CANADIAN THESES

THÈSES CANADIENNES

Robust Confidence Intervals

By

Maureen Tingley

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
at
Dalhousie University,
Halifax, Nova Scotia.
July, 1986.

ISBN 0-315-33087-2

Table of Contents

Abstract

An independent sample $X_1, \ldots, X_n$ is drawn from a population with distribution involving an unknown multidimensional parameter $\underline{\eta}$. We introduce a procedure for constructing confidence intervals for a real-valued function $\theta = \theta(\underline{\eta})$. The intervals have error $O_P(\frac{1}{n})$ for $\underline{\eta}$ in a compact subset of the parameter space.

The unknown parameter is estimated by an M-estimate $\hat{\underline{\eta}}$, and we work with a modified sample, constructed from the observed values of the score functions for the M-estimation. The modified sample is called the configuration. The confidence interval is constructed as all parameter values θ_0 not rejected by a specified test procedure. The test is based on the observed average of the configuration.

The configuration is constructed by applying results found in Bhattacharya and Ghosh (1978) and Field (1982), and is calculated once for each sample. For each value θ_0 under test, the distribution of the configuration is re-estimated by an exponentially tilted bootstrap. Bootstrap re-sampling is avoided via the Lugannani and Rice tail area approximation (1980).

We apply the technique to two problems: confidence intervals for location when scale is unknown, and confidence intervals for slope in simple linear regression.

Acknowledgements

I wish to thank Chris Field for his help over the past three years. Chris has always encouraged me to ask questions and to look into any question that interests me. Thanks, Chris, for all the times you smiled and said "It's your thesis." I trust that we can continue to find interesting problems (and, hopefully, solve a few of them).

Several members of the department have answered my questions, loaned me their books, and encouraged me in this project. Thanks to all of you. I have enjoyed working in a department where one can expect a friendly "hello" from everybody.

And thanks to Yvonne Gaudelius, who sacrificed a sunny Saturday in July when there was typing to be done.

Chapter 1 Introduction

Our primary concern is the construction of confidence intervals for a real-valued population parameter in the presence of nuisance parameters. Our two examples are location, with scale unknown, and the regression coefficient in simple linear regression.

Our techniques can be applied to either a parametric or a non-parametric model; however, our programs are non-parametric.

This chapter contains background information. Our work has been influenced by ideas of Efron, and these ideas are discussed in Sections 1 and 2. The exponential tilt discussed in Section 3 is an important step in our technique. In Section 4 we introduce the cumulant generating function, and state a result of Daniels. The classical approach to the problem is stated in Section 6, along with modifications suggested by the work of Huber, Kent, and Ronchetti.

1.1 The Bootstrap

Suppose we have an independent identically distributed sample $X_1, \ldots X_n$, taken from an unknown distribution F on the real line, and the observed value is $\underline{X} = \underline{x}$. Let $T = T(\underline{X})$ be a real-valued function of $\underline{X}$, and suppose we wish to estimate the significance level of $T_{\text{obs}} = T(\underline{x})$; that is, the probability under F of observing T more extreme than T_{obs}. Then we need to know something about the distribution of T. Efron has introduced the bootstrap to handle situations where, for one reason or another, it is difficult to use theoretical arguments to estimate the distribution of T. The bootstrap is a non-parametric technique which operates as follows.

Let $\hat{F}$ be the empirical probability distribution of the data, putting mass $1/n$ on each x_i. Take a large number of samples of size n, with replacement, from the observed values $x_1, \ldots, x_n$. For each new sample $X_1^*, \ldots X_n^*$, calculate $T^* = T(\underline{X}^*)$. Then the distribution of T can be estimated by the empirical distribution of T^*. For example, $P(T > t_0)$ is estimated by

$$P_B(T^* > t_0) = \frac{\#\text{ of bootstrap samples for which } T^* > t_0}{\#\text{ of bootstrap samples taken}}.$$

Usually, the number of bootstrap samples taken is large, at least 100, so the bootstrap has become feasible in an era when computers have made massive computations fast and inexpensive.

Efron also defines a parametric bootstrap. Suppose we assume a parametric model with distribution function $F(x, \underline{\theta})$, and that the parameter $\underline{\theta}$ is estimated from the sample by $\hat{\underline{\theta}}$. Then, rather than take a large number of Monte Carlo samples, one could possibly use theoretical methods to estimate the distribution of T under $F(x, \hat{\underline{\theta}})$, or under a parametric model where some parameters are estimated and others are assumed known. Note, in particular, that a parametric bootstrap involves no Monte Carlo resampling.

Suppose we wish to test the null hypothesis $H_0: \mu = \mu_0$, where $\mu = \mu(F)$ is a population parameter for the distribution F. Suppose a sample $\underline{X}$, size n, is taken with observed value $\underline{x}$, giving the empirical distribution $\hat{F}$, with weight $1/n$ at each x_i. In order to test H_0, we use a test statistic, call it $T = T(\underline{X})$, and estimate the probability, under H_0, of observing T more extreme than $T_{\text{obs}} = T(\underline{x})$. We will rarely have $\mu(\hat{F}) = \mu_0$, so that bootstrap samples taken from $\hat{F}$ will not give significance levels under H_0. Efron suggests that the bootstrap samples be taken not from $\hat{F}$, but from a modified distribution $\hat{F}_\alpha$, satisfying $\mu(\hat{F}_\alpha) = \mu_0$. The distribution $\hat{F}_\alpha$ is chosen to be that closest to $\hat{F}$ in Kullback-Leibler distance, while satisfying the null hypothesis. The distribution $\hat{F}_\alpha$ puts weight

$$w_i^\alpha = e^{\alpha x_i} \Big/ \sum_{i=1}^{n} e^{\alpha x_i}$$

on the observed value x_i, $i = 1, \ldots, n$, where α is chosen so that $\mu(\hat{F}_\alpha) = \mu_0$. Then the bootstrap is applied as usual, with the Monte Carlo samples taken from $\hat{F}_\alpha$ rather than from $\hat{F}$. This technique is called the exponentially tilted bootstrap.

Efron's concern is primarily with confidence intervals, calculated by inverting appropriately chosen pivotal quantities. We calculate confidence intervals by running a series of tests, a process which Efron has steadfastly tried to avoid. On the other hand, our tests avoid bootstrap resampling, even for a non-parametric model.

1.2 Bootstrap confidence intervals

If the statistic T estimates a parameter θ, Efron suggests we assume that $T-\theta$ has a distribution independent of θ, and that

$$P_\theta(T-\theta > t_0) \doteq P_B(T^\circ - T_{obs} > t_0), \tag{1.2.1}$$

an approach which leads to the (simple) Bootstrap pivotal confidence interval. Tibshirani (1984) presents various refinements of the pivotal interval. These all keep assumption (1.2.1). The Normal bias corrected percentile interval, for example, is constructed by assuming that there is a transformation g such that $g(T) - g(\theta)$ is normal. The beauty of these intervals is that the transformation g need not be known. All we need is a good estimate of the center of the distribution of $g(T) - g(\theta)$.

In Efron's work we see two recurring ideas: first, that there is a transformation to normality and, second, that a confidence interval for a statistic can be inverted to give a confidence interval for the parameter of interest (that is, there is a pivot).

For the moment, we consider a one-parameter parametric model. Efron (1986) constructs his BC_a intervals as follows. The parameter of interest is $\theta = t(F)$, for which there is an estimate $\hat\theta$, and $\hat\theta$ behaves asymptotically like a maximum likelihood estimate in terms of the orders of magnitude of its bias, standard deviation, skewness and kurtosis. A parametric bootstrap estimate, $\hat H$, of the distribution of $\hat\theta$ is constructed. Assume there is a monotone transformation g such that, to a good approximation, $\hat\phi = g(\hat\theta)$ and $\phi = g(\theta)$ satisfy

$$\begin{aligned} \hat\phi &= \phi + \sigma_\phi(Z - z_0) \qquad Z \sim N(0,1) \\ \sigma_\phi &= 1 + a\phi, \end{aligned} \tag{1.2.2}$$

where $z_0 = \Phi^{-1}(\hat H(\hat\theta))$. Then the BC_a $(1-2\varepsilon)100\%$ confidence interval for θ is

$$\theta \in (\hat H^{-1}(\Phi(z[\varepsilon])),\ \hat H^{-1}(\Phi(z[1-\varepsilon]))), \tag{1.2.3}$$

where

$$z[\beta] = z_0 + \frac{z_0 + z^\beta}{1 - a(z_0 + z^\beta)},$$

z^β is defined by $\beta = \hat{\Phi}(z^\beta)$, and $\hat{\Phi}$ is the distribution function of Z. Efron shows that this interval is second order correct for the one parameter parametric model; that is, the interval has error $O_P(\frac{1}{n})$.

The transformation g need not be known. However, the constant a must be estimated. A good approximation to a is

$$a \doteq \frac{1}{6}\text{skew}_{\theta=\hat{\theta}}\dot{l}_\theta, \tag{1.2.4}$$

where $\text{skew}_{\theta=\hat{\theta}}X$ indicates the skewness of a random variable X, evaluated at the parameter point $\theta = \hat{\theta}$, and $\dot{l}_\theta$ is the score function

$$\dot{l}_\theta(\hat{\theta}) = \frac{\partial}{\partial\theta}\log f_\theta(\hat{\theta}), \tag{1.2.5}$$

assuming $\hat{\theta}$ has a density, f_θ.

The BC_a interval is obtained by refining the bootstrap assumption (1.2.1). There is a bias correction, via z_0, and a skewness correction, via a. The transformation $\theta \to \phi$, $\hat{\theta} \to \hat{\phi}$ leads to a further transformation $\phi \to \xi$, $\hat{\phi} \to \hat{\xi}$, which reduces the model to a standard translation problem:

$$\hat{\xi} = \xi + W,$$

where W has a distribution independent of ξ. There is then a natural central $(1-2c)100\%$ confidence interval for ξ having observed $\hat{\xi}$:

$$\xi \in [\hat{\xi} - w^{(1-c)}, \hat{\xi} - w^{(c)}],$$

where $P(W < w^{(\beta)}) = \beta$. This interval is then transformed back to give the interval of (1.2.3).

As mentioned above, Efron (1986) has established the second-order correctness of the BC_a intervals only for the one-parameter parametric model. However, he indicates how to extend the BC_a interval to multidimensional parameters and to non-parametric models.

In the parametric model, with nuisance parameters, Efron estimates a by following Stein's construction of a least favourable one-parameter family. It is this construction that interests us, and we now present it.

Let the parametric model have density $f(y;\underline{\eta})$, let $\hat{\underline{\eta}}$, be the maximum likelihood estimate of $\underline{\eta}$, and suppose $\theta = t(\underline{\eta})$ is the parameter of interest. Let

$$\dot{\underline{l}}(\underline{\eta}) = \frac{\partial \log f(\underline{y},\underline{\eta})}{\partial \underline{\eta}}$$

(1.2.6)
$$\ddot{l}(\underline{\eta}) = \frac{\partial \dot{\underline{l}}(\underline{\eta})}{\partial \underline{\eta}^T}$$

$$\hat{\underline{\mu}} = (-\ddot{l}(\hat{\underline{\eta}}))^{-1} \frac{\partial \theta}{\partial \underline{\eta}}(\hat{\underline{\eta}}).$$

Then the least favourable one-parameter sub-family passing through $\hat{\underline{\eta}}$ is $f(\underline{y};\ \hat{\underline{\eta}} + \alpha\hat{\underline{\mu}})$, and the maximum likelihood estimate of α is $\alpha = 0$. Efron estimates a by

(1.2.7)
$$a \doteq \frac{1}{6}\text{skew}_{\alpha=0} \frac{\partial \log f(y;\hat{\underline{\eta}} + \alpha\hat{\underline{\mu}})}{\partial \alpha}.$$

In the simple exponential family case, we can write

$$f(\underline{y};\ \underline{\eta}) = e^{\underline{y}^T\underline{\eta} - \psi(\underline{\eta})} f(\underline{y};\ \underline{0})$$

and

$$f(\underline{y};\ \underline{\eta} + \alpha\hat{\underline{\mu}}) = C(\alpha)e^{\alpha\underline{y}^T\hat{\underline{\mu}}} f(\underline{y};\ \underline{\eta}).$$

The nuisance parameters are absorbed in the new "underlying measure" $f(\underline{y};\ \underline{\eta})d\underline{y}$. With respect to this measure, $f(\underline{y};\ \underline{\eta} + \alpha\underline{\mu})$ is a one-parameter exponential family, parameter α. For α_0 a fixed value of α, we call the density $C(\alpha_0)\exp(\alpha_0\underline{y}^T\hat{\underline{\mu}})f(\underline{y};\ \hat{\underline{\eta}})$

an exponential tilt of the density $f(y;\ \eta)$. Least favourable distributions are discussed by DiCiccio and Tibshirani (1986). The next section discusses "optimal" exponential tilts.

1.3 The exponential tilt

Suppose we have a multiparameter model with density $f(x, \underline{\eta})$ and we wish to test $H_0\colon \theta = \theta_0$, where $\theta = \theta(\underline{\eta})$ is a real-valued function of the parameter vector $\underline{\eta}$. Suppose further that, for the given sample $\underline{X} = \underline{x}$ we estimate $\underline{\eta}$ by $\hat{\underline{\eta}}$, with $\theta(\hat{\underline{\eta}}) \neq \theta_0$, and the test statistic is $T_{\text{obs}} = T(\underline{x})$. Then it is not appropriate to test the significance of T_{obs} under $f(x;\ \hat{\underline{\eta}})$, since $f(x;\ \hat{\underline{\eta}})$ does not satisfy H_0. The exponential tilt is a simple device used to circumvent this problem.

If the random variable X has a density $f(x)$, α is real, and $\int e^{\alpha x} f(x)dx < \infty$, then the conjugate or exponentially-tilted density determined by α is

$$f_\alpha(x) = \frac{e^{\alpha x} f(x)}{\int e^{\alpha x} f(x)dx}. \tag{1.3.1}$$

The definition can be extended to probability measures not having a density function, but we postpone the more general definition until the next section.

The exponential tilt has been used by Daniels (1954) and Barndorff-Nielsen and Cox (1979). In particular, Field (1982) estimates the density of an M-estimate $\underline{T}$ by applying an exponential tilt at each possible value $\underline{T} = \underline{t}_0$. This paper is motivated, in part, by the idea that, to test any given null hypothesis, just one well-chosen tilt should "do the trick".

The primary justification for an exponential tilt is Kullback's result (1960) which we now state. Suppose that $f_1(x)$ and $f_2(x)$ are generalized densities of a dominated set of probability measures on the measurable space $(\mathcal{X}, \mathcal{S}, \lambda)$, so that

$$\mu_i(E) = \int_E f_i(x)d\lambda(x), \quad E \in \mathcal{S},\ i = 1, 2. \tag{1.3.2}$$

For a given $f_2(x)$ we seek, subject to a constraint, that member f_1 of the dominated set of probability measures that is "nearest" to or most closely resembles the probability measure μ_2 in the sense of smallest Kullback-Leibler "distance",

$$I(1:2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} d\lambda(x). \tag{1.3.3}$$

If f_1 is close to f_2 in Kullback-Leibler distance then, in general, it is difficult to distinguish between $H_1: f_1$ and $H_2: f_2$ when inference is based on a sample taken from f_2.

Theorem 1.3.1 (Kullback (1960), page 38) *If $f_1(x)$ and a given $f_2(x)$ are generalized densities of a dominated set of probability measures, $Y = T(X)$ is a measurable statistic such that $\theta = \int T(x) f_1(x)\, d\lambda(x)$ exists, and $M_2(\alpha) = \int f_2(x) e^{\alpha T(x)}\, d\lambda(x)$ exists for α in some interval; then*

$$I(1:2) \geq \theta\alpha - \log M_2(\alpha) = I(*:2), \quad \theta = \frac{d}{d\alpha} \log M_2(\alpha), \tag{1.3.4}$$

with equality in (1.3.4) if and only if

$$f_1(x) = f^*(x) = e^{\alpha T(x)} f_2(x)/M_2(\alpha)[\lambda].$$

In the above theorem, $[\lambda]$ means "except on a set of λ-measure zero". This result is restated in the next section, in terms of the cumulant generating function.

In the classical problem of testing for the mean of a univariate normal distribution, $H_0: \mu = \mu_0$, the test statistic $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ has a Student t_{n-1} distribution, and H_0 is rejected if the probability that t_{n-1} is greater than the observed value $\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ is too large or too small. In this case, the T statistic is a pivot and the interval

$$\text{lower acceptance point } t_{n-1} < \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < \text{ upper acceptance point } t_{n-1}$$

can be inverted to give a confidence interval for μ: the set of all possible μ_0 which, for this set of data, would not be rejected at the specified confidence level,

$$\bar{x} - \frac{s}{\sqrt{n}}(\text{upper point } t_{n-1}) < \mu < \bar{x} - \frac{s}{\sqrt{n}} \ (\text{lower point } t_{n-1}).$$

In the sequel, we do not have pivots. We wish to test the null hypothesis $\theta = \theta_0$ where $\theta(\underline{\eta})$ is a one-dimensional function of a parameter vector $\underline{\eta}$. We construct a test statistic $\bar{g} = \frac{1}{n}\sum g_i$ (see Chapter 3), and test, under θ_0, to see whether or not the observed value of $\bar{g}$ is too extreme. The distribution of the g_i is estimated from the sample. This initial estimate, call it $\hat{F}$, gives a distribution for the g_i which does not, in general, satisfy $H_0{:}\,\theta = \theta_0$. We "upgrade" the estimate $\hat{F}$ by applying an exponential tilt. The hypothesis $\theta = \theta_0$ holds under the new distribution, call it $\hat{F}_{\alpha_0}$, and it is under this distribution that we calculate $P(\bar{g} > \bar{g}_{obs})$.

To compare with the classical case, $\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ is t_{n-1} because $\bar{X}$ is normal with mean μ_0. Imagine a situation where each null hypothesis gives a different distribution for the test statistic. This unknown distribution must be estimated *and* it must be estimated in such a way that H_0 holds, since H_0 is under test. Kullback's theorem says that the exponential tilt forces H_0, while altering $\hat{F}$ "as little as possible".

Whether or not Kullback-Leibler distance is the best measure of closeness is an issue we do not discuss in this paper. However, it is clear from the work of Daniels and Field that the consequent exponential tilt produces the desired effect.

Note that the exponentially tilted bootstrap *is* an exponential tilt of the empirical distribution $\hat{F}$.

1.4 The cumulant generating function

For our purposes, the exponential tilt has the advantage that it is applied more readily by considering the cumulant generating function, than by considering the generalized density or the distribution function directly. To approximate tail areas, we use a result (Lugannani-Rice, 1980) which is also most readily described in terms of the cumulant generating function.

Let X be a random variable with distribution F. Then the moment generating function of F (or of X) is the function M, defined for real t by

$$M(t) = E(e^{tX}).$$

The characteristic function $M(it)$ exists for all X and all real t. If F has a density f, then $M(it)$ is the Fourier transform of f. The moment generating function does not exist for all t and X.

Daniels defines the cumulant generating function K by

$$M(t) = e^{K(t)}, \qquad t \text{ real}.$$

(This is *not* the same as Cramér's and Feller's definition of the cumulant generating function as $K(it)$, which is well-defined for all X and all real t.)

We use two facts about characteristic functions (see Feller (1971)). If $E(X)$ exists, then

$$M'(0) = E(X)$$

$$K'(0) = E(X).$$

Distinct probability distributions have distinct characteristic functions.

As mentioned above, there are existence problems with the cumulant generating function. These problems disappear when the random variable X is bounded, as will be the case in our applications. We need the following results from Daniels (1954, Theorems 6.1 and 6.2). In both theorems, a and b are real, $-\infty < a < b < \infty$.

Theorem 1.4.1 (Daniels) *$F(x) = 0$ for $x < a$, and $F(x) = 1$ for $x > b$ if and only if $K(t)$ exists for all real t and $K'(t) = \xi$ has no real root whenever $\xi < a$ or $\xi > b$.*

Theorem 1.4.2 (Daniels) *Let $F(x) = 0$ for $x < a$, $0 < F(x) < 1$ for $a < x < b$, $F(x) = 1$ for $b < x$. Then for every ξ in $a < \xi < b$ there is a unique root t_0 of $K'(t) = \xi$. As t increases from $-\infty$ to ∞, $K'(t)$ increases continuously from $\xi = a$ to $\xi = b$.*

The Kullback theorem of section 1.3 can be stated in terms of the cumulant generating function.

Theorem 1.4.3 (Kullback) *Let X be a random variable and F_1, F_2 distribution functions defined on the range of X. Let $Y = T(X)$ be a measurable statistic with cumulant generating function defined by*

$$e^{K_1(t)} = E(e^{tY}), \qquad \text{under } F_1$$

$$e^{K_2(t)} = E(e^{tY}), \qquad \text{under } F_2.$$

Suppose further that $K_1'(0)$ exists, $K_1'(0) = \theta$, and that $K_2(t)$ exists for t in some interval. Then

$$I(1:2) \geq \theta\alpha_0 - K_2(\alpha_0) = I(*:2),$$

where $t = \alpha_0$ solves $K_2'(t) = \theta$, with equality if and only if

$$K_1(t) = K_2(t + \alpha_0) - K_2(\alpha_0).$$

Note in particular that F_1 is well-defined via its cumulant generating function: $K_1(t) = K_2(t + \alpha_0) - K_2(\alpha_0)$. Note also that $K_1'(0) = K_2'(\alpha_0) = \theta$, and F_1 is that distribution closest in Kullback-Leibler distance to F_2, while satisfying $E_{F_1}(Y) = \theta$.

1.5 Effect of the exponential tilt on cumulants

Efron's transformation to normality for the BC_a intervals is given in Section 1.2. In particular, $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$ with

$$\sigma_\phi = 1 + a\phi, \qquad a = \frac{1}{6}\text{skew}\frac{\partial}{\partial\theta}\log f_\theta(\hat{\theta}). \tag{1.5.1}$$

We now demonstrate that the exponential tilt produces a comparable effect.

Suppose the random variable Y is bounded and therefore has finite cumulants. We expand $K(t)$ in a Taylor series,

$$K(t) = \kappa_1 t + \kappa_2\frac{t^2}{2!} + \kappa_3\frac{t^3}{3!} + \dots, \tag{1.5.2}$$

where the κ_j are the cumulants of Y. In particular, κ_1 is the mean of Y, κ_2 is the variance of Y, and $\gamma_1 = \kappa_3/\kappa_2^{3/2}$ is the skewness. Let $t = \alpha_0$ be a solution of $K'(t) = \mu_0$. Then α_0 is unique, by Theorem 1.4.2, and the α_0-tilted distribution has cumulant generating function

$$K_{\alpha_0}(t) = K'(\alpha_0)t + K''(\alpha_0)\frac{t^2}{2!} + K^{(3)}(\alpha_0)\frac{t^3}{3!} + \dots . \tag{1.5.3}$$

The first cumulant of the tilted distribution is $\kappa_{\alpha_0 1} = \mu_0$, as it was forced to be. However, we can also write

$$\begin{aligned} \kappa_{\alpha_0 1} &= K'(\alpha_0) \\ &= \kappa_1 + \alpha_0\kappa_2 + O(\alpha_0^2) \\ &\doteq \kappa_1 + \alpha_0\kappa_2, \end{aligned}$$

provided α_0 is "small". In fact, for each $j \geq 1$,

$$\kappa_{\alpha_0 j} \doteq \kappa_j + \alpha_0\kappa_{j+1}. \tag{1.5.4}$$

We should think of an exponential tilt as a gentle "distortion" of the underlying distribution. The distortion forces the first cumulant to assume a specified value and each cumulant is altered by a (small) multiple of the next. (Discussion of the smallness of α_0 is delayed until the application.)

Since $\kappa_1 + \alpha_0\kappa_2 \doteq \mu_0$, we can make the approximation

$$\alpha_0 \doteq \frac{\mu_0 - \kappa_1}{\kappa_2}.$$

Then, taking $j = 2$ in (1.5.4),

$$\begin{aligned}\kappa_{\alpha_0 2} &\doteq \kappa_2 + \alpha_0\kappa_3 \\ &\doteq \kappa_2 + \frac{\kappa_3}{\kappa_2}(\mu_0 - \kappa_1).\end{aligned}$$

The standard deviation for the tilted distribution is then estimated using Taylor's Theorem:

$$\begin{aligned}\sigma_{\alpha_0} &\doteq \sqrt{\kappa_2} + \frac{\kappa_3(\mu_0 - \kappa_1)}{2\kappa_2^{3/2}} \\ &= \sqrt{\kappa_2} + \frac{1}{2}\gamma_1(\mu_0 - \kappa_1).\end{aligned}$$

Thus, both Efron's transformation to normality and the exponential tilt alter standard deviation by a multiple of skewness. The two approaches are, however, very different. The BC_a model assumes that the random variable of interest can be *transformed* to give a better behaved random variable. In our applications, we take an estimate of the distribution of a random variable and then *improve* that estimate by applying an exponential tilt.

1.6 Variations on the classical approach to testing

Consider, for simplicity, a model with one-dimensional parameter θ. In the classical theory, we assume that the independent identically distributed sample $X_1, \ldots, X_n$ comes from a population with generalized density $f(x, \theta)$. The maximum likelihood estimate $\hat{\theta}$ is the solution of

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(x_i, \theta) = 0, \tag{1.6.1}$$

and $\hat{\theta}$ has a limiting normal distribution, provided f satisfies certain regularity conditions. The proof that $\hat{\theta}$ is asymptotically normal proceeds by taking a Taylor series expansion for the score function, $U(x, \theta) = \frac{\partial}{\partial \theta} \log f(x, \theta)$, about the value θ:

$$\begin{aligned} 0 &= \sum_{i=1}^{n} U(x_i, \hat{\theta}) \\ &= \sum_{i=1}^{n} U(x_i, \theta) + \sum_{i=1}^{n} \frac{\partial}{\partial \theta} U(x_i, \theta)(\hat{\theta} - \theta) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} U(x_i, \theta^*)(\hat{\theta} - \theta)^2, \end{aligned} \tag{1.6.2}$$

θ^* between $\hat{\theta}$ and θ. Then, consistency of $\hat{\theta}$ and the Weak Law of Large Numbers are used to write

$$i(\theta)(\hat{\theta} - \theta)\sqrt{n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U(x_i, \theta) + o_P(1), \tag{1.6.3}$$

where $i(\theta) = E_\theta(-\frac{\partial U}{\partial \theta}(X, \theta)) = E_\theta(U^2(X, \theta))$. The Central Limit Theorem then gives asymptotic normality for the right hand side of (1.6.3), and hence $\hat{\theta}$ is asymptotically normal.

Equation (1.6.3) also gives

$$E(\hat{\theta}) = \theta + o_P(\tfrac{1}{\sqrt{n}}).$$

The bias can be corrected by careful consideration of equation (1.6.2), see Cox and Hinkley (1974), pages 309, 310, giving a correction $b(\hat{\theta})$ so that

$$E(\hat{\theta} + b(\hat{\theta})) = \theta + o_P(\tfrac{1}{n}). \tag{1.6.4}$$

The techniques we use in Chapter 3 incorporate similar ideas, giving a bias-corrected M-estimate.

The classical result can be generalized. For example, Huber (1967) gives two sets of conditions sufficient to guarantee the (weak) consistency of an M-estimate, then gives conditions which imply that a consistent M-estimate is asymptotically normal.

Classical tests of hypotheses are based on the log likelihood ratio. Asymptotic normality of the maximum likelihood estimate is used to prove that the log likelihood ratio has an asymptotic χ^2 distribution. This result, too, can be generalized.

We now discuss Kent's (1982) generalization. Consider a model with distribution $F(x, \underline{\eta})$ determined by a family of densities $\{f(x, \underline{\eta});\ \underline{\eta} \in \mathcal{N}\}$ and a sample $\underline{x}$ drawn from a population with density $g(x)$, not a member of the parametric family, but for which the parameter $\underline{\eta}(g)$ is well defined. Then an analogue of the maximum likelihood estimate for $\underline{\eta}$ is $\hat{\underline{\eta}}$, solution of

$$\sum_{i=1}^{n} \int \log\{f(x_i, \underline{\eta})\} g(x)\ dx = \max!\ .$$

Define the score function by

$$\underline{U}(x, \underline{\eta}) = \frac{\partial}{\partial \underline{\eta}} \log f(x, \underline{\eta})$$

and define

$$C = C(\underline{\eta}) = \int \underline{U}(x, \underline{\eta}) \underline{U}(x, \underline{\eta})^T g(x) \cdot dx$$

$$A = A(\underline{\eta}) = \int \frac{\partial \underline{U}(x, \underline{\eta})}{\partial \underline{\eta}^T} g(x)\ dx.$$

$$B = -A^{-1}.$$

Then $\hat{\underline{\eta}}$ is the solution of

$$\sum_{i=1}^{n} \underline{U}(x_i, \underline{\eta}) = \underline{0}$$

but, unlike the classical theory, $C \neq -A$. Kent shows that the estimate $\underline{\eta}$ is asymptotically normal

$$\sqrt{n}(\hat{\underline{\eta}} - \underline{\eta}) \sim N_m(\underline{0}, BCB), \quad m = \dim \underline{\eta},$$

a result analogous to Huber's (1967) result for M-estimates.

For $\underline{\eta}^T = (\underline{\theta}^T, \underline{\lambda}^T)$, Kent recommends that we test $H_0 \colon \underline{\theta} = \underline{\theta}_0$ as follows. Write

$$\begin{aligned} \underline{U}(x, \underline{\eta}) &= \begin{pmatrix} \frac{\partial}{\partial \underline{\theta}} \log f(x, \underline{\eta}) \\ \frac{\partial}{\partial \underline{\lambda}} \log f(x, \underline{\eta}) \end{pmatrix} \\ &= \begin{pmatrix} U_\theta(x, \underline{\eta}) \\ U_\lambda(x, \underline{\eta}) \end{pmatrix} \end{aligned}$$

and let $\hat{\underline{\eta}}_0^T = (\underline{\theta}_0^T, \hat{\underline{\lambda}}_0^T)$, where $\hat{\underline{\lambda}}_0^T$ satsifies

$$\sum_{i=1}^{n} \underline{U}_\lambda(x_i, (\underline{\theta}_0^T, \underline{\lambda}^T)^T) = \underline{0}.$$

Then the analogue of the likelihood ratio statistic is

$$W = 2 \sum_{i=1}^{n} \log f(x_i, \hat{\underline{\eta}}) - 2 \sum_{i=1}^{n} \log f(x_i, \hat{\underline{\eta}}_0)$$

and Kent gives the following result.

Theorem 1.6.1 (Kent) *Let $X_1, \ldots, X_n$ be independent identically distributed observations from $g(y)$ and suppose that $\underline{\theta}(g) = \underline{\theta}_0$, $\dim \underline{\theta} = p$. Let $V_1, \ldots, V_p$ denote independent χ_1^2 variates. Then asymptotically as $n \to \infty$, W is distributed as*

(1.6.5) $$W \sim \sum_{i=1}^{p} \nu_i V_i$$

where the ν_i are the eigenvalues of the matrix $(B_{\theta\theta})^{-1}(BCB)_{\theta\theta}$, where A and B are partitioned:

$$A = \begin{pmatrix} A_{\theta\theta} & A_{\theta\lambda} \\ A_{\lambda\theta} & A_{\lambda\lambda} \end{pmatrix}$$

$$-A^{-1} = B = \begin{pmatrix} B_{\theta\theta} & B_{\theta\lambda} \\ B_{\lambda\theta} & B_{\lambda\lambda} \end{pmatrix}.$$

Kent's result is a generalization of the classical theory. The classical result is that the log likelihood ratio is asymptotically χ^2_p, a distribution with which it is easier to work than that given in (1.6.5).

We note that the function $\log f(x, \theta)$ in Kent's paper need not be a density function. Kent's results remain valid for any "well behaved" score function $\underline{U}(x, \underline{\eta})$, and that is why the matrices in Kent's result have the same form as Huber's result on the asymptotic normality of M-estimates. We apply Kent's result in Chapter 4, with the score functions appropriate for Huber's "Proposal 2" for joint estimation of location and scale (1981). Kent is concerned with those cases when the matrix $(B_{\theta\theta})^{-1}(BCB)_{\theta\theta}$ of Theorem 1:6.1 is the identity matrix. (In Chapter 4, we force B and C to be diagonal.)

Ronchetti's τ-test (1982) is also an analogue of the log-likelihood ratio test: asymptotic normality of the M-estimate is used to conclude that, under the null-hypothesis, the τ-test statistic (which is quadratic in the scores) is asymptotically a linear combination of χ^2_1 random variables.

In all these classical approaches (Huber, Kent, Ronchetti) there is a common step: application of the Central Limit Theorem to the sum-of-scores statistic to conclude that the estimate has an asymptotically normal distribution. Then tests are based on a quadratic function of the scores.

Our work has been motivated by two beliefs:

1) Almost everything we need to know is in the scores.
2) For small samples, we should be able to improve on the normal approximation to the distribution of the sum-of-scores, and then avoid tests based on a quadratic form in the score statistic.

We do *not* assume that the Central Limit Theorem is applicable. Confidence intervals are constructed from a series of tests. Our procedure to test a null-hypothesis is as follows.

1. Use an M-estimate.
2. Work with the scores directly.
3. Use an exponential tilt to improve the initial estimate of the distribution of the test statistic.
4. Use the Lugannani and Rice tail area approximation (Chapter 2).

Chapter 2 Tail area approximation

In Section 2.1 we give Poincaré's definition (1886) of an asymptotic expansion, and state some basic results. Asymptotic expansions are used to approximate complicated functions. In the application, Section 2.2, uniform asymptotic expansions are used to approximate tail areas. The results of Section 2.2 come from a transformation to normality, which is discussed in Section 2.3.

2.1 Asymptotic expansions

Let $f(z)$ be a function of a complex variable z and $\sum_{j=0}^{\infty} a_j z^{-j}$ a formal power series (which may or may not converge). Let $R_n(z)$ denote the difference between $f(z)$ and the n^{th} partial sum of the power series; that is,

$$f(z) = \sum_{j=0}^{n-1} a_j z^{-j} + R_n(z). \tag{2.1.1}$$

Suppose that, for each fixed value of n,

$$R_n(z) = O(z^{-n}), \quad z \to \infty, \tag{2.1.2}$$

in a certain unbounded region $\mathbb{T}$. Then the series $\sum_{j=0}^{\infty} a_j z^{-j}$ is called an asymptotic expansion of $f(z)$ in the region $\mathbb{T}$. We write

$$f(z) \sim a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots, \quad z \to \infty \text{ in } \mathbb{T}. \tag{2.1.3}$$

We remark that the implied constant in (2.1.2) is not the same for all values of n, and it cannot be less that $|a_n|$.

If the function f has an asymptotic expansion, then that expansion is unique (Olver (1974) page 17), although different functions can have the same asymptotic expansion. On the other hand, if the series $\sum_{j=0}^{\infty} a_j z^{-j}$ converges for $|z|$ sufficiently large, then it is the asymptotic expansion of its sum, and $R_n(z) = \sum_{j=n}^{\infty} a_j z^{-j}$. In general, the remainder term in (2.1.1) is *not* the sum of the tail of the series.

The value of asymptotic expansions lies in the fact that the first few terms can give a good approximation to the function $f(z)$. The error of the approximation has the same order as the first term omitted. In our application, the variable z is sample size. In this chapter, we use an asymptotic expansion to approximate a contour integral of the form

$$(2.1.4) \qquad \int \phi(z) e^{\nu h(z)} dz, \quad z \text{ complex}, \ \nu \text{ large, real.}$$

We now state some results from the theory of asymptotic expansions.

Theorem 2.1.1 (Copson (1965), page 39) The Laplace one-term approximation.

Let $\phi(x)$ and $h(x)$ be two real continuous functions defined in the finite or semi-infinite interval $\alpha \le x \le \beta$, such that

(i) $\phi(x)e^{\nu h(x)}$ is absolutely integrable over the interval for every positive value of ν;

(ii) $h(x)$ has a single maximum in the interval, namely at $x = \alpha$; and the supremum of $h(x)$ in any closed subinterval not containing α is less than $h(\alpha)$;

(iii) $h''(x)$ is continuous; and $h'(\alpha) = 0$, $h''(\alpha) < 0$.

Then, as $\nu \to +\infty$,

$$(2.1.5) \qquad \int_{\alpha}^{\beta} \phi(x) e^{\nu h(x)} dx = \phi(\alpha) e^{\nu h(\alpha)} \left[\frac{-\pi}{2\nu h''(\alpha)}\right]^{\frac{1}{2}} + o(1).$$

The above theorem considers an integral of the form (2.1.4), where the integral and the range of integration are real. The result is that, if $h(x)$ has a maximum at the end-point a then, as ν becomes large, the behaviour of the integrand near a dominates the value of the integral in (2.1.5).

Theorem 2.1.2 (Copson (1965), page 49) Watson's Lemma.

Let $\phi(t)$ be an analytic function of t, regular, apart from a branch-point at 0, when $|t| \leq R + \delta$, $|ph(t)| \leq \Delta < \pi$, where R, δ, Δ are positive; and let

$$\phi(t) = \sum_{m=1}^{\infty} a_m t^{(m/r)-1}, \tag{2.1.6}$$

when $|t| \leq R$, r being positive. Also let $|\phi(t)| < Ke^{bt}$, where K and b are positive numbers independent of t, when t is positive and $t \geq R$. Then

$$\int_0^{\infty} e^{-zt}\phi(t)dt \sim \sum_{m=1}^{\infty} a_m \Gamma(\tfrac{m}{r}) z^{-m/r} \tag{2.1.7}$$

as $|z| \to \infty$ in the sector $|ph(z)| \leq \frac{1}{2}\pi - \epsilon < \frac{1}{2}\pi$. If $r = 1$, $\phi(t)$ does not have a branch-point at the origin and the condition $|ph(t)| \leq \Delta$ is not needed.

There is a real version of Watson's Lemma, see, for example, Olver (1974), page 71. For our purposes, the theorem states that, provided one is careful to avoid branch-cut problems, a power series expansion for $\phi(t)$ in a neighborhood of $t = 0$ leads to an asymptotic expansion for the integral (2.1.7). The proof uses the observation that

$$\int_0^{\infty} e^{-zt} t^{(m/r)-1} dt = z^{-m/r}\Gamma(\tfrac{m}{r}). \tag{2.1.8}$$

Example. One can apply Watson's Lemma to the real integral

$$\mathrm{erfc}(\sigma) = \frac{2}{\sqrt{\pi}} \int_{\sigma}^{\infty} e^{-t^2}\, dt, \quad \sigma > 0. \tag{2.1.9}$$

Making the substitution $t = \sigma\sqrt{1+\tau}$, the integral (2.1.9) becomes

$$\mathrm{erfc}(\sigma) = \frac{\sigma}{\sqrt{\pi}} e^{-\sigma^2} \int_0^{\infty} e^{-\tau\sigma^2} \frac{d\tau}{\sqrt{1+\tau}} \tag{2.1.10}$$

where

$$\frac{1}{\sqrt{1+\tau}} = \sum_{m=1}^{\infty}(-1)^{m-1}\frac{\Gamma(m-\frac{1}{2})\tau^{m-1}}{\Gamma(\frac{1}{2})\Gamma(m)}.$$

Hence, applying Watson's Lemma,

$$\text{(2.1.11)} \qquad \operatorname{erfc}(\sigma) \sim \frac{\sigma}{\sqrt{\pi}}e^{-\sigma^2}\sum_{m=1}^{\infty}(-1)^{m-1}\frac{\Gamma(m-\frac{1}{2})}{\Gamma(\frac{1}{2})}\sigma^{-2m}$$

or, equivalently,

$$\text{(2.1.12)} \qquad \operatorname{erfc}(\sigma) \sim \frac{1}{\sqrt{\pi}}e^{-\sigma^2}\sum_{n=0}^{\infty}(-1)^n(\tfrac{1}{2})_n\sigma^{-(2n+1)}$$

where

$$(\alpha)_0 = 1$$
$$(\alpha)_n = \alpha(\alpha+1)\cdots(\alpha+n-1), \quad n > 0.$$

Note that $\frac{1}{2}\operatorname{erfc}(\sigma) = 1 - \Phi(\sigma\sqrt{2})$, and for $\sigma > 0$

$$\text{(2.1.13)} \qquad \operatorname{erfc}(-\sigma) = 2 - \operatorname{erfc}(\sigma) \sim 2 - \frac{1}{\sqrt{\pi}}e^{-\sigma^2}\sum_{n=0}^{\infty}(-1)^n(\tfrac{1}{2})_n\sigma^{-(2n+1)}.$$

There is no need to restrict attention to real values of σ in the Example. We pursue this example further, in order to demonstrate some of the techniques used to approximate contour integrals. But first we need to discuss paths of steepest descent for $|e^{\nu h(z)}|$.

Any smooth arc in complex space can be described as $z = z(\tau)$, τ real and $|z'(\tau)| = 1$. Then

$$\text{(2.1.14)} \qquad \begin{aligned}\frac{d}{d\tau}|e^{\nu h(z)}| &= \frac{d}{d\tau}e^{\nu Re(h(z))} \\ &= \nu e^{\nu Re(h(z))}\frac{d}{d\tau}Re(h(z)).\end{aligned}$$

Therefore, $|e^{\nu h(z)}|$ increases or decreases as $Re(h(z))$ increases or decreases, $\nu > 0$. Also,

$$\text{(2.1.15)} \qquad h'(z)\frac{dz}{d\tau} = \frac{d}{d\tau}Re(h(z)) + i\frac{d}{d\tau}Im(h(z)),$$

so that, by using $|\frac{dz}{d\tau}| = 1$,

(2.1.16) $$\left\{\frac{d}{d\tau}Re(h(z))\right\}^2 = |h'(z)|^2 - \left\{\frac{d}{d\tau}Im(h(z))\right\}^2.$$

Therefore, from any fixed point $z = w$, the greatest absolute change in $Re(h(z))$, and hence the greatest absolute change in $|e^{\nu h(z)}|$, is along the level curve for $Im(h(z))$; that is, along the curve $Im(h(z)) = Im(h(w))$.

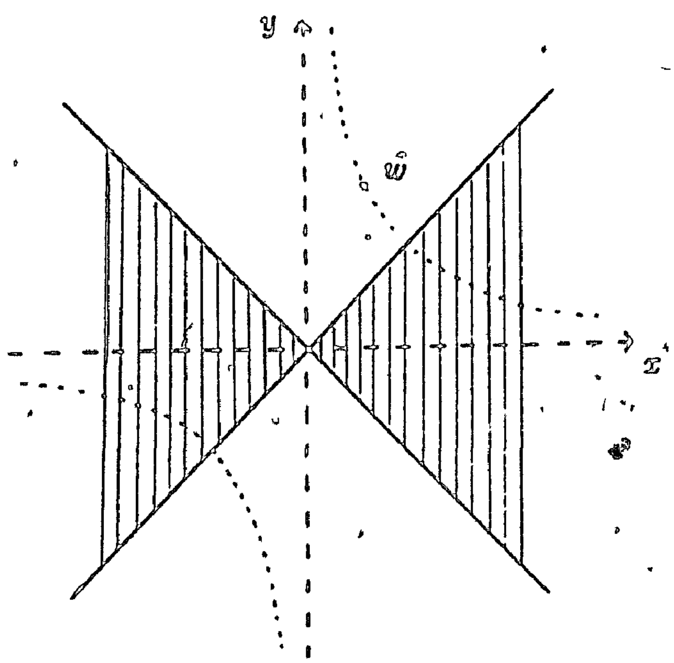
Example (continued). In the Example we considered

(2.1.17) $$\operatorname{erfc}(\sigma) = \frac{2}{\sqrt{\pi}}\int_\sigma^\infty e^{-t^2}\,dt,$$

with σ real. Suppose now that σ is complex and write $\sigma = \sqrt{\nu}e^{\alpha i}$, ν real, $\nu > 0$, $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$. Making the substitution $t = z\sqrt{\nu}$, (2.1.17) can be rewritten

(2.1.18) $$\operatorname{erfc}(\sigma) = \frac{2}{\sqrt{\pi}}\nu^{\frac{1}{2}}\int_w^\infty e^{-\nu z^2}\,dz, \quad w = e^{\alpha i}.$$

Suppose that w is the point indicated in Figure 2.1.1.



Hills and valleys of $Re(h(z))$ (Copson).

Figure 2.1.1.

Writing $h(z) = -z^2$, we need to evaluate

$$J = \int_w^{\infty} e^{\nu h(z)} dz. \tag{2.1.19}$$

If $z = x + iy$, then $Re(h(z)) = y^2 - x^2$ and $Im(h(z)) = -2xy$. There is a saddle-point for $h(z)$ at $z = 0$. The lines $x = \pm y$ are the level curves of $Re(h(z))$ passing through $z = 0$. The shaded regions show the valleys of $Re(h(z))$; that is, regions where $Re(h(z)) < 0$. The other two regions are the hills of $Re(h(z))$. The steepest descent curve from $z = 0$ is the x-axis. The steepest ascent curve from the origin is the y-axis.

For the point $w = e^{i\alpha}$, of integral (2.1.18), with $0 < \alpha < \frac{\pi}{2}$, $Im(h(w)) = -\sin(2\alpha)$, and the branch of the hyperbolic curve $2xy = \sin(2\alpha)$ passing through w is the curve through $z = w$ on which $|e^{\nu h(z)}|$ decreases most rapidly, as $z \to +\infty$ (see Figure 2.1.1). On this steepest descent curve from w to $+\infty$, $Im(h(z))$ is constant, and we can write

$$h(z) = -z^2 = -w^2 - \tau,$$

with τ real, τ increasing from 0 to ∞. Then, using $\nu w^2 = \sigma^2$, we can write

$$-\nu z^2 = -\sigma^2 - \nu\tau$$

on the steepest descent curve from w. Note also that $2z dz = d\tau$ and

$$\begin{aligned} z &= (w^2 + \tau)^{\frac{1}{2}} \\ &= w\left(1 + \frac{\tau}{w^2}\right)^{\frac{1}{2}}, \end{aligned}$$

where the root is uniquely determined so that z lies in the indicated quadrant. Then

$$\begin{aligned} \operatorname{erfc}(\sigma) &= \frac{2}{\sqrt{\pi}} \nu^{1/2} \int_w^{\infty} e^{-\nu z^2} dz \\ &= \frac{\nu^{1/2}}{\sqrt{\pi}} e^{-\sigma^2} \int_0^{\infty} e^{-\nu\tau} \left(1 + \frac{\tau}{w^2}\right)^{-1/2} \frac{d\tau}{w}, \end{aligned}$$

where $|1+\frac{\tau}{w^2}|^{-1/2}$ is bounded, since τ is positive real and $w^2 = e^{2\alpha i}$ is bounded away from the negative real axis. (Remember that $\alpha = ph(\sigma)$ is fixed, $|\alpha| < \frac{\pi}{2}$.) One can then apply Watson's Lemma to obtain

$$\operatorname{erfc}(\sigma) = \frac{|\sigma|}{\sqrt{\pi}} e^{-\sigma^2} \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(n+1/2)}{\Gamma(1/2)} \sigma^{-(2n+1)},$$

which is the result obtained in (2.1.12) for σ real, $\sigma > 0$, but verified now for $|ph(\sigma)| < \frac{\pi}{2}$.

A common method to obtain an asymptotic expansion for a contour integral is to deform the path of integration to pass through a saddle-point, via a steepest ascent/descent curve, then to apply Watson's Lemma to the new contour integral. Clearly, this requires the integrand to be analytic throughout the region of the deformation of the contour. If there is a pole in the region of the deformation, then we must append to the ascent/descent curve a "detour curve" to pass around the pole. (See Figures in the next section.) We state a classic result on the approximation of contour integrals.

Theorem 2.1.3 (Olver (1974), page 127)

Assumptions

(i) $h(t)$ and $\phi(t)$ are independent of z, and single valued and holomorphic in a domain $\mathbb{T}$.

(ii) The integration path P is independent of z. The end-points a and b of P are finite or infinite, and P lies within $\mathbb{T}$.

(iii) $h'(t)$ has a simple zero at an interior point t_0 of P.

(iv) z ranges along a ray or over an annular sector given by $\theta_1 \le \theta \le \theta_2$ and $|z| \ge Z$, where $\theta \equiv ph(z)$, $\theta_2 - \theta_1 < \pi$ and $Z > 0$. The integral $I(z)$,

$$I(z) = \int_0^{\infty} e^{-zt} \phi(t)\,dt$$

converges at a and b absolutely and uniformly with respect to z.

(v) $\mathrm{Re}(e^{i\theta}h(t) - e^{i\theta}h(t_0))$ *is positive on* P*, except at* t_0*, and is bounded away from zero uniformly with respect to* θ *as* $t \to a$ *or* b *along* P.

With the foregoing assumptions,

$$\int_a^b e^{-zh(t)}\phi(t)\,dt \sim 2e^{-zh(t_0)} \sum_{m=0}^{\infty} \Gamma(m+\frac{1}{2}) \frac{a_{2m}}{z^{m+\frac{1}{2}}} \tag{2.1.20}$$

as $z \to \infty$ *in the sector* $\theta_1 \le ph(z) \le \theta_2$*, where*

$$a_0 = \frac{\phi}{(2h^{(2)})^{1/2}},$$
$$a_2 = \left\{2\phi'' - \frac{2h^{(3)}\phi'}{h^{(2)}} + \left(\frac{5h^{(3)^2}}{6h^{(2)^2}} - \frac{h^{(4)}}{2h^{(2)}}\right)\phi\right\}\frac{1}{(2h^{(2)})^{3/2}}$$

and h*,* ϕ *and their derivatives are evaluated at* $t = t_0$*. In forming* $(2h^{(2)})^{1/2}$ *and* $(2h^{(2)})^{3/2}$*, the branch of* $\omega_0 \equiv ph(h^{(2)}(t_0))$ *must satisfy*

$$|\omega_0 + \theta + 2\omega| \le \frac{\pi}{2},$$

where ω *is the limiting value of* $ph(t - t_o)$ *as* $t \to t_0$ *along* P.

This result is stated more precisely in Rice (1968), page 1999, but Theorem 2.1.3 is adequate for our applications. Notice that Theorem 2.1.1 is a special case of Theorem 2.1.3.

2.2 The Lugannani-Rice tail area approximation

The contour integral which concerns us has the form

(2.2.1)
$$J = \int \frac{1}{z} e^{nh(z)} dz,$$

n a positive integer. The classical asymptotic expansion for this integral is an expansion around the saddle-point z_0 of $h(z)$, see (2.2.9) below. Unfortunately, the terms of the expansion become too large for z_0 close to the pole at $z = 0$. We circumvent this problem by using a uniform asymptotic expansion. The expansion is asymptotic in the large variable n, and uniform as $|z_0| \to 0$. Typically, uniform asymptotic expansions have the form of line (2.3.2), below, where the $V_k(n)$ are integrals, less complicated than the original. Uniform asymptotic expansions are discussed in Rice (1968), Bleistein and Handelsman (1975), and Lugannani and Rice (1980). We restrict attention to the particular contour integral that concerns us. The result, (2.2.10), can be found in Lugannani and Rice (1980). We do not attempt to verify the result but, in the next section, we indicate how it is derived.

Let $V_1, \ldots, V_n$ be independent identically distributed random variables. Suppose that V_1 has a density $f(v)$ and let $K(t)$ denote the cumulant generating function of V_1. Then $e^{K(t)} =: E(e^{tV_1})$, and the characteristic function of V_1 is $M(it) = e^{K(it)}$. Let $Y = V_1 + \ldots + V_n$. We wish to approximate

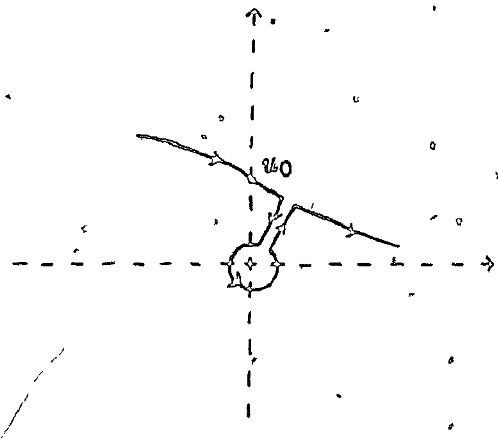
(2.2.2)
$$Q_n(y) = Pr(Y > y).$$

Since V_1 is assumed to have a density, the Fourier inversion formula gives that

(2.2.3)
$$f(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{K(iu) - iuv} du,$$

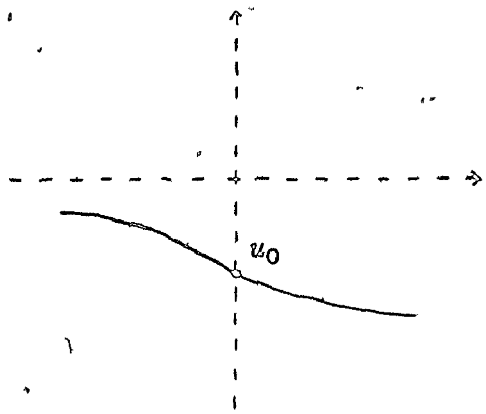
and the density of Y is given by

(2.2.4)
$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{nK(iu) - iuy} du.$$



Steepest Descent Curve with detour about pole.
$Im(u_0) > 0$

Figure 2.2.1



Steepest Descent Curve.
$Im(u_0) < 0$

Figure 2.2.2

Then

$$Q_n(y_0) = \int_{y_0}^{\infty} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{nK(iu)-iuy}\,du\,dy. \tag{2.2.5}$$

Reversing the order of integration, we obtain

$$Q_n(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{nK(iu)-iuy} \frac{du}{iu}. \tag{2.2.6}$$

The path of integration in the integral (2.2.3) cannot pass through the pole at the origin. We choose to indent downwards at the origin, then the integral (2.2.6) is along the same path and is well defined.

We remark that the above argument can be used to prove

$$\int_{-\infty}^{\infty} \frac{1}{v} e^{n(v^2-2v_0 v)}\,dv = i\pi \operatorname{erfc}(v_0\sqrt{n}), \tag{2.2.7}$$

a fact used below.

An asymptotic expansion for $Q_n(y)$ can be obtained by considering the saddle-point u_0 of $nK(iu) - iuy$, solution of $K'(iu) - iy/n = 0$. If V_1 is bounded, then Theorem 1.4.2 establishes the existence and uniqueness of u_0 on the imaginary axis.

We deform the path of integration in (2.2.6), so that u follows a path of steepest ascent up to the point $u = u_0$, then a path of steepest descent away from $u = u_0$. Ascent and descent are with respect to the value of $Re(K(iu) - iuy/n)$. There are two cases to consider, determined by where the steepest path lies in relation to the pole at the origin and the original path of integration (which was chosen to pass *below* the origin). These cases are illustrated in the diagrams. In the case illustrated by Figure 2.2.1, the steepest descent curve must have a "detour" added to it, since we cannot deform through the pole at the origin. The classical aymptotic expansion for $Q_n(y)$ is then obtained by applying Theorem 2.1.3,

$$Q_n(y) \sim \begin{cases} 1 + \sum_{j=0}^{\infty} A_j, & Im(u_0) > 0 \\ \sum_{j=0}^{\infty} A_j, & Im(u_0) < 0, \end{cases} \tag{2.2.9}$$

where
$$t_0 = iu_0 \text{ is real}, \quad K'(t_0) = y/n,$$
$$f_0 = nK(t_0) - t_0 y$$
$$\mu = (t_0(K''(t_0))^{1/2})^{-1}$$
$$\theta_j = K^{(j)}(t_0)$$
$$A_0 = \frac{\mu}{\sqrt{2\pi n}} e^{f_0} = O(\tfrac{1}{\sqrt{n}})$$
$$A_1 = -3\frac{A_0}{n}\{\mu^2/3 + \mu\theta_3 + (5\theta_3^2 - 2\theta_4)/2\} = O(\tfrac{1}{n\sqrt{n}}),$$

and further terms A_j can be calculated by applying Rice (1968, page 1999) but they are not used in the sequel. Our notation is the same as that in Lugannani and Rice (1980), with the exceptions that we use Daniels' symbol, K, for the cumulant generating function and the symbol n for sample size. The series can be used to approximate $Q_n(y)$ with error of order $O(\frac{1}{n^{j+(1/2)}})$ provided all derivatives $K^{(l)}(t_0)$ required to calculate $A_0, A_1, \ldots, A_j$ are finite.

The series (2.2.9) is valid in any region bounded away from $y = nE(V_1)$. As $y/n \to E(V_1)$, the saddle-point u_0 approaches zero and hence the A_j, which involove powers of $1/u_0$, become unbounded. Consequently, a statement that the error has the same order as the first omitted A_j, does not really give a control on the error for y/n near $E(V_1)$. The problem arises because the integrand in equation (2.2.6) for $Q_n(y)$ has a pole at $u = 0$. In our applications, y is the observed value of Y. The Law of Large Numbers states that y/n is close to $E(V_1)$ with high probability as $n \to \infty$. Therefore, it is important to obtain an approximation to $Q_n(y)$ which remains valid for y/n arbitarily close to $E(V_1)$; that is, for u_0 close to the pole at $u = 0$. We need an integral approximation which "picks up" the influence of the pole at the origin. The required approximation is given in Lugannani and Rice (1980):

(2.2.10)
$$Q_n(y) \sim \frac{1}{2}\mathrm{erfc}(\sqrt{-f_0}) + \sum_{j=0}^{\infty}(A_j - B_j),$$

where t_0, f_0, A_j are as defined for the series (2.2.9),

$$B_j = \frac{1}{2}(-\pi f_0)^{-1/2} f_0^{-j} \left(\frac{1}{2}\right)_j e^{f_0}, \quad j \geq 0,$$

and the sign of $\sqrt{-f_0}$ is the same as the sign of t_0. (See Lugannani and Rice).

Note that, for $\sqrt{-f_0} > 0$, (that is, $t_0 > 0$ and $Im(u_0) < 0$), comparison with (2.1.12) shows that

$$\text{(2.2.11)} \qquad \frac{1}{2}\operatorname{erfc}(\sqrt{-f_0}) \sim \sum_{j=0}^{\infty} B_j.$$

On the other hand, for $t_0 < 0$, $\sqrt{-f_0}$ is *negative* and (2.1.13) gives

$$\frac{1}{2}\operatorname{erfc}(\sqrt{-f_0}) = 1 - \frac{1}{2}\operatorname{erfc}(-\sqrt{-f_0})$$

and (2.2.11) gives, since the B_j have also changed sign,

$$\frac{1}{2}\operatorname{erfc}(-\sqrt{-f_0}) \sim -\sum_{j=0}^{\infty} B_j.$$

Then (2.2.10) states that

$$Q_n(y) \sim 1 + \sum_{j=0}^{\infty} B_j + \sum_{j=0}^{\infty}(A_j - B_j).$$

The expansion (2.2.10) is a refinement of the classical expansion (2.2.9).

Consider $r = y/n$ held fixed. Then the series (2.2.10) can be written

$$Q_n(y) \sim \frac{1}{2}\operatorname{erfc}(\sqrt{-f_0}) + e^{n\gamma_0} \sum_{j=0}^{\infty} C_j \left(\tfrac{1}{n}\right)^{j+1/2},$$

where $\gamma_0 = f_0/n < 0$ and the C_j are bounded as $r \to E(V_1)$, that is as $t_0 \to 0$. (The C_j are not functions of n directly, but functions of r.) We demonstrate that C_0 is bounded as $t_0 \to 0$.

$$\begin{aligned} A_0 - B_0 &= \frac{e^{f_0}}{\sqrt{2\pi n}} \left\{ \frac{1}{t_0\sqrt{K''(t_0)}} - \frac{1}{\sqrt{-2f_0/n}} \right\} \\ &= \frac{e^{f_0}}{\sqrt{2\pi n}} \left\{ \frac{1}{t_0\sqrt{K''(t_0)}} - \frac{1}{\sqrt{2}\sqrt{t_0 r - K(t_0)}} \right\}. \end{aligned}$$

For t_0 in a neighbourhood of $t = 0$,

$$K(t_0) = K'(0)(t_0) + K''(0)\frac{t_0^2}{2} + e_1\frac{t_0^3}{3!}$$

$$r = K'(t_0) = K'(0) + K''(0)t_0 + e_2\frac{t_0^2}{2!}$$

$$K''(t_0) = K''(0) + e_3 t_0$$

where $e_1, e_2, e_3 \to K'''(0)$ as $t_0 \to 0$. Then

$$\begin{aligned} t_0 r - K(t_0) &= K'(0)t_0 + K''(0)t_0^2 + e_2\frac{t_0^3}{2} - K'(0)t_0 - K''(0)\frac{t_0^2}{2} - e_1\frac{t_0^3}{6} \\ t_0 r - K(t_0) &= \frac{t_0^2}{2}\{K''(0) + t_0\frac{3e_2 - e_1}{3}\}. \end{aligned} \tag{2.2.12}$$

Next, apply a one-term Taylor approximation to obtain

$$\frac{1}{t_0\sqrt{K''(t_0)}} = \frac{1}{\left\{t_0\sqrt{K''(0) + e_3 t_0}\right\}}$$
$$\doteq \frac{1}{t_0}\left\{\frac{1}{\sqrt{K''(0)}} - \frac{e_3 t_0}{2K''(0)^{3/2}}\right\}.$$

And, similarly, equation (2.2.12) yields

$$\frac{1}{\sqrt{2}\sqrt{t_0 r - K(t_0)}} = 1/\{t_0\sqrt{K''(0) + \frac{3e_2 - e_1}{3}t_0}\}$$
$$= \frac{1}{t_0}\{\frac{1}{\sqrt{K''(0)}} - \frac{t_0}{2K''(0)^{3/2}} \cdot \frac{3e_2 - e_1}{3}\}.$$

Then

$$C_0 = \frac{1}{\sqrt{2\pi}}\left\{\frac{1}{t_0\sqrt{K''(t_0)}} - \frac{1}{\sqrt{-2f_0/n}}\right\}$$

and

$$\lim_{t_0\to 0} C_0 = \frac{1}{\sqrt{2\pi}}\lim_{t_0\to 0}\frac{1}{K''(0)^{3/2}}\left\{\frac{-e_3}{2} + \frac{3e_2 - e_1}{6}\right\}$$
$$= \frac{-1}{\sqrt{2\pi}}\frac{K'''(0)}{K''(0)^{3/2}3!}.$$

Thus, C_0 is bounded as $t_0 \to 0$ and the right hand side of the approximation

$$Q_n(y) \simeq \frac{1}{2}\mathrm{erfc}(\sqrt{-f_0}) + \frac{e^{n\gamma_0}}{\sqrt{n}}C_0$$

approaches $\frac{1}{2} - \frac{K'''(0)}{3!\sqrt{2\pi n}\ K''(0)^{3/2}}$, in agreement with the Lugannani and Rice result (1980), equation (17), *and* in agreement with the Edgeworth expansion at the mean. (Cox and Hinkley (1974), page 464). Remember that $Q_n(y)$ is the tail area to the *right* of y. For n actually equal to $E(V_1)$, the A_j and B_j are infinite, but the limits $\lim_{t_0 \to 0} A_j - B_j$ exist and give the Edgeworth expansion. We see in the applications that, for our purposes, the case $r = E(V_1)$ is never a problem.

We use the approximation (2.2.10) only as far as the $j = 0$ term. In the applications, our concern is with the tail area determined by a sample average. We write

$$P(\bar{V} > r) = Q_n(y), \quad r = \frac{y}{n}.$$

Then the approximation used in the sequel is

$$P(\bar{V} > r) \doteq 1 - \Phi(\sqrt{-2f_0}) + \frac{e^{f_0}}{\sqrt{2\pi}}\left\{\frac{1}{t_0\sqrt{nK''(t_0)}} - \frac{1}{\sqrt{-2f_0}}\right\}, \tag{2.2.13}$$

where Φ is the distribution function of a standard normal random variable, $K'(t_0) = r$, and $f_0 = n(K(t_0) - t_0 r)$, the sign of $\sqrt{-2f_0}$ is taken to be the same as the sign of t_0, and the error is $O(\frac{1}{n\sqrt{n}})$, provided V_1 has finite fourth cumulants.

The derivation of the tail-area approximation relies heavily on the existence of a density function $f(v)$ for V_1. In our application, the distribution of V_1 is estimated by an exponential tilt $\hat{F}_\alpha$ of the empirical distribution $\hat{F}$. The distribution $\hat{F}_\alpha(v)$ is not continuous, and hence does not have a density. However, $\hat{F}_\alpha$ can be approximated to any specified order of accuracy by a continuous distribution, call it $\hat{G}_\alpha$. Then the cumulant generating function of $\hat{F}_\alpha$ approximates the cumulant generating function of $\hat{G}_\alpha$ uniformly in every finite interval (Theorem 2, page 508, Feller (1971)), and so the approximation (2.2.13) remains valid.

The tail area approximation given in (2.2.13) has error of order $O(\frac{1}{n\sqrt{n}})$, provided $K^{(j)}(t_0) < \infty$, $j \leq 4$.

2.3 The approximation as a local transformation to normality

We now discuss how the Lugannani-Rice approximation works. The integral (2.2.6) for $Q_n(y)$ can be written as

$$Q_n(y) = \frac{1}{2\pi i} J, \tag{2.3.1}$$

where

$$J = \int \frac{1}{t} e^{nK(t) - ty} dt.$$

the uniform asymptotic series is constructed by assuming a transformation $t \mapsto v$ such that, in a region enclosing the saddle-point $t = t_0$ and the pole $t = 0$, $0 \mapsto 0$, $t_0 \mapsto v_0$ and

$$K(t) - tr = v^2 - 2v_0 v, \qquad r = y/n.$$

That is,

$$K_r(t) = v^2 - 2v_0 v$$

where

$$K_r(t) = K(t) - tr$$

is the cumulant generating function of $V_1 - r$. Then

$$J = \int v^{-1} f(v) e^{n(v^2 - 2v_0 v)} dv$$

where $f(v) = (\frac{t}{v})^{-1} \frac{dt}{dv}$. The next step is to take a linear approximation $p_{00} + p_{01} v$ to $f(v)$. It can be shown that the error term for the approximation

$$J \doteq \int v^{-1} (p_{00} + p_{01} v) e^{n(v^2 - 2v_0 v)} dv$$

can be written as $1/n$ times an integral with the same form as that in (2.3.1), say

$$J_1 \doteq \int v^{-1} f_1(v) e^{n(v^2 - 2v_0 v)} dv.$$

Next, take a linear approximation $p_{10} + p_{11}v$ to the amplitude $f_1(v)$ in J_1, and so on. Using this technique, we can write

(2.3.2) $$J \sim V_0(n)\sum_{j=0}^{\infty} p_{j0}(\tfrac{1}{n})^j + V_1(n)\sum_{j=0}^{\infty} p_{j1}(\tfrac{1}{n})^j,$$

where

$$V_0(n) = \int \frac{1}{v}e^{n(v^2-2v_0v)}\,dv$$
$$= i\pi\operatorname{erfc}(v_0\sqrt{n}),$$

by (2.2.7), and

$$V_1(x) = \int e^{n(v^2-2v_0v)}\,dv$$
$$= i\sqrt{\frac{\pi}{n}}e^{-nv_0^2}.$$

The coefficients p_{ji} come from approximating $(\frac{t}{v})^{-1}\frac{dt}{dv}$ in an appropriate region, and are independent of the path of integration. The p_{ji} could be calculated directly from an expansion for $f(v)$ – the Bleistein technique. It is more convenient to avoid $f(v)$ and calculate the p_{ji} directly, by choosing well-known paths of integration – the Ursell method. Lugannani and Rice use this latter technique, the "well-known" paths being the steepest descent curve through the saddle-point $v = v_0$, and a loop enclosing $v = 0$.

It turns out that the linear approximations $f_j(v) \doteq p_{j0} + p_{j1}v$ are such that $p_{j0} = 0$, $j \geq 1$ and $p_{00} = 1$. Consequently the first series on the right of (2.3.2) consists of just one term, $j = 0$:

$$V_0(n)\sum_{j=0}^{\infty} p_{j0}(\tfrac{1}{n})^j = V_0(n)p_{00}(\tfrac{1}{n})^0 = V_0(n),$$

giving the erfc term in (2.2.10). The terms $A_j - B_j$ in (2.2.10) are multiples of the p_{j1} in (2.3.2). We emphasize that the derivation of the series (2.2.10) is non-trivial. Our interest is with its application.

For our purposes, there is another perspective to put on the Lugannani-Rice approximation. First, consider what happens when V_1 is normal with mean κ_1, variance κ_2. Then

$$K(t) = \kappa_1 t + \frac{\kappa_2}{2}t^2$$

$$K_r(t) = K(t) - tr$$

$$= (\kappa_1 - r)t + \frac{\kappa_2}{2}t^2, \quad r = \frac{y}{n},$$

and $t_0 = (r - \kappa_1)/\kappa_2$. The transformation $t \mapsto v$ is given by $v = t\sqrt{\kappa_2/2}$, so that

$$K_r(t) = v^2 - 2v_0 v, \quad v_0 = \frac{r - \kappa_1}{\sqrt{2\kappa_2}}.$$

The transformation $t \mapsto v$ amounts to changing scale from $V_1 - r$, normal with mean $\kappa_1 - r$, variance κ_2, to $\sqrt{2/\kappa_2}(V_1 - r)$, normal with mean $-2v_0$, variance 2. In this case, the approximation $f(v) \doteq p_{00} + p_{01}v \equiv 1$ is exact, and $p_{ij} = 0, i \geq 1$. We see that

$$\sqrt{-2f_0} = \frac{r - \kappa_1}{\sqrt{\kappa_2/n}}$$

and

$$\frac{1}{2}\text{erfc}(\sqrt{-f_0}) = 1 - \Phi(\sqrt{-2f_0})$$

$$= P(Y > y).$$

The terms $A_j - B_j$ are multiples of the coefficients p_{j1} (in (2.3.2)), all of which are zero since $(t/v)^{-1}\frac{dt}{dv} \equiv 1$. Thus, when the distribution of V_1 is normal, the Lugannani-Rice approximation is exact, and is given by the erfc term. The net effect of the transformation $t \mapsto v$ in this case is a convenient intermediate, "standardization".

In general, the transformation $t \mapsto v$, $K(t) = v^2 - 2v_0 v$, is assumed well-defined in a region enclosing $t = t_0$ and $t = 0$, contributions to the integral $Q_n(y)$ from outside that region being negligible. Suppose that the transformation $t \mapsto v$ can be extended giving, for characteristic functions,

$$K(it) = -v^2 - 2v_0 iv, \quad \text{for all real } t.$$

Then, by the uniqueness of characteristic functions, this would be equivalent to a change of variables $V_1 \rightarrow W_1$, with W_1 normal, mean $-2v_0$, variance 2. The Lugannani- Rice approximation to $Q_n(y)$, (2.2.10), can be considered to come from a transformation to normality. The erfc term comes directly from the normal variable W_1. The remaining terms $A_i - B_j$ come from the non-linearity of the transformation $V_1 \rightarrow W_1$. (Non-linear in the dummy-variable for the characteristic functions.)

This transformation to normality is difficult to "see", since it is going on at the level of the cumulant generating functions. We quote from Field and Hampel (1982, page 32)

"It has been said that the role of the normal distribution in probability is similar to that of the straight line in geometry."

There is a result of Marcinkiewicz (1938) which states that if a distribution F has polynomial cumulant generating function $K(\imath t)$, then the degree of $K(\imath t)$ is no more than two. Therefore F is either a point mass ($K(\imath t)$ has degree strictly less than 2) or a normal distribution (degree 2). It follows that the usual polynomial approximations to cumulant generating functions are not, in general, the cumulant generating functions of anything. The special feature of the normal distribution is that it is the only (non-trivial) distribution with polynomial cumulant generating function.

The Edgeworth approximation comes from an approximation to the cumulant generating function $K(t)$ which is forced to match the behaviour of $K(t)$ near zero; that is, to match $K(0)$, $K'(0)$, $K''(0)$, etc., up to a finite number of derivatives. The Lugannani-Rice approximation comes from a parabolic approximation to the cumulant generating function near a point t_0 that interests us, and $t_0 \neq 0$.

For $r = y/n$ away from $E(V_1)$, it is best to look at the cumulant generating function of the shifted variable $V_1 - r$,

$$K_r(t) = K(t) - tr.$$

Suppose that the graph of $K(t)$ is given by the curve of Figure 2.3.1. (Note, in particular, that this figure implies that $K'(0) < 0$, that is $E(V_1) < 0$). Then $K'(t_0) = r$ and the lower of the two lines with slope r cuts the vertical axis at $K(t_0) - t_0 r$. The graph of $K_r(t)$ is then given by Figure 2.3.2. Near the point t_0, $K_r'(t_0) = 0$, the graph of $K_r(t)$ looks like a parabola with vertex at t_0 and passing through the origin (but it is generally *not* parabolic).

The point t_0 solves $K'(t) = r$, that is

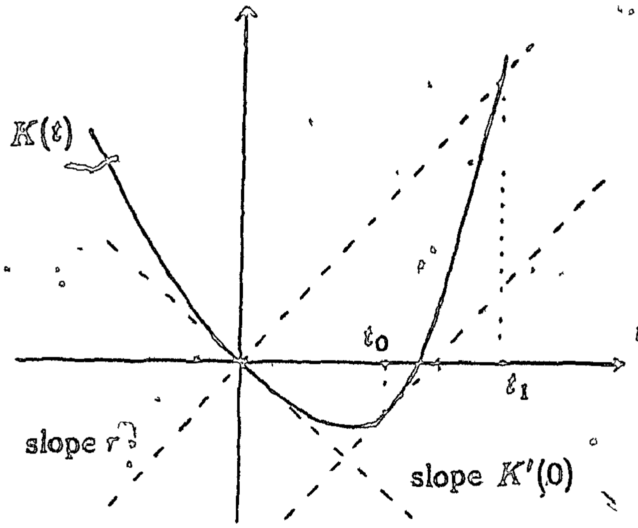
$$\kappa_1 + \kappa_2 t_0 + \kappa_3 \frac{t_0^2}{2!} + \ldots = r,$$

so that

$$t_0 \doteq \frac{r - \kappa_1}{\kappa_2}.$$

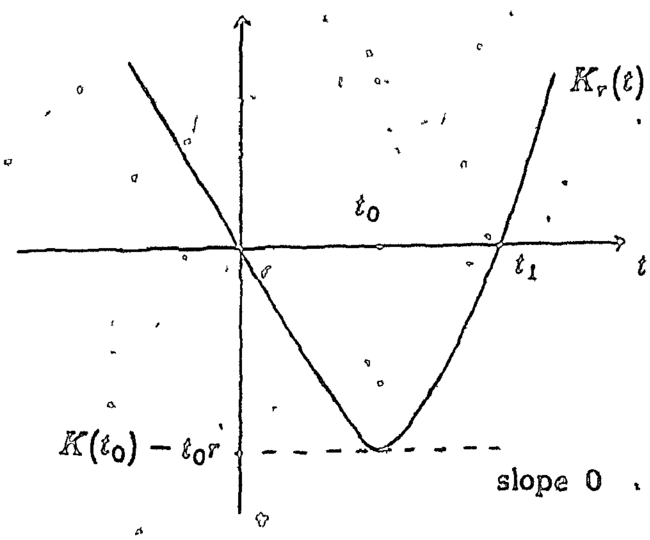
The point t_1 is the non-zero solution of $K_r(t) = 0$, and t_1 is approximately equal to $2t_0$. The Lugannani-Rice approximation proceeds by distorting the approximate parabola of Figure 2.3.2 to give a true parabola $H(v) = v^2 - 2v_0 v$. The transformation $t \mapsto v$ is such that $0 \mapsto 0$, $t_0 \mapsto v_0$, and the point t_1, which is only approximately equal to $2t_0$ is forced: $t_1 \mapsto 2v_0$. The amount by which the curve in Figure 2.3.2 is *not* parabolic is determined by the cumulants $\kappa_3, \kappa_4, \ldots$. The transformed curve *is* parabolic and the terms $(A_j - B_j)$ in the Lugannani-Rice approximation are determined by the amount of distortion needed to make Figure 2.3.2 into a true parabola.

As mentioned above, the Lugannani-Rice transformation is always chosen to give a normal random variable with variance 2, so that the parabola $H(v)$ satisfies $H''(0) = 2$. Scale is unimportant. The same tail-area approximation would



Cumulant generating function of V_1.

Figure 2.3.1.



Cumulant generating function of $V_1 - r$.

Figure 2.3.2.

be obtained by distorting the curve of Figure 2.3.2 to any parabola $(v, H(v))$, so long as $H(0) = 0$, $t_0 \mapsto w_0$, $H'(w_0) = 0$ and $H(w_0) = K(t_0) - t_0 r$. (That is, origin $\mapsto$ origin, vertex $\mapsto$ vertex, with specified amplitude). Different parabolas have different values of $H''(0)$ and are equivalent to a change of scale for the normal random variable determined by the cumulant generating function $H(v)$. The essential feature of the tail area approximation is that the cumulant generating function of the transformed variable closely matches that of the variable $V_1 - r$ near the vertex, and it is this part of the cumulant generating function which makes the major contribution to the integral (2.2.6) determining the tail area $Q_n(y)$.

Once we specify that $H(0) = 0$, $H''(0) = 2$, and that, at the vertex, H assumes the value $K(t_0) - t_0 r$ (and this is the way the approximation proceeds), the function $H(v) = v^2 - 2v_0 v$ is determined. In particular, $H(v)$ is the cumulant generating function of a normal random variable with mean $H'(0) = -2\sqrt{\frac{-f_0}{n}}$ and variance $H''(0) = 2$.

The transformation $t \mapsto v$ gives $H(v)$ which behaves, in a region enclosing v_0 and the origin, very much as $K(t)$ behaves in the vicinity of t_0 and the origin. The classic Edgeworth expansion, on the other hand, comes from approximating $K(t)$ by a polynomial which behaves like $K(t)$ at the origin: matching $K(0), K'(0), K''(0), \ldots$ up to a finite number of derivatives. When $r = E(V_1)$, the Lugannani-Rice approximation is found by considering $\lim_{r\to 0}(A_j - B_j)$, and gives the classic Edgeworth approximation. This is not surprising, since both procedures are then approximating $K_r(t)$ at the same point, the origin.

The accuracy of the Lugannani-Rice approximation for M-estimators of location, small samples, is demonstrated in Tables 1 and 2 of Daniels (1983). Daniels (1986) compares the Lugannani-Rice tail-area approximation with the classical saddlepoint approaches, which are based on an Edgeworth expansion for the exponentially tilted density. Daniels praises the performance and simplicity of the Lugannani-Rice approach.

Chapter 3 Small sample confidence intervals

In this chapter we introduce a procedure for constructing confidence intervals for a real-valued function θ of a vector parameter $\underline{\eta}$, $\theta = \theta(\underline{\eta})$. This general problem includes, of course, the problem of finding confidence intervals for a real-valued parameter in the presence of unknown nuisance parameters. We construct confidence intervals with error $O_P(\frac{1}{n})$ for bounded $\underline{\eta}$.

Examples are not introduced until Chapters 4 and 5. The reader may wish to refer ahead to Chapter 4 while reading Sections 1, 5 and 8 of this chapter.

The unknown parameter is estimated by an M-estimate, and we work with a "modified ~~sample" constructed from the observed values of the~~ scores. The modified sample is called the configuration. The confidence interval is constructed as all parameter values not rejected by a specified test procedure. The test is based on the observed average of the configuration.

The configuration is constructed by applying results found in Bhattacharya and Ghosh (1978) and Field (1982), and is calculated once for each sample. Associated with each parameter value under test, there is an exponential tilt of the initial estimate of the distribution of the configuration. Tail areas are approximated via the Lugannani and Rice result of Chapter 2.

In Section 7 we compare our approach with other approaches in the literature and discuss some details concerning the tail-area approximation.

Sections 1 and 8 outline the basic test procedure used to construct confidence intervals.

Section 4 contains an interesting extension of the Weak Law of Large Numbers to M-estimates. This result is the cornerstone of our error arguments.

3.1 Notation and the general test procedure

We assume an independent identically distributed sample $X_1, \ldots, X_n$ drawn from a population with distribution involving an unknown p-dimensional parameter $\underline{\eta}$. The parameter $\underline{\eta}$ has value $\underline{\eta}_0$ and is estimated by an M-estimate $\hat{\underline{\eta}}$, solution of

$$\frac{1}{n}\sum_{i=1}^{n}\underline{\Psi}(x_i, \underline{\eta}) = \underline{0},$$

where the score function $\underline{\Psi}$ is p-dimensional. Our primary concern is with the construction of confidence intervals for a real-valued parameter $\theta = \theta(\underline{\eta})$, $\theta\colon \Re^p \to \Re$. The confidence interval is constructed as all values of θ_0 for which one should accept $H_0\colon \theta = \theta_0$. The procedure to test H_0 is as follows.

Step 1. A one-dimensional test statistic $\bar{g} = \frac{1}{n}\sum g_i$ is chosen, and a shift condition is specified.

Step 2. The sample is used to construct a first approximation to the distribution of $\bar{g}$ under $\underline{\eta}_0$, where $\underline{\eta}_0$ is unknown, but assumed to satisfy $H_0\colon \theta(\underline{\eta}_0) = \theta_0$.

Step 3. The distribution of $\bar{g}$ is re-estimated via an exponential tilt chosen so that, under the tilted distribution, $E(\bar{g})$ satisfies the shift condition of Step 1.

Step 4. Under the tilted distribution, estimate the probability that $\bar{g}$ attains a value more extreme than that actually observed, then accept or reject H_0 accordingly.

The g_i are functions of the sample, $g_i = g(X_i, \hat{\underline{\eta}}, \frac{\partial\theta}{\partial\underline{\eta}}(\hat{\underline{\eta}}))$. In the sequel, the vector $\underline{g} = (g_1, \ldots, g_n)^T$ is called the configuration. The shift condition is determined by θ_0 and the observed value $\hat{\theta} = \theta(\hat{\underline{\eta}})$. In Step 4, we use the Lugannani and Rice tail-area approximation.

3.2 The Weak Law of Large Numbers

We begin with two definitions concerning the convergence, in probability, of sequences of random variables. Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables and $f(n)$ a positive-valued function defined on the positive integers. We say that $X_n = o_P(f(n))$ as $n \to \infty$ if, for every $\varepsilon > 0$, there is an N_ε such that

$$P(|\frac{X_n}{f(n)}| < \varepsilon) > 1 - \varepsilon \qquad \text{for every } n > N_\varepsilon.$$

We write $X_n = O_P(f(n))$, $n \to \infty$, if for every $\varepsilon > 0$ there are an A_ε and an N_ε such that

$$P(|X_n| < A_\varepsilon f(n)) > 1 - \varepsilon \qquad \text{for every } n > N_\varepsilon.$$

Note that if $X_n = o_P(f(n))$, $n \to \infty$, then $X_n = O_P(f(n))$, $n \to \infty$. If $X_n = O_P(n^{-a})$, $n \to \infty$, for some $a > 0$, then $X_n = o_p(n^{-b})$, $n \to \infty$, for all b satisfying $0 < b < a$.

The independent, identically distributed sample $X_1, \ldots, X_n$ can be considered to be obtained by truncating a sequence of random variables $\{X_i\}_{i\geq 1}$ at the n^{th} term. In this section we assume that the X_i are 1-dimensional random variables. Khintchin's Weak Law of Large Numbers (Bickel and Doksum (1977)) states that if the X_i have finite mean μ then the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges in probability to μ. That is,

$$\bar{X}_n = \mu + o_P(1). \tag{3.2.1}$$

If the underlying distribution of the X_i is normal, with variance σ^2, then we can say more:

$$\bar{X}_n = \mu + O_P(\frac{1}{\sqrt{n}}). \tag{3.2.2}$$

For the proof of this fact notice that, for $\varepsilon > 0$, we can take $A_\varepsilon = \sigma\Phi^{-1}(1-\varepsilon/2)$ and then

$$P(|\bar{X}-\mu| < A_\varepsilon \frac{1}{\sqrt{n}}) = P(\sqrt{n}|\frac{\bar{X}-\mu}{\sigma}| < A_\varepsilon/\sigma)$$
$$= \Phi(A_\varepsilon/\sigma) - \Phi(-A_\varepsilon/\sigma)$$
$$= 1-\varepsilon.$$

Equation (3.2.2) holds whenever the underlying distribution of the X_i has finite third moments. In order to demostrate this fact, we need the following theorem (Gnedenko, Kolmogorov (1954), page 201). To clarify the notation of the theorem we define the absolute moments of X_i:

$$\beta_s = E|X_1|^s, \qquad s = 1, 2, \ldots \ .$$

Recall from elementary probability theory that the moment of order s exists if and only if the absolute moment of order s exists and is finite.

Theorem 3.2.1 (Cramér, Berry, Esseen) *If the X_i have mean 0, variance 1, and finite third moment then $F_n(t)$, the distribution function of $\frac{1}{\sqrt{n}}(X_1+\ldots+X_n)$, satisfies*

$$|F_n(t) - \Phi(t)| \leq \frac{c\beta_3}{\sqrt{n}\beta_2^{3/2}},$$

where c is a constant.

The following corollary is surely well known.

Corollary 3.2.2 *If the X_i have mean μ, variance σ^2 and finite third moment then*

$$\bar{X}_n = \mu + O_P(\tfrac{1}{\sqrt{n}}).$$

Proof: Let $Y_i = \frac{X_i-\mu}{\sigma}$ and let $F_n(t)$ be the distribution function of $\frac{1}{\sqrt{n}}(Y_1 + \ldots + Y_n)$. We have to show that, for every $\varepsilon > 0$, there are an A_ε and an N_ε such that, for $n > N_\varepsilon$,

$$P\left(|\bar{X}_n - \mu| < \frac{A_\varepsilon}{\sqrt{n}}\right) > 1-\varepsilon,$$

that is,

$$P\left(|\bar{Y}_n| < \frac{A_\varepsilon}{\sigma\sqrt{n}}\right) > 1 - \varepsilon.$$

Let $\varepsilon > 0$ be fixed and take $A_\varepsilon = \sigma\bar{\Phi}^{-1}(1-\varepsilon/4)$. Then, for $n > \left(\frac{4c\beta_3}{c\beta_2^{3/2}}\right)^2$,

$$\begin{aligned}
P\left(|\bar{X}_n - \mu| < \frac{A_\varepsilon}{\sqrt{n}}\right) &= P\left(|\bar{Y}_n| < \frac{A_\varepsilon}{\sigma\sqrt{n}}\right) \\
&= P\left(\tfrac{1}{\sqrt{n}}|Y_1 + \ldots + Y_n| < \frac{A_\varepsilon}{\sigma}\right) \\
&= F_n(\frac{A_\varepsilon}{\sigma}) - F_n(\frac{-A_\varepsilon}{\sigma}).
\end{aligned}$$

Applying Theorem 3.2.1,

$$\begin{aligned}
P\left(|\bar{X}_n - \mu| < \frac{A_\varepsilon}{\sqrt{n}}\right) &> \bar{\Phi}(\frac{A_\varepsilon}{\sigma}) - \bar{\Phi}(\frac{-A_\varepsilon}{\sigma}) - \frac{2c\beta_3}{\sqrt{n}\beta_2^{3/2}} \\
&> 1 - \varepsilon,
\end{aligned}$$

for n sufficiently large.

3.3 Regularity conditions and preliminary results

We now assume that the X_i are m-dimensional, each with distribution $F_{\underline{\eta}_0}$, parameterized by a p-dimensional parameter $\underline{\eta}_0$, belonging to an open subset $\mathcal{N}$ of $\Re^p$. As before, the independent sample $X_1, \ldots, X_n$ can be considered to be obtained by truncating a sequence $\{X_i\}_{i\geq 1}$ at the n^{th} term.

We estimate $\underline{\eta}_0$ by an M-estimate $\hat{\underline{\eta}}$, solution of

$$\text{(3.3.1)} \qquad \frac{1}{n}\sum_{i=1}^{n} \underline{\Psi}(X_i, \underline{\eta}) = \underline{0},$$

where $\underline{\Psi}$ is p-dimensional. Let D_j denote differentiation with respect to η_j, $j = 1, \ldots, p$, and $D_0 f \equiv f$.

In all, we make six assumptions, *A1* to *A6*, and these hold for the rest of the chapter. The assumptions and results gathered in this section are routine in the sense that obvious analogues can be found in the classical theory.

The assumptions are as follows.

A1. The system of equations (3.3.1) has a unique solution.

A2. There is an open subset U of $\Re^m$ such that

(i) for each $\underline{\eta} \in \mathcal{N}$ one has $F_{\underline{\eta}}(U) = 1$ and

(ii) the derivatives $D_j\Psi_r(x, \underline{\eta})$, $D_kD_j\Psi_r(x, \underline{\eta})$, $D_lD_kD_j\Psi_r(x, \underline{\eta})$ exist for $1 \leq r, j, k, l \leq p$.

A3. For each compact $K \subset \mathcal{N}$,

(i) for $0 \leq j, k \leq p$, $1 \leq r \leq p$, $\sup_{\underline{\eta}_0 \in K} E_{\underline{\eta}_0} |D_kD_j\Psi_r(X, \underline{\eta}_0)|^4 < \infty$

(ii) there is an $\varepsilon > 0$ such that for $1 \leq r, j, k, l \leq p$,

$$\sup_{\underline{\eta}_0 \in K} E_{\underline{\eta}_0}\Big(\max_{|\underline{\eta}-\underline{\eta}_0|\leq\varepsilon} |D_lD_kD_j\Psi_r(X, \underline{\eta})|\Big)^3 < \infty.$$

A4. For each $\underline{\eta}_0 \in \mathcal{N}$

$$E_{\underline{\eta}_0}\Psi_r(X, \underline{\eta}_0) = 0$$

and the matrices

$$A(\underline{\eta}_0) = E_{\underline{\eta}_0}[\frac{\partial \underline{\Psi}}{\partial \underline{\eta}^T}(X, \underline{\eta}_0)]$$

$$C(\underline{\eta}_0) = E_{\underline{\eta}_0}[\underline{\Psi}(X, \underline{\eta}_0)\underline{\Psi}^T(X, \underline{\eta}_0)]$$

are non singular.

A5. The functions $A(\underline{\eta})$ and $E_{\underline{\eta}}[(D_{k_1}D_{j_1}\Psi_{r_1})(D_{k_2}D_{j_2}\Psi_{r_2})]$, $0 \leq j_1, j_2, k_1, k_2 \leq p$, $k_1 + j_1 \geq 1$, $k_2 + j_2 \geq 1$, $1 \leq r_1, r_2 \leq p$, are continuous on $\mathcal{N}$.

A6. For each compact $K \subset \mathcal{N}$ and for $0 \leq j, k \leq p$,

$$\sup_{\underline{\eta}_0 \in K} |D_k D_j \theta(\underline{\eta}_0)| < \infty \quad \text{and} \quad \inf_{\underline{\eta}_0 \in K} |D_k \theta(\underline{\eta}_0)| > 0.$$

Assumptions *A2-A5* are listed, in more general form, in Bhattacharya and Ghosh (1978) page 439, and are also listed in Field (1982). We use Field's notation. Bhattacharya and Ghosh apply a result of Bahr (1967) on the Central Limit Theorem in multidimensions to prove the following result.

Theorem 3.3.1 (Bhattacharya and Ghosh) *Assume that A2-A5 hold. There is a sequence of statistics* $\{\hat{\underline{\eta}}_n\}_{n\geq 1}$ *such that for every compact* $K \subset \mathcal{N}$

$$(3.3.2) \quad \inf_{\underline{\eta}_0 \in K} P_{\underline{\eta}_0}(|\hat{\underline{\eta}}_n - \underline{\eta}_0| < d_0 n^{-1/2}(\log n)^{1/2}, \quad \hat{\underline{\eta}}_n \text{ solves } (3.3.1)) = 1 - o(\tfrac{1}{\sqrt{n}}),$$

where d_0 *is a constant which may depend on* K.

Corollary 3.3.2 *With the above assumptions,* $|\hat{\underline{\eta}}_n - \underline{\eta}_0|^m$ *is* $o_p(n^{-a})$ *on every compact set* $K \subset \mathcal{N}$*, provided that* $\frac{m}{2} - a > 0$.

In the light of assumption *A1* (that $\hat{\underline{\eta}}$ is unique), Theorem 3.3.1 implies that $\hat{\underline{\eta}}_n$ is weakly consistent for $\underline{\eta}_0$:

$$(3.3.3) \qquad \hat{\underline{\eta}}_n = \underline{\eta}_0 + o_P(1),$$

a result established by Huber (1967) under weaker conditions. The above equation establishes that the Weak Law of Large Numbers extends to *M*-estimates under fairly general conditions. The Bhattacharya and Ghosh paper implies a result, equation 3.3.10 below, which allows us to refine equation (3.3.3), replacing $o_P(1)$ by $O_P(\frac{1}{\sqrt{n}})$. This refinement is presented in the next section.

For the moment, our interest is with the construction used in the Bhattacharya and Ghosh paper. This construction is also used by Field (1982) in a slightly different setting, and we now present it. For simplicity, we drop the subscript n on $\underline{\hat{\eta}}_n$.

Let $\underline{\eta}_0$ be the true, unknown value of $\underline{\eta}$ and, for each $r = 1, \ldots, p$, consider the second-order Taylor expansion of (3.3.1) about $\underline{\eta}_0$:

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n} \Psi_r(x_i, \underline{\hat{\eta}}) \\
0 &= \frac{1}{n}\sum_{i=1}^{n} \Psi_r(x_i, \underline{\eta}_0) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p} D_j\Psi_r(x_i, \underline{\eta}_0)(\hat{\eta}_j - \eta_{0j}) \\
&\quad + \frac{1}{2n}\sum_{i=1}^{n}\sum_{j,l} D_l D_j \Psi_r(x_i, \underline{\eta}_0)(\hat{\eta}_j - \eta_{0j})(\hat{\eta}_l - \eta_{0l}) \\
&\quad + R_{n,r}(\underline{\hat{\eta}}),
\end{aligned} \tag{3.3.4}
$$

where $R_{n,r}(\underline{\hat{\eta}}) = O_P|\underline{\hat{\eta}} - \underline{\eta}_0|^3 = o_P(\frac{1}{n})$. For the moment we consider the first three terms of (3.3.4). The notation we introduce is that of Field (1982), page 676. For $1 \leq r, j, l \leq p$, let $Z_r = \Psi_r(X, \underline{\eta}_0)$, $Z_{rj} = D_j\Psi_r(X, \underline{\eta}_0)$, $Z_{rjl} = D_l D_j \Psi_r(X, \underline{\eta}_0)$, $\bar{Z}_r = \frac{1}{n}\sum_{i=1}^{n} \Psi_r(X_i, \underline{\eta}_0)$, $\bar{Z}_{rj} = \frac{1}{n}\sum_{i=1}^{n} D_j\Psi_r(X_i, \underline{\eta}_0)$, etc. Let E denote expectation under $\underline{\eta}_0$. Then assumption $A4$ gives $EZ_r = 0$, $1 \leq r \leq p$. Let $E(Z_{rj}) = \mu_{rj}$, $E(Z_{rjl}) = \mu_{rjl}$, and define

$$
\begin{aligned}
\underline{a} &= (\overbrace{0, \ldots, 0}^{p}, \overbrace{\mu_{11}, \mu_{12}, \ldots, \mu_{pp}}^{p^2}, \overbrace{\mu_{111}, \ldots, \mu_{ppp}}^{p^3})^T, \\
\underline{Z}^+ &= (Z_1, \ldots, Z_p, Z_{11}, Z_{12}, \ldots Z_{pp}, Z_{111}, \ldots, Z_{ppp})^T, \\
\underline{\bar{Z}}^+ &= (\bar{Z}_1, \ldots, \bar{Z}_p, \bar{Z}_{11}, \bar{Z}_{12}, \ldots \bar{Z}_{pp}, \bar{Z}_{111}, \ldots \bar{Z}_{ppp})^T,
\end{aligned}
$$

so that $\underline{a}$, $\underline{Z}^+$, $\underline{\bar{Z}}^+$ are k-dimensional, $k = p + p^2 + p^3$.

Referring back to the first 3 terms of (3.3.4) define the function $f: \Re^{k+p} \to \Re^p$, $f = (f_1, \ldots, f_p)$ by

$$f_r(\underline{z}^+, \underline{t}) = z_r + \sum_{j=1}^{p} (t_j - \eta_{0j}) z_{rj} + \frac{1}{2} \sum_{j,l} (t_j - \eta_{0j})(t_l - \eta_{0l}) z_{rjl}, \tag{3.3.5}$$

for $1 \le r \le p$. Note that $f_r(\underline{a}, \underline{\eta}_0) = 0$ and, by assumption A4, the matrix $A(\underline{\eta}_0)$, having (r,j) entry μ_{rj}, is non singular. The Implicit Function Theorem can be applied to prove that there is a unique three-times differentiable function $H: \Re^k \to \Re^p$, $H = (H_1, \ldots, H_p)$, such that $H(\underline{a}) = \underline{\eta}_0$ and $f(\underline{z}^+, H(\underline{z}^+)) \equiv 0$ for $\underline{z}^+$ in a neighborhood of $\underline{a}$.

Theorem 3.3.3 (Bhattacharya, Ghosh) *For* $1 \le r \le p$

$$(H(\underline{\bar{Z}}^+) - \underline{\hat{\eta}})_r = o_P(\tfrac{1}{n}). \tag{3.3.6}$$

This theorem is contained in equations (2.39) and (2.40) of Bhattacharya and Ghosh (1978). The latter equation states, in fact, that for $1 \le r \le p$,

$$P_{\underline{\eta}_0}(|H(\underline{\bar{Z}}^+ - \underline{\hat{\eta}})| \le d(\frac{\log n}{n})^{3/2}) = P_{\underline{\eta}_0}(|R_{n,r}(\underline{\hat{\eta}})| \le d(\frac{\log n}{n})^{3/2}) \to 1, \tag{3.3.7}$$

d a constant depending on $\underline{\eta}_0$.

Theorem 3.3.4 (Field) *For* $1 \le r \le p$

$$(H(\underline{\bar{Z}}^+) - \underline{\eta}_0)_r = \sum_{j=1}^{p} b_{rj} \bar{Z}_j + \frac{1}{2} \sum_{j,l} (\sum_{m=1}^{p} b_{rm} C_{jl}(m)) \bar{Z}_j \bar{Z}_l + \sum_{j,l,m} b_{rl} b_{jm} \bar{Z}_j (\bar{Z}_{lm} - \mu_{lm}) + o_P(\tfrac{1}{n}), \tag{3.3.8}$$

where $B = (b_{rj}) = -A(\underline{\eta}_0)^{-1}$, $C_{jl}(m) = \sum_{i_1, i_2} b_{ji_1} b_{li_2} \mu_{mi_1i_2}$.

This theorem is equation (5) of Field (1982). The terms of equation (3.3.8) are obtained by taking a Taylor expansion for $H(\underline{\bar{Z}}^+)$ about $H(\underline{a})$. The derivatives of H

are obtained by considering appropriate derivatives of $f = (f_1, \ldots, f_p)$. Corollary 3.3.2 implies that $|\underline{\hat{\eta}} - \underline{\eta}_0|^3 = o_P(\frac{1}{n})$. Using analogous arguments, Bhattacharya and Ghosh establish (1978, equation (2.32)) that $|\bar{\underline{Z}}^+ - \underline{a}|^3 = o_P(\frac{1}{n})$. The approximation in Theorem 3.3.4 then has error of order $o_P(\frac{1}{n})$. Note that $|\bar{\underline{Z}} - \underline{a}| = O_P(\frac{1}{\sqrt{n}})$ also follows from Corollary 3.2.2.

The above two theorems give:

$$(\underline{\hat{\eta}} - \underline{\eta}_0)_r = \sum_{j=1}^{p} b_{rj} \bar{Z}_j + \frac{1}{2} \sum_{j,l,m} b_{rm} C_{jl}(m) \bar{Z}_j \bar{Z}_l + \sum_{j,l,m} b_{rl} b_{jm} \bar{Z}_j (\bar{Z}_{lm} - \mu_{lm}) + o_P(\tfrac{1}{n}), \tag{3.3.9}$$

provided $\underline{\eta}_0$ is in a compact subset of $\mathcal{N}$.

The first term on the right of the above equation estimates $(\underline{\hat{\eta}} - \underline{\eta}_0)_r$ with error of order $|\bar{\underline{Z}}^+ - \underline{a}|^2 = O_P(\frac{1}{n})$, and should be compared with the classical approximation of the maximum likelihood estimate as an average of the scores (Cox and Hinkley (1974) page 294, 295). The second and third terms of equation (3.3.9) have order $O_P(\frac{1}{n})$. They can be estimated from the sample, and correct the bias of $\underline{\hat{\eta}} - \underline{\eta}_0$ with error $o_P(\frac{1}{n})$. (Under $\underline{\eta}_0$, the first term on the right of (3.3.9) has expected value zero.)

In the sequel, we do not use the bias-correction terms of equation (3.3.9). We give them here for completeness. If our concern were with bias-reduction for M-estimators, then equation (3.3.9) indicates the nature of the bias, and is to be compared with the bias-correction for maximum likelihood estimates (Cox and Hinkley (1974) page 309, 310). In the remainder of the paper we use only the simplified version of (3.3.9):

$$(\underline{\hat{\eta}} - \underline{\eta}_0)_r = \sum_{j=1}^{p} b_{rj} \bar{Z}_j + o_P(\tfrac{1}{\sqrt{n}}). \tag{3.3.10}$$

3.4 The Weak Law of Large Numbers for M-estimates

The result presented in this section extends Corollary 3.2.2 (that $\bar{X}_n = \mu + O_P(\frac{1}{\sqrt{n}})$) to M-estimators. We repeat equation (3.3.10).

$$(3.4.1) \qquad (\underline{\hat{\eta}} - \underline{\eta}_0)_r = \frac{1}{n}\sum_{i=1}^{n}\{\sum_{j=1}^{p} b_{rj}\Psi_j(X_i, \underline{\eta}_0)\} + o_P(\tfrac{1}{\sqrt{n}}), \quad r = 1, \ldots, p,$$

where the parameter $\underline{\eta}_0$ is fixed (but unknown).

Let r be fixed, $1 \le r \le p$, and write

$$\begin{aligned} h_{ri} &= h_r(X_i, \underline{\eta}_0) \\ &= \sum_{j=1}^{p} b_{rj}\Psi_j(X_i, \underline{\eta}_0). \end{aligned}$$

Then the h_{ri} are independent identically distributed random variables, with finite third moments. To see this, recall that $\underline{\eta}_0$ is assumed bounded (and fixed, as is the matrix B), and that the score functions have bounded fourth moments (assumption $A3$). Also, the h_{ri} have mean zero (assumption $A4$). Hence, Corollary 3.2.2 implies that $\bar{h}_r = O_P(\frac{1}{\sqrt{n}})$. That is, the first term on the right of (3.4.1) is $O_P(\frac{1}{\sqrt{n}})$. We therefore have the following result.

Theorem 3.4.1 *Under assumptions A1-A5,*

$$\underline{\hat{\eta}} = \underline{\eta}_0 + O_P(\tfrac{1}{\sqrt{n}}),$$

on compact subsets of N.

Our concern is with the real-valued parameter $\theta = \theta(\underline{\eta})$. We first expand $\hat{\theta} = \theta(\underline{\hat{\eta}})$ in a Taylor series about $\underline{\eta}_0$.

$$\begin{aligned} \theta(\underline{\hat{\eta}}) - \theta(\underline{\eta}_0) &= (\underline{\hat{\eta}} - \underline{\eta}_0)^T \frac{\partial\theta}{\partial\underline{\eta}}(\underline{\eta}_0) + O|\underline{\hat{\eta}} - \underline{\eta}_0|^2 \\ &= (\underline{\hat{\eta}} - \underline{\eta}_0)^T \frac{\partial\theta}{\partial\underline{\eta}}(\underline{\eta}_0) + O_P(\tfrac{1}{n}), \end{aligned}$$

where we have used assumption $A6$, that θ has bounded second derivatives. Then, substituting (3.4.1) into the above equation, we obtain

$$\begin{aligned}\hat{\theta} - \theta(\underline{\eta}_0) &= (B\bar{\underline{\Psi}})^T \frac{\partial \theta}{\partial \underline{\eta}}(\underline{\eta}_0) + o_P(\tfrac{1}{\sqrt{n}}) \\ &= \frac{1}{n}\sum_{s=1}^{n} \underline{\Psi}^T(X_s, \underline{\eta}_0) B^T \frac{\partial \theta}{\partial \underline{\eta}}(\underline{\eta}_0) + o_P(\tfrac{1}{\sqrt{n}}).\end{aligned} \tag{3.4.2}$$

We define

$$\begin{aligned} g_s &= g(X_s, \underline{\eta}_0) \\ &= \underline{\Psi}^T(X_s, \underline{\eta}_0) B^T \frac{\partial \theta}{\partial \underline{\eta}}(\underline{\eta}_0),\end{aligned}$$

and then (3.4.2) can be written

$$\hat{\theta} - \theta(\underline{\eta}_0) = \bar{g} + o_P(\tfrac{1}{\sqrt{n}}). \tag{3.4.3}$$

Since θ has bounded first derivatives (assumption $A6$), we can apply Corollary 3.2.2 again to conclude that the first term on the right of the above equation is $O_P(\frac{1}{\sqrt{n}})$.

Corollary 3.4.2. *Under assumptions A1-A6*

$$\hat{\theta} = \theta(\underline{\eta}_0) + O_P(\tfrac{1}{\sqrt{n}}),$$

on compact subsets of $\mathcal{N}$.

The random variables $g_1, \ldots, g_n$ are called the configuration. We emphasize that, whether or not we know $\underline{\eta}_0$, we *do* know that $\hat{\theta} = \theta(\underline{\eta}_0) + O_P(\frac{1}{\sqrt{n}})$, so long as $\underline{\eta}_0$ lies in a compact subset of $\mathcal{N}$.

3.5 Further discussion of the Bootstrap

We repeat equation (3.4.3)

$$\hat{\theta} - \theta(\underline{\eta}_0) = \frac{1}{n}\sum_{i=1}^{n} g(x_i, \underline{\eta}_0) + o_P(\tfrac{1}{\sqrt{n}}). \tag{3.5.1}$$

This equation implies Corollary 3.4.2, but as it stands it is *not* a practical tool to help evaluate the distribution of $\hat{\theta}$, since $\underline{\eta}_0$ is assumed unknown.

The classical approaches to testing, discussed in Section 1.6, all require that $\hat{\theta}$ be consistent for θ. That is, $\hat{\theta} = \theta + o_P(1)$. Theorem 3.4.1 and Corollary 3.4.2 state stronger results, $\hat{\theta} = \theta + O_P(\frac{1}{\sqrt{n}})$. Our techniques require this stronger convergence: we need to be able to say that $|\hat{\underline{\eta}} - \underline{\eta}_0|^2$ is smaller than $O|\hat{\underline{\eta}} - \underline{\eta}_0|$. The classical approaches also involve an equation like, but slightly weaker than (3.5.1): the error term is replaced by $o_P(1)$. Then the Central Limit Theorem is applied, etc.

We now discuss the implications of the above equation in a Bootstrap setting. Let $x_1, \ldots, x_n$ be the observed values of $X_1, \ldots, X_n$, giving estimates $\hat{\underline{\eta}}_{\text{obs}}$ and $\hat{\theta}_{\text{obs}} = \theta(\hat{\underline{\eta}}_{\text{obs}})$. The Bootstrap assumption (1.2.1) is that the distribution of $\hat{\theta} - \theta(\underline{\eta}_0)$ is approximated by the distribution of $\hat{\theta}^* - \hat{\theta}_{\text{obs}}$, where $\hat{\theta}^*$ is calculated from a sample $x_1^*, \ldots, x_n^*$ drawn from the multinomial distribution having mass $\frac{1}{n}$ at each x_i, $i = 1, \ldots, n$. Several Bootstrap samples are observed and, on the basis of these, the distribution of $\hat{\theta}^* - \hat{\theta}_{\text{obs}}$ is estimated; call the estimate $\hat{H}$. To construct a confidence interval for θ, there are several options. We could, for example, assume some kind of transformation to normality and construct a bias corrected percentile interval (BC-interval, see Tibshirani, 1986). We could, for example, construct a confidence interval as all θ_0 not rejected by an appropriately chosen test. We could test $H_0: \theta(\underline{\eta}_0) = \theta_0$ as follows. Apply an exponential tilt to $\hat{H}$ so that the tilted distribution $\hat{H}_\alpha$ has mean equal to (a good estimate of) the bias of $\hat{\theta}$ for θ_0. The exponential tilt should be seen as a refinement of the Bootstrap assumption. Then

$\hat{H}_\alpha$ is used as an estimate of the distribution of $\hat{\theta} - \theta_0$ and, under $\hat{H}_\alpha$, we estimate the tail area by

$$P_{\underline{\eta}_0}(\hat{\theta} - \theta(\underline{\eta}_0) > \hat{\theta}_{\text{obs}} - \theta_0) \doteq 1 - \hat{H}_\alpha(\hat{\theta}_{\text{obs}} - \theta_0),$$

and accept or reject H_0 accordingly.

We make two observations about the Bootstrap procedure. First, the advantage of equation (3.5.1) is that, with error $o_P(\frac{1}{\sqrt{n}})$, $\hat{\theta} - \theta(\underline{\eta}_0)$ is a *sample mean*. We do not have to go all the way back to the observed sample $x_1, \ldots, x_n$. Rather, we work with the observed configuaration, $g_{i_{\text{obs}}} = g(x_i, \hat{\underline{\eta}}_{\text{obs}})$. Let $\hat{H}$ be the empirical distribution function, placing mass $\frac{1}{n}$ on each $g_{i_{\text{obs}}}$ then

$$\hat{\theta}^* - \hat{\theta}_{\text{obs}} = \bar{g}^* + o_P(\frac{1}{\sqrt{n}}),$$

where $\bar{g}^*$ is the average of a sample $g_1^*, \ldots, g_n^*$ drawn from $\hat{H}$. Thus, for either a parametric or a non parametric bootstrap, once we have an estimate of the distribution of the configuration, our concern is with the distribution of a sample average.

The BC_a intervals modify the Bootstrap assumption (see Section 1.2) and have been demonstrated to have error $O_P(\frac{1}{n})$ for a one-dimensional parametric model. Our second observation is that one way to modify the Bootstrap assumption is to work directly with $g(X, \hat{\underline{\eta}}_{\text{obs}})$, to obtain a good estimate of the distribution, under $\underline{\eta}_0$, of $g(X, \hat{\underline{\eta}}_{\text{obs}})$, and to then base our conclusions on the observed value $\bar{g}_{\text{obs}}$. That is, *do not* use the bootstrap distribution of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ to approximate the distribution of $g(X, \underline{\eta}_0)$ or of $\hat{\theta}$.

In conclusion, equation (3.5.1), though not practical in itself, leads us to realize that we should work directly with the observed configuration, $g_{i\text{obs}} = g(x_i, \hat{\underline{\eta}}_{\text{obs}})$, and to base tests on the observed value of a *sample mean* $\bar{g}_{\text{obs}}$. There is, of course, a loss of information when we move from the sample to the configuration (see Section 3.7). For each $H_0 = \theta(\underline{\eta}_0) = \theta_0$, we obtain a good estimate for the distribution of

$g(\mathcal{X}, \underline{\hat{\eta}}_{\text{obs}})$, $\underline{\hat{\eta}}_{\text{obs}}$ held fixed, $\underline{\eta}_0$ unknown but assumed to satisfy H_0. On the basis of this estimated distribution we then decide whether or not the observed value $\bar{g}_{\text{obs}}$ is too extreme, and accept or reject H_0 accordingly. We *do not* try to estimate the distribution of $g(X, \underline{\eta}_0)$ under H_0.

As promised at the end of Chapter 1, we now present a test procedure based on the observed values of the scores. The test procedure uses an exponentially tilted Bootstrap distribution of the observed scores. Bootstrap resampling is avoided via the Lugannani-Rice approximation.

3.6 The test statistic and its approximate distribution

The observed sample is $x_1, \ldots, x_n$, observations from a distribution $F_{\underline{\eta}_0}$, where $\underline{\eta}_0$ is unknown but assumed to be in a compact subset of $\mathcal{N}$. The estimates $\hat{\underline{\eta}}_{\text{obs}}$ and $\hat{\theta} = \theta(\hat{\underline{\eta}}_{\text{obs}})$ are calculated and *held fixed*. We wish to test $H_0\colon \theta(\underline{\eta}_0) = \theta_0$, where $\underline{\eta}_0$ is unknown. Our test is based on the observed value $\bar{g}_{\text{obs}} = \frac{1}{n}\sum g(x_i, \hat{\underline{\eta}}_{\text{obs}})$.

What is the distribution of $g(X, \hat{\underline{\eta}}_{\text{obs}})$, for X a random variable drawn from $F_{\underline{\eta}_0}$? Remember that $\hat{\underline{\eta}}_{\text{obs}}$ is held fixed and $\underline{\eta}_0$ is unknown, but assumed to satisfy H_0. A first approximation to the distribution of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ is via the cumulant generating function $K(t, \hat{\underline{\eta}}_{\text{obs}})$. For a parametric model, $K(t, \hat{\underline{\eta}}_{\text{obs}})$ is the cumulant generating function of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ for X drawn from $F_{\hat{\underline{\eta}}_{\text{obs}}}$. For a non-parametric model, let $\hat{H}$ be the empirical distribution function putting mass $\frac{1}{n}$ at each $g(x_i, \hat{\underline{\eta}}_{\text{obs}})$. Then

$$
\begin{aligned}
K(t, \hat{\underline{\eta}}_{\text{obs}}) &= \log E_{\hat{H}}(\exp(tg(X, \hat{\underline{\eta}}_{\text{obs}}))) \\
\exp K(t, \hat{\underline{\eta}}_{\text{obs}}) &= \frac{1}{n}\sum_{i=1}^{n} \exp(tg(x_i, \hat{\underline{\eta}}_{\text{obs}})).
\end{aligned}
$$

Clearly, $K(t, \hat{\underline{\eta}}_{\text{obs}})$ is not a particularly good estimate of the cumulant generating function of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ under $\underline{\eta}_0$. In particular, under this estimate, the first cumulant of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ is zero. (We know that the first cumulant of $g(X, \underline{\eta}_0)$ is zero under $\underline{\eta}_0$, so we suspect that the first cumulant of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ is different from zero, under $\underline{\eta}_0$.)

Take the Taylor expansion

$$
(3.6.1) \qquad g(X, \hat{\underline{\eta}}_{\text{obs}}) = g(X, \underline{\eta}_0) + (\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0)^T \frac{\partial g(X, \underline{\eta}_0)}{\partial \underline{\eta}} + O|\hat{\underline{\eta}}_{\text{obs}} - \eta_0|^2
$$

Notice that

$$\begin{aligned}\frac{\partial g(X,\underline{\eta}_0)}{\partial \underline{\eta}} &= \frac{\partial}{\partial \underline{\eta}}\left[\underline{\Psi}^T(X,\underline{\eta}_0)B^T\frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}\right] \\ &= \left[\frac{\partial \underline{\Psi}}{\partial \underline{\eta}^T}(X,\underline{\eta}_0)\right]^T B^T\frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}} \\ &\quad + \left[\frac{\partial}{\partial \underline{\eta}^T}\left(B^T\frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}\right)\right]^T \underline{\Psi}(X,\underline{\eta}_0)\end{aligned}$$

and recall that $E_{\underline{\eta}_0}\underline{\Psi}(X,\underline{\eta}_0) = \underline{0}$. Then

$$\begin{aligned}E_{\underline{\eta}_0}\frac{\partial g(X,\underline{\eta}_0)}{\partial \underline{\eta}} &= A^T B^T \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}} \\ &= \frac{-\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}},\end{aligned}$$

since $B = -A^{-1}$ (see Theorem 3.3.4). We now take expectations, under $\underline{\eta}_0$, in equation (3.6.1). In that equation, $\hat{\underline{\eta}}_{\text{obs}}$ and $\underline{\eta}_0$ are fixed; X is the variable, drawn from $F_{\underline{\eta}_0}$:

(3.6.2)
$$\begin{aligned}E_{\underline{\eta}_0} g(X,\hat{\underline{\eta}}_{\text{obs}}) &= 0 - (\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0)^T\frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}} + O|\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0|^2 \\ &= -(\hat{\theta}_{\text{obs}} - \theta(\underline{\eta}_0)) + O|\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0|^2 \\ &= \theta(\underline{\eta}_0) - \hat{\theta}_{\text{obs}} + O_P(\tfrac{1}{n}),\end{aligned}$$

where $\hat{\underline{\eta}}_{\text{obs}}$ is held fixed while we take expectations, and then we appeal to Theorem 3.4.1, that $\hat{\underline{\eta}} - \underline{\eta}_0 = O_P(\frac{1}{\sqrt{n}})$. Consequently, under $\underline{\eta}_0$ satisfying H_0: $\theta(\underline{\eta}_0) = \theta_0$, the first cumulant of $g(X,\hat{\underline{\eta}})$ is $\theta_0 - \hat{\theta} + O_P(\frac{1}{n})$. So, our initial estimate of the distribution of $g(X,\hat{\underline{\eta}}_{\text{obs}})$ under $\underline{\eta}_0$, via $K(t,\hat{\underline{\eta}}_{\text{obs}})$, should be modified to give

(3.6.3)
$$Eg(X,\hat{\underline{\eta}}_{\text{obs}}) = \theta_0 - \hat{\theta}_{\text{obs}}.$$

What about the second cumulant determined by the estimated cumulant generating function $K(t,\hat{\underline{\eta}}_{\text{obs}})$? For fixed $\hat{\underline{\eta}}_{\text{obs}}$, the non-parametric estimate of the variance of $g(X,\hat{\underline{\eta}}_{\text{obs}})$ is a sample mean, $\frac{1}{n}\sum g(x_i,\hat{\underline{\eta}}_{\text{obs}})^2$, and estimates $\text{Var}_{\underline{\eta}_0} g(X,\hat{\underline{\eta}}_{\text{obs}})$

with error $O_P(\frac{1}{\sqrt{n}})$, (see Corollary 3.2.2). For the parametric case, equation (3.6.1) implies that

$$E_{\underline{\eta}_0} g^2(X, \underline{\hat{\eta}}_{\rm obs}) = E_{\underline{\eta}_0} g^2(X, \underline{\eta}_0) + O|\underline{\hat{\eta}}_{\rm obs} - \underline{\eta}_0|$$

and hence

$$\mathrm{Var}_{\underline{\eta}_0} g(X, \underline{\hat{\eta}}_{\rm obs}) = \mathrm{Var}_{\underline{\eta}_0} g(X, \underline{\eta}_0) + O|\underline{\hat{\eta}}_{\rm obs} - \underline{\eta}_0|.$$

The assumptions of Section 3.3 guarantee that $\frac{\partial}{\partial \underline{\eta}} \mathrm{Var}_{\underline{\eta}} g(X, \underline{\eta})$ is bounded on compact subsets of $\mathcal{N}$, and hence

$$\mathrm{Var}_{\underline{\eta}_0} g(X, \underline{\hat{\eta}}_{\rm obs}) = \mathrm{Var}_{\underline{\hat{\eta}}_{\rm obs}} g(X, \underline{\hat{\eta}}_{\rm obs}) + O|\underline{\hat{\eta}}_{\rm obs} - \underline{\eta}_0|,$$

so that our initial estimate of the distribution of $g(X, \underline{\hat{\eta}}_{\rm obs})$ has error $O_P(\frac{1}{\sqrt{n}})$ for the second cumulant. For fixed $\underline{\hat{\eta}}_{\rm obs}$, the assumptions of Section 3.3 guarantee that the first four cumulants of $g(X, \underline{\hat{\eta}}_{\rm obs})$, under $\underline{\eta}_0$, are finite. The above argument can be extended to show that $K(t, \underline{\hat{\eta}})$ estimates each cumulant of $g(X, \underline{\hat{\eta}})$, under $\underline{\eta}_0$, with error $O_P(\frac{1}{\sqrt{n}})$.

We also remark that our regularity assumptions imply that, on compact subsets of $\mathcal{N}$, $\mathrm{Var}_{\underline{\eta}_0} g(X, \underline{\hat{\eta}}_{\rm obs})$ is bounded away from zero (so we can divide by the standard deviation of $g(X, \underline{\hat{\eta}}_{\rm obs})$ in the order arguments which we now present).

We now *re-estimate* the distribution of $g(X, \underline{\hat{\eta}}_{\rm obs})$, (under the unknown $\underline{\eta}_0$, assumed to satisfy $H_0: \theta(\underline{\eta}_0) = \theta_0$) by applying an exponential tilt to the distribution $K(t, \underline{\hat{\eta}}_{\rm obs})$. The exponentially tilted distribution has cumulant generating function

$$\begin{aligned} K_{\alpha_0}(t) &= K(\alpha_0 + t; \underline{\hat{\eta}}_{\rm obs}) - K(\alpha_0; \underline{\hat{\eta}}_{\rm obs}) \\ &= \sum_{j=1}^{\infty} K^{(j)}(\alpha_0) \frac{t^j}{j!} \qquad (3.6.4) \\ &= (\theta_0 - \hat{\theta}_{\rm obs})t + \sum_{j=2}^{\infty} \kappa_{\alpha_0 j} \frac{t^j}{j!}, \end{aligned}$$

where α_0 satisfies the shift condition

$$K'(\alpha_0; \underline{\hat{\eta}}_{\rm obs}) = \theta_0 - \hat{\theta}_{\rm obs}.$$

Then $K_{\alpha_0}(t)$ estimates the first cumulant of $g(X, \hat{\underline{\eta}}_{\text{obs}})$, under H_0, with error $O_P(\frac{1}{n})$, (see equation 3.6.2). If the score functions $\underline{\Psi}$ are bounded, then all cumulants of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ are finite. In general, the infinite series of (3.6.4) are truncated, at some $j \geq 4$ (and the last term involves $K^{(j)}(\xi)$ for ξ between 0 and t).

Let $\hat{\kappa}_j = K^{(j)}(0; \hat{\underline{\eta}}_{\text{obs}})$, the initial estimate of the j^{th} cumulant, κ_j, of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ under $\underline{\eta}_0$. Then α_0 solves, approximately,

$$\begin{aligned} \hat{\kappa}_1 + \alpha_0 \hat{\kappa}_2 &\doteq \theta_0 - \hat{\theta}_{\text{obs}} \\ \alpha_0 &\doteq \frac{(\theta_0 - \hat{\theta}_{\text{obs}} - \kappa_1) + (\kappa_1 - \hat{\kappa}_1)}{\hat{\kappa}_2} \\ &\doteq O_P(\tfrac{1}{\sqrt{n}}). \end{aligned}$$

Then, for $j \geq 2$, we have from Section 1.5,

$$\begin{aligned} \kappa_{\alpha_0 j} &\doteq \hat{\kappa}_j + \alpha_0 \hat{\kappa}_{j+1} \\ &= \kappa_j + O_P(\tfrac{1}{\sqrt{n}}), \end{aligned}$$

so that $\kappa_{\alpha_0 j}$ has error no bigger than that of $\hat{\kappa}_j$.

Thus, having observed $\hat{\underline{\eta}}_{\text{obs}}$, we hold it fixed and estimate the distribution of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ under $\underline{\eta}_0$, where $\underline{\eta}_0$ is unknown but assumed to satisfy $H_0 \colon \theta(\underline{\eta}_0) = \theta_0$. The cumulant generating function of $g(X, \hat{\eta})$ is estimated by $K_{\alpha_0}(t)$, with error $O_P(\frac{1}{n})$ in the first cumulant, and error $O_P(\frac{1}{\sqrt{n}})$ in subsequent cumulants. We then estimate the distribution of the *average* of n_1 observations, $\frac{1}{n_1} \sum_{i=1}^{n_1} g(X_i, \hat{\underline{\eta}}_{\text{obs}})$, via the cumulant generating function

$$\bar{K}_{\alpha_0}(t) = (\theta_0 - \hat{\theta}_{\text{obs}})t + \sum_{j=2}^{\infty} \kappa_{\alpha_0 j} \left(\frac{t}{\sqrt{n_1}}\right)^j \frac{1}{j!}, \tag{3.6.5}$$

where the first term has error $O_P(\frac{1}{n})$, and the term in t^j, $j \geq 2$, has order $n_1^{-j/2} O_P(\frac{1}{\sqrt{n}})$, and is $O_P(\frac{1}{n})$ if $n_1 \doteq O(n)$, $n_1 \neq o(n)$ (see below). Then, under $\bar{K}_{\alpha_0}(t)$, we calculate the probability of observing $\bar{g}(X, \hat{\eta}_{\text{obs}})$ more extreme than that value actually observed:

$$\bar{g}_{\text{obs}} = \frac{1}{n} \sum_{i=1}^{n} g(x_i, \hat{\underline{\eta}}_{\text{obs}}) = 0.$$

Since we have estimated the cumulant generating function of $\bar{g}$ with error $O_P(\frac{1}{n})$, the tail areas $P(\bar{g} < 0)$ are also estimated with error $O_P(\frac{1}{n})$. One way to see this is to appeal to the Fisher-Cornish expansion. Let μ, σ, γ be the mean, standard deviation, and coefficient of skewness for $g(X, \hat{\eta}_{\text{obs}})$ under $\underline{\eta}_0$, assumed to satisfy $H_0\colon \theta(\underline{\eta}_0) = \theta_0$, and let $\hat{\mu}$, $\hat{\sigma}$, $\hat{\gamma}$ be the estimates obtained under $K_{\alpha_0}(t)$. That is, $\hat{\mu} = K'_{\alpha_0}(0)$, etc. Then the Fisher-Cornish expansion gives

$$
\begin{aligned}
P_{\underline{\eta}_0}(\bar{g} < 0) &= P_{\underline{\eta}_0}\left(\frac{\bar{g}-\mu}{\sigma} < \frac{-\mu}{\sigma}\right) \\
&= \bar{\Phi}\left(\frac{-\mu}{\sigma}\right) - \varphi\left(\frac{-\mu}{\sigma}\right)\frac{\gamma}{6\sqrt{n}}\left(\left(\frac{\mu}{\sigma}\right)^2 - 1\right) + O\left(\tfrac{1}{n}\right) \\
&= \bar{\Phi}\left(\frac{-\hat{\mu}}{\hat{\sigma}} + O_P\left(\tfrac{1}{n}\right)\right) \\
&\qquad - \varphi\left(\frac{-\hat{\mu}}{\hat{\sigma}} + O_P\left(\tfrac{1}{n}\right)\right)\frac{\hat{\gamma} + O_P\left(\frac{1}{\sqrt{n}}\right)}{6\sqrt{n}}\left(\left(\frac{\hat{\mu}}{\hat{\sigma}}\right)^2 + O_P\left(\tfrac{1}{n}\right) - 1\right) \\
&\qquad + O\left(\tfrac{1}{n}\right) \\
&= P_{K_{\alpha_0}}(\bar{g} < 0) + O_P\left(\tfrac{1}{n}\right).
\end{aligned}
$$

We have used $\hat{\mu} = O_P(\frac{1}{\sqrt{n}})$, $\hat{\mu} - \mu = O_P(\frac{1}{n})$, and $\hat{\sigma} - \sigma = O_P(\frac{1}{\sqrt{n}})$.

For an observed sample $x_1, \ldots, x_n$, with M-estimates $\hat{\underline{\eta}}_{\text{obs}}$, $\hat{\theta}_{\text{obs}}$, the $(1 - 2\varepsilon)100\%$ confidence interval for θ is then

$$
\{\theta_0 | \varepsilon < P_{\theta_0}(\bar{g} < 0) < 1 - \varepsilon\}, \tag{3.6.6}
$$

where P_{θ_0} denotes probability under $\underline{\eta}_0$, unknown but assumed to satisfy H_0: $\theta(\underline{\eta}_0) = \theta_0$. The above interval is estimated by

$$
\{\theta_0 | \varepsilon < \hat{P}_{\theta_0}(\bar{g} < 0) < 1 - \varepsilon\}, \tag{3.6.7}
$$

where $\hat{P}_{\theta_0}$ denotes probability under the model which gives $g(X, \hat{\underline{\eta}}_{\text{obs}})$ the cumulant generating function $K_{\alpha_0}(t)$, where $\alpha = \alpha_0$ solves $K'(\alpha; \hat{\underline{\eta}}_{\text{obs}}) = \theta_0 - \hat{\theta}_{\text{obs}}$.

Theorem 3.6.1 *The interval* (3.6.7) *approximates the interval* (3.6.6) *with error* $O_P(\frac{1}{n})$ *on compact subsets of* $\mathcal{N}$.

Proof: We show that if θ_1 satisfies $P_{\theta_1}(\bar{g} < 0) = \varepsilon$ and θ_2 satisfies $\hat{P}_{\theta_2}(\bar{g} < 0) = \varepsilon$, then $|\theta_1 - \theta_2| = O_P(\frac{1}{n})$.

Let σ_j^2 be the variance of $g(X, \hat{\underline{\eta}}_{\text{obs}})$ under $\underline{\eta}_j$ unknown, but assumed to satisfy $\theta(\underline{\eta}_j) = \theta_j$. The regularity assumptions of Section 3.3 imply that, on compact subsets of $\mathcal{N}$,

$$\sigma_2 = \sigma_1 + O|\underline{\eta}_2 - \underline{\eta}_1|,$$

$$\underline{\eta}_2 - \underline{\eta}_1 = O|\theta_2 - \theta_1|,$$

and hence $\sigma_2 = \sigma_1 + O|\theta_2 - \theta_1|$. Similarly, $\gamma_2 = \gamma_1 + O|\theta_2 - \theta_1|$. We also observe that the assumptions of Section 3.3 are sufficient to guarantee that σ_j is bounded away form zero on compact subsets of $\mathcal{N}$ (since the matrix B is continuous and has non-zero eigenvalues).

We appeal, once again, to the Fisher-Cornish expansion

$$\begin{aligned}
P_{\theta_1}(\bar{g} < 0) = \Phi&\Big(-\frac{\theta_1 - \hat{\theta}_{\text{obs}}}{\sigma_1} + O_P(\tfrac{1}{n})\Big) \\
&- \varphi\Big(-\frac{\theta_1 - \hat{\theta}_{\text{obs}}}{\sigma_1} + O_P(\tfrac{1}{n})\Big)\frac{\gamma_1}{6\sqrt{n}}\Big(\big(\frac{\theta_1 - \hat{\theta}_{\text{obs}}}{\sigma_1}\big)^2 + O_P(\tfrac{1}{n}) - 1\Big) \\
&\quad + O(\tfrac{1}{n}) \\
= \Phi&\Big(-\frac{\theta_2 - \hat{\theta}_{\text{obs}}}{\sigma_2} + O|\theta_1 - \theta_2| + O_P(\tfrac{1}{n})\Big) \\
- \varphi&\Big(-\frac{\theta_2 - \hat{\theta}_{\text{obs}}}{\sigma_2} + O|\theta_1 - \theta_2| + O_P(\tfrac{1}{n})\Big)\frac{\gamma_2 + O|\theta_1 - \theta_2|}{6\sqrt{n}}\Big(\big(\frac{\theta_2 - \hat{\theta}_{\text{obs}}}{\sigma_2}\big)^2 \\
&\quad + O|\theta_1 - \theta_2| + O_P(\tfrac{1}{n}) - 1\Big) + O(\tfrac{1}{n}) \\
= \Phi&\Big(-\frac{\theta_2 - \hat{\theta}_{\text{obs}}}{\sigma_2}\Big) - \varphi\Big(-\frac{\theta_2 - \hat{\theta}_{\text{obs}}}{\sigma_2}\Big)\frac{\gamma_2}{6\sqrt{n}}\Big(\big(\frac{\theta_2 - \hat{\theta}_{\text{obs}}}{\sigma_2}\big)^2 - 1\Big) \\
&\quad + O|\theta_1 - \theta_2| + O_P(\tfrac{1}{n}) \\
= \hat{P}&_{\theta_2}(\bar{g} < 0) + O|\theta_1 - \theta_2| + O_P(\tfrac{1}{n})
\end{aligned}$$

That is,

$$\varepsilon = \varepsilon + O|\theta_1 - \theta_2| + O_P(\tfrac{1}{n})$$

$$\theta_1 - \theta_2 = O_P(\tfrac{1}{n}). \qquad \square$$

The parameter value θ_0 is included in the confidence interval if and only if the tail area is not too large or too small. Tail areas are estimated via the Lugannani-Rice approximation, which does not increase the error (see 2.2.13). Theorem 3.6.1 establishes that the confidence intervals for θ have error $O_P(\frac{1}{n})$ on compact subsets of $\mathcal{N}$ *provided* the effective sample size, n_1, satsifies $n_1 = O(n)$, $n_1 \neq o(n)$.

There is, of course, room for debate in this setting. Is $\bar{g}_{\text{obs}}$ the average of n or $n-p$ or $n-(p-1)$ observations (where p is the dimension of $\underline{\eta}$)? That is, what is the value of the effective sample size n_1? We have used several values of n_1 in our programs. Evidence suggests that a value smaller than $n-(p-1)$ is more appropriate. See Huber (1981), page 150. There is also evidence that the intervals can be improved by slightly "relaxing" the shift condition. That is to say, if we replace the condition $K'(\alpha_0) = \theta_0 - \hat{\theta}_{\text{obs}}$ by

$$K'(\alpha_0) = \theta_0 - \hat{\theta}_{\text{obs}} + \text{ correction},$$

where the correction is $o_P(\frac{1}{n})$ and is chosen so that $|\alpha_0|$ is reduced slightly, then our confidence intervals perform better. The correction is discussed further in the example of Chapter 4.

The ideas in the above presentation are discussed at some length in the next section. We close this section with a remark about the practical application of these results.

We go back to the definition,

$$g(X, \underline{\hat{\eta}}_{\text{obs}}) = \underline{\Psi}^T(X, \underline{\hat{\eta}}_{\text{obs}}) B^T \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}.$$

In those cases where θ is a *linear* function of $\underline{\eta}$, $\frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}$ is *not* a function of $\underline{\eta}_0$. Generally, both B and $\frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}$ are functions of the unknown parameter $\underline{\eta}_0$, and must be estimated from the sample. Suppose we take the obvious sample estimates

$$\hat{A}(\hat{\underline{\eta}}_{\mathrm{obs}}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \underline{\Psi}}{\partial \underline{\eta}^T}(x_i, \hat{\underline{\eta}}_{\mathrm{obs}})$$
$$\hat{B}(\hat{\underline{\eta}}_{\mathrm{obs}}) = -\hat{A}(\hat{\underline{\eta}}_{\mathrm{obs}})^{-1}.$$

Then we work with an estimated configuration

$$\tilde{g}_i = \tilde{g}(x_i, \hat{\underline{\eta}}_{\mathrm{obs}}) = \underline{\Psi}^T(x_i, \hat{\underline{\eta}}_{\mathrm{obs}})\hat{B}(\hat{\underline{\eta}}_{\mathrm{obs}})^T \frac{\partial \theta(\hat{\underline{\eta}}_{\mathrm{obs}})}{\partial \underline{\eta}}.$$

The first estimate of the distribution of the $\tilde{g}_i$ is calculated as before. For example, for the non-parametric model, we take

$$\exp\{\tilde{K}(t; \hat{\underline{\eta}}_{\mathrm{obs}})\} = \frac{1}{n}\sum_{i=1}^{n}\exp\{t\tilde{g}_i\}.$$

Next, we need a good estimate of $E_{\underline{\eta}_0}\tilde{g}(X, \hat{\underline{\eta}}_{\mathrm{obs}})$. We need an argument like that given below equation (3.6.1). First notice that

$$\hat{A}(\underline{\eta}_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \underline{\Psi}}{\partial \underline{\eta}}(x_i, \underline{\eta}_0) = A + O_P(\tfrac{1}{\sqrt{n}})$$

and similarly

$$\hat{B}(\underline{\eta}_0) = B + O_P(\tfrac{1}{\sqrt{n}}),$$

on compact subsets of $\mathcal{N}$. Then

$$\tilde{g}(X, \hat{\underline{\eta}}_{\mathrm{obs}}) = \tilde{g}(X, \underline{\eta}_0) + (\hat{\underline{\eta}}_{\mathrm{obs}} - \underline{\eta}_0)^T \frac{\partial \tilde{g}}{\partial \underline{\eta}}(X, \underline{\eta}_0) + O|\hat{\underline{\eta}}_{\mathrm{obs}} - \underline{\eta}_0|^2, \tag{3.6.8}$$

where

$$\tilde{g}(X, \underline{\eta}_0) = \underline{\Psi}(X, \underline{\eta}_0)\hat{B}(\underline{\eta}_0)^T \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}.$$

Then, as before,

$$\frac{\partial \tilde{g}}{\partial \underline{\eta}}(X, \underline{\eta}_0) = \left[\frac{\partial \underline{\Psi}}{\partial \underline{\eta}^T}(X, \underline{\eta}_0)\right]^T \left[B + O_P(\tfrac{1}{\sqrt{n}})\right]^T \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}$$
$$+ \left[\frac{\partial}{\partial \underline{\eta}^T}\left(\hat{B}(\underline{\eta}_0)^T \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}\right)\right]^T \underline{\Psi}(X, \underline{\eta}_0)$$

so that

$$E_{\underline{\eta}_0} \frac{\partial \tilde{g}(X, \underline{\eta}_0)}{\partial \underline{\eta}} = \frac{-\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}} + O_P(\tfrac{1}{\sqrt{n}}) \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}}.$$

Then, taking expectations in (3.6.8), remembering that $\hat{\underline{\eta}}_{\text{obs}}$ is held fixed,

$$\begin{aligned} E_{\underline{\eta}_0} \tilde{g}(X, \hat{\underline{\eta}}_{\text{obs}}) &= 0 - (\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0)^T \frac{\partial \theta(\underline{\eta}_0)}{\partial \underline{\eta}} \\ &\quad + O_P(\tfrac{1}{\sqrt{n}}) O |\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0| + O |\hat{\underline{\eta}}_{\text{obs}} - \underline{\eta}_0|^2 \\ &= \theta(\underline{\eta}_0) - \hat{\theta}_{\text{obs}} + O_P(\tfrac{1}{n}), \end{aligned}$$

and the rest follows as before.

The above equation was obtained using non-parametric estimates of A and B. Surely, for the parametric model, with parametric estimates of A and B, the error is not increased. In the sequel, we do not distinguish between $\tilde{g}$ and g.

3.7 Discussion

We return now to some ideas introduced in Chapter 1. Stein's least favourable distribution is defined at the end of Section 1.2, and Kullback's result on the optimality of exponential tilts is stated in Theorem 1.3.1 and restated in Theorem 1.4.3.

Before applying the exponential tilt, we move from the original sample $X_1, \ldots, X_n$ to a modified sample, the configuration $g_i = g(X_i, \hat{\underline{\eta}})$. The distribution of the configuration is described by a one-parameter exponential family,

$$h^*(g; \alpha) = e^{\alpha g - K(\alpha, \hat{\underline{\eta}}_{\text{obs}})}\, d\mu(g; \hat{\underline{\eta}}_{\text{obs}}), \tag{3.7.1}$$

where $\mu(g; \hat{\underline{\eta}}_{\text{obs}})$ is our initial guess at the distribution of the g_i. We repeat that, in our programs, $\mu(g; \hat{\underline{\eta}}_{\text{obs}})$ is the non-parametric estimate under which g has cumulant generating function $K(\alpha; \hat{\underline{\eta}}_{\text{obs}})$,

$$e^{K(\alpha; \hat{\underline{\eta}}_{\text{obs}})} = \frac{1}{n} \sum_{i=1}^{n} e^{\alpha g(x_i, \hat{\underline{\eta}}_{\text{obs}})}. \tag{3.7.2}$$

If the statistic $g(X, \underline{\eta})$ is sufficient for $\underline{\eta}$, then there is no loss of information. In general, however, there is a loss of information when we move from the sample to the configuration. In order to discuss this idea further, we need to introduce some notation (see Kullback, 1960).

Suppose the observations X_i take values in a set $\mathcal{X}$ and, for each fixed $\underline{\eta}$, suppose that the transformation $T: X \mapsto g(X, \underline{\eta})$ is measurable, from the measure space $(\mathcal{X}, \mathcal{S}, \lambda)$ to a measure space $(\mathcal{Y}, \mathcal{T}, \mu)$. As in Section 1.3, let $f_1(x)$ and $f_2(x)$ be generalized densities on the measure space $(\mathcal{X}, \mathcal{S}, \lambda)$ giving, under the transformation T, the generalized densities $h_1(g)$, $h_2(g)$ respectively for $g = T(X)$.

A measurable transformation $S: \mathcal{X} \to \mathcal{Y}$ is sufficient for discrimination between f_1 and f_2 if

$$\frac{f_1(x)}{f_2(x)} = \frac{h_1(S(x))}{h_2(S(x))}\ [\lambda], \tag{3.7.3}$$

where $[\lambda]$ means "except on a set of λ-measure zero." Note that this definition includes the usual definition when f_1 and f_2 are of the same functional form, but with different values of a parameter.

We recall the definition of Kullback-Leibler distance (Section 1.3):

$$I(1:2;\mathcal{X}) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)}\, d\lambda(x). \tag{3.7.4}$$

Theorem 3.7.1 (Kullback, 1960, page 21) *$I(1:2;\mathcal{X}) \geq I(1:2;\mathcal{Y})$, with equality if and only if T is sufficient for discrimination between f_1 and f_2.*

Suppose we have a parametric model, with $f_2(x) = f(x;\underline{\eta})$ and $f_1(x) = f(x;\underline{\eta}+\Delta\underline{\eta})$, where $\underline{\eta}$ is a fixed value of the parameter and $\Delta\underline{\eta}$ is assumed "small". Let $\mathcal{I}(\underline{\eta})$ be the information under $\underline{\eta}$. That is, $\mathcal{I}(\underline{\eta})$ is the expected value under $f(x;\underline{\eta})$ of $\frac{-\partial^2 \log f(X;\underline{\eta})}{\partial\underline{\eta}\ \partial\underline{\eta}^T}$. The next result follows immediately from Kullback (1960, page 28).

Theorem 3.7.2 (Kullback)

$$I(\underline{\eta}+\Delta\underline{\eta}:\underline{\eta};\mathcal{X}) = \frac{1}{2}(\Delta\underline{\eta})^T \mathcal{I}(\underline{\eta})\Delta\underline{\eta} + O|\Delta\underline{\eta}|^3. \tag{3.7.5}$$

In our applications, the first estimate of the distribution of the configuration g, is via the empirical cumulant generating function $K(\alpha;\hat{\underline{\eta}}_{\text{obs}})$ given in (3.7.2). This estimate has generalized density $h_2(g;\hat{\underline{\eta}}_{\text{obs}})$, say. Then the tilted distribution defines a one-paramter exponential family $h^*(g;\alpha_0)$, where

$$\begin{aligned} h^*(g;\alpha_0) &\stackrel{\text{d}}{=} h_1(g;\hat{\underline{\eta}}_{\text{obs}}) \\ &= e^{\alpha_0 g - K(\alpha_0;\hat{\underline{\eta}}_{\text{obs}})} h_2(g;\hat{\underline{\eta}}_{\text{obs}}). \end{aligned} \tag{3.7.6}$$

Note that $E_{h_2}(g) = 0$ and $E_{h_1}(g) = K'(\alpha_0,\hat{\underline{\eta}}_{\text{obs}}) = \theta_0 - \hat{\theta}_{\text{obs}}$. We define

$$\begin{aligned} \Delta\theta_0 &= E_{h_1}(g) - E_{h_2}(g) \\ &= \theta_0 - \hat{\theta}_{\text{obs}}. \end{aligned} \tag{3.7.7}$$

Under the one-parameter model $h^*(g;\alpha)$, the derivative (with respect to α) of the log-likelihood is

$$\frac{\partial \log h^*(g;\alpha)}{\partial \alpha} = g - K'(\alpha;\hat{\underline{\eta}}_{\text{obs}}),$$

and $\text{Var}_\alpha(g) = K''(\alpha,\hat{\underline{\eta}}_{\text{obs}})$ is the information for α in one observation from the model $h^*(g;\alpha)$. The information for θ is then $\frac{d\alpha}{d\theta}\text{Var}_\alpha(g)\frac{d\alpha}{d\theta}$. Differentiation of $K'(\alpha;\hat{\underline{\eta}}_{\text{obs}}) = \theta_0 - \hat{\theta}_{\text{obs}}$ yields $\frac{d\alpha}{d\theta} = 1/K''(\alpha;\hat{\underline{\eta}}_{\text{obs}})$ and hence the information for θ_0 in one observation from $h^*(g;\alpha_0)$ is

$$I(\theta_0) = 1/K''(\alpha_0;\hat{\eta}_{\text{obs}}).$$

Then, Theorem 3.7.2 gives

$$\begin{aligned} I(\theta_0 + \Delta\theta_0 : \theta_0; Y) &\doteq \frac{(\theta_0 - \hat{\theta}_{\text{obs}})^2}{2K''(\alpha_0;\hat{\underline{\eta}}_{\text{obs}})} \\ &= \frac{(\theta_0 - \hat{\theta}_{\text{obs}})^2}{2\text{Var}_{\alpha_0}(g)}. \end{aligned}$$

In the light of these remarks and Theorem 1.3.1, the following Theorem is not surprising. We need one more piece of notation: $\alpha = \alpha(\xi)$ is the solution of $K'(\alpha) = \xi$.

Theorem 3.7.3 (Kullback, 1960, page 47).

$$I(1{:}2;X) \geq I(1{:}2;Y) \geq \frac{(\theta_0 - \hat{\theta}_{\text{obs}})^2}{2\text{Var}_{\alpha(\xi)}(g)} \tag{3.7.8}$$

where ξ lies between 0 and $\theta_0 - \hat{\theta}_{\text{obs}}$, with equality between the first pair if and only if $g = T(X;\hat{\underline{\eta}}_{\text{obs}})$ is sufficient for discrimination between f_1 and f_2, and with equality between the second pair if and only if

$$h_1(g;\hat{\underline{\eta}}_{\text{obs}}) = e^{\alpha g - K(\alpha,\hat{\underline{\eta}}_{\text{obs}})} h_2(g;\hat{\underline{\eta}}_{\text{obs}}).$$

Generally speaking, Theorem 3.7.2 leads us to think of $2I(1:2)$ as a measure of the information available as we move away from model 2 in the direction of model 1. Stein's least favourable distribution is characterized by the property that it gives the direction in which information does not change. All other directions produce an increase in information, which Stein considers artificial since one should not expect to increase information once the sample is taken. In our applications, the model 1 satisfies an extra condition, $E(g) = \theta_0 - \hat{\theta}_{\text{obs}}$. That is, *we add more information.*

Theorem 3.7.3 states that our tilting procedure gives equality on the right of (3.7.8) and that there is equality on the left if and only if the configuration is sufficient for the original sample. We believe that, to improve the intervals we have constructed, we need to look at the first inequality of Theorem 3.7.3. We need a notion of "approximate sufficiency." Some configurations are, surely, "more sufficient" than others. We do not investigate these ideas further in the present paper.

The distribution of the $g(X, \hat{\underline{\eta}}_{\text{obs}})$ is first estimated via $K(\alpha; \hat{\underline{\eta}}_{\text{obs}})$ and then tilted to force $E(\vartheta) = \theta_0 - \hat{\theta}_{\text{obs}}$ (plus a correction, see Chapter 4). With respect to the measure $\mu(g; \hat{\underline{\eta}}_{\text{obs}})$ determined by $K(\alpha; \hat{\underline{\eta}}_{\text{obs}})$, the tilted distribution belongs to a one-parameter exponential family with density $h^*(g; \alpha)$, see (3.7.1). Let us hold the measure $\mu(g; \hat{\underline{\eta}}_{\text{obs}})$ fixed and consider the one-parameter exponential family

$$h^*(g; \alpha) = e^{\alpha g - K(\alpha, \hat{\underline{\eta}}_{\text{obs}})} d\mu(g; \hat{\underline{\eta}}_{\text{obs}}). \tag{3.7.9}$$

Suppose we wish to test the null hypothesis $H_0: \theta(\underline{\eta}_0) = \theta_0$. Then, in Section 3.6, we reduced this hypothesis to $H_0: E(g(X, \hat{\underline{\eta}}_{\text{obs}})) = \theta_0 - \hat{\theta}_{\text{obs}}$, with error $O_P(\frac{1}{n})$. We now go one step further. Let α_0 be the solution of $K'(\alpha; \hat{\underline{\eta}}_{\text{obs}}) = \theta_0 - \hat{\theta}_{\text{obs}}$. Then the hypothesis that $\tilde{E}(g) = \theta_0 - \hat{\theta}_{\text{obs}}$ is equivalent (to order $O_P(\frac{1}{n})$) to the hypothesis that, in the one-parameter exponential family described by (3.7.9), α has the true value α_0. That is $\tilde{H}_0: \alpha = \alpha_0$.

In this context of a one parameter exponential family, a uniformly most powerful test of $\tilde{H}_0: \alpha = \alpha_0$ versus $\tilde{H}_1: \alpha > \alpha_0$ is of the form reject $\tilde{H}_0$ if $\bar{g}_{\text{obs}}$ exceeds a critical value, where the critical point is determined under

$$h^*(g; \alpha_0) = e^{\alpha_0 g - K(\alpha_0; \hat{\underline{\eta}}_{\text{obs}})} d\mu(g; \hat{\underline{\eta}}_{\text{obs}}).$$

(Cox and Hinkley, 1974, page 94). Similarly, a uniformly most powerful test of $\tilde{H}_0$ versus $\tilde{H}_1: \alpha < \alpha_0$ is to reject $\tilde{H}_0$ if $P_{\alpha_0}(\bar{g} \leq \bar{g}_{\text{obs}})$ is too small.

The $(1-2\varepsilon)100\%$ confidence interval for θ, constructed in this chapter, consists of all values θ_0 for which we estimate

$$\varepsilon < P_{\alpha_0}(\bar{g} \geq 0) < 1 - \varepsilon.$$

(Remember, $\bar{g}_{\text{obs}} = 0$). Thus, *conditional* on the measure determined by $K(\alpha; \hat{\underline{\eta}}_{\text{obs}})$ *and* the assumed one-parameter exponential family model, $h^*(g; \alpha)$, the confidence intervals we have constructed are optimal. The one parameter exponential family model is forced if we choose to minimize Kullback-Leibler distance.

For completeness, we should mention the role of density functions in our procedures. The Kullback results hold for generalized densities and make no assumptions about the continuity of the underlying cumulative distribution function. The Lugannani-Rice approximation is established for distributions having a true density function. As mentioned at the end of Section 2.2, this presents no problems in our applications. As a point of interest, Daniels (1986) gives a different form of the tail-area approximation for lattice distributions.

3.8 The general test procedure, in more detail

For an observed sample $x_1, \ldots, x_n$ we construct a $(1-2\varepsilon)100\%$ confidence interval for $\theta = \theta(\underline{\eta})$ as follows. First calculate the M-estimate $\underline{\hat{\eta}}$, solution of

$$\frac{1}{n}\sum_{i=1}^{n} \underline{\Psi}(x_i, \underline{\eta}) = 0,$$

and the estimates $\hat{\theta} = \theta(\underline{\hat{\eta}})$ and $\hat{B}$. We can calculate $\hat{B}$ parametrically, using $\underline{\hat{\eta}}$ to estimate $\underline{\eta}$, or non-parametrically, as minus the inverse of $\hat{A}$,

(3.8.1) $$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \underline{\Psi}}{\partial \underline{\eta}^T}(x_i, \underline{\hat{\eta}}).$$

Step 1. The configuration is constructed:

$$g_i = \underline{\Psi}^T(x_i, \underline{\hat{\eta}})\hat{B}^T \frac{\partial \theta(\underline{\hat{\eta}})}{\partial \underline{\eta}}.$$

The shift condition is

$$E(g) = \theta_0 - \hat{\theta},$$

for each θ_0 under test.

Step 2. In our programs we have used exclusively the non-parametric first estimate of the distribution of the g_i, via

$$\exp K(t) = \frac{1}{n}\sum_{i=1}^{n} \exp(t g_i).$$

Note that

$$K'(t) = \sum_{i=1}^{n} g_i \exp(t g_i) \Big/ \sum_{i=1}^{n} \exp(t g_i).$$

Steps 1 and 2 are needed once for each sample. Then, for each θ_0 under test, we repeat the remaining steps.

Step 3. Let $t = \alpha_0$ be the solution of

$$K'(t) = \theta_0 - \hat{\theta}.$$

Then the improved estimate of the distribution of $\bar{g}$ is as the average of $n_1 = O(n)$ observations from a distribution with cumulant generating function

$$K_{\alpha_0}(t) = K(\alpha_0 + t) - K(\alpha_0),$$

but this modification does not complicate the final step of the procedure.

Step 4. Apply the Lugannani-Rice approximation of Chapter 2, noting that $\bar{g}_{\text{obs}} = 0$. Then equation (2.2.13) gives

$$(3.8.2) \quad P(\bar{g} > 0) \doteq 1 - \Phi(\sqrt{-2f_0}) + \frac{\exp(f_0)}{\sqrt{2\pi}} \left\{ \frac{1}{t_0\sqrt{n_1 K''_{\alpha_0}(t_0)}} - \frac{1}{\sqrt{-2f_0}} \right\},$$

where t_0 solves $K'_{\alpha_0}(t_0) = 0$, n_1 is to be specified, $n_1 = O(n)$, and other symbols are defined below. Observe that $K'(0) = 0$ and therefore $t_0 = -\alpha_0$. Then $K''_{\alpha_0}(t_0) = K''(0)$ and

$$\begin{aligned} f_0 &= n_1(K_{\alpha_0}(t_0) - t_0\bar{g}_{\text{obs}}) \\ &= -n_1 K(\alpha_0), \end{aligned}$$

and the sign of $\sqrt{-2f_0}$ is taken to be minus the sign of α_0, so that the approximation (3.8.2) becomes (with $\text{sign}(\frac{1}{\sqrt{2n_1 K(\alpha_0)}}) = -\text{sign}(\alpha_0)$),

$$(3.8.3) \quad \begin{aligned} P(\bar{g} > 0) \doteq\ & \Phi(-\sqrt{2n_1 K(\alpha_0)}) \\ & - \frac{\exp(-n_1 K(\alpha_0))}{\sqrt{2\pi}} \left\{ \frac{1}{\alpha_0\sqrt{n_1 K''(0)}} + \frac{1}{\sqrt{2n_1 K(\alpha_0)}} \right\}. \end{aligned}$$

Finally, θ_0 is included in the interval if and only if the approximation (3.8.3) yields

$$(3.8.4) \qquad \varepsilon \le P(\bar{g} > 0) \le 1 - \varepsilon.$$

Our programs first calculate $\hat{\theta}$ and the configuration g_i. This first part of the program is problem-dependent, and is simpler for the location problem of Chapter

4 than for the regression coefficient of Chapter 5, say. Once $\hat{\theta}$ and g_i have been calculated, the second stage is the same for all problems.

At the second stage, the program systematically tries values θ_0 until two values are found, giving equality on either side of (3.8.4). These two values become the endpoints of the confidence interval.

To test a value θ_1, we first solve for $t = \alpha_0$,

$$\sum_{i=1}^{n} g_i \exp(tg_i) / \sum_{i=1}^{n} \exp(tg_i) = \hat{\theta}_0 - \hat{\theta},$$

then plug into (3.8.3) to estimate $P(\bar{g} > 0)$. If (3.8.4) is not satisfied, then try a new value θ_2, closer to the center of the interval. If (3.8.4) is satisfied with strict inequality, then try a value θ_3 further from the center of the interval. The center of the interval should be somewhere near $\hat{\theta}$. This point is discussed further in the examples.

David Andrews has noted that calculations are reduced if we *first* find α_1 and α_2 satisfying

$$P_{\alpha_1}(\bar{g} > 0) = \varepsilon \quad \text{and} \quad P_{\alpha_2}(\bar{g} > 0) = 1 - \varepsilon$$

and *then* calculate the endpoints of the interval for θ via

$$K'(\alpha_i) = \theta_i - \hat{\theta} + \textit{correction}, \qquad i = 1, 2.$$

In subsequent chapters, we apply this four-step procedure and compare its performance with the classical approaches. Methods based on the above procedure are denoted by R (R for Robust). Methods in which the matrix A is estimated from the sample (via (3.8.1)) are denoted RS (Robust and Sample based A-matrix). We present evidence that the RS methods are *not* robust. The effective sample size, n_1, is always specified explicitly; for example RS7 or R($n-2$), the true sample size, n, being clear from the context.

There is evidence that the shift condition $E(g) = \theta_0 - \hat{\theta}_{\text{obs}}$ should be relaxed to

$$E(g) = \theta_0 - \hat{\theta}_{\text{obs}} + \varepsilon_0,$$

where ε_0 is small and chosen so that $\theta_0 - \hat{\theta}_{\text{obs}} + \varepsilon_0$ is between $\theta_0 - \hat{\theta}_{\text{obs}}$ and $\bar{g}_{\text{obs}} = 0$. Theorem 1.4.2 implies that, if $K'(\alpha_0) = \theta_0 - \hat{\theta}_{\text{obs}}$ and $K'(\alpha_1) = \theta_0 - \hat{\theta}_{\text{obs}} + \varepsilon_0$, then $|\alpha_0| > |\alpha_1|$. The observed value $\bar{g}_{\text{obs}} = 0$ is closer to the mean of the cumulant generating function $K_{\alpha_1}(t)$ than to the mean of $K_{\alpha_0}(t)$, so the tail area $P_{\alpha_1}(\bar{g} > 0)$ is less extreme than the tail area $P_{\alpha_0}(\bar{g} > 0)$. Intervals calculated under a relaxed shift condition tend to be longer and to give better coverage. See the examples in subsequent chapters.

The shift correction we have used in Chapter 4 is a multiple of $\frac{1}{n}$ times an estimate of scale for $\hat{\theta}$. The shift correction used in Chapter 5 is $\frac{1}{n}$ times a multiple of $|\theta_0 - \hat{\theta}_{\text{obs}}|$. We have not been able to justify these corrections. We merely present evidence that they improve the intervals. When we relax the shift condition we are saying that the "minimum Kullback-Leibler distance" approach is too strong. Referring back to the discussion of the preceding section, and Theorem 3.7.3 in particular, a relaxation of the shift condition amounts to saying "we over-estimated the discrimination information". We should compensate in the direction of the least favourable distribution. The shift condition should be investigated further. In this regard, we observe that the bias correction of equation (3.3.4) should be used with caution. For small samples, attempts to relax the shift using sample-based estimates of the bias-correction have not been satisfactory. The problems are similar to those encountered with sample-based estimates of the matrix A. (See RS methods in Chapter 4). Methods using a relaxed shift condition are denoted by *corr*, for correction; for example *Rcorr4*.

Chapter 4. Confidence intervals for location, with unknown scale

We construct small sample confidence intervals using the techniques of Chapter 3. The intervals are compared with two variations on the classical approach.

4.1 Introduction to the problem

In this chapter, we use the techniques of Chapter 3 to construct small sample confidence intervals for location, when scale is unknown. We give several variations on the procedure outlined in Section 3.8. We compare these intervals with the classical t-interval, Method C, and with variations on the classical approach: two versions from Kent (1982), Methods K, and Huber's (1981) approximate t-interval, Method H. We use machine-generated data: standard normal, student-t with three degrees of freedom, slash, and chi-squared with three degrees of freedom.

The classical confidence interval assumes the data is from a normal population and constructs a pivotal interval based on Student's t_{n-1}. The other methods all start with the M-estimates described below.

For the sample $X_1, \ldots, X_n$ from an unknown distribution, location μ and scale σ are estimated by the M-estimates $\hat{\mu}$, $\hat{\sigma}$, solutions of

$$\frac{1}{n}\sum_{i=1}^{n} \Psi(X_i, \eta) = 0$$

where $\underline{\eta} = (\mu, \sigma)^T$ and

$$
\begin{aligned}
\Psi_1(X, \underline{\eta}) &= \Psi_c\left(\frac{x-\mu}{\sigma}\right) \\
\Psi_2(X, \underline{\eta}) &= \Psi_c^2\left(\frac{x-\mu}{\sigma}\right) - \beta_L, \\
\Psi_c(y) &= \begin{cases} -c & \text{if } y \leq -c \\ y & \text{if } |y| < c \\ c & \text{if } y \geq c, \end{cases} \\
\beta_L &= \frac{n-1}{n}\beta \\
\beta &= \int_{-\infty}^{\infty} \Psi_c^2(y)\phi(y)\,dy \\
&= 1 - 2\big(c\phi(c) + (1-c^2)\Phi(-c)\big),
\end{aligned}
$$

and $\phi(y)$, $\Phi(y)$ are the density function and cumulative distribution, respectively, of a standard normal random variable. The above estimate, $\hat{\underline{\eta}}$, is referred to in the literature as Huber's Proposal 2. The constant c is usually between 1 and 2.

For normal data, the classical interval is optimal. It cannot be bettered. Our aim, then, is to construct confidence intervals which perform "almost" as well as the classical intervals for normal data, and better than the classical intervals for non-normal data. Our two measures of performance are length and coverage.

4.2 Robust small sample confidence intervals

With the notation of Chapter 3, we proceed as follows,

$$Z_1 = \Psi_c\left(\frac{X-\mu}{\sigma}\right)$$
$$Z_2 = \Psi_c^2\left(\frac{X-\mu}{\sigma}\right) - \beta_L$$
$$Z_{11} = \frac{-1}{\sigma} I_{\mathrm{trc}}\left(\frac{X-\mu}{\sigma}\right) \qquad Z_{12} = \frac{-1}{\sigma}\left(\frac{X-\mu}{\sigma}\right)_{\mathrm{trc}}$$
$$Z_{21} = \frac{-2}{\sigma}\left(\frac{X-\mu}{\sigma}\right)_{\mathrm{trc}} \qquad Z_{22} = \frac{-2}{\sigma}\left(\frac{X-\mu}{\sigma}\right)^2_{\mathrm{trc}},$$

where $I(y) \equiv 1$ and

$$f_{\mathrm{trc}}(y) = \begin{cases} f(y), & \text{if } |y| < c \\ 0, & \text{if } |y| \geq c. \end{cases}$$

For the R methods (*without* the S), we calculate the matrix A under the normal model:

$$A = E\begin{pmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{pmatrix} = \frac{-1}{\sigma}\begin{pmatrix} \delta & 0 \\ 0 & 2\varepsilon \end{pmatrix},$$

where

$$\delta = \Phi(c) - \Phi(-c) \qquad \text{and}$$
$$\varepsilon = \int_{-c}^{c} y^2 \phi(y)\,dy = 1 - 2\big(\Phi(-c) + c\phi(c)\big) = \beta - 2c^2\Phi(-c).$$

Then

$$B = -A^{-1} = \sigma\begin{pmatrix} \frac{1}{\delta} & 0 \\ 0 & \frac{1}{2\varepsilon} \end{pmatrix}$$

and we estimate B by $\hat{B}$,

$$\hat{B} = \hat{\sigma}\begin{pmatrix} \frac{1}{\delta} & 0 \\ 0 & \frac{1}{2\varepsilon} \end{pmatrix}.$$

The parameter vector is $\underline{\eta} = (\mu, \sigma)^T$ and the real-valued function of $\underline{\eta}$ is, in fact, linear:

$$\theta = (1, 0)\underline{\eta} = \mu.$$

The configuration is, from Section 3.8,

$$g_i = \underline{\Psi}^T(x_i, \underline{\hat{\eta}}) \hat{B}^T \frac{\partial \theta(\underline{\hat{\eta}})}{\partial \underline{\eta}} = \frac{\hat{\sigma}}{\delta} \Psi_c\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right),$$

and we work with the g_i, as described in Section 3.8. For each μ_0 under test, the shift is $\mu_0 - \hat{\mu}$.

For the RS methods (S for sample-based A), we calculate the matrix A non-parametrically,

$$\hat{A} = \frac{-1}{\hat{\sigma} n} \sum_{i=1}^{n} \begin{pmatrix} I_{\text{trc}}\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) & \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)_{\text{trc}} \\ 2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)_{\text{trc}} & 2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^2_{\text{trc}} \end{pmatrix}.$$

We calculate

$$\hat{A} = \frac{-1}{n\hat{\sigma}} \begin{pmatrix} a_{11} & a_{12} \\ 2a_{12} & 2a_{22} \end{pmatrix},$$

$$\hat{B} = \frac{\hat{\sigma}}{2a_{11}a_{22} - 2a_{12}^2} \begin{pmatrix} 2a_{22} & -a_{12} \\ -2a_{12} & a_{11} \end{pmatrix};$$

and

$$g_i = \frac{\hat{\sigma}}{a_{11}a_{22} - a_{12}^2} \left\{ a_{22} \Psi_c\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \frac{a_{12}}{2} \left[\Psi_c^2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \beta_L \right] \right\},$$

then proceed as before. The routine SETLOC (see back pocket) is programmed with the following notation.

$$A_1 = \sum_{i=1}^{n} I_{\text{trc}}\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)$$

$$A_2 = \sum_{i=1}^{n} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)_{\text{trc}}$$

$$A_3 = \sum_{i=1}^{n} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)^2_{\text{trc}}$$

$$g_i = \frac{\hat{\sigma} n}{A_1 A_3 - A_2^2} \left(A_3 \Psi_c\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \frac{A_2}{2} (\Psi_c^2 - \beta_L) \right).$$

At Step 4, application of the tail-area approximation, we must specify the *effective* sample size, n_1. The effect of various values of n_1 is discussed below, in Section 4.4.

For the R*corr* methods, the shift correction is a multiple of $\hat{\sigma}/(n\sqrt{n})$. That is, in the notation of Section 3.8, at Step 1, the shift condition is

$$\mathbb{E}g = \theta_0 - \hat{\theta}_{\mathrm{obs}} + \varepsilon_0,$$

where $|\varepsilon_0|$ is a multiple of $\hat{\sigma}/(n\sqrt{n})$ and $\varepsilon_0(\theta_0 - \hat{\theta}_{\mathrm{obs}}) < 0$.

4.3 Variations on the classical approach

The classical confidence interval, which is appropriate when the distribution of the X_i is approximately normal, is constructed as follows. We calculate

$$\bar{x} = \frac{1}{n}\sum x_i$$
$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

and $t_{n-1}(\alpha)$ such that the probability that a random variable with distribution Student-t on $n-1$ degrees of freedom exceed $t_{n-1}(\alpha)$ is α. Then the classical $(1-2\alpha)100\%$ confidence interval for μ, the expected value of any one of the observations, is

$$\bar{x} - t_{n-1}(\alpha)\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{n-1}(\alpha)\frac{s}{\sqrt{n}}.$$

The classical confidence interval is programmed as Method C.

The K methods come from Kent's (1982) paper, as discussed in Section 1.6. The M-estimate $\hat{\underline{\eta}} = (\hat{\mu}, \hat{\sigma})^T$ is the solution of the minimization problem

$$\frac{1}{n}\sum_{i=1}^{n} \tau(x_i; \mu, \sigma) = \min!$$

or, equivalently, the maximization problem obtained by considering $-\tau$, where

$$\tau(X; \mu, \sigma) = \rho(\frac{X-\mu}{\sigma})\sigma$$

$$\rho(y) = \begin{cases} \frac{1}{2}(y^2 + \beta_L) & \text{if } |y| < c \\ c|y| - \frac{1}{2}c^2 + \frac{1}{2}\beta_L & \text{if } |y| \geq c. \end{cases}$$

Notice that

$$\frac{\partial\tau(X;\mu,\sigma)}{\partial\mu} = -\Psi_c(\frac{X-\mu}{\sigma})$$

and

$$\frac{\partial\tau(X;\mu,\sigma)}{\partial\sigma} = -\frac{1}{2}\left[\Psi_c^2(\frac{X-\mu}{\sigma}) - \beta_L\right].$$

Then, in Kent's notation, the score functions (for the max problem) are

$$U_1(X, \underline{\eta}) = \Psi_c(\frac{X-\mu}{\sigma})$$
$$U_2(X, \underline{\eta}) = \frac{1}{2}[\Psi_c^2(\frac{X-\mu}{\sigma}) - \beta_L].$$

The matrix $A = E\frac{\partial \underline{U}}{\partial \underline{\eta}^T}$ is *not* the same as the matrix A of the preceding section. In this section, A is symmetric:

$$\frac{\partial \underline{U}}{\partial \underline{\eta}^T} = \frac{-1}{\sigma}\begin{pmatrix} I_{\mathrm{trc}}(\frac{X-\mu}{\sigma}) & (\frac{X-\mu}{\sigma})_{\mathrm{trc}} \\ (\frac{X-\mu}{\sigma})_{\mathrm{trc}} & (\frac{X-\mu}{\sigma})^2_{\mathrm{trc}} \end{pmatrix}$$

and

$$A = E(\frac{\partial \underline{U}}{\partial \underline{\eta}^T}) = \frac{-1}{\sigma}\begin{pmatrix} \delta & \xi \\ \xi & \varepsilon \end{pmatrix}.$$

If we assume that the X_i are drawn from a normally distributed population, then δ and ε are as defined in Section 4.2 and ξ is zero. To apply Kent's procedure, Theorem 1.6.1, we also need

$$C = E[\underline{U}(X, \underline{\eta})\underline{U}^T(X, \underline{\eta})] = \begin{pmatrix} \beta & \rho \\ \rho & \kappa \end{pmatrix}. \tag{4.3.1}$$

If we assume a normal model, then β is as defined in section 4 2, ρ is zero, and

$$\kappa = \frac{1}{4}E(\Psi_c^2(Z) - \beta_L)^2, \tag{4.3.2}$$

for Z a standard normal random variable. Finally, we need

$$\nu = (B_{11})^{-1}(BCB)_{11},$$

where $B = -A^{-1}$, and for a normal model, ν is simply $\frac{\sigma\beta}{\delta}$.

For each sample $X_1, \ldots, X_n$, the M-estimates $\hat{\mu}$ and $\hat{\sigma}$ are calculated as in Section 4.2. To test $H_0: \mu = \mu_0$, we then calculate the conditional estimate of scale, $\hat{\sigma}_0$, solution of

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_c^2(\frac{X_i - \mu_0}{\sigma}) - \beta_L = 0. \tag{4.3.3}$$

Theorem 1.6.1 then states that, under H_0,

$$(4.3.4) \qquad W = 2\{\sum_{i=1}^{n} \tau(X_i, \mu_0, \hat{\sigma}_0) - \sum_{i=1}^{n} \tau(X_i, \hat{\mu}, \hat{\sigma})\}$$

should be distributed as $\nu\chi_1^2$. The value μ_0 is included in the estimated $(1-2\alpha)100\%$ confidene interval for μ if and only if w_{obs}, the observed value of W, satisfies

$$(4.3.5) \qquad P(\nu\chi_1^2 > w_{obs}) > 2\alpha.$$

For the K methods (*not* KS), we assume that the data $X_1, \ldots, X_n$ is taken from a normal population, we estimate σ by $\hat{\sigma}$. Then the test statistic W of (4.3.4) is assumed to be $\nu\chi_1^2$, $\nu = \frac{\hat{\sigma}\beta}{\delta}$.

For the KS methods, the matrices A, B and C are estimated from the sample, giving

$$\hat{A} = \frac{-1}{\hat{\sigma}} \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix},$$

not the same as the matrix $\hat{A}$ of the previous section, and

$$\hat{B} = \frac{\hat{\sigma}}{a_{11}a_{22} - a_{12}^2} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{12} & a_{11} \end{pmatrix}.$$

Then

$$\hat{C} = \begin{pmatrix} \beta_L & \hat{\rho} \\ \hat{\rho} & \hat{\kappa} \end{pmatrix},$$

where ρ and κ are estimated from the sample:

$$\hat{\rho} = \frac{1}{2n} \sum_{i=1}^{n} \Psi_c\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) \left(\Psi_c^2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \beta_L\right),$$

$$\hat{\kappa} = \frac{1}{4n} \sum_{i=1}^{n} \left(\Psi_c^2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \beta_L\right)^2.$$

Then we estimate ν by

$$\hat{\nu} = \frac{\hat{\sigma}}{a_{22}(a_{11}a_{22} - a_{12}^2)} (a_{22}^2\beta_L - 2a_{22}a_{12}\hat{\rho} + a_{12}^2\hat{\kappa}).$$

Continuing the notation of the previous section, we write

$$A_4 = \sum_{i=1}^{n} \Psi_c\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right)\left[\Psi_c^2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \beta_L\right]$$

$$A_5 = \sum_{i=1}^{n}\left(\Psi_c^2\left(\frac{x_i - \hat{\mu}}{\hat{\sigma}}\right) - \beta_L\right)^2,$$

then

$$\hat{\nu} = \frac{\hat{\sigma}}{A_3(A_1 A_3 - A_2^2)}\left(nA_3^2\beta_L - A_2 A_3 A_4 + \tfrac{1}{4}A_2^2 A_5\right).$$

In the KS methods we estimate $\hat{\nu}$ once for each sample. For each μ_0 under test, the observed value of W is calculated, (4.3.4), and μ_0 is included in the confidence interval only when (4.3.5) is satisfied.

We repeat that we have tampered with Kent's paper. We have replaced score functions which were, for Kent, legitimate derivatives of log likelihoods, by score functions appropriate for Huber's "Proposal 2" for joint estimates of location and scale. In the K and KS methods, ν is estimated once for each sample. For each μ_0 under test, $\hat{\sigma}_0$ is then re-estimated.

In Method H we use the test-statistic suggested by Huber, see Section 1.6. Let

$$nD_n^2 = \frac{\frac{1}{n-1}\sum_{i=1}^{n} \Psi_c^2\left(\frac{x_i-\hat{\mu}}{\hat{\sigma}}\right)\hat{\sigma}^2}{\left[\frac{1}{n}\sum_{i=1}^{n} I_{trd}\left(\frac{x_i-\hat{\mu}}{\hat{\sigma}}\right)\right]^2} = \frac{\hat{\sigma}^2\beta}{\hat{\delta}^2}$$

and, for each μ_0 under test, reject μ_0 if $|t_{obs}| > t_{n-1}(\alpha)$, where $t_{obs} = \frac{\hat{\mu}-\mu_0}{D_n}$. The pivot gives estimated $(1 - 2\alpha)100\%$ confidence interval:

$$\hat{\mu} - D_n t_{n-1}(\alpha) < \mu < \hat{\mu} + D_n t_{n-1}(\alpha).$$

We program a modification of the above method, with $\hat{\delta}$ replaced by $\delta = \Phi(c) - \Phi(-c)$, for reasons discussed below (Section 4.4).

4.4 Numerical results

We now present results of Monte Carlo trials. All data was machine generated. Normal data was generated using the IMSL routine GGNML. Slash data was generated using GGNML and the Fortran RANF routine. A random variable X has the slash distribution if $X = Z/W$, where Z is standard normal and W is uniform on $(0,1)$. Slash data is interesting since the first moment does not exist, $E(|X|) = \infty$, but the distribution is symmetric about zero, so that zero is the central point for the distribution: the true value of the location parameter is zero. The Student-t and the χ^2 data was generated using the NAGLIB, G05DJF and G05DHF routines respectively. All results for 1000 samples of size 5 from standard normal refer to the *same* set of 1000 samples of size 5, etc.

We recall the methods used. Methods R denote the procedures outlined in Section 4.2. Methods RS use a sample-based estimate of the matrix A, whereas methods R (*without* the S) use a diagonal A matrix, calculated under the normal model. The shift correction is denoted by *corr*. Method R*corr*4, for example, denotes the procedure outlined in Section 4.2, with diagonal A-matrix, a shift correction and *effective* sample size $n_1 = 4$, actual sample size n specified elsewhere. With the exception of Table 4.4.21, the shift correction is always a multiple of $\hat{\sigma}/(n\sqrt{n})$; often, the multiple is z_α, where $\Phi(z_\alpha) = \alpha$ and a $(1-2\alpha)100\%$ confidence interval is desired. The letter K refers to our modification of Kent's test procedure (Section 4.3). Methods KS use sample-based estimates of matrices. Methods K use diagonal matrices, estimated under the normal model. Huber's t-test and the classical t-test are specified by name. For all but the classical interval, it is also necessary to specify the value of Huber's c, the truncation point for the score functions Ψ_c.

In Table 4.4.1 we present the results of 1000 samples of size 5, machine generated standard normal data. In this table, we have recorded the average lengths of the 1000 intervals calculated under each method and the proportion of intervals which covered the true center of the distribution, zero. At a glance, we see that the robust intervals are too short. In fact, for each sample, the robust interval is typically shorter than the classical interval.

For 1000 binomial trials, with probability of success $p = 0.9$, the standard deviation of the average number of successes is 0.0095. Hence, for a true 90% confidence interval, the standard deviation of the average coverage for 1000 intervals (reported in the tables as a percent) is 0.95. Similarly, for a true 80% confidence interval, the standard deviation of average coverage (for 1000 samples) is 1.26, and for a 95% interval the standard deviation is 0.69.

For the standard normal data presentd in Table 4.4.1, the average length of the classical intervals is 1.78, and the estimated standard deviation is 0.681. Hence, the estimated standard deviation of the average length of 1000 intervals is 0.022. For the R*corr*4 intervals of Table 4.4.1, the standard deviation of the average length of 1000 intervals is also estimated to be 0.022.

We first consider Methods RS4 and KS. We see that the average length of the RS4 intervals is longer than the R4 intervals, and yet R4 gives better coverage. Similarly, method K gives shorter intervals and better coverage than KS. The R4 and RS4 methods usually give similar intervals, but the RS4 intervals are occasionally much longer. The same phenomenon is occuring with the two Kent methods. The erratic behaviour of intervals constructed from a sample-based A matrix is illustrated by the following example.

Example 4.4.1

Huber's $c = 1.5$. Confidence level 90%

Sample 1 $-3.0, -2.0, -1.1, 0.0, 6.0$

sample average: $-.02$, standard deviation: 3.54

Huber estimates, location: $-.02$ scale: 4.02

Method	*Interval*
Classical	$-3.40, 3.36$
R4	$-1.92, 3.57$
RS4	$-1.67, 3.12$
K4	$-3.15, 3.04$
KS4	$-2.64, 2.46$

Sample 2 $-3.0, -2.0, -1.2, 0.0, 6.0$

sample average: -0.40, standard deviation: 3.55

Huber estimates, location: -0.49, scale: 4.00

Method	*Interval*
Classical	$-3.43, 3.35$
R4	$-1.93, 3.54$
RS4	$-9.40, 21.94$
K4	$-3.17, 3.02$
KS4	$-13.33, 13.25$

configuration for R4: $-3.41, -2.25, -1.33, .06, 6.93$

RS4: $-11.79, -11.56, -10.85, -7.58, 41.78$

In Example 4.4.1, the first sample has no outliers, in the sense that

$$|\Psi_c(\frac{x_i - \hat{\mu}}{\hat{\sigma}})| < c, \quad i = 1, \ldots, 5, \quad \hat{\mu} = -.02, \hat{\sigma} = 4.02.$$

The second sample is the same, with the exception that the middle point has been changed from -1.1 to -1.2. This change causes a small change in location and a smaller change in scale, but the observation 6.0 is now an outlier. For samples of size $n = 5$,

$$\hat{\delta} = \frac{1}{n} \sum_{i=1}^{n} I_{trc}(\frac{x_i - \hat{\mu}}{\hat{\sigma}})$$

is 1.0 if there are no outliers and 0.8 if there is one outlier. If $n = 5$ and $c = 1.5$, then $\delta = .87$ and $\beta = .78$. If there are no outliers, then

$$\hat{A} = \frac{-1}{\hat{\sigma}}\begin{pmatrix} 1.0 & 0 \\ 0 & 1.24 \end{pmatrix}$$
$$-\hat{A}^{-1} = \hat{\sigma}\begin{pmatrix} 1.0 & 0 \\ 0 & 0.80 \end{pmatrix}.$$

and the largest positive entry in the configuration is less than $\hat{\sigma}c/\delta = (1.7)\hat{\sigma}$. However, if there is one outlier, on the right, then

$$\hat{A} = \frac{-1}{\hat{\sigma}}\begin{pmatrix} 0.80 & -0.30 \\ -0.60 & 0.345 \end{pmatrix}$$
$$-\hat{A}^{-1} = \hat{\sigma}\begin{pmatrix} 3.60 & 3.13 \\ 6.26 & 8.35 \end{pmatrix}$$

and the largest positive entry in the configuration is $\hat{\sigma}((3.6)c + (3.13)(c^2 - \beta_L)) = (10.5)\hat{\sigma}$. For small samples, sample-based estimates of the A matrix are *not* robust. We omit methods RS and KS from further discussion. (More samples are presented in detail at the end of the chapter, before the tables).

It is in the light of these observations about the non-robustness of sample estimates of δ that we have programmed the particular variation of Huber's t-test outlined at the end of Section 4.3. Recall that the Huber intervals use δ calculated under the normal model rather than $\hat{\delta}$ estimated from the sample. It seems safer to make a few assumptions about the form of a matrix than to use sample-based estimates, and inverses thereof. Our techniques are sample-based. The trick is to use the sample robustly.

The histograms in Figure 4.4.2 compare the distribution of lengths of intervals for the classical and R*corr*4 methods, $corr = \hat{\sigma}/(n\sqrt{n})$. We see that the classical method gives lengths centered to the right of the R*corr*4 method. The histogram for the Kent method K is roughly between the two histograms of Figure 4.4.2. The Huber t-intervals have lengths with a flatter distribution, and there are more long intervals than for the Classical method.

Referring back to Table 4.4.1, we see that the robust intervals can be improved by decreasing the effective sample size and by incorporating a shift correction. The R*corr* and Kent methods can give intervals which go beyond the smallest or largest observation. The R methods (*without* a correction) stay within the range of the observations. This short-coming is bound to lead to too low coverage. Field has suggested that we compensate for this fault as follows: for a $(1-2\alpha)100\%$ confidence interval, look for a $(1-2\alpha)100/(1-(\frac{1}{2})^{n-1})\%$ confidence interval; the motivation being that, if the distribution is symmetric, then the proportion $(\frac{1}{2})^n$ of samples will have all observations less than the true location, and the same proportion will have all observations larger than the true location. Note that a correction of this type becomes negligible as n increases.

In this chapter and the next, our primary purpose is to demonstrate the *potential* of the confidence-interval procedure outlined in Chapter 3. The procedure needs to be refined. The two refinements we consider are changes to the effective sample size, and the shift correction, both operating at the order $o(\frac{1}{n})$. We believe that a combination of these two refinements is appropriate. The appropriate adjustments should become more clear when we have tested models involving more nuisance parameters. Simple linear regression is covered in the next chapter. Multiple regression is to appear.

Referring to Tables 4.4.1, 4.4.3, 4.4.4 we see that coverage is better for 80% confidence than for 90%, and coverage is worse for 95% confidence. This suggest that the shift correction should be larger for high confidence and lower for low confidence. It is for this reason that we tried the shift-correction $z_\alpha\hat{\sigma}/(n\sqrt{n})$, where $\alpha = \Phi(z_\alpha)$ and a $(1-2\alpha)100\%$ confidence interval is desired. For the same set of samples of size 5, standard normal, the results are presented in Table 4.4.5; the information in the first column is repeated from previous tables. We see from

Table 4.4.5 that we have *not* made a particularly good guess as to how to adjust the shift correction. In particular, the correction should be increased for higher confidence levels. (Footnote: the *multiplicative* correction introduced in Chapter 5 gives, automatically, an increase in the shift correction for larger confidence levels. Table 4.4.21 gives results for a multiplicative correction).

Before moving on to larger samples and other distributions, we address one more factor which can be adjusted: Huber's truncation point c. Table 4.4.6 shows the effect of changing c. We ran similar trials with skewed data, and agree with Huber that c should be between 1.0 and 2.0 (Huber 1981).

We now consider larger samples, drawn from a standard normal distribution. Table 4.4.7 summarizes our results for 1000 samples of size 10, with Huber's $c = 1.5$ and confidence level 90%. Once again, we see that decreasing the effective sample size increases the coverage, and that a shift correction also increases coverage. We draw attention to the good performance of the Kent intervals, which are discussed later in this section. The histograms in Figure 4.4.8 show the distribution of lengths of intervals for the Classical and the R*corr*9 methods. The two distributions are more alike for the larger sample size (compare Figure 4.4.2).

For standard normal data, samples of size 20, the 90% confidence intervals are displayed in Table 4.4.9. Notice that, for this data, we have used the larger shift correction $(1.645)\hat{\sigma}/(n\sqrt{n})$, $n = 20$. We suspect that this correction is too big, and that the appropriate effective sample size is something less than 19.

Next we consider a heavy-tailed symmetric distribution. Table 4.4.10 displays results for 1000 samples of size $n = 5$, taken from the t_3 distribution. The shift correction for these runs was $z_\alpha\hat{\sigma}/(n\sqrt{n})$ for a $(1 - 2\alpha)100\%$ confidence interval. Table 4.4.10 does not contradict our suspicions that this shift correction is too large for 80% confidence and too low for 95% confidence, and that the appropriate choice for effective sample size is something less than $n - 1$ (n for sample size, and one nuisance parameter).

For the 90% confidence intervals of Table 4.4.10, the estimated standard deviation of the average length of 1000 classical intervals is 0.10, while the estimated standard deviation of the average length of 1000 R$corr$4 intervals is 0.02.

Figure 4.4.11 compares the distribution of lengths for the classical and the Kent interval, $c = 1.5$, confidence level 90%. The histogram for the R$corr$4 intervals, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$ rises faster than either of the histograms of Figure 4.4.11, but its tail lies roughly between the two tails of Figure 4.4.11. The largest R$corr$4 interval had length 11.8.

Table 4.4.12 and 4.4.13 display results for samples of size $n = 10$ and $n = 20$, respectively, taken from t_3. We generated 1000 samples in each case. Huber's c was set to 1.5, and the shift correction was $(1.645)\hat{\sigma}/(n\sqrt{n})$. As for normal data, all the intervals have more accurate coverage for larger sample size. Unlike the results for normal data however, the robust intervals (including Kent) remain shorter than the classical intervals, as is to be expected.

We also ran 1000 samples of slash data, sample size $n = 5$. All methods give some long intervals, as displayed in Table 4.4.14. The classical method gives more long intervals and the long intervals are often as much as ten times the length of intervals calculated under the other methods.

For heavy-tailed symmetric distributions, there are more outliers. The classical interval handles outliers poorly, giving long intervals and too high coverage. The robust intervals sometimes overcompensate, giving shorter intervals with too low coverage.

Before moving on to skewed distributions, we pause to draw attention to the good performance of confidence intervals calculated using our variation of Kent's technique. Recall Kent's main result, that his test statistic W is asymptotically a linear combination of χ_1^2 random variables, with coefficients the eigenvalues of

$$(4.4.1) \qquad (B_{\theta\theta})^{-1}(BCB)_{\theta\theta}$$

(see Section 1.6, $B = -A^{-1}$). Kent writes (1982, page 22):

"*Remark 1* If (4.4.1) equals the identity matrix, then asymptotically, $W \sim \chi_p^2$, as in the classical case. If (4.4.1) reduces to the identity matrix for all underlying densities $g(y)$ satisfying the regularity conditions, we call the ... test robust. Clearly, robustness is a desirable property for a ... test to possess."

We have tampered with Kent's ideas. Firstly, our score functions are *not* the derivative log likelihood of any density function. They are the score functions for Huber's Proposal 2 (1981). Then, having noticed the dangers of sample-estimates of a non-diagonal A-matrix (the KS method of Table 4.4.1, which we promptly abandoned) we *assume* a diagonal A-matrix and a diagonal C-matrix (Section 4.3). In effect, we have altered Kent's Remark 1, saying that so long as (4.4.1) is *diagonal*, then we obtain a robust test. We then force the relevant matrices to be diagonal. The performance of the Kent intervals in all the tables is impressive. This good behaviour should be seen as evidence to support Ronchetti's τ-test (1982).

Another attractive feature of the Kent intervals is that they are *not* symmetric about the initial estimate of location. The classical interval is symmetric about the sample mean, and the Huber interval is symmetric about the M-estimate of location. The R and K intervals, however, reflect the shape of the sample. See the examples at the end of the chapter (before the tables).

The shortcoming of the Kent intervals is that, for each μ_0 under test, a new estimate of the nuisance parameters, $\hat{\sigma}_0$, must be recalculated. If there are several nuisance parameters, then calculations are increased. For the R intervals, on the other hand, the configuration is calculated once for each sample. Then, for each μ_0

under test, we calculate a new one-dimensional α_0. The computation of α_0 does *not* increase in complexity as the number of nuisance parameters increases.

For problems with several nuisance parameters, this is an important issue. For the problem presented in this chapter, the endpoints of the R and K intervals are calculated by first making an initial jump to where we expect the endpoint to be. (We usually jump out from $\hat{\mu}$ by a distance $(\delta\hat{\sigma})/\sqrt{\beta n}$ multiplied by the appropriate number chosen from t-tables. The quantity $(\delta\hat{\sigma})/\sqrt{\beta n}$ estimates the standard deviation of $\hat{\mu}$.) For a problem with several nuisance parameters, it would not be advisable to jump out too far from the initial estimate, since then we might not have good enough first-guesses to feed into the root-finders for the nuisance parameters. Consequently, not only is each test of $H_0\colon \theta = \theta_0$ more complex for the Kent method, but we would be obliged to make *more* such tests in order to construct a confidence interval (since steps would be smaller).

We have also tampered with Huber's t-test, replacing $\hat{\delta}$ (a sample-based estimate of the expected value of $I_{\mathrm{trc}}(\frac{X-\mu}{\sigma})$) by $\delta = \Phi(c) - \Phi(-c)$. It is interesting to note that, with this adjustment, our results suggest that $n-1$ degrees of freedom is *not* too large, in disagreement with Huber (1981, page 150).

We turn now to skewed distributions. For χ^2_k data, the population mean is k, and it seems appropriate to measure coverage of this value. All methods give poor coverage of k for $k = 2, 3$. For χ^2_2 there is no point where the density function has first derivative zero; the population mean is 2, and the median is 1.39. For χ^2_3, the mode is 1.0, the median is 2.37, and the mean is 3. While investingating this type of data, we found ourselves asking: "What *is* the central point of the χ^2_3 distribution?" Certainly, the sample mean $\bar{x}$ estimates the population mean, 3. If Huber's c is set to ∞, then $\hat{\mu} = \bar{x}$, and so estimates the population mean 3. If Huber's c is set to zero, however, then $\hat{\mu}$ estimates the population median, a number less than 3.

Table 4.4.15 shows that, for samples of size $n = 5$, χ^2_3 data, all methods give better coverage of the population median, 2.39, than of the population mean. As sample size increases, the intervals cover a value closer to the population mean, see Tables 4.4.16, 4.4.17. The classical interval moves toward coverage of the population mean "faster" than do the other intervals. For skewed data, there is a problem with definition of "location".

Huber's joint M-estimates of location and scale do not satisfy the Bickel and Lehmann definition of a location estimate (1975, the authors recommend trimmed means). The definition of location for a skewed distribution is an important issue for the R-intervals, since it (the appropriate location) dictates the amount of correction needed for the shift condition.

Our next group of tables (4.4.18, 4.4.20) describe contaminated standard normal data, with 5% symmetric contamination at ± 12.0. The shift corrections are motivated by Table 4.4.5. For 80% confidence, the shift correction is $\hat{\sigma}/(n\sqrt{n})$. For 90% confidence, the shift correction is $(1.7)\hat{\sigma}/(n\sqrt{n})$; and for 95% confidence the correction is $(2.2)\hat{\sigma}/(n\sqrt{n})$. We believe that, for optimal result, both the shift correction and the effective sample size should be decreased.

Table 4.4.21 considers standard normal data with asymmetric contamination and a *multiplicative* shift correction, introduced in Chapter 5. The multiplicative shift correction gives a correction proportional to $\frac{1}{n}(\theta_0 - \hat{\theta})$, and we are pleased with its performance.

In summary, the known shortcomings of the classical estimate of location, the sample mean, can be overcome using Huber's joint estimates of location and scale. In order to give confidence intervals for location, we need a good approximation to the distribution of the score functions. The simplest approach is Huber's t-test, which we have modified, essentially to avoid sample-based estimates of a matrix.

We have applied a similar twist to Kent's test procedure. The Kent intervals are not symmetric, as are the Huber intervals, and the Kent intervals clearly outperform the Huber intervals, giving shorter intervals and more accurate coverage. Indeed, there are days when we think that the biggest contribution of this paper is our slight modification of Kent's approach.

The R intervals are interesting nevertheless. They are sample-based, yet they avoid non-robust sample estimates of the A-matrix. The R intervals have the following advantage over the Kent intervals: once we have an initial estimate of all parameters, calculation of the R confidence interval is no more complex for several nuisance parameters than for one nuisance parameter. Our techniques also give confidence intervals for a real-valued function of the original parameter set, and Kent's approach does not lend itself to this generalization. (Functions of parameters are to appear in multiple regression). For the R intervals, we need to better understand the appropriate *effective sample size* (a degrees of freedom problem) and the best *shift correction*. The questions "What is the best shift condition?" and "What is the center of a χ^2_3 distribution?" are related. An exponential tilt skews the distribution. M-estimates do not estimate the mean of a skewed distribution, but something between the mean and the median or mode.

Before turning to our next problem (simple linear regression) we give some examples, intended to give the reader a feeling for how the various intervals behave. For the R intervals, we give the configuration and the value α_0 at each end point:

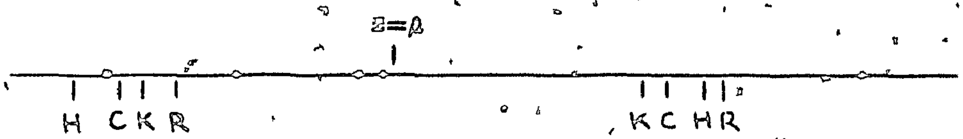
$$K'(\alpha_0) = \mu_0 - \hat{\mu} \pm corr$$

For the Kent intervals, we give the value of $\hat{\sigma}_0/\hat{\sigma}$ at each extreme of the interval. Each *sample* is represented by points on a number line. The classical estimate of location, $\bar{x}$, and the Huber estimate $\hat{\mu}$ are indicated (above the line). When there are no outliers, $\bar{x} = \hat{\mu}$. The symmetry/asymmetry of an interval can be seen by measuring the distances from its endpoints to the appropriate estimate of location. Intervals are indicated below the line: C for classical, R for R$corr(n-1)$, $corr = (1.7)\hat{\sigma}/(n\sqrt{n})$, K for Kent, and H for Huber. In all these examples, Huber's c is set to 1.5 and the desired confidence level is 90%.

Example 4.4.3.

Huber's $c = 1.5$, confidence level 90%, $n = 5$.

Sample: -1.0, -0.5, 0.0, 0.1, 2.0



Method R$corr4$, configuration : -1.3, -0.72, -0.14, -0.02, 2.2

α_0 at extremes : -0.77, 0.66

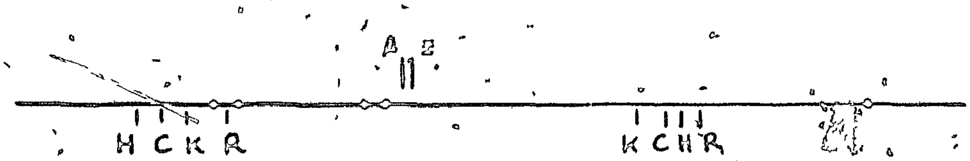
Method K, $\hat{\sigma}_0/\hat{\sigma}$ at extremes : 1.14, 1.39

Example 4.4.4

Huber's $c = 1.5$, confidence level 90%, $n = 5$.

Sample: -0.6, -0.5, 0.0, 0.1, 2.0

This example was obtained from the last, by moving the smallest point to the right.



Method R$corr4$, configuration : -0.88, -0.77, -0.19, -0.08, 1.9

α_0 at extremes : -0.90, 0.76

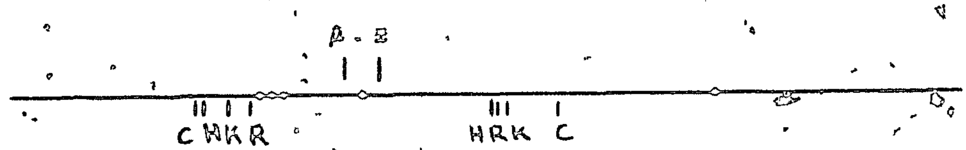
Method K, $\hat{\sigma}_0/\hat{\sigma}$ at extremes : 1.05, 1.47

Example 4.4.5

Huber's $c = 1.5$, confidence level 90%, $n = 5$.

Sample: -0.5, -0.4, -0.3, 0.5, -4.0

This example has a big outlier.



Method $Rcorr4$, configuration : -1.0, -0.88, -0.77, 0.16, 2.5

α_0 at extremes : -0.70, 0.59

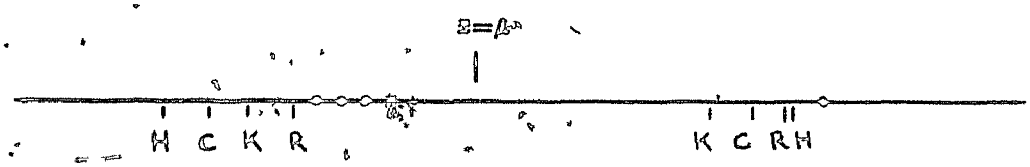
Method K, $\hat{\sigma}_0/\hat{\sigma}$ at extremes : 1.05, 1.88

Example 4.4.6

Huber's $c = 1.5$, confidence level 90%, $n = 4$.

Sample: -0.5, -0.4, -0.3, 0.5,

This example was obtained from the last by dropping the outlier.



Method $Rcorr3$, configuration : -0.38, -0.26, -0.14, 0.78,

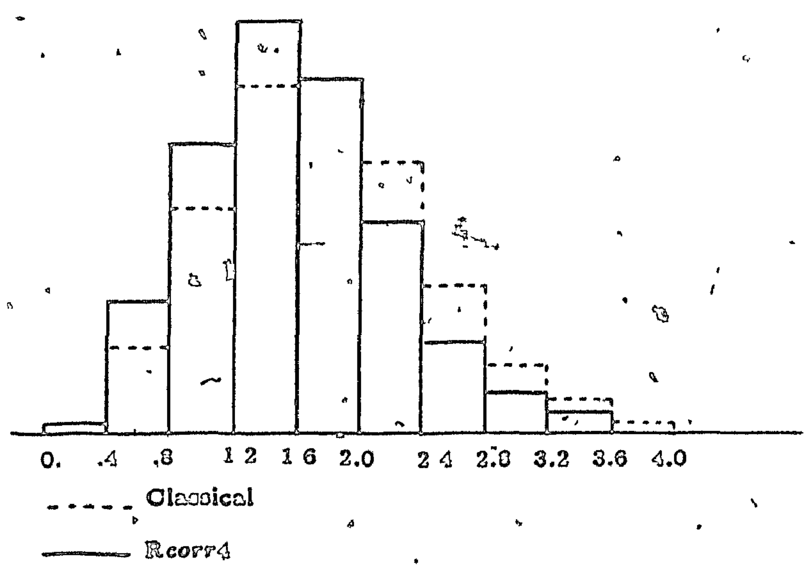
α_0 at extremes : -2.4, 2.0

Method K, $\hat{\sigma}_0/\hat{\sigma}$ at extremes : 1.52, 1.52

Method	Average length	Coverage (percent)
Classical	1.78	91
Rcorr4	1.59	87
R3	1.57	86
R4	1.40	82
RS4	1.60	79
K	1.60	88
KS	1.62	84
Huber	2.03	93

1000 samples of size $n = 5$ standard normal. Confidence level 90%, Huber's $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$

Table 4.4.1



Distribution of lengths of intervals. 1000 samples of size $n = 5$, standard normal. Confidence level 90%, $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$.

Figure 4.4.2

Method	Average length	Coverage (percent)
Classical	1.28	80
$Rcorr4$	1.32	80
R3	1.28	79
R4	1.14	75
K	1.21	78
Huber	1.46	86

1000 samples of size $n = 5$ standard normal. Confidence level 80%, Huber's $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$

Table 4.4.3

Method	Average length	Coverage (percent)
Classical	2.32	95
$Rcorr4$	1.80	91
R3	1.78	90
R4	1.61	87
K	1.97	93
Huber	2.64	96

1000 samples of size $n = 5$, standard normal. Confidence level 95%, $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$

Table 4.4.4

Confidence level	$corr = \hat{\sigma}/(n\sqrt{n})$	$corr = z_\alpha * \hat{\sigma}/(n\sqrt{n})$
80%	1.3, 80	1.4, 82
90%	1.6, 87	1.7, 90
95%	1.8, 91	2.0, 92

Entries: average length, coverage (percent). 1000 samples of size 5, standard normal. Huber's $c = 1.5$, method R*corr4*.

Comparison of shift corrections.

Table 4.4.5

Method	$c = 0.8$	$c = 1.0$	$c = 1.5$	$c = 2.0$
R*corr4*	1.8, 85	1.7, 87	1.6, 87	1.5, 84
Kent	1.7, 87	1.7, 89	1.6, 88	1.5, 87
Huber	2.3, 90	2.2, 92	2.0, 93	1.9, 91

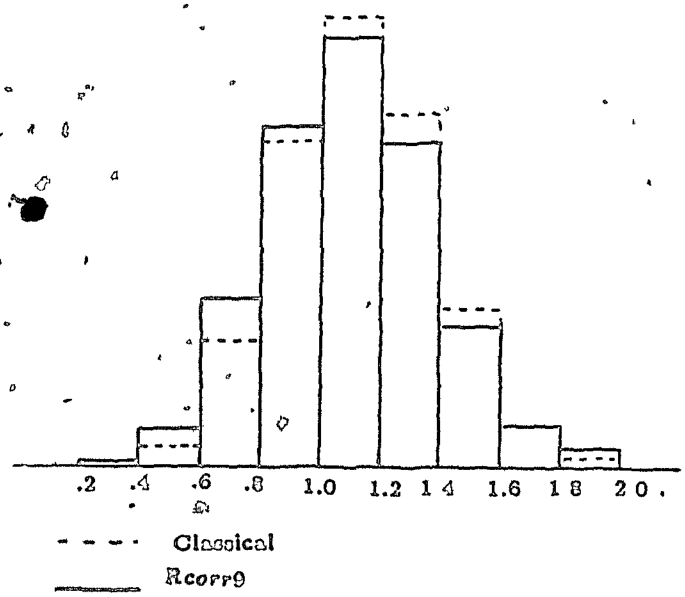
Entries: average length, coverage (percent). 1000 samples of size 5, standard normal. Confidence level 90%, $corr = \hat{\sigma}/(n\sqrt{n})$

Table 4.4.6

Method	Average length	Coverage (percent)
Classical	1.13	89
Rcorr9	1.10	87
R7	1.16	89
R8	1.10	87
R9	1.04	85
Kent	1.09	88
Huber	1.21	90

1000 samples of size 10, standard normal. Confidence level 90%, Huber's $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$

Table 4.4.7



Distribution of lengths of intervals. 1000 samples of size $n = 10$, standard normal. Confidence level 90%, $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$.

Figure 4.4.8

Method	Average length	Coverage (percent)
Classical	0.76	90
Rcorr19	0.77	90
R18.4	0.76	89
R19	0.74	88
Kent	0.76	90
Huber	0.79	90

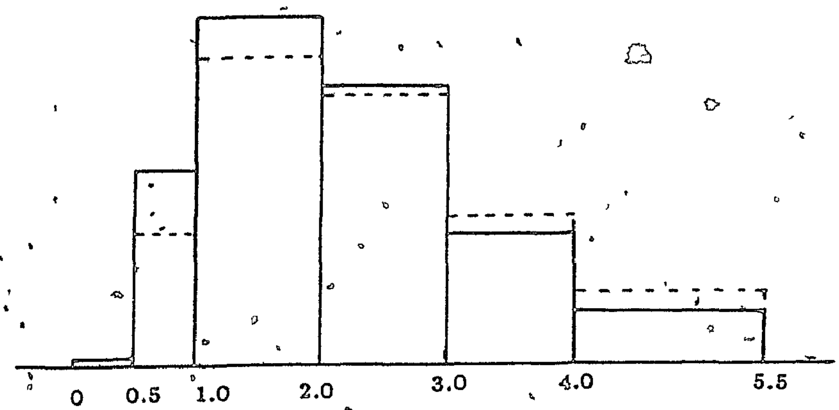
1000 samples of size $n = 20$, standard normal. Confidence level 90%, $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$.

Table 4.4.9

Method	80%	90%	95%
Classical	2.0, 84	2.7, 93	3.6, 97
Rcorr4	2.0, 83	2.5, 90	2.9, 93
R3	1.9, 79	2.3, 86	2.6, 89
R4	1.6, 74	2.0, 82	2.3, 87
Kent	1.8, 81	2.4, 90	2.9, 95
Huber	2.1, 86	2.9, 95	3.8, 97

Entries: average length, coverage (percent). 1000 samples of size $n = 5$, t_3 data. Huber's $c = 1.5$, $corr = z_{\alpha}\hat{\sigma}/(n\sqrt{n})$.

Table 4.4.10



Distribution of lengths of intervals. 1000 samples of size $n = 5$, t_3 data. Confidence level 90%, $c = 1.5$.

Figure 4.4.11

Method	Average length	Coverage (percent)
Classical	1.70	92
R*corr*9	1.46	90
R8	1.40	89
R9	1.33	87
Kent	1.44	91
Huber	1.54	93

1000 samples of size $n = 10$, t_3 data. Confidence level 90%, $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$

Table 4.4.12

Method	Average length	Coverage (percent)
Classical	1.22	91
R$corr$19	0.98	90
R18	0.96	89
R19	0.93	88
Kent	0.98	90
Huber	1.00	91

1000 samples of size $n = 20$, t_3 data, Confidence level 90%, $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$

Table 4.4.13

Method	Average Length	Coverage Percent	Percentage of intervals longer than 40.0
Classical	32.4	94	9
R$corr$4	8.5	87	1
R3	7.8	79	1
R4	7.0	74	1
Kent	10.1	91	2
Huber	9.9	95	2

1000 samples of size $n = 5$, slash data . Confidence level 90%, Huber's $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$.

Table 4.4.14

Method	Average length	Coverage of 2.4 (percent)	Coverage of 3.0 (percent)
Classical	4.00	89	85
Rcorr4	3.69	86	83
Kent	3.53	85	82
Huber	4.38	91	86

1000 samples of size $n = 5$, χ^2_3 data. Confidence level 90%, $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$.

Table 4.4.15

Method	Average length	Coverage of 2.4	Coverage of 2.6	Coverage of 3.0
Classical	2.65	88	90	86
Rcorr4	2.36	86	87	82
Kent	2.33	86	87	81
Huber	2.50	89	89	82

1000 samples of size $n = 10$, χ^2_3 data. Confidence level 90%, $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$

Table 4.4.16

Method	Average length	Coverage of 2.4	Coverage of 2.7	Coverage of 3.0
Classical	1.84	77	88	91
R*corr*4	1.60	80	87	80
Kent	1.62	81	87	81
Huber	1.64	83	87	80

1000 samples of size $n = 20$, χ^2_3 data. Confidence level 90%, $c = 1.5$, $corr = (1.645)\hat{\sigma}/(n\sqrt{n})$

Table 4.4.17

Method	Average length	Coverage (percent)	Percentage of intervals longer than 6.0
Classical	3.66	93	21
R*corr*4	2.54	88	7
Kent	2.73	90	14
Huber	2.97	94	10

1000 samples of size $n = 5$, standard normal with 5% contamination at ± 12. Confidence level 90%, $c = 1.5$, $corr = (1.7)\hat{\sigma}/(n\sqrt{n})$

Table 4.4.18

Method	80%	95%
Classical	2.6, 83	4.8, 96
$Rcorr4$	2.0, 78	3.0, 93
Kent	2.1, 79	3.3, 93
Huber	2.1, 86	3.9, 98

Entries: average length, coverage (percent). 1000 samples of size $n = 5$, standard normal, with 5% contamination at ± 12.0
Huber's $c = 1.5$, $corr = \hat{\sigma}/(n\sqrt{n})$ for 80%, $(2.2)\hat{\sigma}/(n\sqrt{n})$ for 95%.

Table 4.4.19

Method	Average length	Coverage (percent)	Percentage of intervals longer than 4.0
Classical	2.73	94	41
$Rcorr9$	1.41	88	1
Kent	1.45	88	4
Huber	1.48	90	1

1000 samples of size $n = 10$, standard normal with 5% contamination at ± 12.0. Confidence level 90%, $c = 1.5$, $corr = (1.7)\hat{\sigma}/(n\sqrt{n})$.

Table 4.4.20

Method	80%	90%	95%
Classical	71 1.56, 1.45	88 2.07, 1.93	96 2.55, 2.38
R*corr*9	78 1.26, 1.44	89 1.60, 1.82	94 1.88, 2.14
Kent	73 1.07, 1.25	86 1.40, 1.62	94 1.70, 1.95
Huber	75 1.12, 1.28	90 1.48, 1.69	96 1.83, 2.09

1000 samples of size $n = 10$, standard normal with 10% contamination at 5.
Multiplicative shift correction. shift $= \frac{n-2}{n}(\mu_0 - \hat{\mu})$
Entries: Coverage of zero (percent) / median, average length
Huber's $c = 1.5$.

Table 4.4.21

4.5 Programming details

Fortran programs appear in the back pocket of the thesis. The program MATLR calculates several location intervals. Data is generated (or read in) from the routine GENER. The mean, standard deviation, median and mad are calculated in EST1. The M-estimates are calculated using the NAG rootfinder C05NBF. The configuration is set up in SETLOC. The interval routine is ROBCI, which calls the test routine MYTEST. The routine MYTEST calls the rootfinder MYROOT, which incorporates our knowledge of the behaviour of any solution to $K'(\alpha) =$ shift, but eventually calls the NAG rootfinder C05ADF.

The rootfinder MYROOT is also called from the Kent interval routine, KENTCI. The Kent intervals have to find a nested root, $\hat{\sigma}_0$, and the routine CHIRT was written for this purpose. The routine SETLOC (designed to set up the configuration) also takes care of a few details for the Kent intervals.

Optional print statements (so that we can see more details about an interval) are blocked out with CC. Statments blocked out with a single C are options not used in recent runs of the program. For example, in SETLOC, the sample-based A-matrix is blocked out.

Chapter 5. Confidence intervals for slope in simple linear regression

In this chapter we construct confidence intervals for a real parameter (slope) in the presence of *two* nuisance parameters (intercept and mean square error).

5.1 Introduction to the problem.

We fit the model

$$y_i = \mu + \theta x_i + r_i \tag{5.1.1}$$

where the errors r_i are assumed identically and independently distributed, uncorrelated with the predictors x_i. The errors are assumed to have a symmetric distribution with variance σ^2, unknown. The intercept μ is also unknown. We desire a confidence interval for

$$\theta = (0\ 1\ 0)(\mu\ \theta\ \sigma)^T.$$

We estimate μ, θ, σ by M-estimates $\hat{\mu}, \hat{\theta}, \hat{\sigma}$, solutions of the minimization problem:

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i - \underline{x}_i^T\underline{\eta}}{\sigma}\right)\sigma = \min! \tag{5.1.2}$$

where

$$\underline{x}_i^T = (1 \quad x_i)$$

$$\underline{\eta}^T = (\mu \quad \theta)$$

$$\rho(t) = \begin{cases} \frac{1}{2}(t^2 + \beta_R) & \text{if } |t| < c \\ c|t| - \frac{1}{2}c^2 + \frac{1}{2}\beta_R & \text{if } |t| \geq c \end{cases}$$

$$\beta_R = \left(\frac{n-1}{n}\right)^2\beta,$$

see Huber (1981).

Partial differentiation of (5.1.2) with respect to μ, θ, σ leads to the score functions $\underline{\Psi}(x, r, \sigma)$,

$$\begin{aligned} \Psi_1: &\quad \Psi_c\left(\frac{r}{\sigma}\right) \\ \Psi_2: &\quad x\Psi_c\left(\frac{r}{\sigma}\right) \\ \Psi_3: &\quad \frac{1}{2}\left(\Psi_c^2\left(\frac{r}{\sigma}\right) - \beta_R\right) \end{aligned}$$

where $r = y - \mu - \theta x = y - \underline{x}^T\underline{\eta}$. The estimates $\hat{\mu}$, $\hat{\theta}$, $\hat{\sigma}$ solve

$$\frac{1}{n}\sum_{i=1}^{n} \underline{\Psi}(x_i, r_i, \sigma) = \underline{0}.$$

The matrix of derivatives of the score functions is

$$D\underline{\Psi}(x, r, \sigma) = (-1/\sigma) I_{\text{trc}}(r/\sigma) \begin{pmatrix} 1 & x & r/\sigma \\ x & x^2 & xr/\sigma \\ r/\sigma & xr/\sigma & (r/\sigma)^2 \end{pmatrix}. \tag{5.1.3}$$

In order to apply the techniques of Chapter 3, we need to estimate

$$A = E(D\underline{\Psi})$$

and

$$B = -A^{-1}.$$

We could estimate A from the sample:

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} D\underline{\Psi}(x_i, r_i, \hat{\sigma}).$$

Then the configuration is given by

$$g_i = \underline{\Psi}^T(x_i, \hat{r}_i, \hat{\sigma})\hat{B}(0\ 1\ 0)^T,$$

where $\hat{r}_i = y_i - \hat{\mu} - \hat{\theta}x_i$, $\hat{B} = -\hat{A}^{-1}$, and we can construct confidence intervals for θ as outlined in Section 3.8.

Before fitting the model (5.1.1), we recenter the predictors x_i. We do this for two reasons. First, the recentering makes the matrix $\underset{\sim}{A}$ more manageable. Second, the recentering is necessary for *weighted* robust regression (see below). We emphasize that the predictors x_i are assumed fixed. We have *not* attempted a model in which the predictors are also assumed to have a distribution, with location and scale for the predictors to be estimated from the sample (see below).

The simplest way to recenter the predictors is to subtract the average $\bar{x}$ from each x_i. In our programs, we calculate the joint M-estimates of location and scale μ_x, σ_x, solutions of

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_c\left(\frac{x_i-\mu}{\sigma}\right)=0$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\Psi_c^2\left(\frac{x_i-\mu}{\sigma}\right)-\beta_L\right)=0$$

and then replace each x_i by the normalized predictor $\frac{x_i-\mu_x}{\sigma_x}$. *For the rest of this chapter*, x_i denotes *normalized* x_i:

$$\frac{x_i-\mu_x}{\sigma_x}.$$

Note, in particular, that

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_c(x_i)=0.$$

Then the model (5.1.1) has been transformed to

$$y_i=\mu^*+\theta^* x_i+r_i \tag{5.1.4}$$

where $\theta^*=\theta\sigma_x$ and $\mu^*=\mu+\theta\mu_x$. We obtain a confidence interval for θ^*, then divide by σ_x to give a confidence interval for θ.

We now discuss the matrix A, and its estimate $\hat{A}$. First note that A and $\hat{A}$ are symmetric. We write $A=\frac{-1}{\sigma}(a_{jm})$ and $\hat{A}=\frac{-1}{\hat{\sigma}}(\hat{a}_{jm})$. Then, from (5.1.3),

$$\hat{a}_{12}=\frac{1}{n}\sum_{i=1}^{n}x_i I_{\mathrm{trc}}\left(\frac{\hat{r}_i}{\hat{\sigma}}\right)$$

and $\hat{a}_{12}$ should be close to $\frac{1}{n}\sum \Psi_c(x_i) = 0$. ($\hat{a}_{12}$ is actually equal to zero if $I_{trc}(\frac{\hat{r}_i}{\hat{\sigma}}) \neq 0$ and $\Psi_c(x_i) \neq \pm c$ for $i = 1, \ldots, n$; that is, if there are no outliers among the residuals *and* the predictors). Certainly, $\hat{a}_{12}$ should be small when compared with $\hat{a}_{11} = \frac{1}{n}\sum I_{trc}(\frac{\hat{r}_i}{\hat{\sigma}})$. Similarly,

$$\hat{a}_{13} = \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{r}_i}{\hat{\sigma}} I_{trc}\left(\frac{\hat{r}_i}{\hat{\sigma}}\right)$$

should be close to $\frac{1}{n}\sum \Psi_c(\frac{\hat{r}_i}{\hat{\sigma}}) = 0$, and, in particular, $\hat{a}_{13}$ should be small when compared with $\hat{a}_{11}$. The term $\hat{a}_{23}$ is more interesting,

$$\hat{a}_{23} = \frac{1}{n}\sum_{i=1}^{n} x_i \left(\frac{\hat{r}_i}{\hat{\sigma}}\right)_{trc}$$

If $|\hat{a}_{23}|$ is "large", then we do not have a good model: $a_{23} \neq 0$ implies a correlation between the estimated errors $\hat{r}_i$ and the predictors x_i. For the moment, assume that $\hat{a}_{12} \doteq \hat{a}_{13} \doteq 0$, but do not ignore a_{23}.

Then we estimate A by

$$\hat{A} \doteq \frac{-1}{\hat{\sigma}} \begin{pmatrix} \hat{a}_{11} & 0 & 0 \\ 0 & \hat{a}_{22} & \hat{a}_{23} \\ 0 & \hat{a}_{23} & \hat{a}_{33} \end{pmatrix}$$

and

$$\hat{B} = \frac{\hat{\sigma}}{d} \begin{pmatrix} d/\hat{a}_{11} & 0 & 0 \\ 0 & \hat{a}_{33} & -\hat{a}_{23} \\ 0 & -\hat{a}_{23} & \hat{a}_{22} \end{pmatrix}$$

where

$$d = \hat{a}_{22}\hat{a}_{33} - \hat{a}_{23}^2.$$

Then the configuration; giving confidence intervals for θ^* is

(5.1.5) Method R : $$g_i = \frac{\hat{\sigma}}{d}[\hat{a}_{33}\Psi_2(x_i, \hat{r}_i, \hat{\sigma}) - \hat{a}_{23}\Psi_3(x_i, \hat{r}_i, \hat{\sigma})].$$

That is to say, confidence intervals for slope are based on the second score function, with a correction for possible correlation between the predictor x_i and the estimated error $\hat{r}_i$. D is for diagonal $\hat{A}$, giving the following configuration.

(5.1.6) Method RD : $$g_i = \frac{\hat{\sigma}}{\hat{a}_{22}}\Psi_2(x_i, r_i, \hat{\sigma}).$$

We saw in Chapter 4 that we should be wary of sample-dependent estimates of the matrices A and B. For example, rather than using

$$\hat{a}_{33} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{r}_i}{\hat{\sigma}}\right)^2_{\mathrm{trc}}$$

in the configuration (5.1.5), it might be better to use (with the notation of Chapter 4) $a_{33}^* = \varepsilon$, which is the expected value of $(r_i/\sigma)^2$ for normally distributed errors, or some multiple of ε. It is more difficult to decide upon a robust estimate for A in the case of *weighted* robust regression. Our programs perform a weighted regression, which we now describe.

We solve a modification of the minimization problem (5.1.2). We use the hat matrix $H = X(X^TX)^{-1}X^T$ to down-play leverage points. Here, X is the matrix with 1 in every position of the first column and (normalized) x_i in the i^{th} position of the second column. Following Huber (1981, sections 7.2, 7.9) we let h_i be the i^{th} diagonal entry of H and define weights

$$w_i = \sqrt{1 - h_i}.$$

Then, rather than solve the minimization problem (5.1.2), we solve

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i}{w_i\sigma}\right)w_i^2\sigma = 0.$$

where

$$\rho(t) = \begin{cases} \frac{1}{2}(t^2 + \beta) & \text{if } |t| < c \\ c|t| - \frac{1}{2}c^2 + \frac{1}{2}\beta & \text{if } |t| \geq c. \end{cases}$$

Note that we have used β and *not* β_R (see Ψ_3 below). The score functions are then

$\Psi(x, w, r, \sigma)$,

$$\Psi_1: \quad \Psi_c\left(\frac{r}{w\sigma}\right)w$$

$$\Psi_2: \quad \Psi_c\left(\frac{r}{w\sigma}\right)wx$$

$$\Psi_3: \quad \frac{1}{2}w^2\left(\Psi_c^2\left(\frac{r}{w\sigma}\right) - \beta\right).$$

The third score function is equivalent to

$$\Psi_3: \qquad \frac{1}{2}\left(w^2\Psi_c^2\left(\frac{rw}{\sigma}\right) - \beta_R\right),$$

since $\sum w_i^2 = n-2$. (See Huber, 1981.)

The matrix $D\underline{\Psi}(x, w, r, \sigma)$ looks just like (5.1.3) except that, in each position, $I_{\mathrm{trc}}(\frac{r}{\sigma})$ is now replaced by $I_{\mathrm{trc}}(\frac{r}{w\sigma})$.

We estimate A from the sample:

$$\begin{aligned}\hat{A} &= \frac{-1}{\hat{\sigma}}(\hat{a}_{jm}) \\ &= \frac{-1}{n\hat{\sigma}}\sum_{i=1}^{n} I_{\mathrm{trc}}\left(\frac{\hat{r}_i}{w_i\hat{\sigma}}\right)\begin{pmatrix} 1 & x_i & \hat{r}_i/\hat{\sigma} \\ x_i & x_i^2 & x_i\hat{r}_i/\hat{\sigma} \\ \hat{r}_i/\hat{\sigma} & x_i\hat{r}_i/\hat{\sigma} & (\hat{r}_i/\hat{\sigma})^2 \end{pmatrix}.\end{aligned}$$

We base tests on two different configurations. D is for diagonal $\hat{A}$, W is for weighted regression.

$$\text{Method RDW}: \qquad g_i = \frac{\hat{\sigma}}{\hat{a}_{22}}\Psi_2(x_i, w_i, \hat{r}_i, \hat{\sigma})$$

and

$$\text{Method RW}: \qquad g_i = \frac{\hat{\sigma}}{d}\left(\hat{a}_{33}\Psi_2(x_i, w_i, \hat{r}_i, \hat{\sigma}) - \hat{a}_{23}\Psi_3(x_i, w_i, \hat{r}_i, \hat{\sigma})\right)$$

where $d = \hat{a}_{22}\hat{a}_{33} - \hat{a}_{23}^2$.

In all cases, the shift condition is a modification of

$$E(g) = \theta_0^* - \hat{\theta}^*$$

where $H_0\colon \theta = \theta_0$, $\theta_0^* = \theta_0\sigma_x$. For this problem, the shift correction has been incorporated as a multiplicative factor. Recall that, to test $H_0\colon \theta = \theta_0$, we relax the shift condition to

$$E(g) = \theta_0 - \hat{\theta}_{\mathrm{obs}} + corr$$

where the correction is chosen so that

$$|\theta_0 - \hat{\theta}_{\text{obs}} + corr| < |\theta_0 - \hat{\theta}_{\text{obs}}|.$$

In this chapter, the correction is

$$|corr| = \frac{3 - w_{(1)}}{n} |\theta_0^* - \hat{\theta}_{\text{obs}}^*|,$$

where

$$w_{(1)} = \min\{w_i; i = 1, \ldots, n\},$$

w_i the weights used for weighted regression. The same correction is used for both weighted and non weighted regression. The three is just a guess at the appropriate factor.. The minimum weight, $w_{(1)}$, was motivated by test samples. Basically, if one of the predictors x_i has high influence then, in order to improve coverage, we need a larger shift correction. The shift condition is then equivalent to

$$E(g) = \frac{n - 3 + w_{(1)}}{n} (\theta_0^* - \hat{\theta}_{\text{obs}}^*).$$

The confidence interval for θ^* is divided by σ_x to give a confidence interval for θ.

In the next section, we compare the R intervals with the classical, least squares, confidence interval. Recall that the classical estimates of μ, θ and σ^2 are $\tilde{\mu}$, $\tilde{\theta}$ and s^2, where $\underline{\tilde{\eta}}^T = (\tilde{\mu}, \tilde{\theta})$,

$$\begin{aligned} \underline{\tilde{\eta}} &= (X^T X)^{-1} X^T \underline{y} \\ \underline{\tilde{y}} &= X \underline{\tilde{\eta}} \\ s^2 &= \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2, \end{aligned}$$

and the $(1 - 2\alpha)100\%$ confidence interval for θ is

$$\tilde{\theta} - t_{n-2}(\alpha) s \left(\sum (x_i - \bar{x})\right)^{-1/2} < \theta < \tilde{\theta} + t_{n-2}(\alpha) s \left(\sum (x_i - \bar{x})^2\right)^{-1/2}.$$

5.2 Numerical results

We consider the simple model

$$y = \theta x + \mu + \varepsilon$$

$$y = 2x + \mu + \varepsilon$$

and display results for four sets of predictors. The sets are as follows.

Set 1, $n = 7$,

$\{-3.0,\ -2.5,\ -2.0,\ -1.5,\ -1.0,\ -0.5,\ 0.0\}$.

Set 2, $n = 13$,

$\{-3.0,\ -2.5,\ -2.0,\ -1.5,\ -1.0,\ -0.5,\ 0.0,\ 0.5,\ 1.0,\ 1.5,\ 2.0,\ 2.5,\ 3.0\}$.

Set 3, $n = 7$,

$\{0.1,\ 0.2,\ 0.3,\ 1.1,\ 1.2,\ 1.4,\ 2.2\}$.

Set 4, $n = 13$,

$\{0.0,\ 0.1,\ 0.15,\ 0.2,\ 0.25,\ 0.3,\ 1.1,\ 1.2,\ 1.3,\ 1.4,\ 2.1,\ 2.2,\ 2.3\}$.

Notice that Set 1 is a subset of Set 2, and that both these sets have equally spaced predictors. For both Sets 1 and 2 the weights are (of course) heaviest in the center, the minimum weights being 0.73 (12% of the total weight) at each end of Set 1, and 0.85 (7% of the total) at each end of Set 2.

Set 3 is a subset of Set 4. In Set 4, the predictors fall into three groups, 0.0 to 0.3, 1.1 to 1.4, and 2.1 to 2.3, with unequal numbers in each group. In Set 3, the isolated predictor 2.2 is assigned the minimum weight 0.63 (11% of the total weight). In Set 4, the largest predictor is assigned the minimum weight, 0.85 (7% of the total).

We generated standard normal errors, ε, using the IMSL routine GGNML, and thus produced the responses $y = 2x + 1 + \varepsilon$. We generated one set of 500 samples for the predictors in Set 2 ($n = 13$ observed y in each sample) and then used the appropriate subset of these *same* observations to obtain the tables describing Set

1. Similarly, we first obtained the tables describing Set 4, then "threw away" the unwanted predictors and responses to obtain the tables describing Set 3.

In all cases, Huber's c was set to 1.5. Remember that W indicates weighted regression as described in Section 5.1, and D indicates that the A matrix was assumed diagonal. The correction to the shift condition was obtained by multiplying $\theta_0^* - \hat{\theta}^*$ by the factor $\frac{n-3+w_{(1)}}{n}$. The correction is indicated by *corr*.

We compare the R intervals with the classical (least squares) intervals for θ ($\theta = 2$). For standard normal errors, the results for Sets 1 and 2 are displayed in Tables 5.2.1 and 5.2.2 respectively; the results for Sets 3 and 4 are in Tables 5.2.3 and 5.2.4. We see at a glance that the shift correction improves coverage, as does a decrease in the effective sample size. The RDW methods (*without* a shift correction) improve with larger sample size, as is to be expected. Only one method, R$corrn_1$ (*no* weights in the regression), uses a non-diagonal A matrix. The R$corrn_1$ intervals are often the same as the RD$corrn_1$ intervals, and offer no advantage over the latter intervals in our trials. However, the R$corrn_1$ intervals may offer some advantages when there is a correlation between the errors, ε, and the predictors, x. We call attention to the good performance of the RD$corrn_1$ intervals in all the tables.

The most striking feature of the tables is that the R intervals are often *longer* than the classical intervals. (Recall that, in Chapter 4, the R intervals were too short). To illustrate what is going on, we refer the reader to Example 5.2.1 (immediately before the tables). This example was taken from a trial sample. We see that the two extreme points, on either side of the classical regression line, are holding the line "fast" and allow less flexibility for the slope (that is, a small confidence interval for slope). The R intervals, on the other hand, down-play the influence of these points. Example 5.2.2 was obtained from another trial, and shows that the R intervals can also be shorter than the C intervals.

The distributions of the interval lengths are displayed in two histograms. Figure 5.2.5 compares the classical and RD*corr*11 intervals for predictors from Set 2. (The RD*corr* methods use no weights for regression, but *does* use weights to determine the shift correction.) The largest intervals were 0.85 for the classical method, and 1.25 for RD*corr*11. The RDW*corr*11 method (*weighted* regression) gave similar intervals, but the long ones were longer: longest interval 2.16.

Figure 5.2.6 compares lengths for the classical and the RDW*corr*5 intervals. For this data, the largest RD*corr* interval was 5.43.

The tables indicate that the shift correction *is* needed. Weighted regression gives longer intervals (some very long). Non-weighted regression gives shorter intervals. The RD*corr* intervals, in particular, look promising.

The multiplicative shift correction also looks promising. We ran trials with contaminated standard normal errors. The contamination was 10% at 5.0 (asymmetric contamination). The results appear in Table 5.2.7. We draw attention to the negative skewness exhibited by the classical method for $n = 13$.

We wanted some gauge of how the R intervals perform for small samples, with arbitrary sets of "bad" predictors and non-normal errors. We generated 500 samples of size $n = 7$, χ^2_3 data. These become the 500 predictor sets, x_i. Then, for each set of predictors we generated a *new* set of 7 response errors, ε_i, t_4 data. Then we constructed the response variables

$$y_i = 2x_i + 1 + \varepsilon_i.$$

Thus, we formed 500 sets of data (x, y), each with 7 pairs (x, y). The results for this data are shown in Table 5.2.8. The third column of this table is baffling. We suspect the problem is caused by samples in which the two largest (or smallest) terms of the configuration are very close together. Then, for small samples, the solution of $K'(\alpha) =$ shift is inaccurate. Recall that the solution of $K'(\alpha) =$ smallest g_i is $-\infty$

(see Theorem 1.4.2). The routine MYROOT returns a message to shorten the interval if $|\alpha_0|$ is too large. We emphasize that these are *computational* problems. (We have since written a new version of the routine ROBCI, which is not "intimitated" by such situations, producing larger intervals when the two largest (or smallest) terms of the configuration are close together.)

We also wanted to try the multiplicative shift correction for the location problem. We tried the factor $(n-2)/n$, with contaminated standard normal data, contamination 10% at 5. The results are shown in Table 4.4.21, at the very end of Chapter 4. Once again, the Classical intervals have median length larger than average length.

That the R intervals (for regression) are sometimes longer than the C intervals is, we think, reasonable. The classical intervals tend to be short when there are points with high influence. The R intervals down-play such points.

As with the problem of Chapter 4, refinements of the shift condition should be investigated further. The shift condition raises interesting questions about location for skewed distributions, and about loss of information when we move from the sample to the configuration (Section 3.7).

Example 5.2.1

Huber's $c = 1.5$, confidence level. 90%, $n = 6$.

Data: (x, y): (1.0, 1.6), (1 5, 5 8), (2.0, 4.6), (2.5, 6.9), (3.0, 7.5), (3.5, 8.9)
Weights: 0.69, 0.84, 0.91, 0.91, 0.84, 0.69

Model $y = \theta x + \mu + \varepsilon$

Classical estimates and R estimates of θ and μ are equal.

$\hat{\theta} = 2.51, \quad \hat{\mu} = 0.24$

Estimated scale for error ε is 1.14 for Classical, 1.58 for RDW

Classical interval: 1.3, 3.7

Configuration: −2.7, −0.39, −0.30, 0.19, 0.33, 2.9

RDW*corr4* interval : 1.1, 4.0

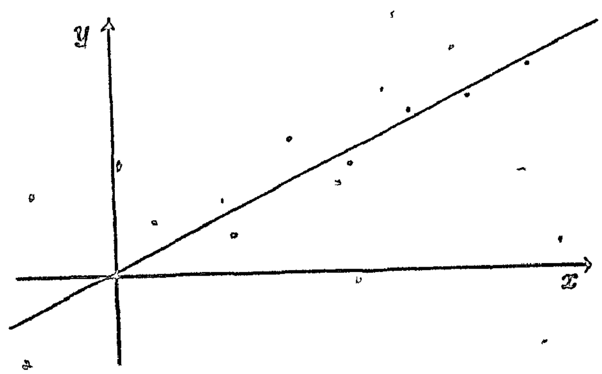
α_0 at extremes : −0.52, 0.51

The RD*corr4* interval is essentially the same.

RDW4 interval : 1.6, 3.4

α_0 at extremes : −0.51, 0.50

The graph shows the fitted regression line.



Example 5.2.2

Huber's $c = 1.5$; confidence level: 90%, $n = 7$.

Data: (x, y): (0.006, 1.73), (0.026, 1.68), (0.17, 0.43), (1.17, 3.21), (1.21, 2.77), (1.85, 3.49), (2.39, 3.93)

Weights: 0.83, 0.83, 0.86, 0.92, 0.92, 0.84, 0.70

Model $y = \theta x + \mu + \varepsilon$, scale of ε is σ

Classical estimates : $\hat{\theta} = 1.17,\ \hat{\mu} = 1.32,\ \hat{\sigma} = 0.59$

RDW estimates : $\hat{\theta} = 1\ 09,\ \hat{\mu} = 1.47,\ \hat{\sigma} = 0.54$

Classical interval : 0.65, 1.69

Configuration : −0.64, −0.55, −0 44, −0.03, −0 02, 0.23, 1.45.

RDW*corr*5 interval : 0.74, 1.76

α_0 at extremes : −1.23, 1.03

RDW5 interval : 0 86, 1.54

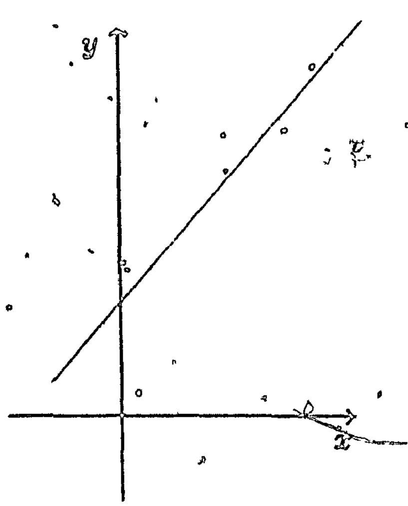
α_0 at extremes : −1.22, 1.03

RD*corr*5 interval : 0.71, 1.82

α_0 at extremes : −1.13, 0.94

and the configuration for non-weighted regression is more spread out: −0.723 to 1.59.

The classical regression line is shown.



Method	80%	90%	95%
Classical	79 1.0, 1.0	89 1.4, 1.4	95 1.8, 1.8
RDW*corr*5	81 1.1, 1.2	88 1.4, 1.6	92 1.7, 1.8
RDW5	66 0.77, 0.84	77 1.0, 1.1	81 1.1, 1.2
RDW4	71 0.85, 0.93	79 1.1, 1.2	84 1.2, 1.3
RD*corr*5 R*corr*5	82 1.1, 1.2	88 1.4, 1.5	92 1.7, 1.8

Regression model $y = 2x + 1 + \varepsilon$, ε standard normal.
$n = 7$ predictors. Set 1 (equally spaced)
Entries: Coverage of 2 (percent) / median, average length
500 samples, three levels of confidence.
Shift correction factor: $\frac{n-3+w_{(1)}}{n}$. Huber's $c = 1.5$.

Table 5.2.1

Method	80%	90%	95%
Classical	81 0.40, 0.39	91 0.52, 0.52	94 0.64, 0.64
RDW*corr*11	85 0.41, 0.44	91 0.53, 0.56	92 0.66, 0.68
RDW11	77 0.34, 0.37	87 0.44, 0.47	88 0.55, 0.57
RDW10	79 0.36, 0.38	88 0.46, 0.49	90 0.58, 0.59
RD*corr*11 R*corr*11	85 0.42, 0.43	91 0.53, 0.55	92 0.66, 0.67

Regression model $y = 2x + 1 + \varepsilon$, ε standard normal.
$n = 13$ predictors: Set 2 (equally spaced)
Entries: Coverage of 2 (percent) / median, average length
500 samples, three levels of confidence.
Shift correction factor: $\frac{n-3+w_{(1)}}{n}$. Huber's $c = 1.5$.

Table 5.2.2

Method	80%	90%	95%
Classical	82 1.4, 1.5	92 2.0, 2.0	96 2.5, 2.6
RDW*corr*5	82 1.6, 1.8	89 2.0, 2.2	92 2.3, 2.6
RDW5	70 1.1, 1.2	77 1.3, 1.5	82 1.5, 1.7
RDW4	72 1.2, 1.3	80 1.5, 1.6	83 1.7, 1.9
RD*corr*5 R*corr*5	83 1.6, 1.7	89 2.0, 2.1	92 2.3, 2.4

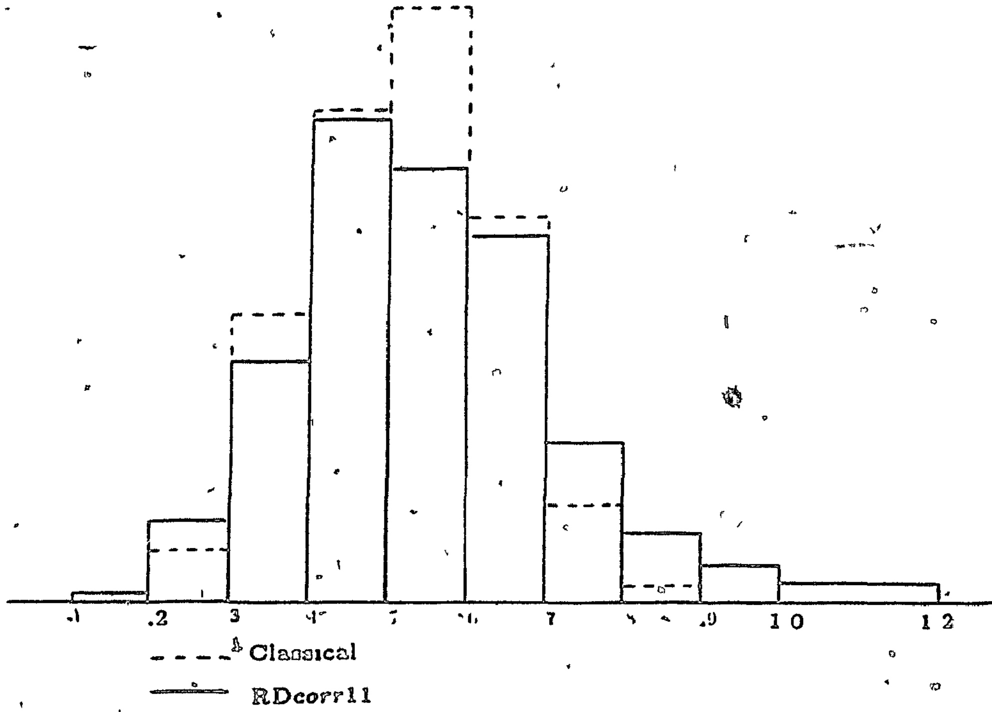
Regression model $y = 2x + 1 + \varepsilon$, ε standard normal.
$n = 7$ predictors: Set 3 (grouped)
Entries: Coverage of 2 (percent) / median, average length
500 samples, three levels of confidence.
Shift correction factor: $\frac{n-3+w_{(1)}}{n}$. Huber's $c = 1.5$.

Table 5.2.3

Method	80%	90%	95%
Classical	82 0.9, 0.9	90 1.2, 1.2	95 1.5, 1.5
RDW*corr*11	82 1.0, 1.0	91 1.3, 1.3	94 1.5, 1.5
RDW11	76 0.8, 0.8	86 1.1, 1.1	91 1.3, 1.3
RDW10	77 0.9, 0.9	87 1.1, 1.1	92 1.3, 1.3
RD*corr*11 R*corr*11	84 / 83 1.0, 1.0	91 1.2, 1.3	94 1.5, 1.5

Regression model $y = 2x + 1 + \varepsilon$, ε standard normal.
$n = 13$ predictors: Set 4 (grouped)
Entries: Coverage of 2 (percent) / median, average length
500 samples, three levels of confidence.
Shift correction factor: $\frac{n-3+w_{(1)}}{n}$. Huber's $c = 1.5$.

Table 5.2.4

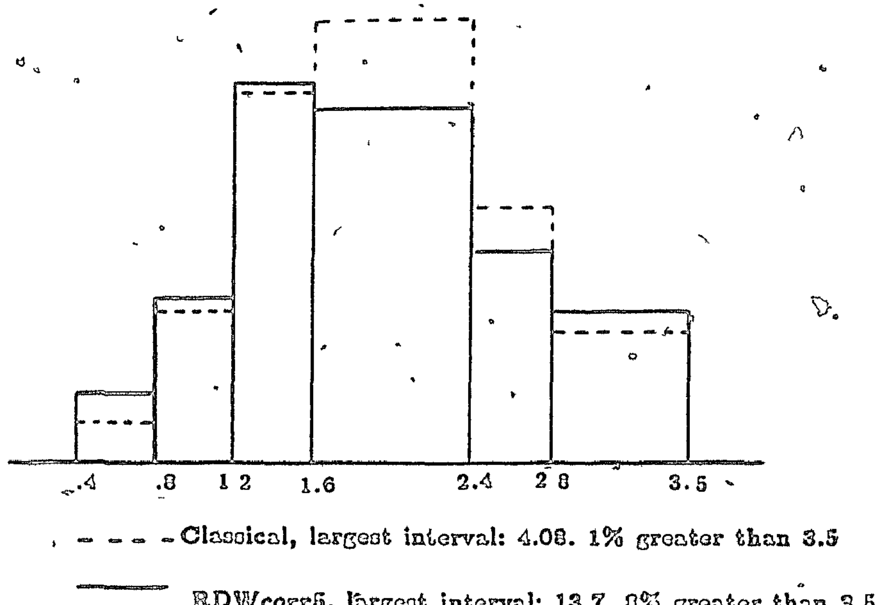


Distribution of lengths of intervals. 500 samples of size $n = 13$.

Predictors: Set 2. Model $y = 2x + 1 + \varepsilon$, with

standard normal errors, ε. Confidence level 90%.

Figure 5.2.5



Distribution of lengths of intervals. 500 samples of size $n = 7$,

Predictors: Set 3. Model $y = 2x + 1 + \varepsilon$, with

standard normal errors, ε. Confidence level 90%

Figure 5.2.6

Method	Set 1 ($n = 7$)	Set 2 ($n = 13$)	Set 3 ($n = 7$)	Set 4 ($n = 13$)
Classical	90 2.3, 2.4	91 0.93, 0.90	89 3.2, 3.4	90 2.1, 2.0
RDW $corr(n-2)$	88 2.1, 2.7	91 0.77, 0.86	87 3.2, 3.8	89 1.7, 1.9
RD $corr(n-2)$	89 2.2, 2.5	91 0.80, 0.89	86 3.3, 3.6	90 1.8, 2.0

Regression model $y = 2x + 1 + \varepsilon$
ε standard normal with 10% contamination at 5.
Entries: Coverage of 2 (percent) / median, average length
500 samples, confidence level 90%.
Shift correction factor: $\frac{n-3+w_{(1)}}{n}$. Huber's $c = 1.5$.

Table 5.2.7

Method	80%	90%	95%
Classical	81 0.68, 0.85	89 0.93, 1.2	94 1.2, 1.5
RDW $corr5$	80 0.70, 0.94	87 0.88, 1.2	89 1.0, 1.4
RD $corr5$	80 0.71, 0.90	87 0.89, 1.1	89 1.0, 1.3

Regression model $y = 2x + 1 + \varepsilon$
Predictors χ^2_3, errors t_4. $n = 7$.
Entries: Coverage of 2 (percent) / median, average length
500 samples, three levels of confidence.
Shift correction factor: $\frac{n-3+w_{(1)}}{n}$. Huber's $c = 1.5$.

Table 5.2.8

5.3 Programming details

Fortran programs appear in the back pocket of the thesis. The program SLREG calculates several R intervals for the slope. Data is generated (or read in) from the routine RGEN. The predictors are normalized in the routine NORMX, and weights are calculated in GETHAT.

The routine LINEST calculates both the classical and the robust regression line and scale estimate. The classical estimates are obtained by a call to the NAG routine G02CAF. For the M-estimates, there is a loop using Huber's iterative procedure to calculate M-estimates (1981, Section 7.8), followed by a call to the NAG rootfinder C05NBF.

The configuration is set up in SLSET1 (diagonal A matrix), with a follow-up call to SLSET2 for a non-diagonal A matrix. When the configuration and the shift condition have been specified, the program calls ROBCI (see Section 4.5).

Chapter 6 Conclusion

6.1 Summary

For a multiparameter problem, we take a sample of size n and estimate all parameters, $\underline{\eta}$, by M-estimates, $\hat{\underline{\eta}}$. To construct a confidence interval for a real-valued function, $\theta = \theta(\underline{\eta})$, we work with the observed scores (score functions for M-estimation), following the procedure described in Section 3.8. We obtain confidence intervals with error $O_P(\frac{1}{n})$, for $\underline{\eta}$ in a compact subset of the parameter space (and under the regularity conditions stated in Chapter 3).

The numerical results of Chapters 4 and 5 demonstrate the performance of the R intervals. We draw attention to those R intervals incorporating a shift correction. (All R intervals use a reduced effective sample size.)

For the location (unknown scale) problem of Chapter 4, we also demonstrate the good performance of intervals based on Kent (1982). These intervals should be seen as evidence to support Ronchetti's τ-test (1982), which we have not programmed.

The shortcoming of the Kent and Ronchetti intervals is that the test procedure increases in complexity as the number of nuisance parameters increases. Once the initial M-estimates have been calculated, calculation of the R intervals is a one-dimensional problem (solve for α_0), no matter how many nuisance parameters are involved. For the Kent and Ronchetti intervals, however, the multidimensional nuisance parameter is re-estimated for every θ_0 under test.

In order to generalize these results to other types of estimates, the essential feature is the score functions. We need a parameter estimate $\tilde{\underline{\eta}}$ defined as the solution of a system of equations

$$\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\Psi}}(X_i,\underline{\eta})=\underline{0},$$

with the score functions $\underline{\tilde{\Psi}}$ satisfying regularity conditions like those listed in Section 3.3. In particular, the matrix $D\underline{\tilde{\Psi}}$ of derivatives of the score functions must have a well-behaved first moment, A. Certainly, A^{-1} must exist.

The R, Kent and Huber intervals demonstrate that the scores hold a lot of information, and should be investigated further.

We repeat the basic procedure:

Use the score functions to construct a modified sample. Approximate the distribution of this modified sample with an exponentially tilted bootstrap. Use the Lugannani-Rice tail area approximation to avoid bootstrap re-sampling.

6.2 Plans for the immediate future

We are presently working on a multiple regression problem, involving three predictors,

$$y = \mu + \eta_1 x_1 + \eta_2 x_2 + \eta_3 x_3 + \varepsilon.$$

In this context, we plan to construct confidence intervals for real valued functions of $\underline{\eta}$, for example $\eta_1 - \eta_2$. We also plan to consider analysis of variance models.

For higher dimensional problems, we need to change our programs considerably. The programs used for Chapters 4 and 5 call packaged routines wherever possible (primarily to give credence to our results). The packages have shortcomings. For example, for multiple regression, most of the relevant NAG and IMSL routines do not allow for variable dimensions. It would be efficient to switch to packages designed for robust regression, for example LINWDR, ROBETH or TROLL (see Hampel et al, 1986). We have not been concerned with efficiency in our programs. We simply want to demonstrate that our techniques work.

We repeat that, for multidimensional problems, the only hurdle is the initial calculation of the M-estimates. After that, it is a relatively simple matter to set up the configuration and then call ROBCI, the interval routine. For several parameters, calculation of the M-estimate is a non-trivial problem.

6.3 Open questions

We have raised more questions than we have answered. The two main problems for the R intervals are (i) the effective sample size, and (ii) the correction to the shift condition.

The first of these problems is, surely, geometric in nature: by moving to the configuration, we have reduced the dimension of the sample. We also ask, how does effective sample size change when Huber's truncation constant c is altered? Closely related to these is the question, what is the loss of information when we move from the sample to the configuration? The discussion in Section 3.7 suggests that the answer to this last question may shed light on our second problem, the appropriate shift condition.

This second problem fascinates us. How much should we skew (i.e., exponentially tilt) our first guess at the distribution of the configuration? What is the appropriate definition of location for a skewed distribution? Bickel and Lehmann (1975) suggest that, possibly, M-estimates are not the appropriate tool.

We have set out to find a balance between the purely computational approach of the Bootstrap and the purely theoretical approach of the classical maximum likelihood ratio test procedure. We believe that we have achieved some measure of success.

References

Bahr, B. von (1967).

On the Central Limit Theorem in $\mathfrak{R}^k$.

Ark. Mat. 7 61–69.

Barndorff-Nielsen, O. and Cox, D.R. (1979).

Edgeworth and saddlepoint approximations with statistical applications.

J. Roy. Statist. Soc. B 41 279–312.

Bhattacharya, R.N. and Ghosh, J.K. (1978).

On the validity of the formal Edgeworth expansion.

Ann. Statist. 6, 434–451.

Bickel, P.J. and Doksum, K.A. (1977).

Mathematical Statistics. Holden-Day, San Francisco.

Bickel, P.J., Lehmann, E.L. (1975).

Descriptive statistics for nonparametric models II. Location.

Annals of statistics, 3,5, 1045–1069.

Bleistein, N. and Handelsman, R.A. (1975).

Asymptotic Expansions of Integrals.

Holt, Rinehart and Winston, N.Y.

Copson, E.T. (1965).

Asymptotic Expansions. Cambridge University Press.

Cox, D.R. and Hinkley, D.V. (1974).

Theoretical Statistics. Chapman and Hall, London.

Daniels, H.E. (1954).

Saddlepoint approximations in statistics. *Ann. Math. Statist.* 25 631–650

Daniels, H.E. (1983).

Saddlepoint approximations for estimating equations. *Biometrika* 70, 89–96.

Daniels, H.E. (1986).

Tail probability approximations. To appear in *IS Review.*

Di Ciccio, T. and Tibshirani, R. (1986).

Approximating the profile likelihood through Stein's least favourable family.

Technical Report 1, Dept. of Statistics, University of Toronto.

Efron, B. (1986).

Better bootstrap confidence intervals. *Technical Report, Dept. of Statistics, Stanford University.*

Feller, W., (1971).

An introduction to probability theory. Vol. 2, 2nd edition. Wiley, New York.

Field, C.A. (1982).

Small sample asymptotic expansions for multivariate M-Estimates.

Annals of Statistics, 10, 3, 672–689.

Field, C.A. and Hampel, F.R. (1982).

Small sample asymptotic distributions of M-estimators of location. *Biometrika* 69, 1, 29–46.

Gnedenko, R.V. and Kolmogorov, A.N., (Chung, K.L. -translator) (1954).

Limit distributions for sums of independent random variables. Addison-Wesley, Cambridge, Mass.

Hampel, F.R., Ronchetti, E., Rousseeuw, P.J., Stahel, W.A. (1986).

Robust statistics. The approach based on Influence Functions. Wiley, New York.

Huber, P.J. (1967).

The behaviour of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley symposium on Mathematical Statistics and Probability,* 1. University of California Press, Berkeley.

Huber, P.J. (1981).

Robust Statistics. John Wiley, N.Y.

Kent, J.T. (1982).

Robust properties of likelihood ratio tests. *Biometrika* 69, 19–27.

Kullback, S. (1960).

Information Theory and Statistics. Wiley, N.Y.

Lugannani, R. and Rice, S.O. (1980).

Saddle point approximation for the distribution of the sum of independent random variables.. *Adv. Appl. Prob.* 12, 475–490.

Marcinkiewicz, J. (1938).

Sur une propriété de la loi de Gauss.

Math. Zeitschr., 44, 612.

Olver, F.W.J. (1974).

Asymptotics and Special Functions. Academic Press, N.Y.

Rice, S.O. (1968).

Uniform asymptotic expansions for saddle point integrals – application to a probability distribution occuring in noise theory.

Bell Syst. Tech. J. 47, 1971–2013.

Ronchetti, E. (1982).

Robust testing in linear models: The infinitessimal approach.

Disertation, Swiss Fed. Institute of Tech., Zurich.

Tibshirani, R. (1986).

Bootstrap confidence intervals.

Technical Report, Dept. of Statistics, Stanford University.

```
      PROGRAM MATLR(INPUT,OUTPUT,TAPE5,TAPE1)
      REAL KLNGTH(12),SCORE(2),THETA(2),W(3),X(100)
     _,LNGTH(12),END1(12),END2(12)
      INTEGER ICOUNT(12),N,M,K,K1,I,L,I,LL,L
     _,ISCOV(12)
      DOUBLE PRECISION DSEED
      EXTERNAL FCN
      COMMON/PIV/PI,RT2PI
      COMMON/CNB/C,F,F1,F2,Z,BETA,ERR
      COMMON/CNSTNT/DELTA,GAMMA,EPSL,ZETA
      COMMON/SAMPNO/II
      COMMON/TOLER/TOL1,TOL2,TOL3,TOL4,TOL5
      COMMON/CHIPP/[illegible]
      COMMON/CONLVL/CLEVEL
      COMMON/HUBEST/I1,C1
      COMMON/TABLE/TABLEN,TABLET
      COMMON/TAOVAL/WT[illegible],E[illegible]
      COMMON/[illegible]/X
      [illegible]T=
      THETA=
      L[illegible]=
      [illegible]AL=
      Z[illegible]=
      TWO=
      PI=X[illegible]F(TWO)
      X[illegible]=
      X[illegible]=
      T[illegible]=
      RT2PI=SQRT(PI*TWO)
      TOL1=
      TOL2=
      TOL3=
      TOL4=
      TOL5=
C     WRITE(1,[illegible])PI,TOL1,TOL2,TOL3,TOL4,TOL5
      CALL PRELIM(DSEED,C,N,M,CLEVEL,I[illegible],X[illegible],SCALE,
     _TABLEN,TABLET,[illegible],ERR,[illegible])
******
*       C IS HUBER'S C
*       N IS SAMPLE SIZE
*       ERR USED TO ADJUST EFFECTIVE SAMPLE SIZE
*       M IS NUMBER OF SAMPLES
*       CLEVEL IS CONFIDENCE LEVEL FOR INTERVAL
*       IDIST KEEPS TRACK OF DISTRIBUTION FROM WHICH
*              SAMPLE IS GENERATED
*       ISTOP  S [illegible] OF DISTRIBUTION, IF RELEVANT
*       TABLEN FROM NORMAL TABLES
*       TABLET FROM T-TABLES
******
      F=FLOAT(N)
      F1=FLOAT(N-1)
      F2=FLOAT(N-2)
      ERR=F1
      ERRN1=F1-ERR
      ERRN2=ERRN1-ERR
      Z=SQRT(F)
C     WRITE(3,37)ERR,ERRN1,ERRN2
C     WRITE(1,37)ERR,ERRN1,ERRN2
      WRITE(3,37)ERRN1
C     WRITE(1,37)ERR,ERRN1,ERRN2
      K=
C     K=
      [illegible]=
******
*       CONFIDENCE K SET TO NUMBER OF TYPES OF CONFIDENCE INTERVALS
*       ICOUNT IS COUNTER FOR COVERING PROPORTIONS
*       LENGTH ACCUMULATES SUMS OF LENGTHS OF INTERVALS
*              EVENTUALLY STORES AVERAGE LENGTH OF INTERVALS
*       ISCOV RECORDS ALL COVERING PROPORTIONS ON TAPE 1)
*       LENGTH RECORDS ALL INTERVAL LENGTHS ON TAPE 1)
******
      DO [illegible] I=1,K
      KLNGTH(I)=
      ICOUNT(I)=
    [illegible] CONTINUE
      CALL CONST(DELTA,BETA,EPSL,ZETA,C)
      BETJLD=BETA
      GAMMA=SQRT(BETA)/DELTA
      BETA=BETA*F1/F
      DO [illegible] II=1,M
******
```

[redacted]

```
      F=FLOAT(N-1)
      ERRN=F1
      ERRN1=F-ERRN
      ERRN2=ERRN1-ERRN
      Z=SQRT(F)
C     WRITE(6,370)ERRN,ERRN1,ERRN2
C     WRITE(1[illegible],37[illegible])ERRN,ERRN1,ERRN2
      WRITE(6,370)ERRN1
C     WRITE([illegible],370)ERRN,ERRN1,ERRN2
      K=3
C     K=
      THOUT=
******
*         COMMENT: K SET TO NUMBER OF TYPES OF CONFIDENCE INTERVALS
*         ICOUNT IS COUNTER FOR COVERING PROPORTIONS
*         LENGTH ACCUMULATES SUMS OF LENGTHS OF INTERVALS
*                 EVENTUALLY STORES AVERAGE LENGTH OF INTERVALS
*         ISCOV RECORDS ALL COVERING INFORMATION, ON TAPE 1)
*         ILNGTH RECORDS ALL INTERVAL LENGTHS, ON TAPE 1)
******
      DO 60 I=1,K
        RLNGTH(I)=0.
        ICOUNT(I)=0
   60 CONTINUE
      CALL CONST(DELTA,BETA,EPSLN,ZETA,C)
      BETOLD=BETA
      GAMMA=SQRT(BETA)/DELTA
      BETA=BETA*F1/F
      DO [illegible] II=1,M
******
*         GENERATE DATA
*         CALCULATE [illegible] [illegible] ORDERED [illegible] AVERAGES [illegible]
******
      CALL GENER(X,[illegible],[illegible],[illegible],[illegible],
     -CONTAM)
      CALL [illegible](X,[illegible],[illegible],[illegible],TABLET,CLASS1,CLASS2,
     -AVERAG)
CC
CC    WRITE(6,[illegible])[illegible]
******
*         INITIAL ESTIMATE FOR CALCULATION OF M-ESTIMATES
******
      THETA(1)=T1
      THETA(2)=(S1*X[illegible]+[illegible])*X[illegible])/[illegible]
      CALL [illegible](FCN,ITHETA,THETA,[illegible],[illegible],[illegible],WA,LWA,IFAIL)
      T1=THETA(1)
      S1=THETA(2)
******
*         T1 IS M-EST OF LOCATION
*         S1 IS M-EST OF SCALE
*         SCALE CANNOT BE NEGATIVE
******
      IF(S1.LT.ZERO)THEN
        THETA(2)=-S1
        CALL [illegible](FCN,ITHETA,THETA,[illegible],[illegible],WA,LWA,IFAIL)
      ENDIF
      S1=THETA(2)
      IF(S1.LT.ZERO)THEN
        WRITE(6,50[illegible])
        STOP
      ENDIF
      ASDT1=S1*G[illegible]/Z
******
*         ASDT1 IS ASYMPTOTIC S.D. OF T1
*         STEPS IS USED FOR FIRST JUMP IN CALC OF C.I.
******
CC
CC    WRITE(6,7[illegible])T1,S1,ASDT1,SCORE(1),SCORE(2)
      [illegible]=(CLASS2-CLASS1)/TABLET/TWL
C     STEPS=ASDT1*TABLET/[illegible]*XNINE
      STEPS=ASDT1*TABLET
      METHOD=1
      CALL CLASS(CLASS1,CLASS2,LNGTH,METHOD,XL,IC,ICOUNT,
     -ISCOV,ILNGTH)
      END1(METHOD)=CLASS1
      END2(METHOD)=CLASS2
      METHOD=2
      ERRN=ERRN1
      ERRN=ERRN1
C     SHCOR=S1/Z/F*[illegible]
C     SHCOR=(F-2.)/F
      CALL SETLOC(T1,S1,SIZE,METHOD)
      CALL FOSCI(STEPS,RLNGTH,ICOUNT,XL,IC,T1,
     -SIZE,METHOD,IF[illegible]ST,ISCOV,ILNGTH,SHCOR,
     -END1(METHOD),END2(METHOD))
      METHOD=3
      IF(K.EQ.4)GO TO [illegible]
C     ERRN=ERRN2
C     SHCOR=ZERO
      SHCOR=X[illegible]
      CALL FOSCI(STEPS,RLNGTH,ICOUNT,XL,IC,T1,
     -SIZE,METHOD,IF[illegible],ISCOV,ILNGTH,SHCOR,
     -END1(METHOD),END2(METHOD))
```

```
      STEP=ALOUT/T,SEST
      METHOD=
      CALL CLASS(CLASSI,CLASS,LENGTH,METHOD,XLOC,ICOUNT,
     1ISOBV,LENGTH)
      END1(METHOD)=CLASSI
      END2(METHOD)=CLASS2
      METHOD=
      EFF1=EFFN
      EFF1=EFFN1
      SHOCK=SQ/Z/F4.0
      SHOCK=(F-1. )/F
      CALL SETLOC(T1, 1,SIZE,METHOD)
      CALL ROBSCL(STEP,LENGTH,ICOUNT,XLOC,T1,
     1SIZE,METHOD,IFROST,ISOBV,SLENGTH,SHOCK,
     2END1(METHOD),END2(METHOD))
      METHOD=
      IF(K.LE.4)GO TO 99
      EFF1=EFFN
      SHOCK=ZERO
      SHOCR=XONE
      CALL ROBSCL(STEP,LENGTH,ICOUNT,XLOC,T1,
     1SIZE,METHOD,IFROST,ISOBV,SLENGTH,SHOCK,
     2END1(METHOD),END2(METHOD))
      METHOD=4
      IF(K.LE.5)GO TO 99
      EFF1=EFFN
C     SHOCK=ZERO
      SHOCR=XONE
      CALL ROBSCL(STEP,LENGTH,ICOUNT,XLOC,T1,
     1SIZE,METHOD,IFROST,ISOBV,SLENGTH,SHOCK,
     2END1(METHOD),END2(METHOD))
C     METHOD=5
C     CALL ROBSCL(STEP,LENGTH,ICOUNT,XLOC,T1,
C    1SIZE,METHOD,IFROST,ISOBV,SLENGTH,SHOCK,
C    2END1(METHOD),END2(METHOD))
C     IF(K.LT.6)GO TO 99
      METHOD=
    9 CALL SETLOC(T1,1,SIZE,METHOD)
C     CALL ROBSCL(STEP,LENGTH,ICOUNT,XLOC,T1,
C    1SIZE,METHOD,IFROST,ISOBV,SLENGTH,SHOCK,
C    2END1(METHOD),END2(METHOD))
C     METHOD=ESTMD
C     METHOD=
      METHD=S.*DELTOLD/DELTA
      CALL KENTC1(STEP,LENGTH,XLOC,ICOUNT,
     1METHOD,ISOBV,SLENGTH,END1(METHOD),
     2END2(METHOD))
C     METHOD=7
C     METHD=METHD1
C     CALL KENTC1(STEP,LENGTH,XLOC,ICOUNT,
C    1METHOD,ISOBV,SLENGTH,END1(METHOD),
C    2END2(METHOD))
C     METHOD=
C     METHOD=
      METHOD=
      CALL HUBCL(LENGTH,ICOUNT,XLOC,METHOD,
     1ISOBV,SLENGTH,END1(METHOD),END2(METHOD))
C  99 WRITE(6,   )(SLENGTH(METH),METH=1,K),
C    1(ISOBV(METH),METH=1,K)
   99 WRITE (1,   )(END1(METH),END2(METH),
     1METH=1,K)
  1 9 CONTINUE
      WRITE(6,  )
      CALL OUTH(LENGTH,K,M)
      CALL COVPRCP(ICOUNT,K,M)
      CPROST=FLOAT(IFROST)/4. /FLOAT( )
      WRITE(6,4 )CPROST
      ISAVE=
CC    CALL PRTLAP(D LED,  N,M,CLEVEL,IOUT,XLOC,SCALE,
CC   1TABLE,TABLEN,   T,ITUR,EFFN,ISAVE)
    5 FORMAT(1X,"MEDIAN:",F9.3,2X," CORRECTED MAD:",F9.3)
    7    FORMAT(" HUBER ESTIMATES, LOC*TION:",F9.3,1X,"SCALE:",F9.3,
     1   3X,"EST ASYMP STD. OF T1:",F9.3,3X,"COR*S:",
     2   F9.3," ,",F9.3)
   2  FORMAT(1X,"AVERAGE LENGTHS OF INTERVALS")
C  3  FORMAT( 2X,"PI=",F10.4,2X,"TOL1=",F10.4," TOL2=",F10.4,
C    1" TOL3=",F5.4," TOL4=",F5.4," TOL5=",F5.4)
  30  FORMAT(" NEGATIVE SE TO 3.8F")
  37  FORMAT(" METHOD 1: CLASSICAL"/
     13X,"2: ROBUST, DIAG A-MATRIX, EFF SAMP SIZE: EFFN"/
     21X,"SHIFT CORRECTION"/
     33X,"3: ROBUST, DIAG A-MATRIX, EFF SAMP SIZE: EFFN"/
     419X,"SHIFT CORRECTION"/
C    53X,"4: ROBUST, DIAG A-MATRIX, EFF SAMP SIZE:",F5.2/
C    63X,"5: ROBUST, PARAMETRIC SET OF T,",
C    7"  EFF SAMP SIZE:",F5.2/
     8"         6: KENT WITH DIAGONAL A-MATRIX"/
C    9"         7: KENT WITH NON-PARAMETRIC SET OF A"/
     A"          : HUBER T-(2-1)"/
     B"   MULTIPLICATIVE SHIFT CORRECTION , (1-2. )/N"/
     C"   EFFN=",F5.2)
   4  FORMAT("  ESTIMATE          ",F5.2)
```

```
      WRITE(3,4 )ERROR
      IREST=1
CC    CALL PRELIM(DIFFUS,N,M,CLEVEL,IDIST,XLOC,SCALE,
CC   -TABLET,TABLEN,IRELT,ITOP,EFFN,ISAVE)
   5  FORMAT(20X,*MEDIAN*,F9.5,2X,* CORRECTED MAD*,F9.3)
   7    FORMAT(* HUBER EST. *ATIS, LOCATION*,F9.3,11X,*SCALE*,F9.3,
     +  5X, *EST. RATIO F.D. OF FIT*,F9.5,5X,*COR*,
     -  F9.3,* *,F9.3)
   2  FORMAT(11X,*RESIDUAL LENGTHS OF THE VALUE*)
CC 3  FORMAT( 5X,*P =*,F10.5,5X,*TLL1=*,F10.5 ,* TLL2=*,F9.5,
     -* TLL3=*,F9.5,* TLL4=*,F9.5,* TLL5=*,F5.4)
  36  FORMAT(* NEGATIVE SE TH E SABE*)
  37  FORMAT(* METHOD OF CLASSICAL*/
     -5X,*1: ADJUST, DIAG A-MATRIX, EFF SAMP SIZE: EFFN*/
     -7X,*SHIFT CORRECTION*/
     -3X,*3: ADJUST, DIAG A-MATRIX, EFF SAMP SIZE: EFFN*/
C    -19X,*SHIFT CORRECTION*/
C    -5X,*4: ADJUST, DIAG A-MATRIX, EFF SAMP SIZE:*,F5.2/
C    -5X,*5: ADJUST, NON-PARAMETRIC,*,
C    -* EFF SAMP SIZE:*,F5.2/
     -*           6: KENT WITH DIAGONAL A-MATRIX*/
C    -*           7: KENT WITH NON-PARAMETRIC EST OF A*/
     -*           8: HUBER T-(N-1)*/
     -*   MULTIPLICATIVE SHIFT CORRECTION , (I=1, )/A*/
     -*   EFF N=*,F5.2)
   4  FORMAT(*  ADJUSTED DIAG APPROXIMATION*,F9.2)
C11   FORMAT(2F9.5,I5)
C1    FORMAT(2F9.5,I5)
 100  FORMAT(2F9.3)
      STOP
      END

      PROGRAM SERIG(INPUT,OUTPUT,TAPE5,TAPE1)
      REAL Y(1  ),X1(1  ),WGHT(1  ),WGHT1(1  )
     -,LENGTH(1  ),THETA(3),FLNGTH(1 ),R(1 ),
     - DZ(1 )
      DOUBLE PRECISION ISEED
******
*         THETA STORES PARA  ESTIMATES: THETA(1)=INTERCEPT
*                                       THETA(2)=SCALE OF ERROR
*                                       THETA(3)=SLOPE
******
      INTEGER ICOUNT(1  ),TDIV(12)
      COMMON/CLEVL/CLEVEL
      COMMON/SW/ITERM/ISTEP
      COMMON/SAMPL/Y,X1
      COMMON/CNB/CN,FSTEP ,Z,BETA,EFF
      COMMON/WGHT1/WGHT
      COMMON/TABLE/TABLE ,TABLET
      ZERO=0.
      XONE=1.
      TWO=2.
      CALL PRELIM(IDIST,PDV ,   ,   ,IFYX,IFXA,X1, ,
     -   TAB)
      EFF = EFF
      EFFN1=EFF -EFF
C     EFFN1=E-1.
C     WRITE(6,3  )EFF ,EFFN ,EFF  IS  EFF  SQIF
      WRITE(6,26)
      DLT =3. ,ATRE/F
      IFRUST=
      N=M+1
      K=0
      DO   I=1,N
        LENGTH( )=ZERO
        ICOUNT( )=
      CONTINUE
      IF(IDIST.EQ.T  )THEN
*****
*         PARAMETRIZE PREDICTORS
*      WEIGHTED REGRESSION
*****
        CALL NORMX(SIZE,XLOC)
        CALL OUTMAT(SHOCKS,WGHT1)
        WRITE(6,  )WGHT1(N1)
          DO 8 J=1,N1
            WGHT(J)=WGHT1(J)
   8    CONTINUE
        CALL GROVEC(WGHT,N,INDY)
        WGHTMN=WGHT(1)
        WRITE(6,27)WGHTMN,(WGHT1(I),I=1,N1)
      ENDIF
      DO 2   LE RP=1,M
        CALL NOR M(SAMP,IDIST,ISAVE,IFY,IFX,
     -  ISEED,INTER)
        IF(IDIS .EQ.  )GOTO 119
*****
*       STANDARDIZE PREDICTORS
*      WEIGHTED REGRESSION
*****
        CALL NORMX( ZL,XL )
        CALL OUTMAT(SHOCKS,WGHT1)
```

```
******
*        [illegible]
*       WEIGHTED REGRESSION
******
         CALL NORMX(SIZE,XLDC)
         CALL GETHAT(SHCLK1,WGHT1)
C        WRITE(6,[illegible])WGHT1(N1)
            DO [illegible] J=1,N1
              WGHT(J)=WGHT1(J)
   [illegible]   CONTINUE
         CALL [illegible]VEC(WGHT,N,INDY)
         WGHT[illegible]=WGHT(1)
         WRITE(6,[illegible])WGHT[illegible],(WGHT1(I),I=1,N1)
      [illegible]
      DO [illegible] IP=1,M
         [illegible]
     -   [illegible]
         IF([illegible].GT.[illegible])GOTO 19
******
*        [illegible]
*       WEIGHTED REGRESSION
******
         CALL [illegible]X([illegible]IZE,XLDC)
         CALL GETHAT(SHCOR1,WGHT1)
            DO 18 J=1,[illegible]
              WGHT(J)=WGHT1(J)
   18       CONTINUE
         CALL [illegible]VEC(WGHT,N,INDY)
         WGHT[illegible]=WGHT([illegible])
CC
CC       WRITE(6,[illegible])WGHT[illegible],(WGHT1(I),I=1,N1)
   19       DO 21 J=1,N1
              WGHT(J)=WGHT1(J)
   21       CONTINUE
CC
CC       WRITE(6,22)(WGHT1(I),I=1,N1)
         METHOD=
         CALL [illegible](STEP,[illegible]IZE,XLDC,[illegible],THETA,TCOUNT,
     -   [illegible]GTH,METHOD,[illegible]
         SHCOR1=SHCOR[illegible]*THETA(2)
      [illegible]TEP=[illegible]*TABLET
         SHCOR1=SHCOR1*1.7
C        SHCOR1=(F2-[illegible])/F[illegible]
C        SHCOR1=(EFF[illegible]-1.2)/EFF[illegible]
C        SHCOR1=(EFF[illegible]-[illegible]+WGHT[illegible])/EFF[illegible]
         SHCOR1=(F-[illegible]+WGHT[illegible])/F
******
*        [illegible] CORRECTION
*        EFFECTIVE SAMPLE SIZE: [illegible]
******
         EFF[illegible]=EFF[illegible]
         SHCOR=SHCOR[illegible]
******
*        SIMPLE CONFIGURATION
******
         CALL [illegible]T1(SCOR,THETA)
         ESTP[illegible]=THETA(3)
CC
CC       WRITE(6,[illegible])SHCOR
         CALL [illegible](STEP,[illegible],TCOUNT,TROVAL,[illegible]
     -   [illegible]ZE,METHOD,AF[illegible]OST,[illegible],LNGTH,SHCOR,
     -   [illegible](METHOD),[illegible](METHOD))
         IF(K.[illegible])GOTO 99
         METHOD=[illegible]
******
*        [illegible] SHIFT CORRECTION
******
         SHCOR=X[illegible]
         CALL ROBC[illegible](STEP,[illegible],TCOUNT,TROVAL,[illegible]STPAR,
     -   SIZE,METHOD,[illegible]COST,[illegible],LNGTH,SHCOR,
     -   [illegible](METHOD),[illegible](METHOD))
         METHOD=[illegible]
******
*        EFFECTIVE SAMPLE SIZE: [illegible]
******
         EFF[illegible]=EFF[illegible]
         CALL ROBC[illegible](STEP,[illegible],TCOUNT,TROVAL,[illegible]STPAR,
     -   SIZE,METHOD,[illegible]COST,[illegible],LNGTH,SHCOR,
     -   [illegible](METHOD),[illegible](METHOD))
         DO [illegible] J=1,[illegible]
            WGHT(J)=X[illegible]
   [illegible]   CONTINUE
         WGHT([illegible])=F
         METHOD=[illegible]
******
*        DROP WEIGHTS
*        USING BACK SHIFT CORRECTION
*        EFFECTIVE SAMPLE SIZE: [illegible]
******
         EFF[illegible]=EFF[illegible]
         SHCOR=SHCOR[illegible]
         CALL [illegible](STEP,[illegible]IZE,XLDC,[illegible],THETA,TCOUNT,
     -   LNGTH,METHOD,[illegible],LNGTH,[illegible])
         CALL [illegible]SET[illegible]([illegible],THETA)
```

```
      CALL   SUBC1(STEP,RLNGTH,ICOUNT,TRUVAL,ESTPAR,
     _SIZE,METHOD,IFROST,ISCUV,SLNGTH,SHCCK,
     _END1(METHOD),END2(METHOD))
      METHOD=4
******
*     EFFECTIVE SAMPLE SIZE: N-2-INCREMENT
******
      EFF4=EFFN1
      CALL  SUBC1(STEP,RLNGTH,ICOUNT,TRUVAL,ESTPAR,
     _SIZE,METHOD,IFROST,ISCUV,SLNGTH,SHCCK,
     _END1(METHOD),END2(METHOD))
      DO 32 J=1,4
        WGHT(J)=XONE
   32 CONTINUE
      WGHT(N1)=F
      METH D=5
****** 
*     DROP WEIGHTS
*     BRING BACK SHIFT CORRECTION
*     EFFECTIVE SAMPLE SIZE: N-2
******
      EFF5=EFFN1
      SHCOK=SHCOK2
      CALL  LINEST(STEP,SIZE,XLOC,IRLV,L,THETA,ICOUNT,
     _RLNGTH,METHOD,ISCUV,SLNGTH,D1,D2)
      CALL  SLOPE1(SCOK,THETA)
      CALL  SUBC1(STEP,RLNGTH,ICOUNT,TRUVAL,ESTPAR,
     _SIZE,METHOD,IFROST,ISCUV,SLNGTH,SHCCK,
     _END1(METHOD),END2(METHOD))
      METHOD=6
******
*     CORRECTION TERM IN CONFIGURATION
******
      CALL  SLOPE1(SCOKR,THETA)
      CALL  SUBC1(STEP,RLNGTH,ICOUNT,TRUVAL,ESTPAR,
     _SIZE,METHOD,IFROST,ISCUV,SLNGTH,SHCCK,
     _END1(METHOD),END2(METHOD))
   99_WRITE(6,100)(END1(METH),END2(METH),
     _METH=1,K)
  500 CONTINUE
      WRITE(6,500)
      CALL  SLNGTH(RLNGTH,K,1)
      CALL  COVPROB(ICOUNT,K,N)
      UFROST=FLOAT(IFROST)/FLOAT(N)/5.
CC
CC    WRITE(6,100)UFROST
   55 FORMAT(/" NEW WGT:",(/F10.5))
  557 FORMAT(/11X,"AVERAGE LENGTHS OF INTERVALS")
  590 FORMAT(" METHOD 1: CLASSICAL"/
     _13X,"2: SIMPLE CONFIG AND SHIFT CORR, EFF SAMP SIZE:",F5.2/
     _11X,"WEIGHTED REGRESSION"/
     _13X,"3: SIMPLE CONFIG, NO SHIFT CORR, EFF SAMP SIZE:",F5.2/
     _11X,"WEIGHTED REGRESSION"/
     _13X,"4: SIMPLE CONFIG, NO SHIFT CORR, EFF SAMP SIZE:",F5.2/
     _11X,"WEIGHTED REGRESSION"/
     _13X,"5: SIMPLE CONFIG, SHIFT CORR, EFF SAMP SIZE:",F5.2/
     _11X,"NO WEIGHTS"/
     _13X,"6: CORR TERM IN CONFIG, EFF SAMP SIZE:",F5.2/
     _11X,"SHIFT CORRECTION, NO WEIGHTS"/
     _=" SHIFT CORRECTION IS (EFFN -2)/EFF  EFFN2=WGHT(N1)-1")
  570 FORMAT(" METHOD 1: CLASSICAL"/
     _13X,"2: SIMPLE CONFIG AND SHIFT CORR, EFF SAMP SIZE:EFFN"/
     _11X,"WEIGHTED REGRESSION"/
     _13X,"3: SIMPLE CONFIG, NO SHIFT CORR, EFF SAMP SIZE:EFFN"/
     _11X,"WEIGHTED REGRESSION"/
     _13X,"4: SIMPLE CONFIG, NO SHIFT CORR, EFF SAMP SIZE:N-2"/
     _11X,"WEIGHTED REGRESSION"/
     _13X,"5: SIMPLE CONFIG, SHIFT CORR, EFF SAMP SIZE:EFFN"/
     _11X,"NO WEIGHTS"/
     _13X,"6: CORR TERM IN CONFIG, EFF SAMP SIZE:EFFN"/
     _11X,"SHIFT CORRECTION, NO WEIGHTS"/
     _" SHIFT FACTOR IS (N-3.*WGHT(N1))/N  EFFN=1-2")
  555 FORMAT(" FRUSTRATION PROPORTION:",F5.2)
   56 FORMAT(" SHIFT CORRECTION:",F6.4)
 1000 FORMAT(1F9.5)
      STOP
      END

      REAL FUNCTION PSIP(Y,C)
      REAL Y,C
******
*     CALCULATES PSI'(Y) WITH HUBER CONSTANT C
******
      D=-C
      ZERO=0.
      IF(Y.GT.D.AND.Y.LT.C)THEN
        PSIP=1.0
        RETURN
      ENDIF
      PSIP=ZERO
      RETURN
      END
```

```
      DO  I=1,K
```

```
     -[illegible] SHIFT CORRECTION [illegible]
     -" SHIFT FACTOR [illegible]"
  355 FORMAT(" FRUSTRATION PROPORTION:",F5.2)
  500 FORMAT(" SHIFT CORRECTION:",F6.4)
 1000 FORMAT(12F9.3)
      STOP
      END

      REAL FUNCTION PSIP(Y,C)
      REAL Y,C
******
*         CALCULATES PSI'(Y) WITH HUBER CONSTANT C
******
      D=-C
      ZERO=0.
      IF(Y.GT.D.AND.Y.LT.C)THEN
        PSIP=1.0
        RETURN
      ENDIF
      PSIP=ZERO
      RETURN
      END

      REAL FUNCTION PSI(Y,C)
      REAL Y
******
*         CALCULATES PSI(Y) WITH HUBER CONSTANT C
******
      D=-C
      IF(Y.LT.D)THEN
        PSI=D
        RETURN
      ENDIF
      IF(Y.GT.C)THEN
        PSI=C
        RETURN
      ENDIF
      PSI=Y
      RETURN
      END

      REAL FUNCTION TRPSI(Y,C)
      REAL Y,C
******
*         CALCULATES TRUNCATED PSI(Y)
******
      ZERO=0.
      D=-C
      IF(Y.GT.D.AND.Y.LT.C)THEN
        TRPSI=Y
        RETURN
      ENDIF
      TRPSI=ZERO
      RETURN
      END

      REAL FUNCTION TAU(Y,C,BETA,S)
******
*         CALCULATES TAU(Y) WITH HUBER-CONSTANTS C AND BETA
******
      HALFB=.5*BETA
      D=-C
      IF(Y.LT.D.OR.Y.GT.C)GOTO 10
      TAU=(.5*Y*Y+HALFB)*S
      RETURN
   10 TAU=(C*ABS(Y)-.5*C*C+HALFB)*S
      RETURN
      END

      REAL FUNCTION PHI(C)
******
*         DENSITY FUNCTION OF STANDARD NORMAL
******
      REAL C
      COMMON/RLVAL/PI,SRT2PI
      PHI=EXP(-.5*C*C)/SRT2PI
      RETURN
      END

      REAL FUNCTION BIGPHI(X)
******
*         DISTRIBUTION FUNCTION OF STANDARD NORMAL
******
      BOUND=[illegible]
      BOUND1=-BOUND
      IF(X.LT.BOUND1)THEN
        BIGPHI=0.
        RETURN
      ENDIF
      IF(X.GT.BOUND)THEN
        BIGPHI=1.
        RETURN
      ENDIF
[illegible]
******
*         CALCULATES [illegible]
```

```
      REAL FUNCTION PHI(C)
******
*      DENSITY FUNCTION OF STANDARD NORMAL
******
      REAL C
      COMMON/PIVAL/ROOT2PI
      PHI=EXP(-.5*C*C)/ROOT2PI
      RETURN
      END

      REAL FUNCTION BIGPHI(X)
******
*      DISTRIBUTION FUNCTION OF STANDARD NORMAL
******
      BOUND=7.
      BOUND1=-BOUND
      IF(X.LT.BOUND1)THEN
        BIGPHI=0.
        RETURN
      ENDIF
      IF(X.GT.BOUND)THEN
        BIGPHI=1.0
        RETURN
      ENDIF
   10 CALL MDNOR(X,BIGPHI)
      RETURN
      END

      SUBROUTINE ESTI(X,N,T1,IQ,IREST,IFLAG,TABLET,CLASS1,CLASS2,
     _AVERAG,SD)
      DIMENSION X(100),Y(100),IOPT(5),STAT(5)
******
*      IF IREST=1 CALCULATES CORRECTED MEDIAN DEVIATION FROM T1
*      OTHERWISE                MEDIAN, CORR MAD, MEAN, S.D.
*      IF IFLAG=   PRINTS MEAN, S.D.
*                  CALCULATES CLASS INT
******
      DO 40 I=1,N
        Y(I)=X(I)
   40 CONTINUE
      DO 50 I=1,5
        IOPT(I)=1
   50 CONTINUE
      IOPT(2)=0
      IOPT(3)=0
      IF(IREST.EQ.1)GOTO 128
   60 CALL VSRTA(Y,N)
      CALL BEIUGR(Y,N,IOPT,STAT,IER)
      T1=STAT(4)
      SD=SQRT(STAT(3))
      SDMEAN=SD/SQRT(FLOAT(N))
      AVERAG=STAT(1)
      IF (IFLAG.EQ.1)THEN
CC
CC        WRITE(6,61)AVERAG,SD,SDMEAN
        CONST1=SDMEAN*TABLET
        CLASS1=AVERAG-CONST1
        CLASS2=AVERAG+CONST1
      ENDIF
  128 IOPT(1)=[illegible]
      IOPT(2)=[illegible]
  129 DO 130 I=1,N
        Y(I)=ABS(Y(I)-T1)
  130 CONTINUE
      CALL VSRTA(Y,N)
      CALL BEIUGR(Y,N,IOPT,STAT,IER)
      SD=STAT(4)/.6745
   61 FORMAT(/22X,"MEAN:",F9.3,12X,"S.D.:",F9.3,
     _14X,"S.D. OF MEAN:",F9.3)
      RETURN
      END

      SUBROUTINE GENER(X,N,DSEED,IDAVE,IDIST,CONTP,ITDF,
     CONTAM)
******
*      GENERATES DATA. IDIST=0: STANDARD NORMAL
*                            1: SLASH
*                            3: STUDENT T, DF=ITDF
*                            4: CHI-SQUARE, DF=ITDF
*                            5: CONTAMINATED NORMAL
*      READS IN DATA  IDIST=2
*      PRINTS SAMPLE NUMBER AND DATA
******
      DIMENSION X(100)
      DOUBLE PRECISION DSEED
      IFAIL=0
      IF(IDIST.EQ.3)THEN
        DO 55 I=1,N
          X(I)=G05DDF([illegible],DF,IFAIL)
    5   CONTINUE
        GOTO 120
      ENDIF
      IF(IDIST.EQ.4)THEN
        DO 65 I=1,N
```

```
*       GENERATES DATA.  IDIST=1: SLASH
*                              3: STUDENT T, DF=ITDF
*                              4: CHI-SQUARE, DF=ITDF
*                              5: CONTAMINATED NORMAL
*       READS IN DATA.  IDIST=2
*       PRINTS SAMPLE NUMBER AND DATA
******
      DIMENSION X(1  )
      DOUBLE PRECISION DSEED
      IFAIL=
      IF(IDIST.EQ.3)THEN
        DO 55 I=1,N
          X(I)=G05DDF(ITDF,IFAIL)
 55     CONTINUE
        GOTO 120
      ENDIF
      IF(IDIST.EQ.4)THEN
        DO 65 I=1,N
          X(I)=G05DHF(ITDF,IFAIL)
 65     CONTINUE
        GOTO 120
      ENDIF
      IF(IDIST.EQ.2)THEN
        READ(1,*)(X(I),I=1,N)
        GOTO 120
      ENDIF
      CALL GGNML(DSEED,N,X)
      IF(IDIST.EQ.1)THEN
        DO 100 I=1,N
          W=RANF()
          X(I)=X(I)/W
        CONTINUE
      ENDIF
      IF(IDIST.EQ.5)THEN
        XONE=1.
        XHALF=0.5
        TWLV=12.
        FUDGE=XONE+CONT/4/(XONE-CONT-4)
        SCONT=XONE-XHALF*CONT
        WRITE(8,2  )FUDGE
        DO 110 I=1,N
          W=RANF()
          WRITE(8,2  )W
          V=W-SCONT
          V=ABS(V)/V
          IW=IFIX(W*FUDGE)
          W=FLOAT(IW)
          W1=XONE-W
C         X(I)=X(I)*W1+W*V*TWLV
          X(I)=X(I)*W1+W*TWLV
 110    CONTINUE
      ENDIF
 120  IF(ISAVE.EQ.1)THEN
        WRITE(8,300)  ,(X(I),I=1,N)
      ENDIF
 300  FORMAT(1X,I X,  F  .4/(5X,  F1 .4))
 21   FORMAT((6X,1 F10.4)
C 21  FORMAT(/6X,  F1 .4)
 22   FORMAT(F5.2)
      RETURN
      END

      SUBROUTINE STUD(X,Y,T,N)
******
*       X IS INPUT DATA
*       Y IS OUTPUT
*       STORES PSI((X(I)-T)/S) IN Y(I)
******
      DIMENSION X(1  ),Y(1  )
      COMMON/CNS/  ,F,FI,  ,Z,  F
      DO   I=1,N
        Y(I)=PSI((X(I)-T)/S  )
      CONTINUE
      RETURN
      END

      SUBROUTINE KENTCI(STEPS,
     1LENGTH,XL  ,ICOUNT,ITERAT,  (NCIV,SL  GTH,
     1TLOW,TUP)
******
*       COMP INTERVAL UNDER KENT'S METHOD
******
      REAL LENGTH(1 ),LINT(2),SINT(2),SLNGTH(12)
      INTEGER ICOUNT(1 ),I,  IV(12)
      EXTERNAL CHITST,MISCOR
      COMMON/LATEST/T(  )
      COMMON/HUBEST/T,  ,S
      COMMON/TOLEFF/TOL1,TOL2,TOL3,TOL4,TOL5
      TWO=2. 
      ZERO= .
      DO 30  JJ=1,2
        XMULT=TWO*(FLOAT(JJ)-1.5)
        T =T1+XMULT*STEP
```

```
      X IS INPUT
*     Y IS OUTPUT
*     STORES PSI((X(I)-T)/S)/IN Y(I)
******
      DIMENSION X(100),Y(100)
      COMMON/CNB/C,N,F,F1,F2,Z,BETA,EFFN
      DO 20 I=1,N
        Y(I)=PSI((X(I)-T)/S,C)
   20 CONTINUE
      RETURN
      END

      SUBROUTINE KENTCI(STEPS,
     -LENGTH,XL C,ICOUNT,METHOD,ISCOV,CLENGTH,
     -TLOW,TUP)
******
*     CONF INTERVAL UNDER KENT'S METHOD
******
      REAL CLENGTH(12),TINT(2),SINT(2),LENGTH(12)
      INTEGER ICOUNT(12),ISCOV(12)
      EXTERNAL CHITST,K,SCOR
      COMMON/EXTEST/T,S
      COMMON/HLSEST/T1,S1
      COMMON/TOLER/TOL1,TOL2,TOL3,TOL4,ICL
      TWO=2.
      ZERO=0.
      DO 5 JJ=1,2
        XMULT=TWO*(FLOAT(JJ)-1.5)
        T2=T1+XMULT*STEPS
        S=SIN(W)
        TIMAX1=50
        IIMAX=0
        CALL ZYPLOT(TOL1,TOL3,TIMAX,ITMAX ,RESULT,S,SCORE?,
     -  IFLAG)
        IF(IFLAG.GT.0)THEN
          WRITE(6,810)
          STOP
        ENDIF
C       WRITE(6,820)T ,S,SCORE
        S1=S
        TNEW=T2-T1
        TIMAX1=50
        ITMAX=50
        CALL ZYROOT(TOL1,TOL3,TIMAX,ITMAX,SCHI,T,TNEW,CHIVL,
     -  IFLAG)
        IF(IFLAG.GT.0)THEN
          WRITE(6,830)
          STOP
        ENDIF
C       WRITE(6,820)TNEW,T,S
        TINT(JJ)=TNEW+T1
        SINT(JJ)=S
    5 CONTINUE
        TLOW=TINT(1)
        TUP=TINT(2)
        TLNGTH=TUP-TLOW
        LENGTH(METHOD)=TLNGTH
        CLNGTH(METHOD)=CLNGTH(METHOD)+TLNGTH
        CALL COVER(TLOW,TUP,METHOD,XLOC,ICOUNT,
     -  ISCOV)
        C1=T1-TLOW
        C2=TUP-T1
CC
CC      WRITE(6,49 )METHOD,TLOW,TUP,SINT(1),
CC   -  SINT(2),C1,C2,TLNGTH
  49 -  FORMAT(/1X,"METHOD",1/,1X,"CONF INTERVAL:",F9.3," ,",
     -  F9.3/1X,"SCALE EST AT EXT:",F9.3,
     -  " ,",F9.3/5X,"LENGTH OF INTERVAL:",F9.3," ,",F9.3,
     -  5X,"TOTAL LENGTH:",F9.3)
  500 CONTINUE
  810 FORMAT(1X," IN KENTCI, TNEW,T,S ",3F12.6)
  820 FORMAT("  IN ZYROOT FROM KENTCI, T,S,SCORE?:", 3F11.5)
  830 FORMAT("  I IS LARGE IN KENTCI")
  840 FORMAT(" I IS LARGE IN KENTCI")
      RETURN
      END

      SUBROUTINE CLASS(TLOW,TUP,LNGTH,K,XLOC,ICOUNT,
     -ISCOV,CLNGTH)
******
*     PRINTS CLASSICAL CONFIDENCE INTERVAL
******
      REAL CLNGTH(12),LNGTH(12)
      INTEGER ICOUNT(12),ISCOV(12)
      RLNGTH=TUP-TLOW
      CLNGTH(K)=CLNGTH(K)+RLNGTH
      LNGTH(K)=RLNGTH
CC
CC    WRITE(6,2 )K,TLOW,TUP,RLNGTH
   2  FORMAT(/1X,"METHOD",I3/" CLASSICAL CONF INTERVAL:",F9.3," ,",
     -F9.3,1X,"LENGTH:",F9.3)
      CALL COVER(TLOW,TUP,K,XLOC,ICOUNT,
     -ISCOV)
      RETURN
```

```
3    FORMAT(" ... LENGTH ...")
5    FORMAT(" ... LENGTH ... KENT...")
     RETURN
     END

     SUBROUTINE CL  (TLENS,TOP,CLENGTH,K,XLOSS,ICOUNT,
    .LOSS,FLNGTH)
******
*       MULTI-STAGE SAMPLING ...
******
     CALL ...LENGTH(...)
     ...COUNT(...)
     FLENGTH=TOP-TLEN
     CLENGTH(K)=CLENGTH(K)+FLENGTH
     FLENGTH(K)=FLENGTH
C
     WRITE(6,...)K,TLENS,TOP,FLENGTH
     FORMAT(/...X,"... TIME ...",7" ... INTERVAL:",F9.2," ...",
    .F9.2,1X,"LENGTH:",F9.3)
     CALL ...(TLENS,TOP,K,XLOSS,ICOUNT,...
    .LOSV)
     RETURN
     END

     SUBROUTINE ...(...,DELT,...,BETA,...)
******
*    ... DEFINED IN CHAPTER 4
******
     XLOSE=...
     TW=2.
     ...=0
     ...=...PHI(...)
     BETA=XLOSE+TW*(...)
     DELTA=XL...-TW*...
     ...=...
     ZETA=...*BETA-...
     ...
     WRITE(6,...)DELTA,BETA,...
     FORMAT(...X,"DELTA:",F9.3,"  BETA:",F9.3,"  ...",F5.3,
    ." ZETA:",F9.3,"  ...",F9.3)
     RETURN
     END

     SUBROUTINE PRECISE(...,XLOSS,...
    .T,DELTA,...)
******
*    PRECISE ... TUBE: ...
*                        SAMPLE SIZE
*                        ... SAMPLES
*                        EFFECTIVE SAMPLE SIZE
*                        ... LEVEL
*                        ... /SCALE
*                        FLAG FOR ...
*                 ...
*                        VALUE FROM ... TABLE
*                        VALUE FROM ... TABLE
*                 PRINT ...
*    ...
******
     DOUBLE PRECISION ...
     ...
     WRITE(6,102)
     CALL ...(...)
       ...=...(...)/...
       CALL ...(...)
     READ...
     READ...
     READ...
     READ...,LEVEL
     ...
     FI=FLOAT(...)
     ...
     ...LEVEL=...LEVEL+...
     CALL ...(TABLE,...,TABLE...)
     CALL ...(...LEVEL,TABLE...)
     READ...LEVEL
     READ...,...ST
     IF(...)THEN
       XLOSS=...
       WRITE (6,13...)
       WRITE(1...)
       GOTO ...
     ENDIF
     IF(...)THEN
       XLOSS=...
       READ ...COUNT...
       WRITE(6,...)COUNT1
       GOTO ...
     ENDIF
     IF(...)THEN
       IF(...FLAG...)GOTO ...
       ...XLOSS
       WRITE(6,...)...
       WRITE(1...)...
```

```
      CALL S[illegible] C[illegible](ISE[illegible])
      READ4,[illegible]
      READ7,N,M
      READ4,EFFN
      READ4,CLEVEL
      I1=N-1
      FI=FLOAT(I1)
      TAIL=1.-CLEVEL
      CLEVEL=CLEVEL+TAIL/2.
      CALL MDSTAT(TAIL,FI,TABLET,I[illegible])
      CALL MDNOR(CLEVEL,TABLEN,I[illegible])
      READ12,IDEVE
      READ1[illegible],I[illegible]ST
   1[illegible] IF(IDIST.EQ.1)THEN
        XLOC=[illegible].
        WRITE(8,130)
        WRITE(1[illegible],130)
        GOTO [illegible]
      ENDIF
      IF(IDIST.EQ.2)THEN
        XLOC=[illegible].
        READ 4,CONTA[illegible]
        WRITE(8,[illegible])CONTA[illegible]
        GOTO 30
      ENDIF
      IF(IDIST.EQ.3)THEN
        IF(IFLAG.EQ.1)GOTO 3[illegible]
        READ4,XLOC
   [illegible]9   WRITE(8,[illegible])XLOC
        WRITE(1[illegible],17[illegible])XLOC
        GOTO [illegible]
      ENDIF
      IF(IDIST.EQ.[illegible])THEN
        IF(IFLAG.EQ.1)GOTO [illegible]
        READ1[illegible],ITDF
        XLOC=[illegible].
   [illegible]    WRITE(8,[illegible])ITDF
        WRITE(1[illegible],[illegible])ITDF
        GOTO [illegible]
      ENDIF
      IF(IDIST.EQ.4)THEN
        IF(IFLAG.EQ.1)GOT[illegible]
        READ1[illegible],ITDF
        XLOC=FLOAT(ITDF)
   [illegible]    WRITE(8,[illegible])ITDF
        WRITE(1[illegible],[illegible])ITDF
        GOTO [illegible]
      ENDIF
      IF(IFLAG.EQ.1)GOTO [illegible]
      XLOC=[illegible].
      [illegible]=[illegible].
   [illegible]7 WRITE(8,[illegible])XLOC,[illegible]
      WRITE([illegible],[illegible])XLOC,[illegible]
   [illegible] WRITE(8,3[illegible]),[illegible]EFFN,[illegible],CLEVEL,TABLET,TABLE[illegible]
      WRITE([illegible],3[illegible]),[illegible]EFFN,[illegible],CLEVEL,TABLET,TABLE[illegible]
    4 FORMAT(F1[illegible].[illegible])
    7 FORMAT(20I[illegible])
   [illegible] FORMAT([illegible]X," DATA: NORMAL  MEAN=",[illegible]/[illegible]X,
     *[illegible]",F1[illegible].[illegible])
   [illegible] FORMAT(2[illegible]X," SAMPLE SIZE:",I[illegible]/
     [illegible]X,"[illegible] FOR [illegible] SAMPLE SIZE:",F[illegible].1/
     [illegible]7X," NUMBER OF SAMPLES:",I4/
     [illegible]X," HUBER'S C:",F[illegible]/
     [illegible]X," CONFIDENCE LEVEL:",F[illegible].2/
     [illegible]X," VALUE FROM T-TABLE:",F[illegible]/1[illegible]X,
     *"VALUE FROM NORMAL TABLE:",F[illegible].4)
   [illegible] FORMAT(I[illegible])
  15[illegible] FORMAT(F8.4)
  1[illegible]    FORMAT([illegible]X," DATA: SLASH")
  1[illegible] FORMAT(/1X,[illegible]("*")//)
  [illegible]7 FORMAT([illegible]X,"DATA: SUPPLIED,LOCATION=",F1[illegible].[illegible])
  [illegible] FORMAT(3X," DATA: STUDENT T, DF=",I[illegible])
  [illegible]1 FORMAT(3X,"DATA: CHI-SQUARE, DF=",I[illegible])
  [illegible] FORMAT(" STANDARD NORMAL DATA WITH",
     *" ASYMMETRIC CONTAMINATION",F4.[illegible])
      RETURN
      END

      SUBROUTINE ALNTH([illegible]NGTH,K,M)
      REAL LENGTH(1[illegible])
      DO 1[illegible] I=1,K
        XLNGTH([illegible])=XLENGTH(I)/FLOAT(M)
        WRITE(8,1[illegible])I,XLNGTH(I)
  1[illegible]    FORMAT(3[illegible]X,"LTH(D",I[illegible],":",F9.[illegible])
  1[illegible] CONTINUE
      RETURN
      END

      REAL FUNCTION CHISI(T)
******
*        CHI-SQUARE TEST USED [illegible]
******
      REAL X([illegible]),[illegible]([illegible]),[illegible],[illegible],[illegible]ST
      INTEGER [illegible]
```

```
      DO 150 I=1,K
         XLENGTH(I)=XLENGTH(I)/FLOAT(M)
         WRITE(6,150)I,XLENGTH(I)
150      FORMAT(34X,"METHOD",I3,":",F9.3)
15+   CONTINUE
      RETURN
      END

      REAL FUNCTION CHISST(T)
******
*        CHI-SQUARE TEST USED IN KENT'S METHOD
******
      REAL X(10 ),SCORE(7),BETA,C,S,T,W ,WIFST
      INTEGER METHOD
      EXTERNAL RESCOR
      COMMON/SAMP/X
      COMMON/CHD/C,N,F,F1,F2,Z,BETA,EFF
      COMMON/TOLER/TOL1,TOL2,TOL3,TOL4,TOL5
      COMMON/EXTEST/IN,S
      COMMON/HUBEST/T1,S1
      COMMON/TZOVAL/WTEST,FSTIU
      COMMON/CONLVL/CLEVEL
      I=0
      T =T+T1
      S=S
      EE=1.
      MAX1=50
      MAX2=50
      CALL CHIFT(TOL1,TOL2,ITMAX1,ITMAX ,RESCOR,S,SCORE2)
      S =S
C     WRITE(6,3 )T ,S
      W =0.
      DO 3  I=1,N
         W=(X(I)-T )/S
         W =W +T(W,C,BETA,  )
 3    CONTINUE
      W =1.0*(W -WIFST)/EFF
 2    CALL MDCH(W,EF,P,IER )
      CHITST=CLEVEL-P
C     WRITE(6,2 )CHITST
 39   FORMAT(" CHITST:",F8.3)
 3    FORMAT("  ,  CHI  CHITST:",
     *F9.3)
      RETURN
      END

      SUBROUTINE COVER(TLOW,TOP,I,XLOC,ICOUNT,
     *ICOV)
******
*        IF INTERVAL COVERS TRUE LOCATION, ADD 1 TO COUNTER
*                                       STORE 1 IN ISCOV
******
      INTEGER ICOUNT(10),ISCOV(10)
      IF(TLOW.LT.XLOC)THEN
         IF(XLOC.LT.TOP)THEN
            ICOUNT(I)=ICOUNT(I)+1
            ISCOV(I)=1
            RETURN
         ENDIF
      ENDIF
      ISCOV(I)=
      RETURN
      END

      SUBROUTINE COVPRP(ICOUNT,K,M)
******
*        CALCULATE AND PRINT COVERING PROPORTIONS
******
      INTEGER ICOUNT(10)
      XM=FLOAT(M)
      WRITE(6,4 )
      DO 10 I=1,K
         COUNT=FLOAT(ICOUNT(I))/XM
         WRITE(6,2 )I,COUNT
 2       FORMAT(34X,"METHOD",I3,":",F9.3)
10    CONTINUE
 4    FORMAT(1 X,"COVERING PROPORTIONS")
      RETURN
      END

      REAL FUNCTION RESCOR(S)
******
*        CALCULATES SECOND SCORE(T,S)
******
      REAL X(10 ),C,F,BETA
      COMMON/CHD/C,F,F1,F2,Z,BETA,EFF
      COMMON/SAMP/X
      COMMON/EXTEST/T,S
      SCORE=0.
      DO 10 I=1,N
         W=PSI((X(I)-T)/S,C)
         SCORE=SCORE+W*W
10    CONTINUE
      RESCOR=SCORE/
```

```
******
*      CALCULATES AND PRINTS COVERING PROPORTIONS
******
      INTEGER ICOUNT(12)
      XM=FLOAT(M)
      WRITE(6,40)
      DO 10 I=1,K
        COUNT=FLOAT(ICOUNT(I))/XM
        WRITE(6,20)I,COUNT
        FORMAT(34X,"METHOD",I2,":",F9.3)
  10  CONTINUE
  40  FORMAT(  X,"COVERING PROPORTIONS")
      RETURN
      END
      REAL FUNCTION RESCOR(S)
******
*       CALCULATES SECOND SCORE(T0,S)
******
      REAL X(1  ),C,F,BETA
      COMMON/CNS/C,F,FA,FB,Z,BETA,EFF
      COMMON/DATA/X
      COMMON/EXPEST/T0,D
      SCORE=0.
      DO 10 I=1,N
        W=PSI((X(I)-T0)/S,C)
        SCORE=SCORE+W*W
  10  CONTINUE
      RESCOR=SCORE/F-BETA
      RETURN
      END
      SUBROUTINE GSEED(DSEED)
******
*       USES CLOCK TO GET SEEDS FOR RANDOM NUMBER GENERATION
******
      DOUBLE PRECISION DSEED
      INTEGER ISEC,IHOUR,IMIN
      CHARACTER*10 XCLOCK,CLOCK
      XCLOCK=CLOCK()
      FIFTY=50.
      READ(XCLOCK,'(1X,3(I2,1X))')IHOUR,IMIN,ISEC
      ISEC=ISEC+1
      XLIMIT1=2.**46
      XLIMIT2=2147483646
      ANUMB=ALOG(FLOAT(IMIN*60+ISEC))
      NSEED=IFIX((ANUMB-IFIX(ANUMB))*XLIMIT2)+1
      CALL RANSET(NSEED)
        W=RANF()
      DSEED=W*XLIMIT1
      RETURN
      END
      SUBROUTINE MYROOT(TOLFN,TOLVAR,ITMAX1,ITMAX2,FN,VAR
     *VALUE,IFLAG)
******
*       SOLVES FN(VAR)=0
*       FN(ZERO) IS ASSUMED POSITIVE
*       ROUTINE NEVER EVALUATES FN(ZERO)
*       SIGN OF VAR IS ASSUMED DIRECTION OF DECREASING FN(VAR)
*              IFLAG IS SET TO -1 IF THIS ASSUMPTION IS SUSPECT,
*                 RETURN TO CALLING PROGRAM WITHOUT FINDING ROOT
*       SUBROUTINE FIRST CHECKS FOR VAR SUCH THAT FN(VAR) IS POSITIVE
*       10-LOOP ALLOWS BACK-UP IF INITIAL GUESS GIVES NEGATIVE FN(VAR)
*       20-LOOP MOVES AWAY FROM ZERO LOOKING FOR INTERVAL CONTAINING R
*       THEN CALL NAGLIB ROUTINE C05ADF
*       IFLAG IS SET TO 1 IF ABS(VAR) IS LARGE, AND RETURN TO
*                      CALLING PROGRAM WITHOUT FINDING ROOT
******
      REAL VAREND(3),VAL(3)
      EXTERNAL FN
      IFLAG=0
      ZERO=0.
      TWO=2.
      VALUE=FN(VAR)
      DIFF=ABS(VALUE)
      IF(DIFF.LT.TOLFN)THEN
        RETURN
      ENDIF
      IF(VALUE.LT.ZERO)THEN
        DO 10 I=1,ITMAX
          VAR=VAR/TWO
          VALUE=FN(VAR)
          DIFF=ABS(VALUE)
          IF(DIFF.LT.TOLFN)RETURN
          IF(VALUE.GT.ZERO)THEN
            VAREND(1)=VAR
            VAREND(2)=VAR*TWO
            GOTO 30
          ENDIF
  10    CONTINUE
        WRITE(6,210)
        STOP
******
```

```
      REAL VAREND(3),VAL(3)
      EXTERNAL FN
      IFLAG=0
      ZERO=0.
      TWO=2.0
      VALUE=FN(VAR)
      DIFF=ABS(VALUE)
      IF(DIFF.LT.TOLFN)THEN
        RETURN
      ENDIF
      IF(VALUE.LT.ZERO)THEN
        DO 10 I=1,ITMAX
          VAR=VAR/TWO
          VALUE=FN(VAR)
          DIFF=ABS(VALUE)
          IF(DIFF.LT.TOLFN)RETURN
          IF(VALUE.GT.ZERO)THEN
            VAREND(1)=VAR
            VAREND(2)=VAR*TWO
            GOTO 50
          ENDIF
   10   CONTINUE
        WRITE(6,210)
        STOP
C
******
*         IF THERE IS NO JUMP OUT OF 10-LOOP, THEN SOLUTION TO
*         FN(VAR) [illegible] SMALLER [illegible] IS SUSPECT
*         SET IFLAG=-1 AND RETURN TO CALLING PROGRAM
*         NO ROOT IS FOUND
******
        IFLAG=-1
        RETURN
      ENDIF
   15 VAREND(1)=VAR
      VAL(1)=VALUE
      STEP=VAR
      DO 20 I=1,ITMAX1
        VAREND(2)=VAREND(1)+STEP
        VAL(2)=FN(VAREND(2))
C       WRITE(6,[illegible])VAREND(1),VAL(1),VAREND(2),VAL(2)
        PROD=VAL(1)*VAL(2)
        IF(PROD.LT.ZERO)GOTO 35
        DIFF=ABS(VAL(2))
        IF(DIFF.LT.TOLFN)THEN
          VAR=VAREND(2)
          VALUE=VAL(2)
          RETURN
        ENDIF
        VAREND(1)=VAREND(2)
        VAL(1)=VAL(2)
   20 CONTINUE
      WRITE(6,221)
C
******
*         IF THERE IS NO JUMP OUT OF 20-LOOP THEN SOLUTION TO
*         FN(VAR)=0 IS LARGE IN ABSOLUTE VALUE. SET IFLAG=1
*         AND RETURN TO CALLING PROGRAM
******
      IFLAG=1
      RETURN
   35 IF(VAR.LT.ZERO)THEN
        ENDL=VAREND(2)
        ENDR=VAREND(1)
        GOTO 40
      ENDIF
      ENDL=VAREND(1)
      ENDR=VAREND(2)
   40 VAR=(ENDL+ENDR)/TWO
      IFAIL=0
      CALL [illegible](ENDL,ENDR,TOLVAR,TOLFN,F [illegible],VAR,IFAIL)
      VALUE=FN(VAR)
      IF(IFAIL.NE.0)THEN
        WRITE(6,220)IFAIL,VALUE
  300 FORMAT(" [illegible]",[illegible])
        STOP
      ENDIF
  200 FORMAT(" TOO MANY ITERATIONS IN [illegible] 10-LOOP")
  210 FORMAT(" INITIAL ESTIMATE BAD IN [illegible]")
  220 FORMAT(" ERROR IN [illegible]. IFAIL, VALUE",I4,F9.5)
      RETURN
      END

      SUBROUTINE [illegible](TOLFN,TOLVAR,ITMAX1,ITMAX2,FN,VAR,VALUE)
******
*         WORKS JUST LIKE [illegible], EXCEPT THAT
*         [illegible]-LOOP USES BISECTION TO REPLACE CALL TO NAGLIB
*         THIS ROUTINE IS CALLED FROM [illegible] FROM [illegible]
******
      REAL VAREND(2),VAL(3)
      EXTERNAL FN
      ZERO=0.
      TWO=2.
      VALUE=FN(VAR)
      DIFF=ABS(VALUE)
```

```
      [illegible]
          STOP
        ENDIF
  200 FORMAT(" TOO MANY ITERATIONS IN MYROOT 200-LOOP")
  210 FORMAT(" INITIAL ESTIMATE BAD IN MYROOT")
  220 FORMAT(" ERROR IN MYROOT, FINAL VALUE",I4,F9.5)
      RETURN
      END
      SUBROUTINE CHIRT(TOLF,TTLEVAR,ITMAX1,ITMAX2,FN,VAR,VALUE)
******
*          WORKS JUST LIKE MYROOT, EXCEPT THAT
*          300-LOOP USES BISECTION TO REPLACE CALL TO MASLIB
*          THIS ROUTINE IS CALLED FROM MYROOT FROM KRITCH
*******
      REAL VAREND(3),VAL(3)
      EXTERNAL FN
      ZERO=0.
      TWO=2.
      VALUE=FN(VAR)
      DIFF=ABS(VALUE)
      IF(DIFF.LT.TOLF)THEN
        RETURN
      ENDIF
      IF(VALUE.LT.ZERO)THEN
        DO 10 I=1,ITMAX
          VAR=VAR/TWO
          VALUE=FN(VAR)
          DIFF=ABS(VALUE)
          IF(DIFF.LT.TOLF)RETURN
          IF(VALUE.GT.ZERO)THEN
            VAREND(1)=VAR
            VAREND(2)=VAR*TWO
            GOTO 12
          ENDIF
   10   CONTINUE
        WRITE(6,210)
        STOP
      ENDIF
   12 VAREND(1)=VAR
      VAL(1)=VALUE
      STEP=VAR
      DO 20 I=1,ITMAX1
        VAREND(2)=VAREND(1)+STEP
        VAL(2)=FN(VAREND(2))
        PROD=VAL(1)*VAL(2)
        IF(PROD.LT.ZERO)GOTO 5
        DIFF=ABS(VAL(2))
        IF(DIFF.LT.TOLF)THEN
          VAR=VAREND(2)
          VALUE=VAL(2)
          RETURN
        ENDIF
        VAREND(1)=VAREND(2)
        VAL(1)=VAL(2)
   20 CONTINUE
      WRITE(6,220)
      STOP
    5 DO 100 II=1,ITMAX2
        VAREND(3)=(VAREND(1)+VAREND(2))/TWO
        VAL(3)=FN(VAREND(3))
        DIFF=ABS(VAREND(1)-VAREND(2))
        IF(DIFF.LT.TLEVAR)GOTO 201
        DIFF=ABS(VAL(3))
        IF(DIFF.LT.TOLF)GOTO 201
        IF(VAL(3).LT.ZERO)THEN
          VAREND(2)=VAREND(3)
          VAL(2)=VAL(3)
          GOTO 100
        ENDIF
        VAL(1)=VAL(3)
        VAREND(1)=VAREND(3)
  100 CONTINUE
      WRITE(6,200)
      STOP
  201 VAR=VAREND(3)
      VALUE=VAL(3)
  200 FORMAT(" TOO MANY ITERATIONS IN CHIRT 100-LOOP")
  210 FORMAT(" INITIAL ESTIMATE BAD IN CHIRT")
  220 FORMAT(" TOO MANY ITERATIONS IN CHIRT 200-LOOP")
      RETURN
      END
      SUBROUTINE FUN(ITHETA,THETA,SCORE,IFLAG)
******
*       2 DIMENSIONAL SCORE FUNCTION FOR LOCATION, SCALE
******
      INTEGER IFLAG,ITHETA
      REAL THETA(2),SCORE(2),X(...)
      COMMON/SAMP/X
      COMMON/CNS/[illegible]
      ZERO=0.
      THETA1=THETA(1)
      THETA2=THETA(2)
```

[redacted]

```
      [illegible]
      V[illegible]L(I)=V[illegible]
      VARE[illegible](I)=VARE[illegible](3)
1[illegible] CONTINUE
      WRITE(6,2[illegible])
      STOP
[illegible] V[illegible]=VAR[illegible](3)
      V[illegible]LO[illegible]=V[illegible]L(3)
[illegible] FORMAT(" [illegible] MANY ITERATIONS IN CHI[illegible]T [illegible]-LOOP")
[illegible] FORMAT(" INITIAL ESTIMATE BAD IN CHIST")
[illegible] FORMAT(" TOO MANY ITERATIONS I[illegible] CHART [illegible]-LOOP")
      RETURN
      END

      SUBROUTINE [illegible]FUN(ITHETA,THETA,SCORE,IFLAG)
******
*       [illegible] SCORE FUNCTION FOR LOCATION, SCALE
******
      INTEGER IFLAG,ITHETA,[illegible]
      REAL THETA(2),SCORE(2),X(1[illegible])
      COMMON/[illegible]P/X
      COMMON/[illegible]/[illegible],F,[illegible],Z,BETA,[illegible]F[illegible]
      ZERO=[illegible].
      THETA[illegible]=THETA(1)
      THETA[illegible]=THETA(2)
      SCORE1=ZERO
      SCORE[illegible]=ZERO
      DO [illegible] I=1,N
        [illegible]=P[illegible](X(I)-THETA1)/THETA2,C)
        SCORE1=SCORE1+[illegible]
        SCORE[illegible]=SCORE[illegible]+[illegible]*[illegible]
[illegible] CONTINUE
      SCORE(1)=SCORE1/F
      SCORE(2)=SCORE[illegible]/F-[illegible]ET[illegible]
      RETURN
      END

      SUBROUTINE MYTEST(A,DIR,FLAG)
******
*       [illegible] ENTRY [illegible] STORES ESTIMATE OF [illegible]K[illegible](1)=SHIFT
*         [illegible] EXIT [illegible] IMPROVED FOR [illegible] UNDER TEST
*       ON EXIT DIR=1 IF TEST ACCEPTS [illegible] AT LEVEL CORRESP TO TABLE1
*                DIR=-1 IF TEST REJECTS [illegible] AT LEVEL CORRESP TO TABLE[illegible]
*                FLAG=1 IF [illegible] IS AN END POINT OF THE CONFIDENCE INTERVAL
*                FLAG=-1 OTHERWISE
******
      REAL Y(1[illegible]),[illegible](4)
      INTEGER [illegible] DY
      EXTERNAL [illegible]PT
      COMMON/[illegible]IVAL/P[illegible],TOPI
      COMMON/SHIFTY/SHIFT,XFACT
      COMMON/[illegible]/[illegible],F,F1,F[illegible],Z,BETA,[illegible]F[illegible]
      COMMON/CONLVL/CLEVEL
      COMMON/[illegible]/[illegible]
      COMMON/TOLER/T[illegible],T[illegible],TOL3,TOL4,[illegible]
      COMMON/CO[illegible]F[illegible]/Y,[illegible]DY
      [illegible]=CON[illegible]
      T[illegible]=[illegible]T[illegible]
      X[illegible]=1.
      T[illegible]=[illegible].
      F[illegible]=[illegible].
      ZERO=[illegible].
      FLAG=-1.
      [illegible]
******
*       [illegible]T [illegible]EVE[illegible] K[illegible](4)=SHIFT [illegible]
*         Y(1) < SHIFT < Y(4) [illegible]
******
      [illegible]=Y(1)+[illegible]
      [illegible]=Y([illegible])-[illegible]
      IF([illegible].GT.[illegible].OR.[illegible].GT.[illegible]G)THEN
        [illegible]
        WRITE(6,[illegible])[illegible]
        RETURN
      [illegible]
      [illegible]=1
      [illegible]=[illegible]
      [illegible] ZERO[illegible]T([illegible],[illegible]3,[illegible],X[illegible],[illegible]VALUE,
     [illegible])
      IF(IFL[illegible].[illegible])TH[illegible]
        DIR=-1
        RETURN
      [illegible]
      IF([illegible])[illegible]
      IF([illegible](-[illegible])[illegible]
******
*       IF [illegible] SMALL [illegible]
******
      [illegible]
        WRITE(6,[illegible])[illegible]
        DIR=1
        RETURN
      ENDIF
      [illegible]
```

```
******
      CG1=Y(1)+TOL2
      CG2=Y(N)-TOL2
      IF(SHIFT.LT.CG1.OR.SHIFT.GT.CG2)THEN
         DIR=-1.
C        WRITE(1,993)NSAMP
         RETURN
      ENDIF
      [illegible] X1=1.
      [illegible] X2=2.
      CALL [illegible]T(T[illegible],TOL3,[illegible]TMAX,[illegible],COEFFI,N,VALUE,
     +IFLAG)
      IF(IFLAG.GE.[illegible])THEN
         DIR=-1
         RETURN
      ENDIF
      IF(IFLAG.[illegible])[illegible]
C     IF(IFAIL.NE.[illegible])WRITE(1,995)[illegible],IFAIL,[illegible]
******
*        IF [illegible] IS SMALL [illegible]
******
      IF(ABS(F).LT.TOL1)THEN
C        WRITE(1,993)METHOD,T[illegible],NSAMP
         DIR=1
         RETURN
      ENDIF
      [illegible]=SQRT(COEF+[illegible])
      F[illegible]=COF(A)*[illegible]
******
*        F 1 MUST BE POSITIVE
******
      IF(FO1.LT.TOL2)THEN
         WRITE(8,999)F,F[illegible],METHOD
         STOP
      ENDIF
      CUT1=(XLIM-CL[illegible]VEL)/[illegible]
      CUT2=XMAX-CUT1
      R[illegible]D=SQRT([illegible]*FO1)*XFACT
******
*        [illegible] TAIL [illegible] APPROXIMATION
******
C     W=[illegible]GPHI(-RAD)-EXP(-FO1)/RT2PI*(XLIM/[illegible]/CON[illegible]+X[illegible]/RAD)
      W=[illegible]GPHI(-RAD)-EXP(-F01)/RT2PI*(([illegible])+A*[illegible])/(A
     +*[illegible]+CON[illegible]))
C     WRITE(1,995)W,NSAMP
******
*        SET FLAG AND DIR
******
      DECIDE1=ABS(W-CUT1)
      DECIDE2=ABS(W-CUT2)
C
C     WRITE(8,[illegible])METHOD,CUT1,W,CUT2
      IF(MIN(DECIDE1,DECIDE2).LT.TOL4)THEN
         FLAG=1.
         RETURN
      ENDIF
      IF(W.LT.CUT2)THEN
         IF(W.GT.CUT1)THEN
            DIR=1.
            RETURN
         ENDIF
      ENDIF
      DIR=-1.
      RETURN
 53   FORMAT([illegible])
 [illegible] FORMAT(' *** [illegible]LUTION [illegible] SHIFT [illegible]PLE',I4)
 9[illegible] FORMAT(' [illegible] ZERO METHOD, [illegible]',F9.3,' SAMPLE',[illegible])
 9[illegible] FORMAT(' [illegible]SAMPLE[illegible]',F9.3,I5)
 95[illegible] FORMAT(' [illegible] IFAIL',F9.3,[illegible])
 999 FORMAT(' STOP. [illegible]METHOD',F9.3,[illegible])
      END

      SUBROUTINE [illegible]( [illegible]
     +[illegible]COUNT,XL[illegible]T,S[illegible]FROLI,
     +[illegible])
******
*        [illegible] IS USED T[illegible]
*        [illegible] THEN [illegible] FOR [illegible]
*        [illegible] SHIFT [illegible] LESS THAN OR EQUAL TO 1.
*                                    GREATER THAN IT EQUAL [illegible]
******
      REAL AVAL(2),[illegible]
      INTEGER [illegible]COUNT([illegible])
      COMMON/[illegible]/[illegible]
      COMMON/[illegible]SHIFT,XFACT
      [illegible]=1.
      TWO=2.
      [illegible]=[illegible]/[illegible]
      [illegible]=[illegible]
******
*        [illegible]
*          [illegible]
*          [illegible]ACC[illegible]
```

[redacted] [redacted]

```
      SUBROUTINE ROBCI(STEPS,LENGTH,
     1ICOUNT,XLJC,TI,SIZE,METHOD,IFROST,
     2ACCDV,LENGTH,SHCOR,TLOW,TUP)
******
*        SIZE IS USED TO SCALE DATA.  FIRST FIND DATA FOR T VIA THEN
*              THEN DATA FOR T*SIZE
*        SHOULD BE SHIFT CORRECTION, LESS THAN OR EQUAL TO 1.0
*                                    GREATER THAN OR EQUAL TO
******
      REAL AVAL(2),CI(2),LENGTH(12),LENGTH(12)
      INTEGER ICOUNT(12),ISCDV(12)
      COMMON/ZCP1/ZCP
      COMMON/SHIFTY/SHIFT,XFACT
      ZERO=0.
      TWO=2.
      STEP2=STEPS/4.
      DO 300 JJ=1,2
******
*        JJ=1 LOOKING FOR LEFT END-POINT OF INTERVAL
*              XMULT<0
*              SIGN(ACCNJ)<0
*        JJ=2 LOOKING FOR RIGHT END-POINT OF INTERVAL
*              XMULT>0
*              SIGN(ACCNJ)>0
******
         STEP=STEPS
         XMULT=TWO*(FLOAT(JJ)-1.5)
         ACCNJ=XMULT
******
*        XFACT=-SIGN(ACCNJ)
******
         XFACT=-XMULT
         TNEW=TI+XMULT*STEP
         DO 30  I=1,?
******
*        LOOP LOOKS FOR POINT OUTSIDE INTERVAL
******
C           SHIFT=TNEW-TI-SHCOR*XMULT
            SHIFT=(TNEW-TI)*SHCOR
C
C           WRITE(6,890)TNEW,SHIFT
            CALL MYTEST(ACCNJ,DIR,FLAG)
            IF(FLAG.GT.ZERO)GOTO 30
            IF(DIR.LT.ZERO)GOTO 30
            TNEW=TNEW+XMULT*STEP
            STEP2=STEP
            ACCNJ=XMULT
   30    CONTINUE
         WRITE(6,800)NSAMP
         STOP
   3     STEP2=STEP2/TWO
         DO 40  I=1,20
******
*        THIS LOOP CLOSES IN ON ENDPOINT
******
            ACCNJ=XMULT
            DIR=DIR*XMULT
            TNEW=TNEW+DIR*STEP2
C           SHIFT=TNEW-TI-SHCOR*XMULT
            SHIFT=(TNEW-TI)*SHCOR
C
C           WRITE(6,890)TNEW,SHIFT
            CALL MYTEST(ACCNJ,DIR,FLAG)
            IF(FLAG.GT.ZERO)GOTO 40
            STEP2=STEP2/TWO
   40    CONTINUE
         IFROST=IFROST+1
         WRITE(6,801)NSAMP,JJ,METHOD
C        STOP
   4     AVAL(JJ)=ACCNJ
         CI(JJ)=TNEW*SIZE
  300 CONTINUE
      TLOW=CI(1)
      TUP=CI(2)
      LENGTH=TUP-TLOW
      LENGTH(METHOD)=LENGTH
      LENGTH(METHOD)=LENGTH(METHOD)+LENGTH
      CALL COVER(TLOW,TUP,METHOD,XLJC,ICOUNT,
     1ICCDV)
      C1=TI*SIZE-TLOW
      C2=TUP-TI*SIZE
C
CC
CC       WRITE(6,900)METHOD,TLOW,TUP,C1,C2,LENGTH
CC      1AVAL(1),AVAL(2)
CC  900 FORMAT(1X,I2,2X,'METHOD',1X,'/',2X,'CONF INTERVAL=',F9.3,' ,',
C     1F9.3,/2X,'LENGTH',' IS ',F9.3,' ,',F9.3,2X,'T TAIL LENGTH=',
C     2F9.3/7X,'ACCNJ AT EXTREMES=',F9.3,' ,',F9.3)
  800 FORMAT(' TROUBLE IN ROBCI 30-LOOP ',I5,'I=',I5)
  801 FORMAT(' ROBCI ',I5,'-LOOP,',
     1' ',2X,'METHOD',I4)
```

```
      IFROUT=IFROUT+1
      WRITE(6,801) IFROUT,JJ,METHOD
      STOP
  497 AVAL(JJ)=ACONJ
      CI(JJ)=TNEW*SIZE
  500 CONTINUE
      TLOW=CI(1)
      TUP=CI(2)
      TLNGTH=TUP-TLOW
      LNGTH(METHOD)=TLNGTH
      CLNGTH(METHOD)=CLNGTH(METHOD)+TLNGTH
      CALL COVER(TLOW,TUP,METHOD,XLOC,ICOUNT,
     1 ICOV)
      C1=T1*SIZE-TLOW
      C2=TUP-T1*SIZE
CC
CC
CC    WRITE(6,49) METHOD,TLOW,TUP,C1,C2,TLNGTH
CC   1,AVAL(1),AVAL(2)
   49 FORMAT(/11X,"METHOD",I2/11X,"CONF INTERVAL:",F9.3," ,",
     1F9.3/6X,"LENGTH OF INTERVAL:",F9.3," ,",F9.3,5X,"TOTAL LENGTH:"
     2,F9.3/7X,"ACONJ AT EXTREMES:",F9.3," ,",F9.3)
  800 FORMAT(" TROUBLE IN R33CI 33-LOOP, SAMPLE:",I5)
  801 FORMAT(" TROUBLE IN -LOOP,"
     1," SAMPLE:",I4)
  333 FORMAT(" METHOD:",I3," TAU:",F3.2)
  888 FORMAT(I5,F9.3)
  897 FORMAT(" TNEW, START IN R33CI:",2F9.3)
      RETURN
      END

      SUBROUTINE SETLOC(T1,S1,SIZE,METHOD)
      REAL X(100),Y(100),COF(4)
      INTEGER INDY
      COMMON/CNS/C,N,F,N1,F,Z,BETA,EFF
      COMMON/CONFIG/Y,CON,INDY
      COMMON/SAMP/X
      COMMON/CNSTNT/DELTA,GAMMA,EPSLN,ZETA
      COMMON/TAUVAL/WTEST,ESTNO
      SIZE=1.
      ZERO=0.
      TWO=2.
      FOUR=4.
      CON2=ZERO
      IF(METHOD.GT.2)GOTO 12
      DO 10 I=1,N
        W=PSI((X(I)-T1)/S1,C)*S1/DELTA
        Y(I)=W
        CON2=CON2+W*W
   10 CONTINUE
      GOTO 20
   12 A1=ZERO
      A2=ZERO
      A3=ZERO
      A4=ZERO
      A5=ZERO
      WTEST=ZERO
      DO 15 I=1,N
        W=(X(I)-T1)/S1
        WTEST=WTEST+TAU(W,C,BETA,Z1)
        W=PSI(W,C)
        Y(I)=W
        A1=A1+PSIP(W,C)
        WT=T*PSI(W,C)
        WB=W*W-BETA
        A2=A2+WT
        A3=A3+WT*WT
        A4=A4+WB*WB
        A5=A5+WB*A3
   15 CONTINUE
      RETURN
      ALSQ=A1*A
      DENOM=A1*A...
      EFTCO=A1*(A3*(F*A4*B-T1-A5*A4)+A ...*A5/FOUR)
     1 /A3/DENOM
      XCONF=A1*F/DENOM
      DO 18 I=1,N
        W=Y(I)
        W=XCONF*(A3*W-...*(W*W-BETA)A1,A5)
        Y(I)=W
        CON=CON+W*W
   18 CONTINUE
   20 COF(1)=CON/F
      CALL GROVEC(Y,N,...)
CC
CC    WRITE(6,99)(Y(I),I=1,N)
   99 FORMAT(" COEFFICIENTS",F9.3/(...X,...F9.3))
      RETURN
      END

      SUBROUTINE ROBCI(CLNGTH,ICOUNT,XLOC,METHOD,
     1ICOV,LNGTH,TLOW,TUP)
```

```
C           A4=A4+W*W 3
C           A5=A5+W*B*W*3
   10 CONTINUE
C
      RETURN
C     ALSQ=AL*A
C     DENOM=AL*AL-A 3D
C     ESTFU=S1*(A3*(F*AL*BETA-AL*A4)+A2*A5/FOUR)
C          /A3/DENOM
C     XCONF=S1*F/DENOM
C     DO 20 I=1,F
C        W=Y(I)
C        W=XCONF*(A3*W-  *(W*W-BETA)/TWO)
C        Y(I)=W
C        CUML=CUML+W*W
C     CONTINUE
   25 CUM(3)=CUM(3)/F
      CALL ORDV C(Y,F,IDY)
CC
CC    WRITE(3,99)(Y(I),I=1,4)
   99 FORMAT(/" CONFIG:",5X,1 F9.3/(11X,1F9.3))
      RETURN
      END

      SUBROUTINE HULL1(CLNGTH,ICOUNT,XLOC,METHOD,
     1ICONV,LENGTH,TLOW,TUP)
      REAL CLNGTH(12),X(10),Y(100),CUM(4)
     1,LENGTH(12)
      INTEGER ICOUNT(12),ICONV(12),IDY
      COMMON/TABLE/TABLE4,TABLET
      COMMON/SAMP/X
      COMMON/CONFIG/Y,CUM,IDY
      COMMON/CNB/CNT,F,F1,F2,Z,BETA,EFF
      COMMON/HUBEST/T1,S1
      COMMON/CNSTNT/DELTA,GAMMA,EPSLN,Z T
      CALL STUD(X,Y,T1,S1)
      DHAT=0.
      DO 10 I=1,N
         DHAT=DHAT+PSIP(Y(I),S)
   10 CONTINUE
      DHAT=(DHAT/F)**2
      DN=SQRT(S1*S1+BETA/DELTA/DELTA/F))
      CONST1=DN*TABLET
      TLOW=T1-CONST1
      TUP=T1+CONST1
      LNGTH=2.*CONST1
CC
CC    WRITE(6,2 )METHOD,TLOW,TUP,LNGTH
      CALL COVER(TLOW,TUP,METHOD,XLOC,ICOUNT,
     1ICONV)
      LENGTH(METHOD)=CLNGTH(METHOD)+LNGTH
      LENGTH(METHOD)=CLNGTH
    2 FORMAT(/1X,"METHOD",I3/11X,"CONF INTERVAL:",F9.2," ,",
     1F9.3,11X,"LENGTH:",F9.3)
      RETURN
      END

      REAL FUNCTION COF(A)
******
*     CONSTANT GENERATING FUNCTION, K(A), OF THE CONFIGURATION
******
      REAL Y(100),CUM(4)
      INTEGER IDY
      COMMON/CONFIG/Y,CUM,IDY
      COMMON/CNB/CNT,F,F1,F2,Z,BETA,EFF
      ZERO=0.
      COFP=ZERO
      S1=ZERO
      IF(IDY.EQ.1)GO TO 20
      DO 5 I=1,IDY-1
         S1=S1+EXP(A*Y(I))
    5 CONTINUE
   20 IF(IDY.GT.N)GO TO 15
      DO 10 I=IDY,N
         COFP=COFP+EXP(A*Y(I))
   10 CONTINUE
   15 COF=ALOG(COFP+S1)/F)
      RETURN
      END

      REAL FUNCTION COFPRT(A)
******
*     CONVERTS COFP TO FORM COMPATIBLE WITH ZEROIN
******
      COMMON/SHIFTY/START,XFACT
      COFPRT=XFACT*COFP(A)
      RETURN
      END

      REAL FUNCTION COFP(A)
******
*     CALCULATES KF(A) - START
```

```
      IF(INDY.GT.N)GOTO 13
    8 DO 10 I=INDY,N
        CGFP=CGFP+EXP(A*Y(I))
   10 CONTINUE
   13 CGF=ALOG((CGFP+S1)/F)
      RETURN
      END

      REAL FUNCTION CGFPRT(A)
******
*       CONVERTS CGFP TO FORM COMPATIBLE WITH MYROOT
******
      COMMON/SHIFTY/SHIFT,XFACT
      CGFPRT=XFACT*CGFP(A)
      RETURN
      END

      REAL FUNCTION CGFP(A)
******
*       CALCULATES K'(A) - SHIFT
******
      REAL Y(2  ),CUM(4)
      INTEGER INDY
      COMMON/CONFIG/Y,CUM,INDY
      COMMON/CMB/C,N,F,F1,F2,Z,BETA,EFF
      COMMON/SHIFTY/SHIFT,XFACT
      ZERO=0.
      DENOM=ZERO
      CGFP=ZERO
      S1=ZERO
      S2=ZERO
      IF(INDY.LE.1)GOTO 8
      I=INDY-1
      DO 5 I=1,I
        W1=Y(I)
        W2=EXP(A*W1)
        S1=S1+W1*W2
        S2=S2+W2
    5 CONTINUE
      IF(INDY.GT.N)GOTO 13
    8 DO 10 I=INDY,N
        W1=Y(I)
        W2=EXP(A*W1)
        CGFP=CGFP+W1*W2
        DENOM=DENOM+W2
   10 CONTINUE
   13 CGFP=(CGFP+S1)/(DENOM+S2)-SHIFT
C     WRITE(6,10)A,CGFP
C   1 FORMAT(' A,CGFP',2F10.4)
      RETURN
      END

      REAL FUNCTION CGFPP(A)
******
*       CALCULATES K''(A)
******
      REAL Y(2  ),CUM(4)
      INTEGER INDY
      COMMON/CONFIG/Y,CUM,INDY
      COMMON/CMB/C,N,F,F1,F2,Z,BETA,EFF
      COMMON/SHIFTY/SHIFT,XFACT
      ZERO=0.
      DENOM=ZERO
      CGFPP=ZERO
      S1=ZERO
      S2=ZERO
      IF(INDY.LE.1)GOTO 8
      I=INDY-1
      DO 5 I=1,I
        W1=Y(I)
        W2=EXP(A*W1)
        S1=S1+W1*W1*W2
        S2=S2+W2
    5 CONTINUE
      IF(INDY.GT.N)GOTO 13
    8 DO 10 I=INDY,N
        W1=Y(I)
        W2=EXP(A*W1)
        CGFPP=CGFPP+W1*W1*W2
        DENOM=DENOM+W2
   10 CONTINUE
   13 CGFPP=(CGFPP+S1)/(DENOM+S2)
C     WRITE(6,1)A,CGFPP
C   1 FORMAT(' A,CGFPP',2F10.4)
      RETURN
      END

      SUBROUTINE ORDVEC(Y,N,INDY)
******
*       ORDERS THE VECTOR Y
*       INDY RECORDS POSITION OF FIRST NONNEGATIVE ENTRY
******
      REAL Y(100)
      INTEGER N
```

```
      CGFPP=ZERO
      S1=ZERO
      S2=ZERO
      IF(INDY.EQ.1)GOTO 8
      I1=INDY-1
      DO 5 I=1,I1
        W1=Y(I)
        W2=EXP(-*W1)
        S1=S1+W1*W1*W2
        S2=S2+W2
    5 CONTINUE
      IF(INDY.GT.N)GOTO 10
    8 DO 9 I=INDY,N
        W1=Y(I)
        W2=EXP(-*W1)
        CGFPP=CGFPP+W1*W1*W2
        DEN=DEN+W2
    9 CONTINUE
   10 CGFPP=(CGFPP+S1)/(DEN+S2)
C     WRITE(6,10)CGFPP
      FORMAT(' CGFPP=',F12.5)
      RETURN
      END

      SUBROUTINE ORDVEC(Y,N,INDY)
******
*         ORDERS THE VECTOR Y
*         INDY, LOCATION POSITION OF FIRST POSITIVE ENTRY
******
      REAL Y(N)
      ZERO=0.0
      IFAIL=0
      CALL M01ANF(Y,1,N,IFAIL)
      DO 30 I=1,N
        IF(Y(I).GT.ZERO)THEN
          INDY=I
          RETURN
        ENDIF
   30 CONTINUE
      INDY=N+1
      RETURN
      END

      SUBROUTINE PREP(IDIST,TRUVAL,N,NSAV,STUDY,IFIX,
     _XI,DSEED,CONTR)
      REAL THETA(2),CLEVEL(2),X(N),XI(1 )
     _,XORIG(1 )
      DOUBLE PRECISION DSEED
      COMMON/CNS/CONS,PI,P2,Z,BETA,EPP
      COMMON/CLEVEL/CLEVEL
      COMMON/PARAM/
      COMMON/PIVAL/PI,SQ2PI
      COMMON/ISEED/ISEED,IOL2,IOL3,IOL4,IOL
      COMMON/TABLE/TABLEN,TABLEPT
      COMMON/PRODCT/XPROD
      PI=X02AAF()
      SQ2PI=SQRT(PI*2.0)
      IOL1=...
      IOL2=...
      IOL3=...
      IOL4=...
      IOL5=...
      ZERO=0.
      IFAIL=0
      READ ...,CLEVEL,CLT
      IF(IDIST.NE.4)THEN
        READ 11,CONS,EPP,...,CLEVEL,TRUVAL
        WRITE(6,119)
        WRITE(6,120)...CLEVEL,TRUVAL
        GOTO
      ENDIF
      CALL DSEED(DSEED)
      DSEED=...
      CALL G05CBF(DSEED)
      IF(IDIST.EQ.4.)THEN
        READ ...CONS,EPP,...,CLEVEL,TRUVAL
        WRITE(6,220)CONS,EPP,...,CLEVEL,TRUVAL
        WRITE(6,230)...
        IF(...)THEN
          WRITE(6,240)
        ENDIF
        X(1)=TRUVAL
        READ 1,...(2),CONTRAN
        READ 95,IOPY
        READ...(XI(I),I=1,N)
        DO 2 I=1,N
          XORIG(I)=XI(I)
        CONTINUE
        WRITE(6,...)(XI(I),I=1,N)
        GOTO
      ENDIF
      IF(IDIST.EQ.3)THEN
*****
*       P(2)=SLOPE
```

```
      WRITE(6,2 )IDI 1
      IF(IDIST.EQ.2)THEN
        WRITE(6,240)
      ENDIF
      B(1)=TRUVAL
      READ 122,B(2),CONTAM
      READ 95,IDFY
      READ [illegible],(XI(I),I=1,N)
      DO   I=1,N
        XORIG(I)=XI(I)
      CONTINUE
      WRITE(6,2 9)(XI(I),I=1,N)
      GOTO
      ENDIF
      IF(IDIST.EQ.3)THEN
******
*     B(2)=SLOPE
*     ~~B(2)=INTERCEPT~~
******
      READ 12 ,N,S,ERR,NS,IC,CLEVEL,TRUVAL
      WRITE(6,1 )N,S,ERR,NS,IC,CLEVEL,TRUVAL
      B(1)=TRUVAL
      READ   ,B(2)
      READ 95,IDFY,IDFX
      WRITE(6,2 )IDFX,IDFY,B(2)
      GOTO
      ENDIF
      READ 1  ,N,S,ERR,NS,IC,CLEVEL,AMEAN(1),AMEAN(2),
     1 COVAR(1,1),COVAR(1,2),COVAR(2,2)
      TRUVAL=COVAR(1,2)/COVAR(2,2)
      CALL GENER(AMEAN,COVAR,[illegible])
      WRITE(6,110)N,S,ERR,NS,IC,CLEVEL,AMEAN(1),COVAR(1,1),
     1 COVAR(1,2),[illegible],COVAR(1,2),COVAR(2,2)
      B(1)=TRUVAL
      CALL [illegible](DELT,DEST,[illegible],ZET,[illegible])
      F=FLOAT(N)
      F1=FLOAT(N-1)
      F2=FLOAT(  -  )
      Z=SQRT(F)
      [illegible]=[illegible]/F
      [illegible]L=1.-CLEVEL
      CALL [illegible](TA [illegible],TABLET,[illegible])
      [illegible]L=CLEV L+(A L/ . )
      CALL [illegible](T[illegible],TABLE,[illegible])
      [illegible](24)
      FORMAT(F1 . /F . /I /F /F   . / .)
  119 FORMAT(/" DATA SUPPLIED")
      FORMAT( X," HUBER'S C:",F6.4/
     1   X," SAMPLE SIZE:",I3/
     1 4X," INCREMENT IN SAMPLE SIZE:",F .1/
     1 7X," NUMBER OF SAMPLES:",I3/
     1 3X," CONFIDENCE LEVEL:",F5.2/
     1" TRUE VALUE OF REGRESSION PARAMETER:",F5. )
  12  FORMAT(F . /F . / . / /(F . ))
  122 FORMAT(F .  )
      FORMAT(7 X," HUBER'S C:",F5.2/
     1 X," INCREMENT SAMPLE SIZE:",F . /
     1  X," SAMPLE SIZE:",I3/
     1 7X," NUMBER OF SAMPLES:",I3/
     1 3X," CONFIDENCE LEVEL:",F5.2/
     1 X," DIST: BIVARIATE NORMAL",
     1 X," MEAN:",F . ," COVAR:",F . /
     1 X,F . , X,F . /" TRUE VALUE OF",
     1" REGRESSION PARAMETER:",F5. )
      FORMAT(" PREDICTORS CHI-SQUARE, DF=",I3/
     1" ERROR STUDENT T, DF=",I3/
     1" TRUE VALUE OF INTERCEPT:",F5. )
      FORMAT(" X SUPPLIED, Y [illegible]SUT[illegible]:", )
      FORMAT(" Y SUPPLIED")
      FORMAT( X,7F . )
      FORMAT("  XI  "/( X,7F1 .4))
      RETURN
      END

      SUBROUTINE GENRAT( N,B,[illegible],I,[illegible],IDFY,IDFX,
     1 DBLED,CONTAM)
******
*      IDIST=1 READ DATA
*            1 X TO CHI-SQUARE, DF=IDFX
*              ERROR TO STUDENT T, DF=IDFY
*              B(2) HOLDS TRUE VALUE OF INTERCEPT
*              B(1) HOLDS TRUE VALUE OF SLOPE
*            DEFAULT: BIVARIATE NORMAL
*            2 X SUPPLIED
*                B(2) INTERCEPT (IF REQUIRED)
*                B(1) SLOPE
*                     ERROR IS [illegible] NORMAL
*                1        [illegible]
*                   Y SUPPLIED
*                     [illegible]   STUDENT T-1, IDFY
*                4                CHI-SQUARE, IDFY
*                5                [illegible]
******
      [illegible]
```

```
      RETURN
      END

      SUBROUTINE RGEN(NSAMP,IDIST,ICURVE,IDFY,IDFX,
     _DSEED,CONTAM)
******
*          IDIST=1 READ DATA
*              2 X ETC CHI-SQUARE, DF=IDFX
*                ERROR ST STUDENT-T, DF=IDFY
*                R(1) HOLDS TRUE VALUE OF INTERCEPT
*                R(2) HOLDS TRUE VALUE OF SLOPE
*              DEFAULT: BIVARIATE NORMAL
*              >5 X SUPPLIED
*                 R(1) INTERCEPT (IF REQUIRED)
*                 R(2) SLOPE
*                 =5 ERROR IS STANDARD NORMAL
*                 =1        SLASH
*                 =  Y SUPPLIED
*                 =5 ERROR IS STUDENT-T, IDFY
*                 =4           CHI-SQUARE, IDFY
*                 =5           CONTAMINATED NORMAL
******
      REAL Y(N),X(N),XSAMP(2),R(2)
     _,XORIG(N)
      DOUBLE PRECISION DSEED
      COMMON/PARAM/N
      COMMON/SAMPL/Y,X
      COMMON/CNS/...,F,Z,BETA,EPS
      COMMON/PLOT/XORIG
      IF(IDIST.GT.4)GOTO 4
      IF(IDIST.EQ.1)THEN
        READ(5,*)(Y(I),X(I),I=1,N)
        GOTO 31
      ENDIF
      IF(IDIST.EQ.1)THEN
        IFAIL=0
        DO 10 I=1,N
          X(I)=G05DHF(IDFX,IFAIL)
          Y(I)=R(1)+R(2)*X(I)+G05DDF(IDFY,IFAIL)
        CONTINUE
        GOTO 30
      ENDIF
      DO  I=1,N
        IFAIL=
        CALL G05EZF(XSAMP,2,X,IOUT,IFAIL)
        Y(I)=XSAMP(1)
        X(I)=XSAMP(2)
      CONTINUE
30    IF(ICURVE.NE.  )THEN
        WRITE(6,1 )NSAMP,(Y(I),X(I),I=1,N)
      ENDIF
      RETURN
4     I   =IDIST-4
      CALL GENER(Y, DSEED, I   , SAMP, IDFY,CONTAM)
      IF(       .EQ.  )THEN
        DO    I=1,N
          Y(I)=R(1)+R(2)*XORIG(I)+Y(I)
        CONTINUE
      ENDIF
      IF(ICURVE.EQ.  )THEN
        WRITE(6,13 )NSAMP,(Y(I),I=1, )
      ENDIF
      FORMAT("(Y,X)",   SUBSAMPLE  ,/(5X,  F   .4))
      FORMAT(5X,    F   .4,/(5X,3F   .4))
13    FORMAT(   ,   F  .4/(5X,  F  .4))
      RETURN
      END

      SUBROUTINE LE   (STEPS,SIZE,XL,  ,IMOVE,THETA,ICOUNT,
     _LENGTH,METHOD,   V,LNGTH,   ,  L )
      REAL Y(   ),X(   ),RESULT(  ),THETA(3),REST(1  )
     _,   (4),SCORE(3),WGHT(  ),LNGTH(  ),LNGTH(12)
      INTEGER ICOUNT(12),ISCOV(12)
      EXTERNAL FUNKG
      COMMON/SAMPL/Y,X
      COMMON/TABLE/TABLE,ITABLE
      COMMON/CNS/  ,  ,F,Z,BETA,EPS
      COMMON/TOLER/TOL1,TOL2,TOL3,TOL4,TOL5
      COMMON/WGHTER/WGHT
      N=N+
      IW=
      IFAIL=
      ZERO=.
      NUM=1
      IF(METHOD.GT.  )THEN
        THETA1=THETA(1)
        THETA2=THETA(2)
        BEST=THETA( )
        GOTO 35
      ENDIF
******
*          CLASSICAL REGRESSION LINE
******
      CALL G02CAF(N,X,Y,RESULT,IFAIL)
```

```
  [illegible]
      FORMAT(5X,CF1 [illegible]/(5X,5F[illegible]))
  1[illegible] FORMAT([illegible],11F1[illegible].4/(5X,11F1[illegible].4))
      RETURN
      END

      SUBROUTINE LI[illegible](STEP,SIZE,XL[illegible]C,IROVAL,THETA,ICOUNT,
     1 LENGTH,METHOD,[illegible]V,LNGTH,E[illegible])
      REAL Y([illegible]),X1([illegible]),RESULT(20),THETA(3),R[illegible]([illegible])
     1,W[illegible](4[illegible]),[illegible]([illegible]),WGHT([illegible]),LNGTH([illegible]),LNGTH(12)
      INTEGER [illegible]([illegible]),I[illegible]V(12)
      EXTERNAL [illegible]
      COMMON/[illegible]/Y,X1
      COMMON/TABLE/TABL[illegible]
      COMMON/CRO/[illegible]
      COMMON/TOLER/TOL1,TOL2,TOL3,TOL4,T[illegible]
      COMMON/WGHT[illegible]/WGHT
      [illegible]=[illegible]
      [illegible]=[illegible]
      [illegible]=
      ZERO=0.
      [illegible]=1
      IF(METHOD.GT.[illegible])THEN
        THETA1=THETA(1)
        THETA2=THETA(3)
        [illegible]=THETA([illegible])
        GOTO [illegible]
      ENDIF
******
*       CLASSICAL REGRESSION LINE
******
      CALL G[illegible](N,X1,Y,RESULT,[illegible])
      THETA1=RESULT(7)
      THETA3=RESULT(8)
      [illegible]=SQRT(RESULT(1[illegible]))
      [illegible]=RESULT(3)*TABLE[illegible]
CC
CC    WRITE(6,1[illegible])STEP
      [illegible]=(THETA3-STEP)*SIZE
      [illegible]=(THETA3+STEP)*SIZE
      [illegible]=[illegible]-[illegible]
      CALL [illegible]([illegible],[illegible],[illegible],[illegible]V)
      LENGTH(NW[illegible])=LENGTH(NW[illegible])+LENGTH
      LENGTH(NW[illegible])=LENGTH
      [illegible]=THETA[illegible]*SIZE
      [illegible]=THETA1-[illegible]*XL[illegible]
CC
CC    WRITE(6,1[illegible])[illegible]
******
*       B-LOOP IS HUBER'S ITERATIVE PROCEDURE TO CALCULATE
*       ROBUST REGRESSION LINE
******
      DO [illegible] J=1,1[illegible]
******
*       IMPROVE SCALE ESTIMATE
******
        [illegible]=ZERO
        DO [illegible]
          W1=WGHT([illegible])
          R1=P[illegible](([illegible]-THETA1-THETA3*X1([illegible]))/[illegible]/W1,[illegible])
          [illegible]=R1*[illegible]
          [illegible]([illegible])=R1
          [illegible]=[illegible]+[illegible]
        CONTINUE
        [illegible]=[illegible]/BETA/WGHT([illegible])
C       WRITE([illegible])([illegible]([illegible]),I=1,N)
******
*       REGRESS RESIDUALS ON X1
******
        CALL [illegible]CAT(N,X1,[illegible],RESULT,[illegible])
        THETA1=THETA1+RESULT(7)
        THETA3=THETA3+RESULT(8)
        [illegible]=[illegible]T([illegible])
C       WRITE(6,1[illegible])THETA1,[illegible],THETA[illegible]
      CONTINUE
      THETA(1)=THETA1
      THETA(3)=THETA[illegible]
      THETA([illegible])=[illegible]
******
*       CALCULATE ROBUST REGRESSION LINE
******
      CALL [illegible]([illegible],THETA,[illegible],[illegible],
     1 [illegible],[illegible])
      IF(THETA([illegible]).LT.ZERO)THEN
        THETA([illegible])=-THETA([illegible])
        CALL [illegible](FC[illegible],THETA,[illegible],[illegible],
     1 [illegible],[illegible])
      ENDIF
      [illegible]=THETA(3)*SIZE
      [illegible]=THETA(1)-[illegible]*XL[illegible]C
CC
CC    WRITE(6,1[illegible])[illegible],THETA(2),[illegible]
      IF(METHOD.GT.[illegible])RETURN
CC    [illegible]
```

```
      WRITE(6,1..) ..'J(1),Y=1,..)
******
*     REGRESS NEW RESIDUALS ON X1
******
      CALL GLCAF(N,X1,RESI),RESULT,IFAIL)
      THETA1=THETA1+RESULT(7)
      THETA3=THETA3+RESULT(5)
      BEST=.. T(..EW)
      WRITE(6,1..)THETA1,BEST,THETA..
    9 CONTINUE
      THETA(1)=THETA1
      THETA(3)=THETA3
      THETA(2)= EST
******
*     CALCULATE ROBUST REGRESSION LINE
******
      CALL C..NBF(FC..R..,3,THETA,SCORE,TOLL,
     _WS,LW,IFAIL)
      IF(THETA(2).LT.ZERO)THEN
      THETA(2)=-THETA(2)
      CALL C..NBF(FCORLG,3,THETA,SCORE,TOLL,
     _WS,LW,IFAIL)
      ENDIF
      C..SL=THETA(3)*SIZE
      C..RC..T=THETA(1)-C..RSL*XLOC
CC
CC    WRITE(6,1..)C..I.,THETA(2),C..RSL
      IF(MEIH .. .GT...)RETURN
CC
CC    WRITE(6,1..)..LI..L..D..,TENGTH
  1.. FORMAT(7..X,"METHOD  1"/" CLASSICAL ...NT-RV'L:",
     _F9.3," ,",F9.3,"     LEAST ..:",F9.3)
  1.. FORMAT(75X,"  ROBUST ESTIMATES:",3F9.3)
  1.. FORMAT(" REGR... .. ESTIMATES:",3F9.3)
  1.. FORMAT(" WEIGHTED ..  FOR CL.. ...:",7(..)
  1.. FORMAT(" NEGATIVE SNEW")
  14. FORMAT(97X,"COMPUT.. NT    SCALE    ..L..:"/75X,
     _" CLASSICAL ..T.S:",3F9.3)
  15. FORMAT("  ..PT",F9.3)
      RETURN
      END

      SUBROUTINE FCOR..G(ITHETA,THETA,SC..,NSC,FLAG)
      INTEGER IFLAG,ITHETA,I
      REAL THETA(3),SCORE(3),X1(1..),Y(1..)
     _,WGHT(1..)
      COMMON/SAMPLE/Y,X1
      COMMON/CNS/C..,F,..,H,F ,Z,BETA,EFF
      COMMON/WGHTEN/WGHT
      ZERO=0.
      SCORE1=ZERO
      SCORE2=ZERO
      SCORE3=ZERO
      DO 1 I=1,N
      W1=WGHT(I)
      RESID=PSI((Y(I)-THETA(1)-THETA(3)*X1(I))/THETA(2)/W1,C)
      W2=RESID*W1
      SCORE1=SCORE1+W..
      SCORE2=SCORE2+X1(I)*W2
      SCORE3=SCORE3+(W2*W2-BETA)
    1 CONTINUE
      SCORE(1)=SCORE1/F
      SCORE(3)=SCORE2/F
      SCORE(2)=SCORE3/F/2.
      RETURN
      END

      SUBROUTINE SLOT1(SCORR,THETA)
      REAL X1(1..),Y(1..),G(1..),THETA(3),WGHT(1..),COV(4)
      COMMON/COMFIG/..,..,..,DY
      COMMON/WGHTEN/WGHT
      COMMON/CNS/C..,F,H,F ,Z,BETA,EFF
      COMMON/SAMPLE/Y,X1
      ZERO=0.
      A11=ZERO
      A12=ZERO
      A23=ZERO
      A33=ZERO
      DO 1.. I=1,N
      GI=(Y(I)-THETA(1)-THETA(3)*X1(I))/THETA(2)
      GI..=PSIP(GI/WGHT(I),C)
      GTX=GI..*X1(I)
      A11=A11+GI..
      A22=A22+GTX*GTX
      A23=A23+GTX*G..
      A33=A33+GI*GI*GI..
      G(I)=PSI(GI/WGHT(I),C)
    . CONTINUE
      DET=(A11*A33-A23*..3)/F
      SCORR=THETA(2)/DET
      CORR=ZERO
      DO 2.. I=1,N
      GI=SCORR*A33*G(I)*WGHT(I)*X1(I)
```

```
      C[illegible]/CONFIG/[illegible]
      COMMON/WGHTER/WGHT
      COMMON/CNO/C[illegible],F,F1,F2,Z,BETA,EFF
      COMMON/SAMPLE/Y,XI
      ZERO=0.
      [illegible]=ZERO
      [illegible]=ZERO
      [illegible]=ZERO
      [illegible]=ZERO
      DO [illegible] I=1,N
      GI=(Y(I)-THETA(1)-THETA(4)*XI(I))/THETA([illegible])
      GI =[illegible](GI/WGHT([illegible]),C)
      GTX=GTX*XI(I)
      [illegible]=[illegible]+GTI
      [illegible]=[illegible]+GTX*GI X
      [illegible]=[illegible]+GTX*G
      [illegible]=[illegible]+GI*GI*GIX
      G(I)=[illegible](GI/WGHT(I),C)
      CONTINUE
      DET=([illegible]*[illegible]-[illegible]*[illegible])/F
      [illegible]=THETA(2)/DET
      [illegible]=ZERO
      DO [illegible] I=1,N
      GI=[illegible]*[illegible]*G(I)*WGHT(I)*XI(I)
      G(I)=GI
      [illegible]=[illegible]+GI*GI
      CONTINUE
      [illegible]=[illegible]/F
      CALL [illegible](G,[illegible])
      WRITE(6,99)(G(I),I=1,N)
99    FORMAT(/[illegible] CONFIG[illegible],[illegible]X,[illegible]F9.3/([illegible]X,[illegible]F9.3))
      [illegible]=[illegible]**2
      RETURN
      END

      SUBROUTINE SLOET[illegible](G[illegible],THETA)
      REAL XI(1[illegible]),Y(1[illegible]),G(1[illegible]),THETA([illegible]),WGHT([illegible]),[illegible]([illegible])
      COMMON/CO[illegible]FIG/G
      COMMON/WGHT[illegible]/WGHT
      COMMON/CNO/C[illegible],F,F[illegible],F2,Z,BETA,EFF
      COMMON/HALFY/[illegible]Y
      COMMON/SAMPLE/Y,XI
      COMMON/CONS[illegible]/CON
      [illegible]=[illegible].
      TWO=[illegible].
      DO [illegible] I=1,N
      W=WGHT(I)
      K=P[illegible](((Y(I)-THETA(1)-THETA(4)*XI(I))/THETA(2)/W,C)
      G[illegible]=G(I)*GCON[illegible]*(W*W*[illegible]-BETA)/TWO
      G(I)=GI
      [illegible]=[illegible]+GI*GI
10    CONTINUE
      CON(I)=CON[illegible]/F
      CALL [illegible](G,[illegible])
      WRITE(6,99)(G(I),I=1,N)
99    FORMAT(/[illegible] CONFIG[illegible],[illegible]X,[illegible]F9.3/(1[illegible]X,[illegible]F9.[illegible]))
      RETURN
      END

      SUBROUTINE GETHAT(SH[illegible]R,XHAT)
      REAL XI(1[illegible]),Y(1[illegible]),XHAT(1[illegible])
      COMMON/SAMPLE/Y,XI
      COMMON/CNO/C[illegible],F,F[illegible],F[illegible],Z,BETA,EFF
      [illegible]=[illegible].
      X[illegible]=[illegible].
      TWO=[illegible].
      [illegible]=ZERO
      S4=ZERO
      DO [illegible] I=1,N
      WI=XI(I)
      [illegible]=[illegible]+WI
      S4=S4+WI*WI
1[illegible]  CONTINUE
      DET=F*S4-[illegible]*S[illegible]
      DO [illegible] I=1,N
      WI=XI(I)
      XHAT(I)=[illegible](X[illegible]-[illegible]+WI*(W[illegible]*F-TWO*S[illegible]))/DET)
[illegible]   CONTINUE
      W[illegible]=ZERO
      DO [illegible] J=1,N
      W[illegible]=WI+XHAT(J)
[illegible]   CONTINUE
      XHAT(N+1)=W[illegible]
      [illegible]=[illegible]*[illegible](X[illegible]*F/F/DET)
      WRITE(6,1[illegible])(XHAT(I),I=1,[illegible])
1[illegible]  FORMAT(/[illegible] HAT WGT[illegible],(1[illegible]F1[illegible].[illegible]))
      RETURN
      END

      SUBROUTINE [illegible]([illegible]Z[illegible],T[illegible])
      REAL XI(1[illegible]),W[illegible]([illegible]),THETA([illegible]),S[illegible]([illegible])
     -,Y(1[illegible]),SAM(1[illegible])
      INTEGER ISORT(1[illegible])
```

```
      ZERO=0.
      ONE=1.
      TWO=2.0
      S2=ZERO
      S4=ZERO
      DO 10 I=1,N
        WI=XI(I)
        S2=S2+W
        S4=S4+WI* WI
      ...
      DET=F*S4-S2*S2
      DO  I=1,N
        WI=XI(I)
        XHAT(I)=C...(X...-(S4+WI*(WI*F-TWO*S2))/DET)
   20 CONTINUE
      WI=ZERO
      DO 30 J=1,N
        WI=WI+XHAT(J)
   30 CONTINUE
      XHAT(N+1)=W
      ...=SQRT(X...*F/F/DET)
      WRITE(3,1  )(XHAT(I),I=1,N)
  1   FORMAT(/" HAT WGTS",(10F10.3))
      RETURN
      END

      SUBROUTINE  ...X(..,IZ,T1)
      REAL XI(I  ),W  (  ),THETA(2),SCOR  (2)
     ,Y(1  ),S  (3  )
      INTEGER IDUM2(  )
      EXTERNAL FCN
      COMMON/SAMPL/Y,X
      COMMON/TOLER/TOL,TOL,TOL3,TOL4,...
      COMMON/CHB/...,F,...,Z,BETA,EFF
      COMMON/SAMP/S,M
      Z...=0.
      X...=0.
      X...=1.
      ...=1.
      DO    I=1,
        ...(I)=X1(I)
      CONTINUE
      CALL  ...(...)
      THETA(1)=T...
      THETA(2)=(S1*X...+...*X...)/T...
      IFAIL=
      CALL ...(FCN,..THETA,..SCOR,...TOL...,IFAIL)
      IF(THETA(2).LT.ZERO)THEN
        THETA(2)=-THETA(2)
        CALL ...(FCN,2,THETA,SCOR ,TOL...,IFAIL)
      ...IF
      ...=THETA(1)
      ...=THETA(2)
      ...Z...=...
      DO    ...
        X...(I)=(XI(I)-T...)/...
      CONTINUE
      WRITE(3,...)(Y(I),XI(I),I=1,M)
      IFAIL=
      WRITE(3,1  )
CC
CC
C     WRITE(3,...)...ZE
C     CALL XC4ADF(...)
C     CALL X...AGF(1,3)
C     CALL ...AGF(XI,Y,...,IFAIL)
C     CALL X...ZBF(...)
C     DO    I=1,N
C       ...(I)=XI(I...T(I))
C     CONTINUE
C     DO   I=1,
C       XI(I)=...(I)
C     CONTINUE
CC
CC    WRITE(3,1  )(Y(I),XI(I),I=1,M)
 130  FORMAT(/10X,"PLOT OF Y V R...  ...ZED X")
 131  FORMAT("  ...LL",F9.4)
 1    FORMAT("  ...LIZED  SAMPLE"/(6X,6F1 .4))
      RETURN
      END
```

... KL... SAMP.

[redacted]

```
C       [illegible]
C
CC      WRITE(5,2[illegible])(Y(I),X1(I),I=1,N)
  150 FORMAT(/1X,"PLOT OF Y VERSUS NORMALIZED X")
  151 FORMAT(" SIZE",F9.3)
  1[illegible] FORMAT(" NORMALIZED ",A"PLOT"/(6X,6F1[illegible].4))
        [illegible]
```

[illegible]