
Estimating Tail Probabilities

D. B. Carr
H. D. Tolley

December 1982

Prepared for the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Operated for the U.S. Department of Energy
by Battelle Memorial Institute



DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC06-76RLO 1830

Printed in the United States of America
Available from
National Technical Information Service
United States Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151

NTIS Price Codes
Microfiche A01

Printed Copy

Pages	Price Codes
001-025	A02
026-050	A03
051-075	A04
076-100	A05
101-125	A06
126-150	A07
151-175	A08
176-200	A09
201-225	A010
226-250	A011
251-275	A012
276-300	A013

3 3679 00058 9988

PNL-4580
UC-32

ESTIMATING TAIL PROBABILITIES

D. B. Carr
H. D. Tolley

December 1982

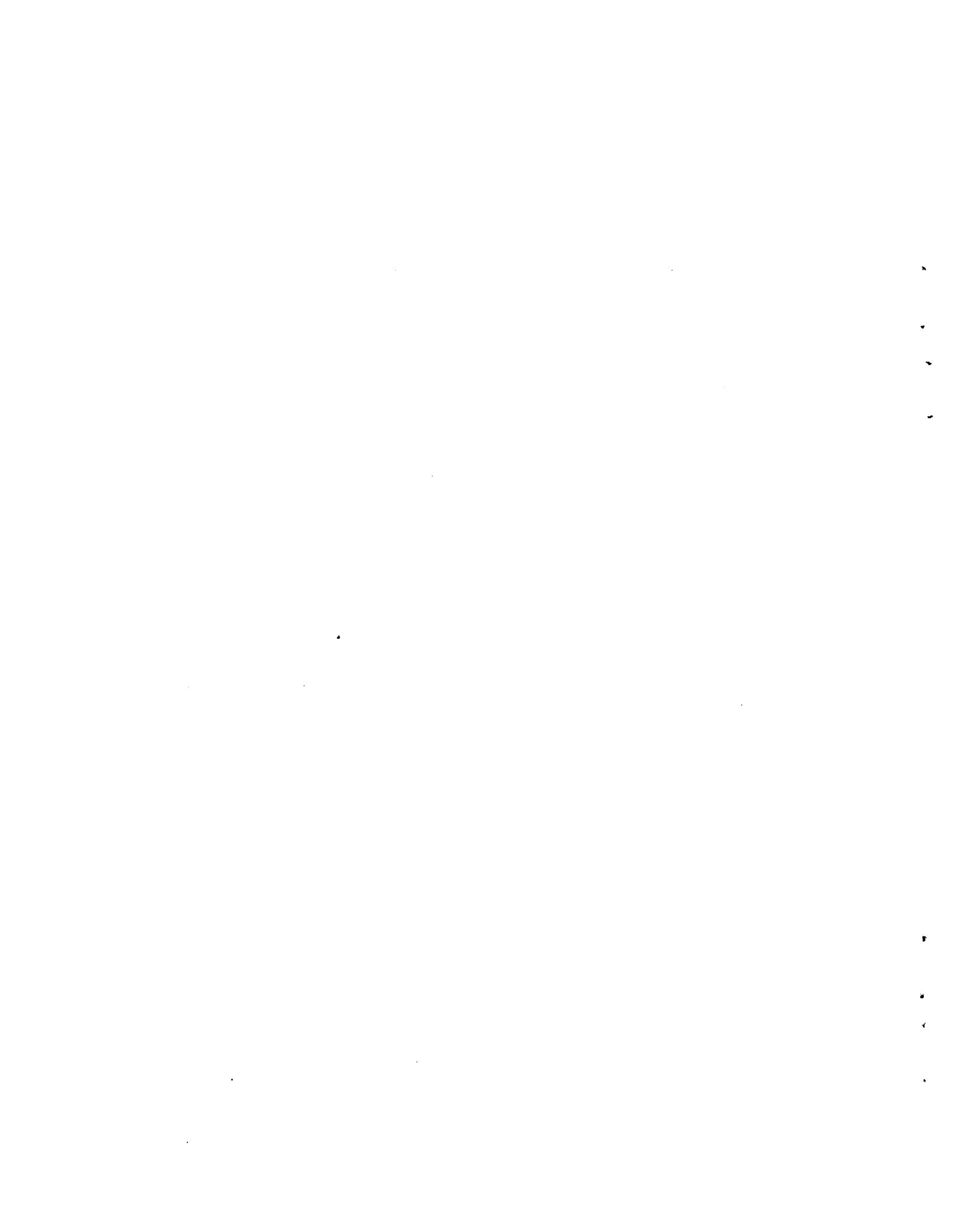
Prepared for
the U.S. Department of Energy
under Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352



ABSTRACT

This paper investigates procedures for univariate nonparametric estimation of tail probabilities. Extrapolated values for tail probabilities beyond the data are also obtained based on the shape of the density in the tail. Several estimators which use exponential weighting are described. These are compared in a Monte Carlo study to nonweighted estimators, to the empirical cdf, to an integrated kernel, to a Fourier series estimate, to a penalized likelihood estimate and a maximum likelihood estimate. Selected weighted estimators are shown to compare favorably to many of these standard estimators for the sampling distributions investigated.



CONTENTS

ABSTRACT	iii
1.0 INTRODUCTION	1.1
2.0 ESSCHER'S WEIGHTING PROCEDURE	2.1
3.0 WEIGHTED ESTIMATION PROCEDURES	3.1
3.1 CHOICE OF A WEIGHTING FUNCTION	3.1
3.2 CHOICE OF A NONPARAMETRIC DENSITY ESTIMATION PROCEDURE	3.2
3.3 A WEIGHTED DATA FOURIER ESTIMATION PROCEDURE	3.3
3.4 AN INTEGRATED KERNEL ESTIMATE	3.4
3.5 OTHER ESTIMATION PROCEDURES	3.6
4.0 THE MONTE CARLO STUDY	4.1
4.1 DISTRIBUTIONS INVESTIGATED AND GENERATION METHODS	4.2
4.2 RESULTS	4.2
5.0 SUMMARY AND CONCLUSIONS	5.1
6.0 REFERENCES	6.1

FIGURE

3.1 Truncated Exponential Weight Function	3.2
---	-----

TABLES

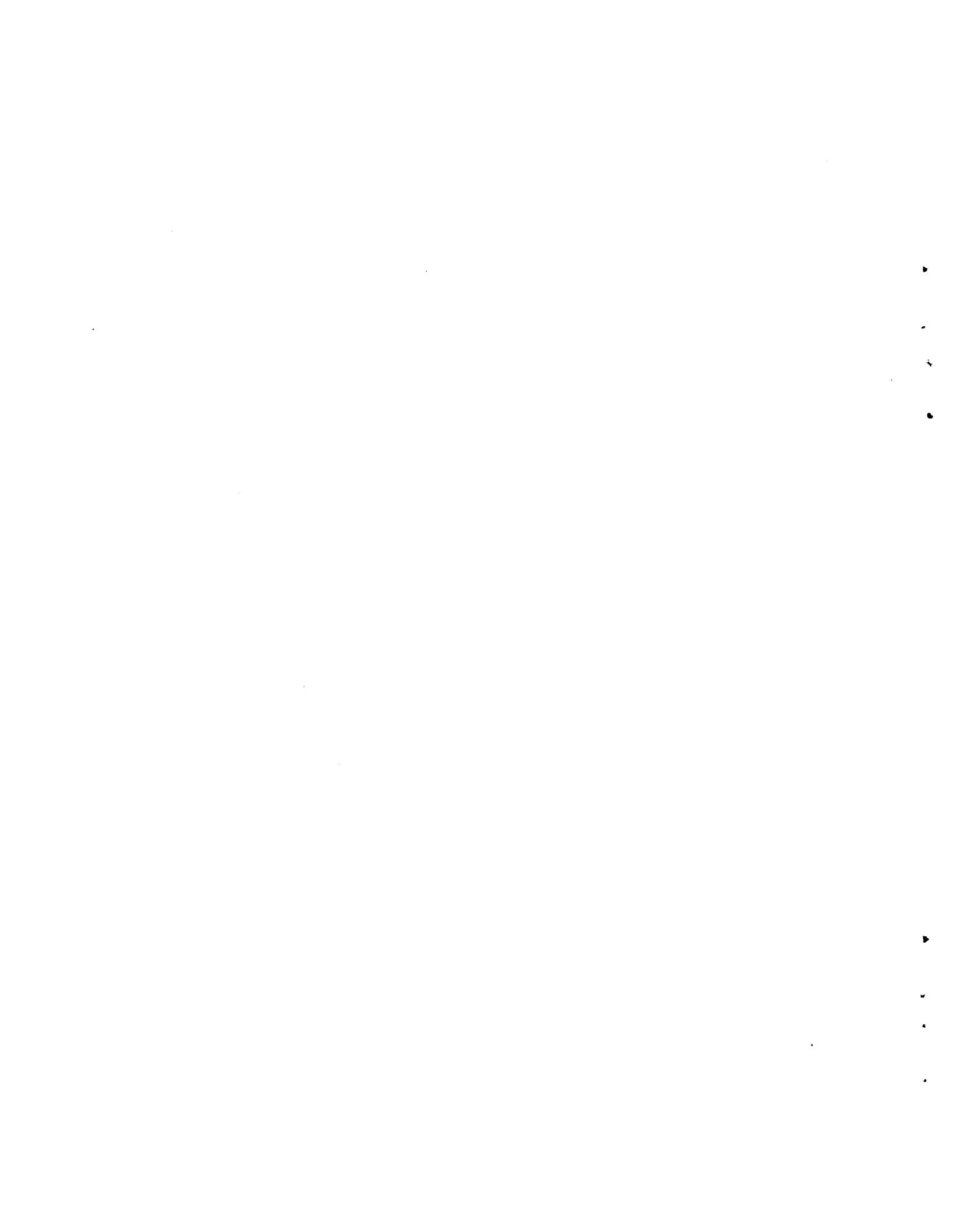
2.1 Relative Errors in Approximating a χ^2 Distribution	2.5
2.2 Relative Error in Estimating a Standard Normal with 5% Contamination from $N(\mu_2, 0.04)$	2.6
4.1 Mean-Square Relative Error for a T_{30} Distribution	4.3
4.2 Mean-Square Relative Error for a T_{10} Distribution	4.3
4.3 Mean-Square Relative Error for a T_3 Distribution	4.4
4.4 Mean-Square Relative Error for a Standard Normal with 5% Contamination from $N(2, 0.25)$	4.4
4.5 Mean-Square Relative Error for a Standard Normal with 1% Contamination from $N(2, 0.25)$	4.5
4.6 Mean-Square Relative Error for a Standard Normal with 5% Contamination from $N(4, 0.25)$	4.5
4.7 Mean-Square Relative Error for a Standard Normal with 1% Contamination from $N(4, 0.25)$	4.6

ESTIMATING TAIL PROBABILITIES

1.0 INTRODUCTION

This paper discusses the nonparametric estimation of tail probabilities. Statistical data analysts frequently evaluate tail probabilities. Typically, tail probabilities are found by making assumptions and looking up results in tables of theoretical distributions. However, if a substantial amount of data is available as a reference distribution, assumptions can be avoided and tail probabilities can be evaluated more directly. Large data sets frequently contain relevant reference distributions. For many problems the empirical cdf based on such reference distributions will answer the questions about tail probabilities. However, the empirical cdf or binomial estimate of tail probability is somewhat rough. We believe that tail estimates can be improved by smoothing. Density estimates are smoothed estimates that correspond to a smoothed cdf and provide a starting place for estimating tail areas. The exact choice of a density estimation procedure and the corresponding smoothing is not crucial in the center distribution where there is a lot of data. The smoothing does become important when we are so far out in the tails that very little data is directly relevant, and especially important in the extrapolation problem, when the critical point is beyond all the data.

The major method developed for empirical cdf smoothing in this paper is a weighted-data Fourier series procedure. To motivate the use of weighted data, this paper first presents Esscher's procedure for estimating and approximating tail probabilities. Then weighted procedures are discussed in more detail and the Fourier series smoothing method is developed. To evaluate this smoothing method and other smoothing methods a Monte Carlo study is performed and those results are described.



2.0 ESSCHER'S WEIGHTING PROCEDURE

The problem of estimating tail areas is a major concern to actuaries trying to control the probability of company ruin due to accumulation of claims in too short a period. To specifically address the actuarial problem, Esscher (1932) proposed weighting the data before trying to represent the density using an Edgeworth series. Although the technique does not globally improve the estimation of the density, it considerably improves the density estimate in the tail. In the actuarial literature, the technique is used in estimating the right-hand tail of the compound Poisson distribution, and the estimate is amazingly accurate (Bohman 1963; Bohman and Esscher 1963-1964).

For the general Esscher procedure, define the following:

$$f(x) = \text{density for which tail area is desired} \quad (2.1)$$

$$S(t) = \int_{t_0}^{\infty} f(x)dx \quad (2.2)$$

$$c_k(s) = \int_{-\infty}^{\infty} x^k e^{sx} f(x)dx \quad k = 0, 1, \dots, \quad (2.3)$$

$$g(x,s) = c_0^{-1}(s) e^{sx} f(x) \quad (2.4)$$

$$t_0 = \text{critical point in the right-hand tail} . \quad (2.5)$$

The Esscher technique transforms the problem from a question regarding the tail of $f(x)$ to one regarding the tail of $g(x,s)$, for a suitable value of s . After weighting the $f(x)$ by e^{sx} , the constant $c_0^{-1}(s)$ is used to make $g(x,s)$ a density.

The Esscher procedure is summarized as follows:

1. Determine s^* such that

$$\int_{-\infty}^{\infty} x g(x, s^*) dx = t_0 . \quad (2.6)$$

2. Replace $g(x, s^*)$ by an Edgeworth expansion $\hat{g}(x, s^*)$ with moments

$$\mu'_k = \frac{c_k(s^*)}{c_0(s^*)} . \quad (2.7)$$

3. Approximate $S(t_0)$ by

$$S(t_0) \approx \int_{t_0}^{\infty} c_0(s^*) e^{-s^* x} \hat{g}(x, s^*) dx . \quad (2.8)$$

The Edgeworth expansion for $g(x, s^*)$ is given by

$$\begin{aligned} \hat{g}(x, s^*) = & \left[Z(y) - \frac{1}{6} \sqrt{B_1} Z^{(3)}(y) + \frac{1}{24} (B_2 - 3) Z^{(4)}(y) \right. \\ & \left. + \frac{1}{72} B_1 Z^{(6)}(y) \right] / \sigma , \end{aligned} \quad (2.9)$$

where

$$Z^{(j)}(y) = \frac{d^j}{dx^j} e^{-y^2/2} / \sqrt{2\pi} , \quad (2.10)$$

and

$$y = (x - t_0) / \sigma . \quad (2.11)$$

The standard deviation, skewness and kurtosis in (2.9) is a function of s^* through the central moments, i.e.,

$$\sigma = \sqrt{\mu_2} \quad (2.12)$$

$$\sqrt{B_1} = \mu_3 \mu_2^{-3/2} \quad (2.13)$$

and

$$B_2 = \mu_4 \mu_2^{-2} \quad (2.14)$$

These are found from μ_k in (2.7) using the relation

$$\mu_r = \sum_{j=0}^r (-1)^j \binom{r}{j} (\mu'_1)^j (\mu'_{r-j})^{r-j} \quad (2.15)$$

Changing variables in (2.8) gives

$$S(t_0) = c_0(s^*) e^{-s^* t_0} \left[\sum_{j=0}^6 e_j \int_0^\infty e^{-s^* \sigma y} d\Phi^{(j)}(y) \right]. \quad (2.16)$$

Φ is the Gaussian cdf. The e_j 's are 0 except for

$$e_0 = 1; \quad e_3 = -\frac{1}{6} \sqrt{B_1}; \quad e_4 = \frac{1}{24} (B_2 - 3); \quad e_6 = \frac{1}{72} B_1. \quad (2.17)$$

Following Beard et al. (1969), the integral in (2.16) is derived recursively using

$$E_j(x) = \int_0^\infty e^{xy} d\Phi^{(j)}(y) = -\Phi^{(j)}(0) + x E_{j-1}(x). \quad (2.18)$$

Since

$$\sqrt{2\pi} E_0(x) = \frac{1 - \Phi(x)}{\sqrt{2\pi} \Phi'(1)(x)} = A(x)/\sqrt{2\pi} , \quad (2.19)$$

then

$$\begin{aligned} \sqrt{2\pi} E_1(x) &= xA(x) - 1 , \\ \sqrt{2\pi} E_2(x) &= x^2A(x) - x , \\ \sqrt{2\pi} E_3(x) &= x^3A(x) - x^2 + 1 , \\ \sqrt{2\pi} E_4(x) &= x^4A(x) - x^3 + x , \\ \sqrt{2\pi} E_5(x) &= x^5A(x) - x^4 + x^2 - 3 , \end{aligned} \quad (2.20)$$

and

$$\sqrt{2\pi} E_6(x) = x^6A(x) - x^5 + x^3 - 3x .$$

Thus

$$\hat{s}(t_0) = c_0(s^*) e^{-s^* t_0} \sum_{j=0}^6 e_j E_j(s^*) . \quad (2.21)$$

In many applications, only the first two nonzero terms are used in (2.21).

With asymptotic evaluation in mind, we illustrate the procedure when $f(x)$ is known to have a gamma density

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} . \quad (2.22)$$

Then

$$g(x, s) = \frac{\beta^\alpha x^{\alpha-1} e^{-(\beta-s)x}}{c_0(s)\Gamma(\alpha)} . \quad (2.23)$$

From Step 1, knowing the moments of $g(x, s)$ gives

$$s^* = \beta - \alpha/t_0 , \quad (2.24)$$

with standard deviation, skewness and kurtosis of $g(x, s^*)$ given by

$$\sigma = \alpha^{1/2}/(\beta = s^*) = \alpha^{-1/2}t_0 , \quad (2.25)$$

$$\sqrt{B_1} = 2\alpha^{-1/2} \quad (2.26)$$

and

$$B_2 = 3 + 6/\alpha \quad (2.27)$$

For the values $\alpha = n/2$ and $\beta = 1/2$, a chi-square with n degrees of freedom is obtained. The results are shown in Table 2.1.

Examination of Table 2.1 illustrates several points. First, the relative errors are small [see for example, Gross and Hosmer (1978) and Andrews (1973)]. This is expected for large degrees of freedom, when the central limit theorem

TABLE 2.1. Relative Errors $(p - \hat{p})/p$ in Approximating a χ^2 Distribution

Target p	Degrees of Freedom			
	10	5	2	1
0.05	-0.003	-0.007	-0.024	-0.056
0.01	-0.002	-0.004	-0.008	-0.002
0.001	-0.001	-0.001	0.004	0.032
0.0005	-0.001	0	0.006	0.036
0.0001	0	0.001	0.009	0.040

applies. Surprisingly, the estimate works well for the thicker tails corresponding to fewer degrees of freedom. The Esscher density estimate at t_0 is a saddlepoint approximation (Daniels 1954) for the density of a *mean*, but without large degrees of freedom the ordering of terms for the series expansion is not established. Consequently, the good results for one degree of freedom are unexpected.

Values for the tables are easily computed using a continued fraction expansion [see Bouver and Bargmann (1979); and Abramowitz and Stegun (1968)] for Mills ratio in (2.19). Were the bias smaller, an implementation of (2.21) for the incomplete gamma distribution might be considered. The bias is relatively small in this case, but can be large in others.

Table 2.2 shows the relative errors from the Esscher procedure from approximating the tails of a contaminated Gaussian distribution. The Esscher procedure is unable to follow the bump in the tail added by the contamination. Thus, the Esscher procedure, which works well for its intended application involving the compound Poisson and other situations involving the central limit theorem, should not be routinely used in nonparametric situations. In fact, for thick-tailed distributions, the theoretical moments in (2.3) do not exist. The procedure, however, suggests the use of weighted samples in conjunction with alternative density estimates. The La Guarre series, the Fourier series, and a host of other density estimates are candidates.

TABLE 2.2. Relative Error in Estimating a Standard Normal with 5% Contamination from $N(\mu_2, 0.04)$

Critical Point t_0	μ_2			
	0	1	2	3
0	0.0	0.0	0.0	0.0
1	0.01	-0.03	0.0	0.2
2	-0.07	0.18	-0.2	-0.2
3	-0.18	0.23	1.6	0.7
4	-0.31	0.29	3.4	16.0

3.0 WEIGHTED ESTIMATION PROCEDURES

In the previous section, the Esscher procedure improved tail estimates by weighting the data, approximating the density for the weighted data, and adjusting for the weights during integration. In this section, we investigate the same strategy using different weights and/or different estimation procedures.

3.1 CHOICE OF A WEIGHTING FUNCTION

The exponential weights used in the Esscher procedure are ideally suited to the Gaussian distribution, because the weighted density has a Gaussian distribution. Only the mean is changed. Similarly, the weight function changes only the scale when applied to the family of gamma densities. However, the general choice of an exponential weight function is unappealing since it dominates the tails of thick-tailed unimodal distributions. That is, for $s > 0$, $\lim_{x \rightarrow \infty} e^{sx} f(x) \rightarrow \infty$. Consequently, the proposed procedure uses the exponential function only up to the critical point and uses a constant weight thereafter. Further, zero weight is given to data less than the sample mean. Omitting the data below the mean focuses the modeling effort on the relevant portion of the distribution. Other cutoff points could be used. The resulting weight function is illustrated in Figure 3.1.

Because of the shape of the weight function, the weight parameter s in part controls the degree of smoothing that takes place. If s is large, the weight function approximates the step function corresponding to the binomial estimate. An advantage of the weight function given in Figure 3.1 is the simplicity in inverting it on the interval $[t_0, \infty)$. A minor disadvantage is that the discontinuity in the derivative of weight function leads to a corresponding discontinuity in the weighted density to be modeled. When smooth functions are used in the modeling, some lack of fit can be expected. Various values of s are considered in Section 4.

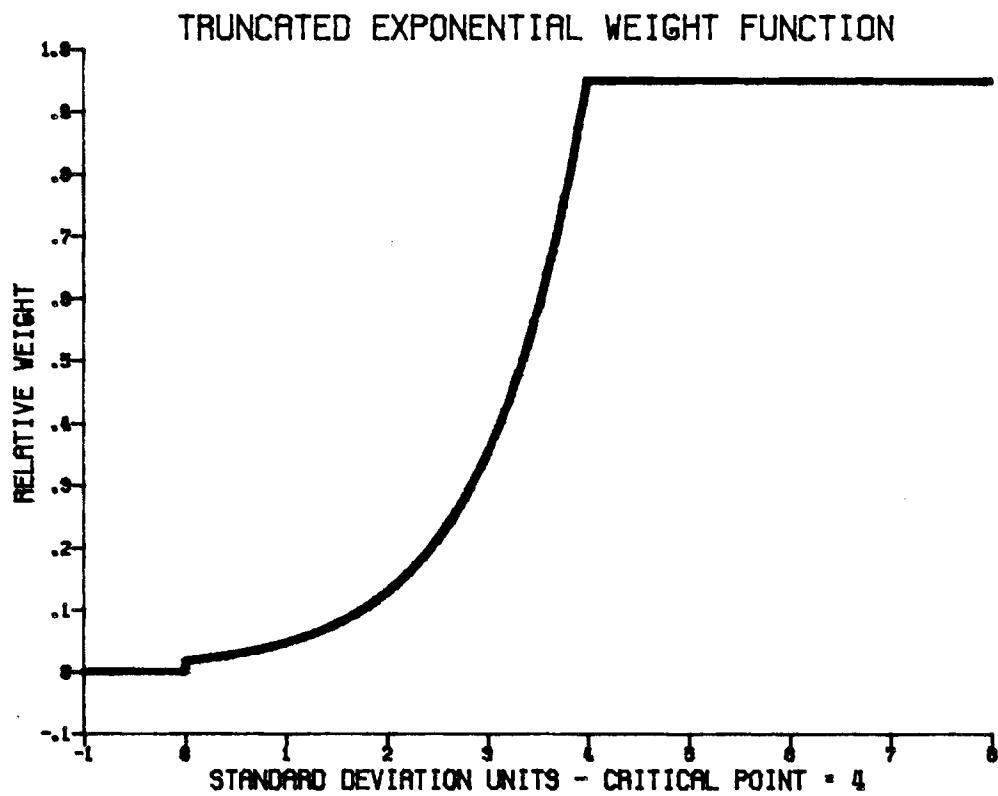


FIGURE 3.1. Truncated Exponential Weight Function

3.2 CHOICE OF A NONPARAMETRIC DENSITY ESTIMATION PROCEDURE

A variety of nonparametric procedures for estimating the density are available. Considerations in this choice include prior knowledge about the distribution to be modeled, theoretical optimality criteria, preferences of form for an estimation procedure and preference of form for smoothing parameter estimation. For illustrative purposes, we select two density estimation procedures: a Fourier series estimator (F) [see Kronmal and Tarter (1968); Tarter and Kronmal (1970, 1976); and Tarter (1979)] and a kernel estimator (K) [see Parzen (1962)]. We select the F estimator because its theoretical properties for unweighted samples are known and because the smoothing parameter, the number of terms in series expansions, is readily determined. The kernel estimate is chosen for comparison purposes. Technically, the F procedure can be expressed as a particular kernel procedure, but the choice of the smoothing parameter leads it to be viewed differently.

3.3 A WEIGHTED DATA FOURIER ESTIMATION PROCEDURE

Modifying the Fourier series estimator to accommodate the weight function in Figure 3.1 is straight forward. Let the sample be y_i , $i = 1, \dots, n$ and let the subsample x_i , $i = 1, \dots, n_1$ be observations with $x_i > \bar{y}$ where $\bar{y} = n^{-1} \sum y_i$. The estimates of the coefficients for a weighted Fourier procedure are given by

$$\hat{c}_k = \sum_{i=1}^{n_1} w_i \cos \left[\frac{k\pi(x_i - a)}{b - a} \right] \quad (3.1)$$

and

$$\hat{s}_k = \sum_{i=1}^{n_1} w_i \sin \left[\frac{k\pi(x_i - a)}{b - a} \right], \quad (3.2)$$

where

$$w_i = \frac{\min \left[e^{s(x_i - t_0)}, 1 \right]}{\sum_{j=1}^{n_1} \min \left[e^{s(x_j - t_0)}, 1 \right]} \quad \text{and } a = \bar{y}. \quad (3.3)$$

Then $\bar{x} = \sum_{i=1}^{n_1} w_i x_i$ and

$$s(t_0) = \frac{n_1}{n} \cdot \frac{\sum_{i=1}^{n_1} \min \left[e^{s(x_i - t_0)}, 1 \right]}{n_1 \pi} \cdot \left\{ 0.5\pi \left(1 + \frac{\bar{x} - t_0}{b - a} \right) \right. \\ \left. + \sum_{k=1}^m \frac{\hat{s}_k}{k} \cos \left[\frac{k\pi(t_0 - a)}{b - a} \right] - \sum_{k=1}^m \frac{\hat{c}_k}{k} \sin \left[\frac{k\pi(t_0 - a)}{b - a} \right] \right\}. \quad (3.4)$$

Note that the expression in brackets in (3.4) has the same form as $1 - F(t)$ in the unweighted Fourier series estimate. The factor n_1/n in (3.4) adjusts for omitting values below \bar{y} . The factor

$$n_1^{-1} \sum_{i=1}^{n_1} \min \left[e^{s(x_i - t_0)} , 1 \right]$$

adjusts for the weights and the weight inversion. The undetermined constants in (3.3) and (3.4) are the upper bound b , the weighting factor s , and the number of terms m . Selected values for b and s are investigated in Section 4.

The derivation of the optimal stopping rule to determine m in the unweighted F estimate does not hold for weighted samples. However, as sample sizes get large, distinguishing between a weighted sample and samples from the correspondingly weighted density becomes more difficult. This suggests that stopping rules for a sample from a weighted density would approximate the rules for a weighted sample from a nonweighted density. Therefore, based upon the results of Kronmal and Tarter, we use the rule to stop if

$$(\hat{s}_k^2 + \hat{c}_k^2) < \frac{2}{n_1 + 1} \quad (3.5)$$

for two consecutive terms, or if $m > n_1^{0.4375}$. The value of $0.4357 = 7/16$ is an arbitrary choice. This is to keep $m = o(n^{1/2})$ which ensures a variance of 0 as m and n go to infinity in the unweighted case.

In the following Monte Carlo study, the maximum number of terms is typically reached on the thicker-tailed distributions. This and the results of Kronmal and Tarter (1968) for the Cauchy distribution suggests that less than the optimal number of terms are being used. Hence, better results may be obtained than demonstrated in this paper.

3.4 AN INTEGRATED KERNEL ESTIMATE

The weighting concept is also applied to kernel estimates. Kernel density estimates have the form

$$f(\hat{y}) = \int \frac{1}{h} K\left(\frac{y-z}{h}\right) d\hat{F}_e(z) \quad (3.6)$$

where K is a kernel. A tail estimate is obtained by integration of $\hat{f}(y)$.

Thus

$$\hat{S}(t_0) = \int_{t_0}^{\infty} \hat{f}(y) dy = \int_{-\infty}^{\infty} H(z - t_0) d\hat{F}_e(z), \quad (3.7)$$

where

$$H(x) = \int_{-x/h}^{\infty} K(y) dy. \quad (3.8)$$

Following Tapia and Thompson (1978), the kernel

$$K(y) = \begin{cases} \frac{15}{16} (1 - y^2)^2 & \text{for } |y| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.9)$$

is selected. Thus

$$H(x) = \begin{cases} 1 & \text{for } x > h > 0 \\ \frac{1}{2} + \frac{15}{16} \left[\frac{x}{h} - \frac{2}{3} \left(\frac{x}{h} \right)^3 + \frac{1}{5} \left(\frac{x}{h} \right)^5 \right] & \text{for } -h < x < h \\ 0 & \text{for } x < -h \end{cases} \quad (3.10)$$

This kernel was chosen as one likely to be used in practice (at least prior to Silverman's 1981 fast Gaussian kernel algorithm). The choice is not based on knowledge of the density to be estimated. However, we exploit our knowledge of the sampling distribution to obtain optimal smoothing parameters. If we choose to minimize the mean-square error of the density over the tail, the optimal smoothing parameter becomes [Tapia and Thompson (1978)]

$$h = n^{-1/5} \frac{\int K^2(y) dy}{4 \left[\int y^2 K(y) dy / 2! \right]^2} \left[\frac{\int_{t_0}^{\infty} f(y) dy}{\int_{t_0}^{\infty} |f''(y)|^2 dy} \right]^{1/5} \quad (3.11)$$

$$= n^{1/5} \alpha(K) \beta(f) = n^{-1/5} 2.0362 \beta(f) . \quad (3.12)$$

Algorithms for computing $\beta(f)$ for the t distribution and the mixture of normals are given in the appendix.

Given h , modification of the integrated kernel estimate yields

$$s(t_0) = \left\{ \sum_{i=1}^{n_1} \frac{\min \left[e^{\frac{s(x_i - t_0)}{n}}, 1 \right]}{n} \right\} \sum_{i=1}^{n_1} w_i H(x_i - t_0) . \quad (3.13)$$

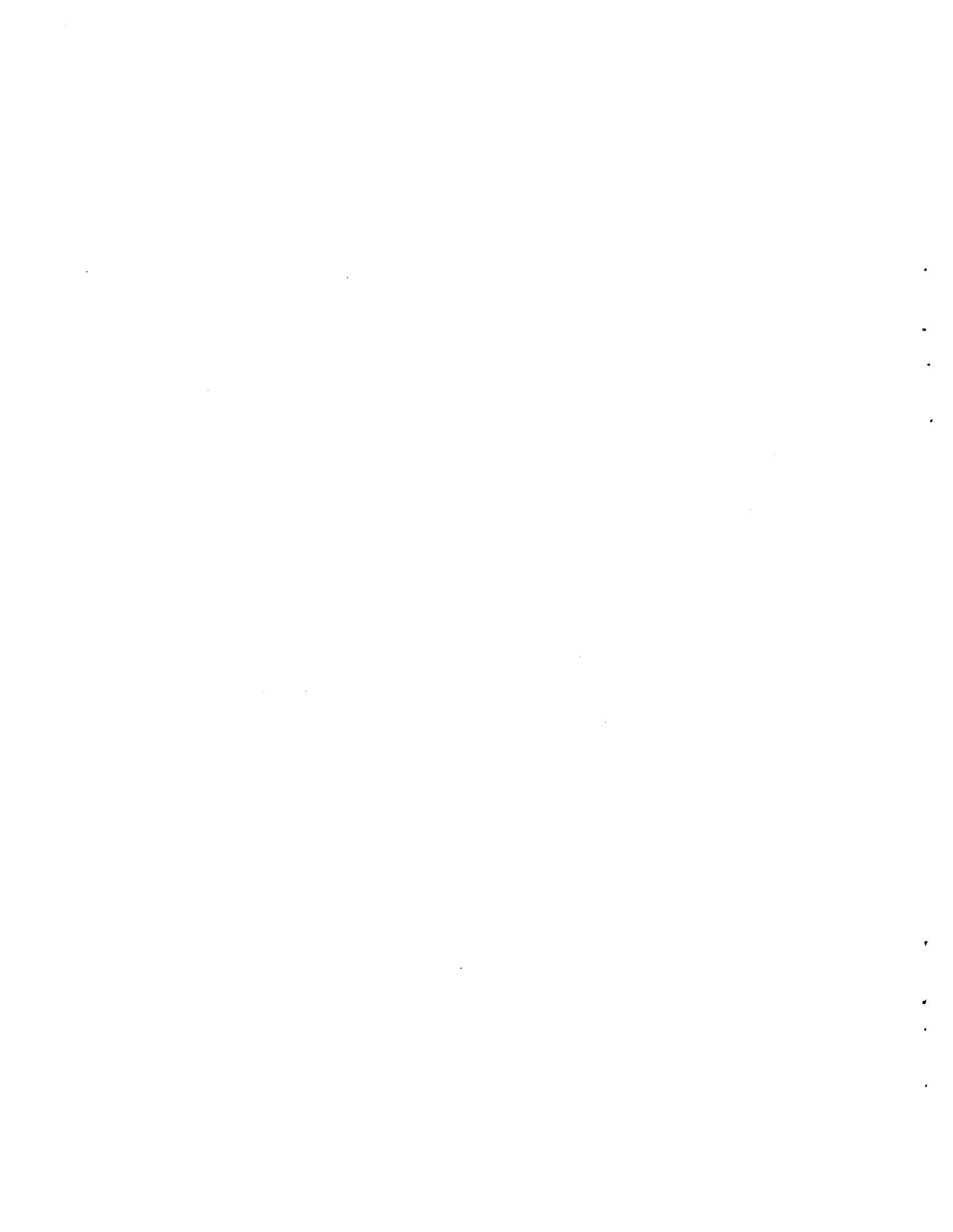
The weighting modifies the shape of the function $H(\cdot)$ and consequently of the kernel $K(\cdot)$. Since the shape of the kernel is of importance in the tails, the results are of interest.

3.5 OTHER ESTIMATION PROCEDURES

A variety of procedures can be used to estimate tails directly or modified to accommodate weighted samples. The interplay between the weighting, the smoothing parameter and the choice of density estimate can be explored for the numerous combinations. It is beyond the scope of this paper to look at this relationship in any detail.

Even with the domain of Fourier series estimates, there are numerous choices and variations. The specification of the modeling interval is one such choice. A significant variation involves Watson's (1969) extension to the orthogonal series estimate that includes a "damping" function (applied to the coefficients) parameterized by a smoothing parameter. Fellner and Tarter (1971) and Tarter and Raman (1972) develop the criterion of mean integrated weighted squared error. Using this the damping function can be determined so

the emphasis is placed on a chosen region of the density's support. Wahba (1978) considers the adaptive choice of the smoothing parameter. Anderson and DeFiqueiredo (1980) consider the adaptive choice of the orthogonal series and note some improvement in the tails. Thus in the Monte Carlo study below, only a few possibilities are being considered.



4.0 THE MONTE CARLO STUDY

To see how the weighted Fourier estimator (WF) performs, a Monte Carlo study is conducted. Several variations on the proposed estimate are considered. Three values of s in the weighting are used, $0.5/\hat{\sigma}$, $1/\hat{\sigma}$ and $2/\hat{\sigma}$. Thus the drop in weight to the left of the critical point is in sigma units and the estimator presumes the existence of two moments. For measured quantities this is a reasonable assumption. Three values were also used as the upper bound on the modeling interval. These are $k*\max(z_i)$, where k is 1.2, 1.3 and 1.4. z_i represents the standardized data. This impacts the number of terms used in the estimate as well as the coefficients. When the standardized critical points were beyond the upper bound, the estimate described below occasionally produced negative values. These estimates are also taken as 0.

For comparison purposes several other estimates are computed. The empirical cdf, denoted CDF is an obvious choice. An ordinary Fourier series estimate (F) is also used. The modeling interval is taken to be k times the extremes of the standardized data. Values of k used were 1.2, 1.3 and 1.4. Estimates from the optimally smoothed kernel (K) and the weighted variation (WK) as described in Section 3 are computed.

To consider another major approach to density estimation, tail estimates based on the penalized likelihood (PL) described in Tapia and Thompson (1978) are evaluated. The double precision IMSL routine NDMPLE is used. Here the data is scaled into the interval -2.75 to 2.75. The modeling interval was taken to be -3 to 3 with 29 internal partitions. The smoothing parameter α was set at a single number depending on the distribution being simulated. When using the same smoothing parameter the routine does not always converge. Thus the number of convergent results and the smoothing parameter are also reported.

Finally as a comparison against the parametric estimator, the maximum likelihood estimates (ML) for the mixture of normals are obtained when those distributions are studied. The maximum likelihood estimates are then used to obtain tail probabilities. The algorithm described by Hosmer (1973) is used

with iterations limited to 1000 and a change in the log likelihood convergence criterion of 0.0001. Again the algorithm did not always converge so the number of convergent results is reported.

4.1 DISTRIBUTIONS INVESTIGATED AND GENERATION METHODS

For sampling distributions, the t distribution with 30, 10 and 3 degrees of freedom is used. We also investigated mixtures of normals of the form $(1 - p)N(0, 1) + pN(\mu, \sigma^2)$, where $p = 0.05$ and 0.01 , $\mu = 2$ and 4 , and $\sigma^2 = 0.25$. The t distribution is used to investigate progressively thicker tails and the contaminated normals are used to investigate the effects of bumps.

To generate random numbers, the multiplicative congruential generator

```
SEED = 69069*SEED + 1 (MOD 2**32)
```

is used to generate uniform $[0,1]$ numbers. The 24 high-order bits of the 32 bit numbers are converted to floating point and are used as the next seed. Congruential generators of this type tend to have triples lying in planes, but are adequate for purposes here. For further details on congruential generators see Marsaglia (1976). The inversion to produce the t distribution uses the IMSL routine MDSTI and is considered sufficiently accurate (Hill 1970).

The study is performed by determining critical points corresponding to right-tail probabilities of 0.1, 0.005, 0.0001 and 0.0005. Results are summaries of 500 replicates of samples size 1000.

4.2 RESULTS

Only selected estimators are presented. For the WF estimate, $s = 2/\hat{\sigma}$ and $k = 1.4$ generally gives slightly better results. Consequently, we present only this variation.

The weighted kernel with $s = 0.5/\hat{\sigma}$ was usually the better variation and is reported here. Only the Fourier estimate with $k = 1.2$ is presented. Results not reported here are available from the authors on request. Since the estimators tend to have small bias, relative to their variability, we are reporting the mean-square relative errors MSRE. The CDF estimate typically has the smallest bias and confirms the consistency between the critical points

computed and the random numbers generated. Note that in the last column of Tables 4.1 through 4.7 the MSRE is greater than 1. Here the estimate "0" provides an improvement. Other error criteria may be more informative.

Some results are similar across the tables. The weighted data estimates WF and WK always gave smaller mean-squared error than the CDF estimate. With one exception the kernel estimate K gave smaller mean-squared error. If the underlying distribution is smooth, a smoothed estimate can be as good as or better cdf. Thus while the empirical cdf maintains the practical advantages of simplicity and availability of confidence bounds, for some situations smoothing procedures are warranted.

TABLE 4.1. Mean-Square Relative Error for a T_{30} Distribution

<u>Estimator</u>	Target p			
	<u>0.01</u>	<u>0.005</u>	<u>0.001</u>	<u>0.0005</u>
CDF	0.100	0.21	0.96	2.0
WF	0.092	0.19	0.88	1.6
WK	0.088	0.18	0.85	1.6
F	0.100	0.22	1.60	2.6
K	0.095	0.19	0.88	1.6
PL ($\alpha = 5$)	0.090	0.18	0.68	1.3

TABLE 4.2. Mean-Square Relative Error for a T_{10} Distribution

<u>Estimator</u>	Target p			
	<u>0.01</u>	<u>0.005</u>	<u>0.001</u>	<u>0.0005</u>
CDF	0.094	0.18	0.91	2.1
WF	0.085	0.16	0.84	1.9
WK	0.080	0.15	0.78	1.8
F	0.088	0.20	1.00	3.0
K	0.083	0.15	0.77	1.8
PL ($\alpha = 1$)	0.15	0.36	0.83	1.7

TABLE 4.3. Mean-Square Relative Error for a T_3 Distribution

Estimator	Target p			
	0.01	0.005	0.001	0.0005
CDF	0.091	0.19	0.81	1.7
WF	0.084	0.17	0.80	1.7
WK	0.083	0.17	0.74	1.5
F	(a)	(a)	(a)	(a)
K	0.084	0.17	0.74	1.5
PL ($\alpha = 0.1$)	1.278 ^(b)	0.37	1.3	2.2

(a) Very large.

(b) 486 of 500 cases.

TABLE 4.4. Mean-Square Relative Error for a Standard Normal with 5% Contamination from $N(2, 0.25)$

Estimator	Target p			
	0.01	0.005	0.001	0.0005
CDF	0.098	0.21	0.99	2.1
WF	0.085	0.17	0.79	1.7
WK	0.084	0.18	0.82	1.8
F	0.110	0.27	1.50	2.9
K	0.092	0.20	0.95	2.1
PL ($\alpha = 0.5$)	0.098 ^(a)	0.22	0.79	1.7
ML	0.077 ^(b)	0.13	0.40	0.7

(a) 497 of 500 cases.

(b) 484 of 500 cases.

The ordinary Fourier series and penalized likelihood estimates on the surface appear very desirable in that after a pass at the data, they provide a density estimate and readily computed tail estimates for any desired critical point. The occasional poor results, for example in Table 4.3 with the thick-tailed distribution suggests that care is required in selecting a modeling interval. Kronmal and Tarter are very careful in their selection of a

TABLE 4.5. Mean-Square Relative Error for a Standard Normal with 1% Contamination from $N(2,0.25)$

Estimator	Target p			
	0.01	0.005	0.001	0.0005
CDF	0.094	0.19	0.97	2.0
WF	0.084	0.17	0.82	1.5
WK	0.081	0.16	0.80	1.6
F	0.088	0.22	2.10	5.0
K	0.095	0.19	0.92	1.8
PL ($\alpha = 0.5$)	0.092 ^(a)	0.18	0.82	1.6
ML	0.081 ^(b)	0.15	0.41	0.6

(a) 492 of 500 cases.

(b) 471 of 500 cases.

TABLE 4.6. Mean-Square Relative Error for a Standard Normal with 5% Contamination from $N(4,0.25)$

Estimator	Target p			
	0.01	0.005	0.001	0.0005
CDF	0.077	0.19	1.3	2.1
WF	0.059	0.14	0.9	1.8
WK	0.070	0.17	1.1	2.0
F	0.067	0.19	1.1	1.7
K	0.067	0.17	1.1	1.9
PL ($\alpha = 0.05$)	0.067	0.17	0.9	1.7
ML	0.052	0.11	0.4	0.7

modeling interval when looking at a thick-tailed Cauchy distribution. They provide for not including all of the data. This seems inappropriate for the problem at hand.

That the penalized likelihood estimate was occasionally poor is a bit surprising. The subroutine documentation suggests throwing out outliers. When this is not done as in the above Monte Carlo, many consecutive intervals can have no data and the routine sometimes fails to converge. For the convergent situations with the thick-tailed t distribution, a closer look reveals

TABLE 4.7. Mean-Square Relative Error for a Standard Normal with 1% Contamination from $N(4,0.25)$

<u>Estimator</u>	Target p			
	<u>0.01</u>	<u>0.005</u>	<u>0.001</u>	<u>0.0005</u>
CDF	0.013	0.091	0.85	1.9
WF	0.011	0.078	0.68	1.3
WK	0.010	0.081	0.76	1.6
F	0.025	0.089	0.96	2.8
K	0.009	0.074	0.71	1.5
PL ($\alpha = 0.05$)	0.012	0.091	0.75	1.4
ML	0.011	0.084	0.51	0.8

that a few large tail estimates make almost all of the contribution to the mean-square error. In about half of these problem cases, an extreme data point (maximum or minimum) was present. Consequently the scaling bunches most of the data into a few intervals. The coarse binning in the region of the critical point leads to poor estimates. Thus omitting data not only helps with convergence but can improve the modeling in the region of interest. Of course the tail estimates would have to be adjusted to account for the omitted data.

In the remaining problem cases the poor estimates are due to some other source. For example, in the single problem case for the t distribution was with ten degrees of freedom, the discrete penalized likelihood estimate was 0.0618 and the binomial estimate was 0.011. The density estimates in the tail all seemed to be inflated relative to those in the center of the distribution. The reason for this is not self evident. A possibility is that a false convergence has been reached that is not clearly indicated by the log likelihood. Whatever the situation, the existence of a problem is clearly evident from the mismatch with the binomial estimate. With a little bit of care given to the data modeled, the number of intervals and the smoothing parameter, much better estimates can be obtained.

For the contaminated normal distribution, the penalized likelihood does not differ much from the binomial estimate at the 0.01 and 0.005 critical points. It provides improvement for the more extreme critical points. The situation may change with different selection of algorithm parameters.

Silverman (1982) points out that the penalty function can be chosen to yield a particular distribution when the smoothing is heavy. Thus the analyst has several decisions to make in using penalized likelihood procedures. Some choices can be made by looking at the data, but some seem arbitrary. How much the arbitrary choices impact the tail estimates remains an open question.

The kernel estimates performed fairly well in this study. This is expected since the optimal smoothing parameter was derived from the distributions used to generate the samples. We have no explanation why they did not perform better for the contaminated normals. The WK estimator is sometimes a little better than the K estimator. This suggests that the chosen kernel does not have the best shape for estimating tail values and that an asymmetric weighting in the window is helpful. To use a kernel procedure, the analyst must specify both the kernel and the smoothing parameter. For general density estimation purposes the exact shape of the kernel is not considered crucial. Unfortunately this result may not hold in the tails. Additionally, it is not well established how to get a good smoothing parameter for the tails. Methods like those in Silverman (1978) are not well suited to the tail problem.

The data weighted Fourier series estimator performed well. Results are biased somewhat in its favor since different settings of the arbitrary parameters s and b were investigated. However, since the one set reported performed reasonably well for all distributions investigated, there is hope that this set is a reasonable selection for many other distributions.

Perhaps the biggest criticism of Fourier series estimates is the possibility of negative tail estimates and of tail estimates that are not monotone decreasing as the critical point moves to the right. Trade offs are involved. Kernel density estimates that can only be positive cannot reduce the bias as much as estimates that can take on negative values. Consequently guaranteeing positive estimates costs in terms of increased bias. Fejer weights are sometimes used to guarantee that the resulting Fourier based estimates are positive. Our results and others indicate that such estimates are typically much poorer.

A minor criticism might be the computational cost, since each critical point requires a separate computation with the full data set. The Chebyshev recursive relationships for the sine and cosine are fast and have good accuracy

for a substantial number of terms [see Tarter et al. (1967) and Tarter and Kronmal (1968)]. Certainly FFT methods could be applied to binned weighted data. Thus the computational costs should not be prohibitive.

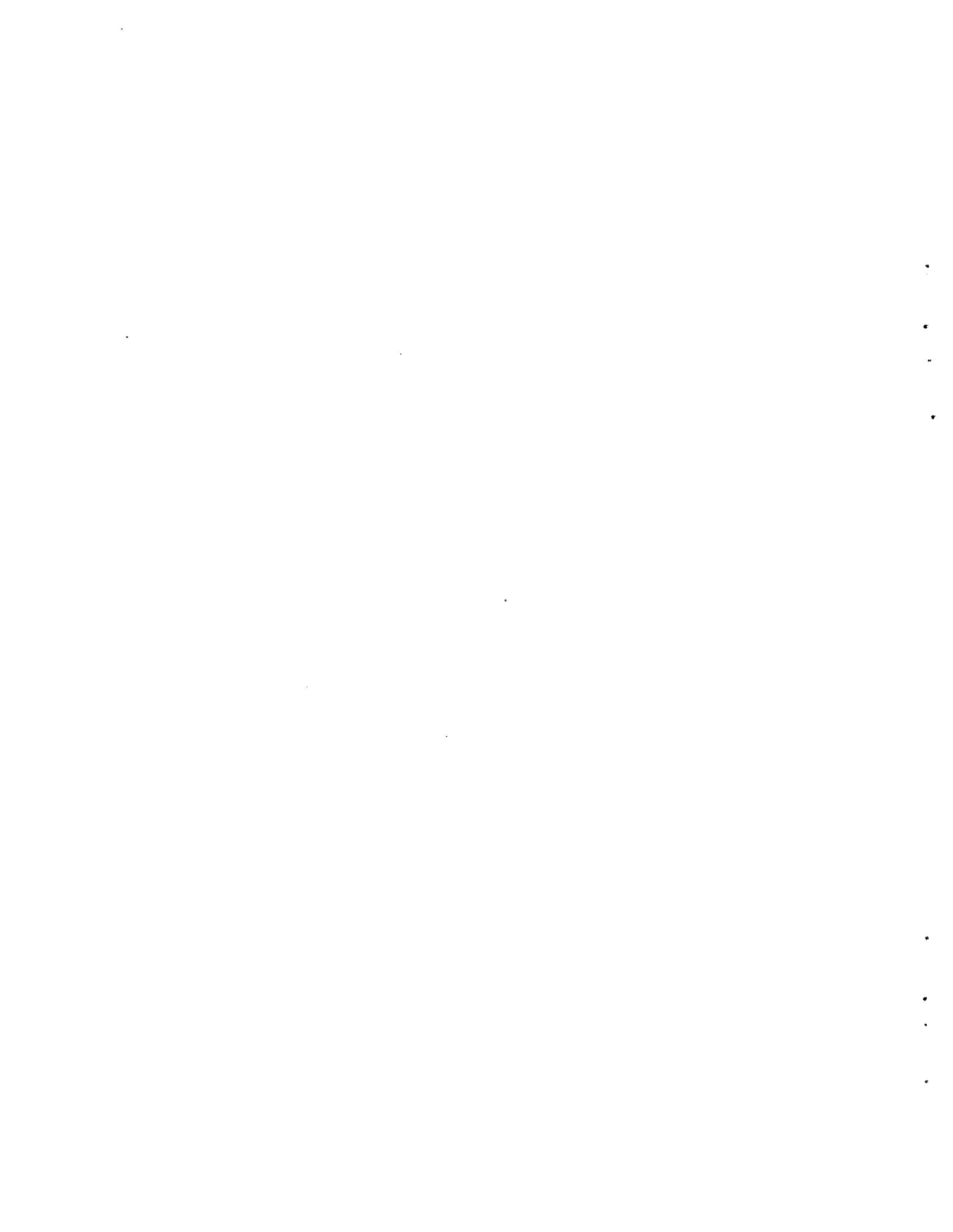
A third criticism concerns the lack of theoretical results about the properties of the estimator that do not carry over from the unweighted situation. The lack of theoretical results for smoothing is a general problem. Generally, density estimation procedures require the analyst to choose the smoothing and modeling parameters. Theoretical results do not typically incorporate the information that goes into such decisions.

What has the weighting bought? Basically it focuses the modeling attenuation on the part of the distribution of interest. It helps guarantee that bumps to the right of the critical point are modeled while still letting points to the left of the critical point have some contribution. The weighting has made the specification of the modeling interval much less crucial. Here the modeling interval is chosen so that the weighted data is not bunched up at the end of the interval. The naive algorithm works pretty well in the Monte Carlo study. Certainly better results could have been obtained by judiciously selecting the interval after looking at the data.

The mixture of normals maximum likelihood results address the question, how much could be gained if a parametric model could be specified? Conditional on convergence of the algorithm, the answer in this study is quite a lot for extreme critical points. However if one percent or more of the data is beyond the critical point the improvement is marginal. The fall-back position when the algorithm fails to converge is another issue. Estimating the parameters of the mixture of normals is not a trivial problem and better results may be obtained for more tractable distributions. However for almost all practical situations the question is irrelevant. Anyone who is mildly concerned about the *ad hoc* nature of the smoothing above should be horrified at the specification of a parametric model appropriate for the tail.

5.0 SUMMARY AND CONCLUSIONS

The empirical cdf estimator of the tail probability is simple and its properties are well known. However for many situations the underlying distribution is believed to be smooth and the empirical cdf estimate of the tail probability can be improved by smoothing. Generally available smoothing methods tend to be *ad hoc* and cannot always be applied naively while still obtaining improved estimates. As a part of smoothing process, weighting the sample is a reasonable way of placing emphasis on the tail of interest. In Section 3, two tail estimators, the Fourier series and the integrated kernel were modified to use weighted data. For the Fourier series, the weighting simplified the modeling process to the point where a very simple algorithm could produce consistently good tail estimates as demonstrated in a Monte Carlo study. The weighting did not markedly change the kernel estimates which had made explicit use of the optimal smoothing parameter. For practical situations in which the optimal smoothing parameter is not known, the weighted data Fourier series estimate provides one relatively safe method of obtaining smoothed tail estimates. It appears that weighting may also be used advantageously with the penalized likelihood and other density estimates, particularly when part of the modeling problem is associated with the tail that is not of interest. The best method for obtaining smoothed tail estimates remains an open issue.



6.0 REFERENCES

Abramowitz, M., and I. A. Stegun (ed.) 1968. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. National Bureau of Standards, Washington, D.C. p. 932.

Andrews, D. F. 1973. "A General Method for the Approximation of Tail Areas." Annals of Statistics 1:367-372.

Anderson, G. L., and R. J. P. DeFigueiredo. 1980. "An Adaptive Orthogonal-Series Estimator for Probability Density Functions." Ann. Math. Statist. 51:347-376.

Bean, S. J., and C. P. Tsokos. 1980. "Developments in Nonparametric Density Estimation." Int'l. Statist. Review. 48:267-287.

Beard, R. E., T. Pentikäinen and E. Pesonen. 1969. Risk Theory. Methuen, London.

Bohman, H. 1963. "What is the Reason That Esscher's Method of Approximation Is As Good As It Is?" Skand. Aktu. Tidekr. 46:87-94.

Bohman, H., and F. Esscher. 1963-64. "Studies in Risk Theory With Numerical Illustrations Concerning Distribution Functions and Step-Loss Premiums." Skand. Aktu. Tidekr. 46:173-225; 47:1-40.

Bouver, H. and R. E. Bargmann. 1979. "Comparison of Computational Algorithms for the Evaluation of the Univariate and Bivariate Normal Distributions. In Proceedings of the Computer Science and Statistics 12th Annual Symposium on the Interface, ed. J. F. Gentleman, pp. 344-348. University of Waterloo, Waterloo, Ontario.

Daniels, H. E. 1954. "Saddlepoint Approximations in Statistics." Ann. Math. Statist. 25:631-650.

Esscher, F. 1932. "On the Probability Function in the Collective Theory of Risk." Skand. Aktu. Tidekr. 15:175-195.

Fellner, W. H., and M. E. Tarter. 1971. "Some New Results Concerning Density Estimates Based Upon Fourier Series." In Proceedings of Computer Science and Statistics: Fifth Annual Symposium on Interface, ed. M. O. Locks, pp. 54-64. Oklahoma State University, Stillwater, Oklahoma.

Gross, A. J., and D. W. Hosmer, Jr. 1978. "Approximating Tail Areas of Probability Distributions." Annals of Statistics 6:1352-1359.

Hill, G. W. 1970. "Algorithm 396 Student's t Quantiles." Communications of the ACM 13:619.

Hosmer, D. W. 1973. "On MLE of the Parameters of 'Two Normal Distributions when the Sample Size Is Small'." Communications in Statistics 1(3):217-227.

Johnston, N. L., and S. Kotz. 1970. Continuous Univariate Distributions-1. Distributions in Statistics. John Wiley & Sons, New York, New York.

Kronmal, R. A., and M. E. Tarter. 1968. "The Estimation of Probability Density and Cumulatives by Fourier Series Methods." J. of the Amer. Statist. Assoc. 63:925-952.

Kronmal, R. A., and M. E. Tarter. 1976. "An Introduction to the Implementation and Theory of Nonparametric Density Estimation." Amer. Stat. 30:105-112.

Marsaglia, S. 1976. "Random Number Generators." In The Encyclopedia of Computer Science, ed. Anthony Rolston, pp. 1192-1197. Petrocelli/Charter.

Parzen, E. 1962. "On Estimation of a Probability Density Function and Mode." Ann. Math. Statist. 33:1065-1076.

Silverman, B. W. 1978. "Choosing the Window When Estimating a Density." Biometrika 65(1):1-11.

Silverman, B. W. 1981. "Density Estimation for Univariate and Bivariate Data." In Interpreting Multivariate Data, ed. V. Barnett, Chapter 3. Wiley, Chichester.

Silverman, B. W. 1982. "Density Estimation of Orthogonal Series." Ann. Math. Statist. 40:1496-1488.

Tapia, R. A., and J. R. Thompson. 1978. Nonparametric Probability Density Estimation, pp. 59-60. Johns Hopkins University Press, Baltimore, Maryland.

Tarter, M. E. 1979. "Biocomputational Methodology, an Adjunct to Theory and Applications." Biometrika 35:9-24.

Tarter, M. E., R. L. Holcomb and R. A. Kronmal. 1967. "A Description of New Computer Methods for Estimating the Population Density." Proceedings of the 23rd National Conference of Association for Computing Machinery. Brandon Systems Press, Inc., Princeton, New Jersey.

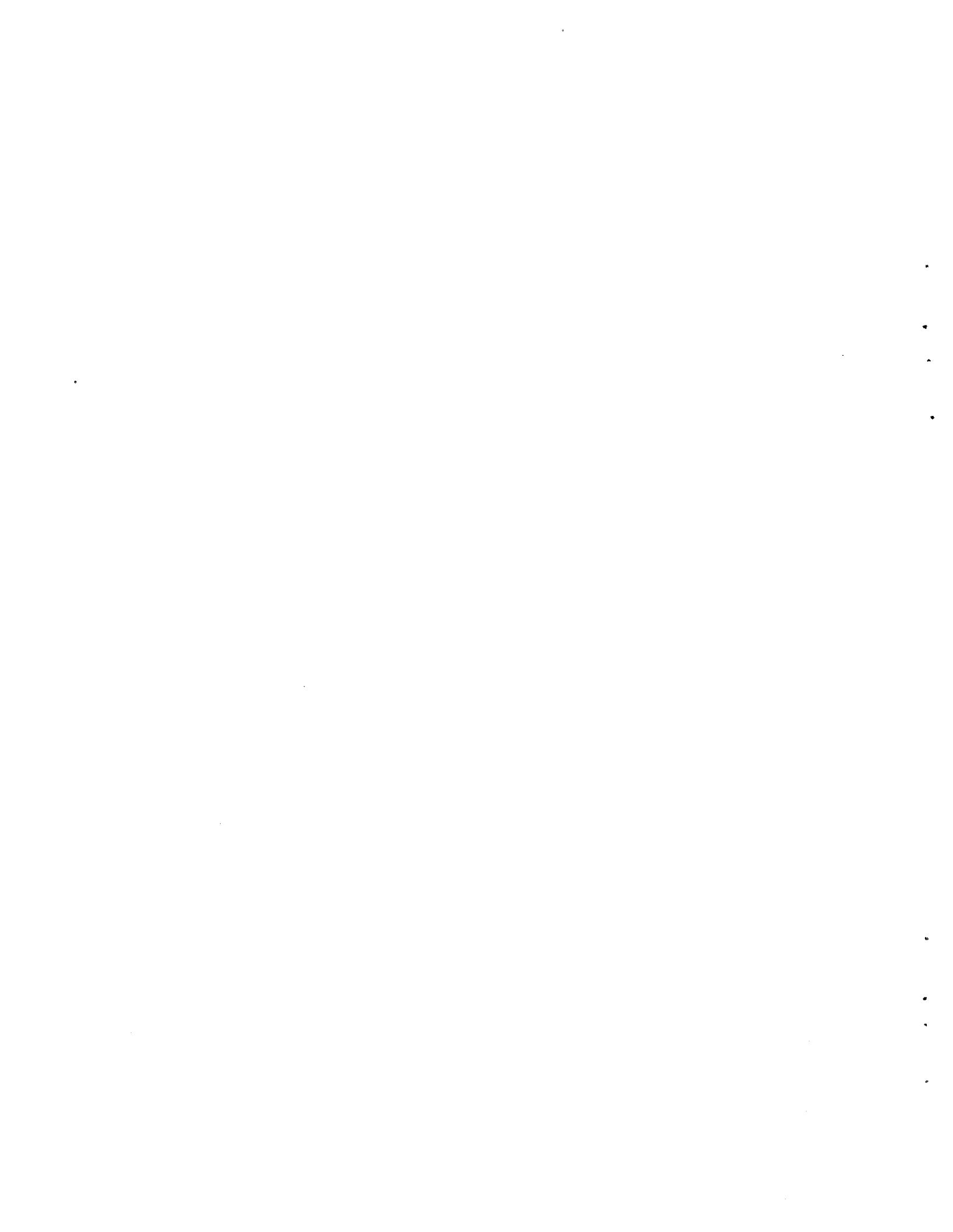
Tarter, M. E., and S. Raman. 1972. "A Systematic Approach to Graphical Methods in Biometry." In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4, pp. 192-221. University of California Press, Berkeley and Los Angeles, California.

Wahba, G. 1978. Data-Based Optimal Smoothing of Orthogonal Series Density Estimators. Dept. Statist. University of Wisconsin Report No. 509.

Watson, G. S. 1969. "Density Estimation of Orthogonal Series." Ann. Math. Statist. 40:1496-1488.

Wegman, E. J. 1972. "Nonparametric Probability Density Estimation. I. A Summary of Available Methods. Technometrics 14:533-546. "II. A Comparison of Density Estimation Methods." J. Statist. Comput. Simul. 1:225-245.

Wertz, W. 1978. Statistical Density Estimation. A Survey. Vandenhoeck and Ruprecht, Gottingen.



APPENDIX A

EVALUATING THE SMOOTHING PARAMETER $\beta(f)$ FOR A T DISTRIBUTION

APPENDIX A

EVALUATING THE SMOOTHING PARAMETER $\beta(f)$ FOR A T DISTRIBUTION

The purpose of this appendix is to show how

$$\beta(f) = \left(\frac{\int_{t_0}^{\infty} f(x) dx}{\int_{t_0}^{\infty} [f^{(2)}(x)]^2 dx} \right)^{1/5} \quad (A.1)$$

can be reduced to a form that can be numerically evaluated when $f(x)$ is the student t density. Expression (A.1) results when trying to determine the optimal smoothing parameter in Equation (3.11).

Let

$$T(t_0, r) = \int_{t_0}^{\infty} f(x) dx \quad (A.2)$$

where $f(x)$ has a t density with r degrees of freedom. T is readily evaluated using numerical methods so only the denominator in (A.1) need be considered. Now

$$f(x) = c(r) \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2} \quad (A.3)$$

$$f^{(1)}(x) = -c(r) \left(\frac{r+1}{2}\right) \left(1 + \frac{x^2}{r}\right)^{-(r+3)/2} \cdot 2\frac{x}{r} \quad (A.4)$$

and

$$f^{(2)}(x) = c(r) \frac{r+1}{r} \left(1 + \frac{x^2}{r}\right)^{-(r+5)/2} \left(\frac{r+2}{r} x^2 - 1\right) \quad (A.5)$$

where

$$c(r) = \left[\sqrt{r} \beta\left(\frac{1}{2}, \frac{r}{2}\right) \right]^{-1} \quad (A.6)$$

$c(r)$ involves the beta coefficient which is directly computable when r is integer valued.

Squaring and integrating (A.5) yields three terms denoted T_1 , T_2 , and T_3 .

$$T_1 = c^2(r) \left(\frac{r+1}{r}\right)^2 \int_{t_0}^{\infty} \left(1 + \frac{x^2}{r}\right)^{-(r+5)} dx \quad (A.7)$$

The integral has the form of the t so we have

$$T_1 = c^2(r) \left(\frac{r+1}{r}\right)^2 c^{-1}(2r+9) T(t_0, 2r+9) \quad (A.8)$$

The second term is given by

$$T_2 = C_2 \int_t^{\infty} \frac{x^2}{(1 + x^2/r)^{r+5}} dx, \quad (A.9)$$

where C_2 is defined as

$$C_2 = -2 \frac{r+2}{r} c^2(r) \left(\frac{r+1}{r}\right)^2. \quad (A.10)$$

Recall the following mathematical identity

$$\int \frac{x^m}{(1+x^2/r)^{n+1}} = \frac{-1}{2n-m+1} \frac{rx^{m-1}}{(1+x^2/r)^n} + \frac{r(m-1)}{2n-m+1} \int \frac{x^{m-2}}{(1+x^2/r)^{n+1}}. \quad (A.11)$$

Thus we find that

$$T_2 = C_2 \frac{r^{3/2}}{2r+7} \left[\frac{t_0/\sqrt{r}}{(1+t_0^2/r)^{r+4}} + r^{-1/2} c^{-1}(2r+9) T(t_0, 2r+9) \right]. \quad (A.12)$$

For the T_3 term, let

$$C_3 = \left(\frac{r+2}{r} \right)^2 c^2(r) \left(\frac{r+1}{r} \right)^2. \quad (A.13)$$

Then

$$T_3 = C_3 \int_{t_0}^{\infty} \frac{x^4}{(1+x^2/r)^{r+4+1}} dx = \frac{C_3}{2r+5} \left[\frac{t_0^3}{(1+t_0^2/r)^{r+4}} + 3 \int_{t_0}^{\infty} \frac{x^2}{(1+x^2/r)^{r+4+1}} dx \right]. \quad (A.14)$$

Applying (A.11) a second time yields

$$T_3 = \frac{C_3 r}{2r+5} \left[\frac{t_0^3}{(1+t_0^2/r)^{r+4}} + 3 \frac{r^{3/2}}{2r+7} \left[\frac{t_0/\sqrt{r}}{(1+t_0^2/r)^{r+4}} + r^{-1/2} c^{-1}(2r+9) T(t_0, 2r+9) \right] \right] \quad (A.15)$$

Thus if we let

$$K = r^{-1/2} c^{-1}(2r+9) T(t_0, 2r+9) \quad (A.16)$$

the denominator becomes

$$c^2(r) \left(\frac{r+1}{r}\right)^2 r^{1/2} \left[K - \frac{2(r+2)}{2r+7} \left(\frac{t_0/\sqrt{r}}{(1+t_0^2/r)^{r+4}} + K \right) + \frac{(r+2)^2}{2r+5} \left(\frac{t_0/\sqrt{r}^3}{(1+t_0^2/r)^{r+4}} + \frac{3}{2r+7} \left[\frac{t_0/\sqrt{r}}{(1+t_0^2/r)^{r+4}} + K \right] \right) \right] \quad (A.17)$$

Thus (A.1) can be readily evaluated for the t distribution.

APPENDIX B

EVALUATING THE SMOOTHING PARAMETER $\beta(f)$ FOR A MIXTURE OF NORMAL DENSITIES

APPENDIX B

EVALUATING THE SMOOTHING PARAMETER $\beta(f)$ FOR A MIXTURE OF NORMAL DENSITIES

The purpose of this appendix is to show how

$$\beta(f) = \left(\frac{\int_0^\infty f(x) dx}{\int_0^\infty |f^{(2)}(x)|^2 dx} \right)^{\frac{1}{2r+1}} \quad (8.1)$$

where $r = 2$ can be reduced to a form that can be numerically evaluated. Let f be given by

$$f(x) = p f_1(x) + (1 - p) f_2(x) \quad (B.2)$$

where f_i are normal densities with means μ_i and standard deviations σ_i . The numerator in (B.1) is simply evaluated using a Gaussian cdf algorithm. The denominator will be evaluated as follows.

First note that

$$f^{(1)}(x) = \frac{-p}{\sigma_1^2} (x - \mu_1) f_1(x) - \frac{(1 - p)}{\sigma_2^2} (x - \mu_2) f_2 \quad (B.3)$$

and

$$f^{(2)}(x) = \left[\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 1 \right] \frac{p f_1(x)}{\sigma_1^2} + \left[\left(\frac{x - \mu_2}{\sigma_2} \right)^2 - 1 \right] (1 - p) \frac{f_2(x)}{\sigma_2^2} \quad (B.4)$$

Squaring (B.4) and integrating requires the evaluation of three integrals designated here as T1, T2 and T3:

$$T_1 = \int_{t_0}^{\infty} \left[\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 1 \right]^2 \frac{p^2 f_1^2(x)}{\sigma_1^4} dx$$

$$T_2 = 2 \int_{t_0}^{\infty} \left[\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 1 \right] \left[\left(\frac{x - \mu_2}{\sigma_2} \right)^2 - 1 \right] \frac{p(1-p) f_1(x) f_2(x)}{\sigma_1^2 \sigma_2^2} dx$$

$$T_3 = \int_{t_0}^{\infty} \left[\left(\frac{x - \mu_2}{\sigma_2} \right)^2 - 1 \right]^2 \frac{(1-p)^2 f_2^2(x)}{\sigma_2^4} dx$$

Deriving the expressions for T1 and T3 are similar and are as follows:

$$T_1 = \frac{p^2}{\sigma_1^5} \frac{1}{2\pi} \int_{t_0}^{\infty} \left[\left(\frac{x - \mu_1}{\sigma_1} \right)^2 - 1 \right]^2 \frac{e^{-(x - \mu_1)^2/\sigma_1^2}}{\sigma_1} dx \quad (B.5)$$

Then changing variables using $z = [(x - \mu_1)/\sigma_1]^2$ yields

$$T_1 = \frac{p^2}{\sigma_1^5} \frac{1}{4\pi} \int_{\left(\frac{t_0 - \mu_1}{\sigma_1}\right)^2}^{\infty} (z^{3/2} - 2z^{1/2} + z^{-1/2}) e^{-z} dz \quad (B.6)$$

For this study t_0 is positive so the transformation is 1 to 1. Then if $G(\alpha, x)$ denotes the incomplete gamma integral, T1 can be found using numerical algorithms from

$$\tau_1 = \frac{p^2}{\sigma_1^5} \frac{1}{4\pi} \left\{ \Gamma\left(\frac{5}{2}\right) \left[1 - G\left(\frac{5}{2}, t^*\right) \right] - 2 \Gamma\left(\frac{3}{2}\right) \left[1 - G\left(\frac{3}{2}, t^*\right) \right] + \Gamma\left(\frac{1}{2}\right) \left[1 - G\left(\frac{1}{2}, t^*\right) \right] \right\} \quad (B.7)$$

$$\text{where } t^* = [(t_0 - \mu_1)/\sigma_1]^2$$

The third term follows immediately by replacing p by $1 - p$ and the subscript 1 by the subscript 2.

The cross product term is

$$\begin{aligned} \tau_2 = \frac{2p(1-p)}{2\pi \sigma_2^3 \sigma_2^3} \int_0^\infty & \left[1 - \left(\frac{x - \mu_1}{\sigma_1} \right)^2 - \left(\frac{x - \mu_2}{\sigma_2} \right)^2 + \left(\frac{x - \mu_1}{\sigma_1} \right)^2 \left(\frac{x - \mu_2}{\sigma_2} \right)^2 \right] \\ & \exp \left[-\frac{1}{2} \sum_{i=1}^2 \left(\frac{x - \mu_i}{\sigma_i} \right)^2 \right] dx \end{aligned} \quad (B.8)$$

If the exponent currently has form $-(ax^2 + 2bx + c)$ we complete the square to get a form $-(ax^2 + 2\sqrt{a}dx + d^2 + c - d^2)$.

Applying this to the exponential term in (B.8) we get

$$a = \frac{1}{2} \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) \quad (B.9)$$

$$b = -\frac{1}{2} \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right) \quad (B.10)$$

$$c = \frac{1}{2} \left(\frac{\mu_1^2}{\sigma_1^2} + \frac{\mu_2^2}{\sigma_2^2} \right) \quad (B.11)$$

$$d = -\frac{1}{2} \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right) \left/ \sqrt{\frac{1}{2} \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)} \right. \quad (B.12)$$

Then transforming with $y = \sqrt{a}x + d$ yields

$$\tau_2 = \frac{e^{-(c-d)^2} p(1-p)}{\sqrt{a} \pi \sigma_1^3 \sigma_2^3} \int_{\sqrt{a}t_0+d}^{\infty} \sum_{i=0}^4 g_i y^i e^{-y^2} dy \quad (B.13)$$

Postponing for the moment the determination of constants g_i , making a change of variables in (B.13) of $z = y^2$ yields

$$\tau_2 = \frac{e^{-(c-d)^2} p(1-p)}{\sqrt{a} 2\pi \sigma_1^3 \sigma_2^3} \left(\sqrt{a} t_0 + d \right)^2 \sum_{i=0}^4 g_i z^{i/2-1/2} e^{-z} dz \quad (8.14)$$

$$= \frac{e^{-(c-d)^2} p(1-p)}{\sqrt{a} 2\pi \sigma_1^3 \sigma_2^3} \sum_{i=0}^4 g_i \Gamma\left(\frac{i}{2} + \frac{1}{2}\right) \left(1 - G\left[\frac{i}{2} + \frac{1}{2}, (\sqrt{a} t_0 + d)^2\right]\right) \quad (B.15)$$

which is readily evaluated using numerical algorithms. The transformation from (B.13) to (B.14) is 1 to 1 provided $\sqrt{a} t_0 + d > 0$. In terms of the parameters of our model, this is the same as requiring

$$t_0 > \frac{1}{2} \left(\frac{\mu_1}{\sigma_1^2} + \frac{\mu_2}{\sigma_2^2} \right) \left/ \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right) \right. \quad (B.16)$$

For the Monte Carlo study $\mu_1 = 0$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 0.25$ so that we must have

$$t_0 > 0.4 \mu_2$$

This is satisfied for the μ_2 's and critical points investigated. To find the constants g_i , expressions are simplified by letting

$$\begin{aligned}
 s_1 &= \sqrt{a} \sigma_1 \\
 s_2 &= \sqrt{a} \sigma_2 \\
 c_1^* &= (\mu_1 + d/\sqrt{a})/\sigma_1 \\
 c_2^* &= (\mu_2 + d/\sqrt{a})/\sigma_2
 \end{aligned} \tag{B.17}$$

Thus when $x = (y - d)/\sqrt{a}$, $(x - \mu_i)/\sigma_i$ becomes $y/s_i - c_i$.

Table B.1 shows the coefficients that sum into g_i .

TABLE B.1. Selected Coefficients in Finding the Mixture of Normal Smoothing Parameter

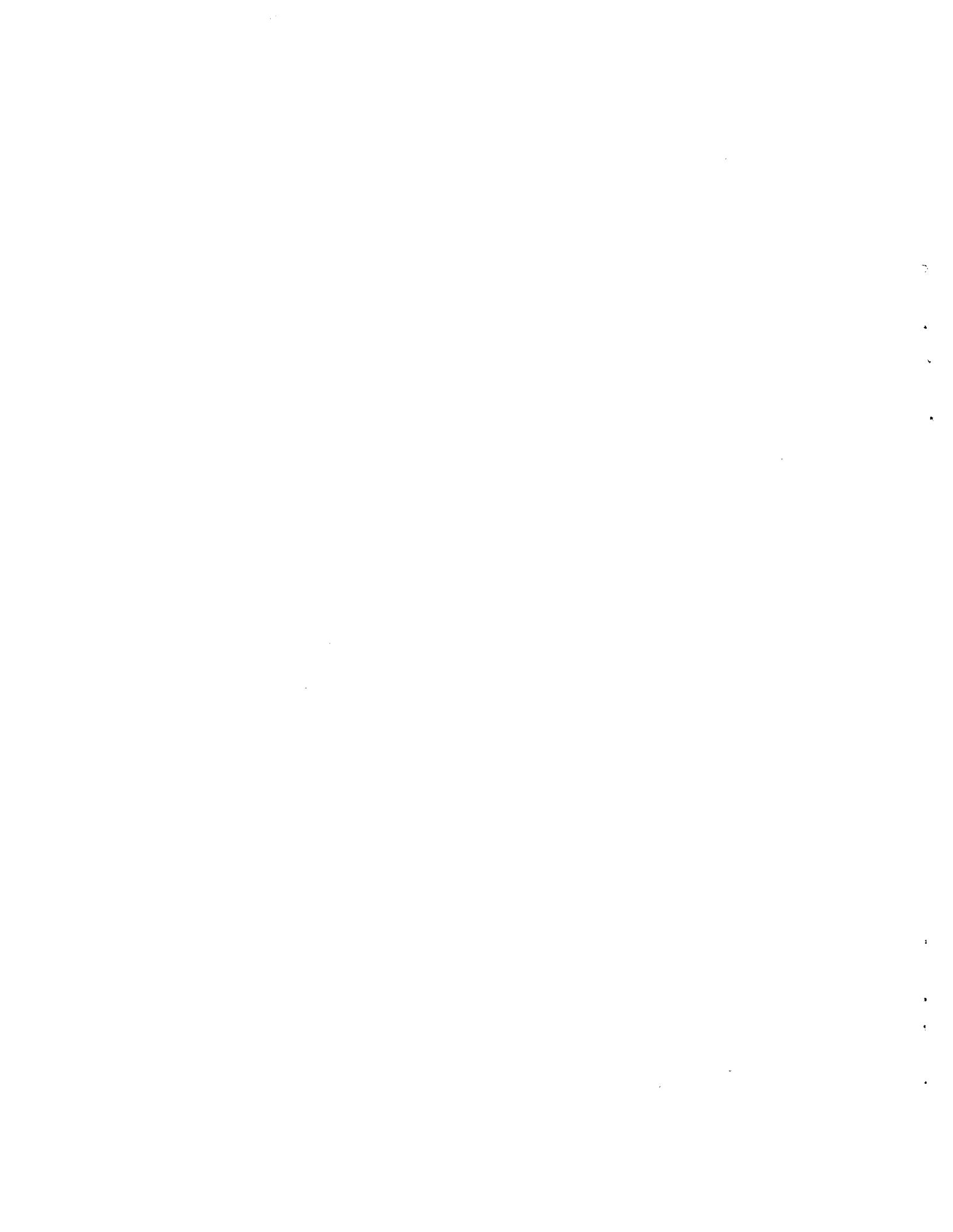
Numbered Terms	Powers of γ				
	γ^4	γ^3	γ^2	γ	1
1					1
2			$-\frac{1}{s_1^2}$	$+2 \frac{c_1}{s_1}$	$-c_1^2$
3			$-\frac{1}{s_2^2}$	$+2 \frac{c_2}{s_2}$	$-c_2^2$
4	$\frac{1}{s_1^2 s_2^2}$	$\frac{-2}{s_1 s_2} \left[\frac{c_2}{s_1} + \frac{c_1}{s_2} \right]$	$\frac{c_1^2}{s_2^2} + \frac{c_2^2}{s_1^2}$	$-2 c_1^2 \frac{c_2}{s_2}$	$+c_1^2 c_2^2$
4			$4 \frac{c_1 c_2}{s_1 s_2}$	$-2 c_2^2 \frac{c_1}{s_1}$	

Summing the coefficients in Table B.1 yields:

$$\begin{aligned}
 g_0 &= 1 - c_1^2 - c_2^2 + c_1^2 c_2^2 \\
 g_1 &= 2 \left[\frac{c_2}{s_2} (1 - c_1^2) + \frac{c_1}{s_1} (1 - c_2^2) \right] \\
 g_2 &= \frac{c_1^2 - 1}{s_2^2} + \frac{c_2^2 - 1}{s_1^2} + \frac{4 c_1 c_2}{s_1 s_2} \\
 g_3 &= \frac{-2}{s_1 s_2} \left(\frac{c_2}{s_1} + \frac{c_1}{s_2} \right) \\
 g_4 &= \frac{1}{s_1^2 s_2^2}
 \end{aligned} \tag{B.18}$$

Thus $B(f)$ can be found from

$$B(f) = \left[\frac{\int_0^\infty f(x) dx}{T_1 + T_2 + T_3} \right]^{0.2} \tag{B.19}$$



DISTRIBUTION

No. of
Copies

OFFSITE

D. M. Austin
DOE Division of Engineering,
Mathematical and Geosciences
Washington, DC 20545

27 DOE Technical Information Center

D. Bates
University of Wisconsin
Statistics Department
1210 W. Dayton Street
Madison, WI 53706

D. J. DeWitt
Computer Sciences Department
University of Wisconsin
Madison, WI 53706

W. Dixon
Department of Mathematics
UCLA Medical Center
Los Angeles, CA 90024

J. H. Friedman
Stanford Linear Accelerator
Center
P.O. Box 4349
Stanford, CA 94305

J. Foley
George Washington University
School of Engineer & Applied
Science
Washington, DC 20052

J. Heller
Department of Computer Science
SUNY at Stony Brook
Stony Brook, NY 11794

B. J. Joiner
Statistics Department
University of Wisconsin
Madison, WI 53706

No. of
Copies

V. E. Kane
Mathematics & Statistics
Research Department
Union Carbide Corp.--
Nuclear Division
P.O. Box Y
Oak Ridge, TN 37830

A. Marcus
Department of Pure and
Applied Mathematics
Washington State University
Pullman, WA 99163

H. Marks
U.S. Department of Agriculture
300 12th St. SW, Room 612
Washington, DC 20250

R. W. Mensing
Numerical Mathematics Group
Lawrence Livermore Laboratory
P.O. Box 808
Livermore, CA 94550

M. Pagano
Harvard School of Public Health
Department of Biostatistics
677 Huntington Avenue
Boston, MA 02115

E. Parzen
Institute of Statistics
Texas A&M University
College Station, TX 77843

D. W. Scott
Department of Mathematical
Sciences
Rice University
Houston, TX 77001

No. of
Copies

A. Shoshani
Computer Science & Applied
Mathematics Department
Lawrence Berkeley Laboratory
Berkeley, CA 94720

D. L. Wallace
Department of Statistics
University of Chicago
1118 E. 58th Street
Chicago, IL 60637

R. A. Waller
Group Q-12, MS-608
Los Alamos National Laboratory
Los Alamos, NM 87545

H. Wong
Lawrence Berkeley National
Laboratory
University of California
Berkeley, CA 94720

No. of
Copies

ONSITE

DOE Richland Operations Office

H. E. Ransom

47 Pacific Northwest Laboratory

R. A. Burnett
D. B. Carr (22)
K. R. Chase
M. D. Erickson
D. L. Hall (3)
P. G. Heasler
R. E. Kleinknecht
R. J. Littlefield
W. L. Nicholson
A. R. Olsen
J. J. Thomas
H. D. Tolley (8)
Publishing Coordination (2)
Technical Information (3)