NECESSARY ANALYSIS AND ITS IMPLEMENTATION

by
D.A.S. Fraser and G.H. Fick
University of Toronto [March 1975]

1. Introduction

Consider an experimenter at an early stage of analysis of his data. Suppose that his data are values of a real valued response and that no input variables to the system have been identified; thus

$$\boldsymbol{y}=(y_1\ ,\ \ldots\ ,\ \ y_n)'$$

And suppose that he wants to use a statistical model such as

$$f_0(\boldsymbol{y}\mid \boldsymbol{\theta_0})\,d\,\boldsymbol{y}$$

where $\boldsymbol{\theta_0}$ is a parameter that indexes a class of probability density functions. Perhaps the dimension of the parameter is too large to analyze in some way component by component; he may then consider some of the many nonparametric tests. Or perhaps the dimension of $\boldsymbol{\theta_0}$ is small and that with specific characteristics for analysis at hand; he may then consider the methods in this paper. Suppose that early investigations support this alternative situation.

Of primary interest in his investigation is likely to be the general level of $\boldsymbol{y}$. This suggests viewing the response $\boldsymbol{y}$ as

$$\boldsymbol{y}=\mu\,\boldsymbol{1}+\boldsymbol{\epsilon}$$

where μ is the general level or location of y and $\boldsymbol{\epsilon}$ does not reflect the location of $\boldsymbol{y}$. To describe $\boldsymbol{\epsilon}$, we offer a class of probability density functions that does not reflect location characteristics, say

$$f_1(\boldsymbol{\epsilon}\mid \boldsymbol{\theta_1})\,d\,\boldsymbol{\epsilon}$$

Also of interest is likely to be the scaling of the response $\boldsymbol{y}$ and this suggests viewing the response as

$$\boldsymbol{y}=\mu\,\boldsymbol{1}+\sigma\,\boldsymbol{z}$$

where σ indicates the scale in which $\boldsymbol{y}$ was obtained and $\boldsymbol{z}$ does not reflect the location or scale of y . To describe $\boldsymbol{z}$ we offer a class of probability density functions that does not reflect location or scaling characteristics, say

$$f_2(\boldsymbol{z}\mid \boldsymbol{\theta_2})\,d\,\boldsymbol{z}$$

The experimenter is then left with deciding what other characteristics of his data can be extracted in the manner out lined and with contemplating a class of density functions for the remaining differences.

Suppose he has decided on the preceding model and in addition has elected to allow only a one

dimensional parameter λ for the class of densities for $\boldsymbol{z}$. He then has

$$\boldsymbol{y}$$

$$\boldsymbol{y} = \mu \mathbf{1} + \sigma \boldsymbol{z} \qquad (1)$$

$$f_2(\boldsymbol{z} \mid \lambda)\, d\boldsymbol{z}$$

And further to this, suppose he views each z_i , as independent and identically distributed so that

$$f_2(\boldsymbol{z} \mid \lambda)\, d\boldsymbol{z} = \prod f_\lambda(z_i) \prod dz_i$$

The common procedure at this stage is to then use the response distribution for y based on the above identifications; that is

$$\prod f_\lambda\left(\frac{y_i - \mu}{\sigma}\right) \frac{1}{\sigma^n} \prod dy_i$$

This is then viewed as the model for further study, the classical model of statistics.

$$\boldsymbol{y}$$

$$\prod f_\lambda\left(\frac{y_i - \mu}{\sigma}\right) \frac{1}{\sigma^n} \prod dy_i$$

The observed likelihood function could then be examined examined and inferences made about μ , σ and λ .

Plausible values of λ might be considered and studies concerning μ and σ conditional on λ might be undertaken.

The question to be asked is, Is there further information available in the model outlined in (1) ? Can we study the model more and logically deduce additional facts rather than immediately infer from the response $\boldsymbol{y}$?

The remainder of this paper is concerned with demonstrating a procedure that does reconsider the model at the stage of equation (1). This paper illustrates the kinds of inferences that can be made as a result of the further study. The conclusions rely upon the data and the model alone with no recourse to the usual additional assumptions in conventional statistics: The analysis necessarily follows from the data and the model alone. Thus we speak of necessary analysis.

The analysis is illustrated with the aid of a computer program that can compare the usual normal distribution theory analysis with the analysis to be described next in this paper. The program offers estimates and confidence intervals for the parameters of primary interest. Section 2 is concerned with detailing the analysis used and subsequent sections are concerned with the implementation of the methods. The reader is invited to view what can be done using this procedure with examples and perhaps then to return later to examine and study the derivations. The procedures offer greatly

increased flexibility. And they may be very important to the experimenter in obtaining strong and valid conclusions from his experiment.

2. Derivations (The Location Scale Model)

The classical model describes the response variable $\boldsymbol{y}$ by a class of density functions indexed by a class of parameters. The model of equation (1) describes what will be called the error variable $\boldsymbol{z}$ by a class of density functions indexed by the parameter λ. Accordingly, inference concerning λ will be made based on the classical model but what about μ and σ? How can we use the observed $\boldsymbol{y}$ to gain information about $\boldsymbol{z}$, μ and σ?

Since $\boldsymbol{y}$ was obtained from $\boldsymbol{z}$ by a linear transformation we can solve for $\boldsymbol{z}$ in terms of $\boldsymbol{y}$ by the inverse transformation

$$\boldsymbol{z}=-\frac{\mu}{\sigma}\mathbf{1}+\frac{1}{\sigma}\boldsymbol{y}$$

But since μ and σ are unknown, all we can say is that :

$$\boldsymbol{z}\in L^{+}(\mathbf{1},\boldsymbol{y})=\{a\mathbf{1}+b\,\boldsymbol{y}:a\in\mathbb{R},b\in\mathbb{R}^{+}\}\quad .$$

While $\boldsymbol{z}$ could initially be viewed as any point in $\mathbb{R}^{n}$, we find by using our assumed model (1) that $\boldsymbol{z}$ is constrained to a particular two dimensional subspace of $\mathbb{R}^{n}$, the half plane generated by the $\mathbf{1}$ vector and the response vector $\boldsymbol{y}$. By studying where $\boldsymbol{z}$ could be on this half plane $L^{+}(\mathbf{1},\boldsymbol{y})$ in relation to $\boldsymbol{y}$ we will be able to infer about μ and σ because of the way in which $\boldsymbol{z}$ and $\boldsymbol{y}$ are linked by μ and σ.
Notice that this identification gave us no information concerning λ. It will be the distribution of the possible half planes that will give us information concerning λ. This distribution will be shown to not depend:on μ and σ.

Specifically now we are interested in obtaining the conditional distribution for $\boldsymbol{z}$ Ze given that $\boldsymbol{z}$ is constrained to $L^{+}(\mathbf{1},\boldsymbol{y})$. This is a familiar procedure. We derive the distribution for $\boldsymbol{z}$ expressed in terms of coordinates on the half plane and a set of complementary coordinates and then appropriately normalize, given the complementary coordinates.

The particular coordinates that we choose for $\boldsymbol{z}$ on the half plane are not important and the remainder of the analysis does not depend on the choice as we shall see later. We might as well choose convenient and familiar coordinates. We choose orthogonal vectors in $L^{+}(\mathbf{1},\boldsymbol{y})$. Specifically the $\mathbf{1}$ vector and the vector in $L^{+}(\mathbf{1},\boldsymbol{y})$ that is orthogonal to the $\mathbf{1}$ vector labelled $\boldsymbol{d}(\boldsymbol{z})$, the vector $\boldsymbol{d}(\boldsymbol{z})$ that has
unit length is of course $\frac{(\boldsymbol{z}-\bar{\boldsymbol{z}})}{s(\boldsymbol{z})}$ where $s^{2}(\boldsymbol{z})=\sum_{i=1}^{n}(z_i-\bar{z})^{2}$. The coordinates on the half plane are then $\bar{z},s(\boldsymbol{z})$ and we have $\boldsymbol{z}=\bar{z}\mathbf{1}+s(\boldsymbol{z})\boldsymbol{d}(\boldsymbol{z})$. The complementary coordinates are given by $\boldsymbol{d}=\boldsymbol{d}(\boldsymbol{z})$. We will not need complementary coordinates without constraints but they could be found in a convenient form by using say the Gram Schmidt method to find n-2 complementary vectors. Now notice that by identifying $\boldsymbol{d}(\boldsymbol{z})$ we have identified $L^{+}(\mathbf{1},\boldsymbol{y})$ and conversely; thus
$\boldsymbol{d}(\boldsymbol{z})=\boldsymbol{d}(\boldsymbol{y})=\boldsymbol{d}$.

Notice that $\boldsymbol{d}$ indexes the half planes $L^{+}(\mathbf{1}, .)$ and can be called a reference point for the planes. Also notice that studying the distribution of possible $\boldsymbol{d}'s$ is equivalent to studying the distribution for the possible half-planes.

The information we have extracted thus far is to identify $\boldsymbol{z}$ on the half plane $L^{+}(\mathbf{1}, \boldsymbol{y})$. And we have no information concerning the position of $\boldsymbol{z}$ on the plane. This is the same thing as saying that $\bar{z}, s(\boldsymbol{z})$ is unknown and that we have valid grounds for studying the conditional distribution for $\bar{z}, s(\boldsymbol{z})$ given $\boldsymbol{d}$ (the identification of the half plane). More specifically, we can say that some $\boldsymbol{z}$ on $L^{+}(\mathbf{1}, \boldsymbol{y})$ has obtained. In fact any $\boldsymbol{z}$ on $L^{+}(\mathbf{1}, \boldsymbol{y})$ could have obtained if our density f_{λ} is nonzero, and any $\boldsymbol{z}$ on $L^{+}(\mathbf{1}, \boldsymbol{y})$ would lead to the same information (that is $\boldsymbol{d}(\boldsymbol{z})=\boldsymbol{d}(\boldsymbol{y})=\boldsymbol{d}$). These conditions may seem to be rather obtuse but such requirements are essential for the proper use of conditional probability theory.

From the initial linking relationship between $\boldsymbol{y}$ and $\boldsymbol{z}$, we can now observe that

$$\bar{y}=\mu+\sigma \bar{z}$$

$$s(\boldsymbol{y})=\sigma s(\boldsymbol{z})$$
$$(\text{ label } \quad s(\boldsymbol{z})=s \quad)$$

The next step in the analysis is to find the Jacobian necessary to transfer from $\boldsymbol{z}$ to $(\bar{z}, s, \boldsymbol{d})$.

There are a number of ways of doing this. One would be to calculate the Jacobian matrix and evaluate its determinant.

$$\left|\frac{\partial \boldsymbol{z}}{\partial(\bar{z}, s, \boldsymbol{d})}\right|_{+}$$

Another method involves the powerful tool of invariant differentials (see Fraser (1968), p. 28-32).

More geometrically and perhaps more intuitively than the first two methods is to directly reexpress Euclidean volume in terms of the new coordinates. To measure length in the direction of the first basis vector $\mathbf{1}$ we have $d\bar{z}$ but because $\mathbf{1}$ has length $\sqrt{n}$, $d\bar{z}$ measures length by a factor $\sqrt{n}$ smaller than Euclidean length and therefore $\sqrt{n}\, d\bar{z}$ measures Euclidean length along $\mathbf{1}$. Since $\boldsymbol{d}$ has length 1 then ds measures Euclidean length directly. Euclidean volume at $\boldsymbol{d}$ orthogonal to $L^{+}(\mathbf{1}, \boldsymbol{y})$ will be labelled as da and it will measure the surface area on the unit sphere in $L^{\perp}(\mathbf{1})$; then Euclidean volume at $s\boldsymbol{d}$ orthogonal to $L^{+}(\mathbf{1}, \boldsymbol{y})$ is $s^{n-2} da$

Notice of course that our choice to measure volume at $\boldsymbol{d}$ was essentially arbitrary and other choices would work as easily. The required Jacobian is then $\sqrt{n} s^{n-2}$. Thus we have

$$\prod f_{\lambda}(z_i) \prod dz_i=\prod f_{\lambda}(\bar{z}+sd_i) \sqrt{n}\, d\bar{z}\, s^{n-2} ds\, da$$

Actually this method of obtaining the Jacobian has many powerful extensions and amounts to a descriptive explanation that has its foundations in invariant differentials.

(The reader is invited to study Fraser (1968), Chapter 2).

The marginal density for $\boldsymbol{d}$ (which indexes the half planes) is obtained by integration

$$k_\lambda(\boldsymbol{d}) = \int_{-\infty}^{\infty}\int_{0}^{\infty} \prod f_\lambda(\bar{z} + sd_i)\sqrt{n}\, d\bar{z}\, s^{n-2} ds$$

The distribution for $\boldsymbol{z}$ can now be expressed as the marginal density for the observed $\boldsymbol{d}$ multiplied by the conditional density for the unobserved $(\bar{z}, s)$ given $\boldsymbol{d}$:

$$k_\lambda(\boldsymbol{d})\, da \;\; k_\lambda^{-1}(\boldsymbol{d}) \prod f_\lambda(\bar{z} + sd_i)\sqrt{n}\, d\bar{z}\, s^{n-2} ds$$

The quantity $k_\lambda(\boldsymbol{d})\, da$ gives the probability of the observed $\boldsymbol{d}$. It is not dependent on what μ and σ are.
As was mentioned earlier, the coordinates $(\bar{z}, s)$ on the half plane being unobserved cannot give information concerning λ . Indeed the observation of the half plane together with its model is the only information available concerning λ . We thus have a classical statistical model for $\boldsymbol{d}$

$$\begin{matrix} \boldsymbol{d} \\ k_\lambda(\boldsymbol{d})\, da \end{matrix}$$

All that is readily available is the observed likelihood function which is

$$L(\boldsymbol{d} \mid \lambda) = c\, k_\lambda(\boldsymbol{d})$$

This can be called the marginal likelihood for λ and it gives us the basis for inference concerning λ .

Based on the likelihood analysis one might consider several values of λ and consider the conditional densities for $(\bar{z}, s)$ for those values of λ :

$$g_\lambda(\bar{z}, s \mid \boldsymbol{d}) = k_\lambda^{-1}(\boldsymbol{d}) \prod f_\lambda(\bar{z} + sd_i)\sqrt{n}\, s^{n-2}$$

We have in a sense reduced our original statistical model to two models, the first being a classical model.

$$(1) \quad \begin{matrix} \boldsymbol{d} \\ k_\lambda(\boldsymbol{d})\, da \end{matrix}$$

$$(2) \quad \begin{matrix} \bar{y} = \mu + \sigma\bar{z} \\ s(\boldsymbol{y}) = \sigma s \\ g_\lambda(\bar{z}, s \mid \boldsymbol{d})\, d\bar{z}\, ds \end{matrix}$$

For the parameter μ , we can use the marginal distribution for t_z

$$t_z = \frac{\sqrt{n}\,\bar{z}}{s/\sqrt{n-1}} = \frac{\sqrt{n}(\bar{y} - \mu)}{s(\boldsymbol{y})/\sqrt{n-1}}$$

as a means for constructing tests or confidence intervals for μ .

For the parameter σ, we can use the marginal distribution for s_z

$$s_z = s(\boldsymbol{z})/\sqrt{n-1} = \frac{s(\boldsymbol{y})/\sqrt{n-1}}{\sigma}$$

as a means for constructing tests or confidence intervals for σ.

The remaining sections of this paper illustrate just how such procedures can be implemented; they indicate the possibilities and flexibility of the procedure.

Notice that our labelling of $\boldsymbol{z}$ by $(\bar{z}, s)$ was merely a convenient one. Our choice happened to lead to the familiar t and s statistics. We note first that the essential ingredients are not dependent on the labelling of $\boldsymbol{z}$. Rather the ingredients are an angular measure statistic which may be called t_z that gives the angle $\boldsymbol{z}$ makes with the $\mathbf{1}$ vector (for information concerning μ) and a distance measure statistic which may be labelled s_z that gives the distance $\boldsymbol{z}$ is from the $\mathbf{1}$ vector (for information concerning σ). Any bijective transformation on these fundamental statistics will lead to the same inferences. The distributions will differ and the linking equations with μ and σ will differ, but the conclusions will remain the same (see Figure (1)[1975]).

3. Implementation: The Distributions Involved

We are faced with finding a class of densities that describe $\boldsymbol{z}$. The vector $\boldsymbol{z}$ represents in a sense the identified variation in the system and the parameter λ expresses our uncertainty concerning the form of the density function for $\boldsymbol{z}$.

In an applied situation,the experimenter will have to judge what additional characteristics aside from location and scaling are present in his system. The shape and symmetries in his choice of family of densities should reflect his judgment of the additional characteristics. Normal theory analysis totally restricts his ability to allow for these additional characteristics in the analysis.

In some situations, a longer tailed distribution may be appropriate. If the family of error distributions allows for such a characteristic, then the realization of large z_i values are tolerated by the model. It will be consistent in our analysis if our procedure can accept values that are perhaps uncharacteristic for the normal distribution but are appropriate for longer tailed distributions. Such characteristics typically inject bias in a normal theory analysis.

A convenient family that allows for such a characteristic is: what shall be called the standardized Student family of densities.

$$f_\lambda(z) = \frac{\Gamma\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\sqrt{\pi}\,\omega(\lambda)}\left(1+\left(\frac{z}{\omega(\lambda)}\right)^2\right)^{-\frac{\lambda+1}{2}} \qquad \lambda \in (0, \infty]$$

The quantity $\omega(\lambda)$ standardizes the ordinary Student family so that

$$\int_{-1}^{1} f_\lambda(z)\,dz = 0.6826$$

Each member of the family then has median 0 and standard error 1 like the standard normal. Hence μ and σ represent the familiar median measure of location and standard error measure of scaling.

The family contains the standard normal($\lambda=\infty$) and also contains the scaled Cauchy ($\lambda=1$) and even sub-Cauchy ($\lambda<1$) densities. They are shown in Figure (2). For small values of λ , the densities allow with increasing probability for values more and more distant from the origin.

This is the family of densities used in this paper. Numerous other families can be used to exhibit other characteristics such as varying shape and asymmetry.

Families indexed by higher dimensional parameters have also been considered, but the assessment based on the marginal likelihood function becomes more involved.

The computer program mentioned earlier standardizes a given density with respect to the median and standard error before the density is used for the remainder of the calculation. It then uses numerical quadrature procedures to determine the distributions for the t and s statistic functions. Typically these distributions are dependent on the deviation vector $\boldsymbol{d}$ and must be recalculated for each set of data. This is not a drawback though. It indicates that the marginal densities for the usual t and s statistics are not adequate for a thorough understanding and analysis of a data set. And it agrees with the statement that any analysis should involve viewing the residuals and incorporating characteristics of the residuals in a proper analysis [see Tukey (1960)].

There are also examples of densities for which the integration is analytically accessible (for example see Fraser (1975)). A family of densities for which the integration could be explicitly handled would be very advantageous but the primary criteria for a family should be its ability to describe the characteristics present in the system under study.

An Example and Evaluation

4. Implementation:

We now illustrate with an example how to construct tests and confidence intervals using the analysis of section 2. The conclusions obtained will be compared with the classical method and also with the Bayesian method. Comments concerning robustness and other criteria will be made.

We consider a familiar data set as offered by Fisher (1960, p. 37) in connection with an analysis made by Darwin on the difference in heights of self- and cross-fertilized plants. The design of the experiment makes a paired comparison analysis appropriate since there were fifteen pairs of measurements, each pair giving the heights of a cross- and a self-fertilized plant grown in the same pot from the same seed. Accordingly we offer as our observations the differences $\boldsymbol{y}=(y_1\ ,\ \ldots\ ,\ y_{15})'$ in the heights of the two plants.

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

the mean response $\bar{y}$ = 20.933 and the standard deviation s_y = 37.744 .

The standardized deviation vector $\boldsymbol{d}$ which references $L^+(\mathbf{1},\boldsymbol{y})$ 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)' .

It has two somewhat large negative values (-0.6224 and -0.4881) and the largest positive deviation is . 3828 .

We begin by viewing the observed marginal likelihood function for this data. To do this, we consider a particular standardization to remove the arbitrary multiplicative constant. There is a very convenient and useful choice to be made. for $\lambda=\infty$ (normal(0,1)) the distribution of $\boldsymbol{d}$ is uniformly

distributed on the unit sphere in $L^{\perp}(\mathbf{1})$, in fact

$k_{\infty}(\boldsymbol{d})=\frac{1}{A_{n-1}}$ on the unit sphere in $L^{\perp}(\mathbf{1})$

= 0 otherwise,

where $A_{n-1}=2\pi^{(n-1)/2}/\Gamma((n-1)/2)$ is the surface area of this sphere. This result can be shown easily by substituting $\lambda=\infty$ in :

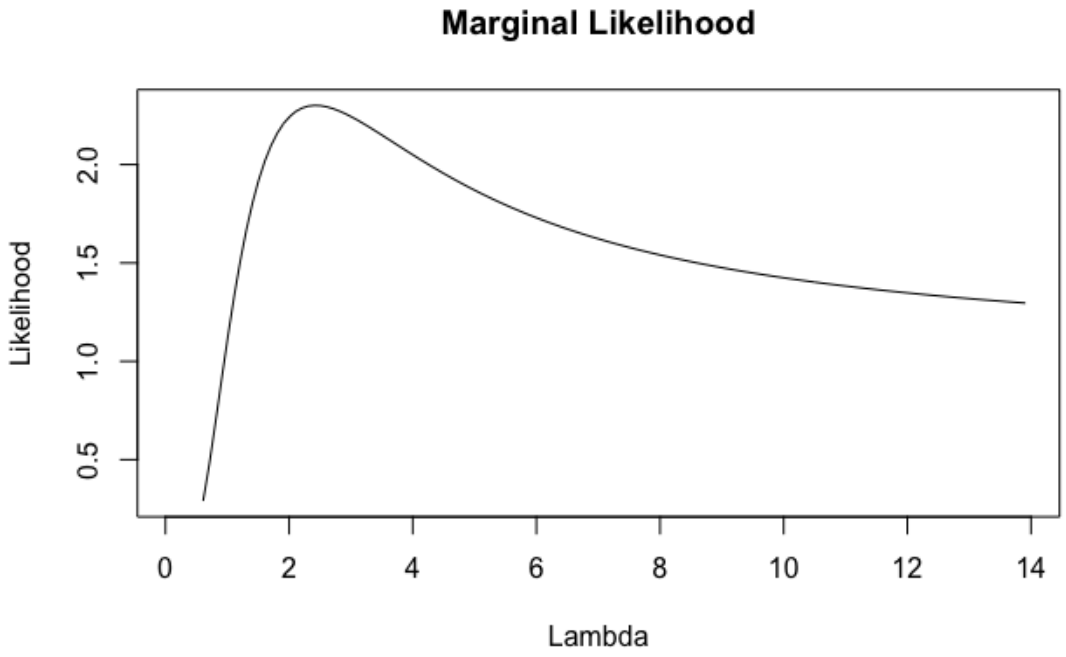
$k_{\lambda}(\boldsymbol{d})=\int_{-\infty}^{\infty}\int_{0}^{\infty}\prod f_{\lambda}(\bar{z}+sd_i)\sqrt{n}\,d\,\bar{z}\,s^{n-2}ds$ and recalling that $\sum d_i=0$ and $\sum d_i^2=1$.

For other values of λ , the distribution is not uniform. A realized $\boldsymbol{d}$ may have a much higher probability for certain values of λ than for $\lambda=\infty$. Of course the opposite can be true. The realized $\boldsymbol{d}$ may be in a region of the sphere of rather small density for certain values of λ . Since $\lambda=\infty$ assigned equal weight to each $\boldsymbol{d}$, it serves as an excellent choice for comparative studies with other λ values.

Accordingly we use the marginal likelihood function
$L(\boldsymbol{d}\,|\lambda)=A_{n-1}k_{\lambda}(\boldsymbol{d})$

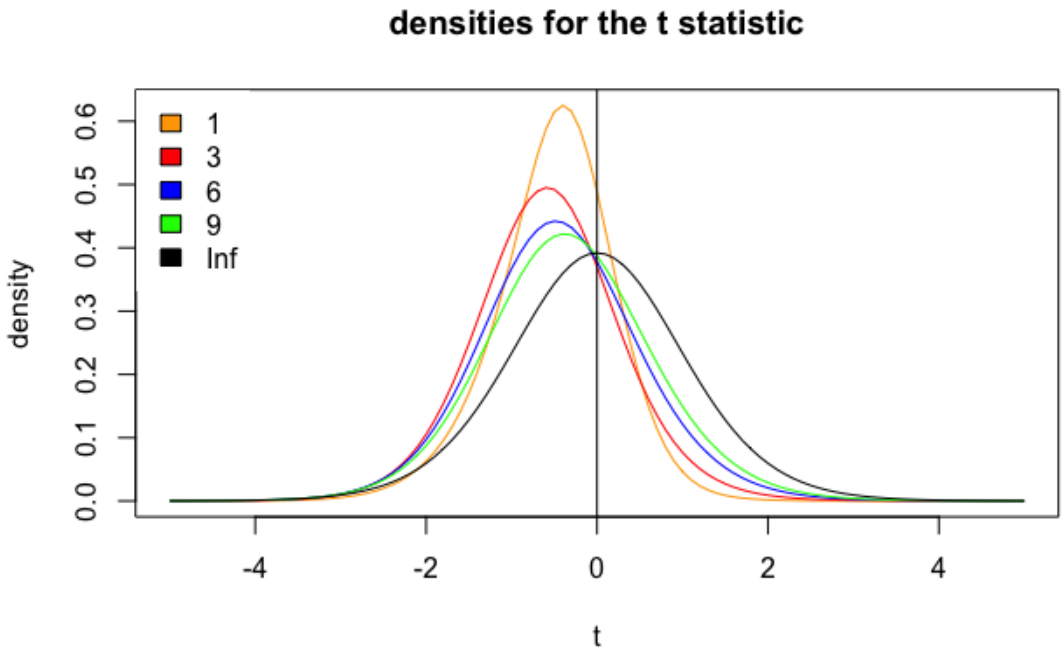
And we know that for all data sets $L(\boldsymbol{d}\,|\infty)=1$.

For the Darwin data the marginal likelihood has a mode of approximately 2.4 at λ = 2.3. It is plotted in Figure (3).



It is greater than 1.5 from λ = 1.2 to λ = 8.5. It therefore seems reasonable to consider λ values from 1 to 9 for the remainder of the analysis since the observed $\boldsymbol{d}$ has a probability that is more than 1.5 times higher for this range of values than for the normal case. For comparative purposes, we consider $\lambda=\infty$ also in the remainder of the analysis. Further comments concerning the observed likelihood are made in the next section.

For inference concerning μ , the conditional distributions for t_z are plotted in Figure (4) for λ =1,3,6,9 and ∞ .



A great deal can be learned from these plots. We note that they are all asymmetric except for $\lambda=\infty$ (which is a Student density on 14 degrees of freedom). The modes of these distributions shift off to the left for decreasing values of λ and become more tightly concentrated about the mode.

It is an expected phenomenon. The use of a Student model provides a tolerance for the extreme negative values in the data and compensates for the biases coming from these values by shifting to the left.

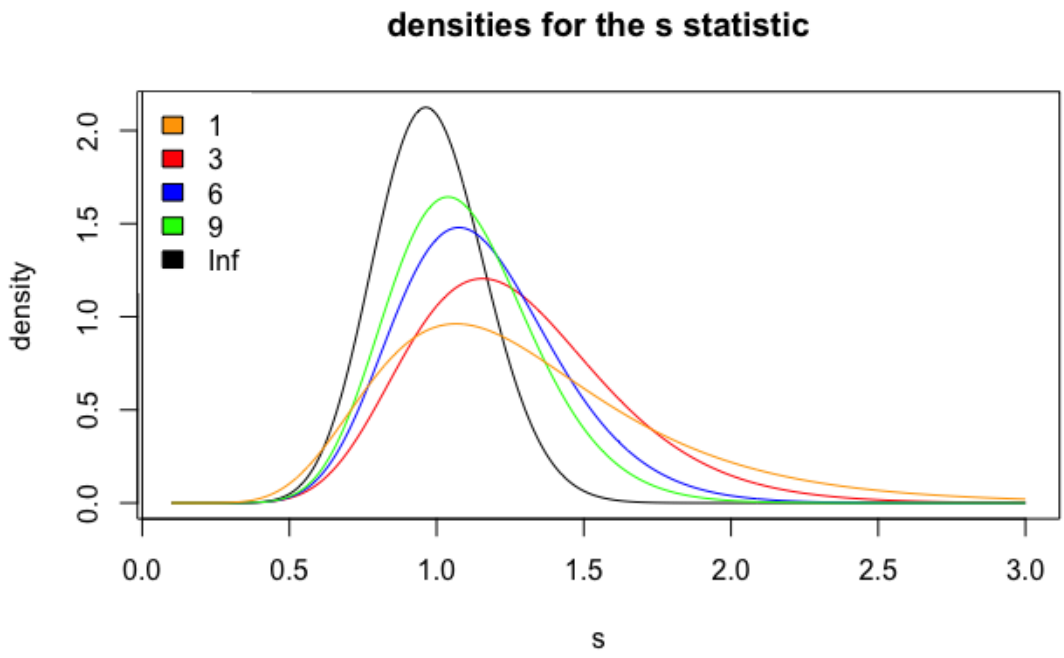
For inferences concerning μ , we offer $1-\alpha$ central probability intervals for t_z

$$P(t_{1,\lambda} < t_z < t_{2,\lambda} \mid \lambda) = 1-\alpha$$

We then obtain central $1-\alpha$ confidence intervals for μ

$$(\bar{y} - t_{2,\alpha} s_y/\sqrt{n},\ \bar{y} - t_{1,\alpha} s_y/\sqrt{n})$$

The distributions for s_z are plotted in Figure (5).



For inferences concerning σ ,. we offer $1-\alpha$ central probability intervals for s_z ,

$$P(s_{1,\lambda} < s_z < s_{2,\lambda} \mid \lambda) = 1-\alpha$$

and then obtain central $1-\alpha$ confidence intervals for σ

$$\left(\frac{s_y}{s_{2,\lambda}}, \frac{s_y}{s_{1,\lambda}}\right).$$

These intervals are obtained directly from the linking equations given in section 2.

Recall that the classical Neyman Pearson theory gives specific confidence intervals only when $\lambda=\infty$.

For the Darwin data, the probability and confidence intervals are in Tables 1 and 2 [1975].

Notice, say, for $\lambda=3$, that the confidence interval for μ is narrower and is centred about the median estimate μ = 26.7 . This contrasts with the normal theory interval centred at μ = 20.9

The estimate of σ is smaller for $\lambda=3$, than for $\lambda=\infty$ and the confidence intervals are also narrower.

For more information concerning these terminal stages of our analysis, the reader is invited to consult (Fraser (1975) and Fraser and Mackay (1975)).

Notice that the analysis based on the traditional model would be based on the observed likelihood function from the full response that is :

$$L(\boldsymbol{y}|(\mu,\sigma,\lambda)')=\frac{c}{(\sigma\omega(\lambda))^{15}}a^{15}(\lambda)\prod_{i=1}^{15}(1+\frac{(y_i-\mu)^2}{\sigma^2\omega^2(\lambda)})^{\frac{\lambda+1}{2}}$$

The viewing of this function might be contemplated along λ sections; or iterative procedures to obtain maximum likelihood estimates could be considered. Likelihood analysis with more than a one dimensional parameter faces many complications (see for example Edwards (1972)). This illustrates the importance of the choice of the statistical model for the analysis.

A Bayesian analysis of this data set is considered in Box and Tiao (1973). They offer noninformative priors for μ and σ and consider a collection of priors for a parameter β that indexes the exponential power family. This parameter β offers a measure of kurtosis as the additional characteristic under study. Conditional posterior distributions for μ and σ dependent on β are offered along with a marginal posterior for β once has one has chosen one of the prior distributions.

Apart from its justifications, this Bayesian procedure does not offer the clearly defined conclusions that are attainable in the present paper. Indeed, one is left with doubt as to how to interpret the wide class of conditional posteriors and even more so with how to justify particular marginal posteriors.

Notice that our necessary analysis together with its terminal analysis need not be concerned with the usual forms of robustness studies; for example, robustness relative to the statistic used or the results obtained. We have selected a broadly based model, and, from the subsequent analysis, we can determine whether strong conclusions are available and how the conclusions depend on the distribution form.

5. Implementation: A Numerical Study

The computer program mentioned earlier has been used with a large number of data collections. A serum study is analyzed in Fraser (1975) together with samples of computer generated data.

The program is now in a form that easily is used by students and researchers. It outputs tables similar to those of Tables 1 and 2. And gives accurate plots of the marginal likelihood function and of the conditional distributions for the t and s statistics via the [1975] {Calcomp or Gould offline systems}.

It can handle any family of error distributions indexed by a single parameter.

Programs to handle general linear models are now being developed. The analysis needed for the linear model very closely parallels the location scale analysis and is outlined in Fraser (1975 or 1968).

One study that has been made consisted of the generation of approximately 20 samples of 30 from the Student distribution on 6 degrees of freedom and on other low degrees of freedom.

From these examples there are four different cases that arise.

1) The likelihood function can be sharply discriminating for particular values of λ and the conditional distributions vary considerably.

2) The conditional distributions may be very similar in spite of discriminating likelihood.

3) The likelihood function may not be sharply discriminating and the conditional densities may not be very different.

4) The likelihood function is not discriminating and yet the conditional distributions do vary substantially.

These four cases can be traced to where the observed $\boldsymbol{d}$ is on the unit sphere in $L^{\perp}(\mathbf{1})$. The form of the marginal likelihood function is determined by where the realized $\boldsymbol{d}$ is in relation to its distribution $k_{\lambda}(\boldsymbol{d})$.

Clearly, an understanding of the model for the observed marginal likelihood functions would be an aid in the interpretation of the results.
The only unsatisfactory situation appears to be with case 4). This case will require cautious conclusions and perhaps the need for reevaluation. It may well reflect that the chosen parameter λ is not informative.

The location scale program has been developed by the authors and is available to interested users at the following address:

[in 1975 : Fortran] Statistics Section (Att: Gordon Fick)
Department of Mathematics
University of Toronto
Toronto, Canada
M5S 1Al
[in 2020 : R] Contact Gordon Fick at ghfick@ucalgary.ca

The motivation and analysis of this procedure have strong and valid foundations. But the most important point at the moment is its effective implementation. The reader is invited to consider the procedure and try the program.

References

Box, G.E.P. and Tiao, G.C. (1973). Bayesian Inference In Statistical Analysis, Reading, Massachusetts: Addison-Wesley.

Edwards, A.W.F. (1972). Likelihood, Cambridge: Cambridge Press.

Fisher, R.A. (1960). The Design of Experiments, seventh edition, New York: Hafner.

Fraser, D.A.S. (1968). The Structure of Inference, New York: Krieger.

Fraser D.A.S. (1975). Necessary Analysis and Adaptive Inference, JASA,

Fraser D.A.S. and MacKay, J. (1975). Parameter factorization and inference based on significance, likelihood, and objective posterior, Annals of Statistics.

Tukey, J.W. (1960). A survey of sampling from contaminated Distributions, in I. Olkin, ed., Contributions to Probability and Statistics, Stanford: Stanford University Press.