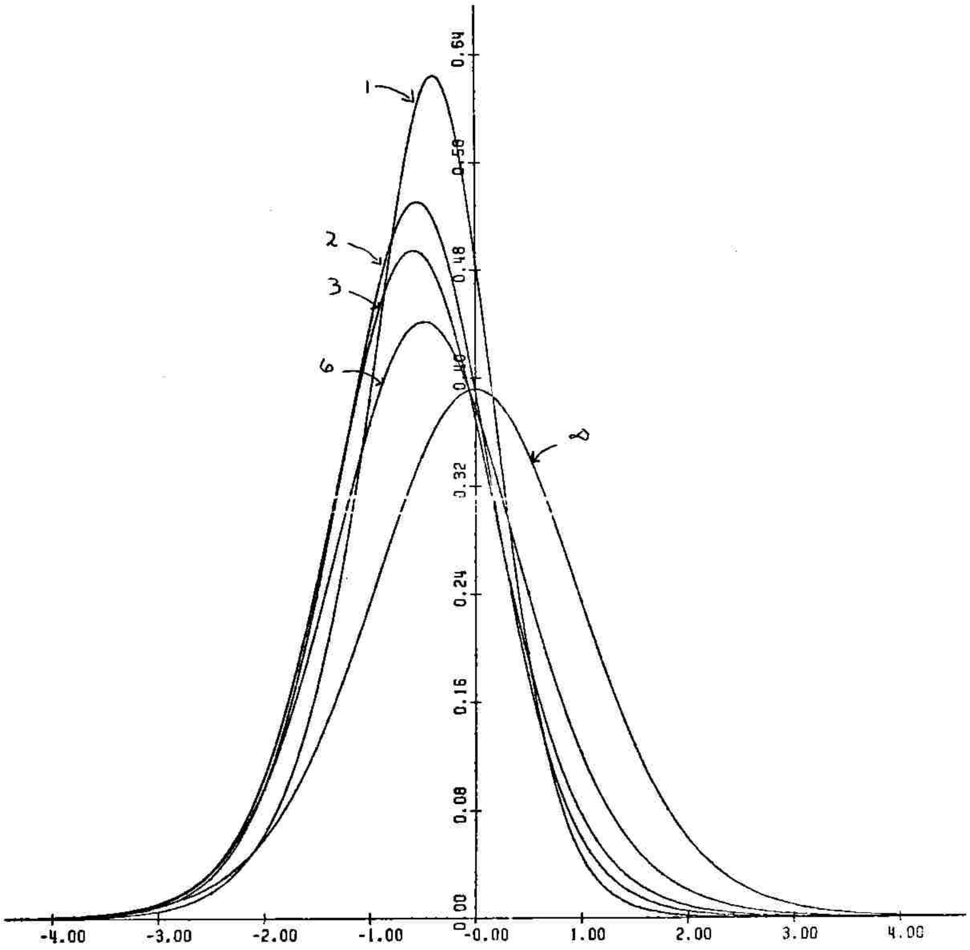


Figure 3.3 The t-statistic densities
(Student analysis)
(λ = 1 , 2 , 3 , 6 , ∞)

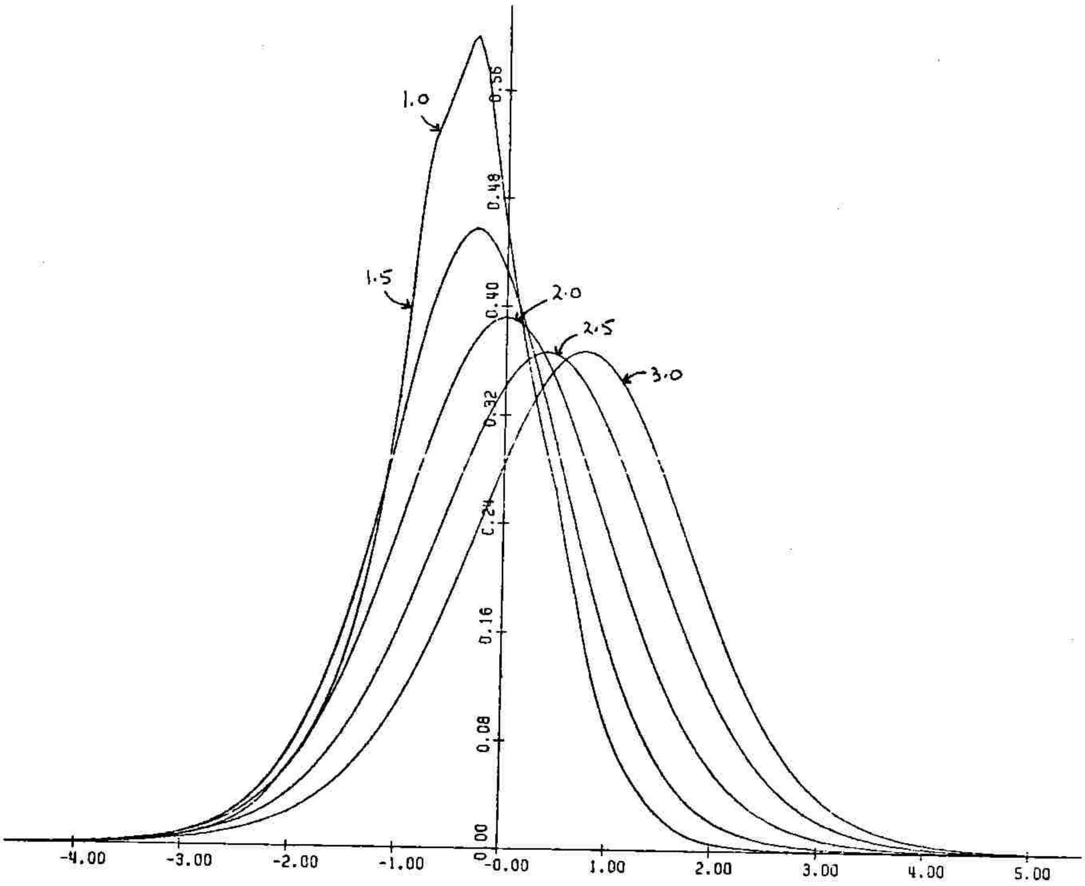


Figure 3.4 The t-statistic densities (Exponential Power analysis) (1.0 , 1.5 , 2.0 , 2.5 , 3.0)

$$t = \frac{a}{s/\sqrt{n-1}} = \sqrt{n}\,\bar{z}/s_{\underset{\sim}{z}} = \frac{\bar{z}}{s_{\underset{\sim}{z}}/\sqrt{n}} .$$

Now consider the formation of confidence intervals for the location parameter μ . The central $1-\alpha$ confidence interval has the form

$$(\bar{y} - t_2 s_{\underset{\sim}{y}}/\sqrt{n} \;,\; \bar{y} - t_1 s_{\underset{\sim}{y}}/\sqrt{n})$$

where (t_1 , t_2) is the central $1-\alpha$ probability interval for the t-statistic based on the appropriate conditional distribution. The computer program computes these intervals for any chosen λ values and confidence levels.

The 95% intervals for the Student analysis are as follows:

	t_1	t_2	$\hat{\mu}_1$	$\hat{\mu}_2$
$\lambda = 1$	-2.25	0.53	14.8	42.9
3	-2.28	1.17	9.5	43.1
6	-2.28	1.55	5.8	43.2
9	-2.27	1.73	4.1	43.0
∞	-2.14	2.14	-0.03	41.8

The 95% intervals for the exponential power analysis are as follows:

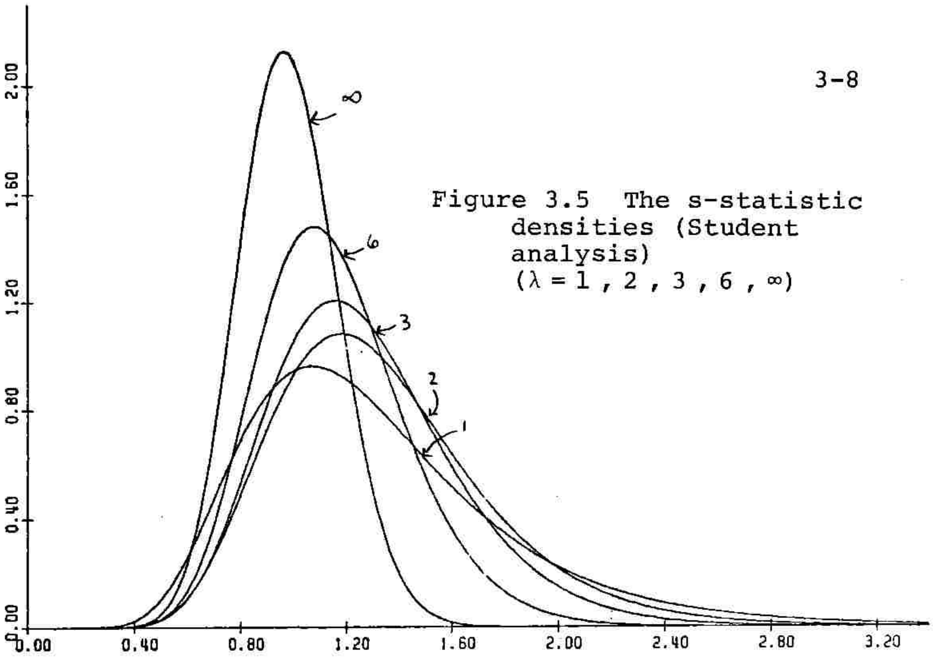


Figure 3.5 The s-statistic densities (Student analysis) (λ = 1 , 2 , 3 , 6 , ∞)

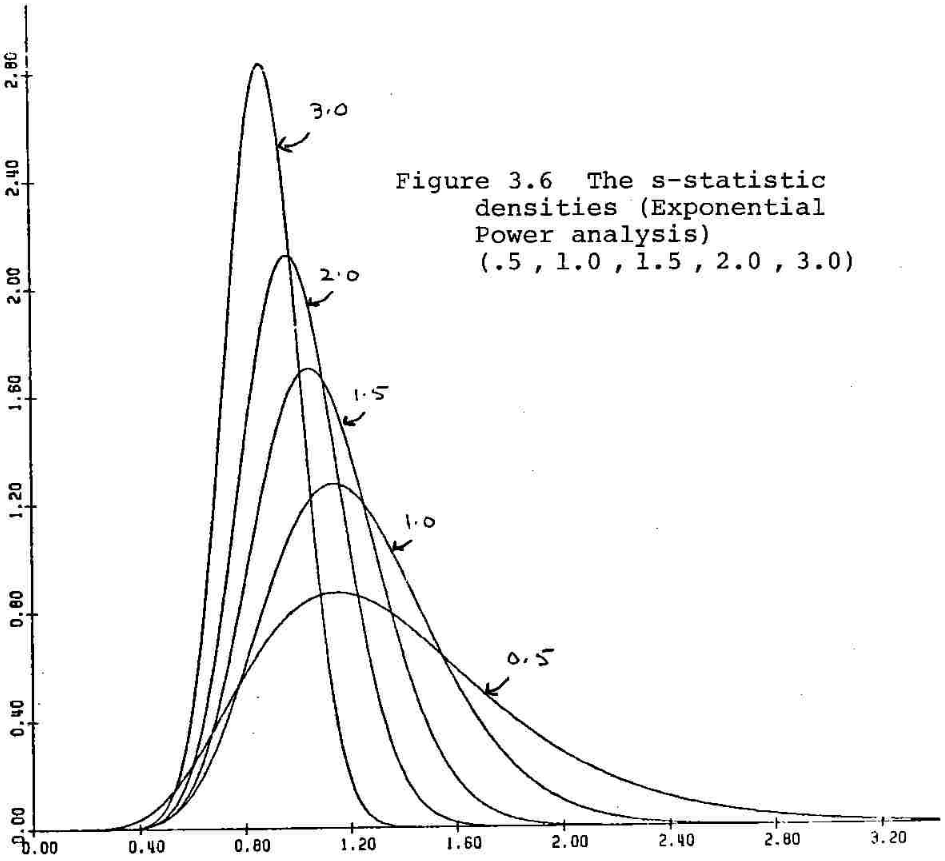


Figure 3.6 The s-statistic densities (Exponential Power analysis) (.5 , 1.0 , 1.5 , 2.0 , 3.0)

	t_1	t_2	$\hat{\mu}_1$	$\hat{\mu}_2$
$\lambda = 1.0$	-2.05	1.13	9.89	40.94
$\lambda = 1.5$	-2.18	1.58	5.51	42.21
$\lambda = 2.0$	-2.14	2.14	-0.03	41.84
$\lambda = 2.5$	-1.98	2.61	-4.50	40.23
$\lambda = 3.0$	-1.74	2.93	-7.58	37.86

Figures 3.5 and 3.6 display the s-statistic distributions for the selected values of λ where $s_{\underset{\sim}{z}} = s/\sqrt{n-1}$.

Central $1-\alpha$ confidence intervals for σ have the form

$$\left(\frac{s_{\underset{\sim}{y}}}{s_2}, \frac{s_{\underset{\sim}{y}}}{s_1}\right) .$$

The 95% intervals for the Student analysis are as follows:

	s_1	s_2	$\hat{\sigma}_1$	$\hat{\sigma}_2$
$\lambda = 1$	0.602	2.544	14.836	62.66
$\lambda = 2$	0.695	2.277	16.575	54.315
$\lambda = 3$	0.698	2.075	18.191	54.110
$\lambda = 6$	0.675	1.772	21.303	55.905
$\lambda = \infty$	0.634	1.366	27.633	59.527

And for the exponential power analysis:

	s_1	s_2	$\hat{\sigma}_1$	$\hat{\sigma}_2$
$\lambda = 1.0$	0.676	1.934	19.52	55.86
$\lambda = 1.5$	0.661	1.582	23.86	57.08
$\lambda = 2.0$	0.634	1.366	27.63	59.53
$\lambda = 2.5$	0.612	1.231	30.67	61.66
$\lambda = 3.0$	0.598	1.144	32.99	63.11

The intervals that have the strongest support are those near the maximum likelihood values. All the intervals of course have the property of marginal as well as conditional confidence since

$$E(1-\alpha : \underset{\sim}{d}) = 1-\alpha \ .$$

They are taking advantage of more information than intervals with a strictly marginal confidence property.

The computer program that handles this analysis has been used hundreds of times now. It has been demonstrated that even beginning students of statistics can understand and interpret the output. It has been distributed (in card deck form) to over 40 centres around the world by the author (i.e., Fick (1975)).

More comments will be made in later sections.

B. Hardware and Software

Most of the work with nonnormal error distributions demands that the user have easy access to a large-scale computing system. This prerequisite is now satisfied by virtually all academic centres where statistical methods are used. Systems with the availability of both batch and interactive systems would be ideal for the use of the methods described in this thesis.

(i) Graphics

The key ingredients for the terminal stages of inference are the distributions themselves. Of course, summary statistics derived from these distributions can be very useful but there is no substitute for a display of a probability distribution or an observed likelihood function.

For displaying one dimensional distributions and likelihoods, CALCOMP and GOULD were used extensively. The use of CRT terminals is increasing all the time. A very useful situation would involve the use of a CRT with an optional device for obtaining hard copy when the display is in an appropriate form.

For two dimensional displays, a contour plot routine was developed for use directly off a line printer. It is illustrated in Figures 2.2 and 2.4. The function is evaluated in a grid and then a coded value is assigned depending on the function value. One particularly useful scheme of coded values was

9 blank 8 blank 7 blank etc.

giving from highest to lowest value. Often this scheme gives a very clear idea of the shape of the distribution. With non-normal error forms the patterns observed are often very interesting (see Figure 2.2). A plot in this coded form can often be used directly to determine approximate function values.

If the computer programs are used interactively, one obtains added flexibility and speed at the terminal stages of inference. For example, one could consult the marginal likelihood function first and then, based on its form, select a range of λ values for the remainder of the analysis involving θ .

(ii) Integration

Many numerical integration methods were tried over the last few years. Indeed several very sophisticated adaptive rules were tried.

These rules were not devised with statistical distributions in mind but were in fact designed to handle pathological functions with continuity and differentiability problems.

For the integration problems considered here, these routines are not recommended. For a majority of situations considered, simple nonadaptive rules such as Gaussian quadrature and Simpson's rule offered excellent results at low cost. Consult the SSP (IBM Scientific Subroutine Package) manual for reliable and accurate code (DQGn and DQSF).

Most of the integration that was carried out involved infinite or semi-infinite regions. With some information about the distributions involved, it does not seem necessary to transform these regions to bounded regions. Generally, truncation at some reasonable point does not appear to cause errors.

Even rules that involve some transformation often lead to an implicit truncation that (in this author's opinion) can occasionally lead to hidden errors.

Clearly, boldly generalizing statements such as these must be taken with a grain of salt. This whole area demands the input of the special considerations needed in individual situations.

The area of numerical multiple integration is a very active topic of research for numerical analysts and statisticians. At this time, there are no methods available that offer consistent, reliable results for the integration of general multivariable functions.

A number of techniques have been developed to handle specialized situations. The so-called product rules, that are extensions of the single variable rules, (such as the Product Gaussian rules) have a disadvantage with statistical distributions. Generally, densities have contours that are more like spheres than cubes and in high dimensions. This can lead to considerable waste in function evaluation. Most of programs developed spend a majority of their time with function evaluation and so it is imperative to ensure that such evaluations are made in as efficient a way as is possible.

With the location-scale analysis, all double integration was carried out in the following way. One coordinate was integrated out with a Gaussian rule (DQGn) then the second coordinate was handled by a Simpson's rule(DQSF) that carried along both the value of density function for the marginal distribution of that coordinate along with the distribution function for that coordinate. In one step, we obtain the marginal density and distribution function. The density can be plotted and the distribution function can be used to compute percentage points needed for estimates and confidence intervals. (See Fick (1975) for additional information.)

For higher dimensional integration, a combination of numerical quadrature and importance sampling monte carlo methods were found to be very useful. (A more detailed discussion of this idea is given in Section 4D.)

(iii) Random Number Generation

With the monte carlo integration and other simulation studies done in this chapter and also in Chapter 4, extensive random number generation was used. A very reliable package has been developed at McGill University by G. Marsaglia called SUPER DUPER (Marsaglia (1973)). It uses a combination of a shift register and linear congruential generator. See Knuth (1969) for a very interesting presentation of these generators and their properties.

The rest of this chapter and Chapter 4 are concerned with the extensive use of the tools described in this section.

C. The Performance of the Marginal Likelihood Function

The analysis of the Darwin data in Section A began with the calculation of the observed marginal likelihood function. We anticipated that this function would give us some indication of the range of λ values to be used for inferences concerning μ and σ . Of course the knowledge of the true value of λ would be ideal as the inferential statements about μ and σ would be very firmly based. Clearly the ability of the marginal likelihood function to accurately predict the actual value could be important. We now consider the sampling properties of the likelihood function. Specifically we wish to study the model for possible likelihood functions.

If λ_0 is the true value of λ in a given situation then the probability of the observed $\underset{\sim}{d}$ is determined by its distribution $h_{\lambda_0}(\underset{\sim}{d})$. The likelihood function is computed by

$$L(\underset{\sim}{d} \mid \lambda) = ch_{\lambda}(\underset{\sim}{d}) \quad .$$

In other words, for every λ_0 , there is a probability measure P_{λ_0} on the space of possible likelihood functions. If there is a simple real or vector valued statistic to index the likelihood functions then the study of P_{λ} could be based on the study of the distribution of such statistics.

Typically such statistics are not available. The understanding of some of the properties of these probability measures can be of considerable help in the evaluation of an observed likelihood function.

The analytical properties of $h_\lambda(\underset{\sim}{d})$ and correspondingly of $L(\underset{\sim}{d} \mid \lambda)$ appear to be quite complex and intractible. Accordingly, a simulation study was carried out to assess some of the properties

Attention was restricted to the situation with independent Student(λ) errors. First, some ideas on what is to be expected.

We expect P_{λ_0} to assign higher probability to likelihood functions with modes near λ_0 .

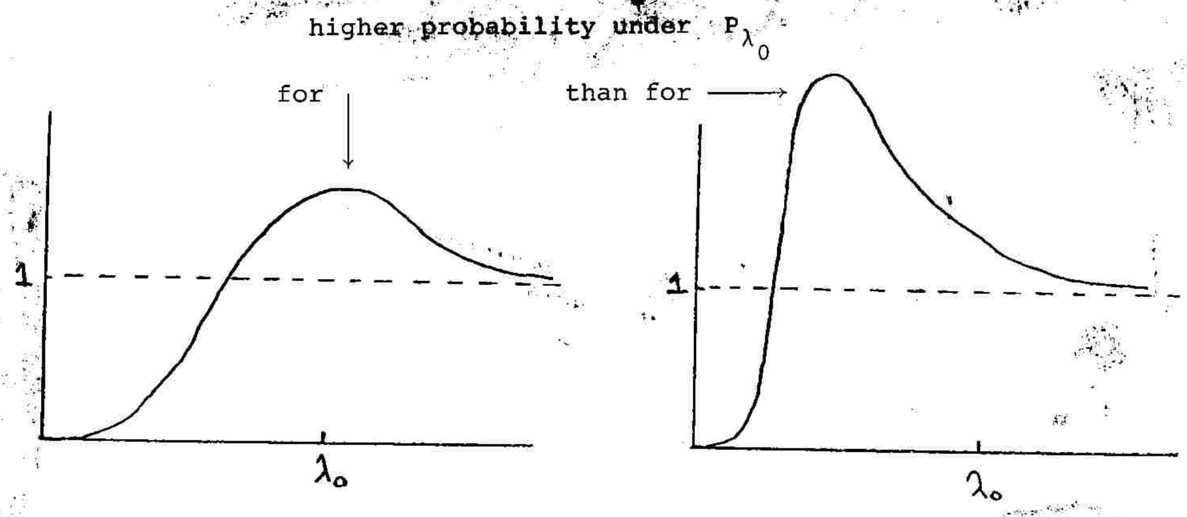


Figure 3.7

We do not expect P_λ to be of a form that will enable us to always make sharp distinctions between candidates for λ . In other words P_λ may be highly nonuniform for some λ values and nearly uniform for others. For example, with a normal error form $h_\infty(\underset{\sim}{d})$ is uniformly distributed on the unit sphere in $L^\perp(\underset{\sim}{1})$ (i.e., all d's are equally likely). Accordingly we expect that for large λ the likelihood function will always be fairly close to 1 — sometimes larger, sometimes smaller since the Student(λ) distributions have contours very similar to the normal distributions (particularly for $\lambda > 30$) .

For low λ values, the contours are considerably different than the normal (recall Figure 2.2, $\lambda = 1 , 2$) . The likelihood function at small λ values can be very large or very small since $h_\lambda(\underset{\sim}{d})$ becomes highly non-uniform.

These qualitative ideas give us guidelines as to the character of the measures P_λ for the possible likelihood functions.

The likelihood function can only be replaced by a set of statistics when the statistics index the functions. There is, however, a pair of statistics that offer considerable insight into character of P_λ .

An estimate of λ can be obtained as the maximizing value $\hat{\lambda}$. We now examine how this maximizing value succeeds in providing inferences approximating those that would be available from the correct true λ value.

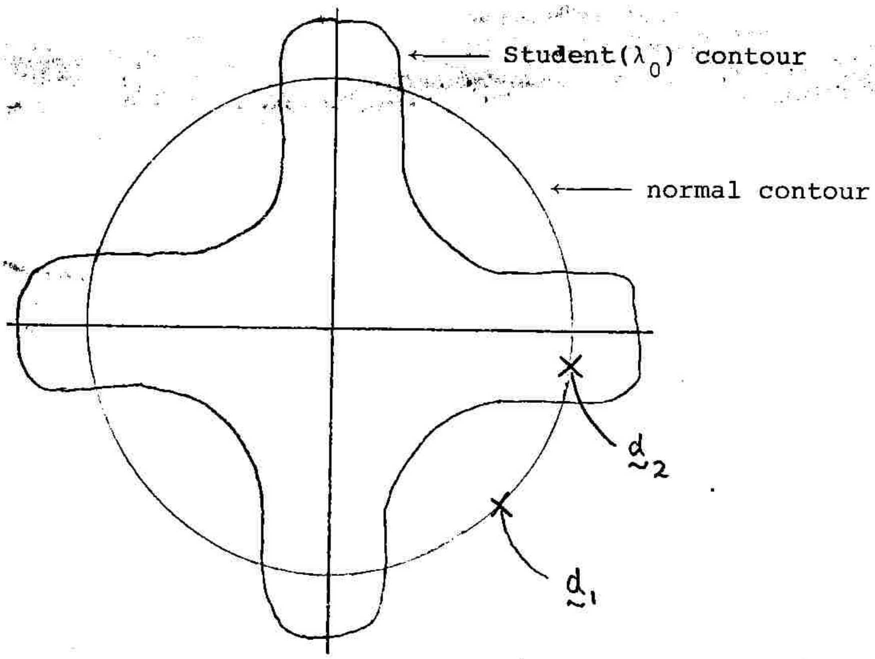


Figure 3.8

With low λ_0 , $h_{\lambda_0}(\underset{\sim}{d}_2)$ would be greater than $h_{\lambda_0}(\underset{\sim}{d}_1)$.

Specifically, 50 samples of size 30 were generated from Student distributions with $\lambda = 1 , 2 , 5 , 8$ and ∞ (the normal distribution). The parameters μ and σ were set at 0 and 1 respectively but these values were not the concern for this part of the study. For each sample, we determined the maximizing value $\hat{\lambda}$ and the apparent precision of $\hat{\lambda}$ as indicated by the likelihood function curvature. The types of likelihood functions depend in a nontrivial way on the configuration $\underset{\sim}{d}$. Typical functions range from very sharply discrimating functions (particularly for low λ values) to practically flat curves giving little indication of the appropriate λ value. One indicator is given by an estimate of Fisher's information function

$$J(\lambda) = -E\left\{\frac{\partial^2}{\partial\lambda^2} \log L(\underset{\sim}{d} \mid \lambda) \mid \lambda\right\} ;$$

the usual estimate is

$$\hat{J} = -\frac{\partial^2}{\partial\lambda^2} \log L(\underset{\sim}{d} \mid \lambda)\bigg|_{\lambda=\hat{\lambda}}$$

and the corresponding estimate of the standard deviation of $\hat{\lambda}$ is

$$\hat{s}_{\hat{\lambda}} = \sqrt{1/\hat{J}} .$$

We are merely estimating the curvature of the log likelihood at $\hat{\lambda}$ anticipating that the shape of this curve will be approximately quadratic.

Typically the likelihood function itself displays far more information than just $\hat{\lambda}$ and $\hat{s}_{\hat{\lambda}}$ with moderate sample sizes. The curves can take widely varying shapes that may give additional clues and directions for the choice of λ .

The Student(λ) distributions change moderately from $\lambda = 5$ to $\lambda = 10$ and then more moderately for $\lambda = 10$ to $\lambda = \infty$. Accordingly we mapped the range $\lambda = 5$ to ∞ into the range 5 to 10 and let λ designate the modified

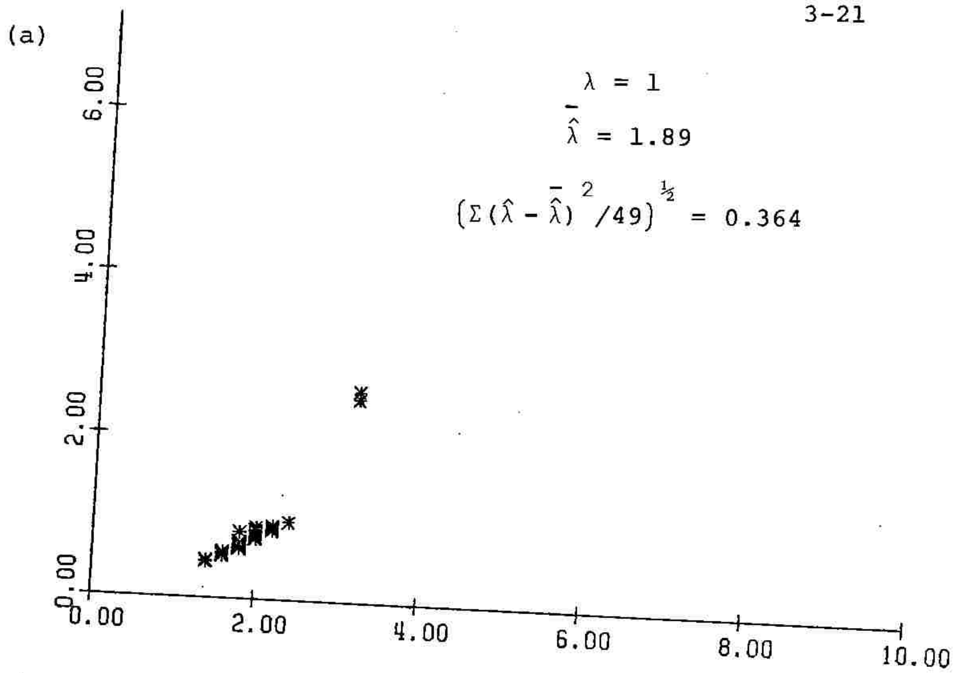
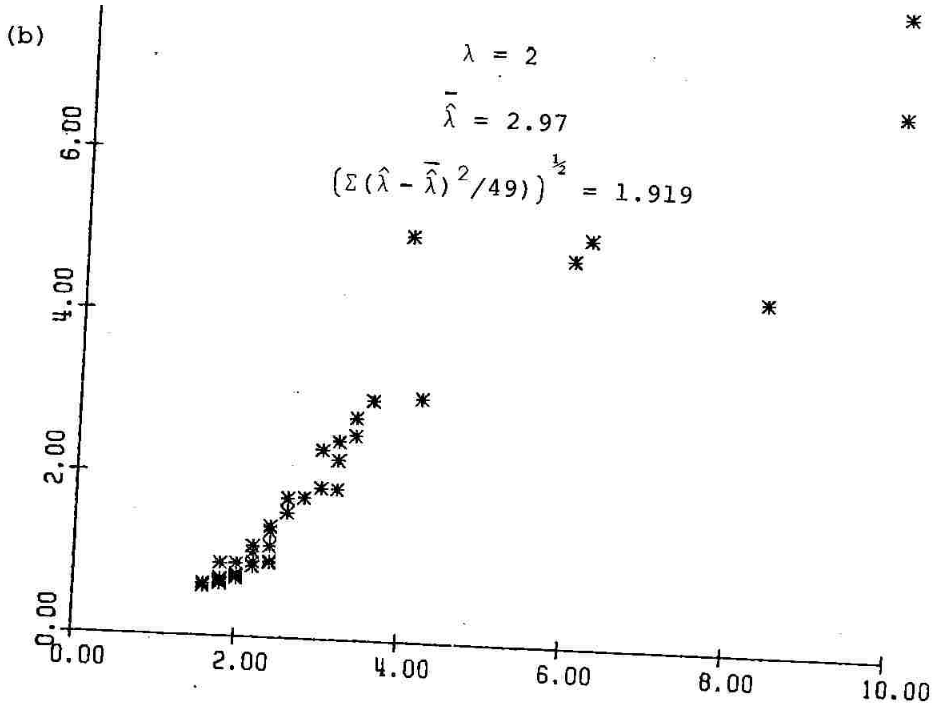
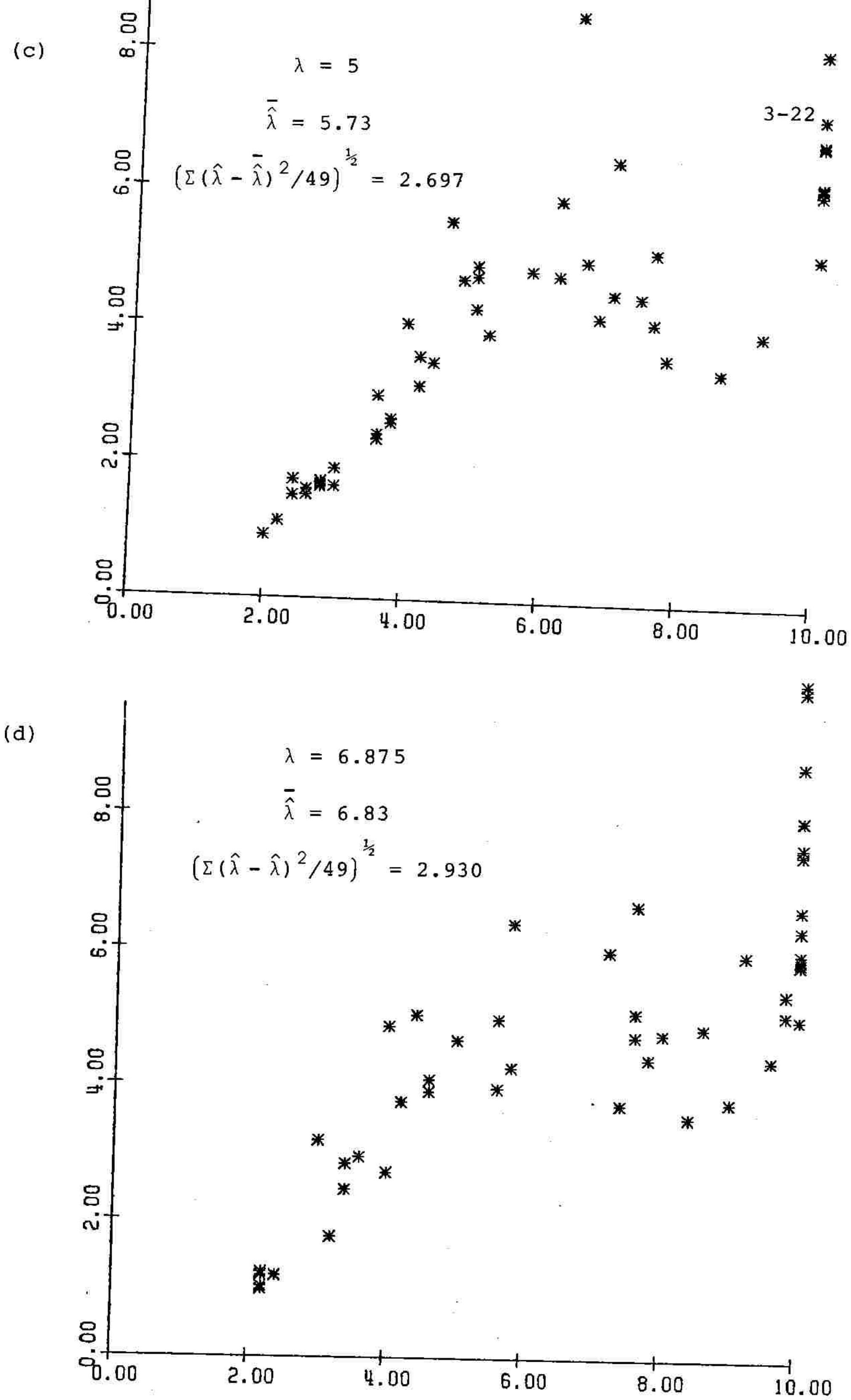
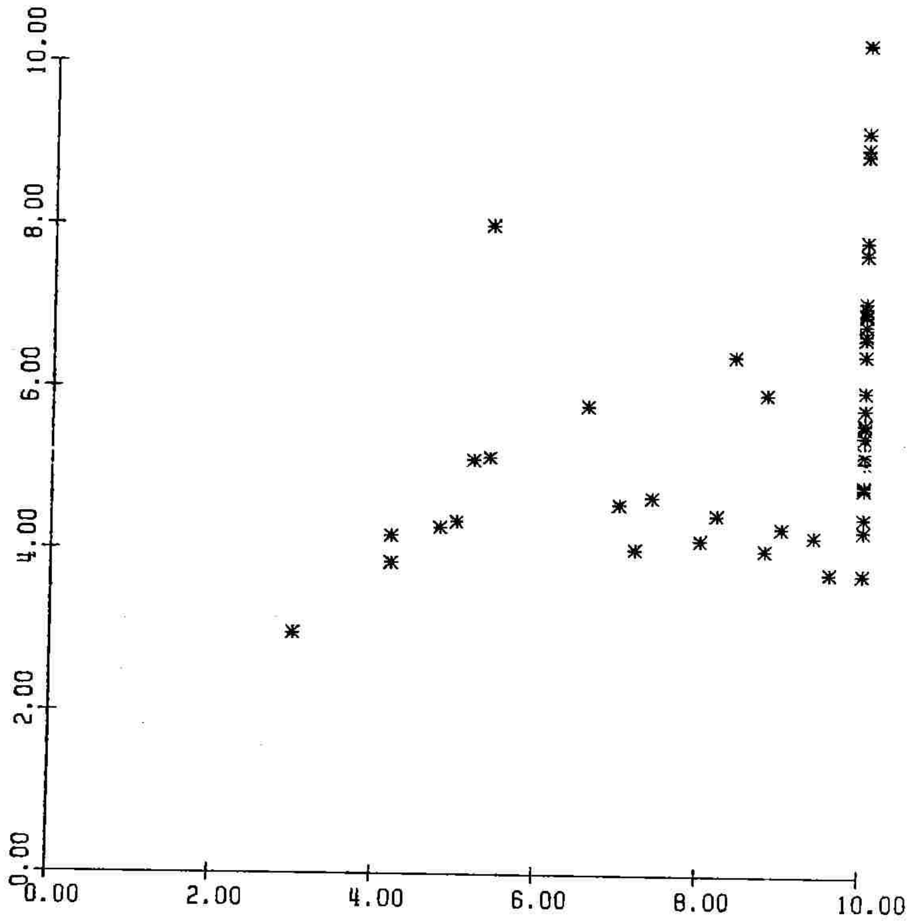


Figure 3.9 Standard error of maximum marginal likelihood estimate vs. estimate (λ = 1 , 2 , 5 , 6.875 , 10)




(c)
$\lambda = 5$
$\bar{\hat{\lambda}} = 5.73$
$\left(\Sigma(\hat{\lambda} - \bar{\hat{\lambda}})^2/49\right)^{\frac{1}{2}} = 2.697$
0.00
2.00
4.00
6.00
8.00
0.00
2.00
4.00
6.00
8.00
10.00
(d)
$\lambda = 6.875$
$\bar{\hat{\lambda}} = 6.83$
$\left(\Sigma(\hat{\lambda} - \hat{\lambda})^2/49\right)^{\frac{1}{2}} = 2.930$
0.00
2.00
4.00
6.00
8.00
0.00
2.00
4.00
6.00
8.00
10.00

(e)



$\lambda = 10$

$\bar{\hat{\lambda}} = 8.59$

$\left(\Sigma(\hat{\lambda} - \bar{\hat{\lambda}})^2/49\right)^{\frac{1}{2}} = 2.110$

parameter and λ_* designate the original parameter. The context will make clear the particular choice. The transformation used from old to new is

$$\begin{aligned} \lambda &= \lambda_* & 0 < \lambda_* \le 5 \\ &= 25 - \frac{10}{\lambda_*} & 5 \le \lambda_* < \infty . \end{aligned}$$

Notice that selected λ values go from $(1 , 2 , 5 , 8 , \infty)$ to $(1 , 2 , 5 , 6.875 , 10)$.

In Figure 3.9 we have plotted $\hat{s}_\lambda$ versus $\hat{\lambda}$ for these five cases; included with each plot is the standard deviation of the estimate

$$\sqrt{\Sigma(\hat{\lambda} - \bar{\hat{\lambda}})^2 / 49} .$$

Clearly if the data came from a Student density with low λ_* value, the likelihood function has a much better chance of suggesting a fairly narrow range of λ values than at high λ_* values.

Approximate 95% intervals could be based on

$$(\hat{\lambda} - 2\hat{s}_{\hat{\lambda}} , \hat{\lambda} + 2\hat{s}_{\hat{\lambda}}) .$$

The actual levels based on this study were

λ	1	2	5	6.875	10
levels	$\frac{50}{50}$	$\frac{50}{50}$	$\frac{48}{50}$	$\frac{43}{50}$	$\frac{49}{50}$.

With $\lambda = 6.875$ and 10 these intervals almost always completely cover $[0 , 10]$ and so for a large number of situations the likelihood is noninformative if data came from a distribution that is either normal or close to normal.

This appears to be very dangerous. In such situations we would likely consult several Student analysis for μ and σ ; perhaps, even ones with low λ values. Fortunately the use of a Student analysis with normal data leads to inferences that are very similar to those derived from a normal analysis. This and several other ideas are addressed in the next section.

TABLE 3.1 STUDENT(3) DATA

$\underset{\sim}{y}$

9.6	9.7	8.9	10.5	11.7
10.1	9.8	9.1	7.6	9.9
8.5	9.6	11.1	8.7	10.9
8.9	9.9	8.3	11.2	11.1
10.3	9.2	11.1	11.9	10.3
11.2	11.8	16.5	10.4	9.5

$\bar{y} = 10.243$

$s_{\underset{\sim}{y}} = 1.6115$

$\underset{\sim}{d}$

-0.0741	-0.0626	-0.1548	0.0296	0.1679
-0.0165	-0.0511	-0.1317	-0.3046	-0.0396
-0.2009	-0.0741	0.0987	-0.1778	0.0757
-0.1548	-0.0396	-0.2239	0.1124	0.0987
0.0065	-0.1202	0.0987	0.1909	0.0065
0.1102	0.1794	0.7209	0.0181	-0.0857

TABLE 3.2 NORMAL DATA

$\underset{\sim}{y}$

10.0003	8.6594	9.7119	9.0439	10.2524
8.0199	10.0567	10.0853	8.9361	10.5039
10.3493	9.9531	11.2711	10.4885	9.7510
8.6350	9.4240	8.1329	11.0372	9.7933
12.1158	9.3683	8.7775	11.1030	10.2973
11.4596	10.5267	11.3192	10.0750	9.1870

$\bar{y} = 9.9445$

$s_{\underset{\sim}{y}} = 1.0071$

$\underset{\sim}{d}$

0.0103	-0.2369	-0.0429	-0.1661	0.0568
-0.3549	0.0207	0.0260	-0.1859	0.1032
0.0746	0.0016	0.2446	0.1003	-0.0357
-0.2414	-0.0960	-0.3340	0.2015	-0.0279
0.4004	-0.1062	-0.2152	0.2136	0.0651
0.2794	0.1074	0.2535	0.0241	-0.1397

D. The Performance of a Student Analysis

In this section, we make a direct comparison between two analyses, the classical normal analysis based on normal errors, and the Student(3) analysis based on standardized Student(3) errors. Both analyses are examined with generated normal data and generated Student(3) data.

The differences between the analyses are displayed by the conditional distributions $g_{\lambda}^{L}(t \mid \underset{\sim}{d})$ and $g_{\lambda}^{S}(s \mid \underset{\sim}{d})$ since all tests and confidence intervals can be derived from them. Many data sets were examined with various sample sizes 20 , 30 , 50 from the Student(3) and the normal distribution. Similar results were found with each sample size. The examples here are for sample size 30 .

For the normal data we took $\mu = 10$ and $\sigma = 1$ and for the Student(3) data we took $\mu = 10$ and $\sigma = 1$.

Representative samples for each case are shown in Tables 3.1 and 3.2.

The distributions for the Student(3) sample are displayed in Figures 3.10 and 3.11.

The distribution for t based on the correct Student(3) model is to the right of the origin and is concentrated, while the distribution for t based on the normal model is of course just the ordinary Student(29) distribution. The Student(3) analysis adjusts depending on the form of the data. The large positive observation (16.5)

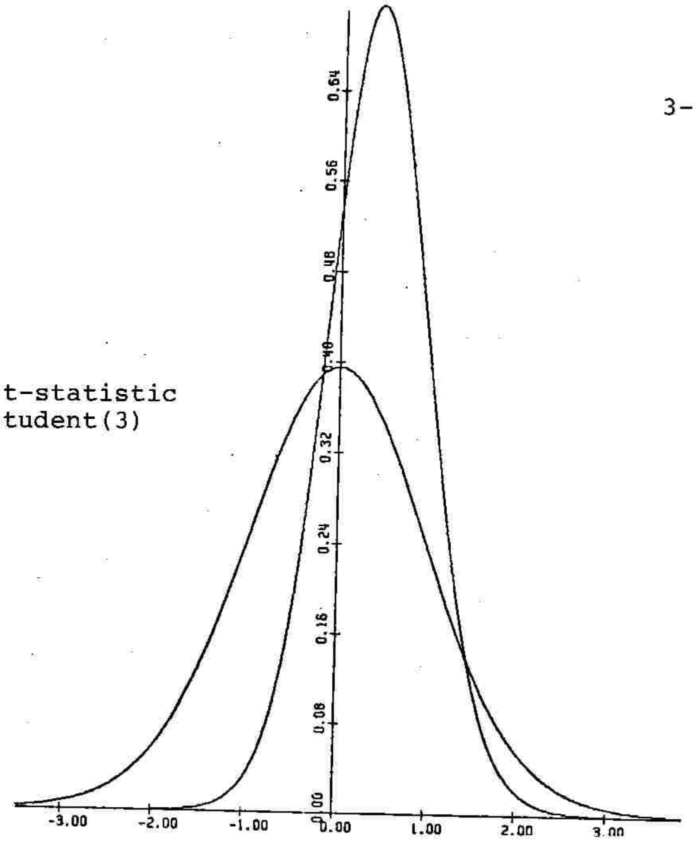


Figure 3.10 The t-statistic densities (Student(3) data)

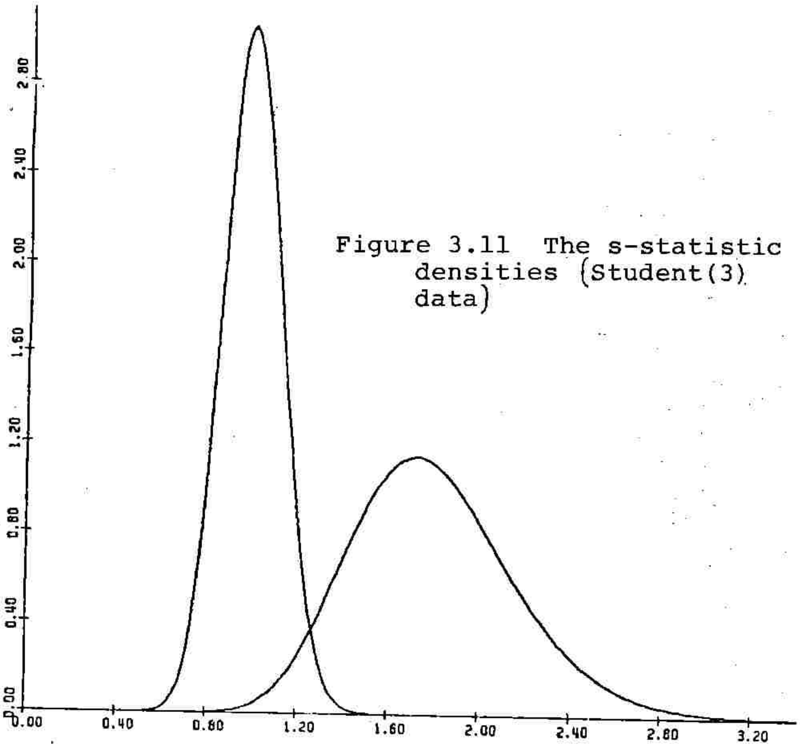


Figure 3.11 The s-statistic densities (Student(3) data)

has contributed to the fact that the sample average is too large and the standard deviation is inflated.

Inferential statements about μ are based on

$$\bar{y} - t s_{\underset{\sim}{y}}/\sqrt{n} = 10.243 - 1.6115\, t/\sqrt{30} \text{ .}$$

Since the conditional distribution for t with the Student(3) model is to the right of the origin, it is correcting for the enlarged sample average. Also, since it is more concentrated it is correcting for the inflated standard deviation.

The distribution for s based on Student(3) is somewhat diffuse in comparison with the $\frac{1}{\sqrt{29}} \chi_{(29)}$ distribution obtained with a normal model.

Inferential statements about σ are based on $s_{\underset{\sim}{y}}/s = 1.6115/s$. Once again the Student(3) analysis is correcting for the inflated standard deviation.

The use of a normal analysis for the data set would have led to very misleading inferences. The appropriate Student(3) analysis appears to correct for the biases that are introduced into $\bar{y}$ and $s_{\underset{\sim}{y}}$ that naturally occur with Student(3) data.

Now let us consider the normal data set. It is certainly reasonable to anticipate that the Student analysis handles Student data properly but what about normal data?

The distributions for t and s are plotted in

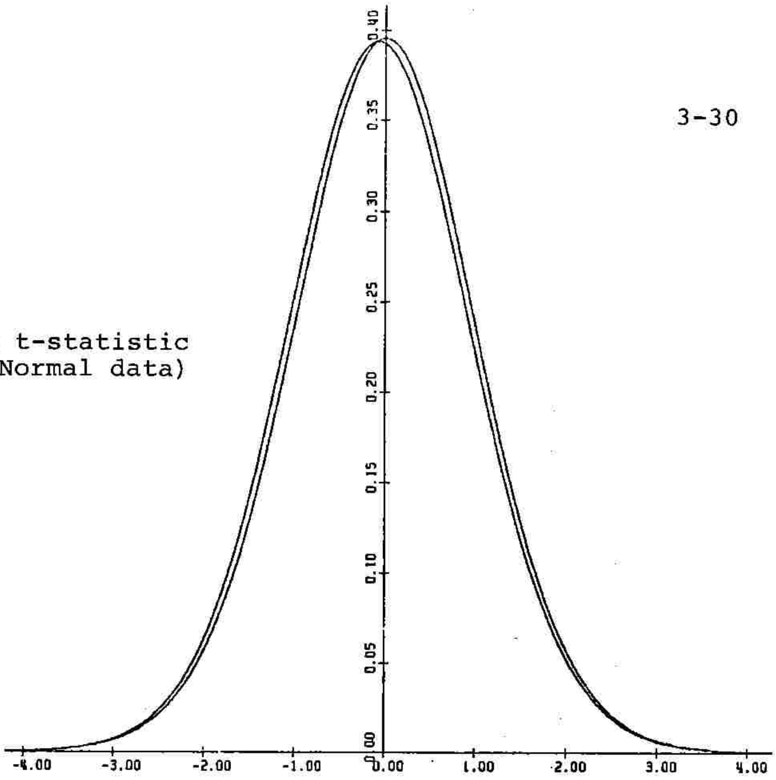


Figure 3.12 The t-statistic densities (Normal data)

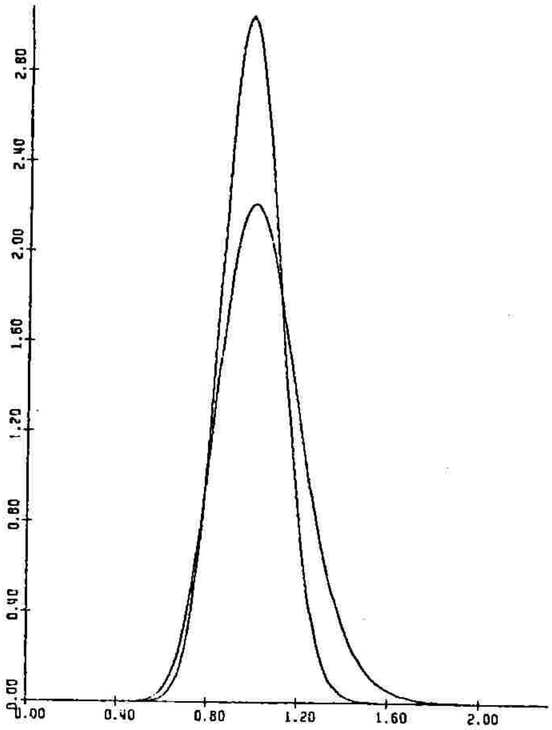


Figure 3.13 The s-statistic densities (Normal data)

Figures 3.12 and 3.13. In both cases the distributions based on the normal model and the Student model are very close. Indeed the estimates and confidence intervals obtained are very close.

The phenomena in these two examples has been observed repeatedly in our study.

Of course, if a sample is known to have come from a particular distribution, we should use the analysis appropriate to that distribution. The computer program that does this is easy and inexpensive to use.

More generally, if a sample is known to have come from the Student family, then we can consider the estimation of λ based on the marginal likelihood function, or we can even contemplate using with moderate confidence a single λ value, say $\lambda = 3$.

It was mentioned that the Student analysis corrects for biases introduced by extreme values that occur naturally with nonnormal data sets. In the next section we consider a method of measuring how accurately Student analysis corrects for extreme values.

E. The Influence of Outlying Observations

We now examine the resistance of the Student location-scale analysis to the presence of outlying observations.

With the Darwin data, there were two observations to the left of the bulk of the data. In the generated Student(3) sample there was one observation off to the right of the majority of the data.

In an applied context, we should ask whether such outlying observations should affect the analysis chosen for it. From a realistic viewpoint, they might have been the most important observations, clues for further experimental investigations.

The viewpoint taken here is that the analysis should routinely handle such extreme values in a way that is reasonable relative to the majority of the data. In fact we focus our concern and examine what affect a deviant observation has on the location-scale analysis.

For this we generated a random sample of observations from the normal(10 , 1) distribution and then carried out the location-scale analysis using various pertubations of one of the initially central observations.

The initial reference data set is as follows:

10.81	9.72	8.64	9.42	9.60
9.18	10.02	8.11	11.11	11.49
9.66	8.15	8.36	9.02	11.09
9.70	10.24	9.89	10.26	8.68
11.98	8.89	10.69	10.45	11.52
8.91	8.23	9.91	9.70	10.21

Statistics calculated from this set will carry the superscript R ; thus $\bar{y}^R$, $s^R_{\underset{\sim}{y}}$, $\underset{\sim}{d}^R$; the observation $y_{29} = 9.70$ was then moved by each of the following amounts:

$$0 - 2\,s^R_{\underset{\sim}{y}} - 4\,s^R_{\underset{\sim}{y}} - 6\,s^R_{\underset{\sim}{y}} - 8\,s^R_{\underset{\sim}{y}} \,.$$

We designate an altered sample by $\underset{\sim}{y}$ with statistics $\bar{y}$, $s_{\underset{\sim}{y}}$, $\underset{\sim}{d}$. We now examine such a data set as one that we might be confronted with in application.

From our data, $\underset{\sim}{y}$, we can make inferences in the direction of confidence intervals and obtain

$$\mu : (\bar{y} - t_2\, s_{\underset{\sim}{y}}/\sqrt{n} \;,\; \bar{y} - t_1\, s_{\underset{\sim}{y}}/\sqrt{n})$$

$$\sigma : (s_{\underset{\sim}{y}}/s_2 \;,\; s_{\underset{\sim}{y}}/s_1)$$

where (t_1 , t_2) , (s_1 , s_2) are appropriate probability intervals from the distributions g^L_λ and g^S_λ which depend on $\underset{\sim}{d}$ except in the normal case.

We now consider how to compare these intervals as the observation is moved out on the left tail of the sample distribution.

We make the comparison by rewriting the confidence intervals in terms of the original inference data set and then examining the intervals or even the corresponding distributions for t and s. The rewritten intervals are

$$\mu : (\bar{y}^R - T_2\, s_{\underset{\sim}{y}}^R / \sqrt{n}\ ,\ \bar{y}^R - T_1\, s_{\underset{\sim}{y}}^R / \sqrt{n})$$

$$\sigma : (s_{\underset{\sim}{y}}^R / S_2\ ,\ s_{\underset{\sim}{y}}^R / S_1)$$

from which we obtain

$$\bar{y}^R - T s_{\underset{\sim}{y}}^R / \sqrt{n} = \bar{y} - t s_{\underset{\sim}{y}} / \sqrt{n}$$

$$s_{\underset{\sim}{y}}^R / S = s_{\underset{\sim}{y}} / s$$

which gives

$$T = \frac{s_{\underset{\sim}{y}}}{s_{\underset{\sim}{y}}^R}\, t + \frac{\bar{y}^R - \bar{y}}{s_{\underset{\sim}{y}}^R / \sqrt{n}} \qquad (3.1)$$

$$S = \frac{s_{\underset{\sim}{y}}^R}{s_{\underset{\sim}{y}}}\, s\ . \qquad (3.2)$$

As the observation is moved out on the left tail, $\bar{y}$ moves out at a rate proportional to $s_{\underset{\sim}{y}}^R$ and becomes a poor estimate of the centre of the distribution and correspondingly the standard deviation $s_{\underset{\sim}{y}}$ becomes inflated in a somewhat obvious pattern. If the intervals based on $\underset{\sim}{y}$ are to remain relatively close to those based on the original $\underset{\sim}{y}^R$ then the conditional distribution for t will have to shift to the left and become more concentrated and the conditional for s will have to become inflated relative to the initial distributions. This phenomena was seen with the Darwin data and also with the computer generated Student(3) data. The question here is whether the distributions shift appropriately to give reasonable resistance to the outlying observation.

We can investigate the proceeding by examining the percentage points of the conditional distributions or even the conditional distributions themselves. For this, we examine the two variables T and S in (3.1) which contain the appropriate corrections to relate their conditional distributions to the original reference data set.

The distributions for T and S are now recorded for several location-scale analyses. Specifically, we record the normal analysis in Figure 3.14, the Student(3) analysis in Figure 3.15 and the Student(1) (Cauchy) analysis in Figure 3.16.

Note that under the normal analysis, the inferences are dramatically altered as the observation shifts to the left tail. By contrast the Student(3) analysis and Student(1)

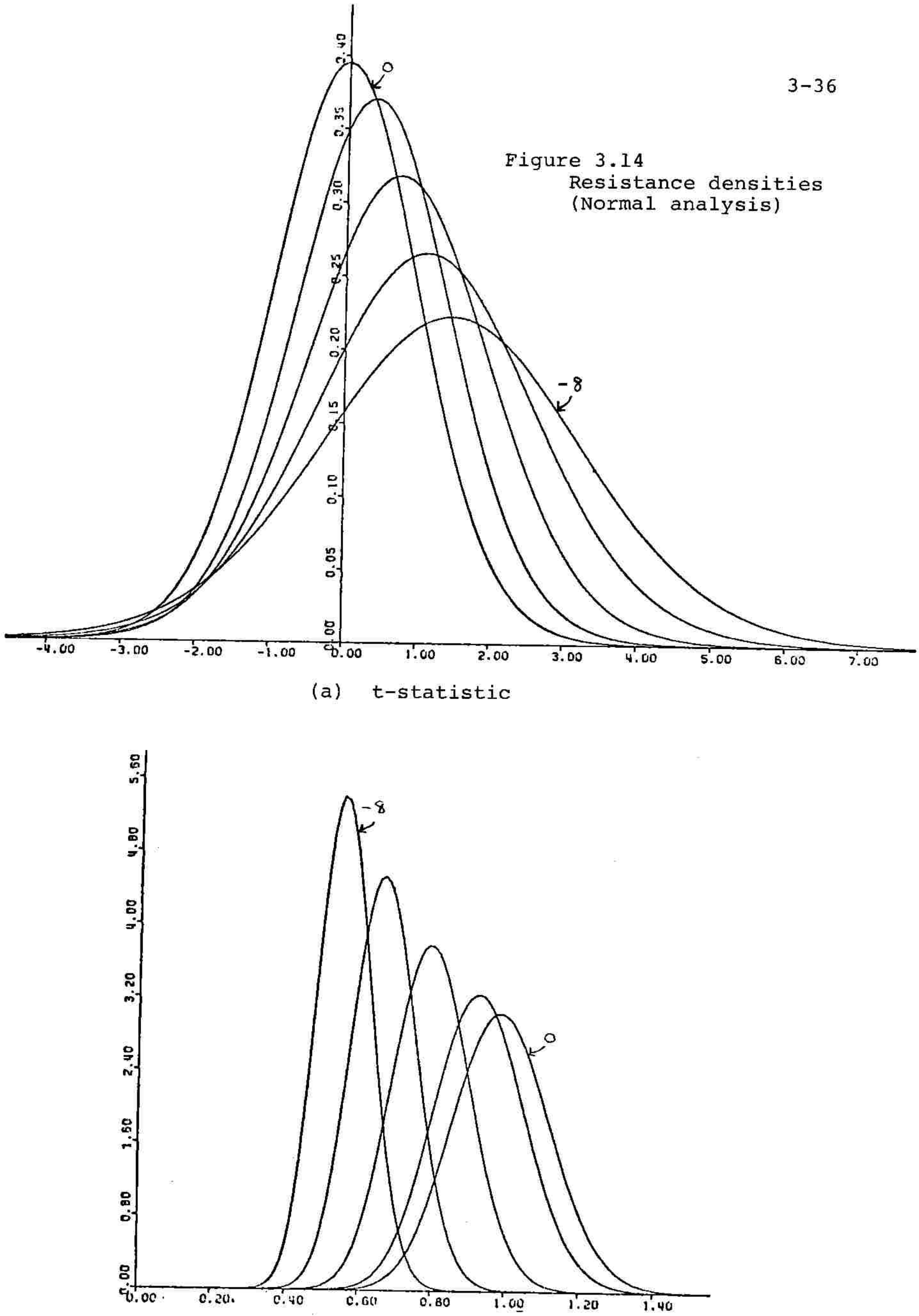


Figure 3.14
Resistance densities
(Normal analysis)

(a) t-statistic

(b) s-statistic

Figure 3.15
Resistance densities
(Student(3) analysis)

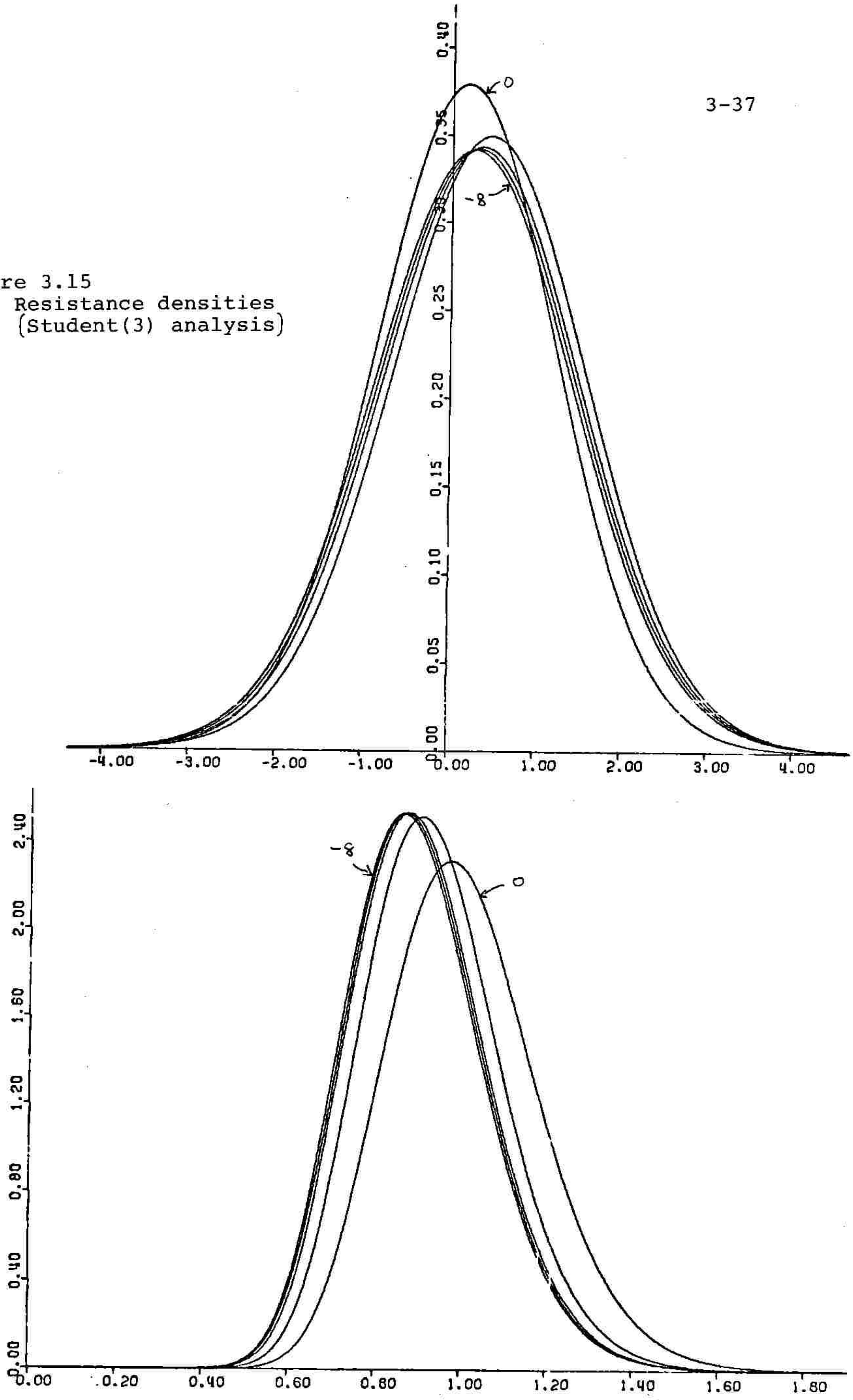
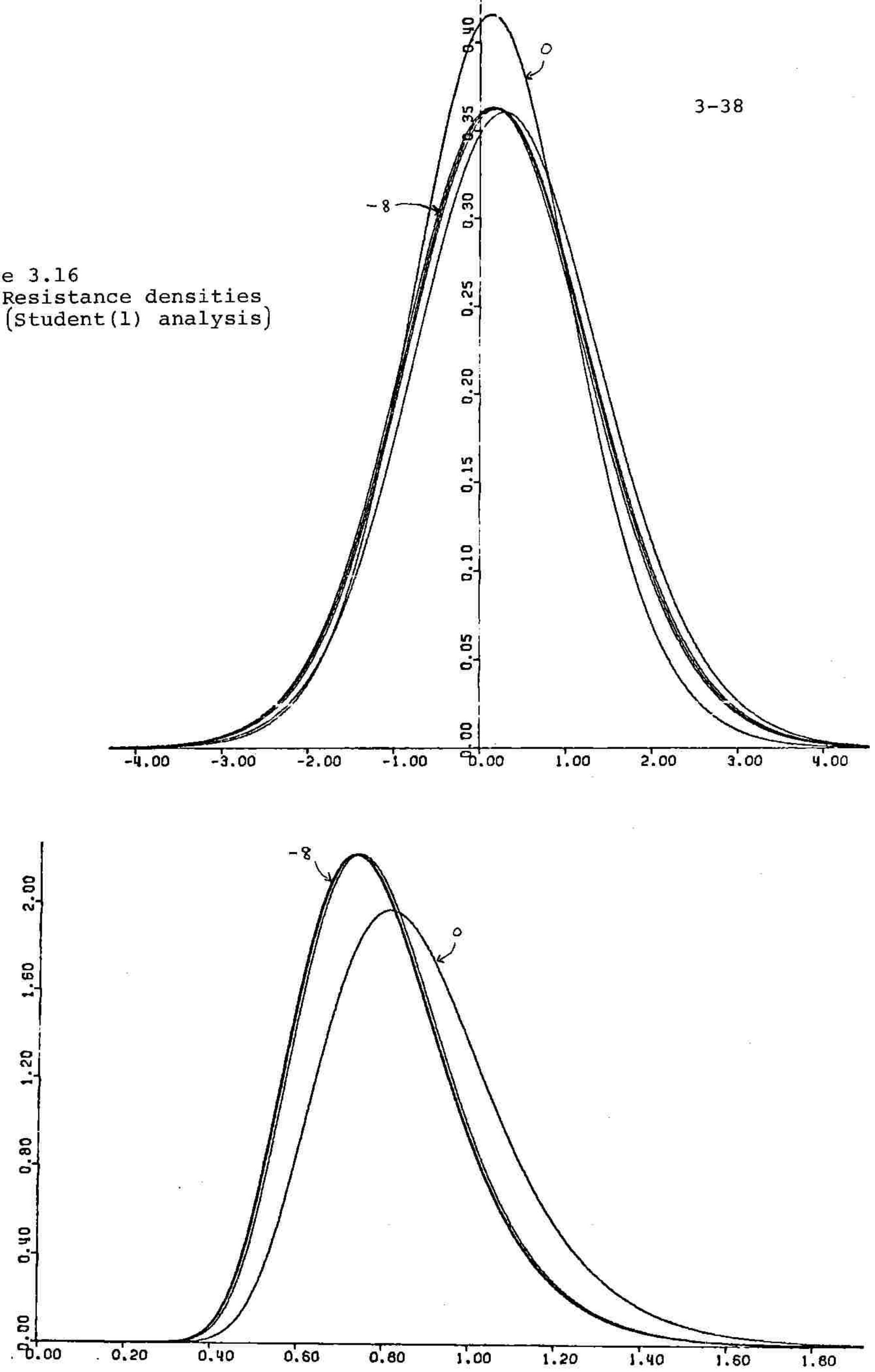


Figure 3.16
Resistance densities
(Student(1) analysis)



analysis are relatively stable. There is of course an initial effect as the observation moves from the centre of the data set changing its primary configuration. But once beyond the centre, the additional effect on the location and scale distributions remains small.

These results were obtained repeatedly with various initial samples.

The displays are of course closely analogous to the influence curves of traditional robust analysis (see Andrews et al (1971)). If we plot the medians of the conditional distributions against the amount of perturbation then the resulting influence curves are remarkably stable.

The plots here however contain much more information than influence curves. Confidence intervals and tests of significance are available from the plots. Asymmetry of various effects is also apparent.

Observations were also moved as far as forty standard deviations from the centre of the distribution with little effect on the Student analyses (see Figures 3.17 and 3.18).

The discussion of this section and Section D have ignored the information that would be supplied by the observed marginal likelihood function. The Student data analyzed in Section D had a sharply discriminating likelihood function strongly indicating a λ range from 2 to 4 . In this section the marginal likelihood function began as essentially flat for

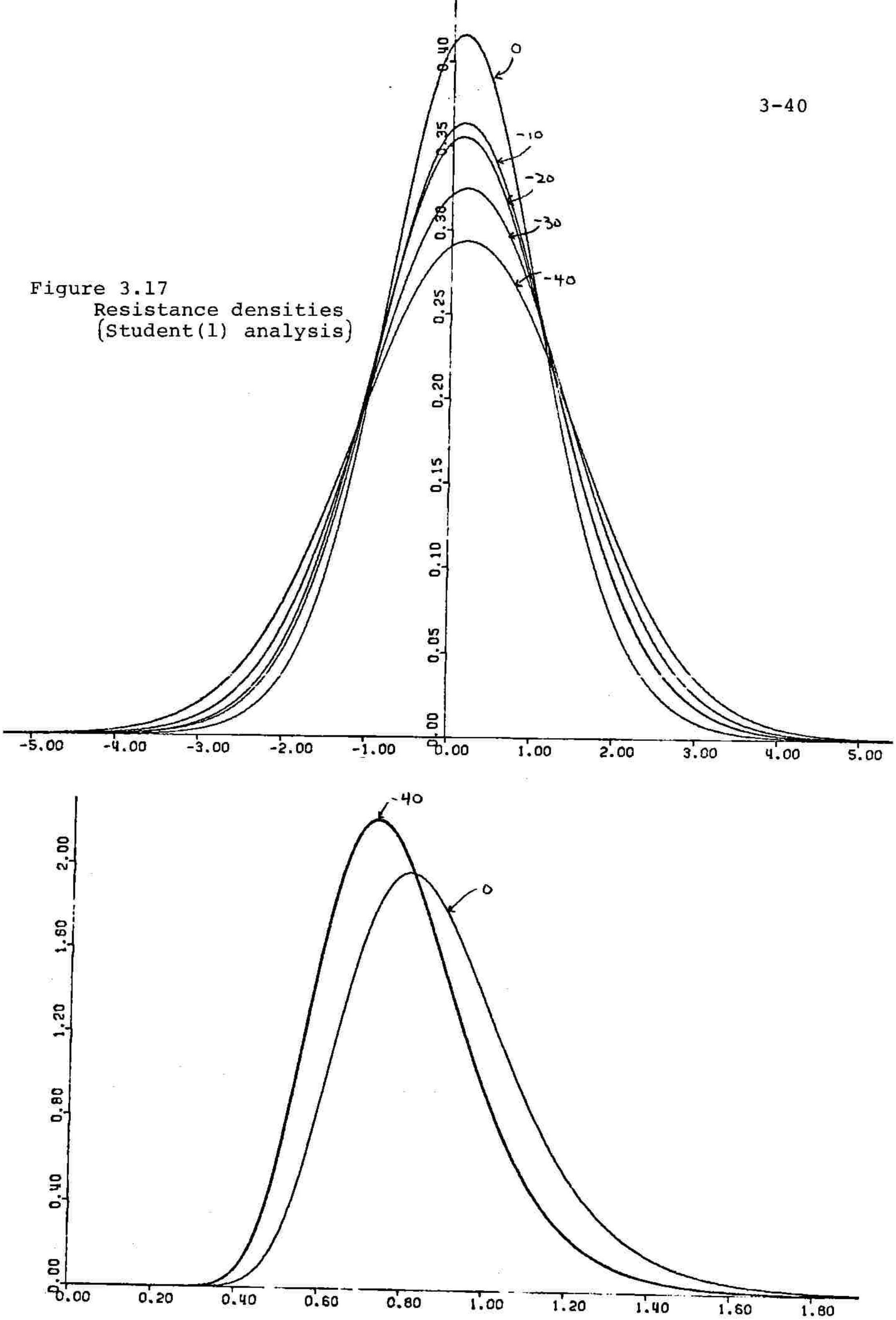
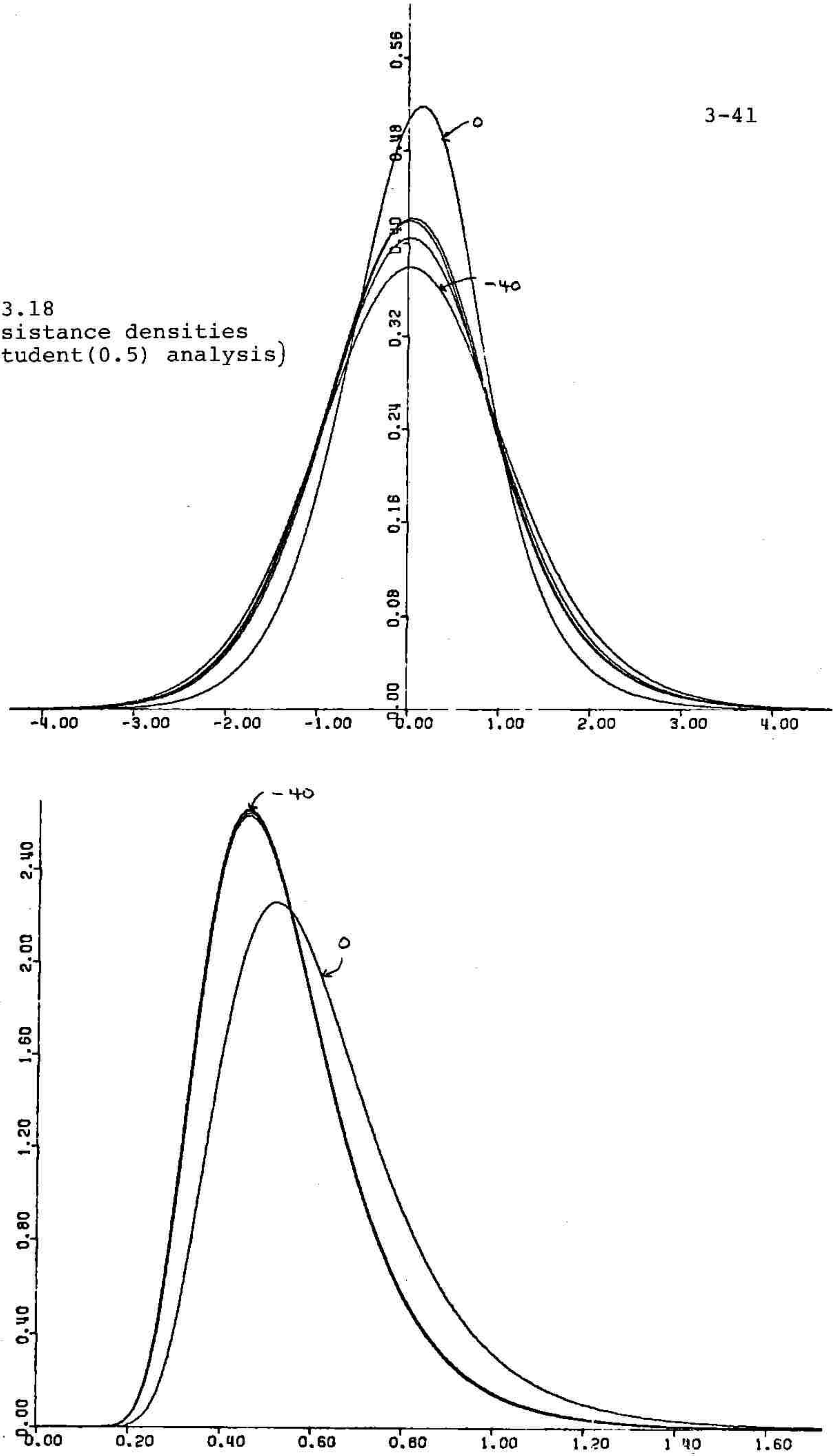


Figure 3.17
Resistance densities
(Student(1) analysis)

Figure 3.18
Resistance densities
(Student(0.5) analysis)

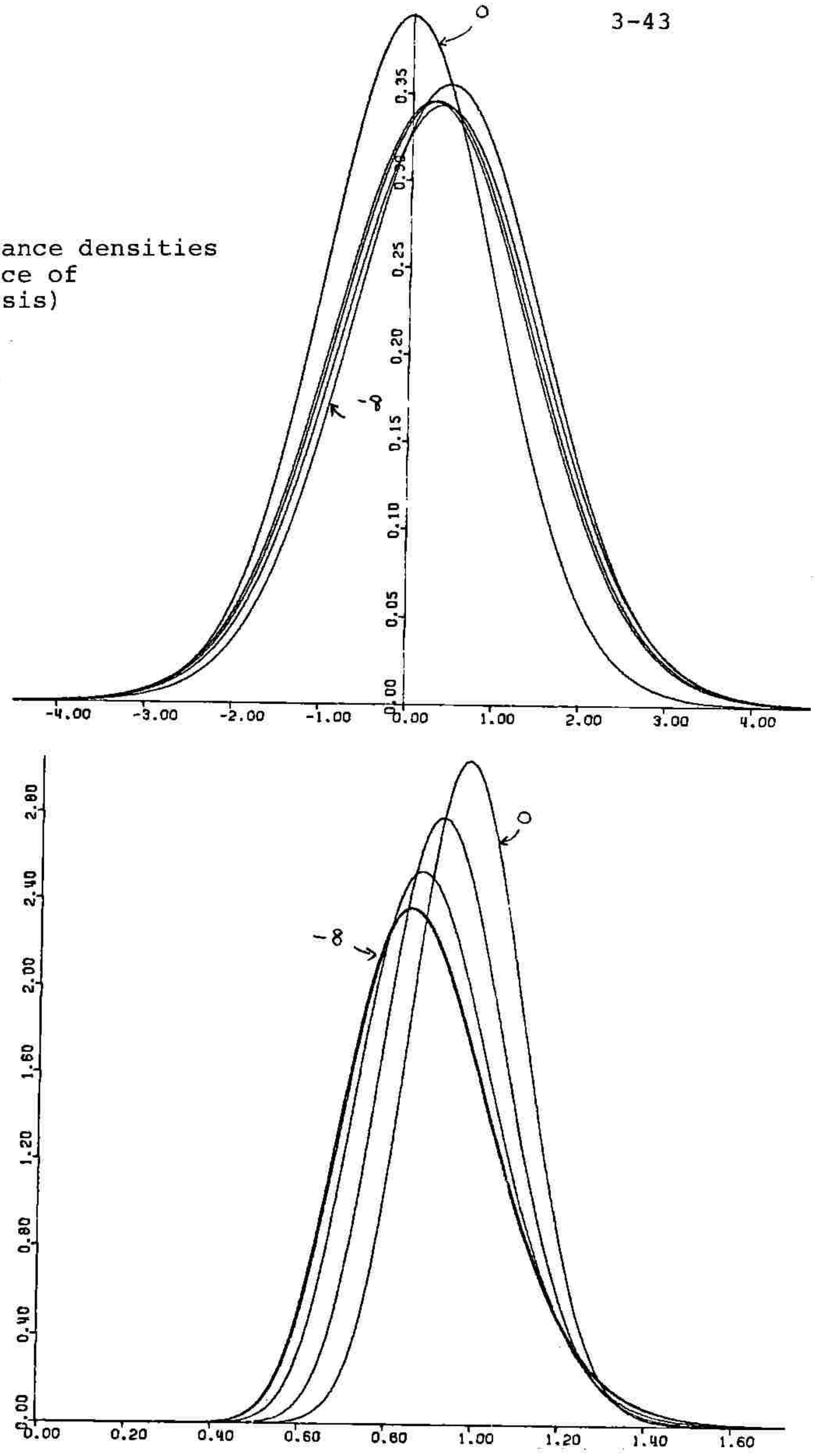


the reference data and then developed modes at lower and lower λ values as the observation was moved out. In fact the marginal likelihoods have modal values according to the following table:

Deviation	0	$-2s_{\underset{\sim}{y}}$	$-4s_{\underset{\sim}{y}}$	$-6s_{\underset{\sim}{y}}$	$-8s_{\underset{\sim}{y}}$
Mode	large	6	3	2	2

If the appropriate analyses were chosen adaptively based on the maximum marginal likelihood estimates we would obtain the adjusted conditional distributions plotted in Figure 3.19. Again, remarkable stability is found. The longer tailed Student distributions can (in a sense) still accommodate a deviant observation without substantially affecting the estimates of the two unknowns needed to describe a fit to the data.

Figure 3.19 Resistance densities
(adaptive choice of
Student analysis)



F. The Dependence on the Deviation Vector

It has been mentioned in several instances that the conditional distributions depend on the deviation vector in a nontrivial way. As soon as the joint error distribution becomes nonspherical, the distributions adjust accordingly based on the data set through $\underset{\sim}{d}$. The attributes of this phenomenon have been discussed in earlier sections, but the direct analytical study of these characteristics appears to be quite complex. To understand the properties of the marginal likelihood function appears to be complicated also.

In this section we discuss one attempt to understand how a nonnormal analysis is affected by different configurations. The method is based on an attempt to emulate the actual contours of a joint long tailed error distribution by a distribution in which the analysis can be handled analytically.

The contours for a Student error form were displayed in Figure 2.2. The distinctive characteristics of this pattern are the lobes extending along each coordinate axis. We can try to copy these lobes with normal distributions having one coordinate rescaled. The scaled normal distributions were discussed in Section 2G.

Consider the following distribution.

$$f_{\Sigma_1}(z_1\ , z_2) = \frac{1}{2\pi\sqrt{\tau_1}}\ \exp\left\{-\frac{z_1^2}{2\tau_1}\right\}\exp\left\{-\frac{z_2^2}{2}\right\} \qquad \Sigma_1 = \begin{pmatrix}\tau_1 & 0\\ 0 & 1\end{pmatrix}\ .$$

If τ_1 is suitably chosen (greater than 1) then the ellipsoidal contours will approximate the lobe along z_1 found with 2 dimensional Student contours.

To gain an approximation to the lobe along z_2 we can place a second rescaled normal on the first.

$$c_1 f_{\Sigma_1}(z_1\ , z_2) + c_2 f_{\Sigma_2}(z_1\ , z_2) = \frac{c_1}{2\pi\sqrt{\tau_1}}\ \exp\left\{-\frac{z_1^2}{2\tau_1}\right\}\exp\left\{-\frac{z_2^2}{2}\right\}$$

$$+\ \frac{c_2}{2\pi\sqrt{\tau_2}}\ \exp\left\{-\frac{z_1^2}{2}\right\}\exp\left\{-\frac{z_2^2}{2\tau_2}\right\} \qquad \begin{array}{l} c_1 + c_2 = 1 \\ c_1\ , c_2 \geq 0\end{array}$$

$$\Sigma_1 = \begin{pmatrix}\tau_1 & 0\\ 0 & 1\end{pmatrix} \qquad \Sigma_2 = \begin{pmatrix}1 & 0\\ 0 & \tau_2\end{pmatrix}\ .$$

An example is plotted in Figure 3.20.

In n dimensions, we can consider the distribution

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

where

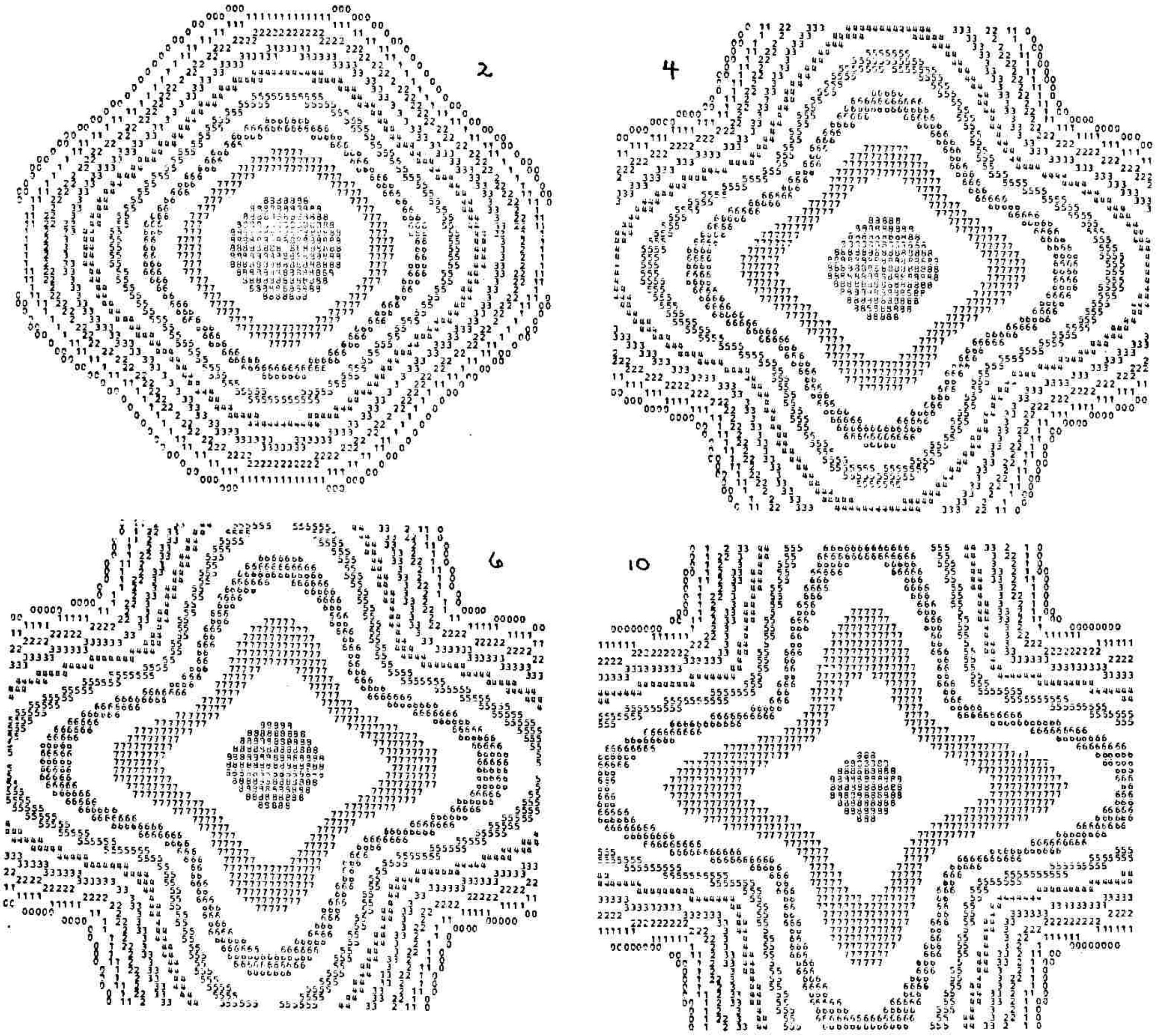


Figure 3.20 Contours of a sum of two rescaled normal densities

$c_1 = c_2 = 0.5$, $\tau_1 = \tau_2 = \tau$; $\tau = 2 , 4 , 6 , 10$

$$f_{\Sigma_k}(\underset{\sim}{z}) = \frac{1}{(2\pi)^{n/2}\sqrt{\tau_k}} \prod_{i\neq k} \exp\left\{-\frac{z_i^2}{2}\right\}\exp\left\{-\frac{z_k^2}{2\tau_k}\right\} . \tag{3.3}$$

$$\Sigma_k = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & \mathcal{O} & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ & & & & \tau_k & & & \\ & & & & & 1 & & \\ & \mathcal{O} & & & & & \ddots & \\ & & & & & & & 1 \end{pmatrix} . \tag{3.4}$$

We can compute the conditional distributions and marginal likelihood analytically for this case. We begin by displaying the distributions and likelihood for an error distribution with one coordinate rescaled $f_{\Sigma_k}(\underset{\sim}{z})$. This only involves the specialization of the results from Section 2G to the location-scale model and the matrix $\Sigma = \Sigma_k$. We require $(V'\Sigma^{-1}V)^{-1}V'\Sigma^{-1}\underset{\sim}{d}$, $\underset{\sim}{d}'R\underset{\sim}{d}$ and W to specialize the expressions

$$V'\Sigma^{-1}\underset{\sim}{d} = \frac{1}{\sqrt{n}}\,\underset{\sim}{1}'(d_1 , \ldots , d_k/\tau_k , \ldots , d_n)$$

$$= \frac{1}{\sqrt{n}} \sum_{i\neq k} d_i + d_k/\tau_k \;=\; \frac{1}{\sqrt{n}}\left(\frac{1}{\tau_k} - 1\right) d_k .$$

$$V'\Sigma^{-1}V = \frac{1}{n}\,\underset{\sim}{1}'\left(1\,,\,\ldots\,,\,\frac{1}{\tau_k}\,,\,\ldots\,,\,1\right)' = \frac{1}{n}\left(n + \left(\frac{1}{\tau_k} - 1\right)\right)$$

$$= 1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)$$

$$(V'\Sigma^{-1}V)^{-1}\,V'\Sigma^{-1}\underset{\sim}{d} = \frac{\frac{1}{\sqrt{n}}\left(\frac{1}{\tau_k} - 1\right)d_k}{1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)}\;. \tag{3.5}$$

$$\underset{\sim}{d}'\Sigma^{-1}\underset{\sim}{d} = 1 + \left(\frac{1}{\tau_k} - 1\right)d_k^2\;. \tag{3.6}$$

$$\underset{\sim}{d}'\left(\Sigma^{-1}V(V'\Sigma^{-1}V)^{-1}\,V'\Sigma^{-1}\right)\underset{\sim}{d} = \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)^2 d_k^2 \Big/ \left(1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right)$$

$$\underset{\sim}{d}'R\underset{\sim}{d} = \left(1 + \left(\frac{1}{\tau_k} - 1\right)d_k^2\right) - \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)^2 d_k^2 \Big/ \left(1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right)$$

$$= 1 + \frac{\left(\frac{1}{\tau_k} - 1\right)d_k^2}{\left(1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right)}\;.$$

$$W = \left(1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right)^2 \Big/ \left(1 + \left(\frac{1}{n} + d_k^2\right)\left(\frac{1}{\tau_k} - 1\right)\right)\;. \tag{3.7}$$

From Section 2G, we see that if the error form is $f_{\Sigma_k}(\underset{\sim}{z})$ then the conditional distribution for T given $\underset{\sim}{d}$ is $\text{Student}_{n-1}(\nu\,,\,W)$ on $\mathbb{R}^1$ where

$$\nu = -\frac{1}{\sqrt{n}}\left(\frac{1}{\tau_k} - 1\right)d_k \Bigg/ \left[1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right]$$

$$W = \left[1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right]^2 \Bigg/ \left[1 + \left(\frac{1}{n} + d_k^2\right)\left(\frac{1}{\tau_k} - 1\right)\right] .$$

We shall call this density $g^L_{\Sigma_k}(T \mid \underset{\sim}{d})$.

The conditional for s given $\underset{\sim}{d}$ can be described as

$$\left(1 + \frac{\left(\frac{1}{\tau_k} - 1\right)d_k^2}{\left[1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right]}\right)^{\frac{1}{2}} s \sim \chi_{n-1} .$$

We shall call this density $g^S_{\Sigma_k}(s \mid \underset{\sim}{d})$.

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

$$h_{\Sigma_k}(\underset{\sim}{d}) = \frac{1}{A_{n-1}} \cdot \frac{\tau_k^{-\frac{1}{2}}\left[1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right]^{-\frac{1}{2}}}{\left(1 + \frac{\left(\frac{1}{\tau_k} - 1\right)d_k^2}{\left[1 + \frac{1}{n}\left(\frac{1}{\tau_k} - 1\right)\right]}\right)^{\frac{n-1}{2}}} . \tag{3.8}$$

When the error distribution is a sum of rescaled normals $\sum_{i=1}^{n} c_i f_{\Sigma_i}$, the conditional distribution for T given $\underset{\sim}{d}$ is then

$$\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d}) g^L_{\Sigma_i}(T \mid \underset{\sim}{d}) \Big/ \left[\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d})\right] ,$$

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

$$\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d}) g^S_{\Sigma_i}(s \mid \underset{\sim}{d}) \Big/ \left(\Sigma c_i h_{\Sigma_i}(\underset{\sim}{d})\right)$$

and the marginal likelihood for $(\tau_1 , \ldots , \tau_n)$ is proportional to

$$\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d}) .$$

From extensive study of the conditional distributions derived from the Student models, we have seen several types of resulting distributions. Depending on the configuration $\underset{\sim}{d}$, the t densities can be highly concentrated or even diffuse. They can be substantially shifted from the origin and are often clearly asymmetric.

The densities $g^L_{\Sigma_k}$ are all symmetric but the densities based on $\Sigma c_i f_{\Sigma_i}$ will typically be asymmetric. For example, let us suppose that y_k is an outlying observation in the right tail, then we would find that d_k will be large and positive. If τ_k is greater than 1 , then $g^L_{\Sigma_k}$ will have its mode to the right of the origin (positive ν) and be concentrated ($W < 1$) [See equations (3.5) and (3.7).]

This is precisely the phenomenen to be found with a Student analysis based on low λ values.

With location-scale models, it is often the presence of one or two outlying observations on one tail that can make the use of an appropriate Student analysis critical for stable inferences. In these situations, it has been found that the use of a single f_{Σ_k} can offer very reasonable approximations where k is chosen to correspond to the most deviant observation, y_k (with corresponding d_k). This amounts to the use of a relocated and rescaled Student distribution for t. Approximations somewhat similar to these and derived from another point of view have been suggested by Lund (1967). When the error form is exponential power (Section 2E), he suggested a relocated and rescaled Student$\left\{n(\frac{\lambda}{2}) - 1\right\}$ distribution. In his investigation, the mode of the t density derived from an exponential power analysis is found numerically. The appropriate approximating density is located at this modal value. Unfortunately, all of these distributions are symmetric.

The distributions derived in this section have the τ and c values unspecified. The choice of these values would typically be made in some iterative way and would depend on the type and form of approximation desired.

One reasonable method would be to match chosen contours based on $\prod_{i=1}^{n} f_{\lambda}(z_i)$ with those based $\sum_{i=1}^{n} c_i f_{\Sigma_i}(\tilde{z})$ say along coordinate axes. Another method is discussed in Section G.

Notice that the conditional distributions derived from $\sum_{i=1}^{n} c_i f_{\Sigma_i}$ are sums of densities weighted by the values of $h_{\Sigma_i}(\tilde{d})$. Each probability $h_{\Sigma_k}(\tilde{d})$ depends only on d_k^2 . It is clear that if one or two d_k^2 are large then they will have a dominating effect on the form of the resulting densities. This give further support for the use of a single rescaled normal $f_{\Sigma_k}(\tilde{z})$ in certain situations.

We now turn our attention to the study of the marginal likelihood function. In Section C we attempted to understand some of the sampling properties of the likelihood function through simulation. The model for possible likelihood functions is difficult to study since there is no simple statistic available to index such likelihood functions.

We now study the model for possible likelihood functions when the error form is $f_{\Sigma_k}(\tilde{z})$. In this case there is a simple real valued statistic available to index the functions and its distribution is now derived.

The marginal distribution $h_{\Sigma_k}(\tilde{d})$ is displayed in expression (3.8). Notice that the distribution for $\tilde{d}$ depends only on the square of its kth coordinate, d_k^2 .

The marginal likelihood for τ_k is then

$$L(\underset{\sim}{d} \mid \tau_k) = c h_{\Sigma_k}(\underset{\sim}{d}) \ .$$

We now choose the representative curve that enables direct comparisons with $\tau_k = 1$ corresponding to identically distributed normal variables

$$L(\underset{\sim}{d} \mid \tau_k) = A_{n-1} h_{\Sigma_k}(\underset{\sim}{d}) \ .$$

This likelihood function depends only on d_k^2 and its associated marginal distribution. There is a one to one correspondence between d_k^2 and $L(\underset{\sim}{d} \mid \tau_k)$.

If we can find the distribution for d_k^2 for general τ_k , then we will, in effect, have described the model for possible likelihood functions for this case.

The likelihood function derived from the joint distribution for $\underset{\sim}{d}$ is the same as the likelihood function derived from the marginal distribution for d_k^2 , $\left[\text{say } f_{\tau_k}^{k}(d_k^2)\right]$

$$L(\underset{\sim}{d} \mid \tau_k) = L_k(d_k^2 \mid \tau_k) \ .$$

The distribution for d_k^2 can then be written as

$$f_{\tau_k}^{k}(d_k^2) = g(d_k^2) L_k(d_k^2 \mid \tau_k)$$

for some function g dependent only on d_k^2 . But

$$f_1^k(d_k^2) = g(d_k^2)L_k(d_k^2 \mid 1)$$

and since $L_k(d_k^2 \mid 1) = 1$ we have that

$$f_{\tau_k}^k(d_k^2) = f_1^k(d_k^2)L_k(d_k^2 \mid \tau_k) \ . \tag{3.9}$$

This is called the likelihood modulation technique. (See Fraser (1968), p. 196 for another example.)

To display the distribution $f_{\tau_k}^k(d_k^2)$, all we require is $f_1^k(d_k^2)$ to complete the formula.

Cramér (1946) derives the density $f_1^k(d_k^2)$ (p. 389) in a different context. It can be described as

$$\frac{d_k^2}{1-\frac{1}{n}} \sim \text{beta}\left(\frac{1}{2}, \frac{n-2}{2}\right) \quad (\text{when} \ \ \tau_k = 1) \ .$$

We now include its simple derivation. Let $\underset{\sim}{z}$ denote a sample from $N(0 , 1)$. Replace $\underset{\sim}{z}$ by new variables $\underset{\sim}{x}$ by means of an orthogonal transformation where

$$x_1 = \sqrt{n}\ \bar{z}$$

$$x_2 = \frac{1}{\sqrt{1-\frac{1}{n}}}(z_k - \bar{z})$$

then

$$\frac{d_k^2}{1-\frac{1}{n}} = \frac{(z_k - \bar{z})^2 \Big/ \left(1-\frac{1}{n}\right)}{s^2(\underset{\sim}{z})}$$

which is

$$\frac{x_2^2}{x_2^2 + \sum\limits_{i=3}^{n} x_i^2}$$

which is distributed as a

$$\chi^2_{(1)} \Big/ \left(\chi^2_{(1)} + \chi^2_{(n-2)}\right) \text{ variable}$$

and which is $\text{beta}\left(\frac{1}{2}, \frac{n-2}{2}\right)$.

Notice that

$$-\sqrt{1-\frac{1}{n}} \le d_k \le \sqrt{1-\frac{1}{n}} .$$

The deviation vector of the form

$$\left(\mp 1/\sqrt{n}\sqrt{n-1} , \ldots , \underset{\substack{\uparrow \\ k^{\text{th}} \text{ coordinate}}}{\pm \sqrt{1-\frac{1}{n}}} , \ldots , \mp 1/\sqrt{n}\sqrt{n-1} \right)$$

corresponds to a data set $\underset{\sim}{y}$ with all coordinates equal but y_k .

[One can also show that d_k has a relocated and rescaled symmetric beta$\left(\frac{n-2}{2}, \frac{n-2}{2}\right)$ distribution on $\left(-\sqrt{1-\frac{1}{n}}, \sqrt{1-\frac{1}{n}}\right)$.]

Using equation (3.8), we can now display the distribution for d_k^2.

$$f_{\tau_k}^{k}(d_k^2) = \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \cdot \frac{\tau_k^{-\frac{1}{2}}\left\{1+\frac{1}{n}\left(\frac{1}{\tau_k}-1\right)\right\}^{-\frac{1}{2}}}{\left\{1+\frac{\left(\frac{1}{\tau_k}-1\right)d_k^2}{1+\frac{1}{n}\left(\frac{1}{\tau_k}-1\right)}\right\}^{\frac{n-1}{2}}}$$

$$\cdot \frac{\left(1-\frac{1}{n}\right)^{-\frac{1}{2}}}{\left(d_k^2\right)^{\frac{1}{2}}}\left(1-\frac{d_k^2}{1-\frac{1}{n}}\right)^{\frac{n-2}{2}-1} \qquad 0 \le d_k^2 \le 1-\frac{1}{n}.$$

This distribution describes the sampling properties of the marginal likelihood function when the error form is $f_{\Sigma_k}(\underset{\sim}{z})$.

This distribution can be displayed in terms of a beta distribution also. In fact if $u = \frac{d_k^2}{1-\frac{1}{n}}$ then the density for u is

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)}\left(1-\frac{(1-a)u}{1-au}\right)^{\frac{n-2}{2}-1}\left(\frac{(1-a)u}{1-au}\right)^{-\frac{1}{2}}\frac{1}{(1-au)^2}$$

where

$$a = \frac{\left(1-\frac{1}{\tau_k}\right)\left(1-\frac{1}{n}\right)}{1+\frac{1}{n}\left(\frac{1}{\tau_k}-1\right)} .$$

In other words, $\frac{(1-a)u}{1-au} \sim \text{beta}\left(\frac{1}{2}, \frac{n-2}{2}\right)$ or

$$\frac{\left(\frac{1}{\tau_k}\right) \Big/ \left(1+\frac{1}{n}\left(\frac{1}{\tau_k}-1\right)\right)\frac{d_k^2}{1-\frac{1}{n}}}{1-\frac{\left(1-\frac{1}{\tau_k}\right)}{1+\frac{1}{n}\left(\frac{1}{\tau_k}-1\right)}d_k^2} \sim \text{beta}\left(\frac{1}{2}, \frac{n-2}{2}\right) .$$

If the error form is $\sum_{i=1}^{n} c_i f_{\Sigma_i}(\underset{\sim}{z})$, then the marginal likelihood function will be

$$\sum_{i=1}^{n} c_i L_i(d_i^2 \mid \tau_i) .$$

G. Importance Sampling Monte Carlo

Several forms of monte carlo integration have received considerable attention in the literature (see, for example, Hammersley and Handscomb (1964) for an extensive summary). In certain situations, they can give results that are more accurate and more efficiently obtained than the so-called fixed point numerical quadrature rules. Used properly, they can lead to substantial reduction in the function evaluations necessary to evaluate an integral. Unless a substantial amount of information is available about a given integrand, most nonadaptive fixed point rules can spend considerable amounts of computer time evaluating a function at points that offer little contribution to the resultant approximate integral.

Monte carlo rules can be very useful with multidimensional integration. In many respects, high dimensional integration is a frontier for research for numerical analysts and statisticians. At this time, I am not aware of any high dimensional rules that have proven reliability with general functions. By studying the properties of a given integrand perhaps by making contour plots of sections or other theoretical tools, it can be determined where to devote most of the function evaluations. There are also adaptive rules which search for modes or ridges of functions and then set up some appropriate grid for function evaluation based on the search.

We now briefly describe how importance sampling works. We describe it with a one dimensional integral. Say we desire the integral of a statistical function f

$$\int f(x)dx \ .$$

Say there is some density g that we suspect is very similar to f in shape. If its integral is known and we can generate random samples from it, then we can write

$$\int f(x)dx = \int \frac{f(x)}{g(x)} g(x)dx = E_g\left(\frac{f}{g}\right)$$

where E_g denotes expectation with respect to the distribution with density g .

If we randomly sample from the distribution with density g say, $(x_1, \ldots, x_m)$ then

$$\frac{1}{m} \sum_{i=1}^{m} \frac{f}{g}(x_i)$$

will tend to $\int f(x)dx$ as $m \to \infty$ by the law of large numbers. For any finite m, $\frac{1}{m} \sum_{i=1}^{m} \frac{f}{g}(x_i)$ will approximate the desired integral. The variance of this estimate is

$$\int \left(\frac{f}{g}(x) - \int f(t)dt\right)^2 g(x)dx/m \ .$$

Clearly, the accuracy of this estimate depends very strongly on the choice of function g . The function g is often called the support density. We desire that $\frac{f}{g}$ be as nearly constant as possible over the region for which g has large density values.

In assessing candidates for g , it is reasonable to plot $\frac{f}{g}$ along with g on the same graph.

In the determination of the conditional distributions, the major barrier is the calculation of $h_\lambda(\underset{\sim}{d})$ which, for location-scale analysis, involves a double integral. One dimensional integration can be handled very efficiently with simple Gaussian quadrature rules, so in this section we investigate how this monte carlo method handles the integration of the second coordinate.

If $f_\lambda^{t,s,\underset{\sim}{d}}$ denotes the joint distribution for $(t , s , \underset{\sim}{d})$ then the conditional distribution for t given $\underset{\sim}{d}$ is

$$f_\lambda^t(t \mid \underset{\sim}{d}) = h_\lambda^{-1}(\underset{\sim}{d}) \int_s f_\lambda^{t,s,\underset{\sim}{d}}(t , s , \underset{\sim}{d}) ds .$$

Let us suppose that the one dimensional integral over s is computed, but only $h_\lambda(\underset{\sim}{d}) f_\lambda^t(t \mid \underset{\sim}{d})$ is available.

Suppose also that a density is available that approximates f_λ^t . We shall use g to denote such approximating support densities.

Suppose also that we can sample from this approximating distribution. If $(t_1 , \ldots , t_m)$ denotes a random sample from g then

$$\frac{1}{m} \sum_{i=1}^{m} \frac{h_\lambda(\underset{\sim}{d}) f_\lambda^t(t_i \mid \underset{\sim}{d})}{g(t_i)}$$

should be close to the unknown $h_\lambda(\underset{\sim}{d})$.

We are still faced however with an appropriate choice of g . We now make use of the results from the last section.

The distribution made up of a sum of scaled normal densities appears to have contours similar to the contours of independent Student(λ) variables. The appropriate t-statistic distribution based on $\Sigma c_i f_{\Sigma_i}$ could be similar in shape to f_λ^t if the quantities $c_1 , \ldots , c_n$ and $\tau_1 , \ldots , \tau_n$ are chosen wisely.

Several data sets have been considered both real and computer generated. We now include an example that was considered in D . The data is displayed in Table 3.2. The most influential observation appears to be $y_{28} = 16.5$. Based on the discussion in the last section it seems reasonable to use a single rescaled normal density for the analysis. We accordingly consider $f_{\Sigma_{28}}(\underset{\sim}{z})$ and contemplate the choice of various $\tau_{28} = \tau$ depending on the value of λ that is desired. For illustration we consider $\lambda = 2 , 3 , 6$ and ∞ . The choice of τ in each case could then be based on the

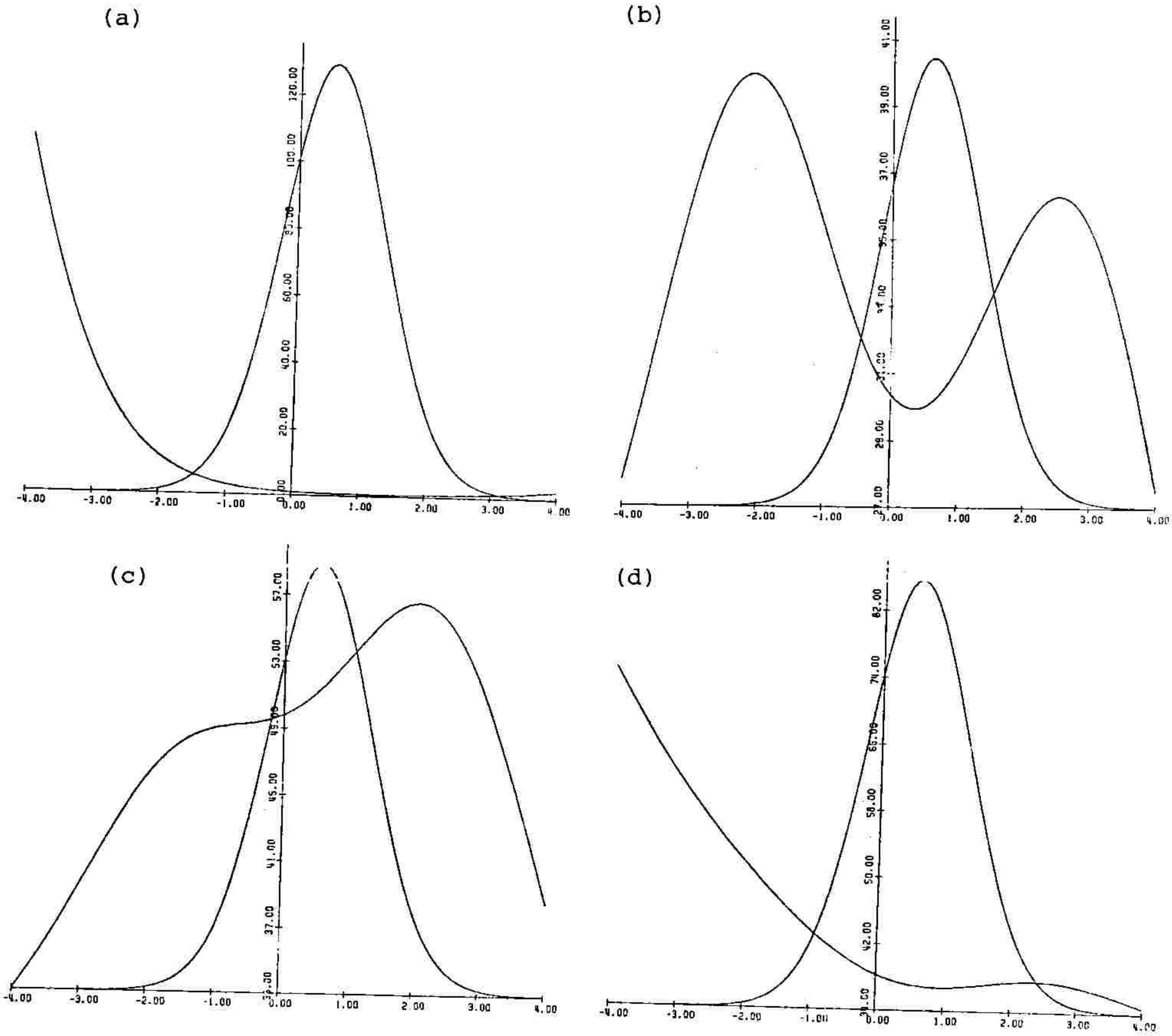


Figure 3.21 The ratio $h_\lambda(\underset{\sim}{d})f_\lambda^t/g_\tau^t$ along with g_τ^t

(a) $\lambda = \infty$, $\tau = 4$ (b) $\lambda = 2$, $\tau = 6$

(c) $\lambda = 3$, $\tau = 4$ (d) $\lambda = 6$, $\tau = 4$

(scaling with respect to the ratio)

straightness of the ratio $f_{\lambda}^{t} \mid g_{\tau}^{t}$ relative to the distribution g_{τ}^{t} .

Figure 3.21 gives an idea of the form that such curves can take. Based on a study of such curves, the following τ values were selected.

λ	2	3	6	∞
τ	6	4	4	1

For the monte carlo, we first considered $m = 100$. Here are the results.

Estimates of $h_{\lambda}(\underset{\sim}{d})$

λ	Monte Carlo	Quadrature
2	31.6	32.6
3	51.7	53.5
6	38.5	39.1
∞	1.0	1.0

The quadrature estimates are based on a very precise rule that required 241 function evaluations. They are exact in the sense that the decimal was rounded correctly. The monte carlo estimates are accurate enough for likelihood assessment of λ values and the normalization of the joint distribution for $\bar{z}$ and s given $\underset{\sim}{d}$.

The accuracy with monte carlo rules increases slowly as m is increased (as $1/\sqrt{m}$) but these results are very encouraging. It has been found that the choice of τ is not extremely critical for this form of accuracy (i.e., τ values between 3 and 8 all lead to fairly close results for the low λ values).

Real gains from the use of monte carlo integration are anticipated for the higher dimensional integration needed with regression analysis. This is discussed in Section 4D.

H. The Normality Assumption

(i) Cautions and Comments

In real experimental situations, there is often strong support for the assumption of normality of the errors. In such situations, it might be argued that the only legitimate significance levels and confidence intervals are those derived from the traditional normal analysis.

The simulation study on the marginal likelihood function in Section 3C demonstrated that in sampling from normal data there is often only weak support for the consultation of only a normal analysis based on the observed marginal likelihood functions from a Student analysis. Typically, such curves are rather flat, suggesting an often wide range of λ values that usually included the normal. In Section 3D, it was observed that, with normal data, the inferences derived from Student analyses are quite close to the correct normal analysis.

It appears to be reasonable to question our assumptions whenever possible. At a minimum, the examination of other analyses may lead to additional insight into a problem.

The marginal likelihood function appears to be a quite sensitive tool for the detection of nonnormality. Based on such a detection, the user could at least consider the results from a nonnormal analysis such as a Student analysis. In such situations, the results of such an analysis may suggest a reappraisal of the assumptions being made and perhaps provide an impetus for a repetition of the experiment.

(ii) Another Comment

Another way to display the fact that the assumption of normality may be appropriate is to display such information in the form of a prior distribution for the parameter λ . If $p(\lambda)$ denotes such a prior density, then a posterior density for λ given $\underset{\sim}{d}$ would be proportional to

$$p(\lambda)L(\underset{\sim}{d} \mid \lambda) \; .$$

Such priors were used by Box and Tiao (1973) in their analysis of the Darwin data with the exponential power distributions. They considered these distributions in the form

$$c(\lambda)\exp\left\{ -\frac{1}{2}|z|^{\frac{2}{1+\lambda}} \right\}$$

where $-1 < \lambda \leq 1$.

They standardized these distributions with respect to mean and standard deviation (see Section 2B). They did not include the distributions for $\lambda > 1$. Although they do allow for longer tails, these members are certainly somewhat unnatural since they develop a sharp peak at the origin. By consulting Figure 3.2, one can see that the marginal likelihood function displayed in their units would peak at the boundary of their parameter space $(\lambda = 1)$.

In my opinion, this should have led to concern that the chosen error family is not adequately describing the structure of the variation in this system.

Recall that the use of the Student family for variation led to a marginal likelihood function with mode near $\lambda = 2.5$. The Student family offers a broader range of descriptions for the variation than the exponential power family when longer tails are suspected.

With their model, the message supplied by the marginal likelihood function is a deficiency in the model. Box and Tiao offer a range of priors to indicate to the degree of support for the normality assumption. They considered a family of relocated and rescaled symmetric beta densities.

$$p(\lambda) = \frac{\Gamma(2a)}{\Gamma^2(a)2^{2a-1}}(1 - \lambda^2)^{a-1} \qquad -1 < \lambda \le 1 \ , \ a \ge 1 \ .$$

The corresponding marginal posterior densities continue to suggest λ values greater than zero (normality) but become increasingly dominated by the prior as a gets large. Except for $a = 1$, all the priors are zero at $\lambda = 1$. They implicitly remove the longer tailed distributions from consideration.

I. Supplement: An Empirical Study

Many of the ideas discussed in this chapter were of a qualitative nature illustrating aspects of the robustness and resistance of non-normal analyses. In sections F and G, attempts at a mathematical description of such properties were discussed.

It has been mentioned in several places that the primary tools for inferences concerning the unknown parameters are the displays and plots of the distributions and likelihood functions. Attempts to reduce such displays to a collection of statistics will result in a loss of information that may be illuminating at a terminal stage of inference. In section C, we considered a simulation study of the marginal likelihood function by studying the maximum marginal likelihood estimate and the log likelihood curvature. The purpose of the study was to supply insight into the highly complex model for possible likelihood functions.

In this section we consider a study of the Student (3) analysis compared with the classical normal analysis. An example from this study was considered in detail in section D. Two important characteristics derived from the conditional distributions are the median estimates and the confidence intervals. To assess these characteristics a large collection of data sets were generated. 100 sets of 25 numbers were generated from a

Student (3) distribution with $\mu = 0$ and $\sigma = 10$ and 100 sets of 25 numbers were generated from a normal distribution with $\mu = 0$ and $\sigma = 10$. All 200 sets were analyzed with a Student (3) analysis and a normal analysis.

The median estimates derived from the Student analysis were denoted $\hat{\mu}_3$ and $\hat{\sigma}_3$ and from the normal analysis, $\hat{\mu}_\infty$ and $\hat{\sigma}_\infty$. See section A for their formulae.

The widths of the 95% confidence intervals provided important clues to the understanding of the analyses, they are denoted $\hat{\ell}_3(\mu)$, $\hat{\ell}_\infty(\mu)$, $\hat{\ell}_3(\sigma)$ and $\hat{\ell}_\infty(\sigma)$ where for example

$$\hat{\ell}_3(\mu) = \hat{\mu}_R - \hat{\mu}_L$$

when $(\hat{\mu}_L, \hat{\mu}_R)$ is the 95% confidence interval for μ derived from a Student (3) analysis.

The observed mean squared error gives an idea of the accuracy of the estimate.

$$\sum(\hat{\mu} - 0)^2/100$$

$$\sum(\hat{\sigma} - 10)^2/100$$

The results are as follows

	Student (3) data		Normal data	
	μ	σ	μ	σ
Student (3) analysis	4.492	3.925	4.554	2.947
Normal analysis	6.993	31.946	3.350	2.354

The average confidence intervals widths were as follows

	Student (3) data		Normal data	
	μ	σ	μ	σ
Student (3) analysis	8.605	8.670	8.139	7.744
Normal analysis	10.812	7.981	8.239	6.082

The observed confidence levels were

	Student (3) data		Normal data	
	μ	σ	μ	σ
Student (3) analysis	$\frac{96}{100}$	$\frac{98}{100}$	$\frac{92}{100}$	$\frac{98}{100}$
Normal analysis	$\frac{97}{100}$	$\frac{60}{100}$	$\frac{96}{100}$	$\frac{96}{100}$

Several points deserve comment. Notice that with the Student data sets, the Student analyses offer tighter intervals for μ than the normal analyses and retain the appropriate confidence level. Notice also that a normal analysis leads to very unreliable intervals for σ.

With the normal data sets, there is no appreciable difference between the two analyses. Table 3.3 displays the results from the Student (3) data sets and table 3.4 displays the results from the normal data sets.

There is one other characteristic of this study that deserves comment here. From the 100 student data sets, 15 of them contained one or more observations that were more than five standard deviations (5σ) from μ. They were sample numbers 1, 5, 15, 16, 19, 35, 42, 49, 75, 80, 85, 87, 91, 99 and 100. In all but 1 case (# 35), $\hat{\mu}_3$ is far closer to zero than $\hat{\mu}_\infty$. All Student (3) intervals for μ are tighter than the normal intervals with these sets and yet they all cover μ. 13 of the 15 normal intervals for σ fail to cover σ while all 15 Student intervals cover σ. These comments offer further evidence of the resistance of Student analyses to observations that can dramatically affect inferences derived from classical normal analysis.

Sample Number	$\hat{\mu}_3$	$\hat{\mu}_\infty$	$\hat{\ell}_3(\mu)$	$\hat{\ell}_\infty(\mu)$	$\hat{\sigma}_3$	$\hat{\sigma}_\infty$	$\hat{\ell}_3(\sigma)$	$\hat{\ell}_\infty(\sigma)$
	-0.656	-3.354	8.689	11.412	10.189	13.928	8.414	8.423
	1.378	0.707	8.176	9.984	9.948	12.185	8.368	7.370
	-0.300	-3.189	7.234	10.224	8.621	12.478	7.697	7.546
	-0.401	-0.802	7.602	7.092	8.799	8.656	6.901	5.235
	0.344	-2.814	5.886	13.241	11.774	16.160	10.051	9.773
	1.811	2.659	8.112	8.969	9.656	10.947	8.116	6.620
	1.425	0.519	8.257	9.581	9.957	11.694	8.613	7.072
	2.098	2.129	8.693	9.425	10.151	10.283	8.096	6.219
	0.575	-0.261	9.332	9.273	10.935	11.318	8.570	6.945
10	-2.015	-0.242	10.403	11.543	12.422	14.088	10.653	8.520
	0.120	-0.905	6.615	7.027	7.916	8.576	6.395	5.187
	-3.327	-2.820	9.340	13.576	11.412	16.570	10.660	10.021
	1.020	1.163	8.429	9.303	10.037	11.355	8.368	6.867
	-1.210	-1.593	7.100	6.591	8.220	8.044	6.423	4.865
	-0.711	-2.271	9.997	14.751	12.112	13.003	10.036	10.388
	2.509	-4.000	12.391	23.127	14.482	30.667	12.407	13.547
	-2.305	-3.866	8.475	10.525	10.222	12.845	9.025	7.769
	-1.774	-2.213	10.847	11.985	12.898	14.627	10.734	8.946
	1.039	-1.980	10.321	17.615	12.448	21.499	10.368	13.002
20	-2.237	-0.491	8.639	10.037	9.995	12.250	7.851	7.409
	3.635	4.207	10.989	10.760	12.835	13.133	10.270	7.942
	-5.327	-5.164	10.983	12.525	13.023	15.286	10.909	9.245
	1.499	1.042	8.428	7.609	9.684	9.287	7.405	5.616
	1.295	-0.181	5.464	6.834	6.473	8.341	5.514	5.045
	-2.166	-2.747	9.103	9.744	10.756	11.892	8.793	7.192
	0.706	-0.916	11.028	12.212	13.097	14.904	10.909	9.014
	1.103	1.820	10.895	12.449	12.956	15.194	10.921	9.199
	-0.711	1.347	8.299	8.562	9.639	10.450	7.922	6.320
	1.817	2.140	7.321	9.222	8.747	10.034	7.398	6.069
30	0.592	0.441	8.299	9.548	9.807	11.653	8.160	7.048
	0.405	0.085	8.035	8.187	9.521	9.922	7.924	6.043
	1.280	3.082	11.517	12.683	13.630	15.479	11.521	9.361
	1.693	1.646	6.934	7.498	8.293	9.151	6.855	5.535
	0.705	0.394	6.320	8.579	7.652	10.470	6.623	6.332
	-3.584	0.991	10.207	19.756	12.109	24.112	10.515	14.582
	4.161	3.724	12.348	13.575	14.719	16.563	12.327	10.020
	-1.449	-3.593	11.675	14.451	13.881	17.637	11.694	10.667
	-1.414	-0.656	8.429	11.732	10.171	14.319	8.859	8.660
	1.865	3.193	9.828	12.890	11.891	15.732	10.589	9.515
40	0.539	-0.155	6.575	8.338	7.832	10.176	6.537	6.154
	2.485	1.826	9.644	10.806	11.488	13.189	9.677	7.976
	-0.423	1.342	8.671	14.939	10.308	18.233	8.870	11.027
	0.906	0.742	8.678	8.236	10.085	10.052	7.970	6.079
	-4.841	-5.900	5.102	6.935	6.124	8.464	5.223	5.119
	1.509	-0.227	7.196	9.143	8.623	11.159	7.475	6.749
	-0.976	-1.770	6.755	7.713	8.129	9.414	7.101	5.694
	0.227	0.493	6.193	7.361	7.401	8.984	6.298	5.434
	-1.122	-2.062	8.206	9.630	9.761	11.314	8.024	7.145
	-0.493	0.653	7.729	14.427	9.383	17.608	8.383	10.649
50	-0.485	-1.981	5.807	6.836	6.945	8.343	5.618	5.046
	1.304	2.279	8.149	9.627	9.764	11.749	8.279	7.106
	4.405	4.606	8.642	9.309	10.204	11.361	8.316	6.971
	0.967	2.467	11.344	13.307	13.580	16.241	12.127	9.822
	-2.167	-2.129	6.775	8.294	8.217	10.123	7.142	6.122
	-3.500	-2.869	9.910	11.030	11.943	13.523	10.055	8.179
	3.106	3.193	8.399	8.909	9.988	10.873	8.178	6.576
	0.866	-0.344	6.422	8.413	7.741	10.268	6.925	6.210
	-2.253	-1.265	5.681	6.293	6.756	7.669	5.772	4.638
	2.523	3.354	7.606	8.506	9.032	10.381	7.544	6.278
60	2.039	1.524	7.940	7.371	9.167	8.996	7.165	5.441
	-1.969	-2.412	8.896	10.770	10.717	13.145	9.286	7.950
	-4.549	-5.407	7.398	8.939	8.684	10.910	7.028	6.598
	1.221	1.142	8.409	11.135	10.125	13.590	9.023	8.219
	1.103	-0.022	7.195	9.503	8.744	11.599	8.123	7.014
	0.867	1.931	6.804	7.792	8.135	9.510	7.227	5.752
	0.787	1.771	6.453	8.199	7.798	10.005	6.387	6.051
	-0.341	-0.039	10.558	11.950	12.618	14.584	10.617	8.820
	-2.114	-1.919	10.468	11.155	12.379	13.615	10.632	8.234
	0.663	-0.775	10.926	12.625	13.121	15.409	11.230	9.319
70	-0.264	-1.407	7.720	9.370	9.308	11.460	7.983	6.931
	0.927	2.392	9.908	12.252	11.916	14.953	10.152	9.043
	3.069	4.645	9.795	11.467	11.624	13.995	9.534	8.464
	-2.160	-3.943	11.708	12.952	13.933	15.807	11.865	9.560
	-0.975	-0.603	5.743	5.618	6.717	6.857	5.401	4.147
	0.634	1.886	8.176	13.224	9.913	16.140	9.145	9.761
	-1.972	-2.448	8.935	8.561	10.431	10.448	8.233	6.319
	-2.059	-3.821	6.353	7.695	7.565	9.391	6.852	5.680
	-3.333	-3.122	7.191	6.845	8.244	8.354	6.420	5.053
	-4.535	-4.170	9.515	10.307	11.243	12.579	9.194	7.608
80	-3.964	-9.045	9.343	19.487	11.182	23.783	10.326	14.394
	-0.970	-1.252	7.891	9.460	9.529	11.546	8.377	6.933
	-0.170	0.076	6.588	8.257	7.905	10.078	6.901	6.095
	-0.128	-0.084	7.746	7.350	8.994	8.970	7.126	5.425
	1.350	-0.683	8.911	12.977	10.312	15.862	9.669	9.593
	-3.313	1.884	9.963	32.895	12.263	40.148	8.477	24.281
	0.675	-0.246	10.533	12.522	12.689	15.283	10.953	9.243
	-0.913	1.404	8.236	14.619	10.030	17.842	8.957	10.791
	0.906	-1.139	9.469	10.745	11.180	13.114	9.408	7.931
	-0.973	-2.881	9.311	10.459	10.946	12.765	9.067	7.720
90	-1.499	-1.301	6.756	7.137	7.990	8.710	6.491	5.268
	0.609	4.303	7.125	14.984	8.507	18.288	7.388	11.060
	0.137	0.419	6.795	8.531	8.146	10.412	7.022	6.297
	-1.497	-1.157	8.919	10.372	10.615	12.659	9.185	7.656
	3.120	5.924	9.301	10.780	10.922	13.157	9.443	7.957
	-2.976	-2.100	8.812	11.956	10.542	14.592	8.982	8.825
	-1.958	0.486	7.098	10.162	8.681	12.402	7.972	7.501
	5.712	5.896	10.634	10.312	12.462	12.585	10.038	7.611
	0.728	0.208	8.888	9.577	10.574	11.689	8.663	7.069
	3.926	4.109	10.944	15.169	13.238	19.514	11.246	11.197
	0.521	-2.241	9.183	13.478	11.048	16.449	9.743	9.948

Table 3.3 Results from the analysis with student (3) data.

Sample Number	$\hat{\mu}_3$	$\hat{\mu}_\infty$	$\hat{\ell}_3(\mu)$	$\hat{\ell}_\infty(\mu)$	$\hat{\sigma}_3$	$\hat{\sigma}_\infty$	$\hat{\ell}_3(\sigma)$	$\hat{\ell}_\infty(\sigma)$
	-1.700	-0.074	9.866	9.990	11.543	12.192	9.651	7.374
	-1.239	0.778	8.749	9.394	10.221	11.465	8.601	6.934
	4.588	3.613	7.978	7.563	9.261	9.231	7.473	5.592
	-3.206	-2.721	5.171	5.913	6.148	7.217	5.149	4.365
	0.244	0.356	6.344	6.294	7.391	7.682	5.880	4.646
	-0.737	0.703	7.048	7.569	8.313	9.239	6.847	5.547
	-2.031	-1.377	6.145	6.372	7.278	7.777	6.065	4.704
	-3.736	-1.600	7.285	7.677	8.411	9.370	7.032	5.667
	-0.817	0.063	8.870	8.873	10.428	10.830	8.529	6.550
10	-2.356	-0.267	10.654	10.434	12.329	12.734	10.203	7.701
	0.380	0.537	7.474	8.684	8.994	10.599	7.807	6.410
	-0.219	-0.423	7.678	7.277	8.901	8.981	7.041	5.371
	0.188	0.149	5.141	6.312	6.175	7.704	5.243	4.659
	0.512	0.018	7.848	7.940	9.150	9.690	7.277	5.961
	-3.545	-2.785	7.862	7.781	9.146	9.496	7.293	5.743
	1.944	0.268	6.673	7.183	7.802	8.766	6.512	5.302
	5.096	4.862	9.972	8.763	11.381	10.695	8.539	6.468
	-0.621	0.013	7.926	8.150	9.327	9.947	7.554	6.016
	0.959	0.444	9.629	9.762	11.449	11.914	9.476	7.205
20	-4.470	-4.470	7.484	7.059	8.702	8.615	6.872	5.210
	-0.202	0.479	11.260	10.951	13.187	13.366	10.613	8.083
	0.350	0.856	7.711	8.102	9.148	9.889	7.524	5.980
	1.674	1.505	10.013	10.789	11.911	13.167	9.777	7.963
	-0.456	-0.324	8.939	10.317	10.694	12.592	9.083	7.615
	4.713	2.988	9.570	9.077	10.969	11.078	8.870	6.700
	0.316	-0.164	8.510	8.105	9.897	9.893	7.960	5.933
	0.124	0.033	7.441	8.273	8.548	10.097	7.655	6.106
	-2.092	-0.859	9.373	9.516	10.959	11.614	8.869	7.024
	0.504	-0.163	5.926	6.545	6.937	7.983	5.935	4.931
30	3.482	2.279	5.919	7.328	7.039	8.944	5.882	5.409
	0.161	0.334	8.139	7.794	9.441	9.512	7.522	5.753
	-2.439	-1.638	9.928	9.105	11.456	11.113	8.843	6.721
	-1.589	-1.908	6.211	6.546	7.367	7.989	6.086	4.332
	0.941	1.687	7.235	7.054	8.336	8.609	6.495	5.207
	-0.195	-0.173	9.253	9.589	10.950	11.704	9.013	7.078
	-1.470	-1.128	5.890	6.876	7.053	8.392	6.024	5.075
	3.764	3.182	9.185	8.905	10.714	10.747	8.608	6.500
	1.427	2.291	8.519	9.141	10.028	11.157	8.156	6.747
	4.132	3.313	8.815	9.133	10.172	9.926	7.924	6.003
40	1.527	0.595	7.344	7.646	8.651	9.331	7.090	5.643
	-0.626	2.099	7.608	8.740	8.994	10.666	8.006	6.451
	-1.464	-1.117	9.512	9.265	11.087	11.307	8.879	6.839
	-0.560	0.330	10.844	11.054	12.795	13.503	10.545	8.166
	-2.287	-2.332	8.766	8.314	10.232	10.147	8.128	6.137
	0.909	1.277	7.853	8.376	9.261	10.223	7.510	6.183
	-0.565	0.041	6.010	5.639	6.949	6.882	5.513	4.162
	0.965	0.233	10.706	10.426	12.545	12.724	10.113	7.695
	0.553	1.460	8.600	9.017	10.163	11.005	8.355	6.556
	1.743	1.072	9.500	9.114	11.102	11.124	8.900	6.727
50	-2.721	-1.873	8.398	8.014	9.723	9.781	7.723	5.915
	-3.042	-2.162	10.861	10.602	12.694	12.939	10.225	7.825
	-3.043	-2.800	7.910	7.215	9.130	8.306	[illegible]	5.226
	-0.945	-1.478	7.231	8.074	8.630	9.854	7.294	5.960
	1.272	1.387	8.676	8.591	10.166	10.486	8.198	6.341
	-3.406	-3.003	6.452	6.086	7.473	7.428	5.887	4.492
	5.638	4.911	7.647	7.446	8.877	9.088	7.169	5.496
	-4.064	-2.869	7.515	9.132	9.064	11.145	7.860	6.741
	0.387	0.159	8.753	8.656	10.299	10.564	8.412	6.389
	0.412	0.548	7.620	7.572	8.994	9.242	7.380	5.589
60	3.557	2.202	11.436	9.939	12.948	12.130	9.342	7.336
	0.388	0.137	9.925	9.601	11.567	11.718	9.253	7.087
	0.337	0.621	6.741	6.741	7.951	9.228	6.518	4.976
	-1.557	0.242	6.329	9.315	7.631	10.148	6.707	5.138
	0.500	1.221	6.405	7.727	9.736	9.431	7.507	5.704
	-1.699	-1.414	8.476	7.970	9.821	9.605	7.688	5.809
	0.593	-1.405	7.969	8.596	9.315	10.492	8.115	6.345
	0.354	0.547	8.039	9.072	9.477	9.952	7.793	5.959
	-1.723	-0.999	6.623	5.349	7.792	8.358	6.273	5.055
	-1.753	-1.828	8.941	8.607	10.501	10.505	8.450	5.353
70	7.321	5.976	5.358	6.882	6.445	8.400	5.659	5.080
	1.970	2.863	6.337	6.680	7.482	8.153	6.065	4.931
	-0.926	-0.216	7.967	8.216	9.316	10.027	7.657	5.064
	2.551	2.260	8.228	8.210	9.643	10.021	7.763	6.060
	1.508	1.841	6.705	6.901	7.939	8.422	6.528	5.094
	-0.760	-0.053	9.269	9.154	10.811	11.172	8.674	6.757
	-1.452	-0.142	7.699	9.244	9.018	10.062	7.339	6.085
	0.451	0.234	8.105	7.850	9.481	9.581	7.642	5.794
	-2.609	-3.090	9.090	8.566	10.549	10.455	8.314	6.323
	1.576	1.630	9.425	8.642	10.508	10.548	8.346	6.379
80	-0.174	0.007	6.687	6.823	7.771	8.328	6.070	5.036
	1.135	-0.353	9.455	9.218	10.878	11.250	8.652	6.804
	0.142	-0.481	9.331	10.193	11.176	12.440	9.327	7.524
	0.324	0.530	9.865	9.451	11.479	11.535	9.156	6.976
	-0.050	0.865	9.681	8.850	11.149	10.802	8.572	6.533
	-2.838	-2.918	9.184	8.725	10.674	10.648	8.462	6.440
	0.088	-0.029	6.521	7.124	7.777	8.694	6.453	5.258
	0.185	0.935	8.019	8.031	9.470	9.862	7.765	5.965
	2.430	2.200	6.317	6.455	7.512	7.878	6.227	4.765
	-0.693	-0.145	9.690	8.905	11.118	10.740	8.516	6.499
90	1.963	1.721	9.355	8.879	10.974	10.836	8.650	6.554
	-0.481	-0.947	9.040	10.049	10.749	12.264	8.962	7.417
	0.408	1.773	10.023	10.268	11.797	12.532	9.708	7.579
	-0.492	-1.886	7.436	7.587	8.706	9.260	7.200	5.600
	0.911	0.995	6.542	6.819	7.748	8.322	6.331	5.033
	-0.561	-1.119	6.286	6.620	7.392	8.079	5.925	4.886
	2.122	-0.022	8.478	9.653	10.069	11.782	8.656	7.125
	1.965	1.916	9.061	7.703	10.221	9.401	7.644	5.686
	-1.980	-2.918	6.903	6.520	7.962	7.957	6.328	4.812
	0.166	-0.003	8.472	8.423	9.974	10.280	9.150	6.217
	1.250	1.387	5.053	6.014	6.145	7.340	5.562	4.439

Table 3.4 Results from the analysis of normal data.

CHAPTER 4

ANALYSIS OF THE REGRESSION MODEL

A. An Example

The family of Student distributions provided a very flexible range of tail length for the analysis of the location-scale model. We anticipated that many of the favourable properties discussed in Chapter 3 would carry over to more complex regression models.

With location-scale data, a deviant observation or observations can often be transparent but with even simple regression problems, detection of such observations is a difficult problem.

It is anticipated that a Student analysis of regression will tolerate such observations in a way similar to location-scale.

We now consider

$$\underset{\sim}{y}$$

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

$$\Pi f_\lambda(z_i)\Pi dz_i$$

where λ indexes the standardized Student family.

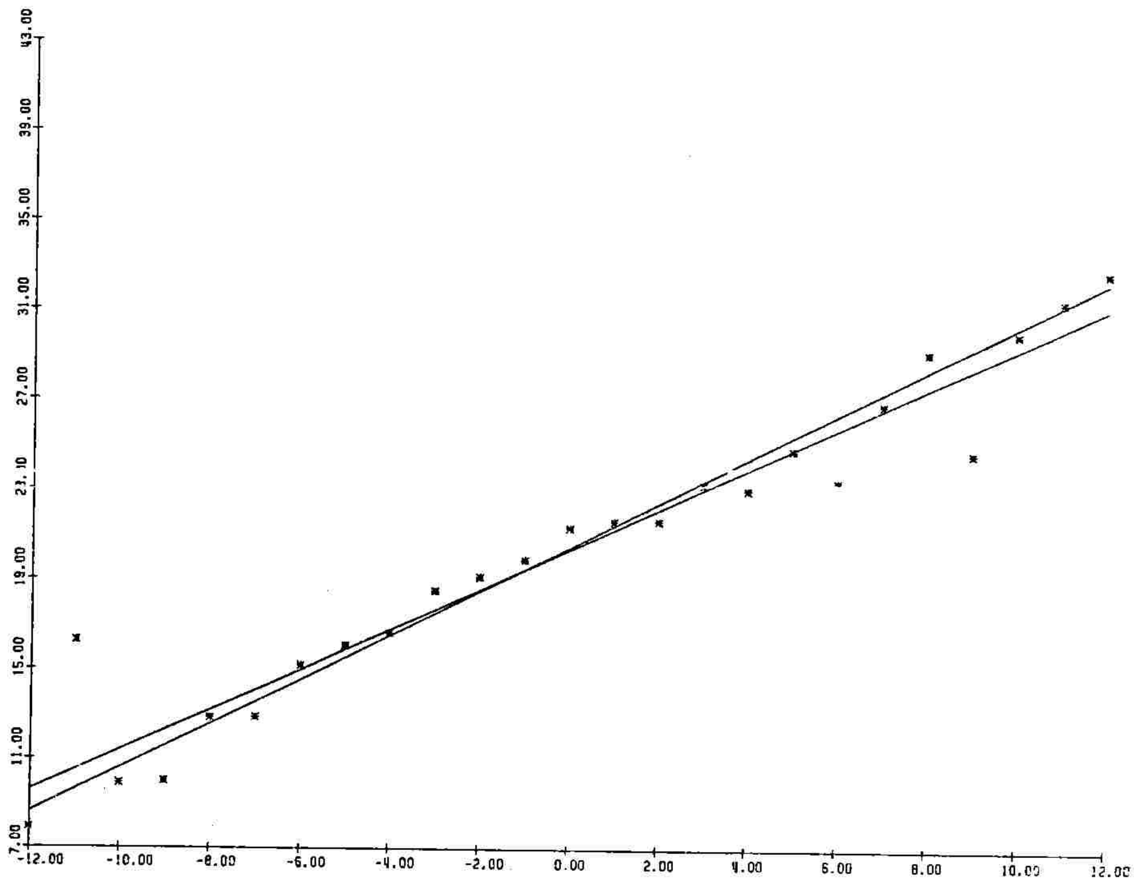


Figure 4.1 A plot of the data, the least squares line and the Student(2) line

Consider the data set displayed in Figure 4.1. The context suggests the model

$$\underset{\sim}{y} = \beta_1 \underset{\sim}{1} + \beta_2 \underset{\sim}{x} + \sigma \underset{\sim}{z}$$

$$= \alpha_1 \underset{\sim}{1}/\sqrt{25} + \alpha_2 \underset{\sim}{x}/\sqrt{\Sigma x^2} + \sigma \underset{\sim}{z} \ .$$

The latter represents the more canonical form specialized from

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

$$V'V = I \qquad \underset{\sim}{\beta} = T^{-1}\underset{\sim}{\alpha}$$

so that

$$\underset{\sim}{b}(\underset{\sim}{y}) = T^{-1}\underset{\sim}{a}(\underset{\sim}{y}) \quad \text{where} \quad \underset{\sim}{a}(\underset{\sim}{y}) = V'\underset{\sim}{y} \ .$$

Clearly $s^2(\underset{\sim}{y})$ and $\underset{\sim}{d}(\underset{\sim}{y})$ do not depend on the basis for $L(X)$.

The assessment of $\underset{\sim}{\alpha}$ and then $\underset{\sim}{\beta}$ is based on the observed value of

$$\underset{\sim}{t}_{\underset{\sim}{z}} = \underset{\sim}{a}(\underset{\sim}{z})/s_{\underset{\sim}{z}} = \sqrt{n-2}\ \underset{\sim}{a}(\underset{\sim}{z})/s(\underset{\sim}{z})$$

together with the conditional distribution given $\underset{\sim}{d}$ as computed for a selection of λ values. For all the distributions except the normal, computer calculations are used.

Discussion of these calculations is made in Section 4D.

Some preliminary calculations yield

$$\underset{\sim}{a}(\underset{\sim}{y}) = (101.746\ ,\ 32.278)'$$

$$\underset{\sim}{b}(\underset{\sim}{y}) = (20.3492\ ,\ 0.8952)'$$

$$\underset{\sim}{s}(\underset{\sim}{y}) = 8.6625$$

$$s_{\underset{\sim}{y}} = 1.80626$$

$$\underset{\sim}{d} = \begin{array}{lllll} (-0.1965 & 0.6627 & -0.1713 & -0.2625 & -0.0406 \\ -0.1411 & 0.0221 & 0.0215 & -0.0186 & 0.0974 \\ 0.0649 & 0.0495 & 0.1111 & 0.0433 & -0.0583 \\ 0.0228 & -0.1061 & -0.0048 & -0.2710 & 0.0207 \\ 0.1867 & -0.4329 & 0.0769 & 0.1397 & 0.1844) \end{array}$$

The largest positive deviation is 0.6627 (corresponding to $y_2 = 16.2425$ and $x_{2,2} = -11.0$) and the largest negative deviation is -0.4329 (corresponding to $y_{22} = 24.6564$ and $x_{2,22} = 9.0$) . These observations seem to be the most influential in determining distributions for $\underset{\sim}{a}(\underset{\sim}{z})$.

We now examine the data using a Student analysis. As with the location scale analysis of Chapter 3, we begin by consulting the likelihood function for λ . In particular we examine

$$L(\underset{\sim}{d} \mid \lambda) = A_{n-2} h_\lambda(\underset{\sim}{d}) = A_{23} h_\lambda(\underset{\sim}{d}) \ .$$

Selected values of the likelihood function are

λ	1	2	3	4	5	6	9	∞
$L(\underset{\sim}{d} \mid \lambda)$	138	302	166	82	45	27	10	1

This rather discriminating likelihood suggests λ values between 1 and 6 . For comparison, suppose we have included the traditional normal analysis corresponding to $\lambda = \infty$.

Contour plots for the distribution of $\underset{\sim}{t} = \underset{\sim}{a}(z)/s_{\underset{\sim}{z}}$ are in Figure 4.2 for $\lambda = 1 , 3 , 6$ and ∞ .

Confidence regions for $\underset{\sim}{\alpha}$ or $\underset{\sim}{\beta}$ are based on these distributions. For $\lambda = \infty$, the distribution of $\underset{\sim}{t}$ is the bivariate Student(23) distribution with density

$$\frac{\Gamma(\frac{25}{2})}{23\pi \ \Gamma(\frac{23}{2})} \left(1 + \frac{t_1^2 + t_2^2}{23} \right)^{-\frac{25}{2}}$$

or using the notation developed for (2.15)

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \sim \text{Student}_{23}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} , \begin{pmatrix} 23 & 0 \\ 0 & 23 \end{pmatrix} \right) \quad \text{on } \mathbb{R}^2 \ .$$

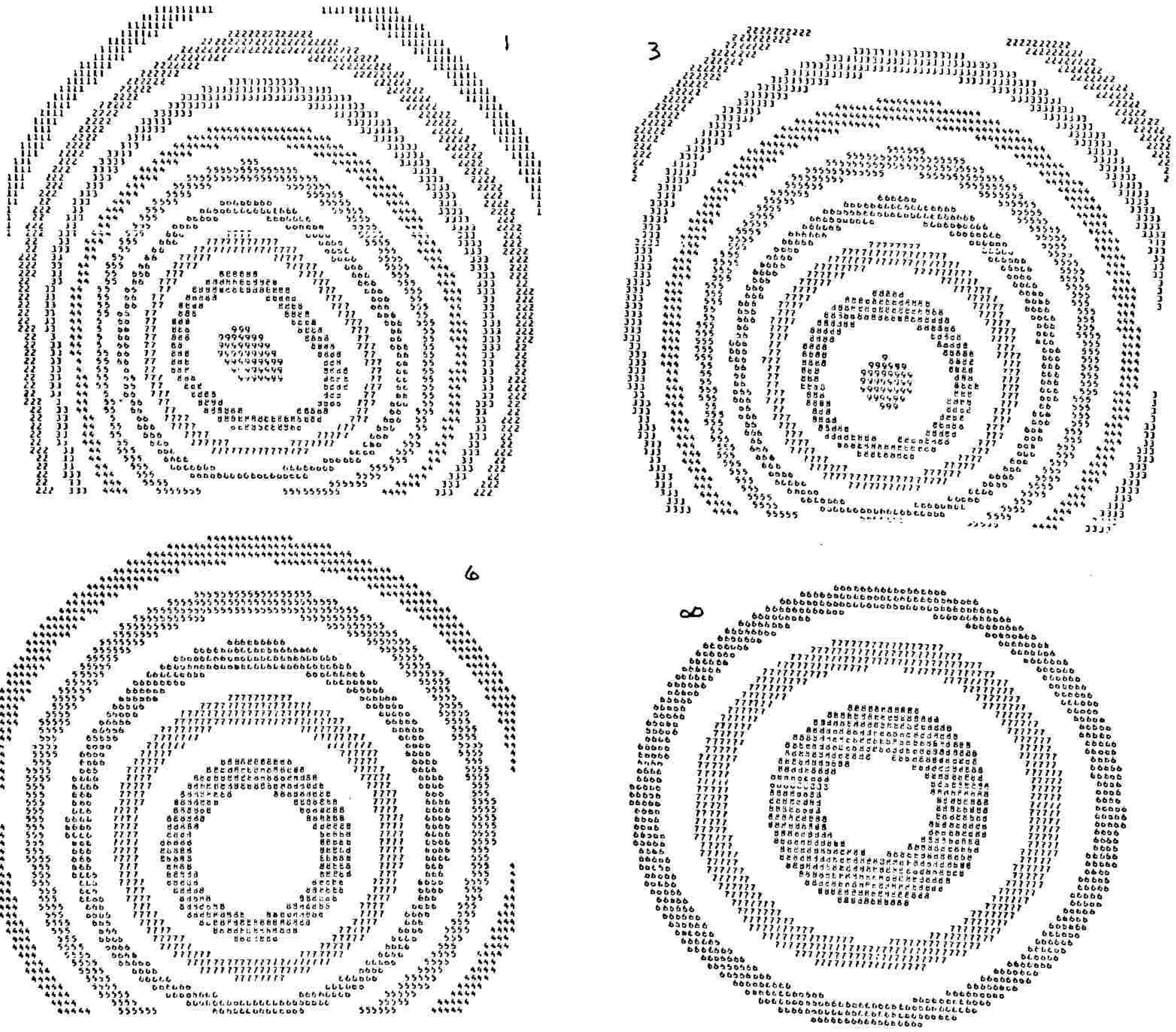


Figure 4.2 Contour plots for $\underset{\sim}{t} = (t_1 , t_2)'$ for $\lambda = 1 , 3 , 6 , \infty$

As λ tends to 1, the distribution shifts and becomes more concentrated. Note under normal analysis, the distribution is independent of $\underset{\sim}{d}$ and does not depend on V.

However, under non-normal analysis, the distribution of $\underset{\sim}{t}$ dramatically changes in shape, location and concentration, depending on the value of $\underset{\sim}{d}$ and V.

To assess β_1 and β_2 individually, we examine the component t statistics

$$t_1 = a_1(\underset{\sim}{z})/s_{\underset{\sim}{z}} \qquad t_2 = a_2(\underset{\sim}{z})/s_{\underset{\sim}{z}}$$

together with their distributions; the distributions are plotted in Figures 4.3 and 4.4 for $\lambda = 1, 3, 6$ and ∞. Under normal analysis, the distributions for both t_1 and t_2 are the ordinary Student(23).

Note that the t_1 densities are more concentrated for $\lambda = 1, 3$ and 6 and do not shift substantially; also note that the t_2 densities are more concentrated and do shift substantially to the left.

Confidence intervals for β_1 and β_2 have the form

$$\beta_1 : \left(b_1(\underset{\sim}{y}) - t_{1U}\, s_{\underset{\sim}{y}}/5 \,,\, b_1(\underset{\sim}{y}) - t_{1L}\, s_{\underset{\sim}{y}}/5\right)$$

$$\beta_2 : \left(b_2(\underset{\sim}{y}) - t_{2U}\, s_{\underset{\sim}{y}}/\sqrt{1300} \,,\, b_2(\underset{\sim}{y}) - t_{2L}\, s_{\underset{\sim}{y}}/\sqrt{1300}\right)$$

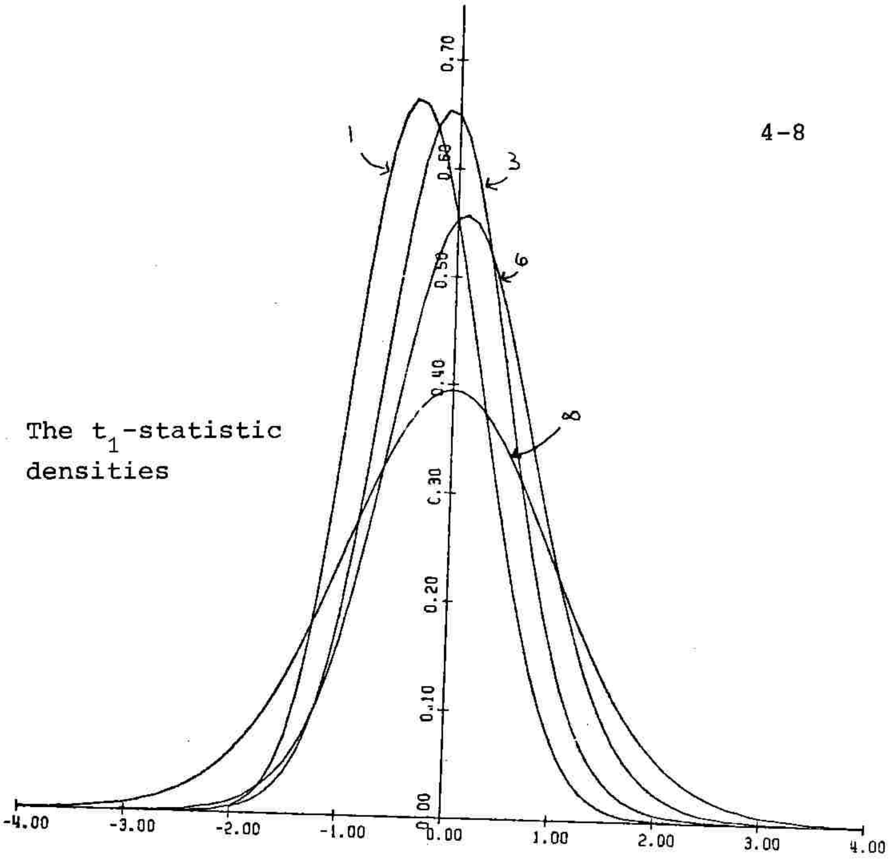
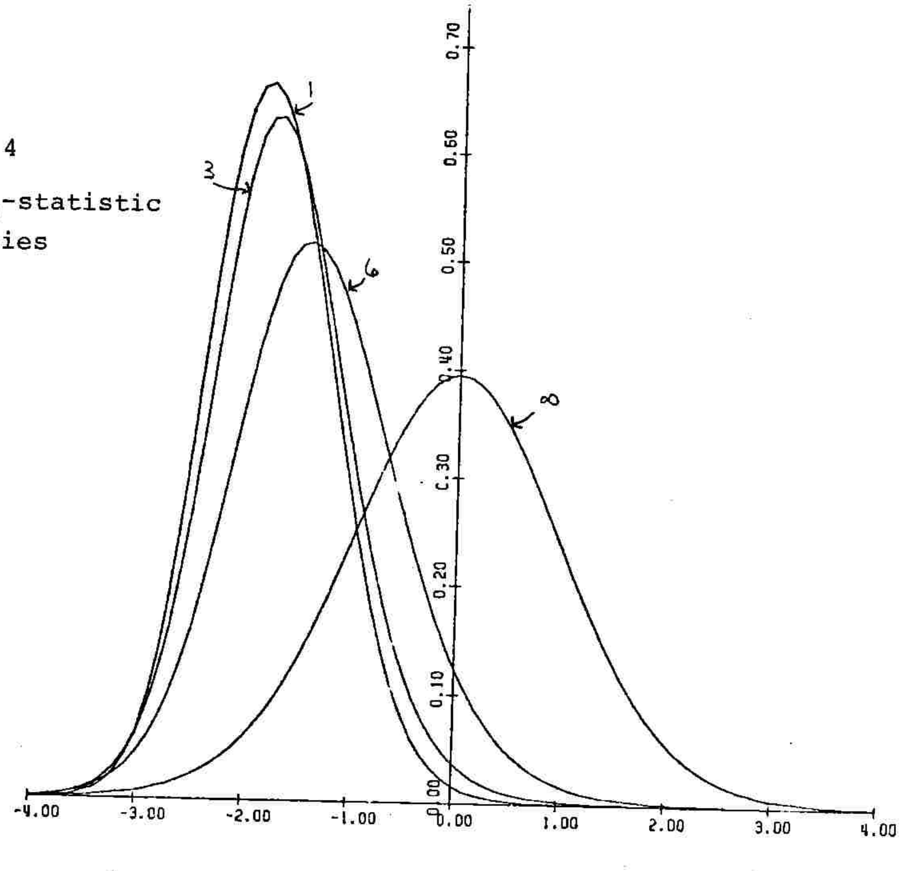


Figure 4.3 The t_1-statistic densities

Figure 4.4

The t_2-statistic densities



where (t_{1L}, t_{1U}) is a central interval for t_1 and (t_{2L}, t_{2U}) is a central interval for t_2. For example, 95% confidence intervals for β_2 are

λ	Confidence Interval
1	(0.92 , 1.04)
3	(0.91 , 1.04)
6	(0.87 , 1.03)
∞	(0.79 , 0.99)

Note that we have included the normal theory least squares line on the data plot. The line based on the Student(2) analysis has the following values:

$$\text{slope} = b_2(\underset{\sim}{y}) - \text{median}(t_2)s_{\underset{\sim}{y}}/\sqrt{1300} = 0.9853$$
$$\text{intercept} = b_1(\underset{\sim}{y}) - \text{median}(t_1)s_{\underset{\sim}{y}}/5 = 20.45 .$$

Notice that the Student(2) line is steeper than the least squares line and is resisting the effect of y_2 and y_{22}.

The Student(λ) analysis provides a more robust and resistant fitting procedure than the usual least squares procedure.

We find that with normal data the analyses are usually similar for various λ, and with non-normal data, the

Student analyses are usually quite different. The Student analyses with smaller λ values seem to have a very broad based reliability in producing the approximately correct analysis whatever the true value of λ ; recall the discussion in Section 2D.

The data set was generated using $\beta_1 = 20$, $\beta_2 = 1$, $\sigma = 1.1966$ and $\lambda = 3$.

The likelihood functions and distributions were obtained by three dimensional integration procedures on the computer. Comments on this and other procedures are made in Section D.

B. Resistance

In the last section, we saw how a Student analysis led to inferences which were resistant to the two somewhat extreme observations. In particular, the Student(2) line was steeper than the least squares line and was resisting the effect that y_2 and y_{22} had on the normal analysis line. The resulting confidence intervals were tighter with the Student analyses than with the normal analyses; correcting for an overinflated residual sum of squares was also affected by those two observations.

We now determine the adjusted conditional distributions similar to those presented in Section 3E. It is these distributions that provide a direct assessment of the resistance of Student analysis to deviant observations. Let $\underset{\sim}{y}^R$ denote the reference data set. Its summary statistics would be

$$\underset{\sim}{a}^R(\underset{\sim}{y}^R) = V'\underset{\sim}{y}^R \qquad s^R(\underset{\sim}{y}^R) \quad \text{and} \quad \underset{\sim}{d}^R .$$

We then contemplate moving one or more observations in $\underset{\sim}{y}^R$ in some systematic way to obtain $\underset{\sim}{y}$; with statistics

$$\underset{\sim}{a}(\underset{\sim}{y}) = V'\underset{\sim}{y} \qquad s(\underset{\sim}{y}) \quad \text{and} \quad \underset{\sim}{d} .$$

The interpretation is similar to location-scale. We imagine being faced with the analysis of $\underset{\sim}{y}$ and wonder whether the inferences derived are close to those that would be obtained if $\underset{\sim}{y}^R$ were available.

Inference concerning $\underset{\sim}{\alpha}$ is based on

$$\underset{\sim}{t} = \underset{\sim}{a}(\underset{\sim}{z})/s_{\underset{\sim}{z}} \ .$$

Confidence regions for $\underset{\sim}{\alpha}$ would be derived from

$$\underset{\sim}{a}(\underset{\sim}{y}) - \underset{\sim}{t}\, s_{\underset{\sim}{y}}$$

based on central probability regions of the conditional distribution of $\underset{\sim}{t}$.

To make direct comparisons, we can reexpress these regions in terms of the original reference data set

$$\underset{\sim}{a}^R(\underset{\sim}{y}^R) - \underset{\sim}{T}\, s^R_{\underset{\sim}{y}} \ .$$

We then inquire as to the stability of the distribution for $\underset{\sim}{T}$ where

$$\underset{\sim}{T} = \frac{s_{\underset{\sim}{y}}}{s^R_{\underset{\sim}{y}}}\, \underset{\sim}{t} + \frac{1}{s^R_{\underset{\sim}{y}}}\left(\underset{\sim}{a}^R(\underset{\sim}{y}^R) - \underset{\sim}{a}(\underset{\sim}{y})\right) \ .$$

The variable $\underset{\sim}{T}$ contains the appropriate corrections to relate its conditional distribution to the reference data set.

For inference concerning σ , we can consult the adjusted conditional distribution for

$$S = \frac{s^R_{\underset{\sim}{y}}}{s_{\underset{\sim}{y}}} s .$$

Once again, if these distributions remain close, then our inference concerning σ is resistant to outlying observations.

We can note that with a normal analysis the distribution for T would be

$$\text{Student}_{n-r}(\underset{\sim}{\nu}, W) \quad \text{on} \quad \mathbb{R}^r$$

where

$$\underset{\sim}{\nu} = [\underset{\sim}{a}^R(\underset{\sim}{y}^R) - \underset{\sim}{a}(\underset{\sim}{y})] / s^R_{\underset{\sim}{y}}$$

$$W = (s_{\underset{\sim}{y}} / s^R_{\underset{\sim}{y}})^2 I .$$

These distributions will shift and become inflated as an observation is moved from the centre of $\underset{\sim}{y}^R$. The degree of shifting will depend on how the projection $a(\underset{\sim}{y}) = V'\underset{\sim}{y}$ is affected by the changing observation.

The distribution for S will be a rescaled chi variable

$$(s^R_{\underset{\sim}{y}} / s_{\underset{\sim}{y}})\chi_{n-r} .$$

These distributions will become progressively concentrated as an observation is moved out.

We saw in the last section, that a Student analysis appears to correct in a manner that would lead to stable adjusted conditional distributions for $\underset{\sim}{T}$ and S .

C. The Dependence on the Deviation Vector: Distributions

We saw in Sections 3F and 3H how scaled normal error forms can be used in the study of the effect of the deviation vector on the distributions with a location-scale model. In Section 2G, the distributions for $\underset{\sim}{T}$ and $\underset{\sim}{s}$ were displayed for the regression model with general symmetric matrix Σ . We now specialize these results to the case where

$$\Sigma = \Sigma_k = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & O \\ & & 1 & & & \\ & & & \tau_k & & \\ & & & & 1 & \\ & O & & & & \ddots \\ & & & & & & 1 \end{pmatrix} .$$

It is convenient to display the matrix V as a column of rows (this is a bit nonstandard, but the resulting formulae can be displayed fairly compactly)

$$V = \begin{pmatrix} \underset{\sim}{v}_1' \\ \vdots \\ \underset{\sim}{v}_n' \end{pmatrix} \qquad \underset{\sim}{v}_i' = (v_{1i}, \ldots, v_{ri}) .$$

With the specialization of Σ to Σ_k , it will be seen that the distributions depend on d_k^2 , $\underset{\sim}{v}_k$ and τ_k only. Several matrix identities are used to display the expressions. An excellent source for these and other identities is Srivastava and Khatri (1978)(Chapter 1).

$$V'\Sigma_k^{-1}V = I + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k\underset{\sim}{v}_k'$$

$$\left(V'\Sigma_k^{-1}V\right)^{-1} = I - \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k\underset{\sim}{v}_k'\left(1 + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k\right)^{-1}$$

$$\underset{\sim}{d}'\Sigma_k^{-1}\underset{\sim}{d} = 1 + \left(\frac{1}{\tau_k} - 1\right)d_k^2$$

$$\underset{\sim}{d}'R_k\underset{\sim}{d} = 1 + \frac{\left(\frac{1}{\tau_k} - 1\right)d_k^2}{\left(1 + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}'\underset{\sim}{v}_k\right)}$$

$$\left(V'\Sigma_k^{-1}V\right)^{-1}V\Sigma_k^{-1}\underset{\sim}{d} = \left[\frac{\left(\frac{1}{\tau_k} - 1\right)d_k}{1 + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k}\right]\underset{\sim}{v}_k$$

$$\left|V'\Sigma_k^{-1}V\right| = \left|I + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k\underset{\sim}{v}_k'\right|$$

$$= 1 + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k \ .$$

It is of interest to note that these expressions reduce to the expressions derived in Section 3F when

$$V = \underset{\sim}{1}/\sqrt{n} \quad \text{so that} \quad v_k = 1/\sqrt{n} \ .$$

When the error form is $f_{\Sigma_k}(\underset{\sim}{z})$, the conditional distribution for $\underset{\sim}{T}$ given $\underset{\sim}{d}$ is Student$_{n-r}(\underset{\sim}{\nu} , W)$ on $\mathbb{R}^r$ where

$$\underset{\sim}{\nu} = -\left[\frac{\left(\frac{1}{\tau_k} - 1\right) d_k}{1 + \left(\frac{1}{\tau_k} - 1\right) \underset{\sim}{v}_k' \underset{\sim}{v}_k}\right] \underset{\sim}{v}_k$$

$$W = \left[I - \left(\frac{1}{\tau_k} - 1\right) \underset{\sim}{v}_k \underset{\sim}{v}_k'\right] \Bigg/ \left[1 + \frac{\left(\frac{1}{\tau_k} - 1\right) d_k^2}{\left[1 + \left(\frac{1}{\tau_k} - 1\right) \underset{\sim}{v}_k' \underset{\sim}{v}_k\right]}\right] ;$$

call it $g^L_{\Sigma_k} (\underset{\sim}{T} \mid \underset{\sim}{d})$.

The conditional distribution for s given $\underset{\sim}{d}$ can be described as

$$\left(1 + \frac{\left(\frac{1}{\tau_k} - 1\right) d_k^2}{\left(1 + \left(\frac{1}{\tau_k} - 1\right) \underset{\sim}{v}_k' \underset{\sim}{v}_k\right)}\right)^{\frac{1}{2}} s \sim \chi_{n-r} ;$$

call it $g^s_{\Sigma_k} (s \mid \underset{\sim}{d})$.

Once again, we note that the familiar statistics would be

$$\underset{\sim}{t}_{\underset{\sim}{z}} = \sqrt{n-r}\, \underset{\sim}{T} \qquad s_{\underset{\sim}{z}} = s/\sqrt{n-r} .$$

The distributions for $\underset{\sim}{t}_{\underset{\sim}{z}}$ and $s_{\underset{\sim}{z}}$ are just rescaled versions of the above displayed distributions.

When the error form is $\Sigma c_i f_{\Sigma_i} (\underset{\sim}{z})$, the distribution for $\underset{\sim}{T}$ given $\underset{\sim}{d}$ will be

$$\frac{\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d}) g^{L}_{\Sigma_i}(\underset{\sim}{T} \mid \underset{\sim}{d})}{\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d})}$$

and the distribution for s given $\underset{\sim}{d}$ will be

$$\frac{\Sigma c_i h_{\Sigma_i}(\underset{\sim}{d}) g^{s}_{\Sigma_i}(s \mid \underset{\sim}{d})}{\Sigma c_i h_{\Sigma_i}(\underset{\sim}{d})}$$

where $h_{\Sigma_k}(\underset{\sim}{d})$ is the marginal probability for $\underset{\sim}{d}$ when the error form is $f_{\Sigma_k}(\underset{\sim}{z})$

$$h_{\Sigma_k}(\underset{\sim}{d}) = \frac{1}{A_{n-r}} \cdot \frac{\tau_k^{-\frac{1}{2}}\left[1 + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k\right]^{-\frac{1}{2}}}{\left[1 + \frac{\left(\frac{1}{\tau_k} - 1\right)d_k^2}{\left[1 + \left(\frac{1}{\tau_k} - 1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k\right]}\right]^{\frac{n-r}{2}}} .$$

Notice that the conditional densities are sums of densities weighted by the probabilities $h_{\Sigma_i}(\underset{\sim}{d})$. Large values of d_k^2 will lead to values of $h_{\Sigma_k}(\underset{\sim}{d})$ that will tend to dominate the form of the conditional densities. In certain situations we may find the use of a single f_{Σ_k} leads to densities that closely approximate the densities derived from Πf_λ .

In Figure 4.2 from Section A, we saw how the use of a Student analysis leads to conditional distributions for $\underset{\sim}{t}$ that are nonellipsoidal and shifted from origin. The distributions $g^L_{\Sigma_k}$ shift along the vector $\underset{\sim}{v}_k$ and develop ellipsoidal shapes as τ_k is varied. The distributions derived from $\Sigma c_i f_{\Sigma_i}$ will have more complicated contours and will shift in more general ways.

We now study the model for possible likelihood functions when the error form is $f_{\Sigma_k}(\underset{\sim}{z})$.

The marginal likelihood for τ_k is then

$$L(\underset{\sim}{d} \mid \tau_k) = c h_{\Sigma_k}(\underset{\sim}{d}) \ .$$

We now choose the representative curve that allows for direct comparisons with $\tau_k = 1$

$$L(\underset{\sim}{d} \mid \tau_k) = A_{n-r} h_{\Sigma_k}(\underset{\sim}{d}) \ .$$

The likelihood function depends only on d_k^2 and its associated marginal distribution. If we can find the distribution for d_k^2 for general τ_k , then we will have described the model for possible likelihood functions.

Let $f^k_{\tau_k}(d_k^2)$ denote the distribution for d_k^2 and let $L_k(d_k^2 \mid \tau_k)$ denote the likelihood function as computed from d_k^2 . Then

$$f^k_{\tau_k}(d_k^2) = f_1^k(d_k^2)L_k(d_k^2 \mid \tau_k) \ . \tag{4.1}$$

To display the distribution $f^k_{\tau_k}$, all we require is f_1^k to complete the formula.

We now show that when $\tau_k = 1$

$$\frac{d_k^2}{1 - \underset{\sim}{v}_k' \underset{\sim}{v}_k} \sim \text{beta}\left(\frac{1}{2}, \frac{n-r-1}{2}\right) \ .$$

Let $\underset{\sim}{z}$ denote a sample from $N(0, 1)$. Replace $\underset{\sim}{z}$ by new variables $\underset{\sim}{x}$ by means of an orthogonal transformation where

$$x_1 = a_1(\underset{\sim}{z})$$

$$\vdots$$

$$x_r = a_r(\underset{\sim}{z})$$

$$x_{r+1} = \left(z_k - \underset{\sim}{v}_k' \underset{\sim}{a}(\underset{\sim}{z})\right) \Big/ \sqrt{1 - \underset{\sim}{v}_k' \underset{\sim}{v}_k}$$ (verified in supplement to this section)

Then

$$\frac{d_k^2}{1 - \underset{\sim}{v}_k' \underset{\sim}{v}_k} = \frac{\left(z_k - \underset{\sim}{v}_k' \underset{\sim}{a}(\underset{\sim}{z})\right)^2 \Big/ (1 - \underset{\sim}{v}_k' \underset{\sim}{v}_k)}{s^2(\underset{\sim}{z})}$$

which is

$$\frac{x_{r+1}^2}{x_{r+1}^2 + \sum_{i=r+2}^{n} x_i^2}$$

which is distributed as a

$$\chi^2_{(1)} \Big/ \left(\chi^2_{(1)} + \chi^2_{(n-r-1)}\right) \text{ variable}$$

and which is

$$\text{beta}\left(\frac{1}{2}, \frac{n-r-1}{2}\right) .$$

Notice that

$$-\sqrt{1 - \underset{\sim}{v}_k'\underset{\sim}{v}_k} \le d_k \le \sqrt{1 - \underset{\sim}{v}_k'\underset{\sim}{v}_k} .$$

(One can also show that d_k has a relocated and rescaled symmetric beta $\left(\frac{n-r-1}{2}, \frac{n-r-1}{2}\right)$ distribution on $\left[-\sqrt{1 - \underset{\sim}{v}_k'\underset{\sim}{v}_k} , \sqrt{1 - \underset{\sim}{v}_k'\underset{\sim}{v}_k}\right]\Big)$.

Using equation (4.1), we can now display the distribution for d_k^2

$$f^{k}_{\tau_k}(d_k^2) = \frac{\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\cdot\Gamma\left(\frac{n-r-1}{2}\right)}\cdot\frac{\tau_k^{-\frac{1}{2}}\left[1+\left(\frac{1}{\tau_k}-1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k\right]^{-\frac{1}{2}}}{\left\{1+\frac{\left(\frac{1}{\tau_k}-1\right)d_k^2}{\left[1+\left(\frac{1}{\tau_k}-1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k\right]}\right\}^{\frac{n-r}{2}}}$$

$$\cdot\ \frac{(1-\underset{\sim}{v}_k'\underset{\sim}{v}_k)^{-\frac{1}{2}}}{(d_k^2)^{\frac{1}{2}}}\left[1-\frac{d_k^2}{1-\underset{\sim}{v}_k'\underset{\sim}{v}_k}\right]^{\frac{n-r-1}{2}-1}$$

$$0 \le d_k^2 \le 1-\underset{\sim}{v}_k'\underset{\sim}{v}_k\ .$$

This distribution describes the sampling properties of the marginal likelihood function when the error form is $f_{\Sigma_k}(\underset{\sim}{z})$.

This distribution can be displayed in terms of a beta distribution also. In fact,

$$\frac{\left[\frac{1}{\tau_k}\cdot\frac{1}{1+\left(\frac{1}{\tau_k}-1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k}\right]\frac{d_k^2}{1-\underset{\sim}{v}_k'\underset{\sim}{v}_k}}{\left[1-\frac{\left(\frac{1}{\tau_k}-1\right)d_k^2}{1+\left(\frac{1}{\tau_k}-1\right)\underset{\sim}{v}_k'\underset{\sim}{v}_k}\right]} \sim \text{beta}\left(\frac{1}{2}, \frac{n-r-1}{2}\right)\ .$$

When the error form is $\sum_{i=1}^{n} c_i f_{\Sigma_i}(\underset{\sim}{z})$ the marginal likelihood function will be

$$\sum_{i=1}^{n} c_i L_i(d_i^2 \mid \tau_i) \ .$$

Supplement

We do have one detail to verify. We need to check that

$$\mathrm{Var}\left(z_k - \underset{\sim}{v}_k' \underset{\sim}{a}(\underset{\sim}{z})\right) = 1 - \underset{\sim}{v}_k' \underset{\sim}{v}_k \ .$$

Here is one way to verify this. Let

$$\underset{\sim}{z}_{/k} = (z_1 , \ldots , z_{k-1} , z_{k+1} , \ldots , z_n)'$$

and let

$$V_{/k} = (\underset{\sim}{v}_1' , \ldots , \underset{\sim}{v}_{k-1}' , \underset{\sim}{v}_{k+1}' , \ldots , \underset{\sim}{v}_n')'$$

and let

$$\underset{\sim}{a}_{/k}(\underset{\sim}{z}_{/k}) = V_{/k}' \underset{\sim}{z}_{/k} \ .$$

Then

$$\underset{\sim}{a}(\underset{\sim}{z}) = \underset{\sim}{a}_{/k}(\underset{\sim}{z}_{/k}) + \underset{\sim}{v}_k z_k$$

so that

$$z_k - \underset{\sim}{v}_k' \underset{\sim}{a}(\underset{\sim}{z}) = (1 - \underset{\sim}{v}_k' \underset{\sim}{v}_k) z_k - \underset{\sim}{v}_k a_{/k}(\underset{\sim}{z}_{/k}) \ .$$

Now note that z_k and $\underset{\sim}{a}_{/k}$ are independent and since $V'_{/k}V_{/k} = V'V - \underset{\sim}{v}_k\underset{\sim}{v}'_k = I - \underset{\sim}{v}_k\underset{\sim}{v}'_k$,

$$\mathrm{Var}\left(\underset{\sim}{a}_{/k}(\underset{\sim}{z}_{/k})\right) = I - \underset{\sim}{v}_k\underset{\sim}{v}'_k$$

and therefore

$$\mathrm{Var}\left(z_k - \underset{\sim}{v}'_k\underset{\sim}{a}(\underset{\sim}{z})\right) = (1 - \underset{\sim}{v}'_k\underset{\sim}{v}_k)^2 - \underset{\sim}{v}'_k(I - \underset{\sim}{v}_k\underset{\sim}{v}'_k)\underset{\sim}{v}_k$$

$$= 1 - \underset{\sim}{v}'_k\underset{\sim}{v}_k$$

as required.

D. Importance Sampling Monte Carlo

Monte carlo integration appears to be the most promising device available for the integration of multi-dimensional statistical functions. We now use the distributions derived in the last section as support densities for importance sampling monte carlo. We suspect that these densities will be quite similar to the actual densities for $\underset{\sim}{t}$ and s desired for a Student analysis. We can be guided to the choice of the τ_k's based on knowledge gained for the work done with the location-scale model.

In Section 3G, quadrature rules were used to integrate out the coordinate for s. Monte carlo work in all $r+1$ variables appears to have problems directly traceable to the s coordinate and empirical work suggests that it is wise to continue integrating out s using a simple quadrature rule towards the construction of the distribution for $\underset{\sim}{t}$.

For inference concerning $\underset{\sim}{\alpha}$ or $\underset{\sim}{\beta}$, we desire

$$f_\lambda^L(\underset{\sim}{t} \mid \underset{\sim}{d})$$

but unfortunately all that is directly available is

$$h_\lambda(\underset{\sim}{d})\, f_\lambda^L(\underset{\sim}{t} \mid \underset{\sim}{d})$$

where $h_\lambda(\underset{\sim}{d})$ is unknown.

Let g denote the distribution for $\underset{\sim}{t}$ given $\underset{\sim}{d}$ when the error form is $\sum_{i=1}^{n} c_i f_{\Sigma_i}(\underset{\sim}{z})$. If we sample $(\underset{\sim}{t}_1 , \ldots , \underset{\sim}{t}_m)$ from g then

$$\frac{1}{m} \sum_{j=1}^{m} \frac{h_\lambda(\underset{\sim}{d}) f_\lambda^L(\underset{\sim}{t}_j \mid \underset{\sim}{d})}{g(\underset{\sim}{t}_j \mid \underset{\sim}{d})}$$

should be close to $h_\lambda(\underset{\sim}{d})$ with appropriate choice of g .

We may however be interested in the assessment of components of $\underset{\sim}{\alpha}$. If we are interested in $\underset{\sim}{\alpha}_2$ where $\underset{\sim}{\alpha} = (\underset{\sim}{\alpha}_1 , \underset{\sim}{\alpha}_2)'$ then an appropriate distribution for inference would be the conditional for $\underset{\sim}{t}_2$ given $\underset{\sim}{d}$ where $\underset{\sim}{t} = (\underset{\sim}{t}_1 , \underset{\sim}{t}_2)'$ on $\mathbb{R}^{r_1} \times \mathbb{R}^{r_2} = \mathbb{R}^r$.

We have that

$$f_\lambda^{\underset{\sim}{t}_2}(\underset{\sim}{t}_2 \mid \underset{\sim}{d}) = \int_{t_1} f_\lambda^L(\underset{\sim}{t}_1 , \underset{\sim}{t}_2 \mid \underset{\sim}{d}) d\underset{\sim}{t}_1 \ .$$

Let g denote a support density for this integration. Then

$$f_\lambda^{\underset{\sim}{t}_2}(\underset{\sim}{t}_2 \mid \underset{\sim}{d}) = \int_{\underset{\sim}{t}_1} \frac{f_\lambda^L(\underset{\sim}{t}_1 , \underset{\sim}{t}_2 \mid \underset{\sim}{d})}{g(\underset{\sim}{t}_1 \mid \underset{\sim}{t}_2 , \underset{\sim}{d})} g(\underset{\sim}{t}_1 \mid \underset{\sim}{t}_2 , \underset{\sim}{d}) d\underset{\sim}{t}_1 \ .$$

If $\underset{\sim}{t}_{11} , \ldots , \underset{\sim}{t}_{1m}$ denotes a sample of size m from g , then

$$\frac{1}{m} \sum_{j=1}^{m} \frac{f_\lambda^L(\underset{\sim}{t}_{1j} , \underset{\sim}{t}_2 \mid \underset{\sim}{d})}{g(\underset{\sim}{t}_{1j} \mid \underset{\sim}{t}_2 , \underset{\sim}{d})}$$

should be close to $f_{\lambda}^{\underset{\sim}{t}_2}(\underset{\sim}{t}_2 \mid \underset{\sim}{d})$.

Based on our earlier discussion, an appropriate support density would be the conditional density for $\underset{\sim}{t}_1$ given $\underset{\sim}{t}_2$ and $\underset{\sim}{d}$ derived from the error from $\sum_{i=1}^{n} c_i f_{\Sigma_i}(\underset{\sim}{z})$.

The conditional distribution for $\underset{\sim}{T}_1$ given $\underset{\sim}{T}_2$ and $\underset{\sim}{d}$ derived from a $N(0\ ,\Sigma)$ error form was displayed in expression (2.17) in Section 2G. It is again just a relocated and relocated Student density on $\mathbb{R}^{r_1}$. To obtain a value for the density for $\underset{\sim}{t}_2$ requires an r_1-dimensional monte carlo. This is assuming that $f_{\lambda}^{L}(\underset{\sim}{t} \mid \underset{\sim}{d})$ is available and we should recall that its norming constant required an r-dimensional monte carlo.

When the error form is $f_{\Sigma_k}(\underset{\sim}{z})$, the conditional for $\underset{\sim}{t}_1$ given $\underset{\sim}{t}_2$ and $\underset{\sim}{d}$ would be computed as the specialization of Σ to Σ_k . Denote it $g_{\Sigma_k}^{c}(\underset{\sim}{t}_1 \mid \underset{\sim}{t}_2\ ,\underset{\sim}{d})$. Similarly the marginal for $\underset{\sim}{t}_2$ (given $\underset{\sim}{d}$) is obtained from expression (2.17) by specializing Σ to Σ_k . Denote it $g_{\Sigma_k}^{m}(\underset{\sim}{t}_2 \mid \underset{\sim}{d})$.

We now display the distribution for $\underset{\sim}{t}_1$ given $\underset{\sim}{t}_2$ and $\underset{\sim}{d}$ when the error form is $\sum_{i=1}^{n} c_i f_{\Sigma_i}(\underset{\sim}{z})$. It would be

$$\frac{\sum_{i=1}^{n} e_i g_{\Sigma_i}^{m}(\underset{\sim}{t}_2 \mid \underset{\sim}{d})\, g_{\Sigma_i}^{c}(\underset{\sim}{t}_1 \mid \underset{\sim}{t}_2\ ,\underset{\sim}{d})}{\sum_{i=1}^{n} e_i g_{\Sigma_i}^{m}(\underset{\sim}{t}_2 \mid \underset{\sim}{d})}$$

where

$$e_k = \frac{c_k h_{\Sigma_k}(\underset{\sim}{d})}{\sum_{i=1}^{n} c_i h_{\Sigma_i}(\underset{\sim}{d})} .$$

This distribution is a weighted sum of the conditional densities derived from f_{Σ_k} error forms. The weights are based on the marginal densities for $\underset{\sim}{t}_2$ when the error forms are f_{Σ_k} .

All of these expressions can be fully displayed analytically. The implementation of these support densities for monte carlo and other comments concerning accuracy and efficiency are made in the next section.

E. Computing Time and Accuracy

Many of the ideas and methods discussed in the last two sections of this chapter and later sections of Chapter 3 have been implemented in the form of a computer program to handle nonnormal regression analysis. The program is set up to integrate using both quadrature and monte carlo to offer confidence intervals and other percentage points, to plot contours of sections of joint distributions and also to examine prospective support densities.

Several examples involving both real and generated data have been examined involving simple polynomial models and small factorial experiments. The results to date are very encouraging but there is still much work to be done.

We now give a brief illustration that gives a flavour of the methods that have been discussed and the direction for future work. The simple regression problem in Section A will suffice for this purpose.

The marginal likelihood and conditional distributions were computed in Section A using a three dimensional numerical quadrature technique. Over 6500 evaluations of the joint density for $(\underset{\sim}{t}, \underset{\sim}{d})$ were made in determining the desired quantities. The computer time needed to perform integration higher than three dimensional becomes unrealistic at the present time if these same techniques are used.

The type of accuracy required will of course depend on the application. For marginal likelihood assessment of λ, we require a reasonable approximation to $h_\lambda(\underset{\sim}{d})$. From earlier comments in this and the last chapter, the likelihood should be determined with an accuracy fine enough to indicate a plausible range of λ values or perhaps just fine enough to give an appropriate indication of nonnormality.

Next, one might wish to consult the joint distribution for $(\underset{\sim}{a}, s)$ or $(\underset{\sim}{t}, s)$ or $\underset{\sim}{t}$ given $\underset{\sim}{d}$. Practically, this will likely involve the plotting of sections of such densities using contour plotting techniques. Such plots could be used to gain information as to the form of confidence regions. It is the shape of the contours that would be of primary interest and in many situations only a course approximation to $h_\lambda(\underset{\sim}{d})$ would be needed.

Quite often, interest is likely to centre on component parameters. In determining the component distributions, (for say, t_1 or t_2 given $\underset{\sim}{d}$) one useful direction involves the use of fairly coarse monte carlo to evaluate the density at a reasonable selection of points.

The resulting approximate density could then be improved with the use of a smoothing algorithm. A one dimensional quadrature of the resulting smoothed function will give an idea of the accuracy of such a process. The most important characteristic of such densities is the degree of shifting relative to normal analysis. Detection of such shifting at a preliminary stage of analysis might suggest recalculation in a more precise manner. This detection would be displaying the importance of the use of a Student analysis.

Methods for choosing an appropriate support density are still under development. The choice of τ values could be based on a number of criteria. Studies with the location-scale model in Section 3G offer some direction. It then seems reasonable to compute $h_{\Sigma_i}(\underset{\sim}{d})$ for $i = 1, \ldots, n$. Recall that the conditional distributions for $\underset{\sim}{t}$ and s are weighted by these quantities. By consulting formula (4.1), each value of $h_{\Sigma_k}(\underset{\sim}{d})$ is heavily influenced by

$$\left(\frac{1}{\tau_k} - 1\right) d_k^2 \Big/ \left[1 + \left(\frac{1}{\tau_k} - 1\right) \underset{\sim}{v}_k' \underset{\sim}{v}_k\right] .$$

It seems to be efficient to use a support density made up of a sum of rescaled normals that involves only the dominating coordinates. With the example in Section A, the most influential coordinate is y_2. Consider now, the use of a single rescaled normal error form as support density $\left(f_{\Sigma_2}(\underset{\sim}{z})\right)$. Although several monte carlo sample sizes were considered, we illustrate here with a coarse one $(m = 100)$. Attention was centred on the λ values 6, 3 and 2. With $\tau_2 = 4$, we obtained

Estimates of $h_\lambda(\underset{\sim}{d})$

λ	Monte Carlo (m = 100)	Quadrature (m > 6500)
2	348	302
3	143	166
6	24	27

These numbers are included only to give the indication that there are obvious gains in efficiency even with this rather coarse support density.

Measures of the variability of such estimates are available with monte carlo. Once again, the most reasonable choice here is not clear; the sample range would be the most conservative and the standard error is also attractive. It does not seem appropriate to include these values here.

It seems to me that most of these ideas can and will be tightened up with future work. The results to date indicate that this method will eventually lead to substantial gains in the analysis of nonnormal data with even highly complex regression models.

CHAPTER 5

ROBUSTNESS AND RESISTANCE: A GENERAL FORMULATION

A. Motivation

In Sections 3D and 4A we saw how the use of a Student analysis appropriately handled nonnormal data by tolerating extreme observations in the tails that are not characteristic of normal data. At the same time, it was observed that a Student analysis also handles normal data in a manner very similar to a normal analysis.

In Sections 3E and 4B, adjusted statistics $\underset{\sim}{T}$ and S were constructed that gave an accurate measure of how well different analyses handle such extreme observations.

In each instance, we placed ourselves in the situation of having to deal with a data set $\underset{\sim}{y}$ that contains observation(s) that could have a substantial effect on classical normal analysis. Comparisons were made relative to a reference data set $\underset{\sim}{y}^R$ that represented the data set for which either form of analysis would lead to appropriate inferences. The data sets labeled $\underset{\sim}{y}$ were deliberately flawed by moving an observation away from the centre of data.

We now present these ideas in the general setting of a general variation-based model.

B. Resistance

Consider the general setting of a structural model

$$\begin{array}{ll} Y & \\ Y = \theta Z & \theta \in G \\ f_\lambda(Z)dZ & \lambda \in \Lambda \, . \end{array}$$

The concern now is that this model may not be precisely appropriate for possibly several reasons. The distribution form may not be adequately displaying the error system from which Y was obtained or Y may be made up of certain members that simply do not belong. Particularly with complex models such as regression, it may be an nontrivial problem to detect the members mentioned above. In any case, the data set Y is viewed as being flawed in that the formulated model is not appropriate. Interest now is with whether an analysis will tolerate such flaws and offer reasonably correct inferences.

By correct, we mean to think of having some standard reference data set, say Y^R . This data set could be viewed as the data that we should have obtained (for the model as presented to be correct), but instead Y was obtained (for reasons that are usually not of direct statistical concern).

Following the standard notation summarized in Chapter 1 we, of course, have

$$Y = [Y]D$$

$$[Y] = \theta[Z] \qquad g_\lambda([Z] : D)$$

$$D(Z) = D \qquad h_\lambda(D) \rightarrow L(D \mid \lambda) \ .$$

The central question now is 'Does g_λ adjust itself accordingly to account for the flaw that is possibly still reflected in $[Y]$?'

The point is that the analysis should have been based on $[Y^R]$ and inference concerning θ through $[Y^R]$. As we have seen in Sections 3D and 4B, it can be very illuminating to study the effect of the flaws in Y relative to $[Y^R]$.

We ask what variable ω satisfies

$$[Y^R] = \theta\omega$$

when in fact

$$[Y] = \theta[Z] \ .$$

Multiplying on the right by $[Y]^{-1}[Y^R]$ we obtain

$$[Y][Y]^{-1}[Y^R] = \theta[Z][Y]^{-1}[Y^R]$$

or

$$[Y^R] = \theta[Z][Y]^{-1}[Y^R]$$

so that

$$\omega = [Z][Y]^{-1}[Y^R] .$$

The distribution that displays the effect of the flaws relative to the correct $[Y^R]$ is

$$g_\lambda(\omega[Y^R]^{-1}[Y] : D)J_L^*([Y^R]^{-1}[Y] : \omega) .$$

Notice that if Y is in fact Y^R then we obtain the conditional distribution for $[Z]$. Otherwise, this distribution can be used to directly construct observed confidence intervals or tests of significance computed from Y^R .

If we contemplate a study to investigate resistance with a particular analysis, we might begin with Y^R and gradually add perturbations of interest. It is closeness of these adjusted conditional distributions that measure the ability of analysis and resist the effect of the perturbations.

In this general setting we might begin with some reference data set Y^R and alter one or more of coordinates of Y^R in some systematic way and assess the adjusted distributions.

C. An Illustration

It is of interest to specialize the general formulae displayed in Section B to the location-scale case. We have

$$\underset{\sim}{y} = [\mu , \sigma]\underset{\sim}{z}$$

where $[\mu , \sigma]\underset{\sim}{z} = \mu\underset{\sim}{1} + \sigma\underset{\sim}{z}$ (on S) .

$$[a_1 , c_1][a_2 , c_2] = [a_1 + c_1 a_2 , c_1 c_2] \qquad (\text{ on } G) .$$

Accordingly we have

$$[\bar{y} , s(\underset{\sim}{y})] = [\mu , \sigma][\bar{z} , s(\underset{\sim}{z})]$$

and

$$\underset{\sim}{d} = [\bar{z} , s(\underset{\sim}{z})]^{-1} \underset{\sim}{z} .$$

Now $\underset{\sim}{y}^R$ denotes the reference data set, then

$$[\bar{y} , s(\underset{\sim}{y})]^{-1}[\bar{y}^R , s^R(\underset{\sim}{y})]$$

$$= \left[\frac{\bar{y}^R - \bar{y}}{s(\underset{\sim}{y})} , \frac{s^R(\underset{\sim}{y})}{s(\underset{\sim}{y})}\right]$$

and

$$[\bar{z}\ ,\ s(\underset{\sim}{z})]\left[\frac{\bar{y}^R - \bar{y}}{s(\underset{\sim}{y})}\ ,\ \frac{s^R(\underset{\sim}{y})}{s(\underset{\sim}{y})}\right]$$

$$= \left[\bar{z} + \frac{\bar{y}^R - \bar{y}}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\ ,\ \frac{s^R(\underset{\sim}{y})}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\right]$$

so that we have

$$[\bar{y}\ ,\ s^R(\underset{\sim}{y})] = [\mu\ ,\ \sigma]\left[\bar{z} + \frac{\bar{y}^R - \bar{y}}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\ ,\ \frac{s^R(\underset{\sim}{y})}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\right]$$

and displaying the components, we have

$$\bar{y}^R = \mu + \sigma\left\{\bar{z} + \frac{\bar{y}^R - \bar{y}}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\right\}$$

$$s^R(\underset{\sim}{y}) = \sigma\, \frac{s^R(\underset{\sim}{y})}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\ .$$

Stability with respect to inference concerning μ would be based on

$$T = \left[\bar{z} + \frac{\bar{y}^R - \bar{y}}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})\right] \Big/ \frac{s^R(\underset{\sim}{y})}{s(\underset{\sim}{y})}\, s(\underset{\sim}{z})$$

$$= \frac{s(\underset{\sim}{y})}{s^R(\underset{\sim}{y})} \cdot \frac{\bar{z}}{s(\underset{\sim}{z})} + \frac{\bar{y}^R - \bar{y}}{s^R(\underset{\sim}{y})}$$

$$= \frac{1}{\sqrt{n}\sqrt{n-1}} \left(\frac{s_{\underset{\sim}{y}}}{s^{R}_{\underset{\sim}{y}}} \cdot \frac{\bar{z}}{s_{\underset{\sim}{z}}/\sqrt{n}} + \frac{\bar{y}^{R} - \bar{y}}{s^{R}_{\underset{\sim}{y}}/\sqrt{n}} \right) .$$

The expression in parentheses is expression (3.1).

Similarly, stability with respect to inference concerning σ would be based on

$$S = \frac{s^{R}(\underset{\sim}{y})}{s(\underset{\sim}{y})} \, s(\underset{\sim}{z})$$

$$= \sqrt{n-1} \left(\frac{s^{R}_{\underset{\sim}{y}}}{s_{\underset{\sim}{y}}} \, s_{\underset{\sim}{z}} \right) .$$

The expression in parentheses is expression (3.2).

D. Robustness

We noted earlier that the central question is the degree to which g_λ adjusts to correct for the flaw in $[Y]$.

We are of course free to choose any transformation variable $[\ ]$ we wish since our inferences concerning θ and λ are not affected by such a choice.

It is of interest to note that, depending on the form of the flaw in Y , $[Y]$ may not reflect it. In other words, although Y may become considerably different from Y^R , $[Y]$ may be very similar to $[Y^R]$.

Many robustness studies are concerned with finding a variable $[\ ]$ so that $[Y]$ is only moderately affected by bad coordinates in Y (Andrews et al [1971]).

From the point of view of this chapter, we observe that if such a $[\ ]$ were chosen, then $[Y^R]^{-1}[Y]$ would likely be close to the identity of the group G and we could consult the unadjusted conditional distributions for $[Z]$ to assess the resistance of the analysis.

Traditional resistance studies, in fact deal with the marginal distributions of such statistics usually through the extensive use of monte carlo (see Relles and Rogers (1977)).

The point of view here is overwhelmingly directed towards the study of the appropriate conditional distributions. Inferential statements of conditional confidence also have the marginal confidence interpretation.

REFERENCES

1. Andrews et al. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton University Press, New Jersey.

2. Barnard, G. (1976). Lecture at the University of Toronto.

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

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

5. Dempster, A.P. (1969). *Elements of Continuous Multivariate Analysis.* Addison Wesley, Reading, Mass.

6. Duncan, D.B. and Jones, R.H. (1966). "Multiple Regression with Stationary Errors", JASA 61, 917-928.

7. Fick, G.H. (1975). "Location Scale Analysis Implementation (LSAI)". Dept. of Statistics, University of Toronto.

8. Fick, G.H. and Fraser, D.A.S. (1976). "Robustness with Structural Methods", Ball State University Technical Report, 75-93.

9. Fisher, R.A. (1971). *The Design of Experiments.* Collier MacMillan, New York.

10. Fraser, D.A.S. (1968). *The Structure of Inference,* Krieger, Huntington, N.Y.

11. ________ (1976). *Probability and Statistics: Theory and Applications.* Duxbury, North Scituate, Mass.

12. ________ (1976) "Necessary Analysis and Adaptive Inference" JASA 71, 99-113.

13. ________ (1978). *Inference and Linear Models*, McGraw Hill, New York.

14. Fraser, D.A.S. and Fick, G.H. (1975). "Necessary Analysis and Its Implementation", Carleton University Technical Report, Ottawa, pp. 5.01 - 5.30.

15. Fraser, D.A.S., Guttman, I., and Styan, G.P.H. (1976). "Serial Correlation and Distributions of the Sphere", Commun. Statis.-Theory and Method, Vol. A5, 97-118.

16. Hammersley, J.M. and Handscomb, D.C. (1964). *Monte Carlo Methods*, Methuen, London.

17. Kagan, A.M., Linnik, Y.V., and Rao, C.R. (1973). *Characterization Problems in Mathematical Statistics*, John Wiley and Sons, New York.

18. Knuth, D.E. (1969). *The Art of Computer Programming*, Vol. 2. Addison Wesley, Reading, Mass.

19. Lund, D.R. (1967). "Parameter Estimation in a Class of Power Distributions", Ph.D. Thesis, University of Wisconsin, Madison.

20. Marsgalia, G. (1974). Super-Duper: Random Number Package. McGill University, Montreal.

21. Prokhorov, Y.V. (1965). "Characterization of a class of distributions through the distribution of certain statistics". Teoriia Veroiatin Prim. X, 479-487.

22. Relles, D.A. and Rogers, W.H. (1977). "Statisticians are fairly robust estimators of location". JASA 72, 107-112.

23. Sprott, D.A. (1977). Lecture at the University of Toronto.

24. Srivastava, M.S. and Khatri, C.G. (1978). *An Introduction to Multivariate Statistics* (to appear).

25. Zinger, A.A. (1956). "On a problem of Kolmogorov's". Vestnik Leningrad Univ. 1, 53-56.

26. Zinger, A.A. and Linnik, Y.V. (1964). "On characterizations of the normal distributions", Teoriia Veroiatin Prim. IX, 692-695.