2000

Likelihood-based inference in some partially non-regular exponential families

Thomas Michael Dubinin
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/13895
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
NOTE TO USERS

This reproduction is the best copy available.

UMI
Likelihood-based inference in some partially non-regular exponential families

by

Thomas Michael Dubinin

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Major Professor: Stephen B. Vardeman

Iowa State University

Ames, Iowa

2000
Graduate College
Iowa State University

This is to certify that the Doctoral dissertation of

Thomas Michael Dubinin

has met the dissertation requirements of Iowa State University

Signatures have been redacted for privacy.
TABLE OF CONTENTS

1 THE MOTIVATING PROBLEM .................................. 1
1.1 Introduction ............................................. 1

2 LITERATURE REVIEW ....................................... 8
2.1 Numerics of Maximum Likelihood Estimation .............. 8
   2.1.1 The Newton-Raphson Method for Function Maximization .... 9
   2.1.2 Cohen's Method for IID Truncated Normal Data ............ 10
   2.1.3 Cohen's Method for IID Left-Censored Normal Data ...... 12
2.2 Asymptotics of Likelihood-Based Inference in "Regular" Problems ... 14
2.3 Asymptotic Distributions of Extreme Order Statistics ........ 15
2.4 Smith (1985) on Maximum Likelihood Estimation in Non-Regular Problems 18
2.5 Chow and Teugels (1978) on Joint Asymptotics for Extreme Order Statistics and Sample Means .......... 19

3 INFERENCE IN TRUNCATED EXPONENTIAL FAMILIES (WITH UNKNOWN TRUNCATION PARAMETER) ..................... 24
3.1 Exponential Families .................................... 25
3.2 Notation ............................................... 26
3.3 Truncated Exponential Families ................................ 27
3.4 Moments in a Truncated Exponential Family .............. 28
3.5 Likelihood-Based Estimation ................................ 35
3.6 Inference in a Truncated Exponential Family .............. 36


LIST OF TABLES

Table 1.1  Data from Tomato Seed Experiment  

Table 3.1  The behavior of \( s^2/(\bar{x} - d)^2 \) based on 500 samples of size 100  

Table 3.2  Summary statistics for the null distribution of terms in the Taylor expansion of the LRT statistic for \( H_0 : (\mu, \sigma, d) = (0, 1, -1) \)  

Table 3.3  Statistics for checking normality of maximum likelihood estimators for \( \mu \) and \( \sigma \)  

Table 3.4  Estimated correlations between sample moments and between maximum likelihood estimators of \( \mu \) and \( \sigma \)  

Table 4.1  Statistics for checking normality of maximum likelihood estimators for \( \mu \) and \( \sigma \)  

Table 4.2  Summary statistics for the null distribution of terms in the Taylor expansion of the LRT statistic for \( H_0 : (\mu, \sigma, d) = (0, 1, -1) \)  

Table 5.1  Statistics for checking normality of the maximum likelihood estimators for \( \mu, \sigma, \) and \( p \)  

Table 5.2  Correlations (with \( p \)-values) between parameter estimators
# LIST OF FIGURES

| Figure 1.1 | Histogram of the 198 bacterial canker log counts | 3 |
| Figure 1.2 | Normal plot of the log (non-zero) bacterial counts | 4 |
| Figure 3.1 | Illustration of Equations (3.12) and (3.13) | 39 |
| Figure 3.2 | Illustration of relationship between \((\hat{d}, \hat{\eta}), (d^*, \eta^*)\), and \((d_0, \eta_0)\) | 44 |
| Figure 3.3 | A plot of \(\lambda(\xi)\) and its first derivative | 51 |
| Figure 3.4 | A log-log plot of \(\theta(g)\) vs \(g\) | 52 |
| Figure 3.5 | The inverse of \(\lambda\) and the fifth degree polynomial generated by fitting points \((g, \lambda^{-1}(g))\) for \(g \in \{0.01, 0.02, \ldots, 0.99\}\) | 53 |
| Figure 3.6 | A typical likelihood-based confidence region for \((\mu, d)\) based on \(n=25\) | 58 |
| Figure 3.7 | The likelihood ratio test statistic vs. the sum of the asymptotically non-negligible terms of the Taylor series for sample sizes \(n=30, 100, 1000,\) and \(10000\) | 63 |
| Figure 3.8 | The sampling distribution of \(\tilde{\mu}\) for \(n=30, 100, 1000,\) and \(10000\) | 64 |
| Figure 3.9 | The sampling distribution of \(\tilde{\sigma}\) for \(n=30, 100, 1000,\) and \(10000\) | 65 |
| Figure 3.10 | Side-by-side \(\chi^2\) plots of \(\hat{d}\) and the first order Taylor term for \(d\), \[ (\hat{d} - d) \sum_{i=1}^{n} \left. \frac{\partial}{\partial d} \log f_{n,d} \right|_{(\hat{\eta}, \hat{d})} \) for \(n=30, 100, 1000,\) and \(10000\) | 66 |
Figure 3.11 \( \chi^2 \) plots, each showing the likelihood ratio test statistic from 500 simulations where the data producing the statistic is generated from a truncated standard normal with truncation parameter \(-1\) and for sample sizes \( n=30, 100, 1000, \) and \( 10000 \) 

Figure 4.1 A 3D plot of \( \theta(x, \hat{p}) \) versus its two arguments 

Figure 4.2 The sampling distribution of \( \hat{\mu} \) for \( n=30, 100, 1000, 10000 \) 

Figure 4.3 The sampling distribution of \( \hat{\sigma} \) for \( n=30, 100, 1000, 10000 \) 

Figure 4.4 Side-by-side \( \chi^2 \) plots of \( \hat{d} \) and the first order Taylor term for \( d, \) 
\[
\left( \hat{d} - d \right) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log f_{\hat{d}}(x_i) \bigg|_{\hat{d}} \] for \( n=30, 100, 1000, 10000 \) 

Figure 4.5 The likelihood ratio test statistic vs. the sum of the asymptotically non-negligible terms of the Taylor series for sample sizes \( n=30, 100, 1000, \) and \( 10000 \) 

Figure 4.6 \( \chi^2 \) Plots, each showing the likelihood ratio test statistic from 500 simulations where the data producing the statistic is randomly generated from a censored standard normal distribution with threshold parameter \(-1\), for sample sizes \( n=30, 100, 1000, \) and \( 10000 \) 

Figure 5.1 The sampling distribution of \( \hat{\mu} \) for \( n=30, 100, 1000, 10000 \) 

Figure 5.2 The sampling distribution of \( \hat{\sigma} \) for \( n=30, 100, 1000, 10000 \) 

Figure 5.3 The sampling distribution of \( \hat{p} \) for \( n=30, 100, 1000, 10000 \) 

Figure 5.4 3D plots of \( \hat{p} \) versus \( (\hat{\mu}, \hat{\sigma}) \) and scatter plots of \( \hat{d} \) versus \( \hat{\mu}, \hat{\sigma} \), and \( \hat{p} \) for \( n=30,100,1000,10000 \)
Figure 5.5 $\chi^2$ Plots, each showing the likelihood ratio test Statistic from 500 simulations where the data producing the statistic is randomly generated from a censored standard normal with extra mass and with threshold parameter $-1$ and for sample sizes $n=30, 100, 1000,$ and $10000$ ................................. 111

Figure 5.6 The likelihood ratio test statistic vs. the sum of the asymptotically non-negligible terms of the Taylor series for sample sizes $n=30, 100, 1000,$ and $10000$ ................................. 112
1 THE MOTIVATING PROBLEM

1.1 Introduction

Bacterial canker is an important problem for agribusinesses who buy tomato seed for planting in contract fields. A means of checking for canker in an incoming lot of seed is to select some portion, grind it up, and perform a lab analysis by counting the number of bacteria colonies in the sample.

Logarithms of colony counts look reasonably normal except for two features. First, there are often zero counts that appear as though they result from the inability of the physical analysis to detect bacteria below a certain threshold. Second, it often is the case that the number of zero-counts exceeds what one would expect from the censored left tail of a normal distribution. Information on 198 samples of tomato seeds infested by bacterial canker was collected by the Plant Pathology Department at Iowa State University over a six month period in 1995. Table 1.1 containing the data and a histogram (Figure 1.1) follow.

One plausible model for this behavior is that the distribution generating an individual count is a mixture of a point mass at zero and a left-censored log-normal distribution. This could represent counts that appear as “detects” only when the equipment does not malfunction, and when the specimen itself contains sufficient number of bacteria to be detectable. The dual possibility of getting a zero-count/non-detect creates a partially continuous, partially discrete distribution with “extra mass” at zero. This is, of course,
Table 1.1 Data from Tomato Seed Experiment

<table>
<thead>
<tr>
<th>COUNT</th>
<th>Frequency</th>
<th>COUNT</th>
<th>Frequency</th>
<th>COUNT</th>
<th>Frequency</th>
<th>COUNT</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>54</td>
<td>450</td>
<td>3</td>
<td>160</td>
<td>1</td>
<td>1280</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>470</td>
<td>1</td>
<td>170</td>
<td>2</td>
<td>1660</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>510</td>
<td>1</td>
<td>180</td>
<td>4</td>
<td>1800</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>9</td>
<td>520</td>
<td>2</td>
<td>190</td>
<td>1</td>
<td>1840</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>540</td>
<td>1</td>
<td>210</td>
<td>1</td>
<td>1880</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>550</td>
<td>1</td>
<td>220</td>
<td>6</td>
<td>1910</td>
<td>1</td>
</tr>
<tr>
<td>50</td>
<td>4</td>
<td>570</td>
<td>1</td>
<td>230</td>
<td>1</td>
<td>1930</td>
<td>1</td>
</tr>
<tr>
<td>60</td>
<td>3</td>
<td>610</td>
<td>1</td>
<td>240</td>
<td>2</td>
<td>1950</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
<td>620</td>
<td>1</td>
<td>270</td>
<td>1</td>
<td>2100</td>
<td>1</td>
</tr>
<tr>
<td>80</td>
<td>1</td>
<td>680</td>
<td>1</td>
<td>290</td>
<td>2</td>
<td>2250</td>
<td>1</td>
</tr>
<tr>
<td>90</td>
<td>5</td>
<td>690</td>
<td>1</td>
<td>300</td>
<td>1</td>
<td>2300</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>1</td>
<td>760</td>
<td>1</td>
<td>310</td>
<td>2</td>
<td>2500</td>
<td>1</td>
</tr>
<tr>
<td>110</td>
<td>4</td>
<td>820</td>
<td>1</td>
<td>330</td>
<td>1</td>
<td>2650</td>
<td>1</td>
</tr>
<tr>
<td>118</td>
<td>1</td>
<td>900</td>
<td>2</td>
<td>340</td>
<td>2</td>
<td>2700</td>
<td>1</td>
</tr>
<tr>
<td>120</td>
<td>6</td>
<td>950</td>
<td>1</td>
<td>360</td>
<td>1</td>
<td>3300</td>
<td>1</td>
</tr>
<tr>
<td>130</td>
<td>1</td>
<td>1030</td>
<td>1</td>
<td>390</td>
<td>1</td>
<td>3400</td>
<td>1</td>
</tr>
<tr>
<td>140</td>
<td>1</td>
<td>1050</td>
<td>1</td>
<td>410</td>
<td>1</td>
<td>3800</td>
<td>1</td>
</tr>
<tr>
<td>150</td>
<td>2</td>
<td>1170</td>
<td>1</td>
<td>420</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

because we cannot distinguish among the zero counts as to whether they were recorded as zero because of equipment malfunction or because of the size of the bacteria colony.

We present a formal probability model that allows for the circumstances outlined above. Suppose that, with probability $p$, an analysis, for some reason unrelated to the actual number of bacteria present, produces a zero-count. With probability $1 - p$, a bacteria count of $X^*$ potentially appears where $\ln(X^*) \equiv X$ is normal with mean $\mu$ and standard deviation $\sigma$. However, for a constant $d$, if $X^* < \exp(d)$ we observe a zero-count rather than $X^*$ because of limitations of the detection equipment.

Mathematically we have that $x_1^*, x_2^*, \ldots, x_n^*$ are iid from a mixed point mass at zero and a left-censored lognormal distribution. If we let $X \equiv \ln(X^*)$ (understanding the logarithm of zero to be $-\infty$), a density for our model with respect to the sum of a
Histogram of the logarithm of Bacteria Counts

Figure 1.1 Histogram of the 198 bacterial canker log counts

unit point mass at $-\infty$ and Lebesgue measure on $\mathbb{R}$ is

$$f(x|p, d, \mu, \sigma) = \begin{cases} 
  p + (1 - p)\Phi \left( \frac{d - \mu}{\sigma} \right) & x = -\infty \\
  \frac{(1-p)}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2\sigma^2} (x - \mu)^2 \right] & x \geq d 
\end{cases}$$

$$= \left\{ p + (1 - p)\Phi \left( \frac{d - \mu}{\sigma} \right) \right\}^{1_{\{x=-\infty\}}} \left\{ \frac{(1-p)}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2\sigma^2} (x - \mu)^2 \right] \right\}^{1_{\{x\geq d\}}}.$$

We can justify, at least qualitatively, the use of the normal distribution here by looking at a normal probability plot of the non-zero log counts in Table 1.1. Maximum likelihood here suggests that $198p \approx 32$, so with ordered (log) non-zero counts,

$$y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(144)}$$
Figure 1.2 Normal plot of the log (non-zero) bacterial counts

Figure 1.2 consists of plotted points

\[
\left( \Phi^{-1} \left( \frac{i + 21.5}{166} \right), y(i) \right).
\]

The plot in Figure 1.2 is quite linear which seems to indicate our model is, at least, somewhat plausible.

If \(x_1, x_2, \ldots, x_n\) is a random sample from the distribution specified by the above density, the joint density of the vector \((x_1, x_2, \ldots, x_n)\) is simply the \(n\)-fold product of the marginal densities. In this present situation let:

\(n_{-\infty}\) the number of non-detects in the sample of size \(n\),

\(n_{\mathbb{R}}\) the number of 'actual' measurements in the sample of size \(n\),
\( p \) the probability of a non-detect resulting from equipment malfunction,
\( d \) the threshold of detection (smallest possible value that can be detected),
\( \mu \) the mean for the population of log-counts (including those less than \( d \)),
\( \sigma \) the standard deviation for the population of log-counts (including those less than \( d \)),
\( \theta \) the vector \((p, d, \mu, \sigma)\).

Then we write
\[
\begin{align*}
  f(x|\theta) &= \prod_{i=1}^{n} \left\{ p + (1-p) \Phi \left( \frac{d-\mu}{\sigma} \right) \right\}^{1_{\{x_i=-\infty\}}} \left\{ \prod_{\{x_i \geq d\}} \left\{ \frac{(1-p)}{\sqrt{2\pi\sigma}} \exp \left( \frac{-1}{2\sigma^2}(x_i - \mu)^2 \right) \right\} \right\} \\
  &= \left\{ p + (1-p) \Phi \left( \frac{d-\mu}{\sigma} \right) \right\} \prod_{\{x_i \geq d\}} \left\{ \frac{(1-p)}{\sqrt{2\pi\sigma}} \exp \left( \frac{-1}{2\sigma^2}(x_i - \mu)^2 \right) \right\}. \quad (1.1)
\end{align*}
\]

The function \((1.1)\) of \( \mu, \sigma, d, \) and \( p \) is the joint likelihood for this problem. It is generally easier to work with the logarithm of the likelihood rather than the likelihood itself. To that end we note that
\[
\begin{align*}
  \sum_{i=1}^{n} \log f(x_i|p, d, \mu, \sigma) &= \log \left\{ p + (1-p) \Phi \left( \frac{d-\mu}{\sigma} \right) \right\} \prod_{\{x_i \geq d\}} \left\{ \frac{(1-p)}{\sqrt{2\pi\sigma}} \exp \left( \frac{-1}{2\sigma^2}(x_i - \mu)^2 \right) \right\} \\
  &= n_{-\infty} \log \left[ p + (1-p) \Phi \left( \frac{d-\mu}{\sigma} \right) \right] + n_{\infty} \left( \log(1-p) + \log(\sigma) - \frac{1}{2} \log(2\pi) \right) \\
  &\quad - \frac{1}{2\sigma^2} (n_{\infty}(\bar{x} - \mu)^2 + (n_{\infty} - 1)s^2), \quad (1.2)
\end{align*}
\]

where in expression \((1.2)\) we employ the abbreviations
\[
\begin{align*}
  \sum_{i=1}^{n_{\infty}} \left( \frac{x_i - \mu}{\sqrt{2\sigma}} \right)^2 &= \frac{1}{2\sigma^2} \sum_{i=1}^{n_{\infty}} (x_i^2 - 2x_i\mu + \mu^2) \\
  &= \frac{1}{2\sigma^2} \left\{ \left( \sum_{i=1}^{n_{\infty}} x_i^2 - n_{\infty}\bar{x}^2 \right) + (n_{\infty}\bar{x}^2 - 2n_{\infty}\bar{x}\mu + n_{\infty}\mu^2) \right\} \\
  &= \frac{1}{2\sigma^2} \left( (n_{\infty} - 1)s^2 + n_{\infty}(\bar{x} - \mu)^2 \right)
\end{align*}
\]
for
\[ \bar{x} = \left( \frac{1}{n_R} \right) \sum_{\{x_i > -\infty\}} x_i \quad \text{and} \quad s^2 = \left( \frac{1}{n_R - 1} \right) \sum_{\{x_i > -\infty\}} (x_i - \bar{x})^2. \]

We approach inference for the motivating problem from the likelihood perspective. Non-standard issues arise when using this approach. Chief among these is how to handle the non-regularity introduced into the problem by the unknown censoring point. Deriving a point estimate for \( d \) is quite trivial. However, inference by the way of confidence intervals or tests of significance is not as easy. And the impact on likelihood-based inference for the whole vector \((p, d, \mu, \sigma)\) of the non-regularity introduced into the problem by \( d \) is not clear.

For a fixed \( d \), the vector of maximum likelihood estimators \((\hat{p}, \hat{\mu}, \hat{\sigma})\), is asymptotically multivariate normal and likelihood ratio test statistics are asymptotically \( \chi^2 \) under null hypotheses. This follows from standard regularity conditions that include three-times differentiability of the density in the parameters for each \( x \).

However, when \( d \) is treated as an unknown parameter, differentiability of the density in \( d \) for fixed \( x \) will fail to hold. In fact it is clear that "standard" results like the asymptotic (joint) normality of the maximum likelihood estimator must fail. This is because \( \hat{d} \) takes the form of the minimum "realized" log count and extreme order statistics are never asymptotically normal. Perhaps we can study \( \hat{d} \) in isolation of \((\hat{p}, \hat{\mu}, \hat{\sigma})\), but to do this requires asymptotic independence between \( \hat{d} \) and \((\hat{p}, \hat{\mu}, \hat{\sigma})\).

In what follows we demonstrate under fairly general circumstances that the sample minimum and the other entries of a vector of maximum likelihood estimators are asymptotically independent. In addition, the impact of this fact on likelihood-based confidence regions and likelihood ratio tests for the combined vector is demonstrated. We show these in various simplifications and modifications of the motivating problem. We will first consider the case of a truncation parameter and a vector of "regular" parameters. We will next consider the problem of a single unknown censoring parameter.
and a vector of "regular" parameters. Finally, we return to the analysis of problems like the motivating problem, that involve a single unknown censoring parameter, a vector of "regular" parameters and "extra mass at $-\infty$".
2 LITERATURE REVIEW

2.1 Numerics of Maximum Likelihood Estimation

In the motivating problem introduced in Chapter 1, since $\Phi(\cdot)$ is monotone increasing and $d < x_i$ for all $i$ corresponding to uncensored observations, it is clear that for any $(p, \mu, \sigma)$ the log-likelihood (1.2) is maximized as a function of $d$ by

$$\hat{d} = \min\{x_i | x_i > -\infty\}.$$

Subsequently finding values $(p, \mu, \sigma)$ that maximize $f(x|p, \hat{d}, \mu, \sigma)$ is not simple. The set of equations produced by setting first partials equal to zero is not solvable in closed form. There are, however, several iterative methods of solution that may be employed.

The approaches vary in terms of the class of functions to which they may be applied, programming difficulty, and the computer time typically required for convergence. When working with a single data set to produce the maximum likelihood estimates corresponding to only that set of data, computing time and, to some extent, programming difficulties are of little importance. However, for purposes of implementing a procedure in a simulation where there are potentially any number of data sets, programming considerations become paramount. For problems related to the motivating example presented in Chapter 1, a modified version of Newton-Raphson has proved to be a good choice. "Modified" here is meant to describe a procedure that is Newton-Raphson in some subset of the parameter space, but, in part, relies on the relationships between the parameters to improve an otherwise blind process.
2.1.1 The Newton-Raphson Method for Function Maximization

The Newton-Raphson method is an explicitly iterative method of maximizing a function. The theoretical requirements of the procedure are that the function being maximized is twice differentiable in the variables of interest. For a likelihood with \( k \) parameters, one simply forms a system of \( k \) equations by setting first partials with respect to single parameters equal to zero. At each iteration, a current approximate maximizer (parameter vector) is updated by adding to it a value computed as a function of the Hessian matrix evaluated at the current parameter vector. Specifically, for independent and identically distributed data, if we define

\[
L(\theta|x) \equiv \sum \log f(x_i|\theta) = \text{the log-likelihood}
\]

\[
\theta_i \equiv (\theta_{1,i}, \theta_{2,i}, \ldots, \theta_{n,i}) = \text{a vector of unknown parameters at iteration } i
\]

\[
\Delta^{(i)} = \text{a step length at iteration } i
\]

then we compute the approximate maximizer at iteration \( i + 1 \) as

\[
\begin{pmatrix}
\theta_{1,i+1} \\
\theta_{2,i+1} \\
\vdots \\
\theta_{k,i+1}
\end{pmatrix}
= \begin{pmatrix}
\theta_{1,i} \\
\theta_{2,i} \\
\vdots \\
\theta_{k,i}
\end{pmatrix}
- \Delta^{(i)}
\begin{bmatrix}
\frac{\partial L}{\partial \theta_1} |\theta_i \\
\frac{\partial L}{\partial \theta_2} |\theta_i \\
\vdots \\
\frac{\partial L}{\partial \theta_k} |\theta_i
\end{bmatrix}
\begin{bmatrix}
\frac{\partial^2 L}{\partial \theta_1^2} |\theta_i & \frac{\partial^2 L}{\partial \theta_1 \theta_2} |\theta_i & \cdots & \frac{\partial^2 L}{\partial \theta_1 \theta_k} |\theta_i \\
\frac{\partial^2 L}{\partial \theta_2 \theta_1} |\theta_i & \frac{\partial^2 L}{\partial \theta_2^2} |\theta_i & \cdots & \frac{\partial^2 L}{\partial \theta_2 \theta_k} |\theta_i \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 L}{\partial \theta_k \theta_1} |\theta_i & \frac{\partial^2 L}{\partial \theta_k \theta_2} |\theta_i & \cdots & \frac{\partial^2 L}{\partial \theta_k^2} |\theta_i
\end{bmatrix}^{-1}
\begin{bmatrix}
\frac{\partial^2 L}{\partial \theta_1 \theta_1} |\theta_i & \frac{\partial^2 L}{\partial \theta_1 \theta_2} |\theta_i & \cdots & \frac{\partial^2 L}{\partial \theta_1 \theta_k} |\theta_i \\
\frac{\partial^2 L}{\partial \theta_2 \theta_1} |\theta_i & \frac{\partial^2 L}{\partial \theta_2 \theta_2} |\theta_i & \cdots & \frac{\partial^2 L}{\partial \theta_2 \theta_k} |\theta_i \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial^2 L}{\partial \theta_k \theta_1} |\theta_i & \frac{\partial^2 L}{\partial \theta_k \theta_2} |\theta_i & \cdots & \frac{\partial^2 L}{\partial \theta_k \theta_k} |\theta_i
\end{bmatrix}
\]

Often the covariance matrix of the score function, also called the Fisher Information,

\[
\mathcal{I}(\theta) = \text{Cov} \left( \frac{\partial}{\partial \theta_a} L(\theta|x), \frac{\partial}{\partial \theta_b} L(\theta|x) \right)_{a,b \in \{1,2,\ldots,k\}}
\]  

is substituted for the matrix of second partial derivatives. Sometimes, the procedure using the Fisher information is called Fisher Scoring.

Newton-Raphson (or some iterative routine incorporating Newton-Raphson) often provides an efficient method of finding the maximum of a likelihood as measured by the number of iterations until convergence. There are, however, difficulties with surfaces
that are "nearly vertical" or form a "ridge" in some sub-collection of the parameters. At the other extreme, surfaces that are "nearly flat" also pose difficulties, since choosing the correct $\Delta^{(i)}$ can be problematic. There are also difficulties in choosing starting values, as the convergence of the algorithm is often highly influenced by the initial guess of the correct parameter vector. There is also the practical difficulty of computing first and second derivatives for complicated likelihoods.

2.1.2 Cohen's Method for IID Truncated Normal Data

The left-truncated normal distribution is considered in what follows, and so a refinement of the straightforward Newton-Raphson approximation technique for this case is appropriate. One meta-result about iterative approximation techniques is that complexity and the possibility for something to go wrong increase with the dimensionality of the problem. Thus, one should reduce the dimension of the problem whenever possible.

For data from a normal distribution with mean $\mu$ and variance $\sigma^2$, it is well known that the maximum likelihood estimators of $\mu$ and $\sigma^2$ are $\hat{x} = \frac{1}{n} \sum x_i$ and $s^2 = \frac{1}{n} \sum (x_i - \hat{x})^2$ respectively. For data that come from a left-truncated version of this distribution, it is clear that $\hat{x}$ will be biased upward from $\mu$ while $s^2$ will be too small for estimating $\sigma^2$. Several techniques for "correcting" these estimators have been developed. In general these involve the use of an auxiliary function whose values can be had by solving a one-dimensional optimization problem.

One technique developed by Cohen (1959) is typical. For purposes of consistency we will temporarily let $d$ be a "known" left truncation point. The truncated normal log-likelihood is

$$f(x_1, \ldots, x_n|\mu, \sigma) = \left[1 - \Phi \left( \frac{d - \mu}{\sigma} \right) \right]^{-n} (\sigma \sqrt{2\pi})^{-n} \exp \left[ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right].$$

The likelihood equations derived from the log-likelihood by setting first partials equal
to zero are

\[ \bar{x} - \mu = \sigma \left[ \frac{\phi(\xi)}{1 - \Phi(\xi)} \right] \]  \hspace{1cm} (2.2)

and

\[ s^2 + (\bar{x} - \mu)^2 = \sigma^2 \left[ 1 + \xi \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) \right], \]  \hspace{1cm} (2.3)

where we define

\[ \xi = \frac{d - \mu}{\sigma}. \]  \hspace{1cm} (2.4)

Substituting \((\bar{x} - \mu)\) from (2.2) into (2.3) and solving (2.4) for \(\mu\) and substituting for \(\mu\) in (2.2) we produce the pair of equations

\[ \sigma^2 = s^2 + \sigma^2 \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) \left[ \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi \right] \]  \hspace{1cm} (2.5)

and

\[ (\bar{x} - d) = \sigma \left[ \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi \right]. \]  \hspace{1cm} (2.6)

Squaring both sides of (2.6) and substituting on the right of (2.5) gives the equation

\[ \sigma^2 = s^2 + \left[ \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi \right] (\bar{x} - d)^2 \]

\[ = s^2 + \theta(\xi)(\bar{x} - d)^2 \]  \hspace{1cm} (2.7)

where

\[ \theta(\xi) \equiv \frac{\left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right)}{\left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi}. \]  \hspace{1cm} (2.8)

We can produce an equation for \(\mu\) involving only the auxiliary function \(\theta(\xi)\) by solving (2.6) for \(\left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi\) and then substituting on the right-hand-side of (2.5). After first subtracting \(s^2\) and then dividing by \((\bar{x} - d)\) on each side,

\[ \sigma \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) = \left[ \frac{\sigma^2 - s^2}{\bar{x} - d} \right]. \]
Since \( \sigma^2 - s_*^2 = \theta(\bar{x} - d)^2 \) from (2.7) we have
\[
\sigma \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) = \theta(\bar{x} - d).
\]
This last result, along with (2.2), gives the equation
\[
\mu = \bar{x} - \theta(\bar{x} - d).
\]
This argument yields the estimating equations for \( \mu \) and \( \sigma \),
\[
\hat{\sigma}^2 = s_*^2 + \theta(\bar{x} - d)^2 \quad (2.9)
\]
and
\[
\hat{\mu} = \bar{x} - \theta(\bar{x} - d). \quad (2.10)
\]
Thus we have reduced the maximization problem (over \( \mu \) and \( \sigma \)) to one of solving for a single parameter, namely \( \theta \). And by equation (2.8) \( \theta \) is a function of \( \xi \), which can be solved for as follows. We can eliminate \( \sigma \) in (2.5) and (2.6) through simple substitution, producing the equation
\[
\left\{ 1 - \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) \left[ \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi \right] \right\} \left\{ \left( \frac{\phi(\xi)}{1 - \Phi(\xi)} \right) - \xi \right\}^{-2} = \frac{s_*^2}{(\bar{x} - d)^2}. \quad (2.11)
\]
The right hand side of (2.11) is a known quantity from the sample data, and \( \theta \) is a simple function of \( \xi \).

### 2.1.3 Cohen’s Method for IID Left-Censored Normal Data

For iid left-censored normal data we can compute maximum likelihood estimates of \( \mu \) and \( \sigma \) by iteratively solving an optimization problem in a single dimension in a manner similar to Cohen’s treatment of the truncated normal case discussed above. The form of the auxiliary function is more complicated than for the truncated case owing to
the fact that there is information coming from censored observations (ones reported as non-detects).

For the moment adhere to the convention that \( n \) is the number of observations, both censored and ones for which we have an actual data value. Let \( n_{-\infty} \) be the number of observations for which we have no value. Let \( n_r \) be the number of observations for which we have measured an actual value. The likelihood for the left-censored normal distribution is

\[
f(n_{-\infty}, x_1, \ldots, x_r | \mu, \sigma) = \frac{(n_{-\infty} + n_r)!}{n_{-\infty}! n_r!} \left[ \Phi \left( \frac{d - \mu}{\sigma} \right) \right]^{n_{-\infty}} \left( \sigma \sqrt{2\pi} \right)^{-n_r} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{n_r} (x_i - \mu)^2 \right].
\]

Letting

\[
h \equiv \frac{n_{-\infty}}{(n_{-\infty} + n_r)},
\]

\[
\xi \equiv \frac{d - \mu}{\sigma},
\]

and \( Y(h, \xi) = \left[ \frac{h}{1 - h} \right] \left[ \frac{\phi(-\xi)}{1 - \Phi(-\xi)} \right] \)

we produce likelihood equations that are identical to (2.4), (2.2), and (2.3) with \( Y(h, \xi) \) substituted for \( \phi(\xi)/[1 - \Phi(\xi)] \). By solving these for \( \hat{\mu} \) and \( \hat{\sigma} \) we get:

\[
\hat{\sigma}^2 = s^2 + \lambda(\bar{x} - d)^2
\]

\[
\hat{\mu} = \bar{x} - \lambda(\bar{x} - d)
\]

with the auxiliary function \( \lambda \) being defined as

\[
\lambda(h, \xi) = \frac{Y(h, \xi)}{Y(h, \xi) - \xi}. \tag{2.12}
\]

In this scenario, \( \xi \) is found by solving the equation

\[
\left\{ 1 - \left( \frac{h}{1 - h} \right) \left[ \frac{\phi(-\xi)}{1 - \Phi(-\xi)} \right] \left[ \frac{h}{1 - h} \right] \left[ \frac{\phi(-\xi)}{1 - \Phi(-\xi)} \right] - \xi \right\} \left\{ \left( \frac{h}{1 - h} \right) \left[ \frac{\phi(-\xi)}{1 - \Phi(-\xi)} \right] - \xi \right\}^2 = \frac{s^2}{(\bar{x} - d)^2}. \tag{2.13}
\]
2.2 Asymptotics of Likelihood-Based Inference in “Regular” Problems

The asymptotic theory for estimators of parameters in models obeying certain regularity conditions is well established. See, for example, Schervish (1995). The two most important results concern the asymptotic distribution of the maximum likelihood estimator and the asymptotic null distribution of a likelihood ratio test statistic. More concretely, for a parameter \( \theta \in \Omega \) and hypotheses

\[
H_0 : \quad \theta = \theta_0 \quad (\equiv \Theta_0) \\
H_a : \quad \theta \in \Omega - \{\theta_0\} \quad (\equiv \Theta_a)
\]

the likelihood ratio test (LRT) statistic is

\[
L_n = \frac{\sup_{\theta \in \Theta_0} f(x|\theta)}{\sup_{\theta \in \Theta_a} f(x|\theta_a)} \frac{\prod_{i=1}^{n} f(x_i|\theta_0)}{\prod_{i=1}^{n} f(x_i|\hat{\theta})}.
\]

Schervish (1995) proves the following two theorems.

**Theorem 1** Let \( \Omega \) be a subset of \( \mathbb{R}^k \), and let \( \{x_i\}_{i=1}^{\infty} \) be IID with density \( f(x_i|\theta_0) \).

Assume the following to be true:

1. \( f_{\theta}(x) > 0 \) for all \( x \) and for all \( \theta \in \Omega \),

2. for all \( x \), \( f_{\theta}(x) \) is thrice differentiable for all \( \theta \in O \), where \( O \) is an open neighborhood of \( \Omega \) containing \( \theta_0 \),

3. differentiation with respect to \( \theta_i \) for all \( i \in \{1, 2, \cdots, k\} \), can be passed under the integral sign, that is, \( \frac{\partial}{\partial \theta_i} \int f_{\theta}(x)dx = \int \frac{\partial}{\partial \theta_i} f_{\theta}(x)dx \),

4. \( \hat{\theta}_n \xrightarrow{P} \theta_0 \) and, with probability approaching one, \( \{\hat{\theta}_n\} \) is a root of the likelihood equations,

5. \( \sup_{1<|\theta-\theta_0|<r} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_k} \log f_{\theta_0}(x) - \frac{\partial^2}{\partial \theta_j \partial \theta_k} \log f_{\theta}(x) \right| \leq H_{r,\theta_0}(x) \) for all \( \theta \in O \), and for all \( j, k \leq n \), where \( \lim_{r \to 0} H_{r,\theta_0}(x) \to 0 \), and

6. the Fisher information matrix, \( I(\theta) \) is finite and nonsingular for all \( \theta \in O \).
Then if $\theta = \theta_0$

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{D}{\rightarrow} N(0, I^{-1}(\theta_0)). \tag{2.15}$$

**Theorem 2** Under the hypotheses of Theorem 1, and with $H_0 : \theta = \theta_0$

$$-2 \log(L_n) \equiv -2 \log \left[ \sup_{\theta \in \Theta} f_{\theta}(x|\theta_0) \right] \overset{D}{\rightarrow} \chi^2_k. \tag{2.16}$$

### 2.3 Asymptotic Distributions of Extreme Order Statistics

Often the maximum likelihood estimator for an unknown truncation (or censoring) point of a left-truncated (or left-censored) distribution is the minimum of the data (or the minimum of those data recorded as 'detects'). Sample minima are not asymptotically normal. For appropriate sequences of real numbers $\{c_n\}_{n=1}^{\infty}$ and $\{d_n\}_{n=1}^{\infty}$, the limiting distribution of the properly standardized minimum, $c_n^{-1}(\min\{x_1, \ldots, x_n\} - d_n)$, of independent and identically distributed data must belong to one of three families (none of which includes the normal) specified by cumulative distribution functions $\Psi$:

**Gumbel** $\Psi\left(\frac{x - d}{c}\right) = 1 - \exp\left[-\exp\left(\frac{x - d}{c}\right)\right] \quad x \in \mathbb{R}$

**Fréchet** $\Psi_\beta\left(\frac{x - d}{c}\right) = \begin{cases} 1 - \exp\left[-\left(\frac{x - d}{c}\right)^{-\beta}\right] & x < d \text{ and } \beta > 0 \\ 1 & x \geq d \end{cases}$

**Weibull** $\Psi_\beta\left(\frac{x - d}{c}\right) = \begin{cases} 1 - \exp\left[-\left(\frac{x - d}{c}\right)^{\beta}\right] & x \geq d \text{ and } \beta > 0 \\ 0 & x < d \end{cases}$

See Galombos (1978, p. 75).

The form of the limiting distribution of a sample minimum depends on the left tail thickness of the marginal distribution. For densities in which there is a value, $\alpha$, such that $f(x) = 0$ for all $x \leq \alpha$, the extreme order statistic is asymptotically Weibull. For
distributions with thick tails, the Fréchet limit is correct. The first order statistics of samples from the normal are asymptotically Gumbel. Exhaustive theorems are given by Galambos for determining the correct family for extreme order statistics (Galambos 1978, pp. 58-59).

For our purposes in this dissertation, a simple direct argument leads to the correct family of limit distributions and to appropriate sequences \( \{c_n\}_{n=1}^{\infty} \) and \( \{d_n\}_{n=1}^{\infty} \). First notice that

\[
\Pr(c_n^{-1}(\min\{x_1, \ldots, x_n\} - d_n) < \epsilon) = 1 - \Pr[\text{all } c_n^{-1}(x_i - d_n) > \epsilon \quad i = 1, 2, \ldots, n]
\]

\[
= 1 - (1 - \Pr[c_n^{-1}(x_1 - d_n) < \epsilon])^n
\]

\[
= 1 - \left(1 - \frac{nF(c_n \epsilon + d_n)}{n}\right)^n.
\]

Then if \( nF(c_n \epsilon + d_n) \) converges to some expression not involving \( n \), say \( \alpha(\epsilon) \), \( \Pr(c_n^{-1}(x_i - d_n)) \) converges to \( 1 - \exp(-\alpha(\epsilon)) \). We state this more formally.

**Theorem 3** Asymptotic Distribution of the Minimum for a Truncated Distribution

Assume \( x_1, x_2, \ldots, x_n \) are independent and identically distributed from a continuous distribution with density \( f(x) \) and cumulative distribution function \( F(x) \). Suppose \( \Pr(x \leq d) = 0 \), \( f(d) > 0 \) and \( f \) is right-continuous at \( d \). Let \( \alpha_n \equiv \min\{x_1, x_2, \ldots, x_n\} \).

There exist \( \{c_n\}_{n=1}^{\infty} \) and \( \{d_n\}_{n=1}^{\infty} \) such that the asymptotic distribution of \( c_n^{-1}(\alpha_n - d_n) \) is exponential with mean \( f(d)^{-1} \). In particular, with \( d_n = d \) and \( c_n = n^{-1} \) the density of the asymptotic distribution of \( c_n^{-1}(\alpha_n - d_n) \) is

\[
\mathcal{F}(x) = f(d) \exp(-f(d)x) 1_{[x > 0]}
\]
PROOF of Theorem 3

Let \( t > 0 \).

\[
\Pr(\alpha_n - d_n < t) = \Pr(\alpha_n < c_n t + d_n)
\]
\[
= 1 - \Pr(\alpha_n > c_n t + d_n)
\]
\[
= 1 - \Pr(\text{all } x_i > c_n t + d_n)
\]
\[
= 1 - \left[ \Pr(x_1 > c_n t + d_n) \right]^n
\]
\[
= 1 - \left( 1 - \frac{nF(c_n t + d_n)}{n} \right)^n
\]

Suppose that \( nF(c_n t + d_n) \) has a finite limit as \( n \to \infty \). The above string of equalities says that \( \lim_{n \to \infty} \Pr(\alpha_n - d_n < t) = 1 - \exp(- \lim_{n \to \infty} nF(c_n t + d_n)) \). Now for \( d_n \equiv d \) and \( c_n \equiv n^{-1} \) we get (through the Mean-Value-Theorem for integrals) that for \( t > 0 \)

\[
nF(c_n t + d_n) = nF(n^{-1} t + d)
\]
\[
= nf(t^*)(\frac{t}{n}) \quad \text{for some } t^* \in \left[ d, d + \frac{t}{n} \right]
\]
\[
= tf(t^*)
\]
\[
\to tf(d).
\]

Hence we see that \( \Pr(\alpha_n < c_n t + d_n) \to 1 - \exp[-f(d)t] \) which is the cumulative distribution function of the exponential distribution with mean \( f(d)^{-1} \). \( \square \)
2.4 Smith (1985) on Maximum Likelihood Estimation in Non-Regular Problems

Smith (1985) considers maximum likelihood estimation for parameters of probability densities of the form

\[ f(x|\theta, \phi) = (x - \theta)^{a-1}g(x - \theta, \phi) \quad x \in (\theta, \infty) \quad (2.17) \]

where \( \lim_{x \to \theta} g(x - \theta, \phi) = \alpha c \). This class includes translated versions of the Weibull, gamma, beta and log gamma distributions.

Smith demonstrates that the classical properties of maximum likelihood estimation apply when \( \alpha \geq 2 \) and works extensively in the case where \( \alpha \in (1, 2) \). For this case Smith demonstrates asymptotic independence of the maximum likelihood estimators for \( \theta \) and \( \phi \). One thing to note is that, for \( \alpha > 1 \), the maximum likelihood estimator for \( \theta \) need not be the minimum observation. The order of convergence for \( \hat{\theta} \) is generally faster than \( O(n^{-\frac{1}{2}}) \), being \( O(n^{-\frac{1}{2}}) \) when \( \alpha \in (1, 2) \).

For the distributions of the form (2.17) and for \( \alpha > 1 \), Smith derives the asymptotic distributions of \([k_1(n)(\hat{\theta}_n - \theta), n^{\frac{1}{2}}(\hat{\phi}_n - \phi)]\) for

\[
k_1(n) = \begin{cases} 
  n^{\frac{1}{2}} & \alpha > 2 \\
  (nc \log n)^{\frac{1}{2}} & \alpha = 2 \\
  (nc)^{\frac{1}{2}} & 1 < \alpha < 2 
\end{cases}
\]

The distribution turns out to be multivariate normal for \( \alpha \geq 2 \), with independence when \( \alpha = 2 \). Otherwise, \( n^{\frac{1}{2}}(\hat{\phi}_n - \phi_0) \) is multivariate normal while the other term has a distribution described by Woodroofe (1972).

The left-truncated log-normal density falls into Smith's paradigm for \( \alpha = 1 \). Following an approach similar to that used by Smith, the maximum likelihood estimator of the truncation parameter \( d \) can be shown to be asymptotically independent of \( \phi \). We
do this and go on to demonstrate how that fact can be used to find asymptotic null distributions of likelihood ratio statistics in Chapter 3.

2.5 Chow and Teugels (1978) on Joint Asymptotics for Extreme Order Statistics and Sample Means

For the sake of inference in truncated (and censored) families, we need to explore the asymptotic relationship between the MLE of the truncation (or censoring) parameter and the MLEs for other parameters. By restricting the class of densities considered to ones where the "non-truncation" (or "non-censoring") parameters can be estimated by solving likelihood equations, it is possible to show asymptotic independence between estimators for these parameters and that of the truncation (or censoring) parameter. This fact will be the key to determining the asymptotic null distributions of the likelihood ratio statistics.

Chow and Teugels (1978) demonstrate that in all but a single class of distributions the asymptotic joint distributions of the sample mean and the sample minimum of iid observations are ones of independence. (The lone class of densities where this is not true is the class for which the domain of attraction for the sample mean is Cauchy.) We next present a fleshed out version of their argument for distributions where the sample mean of iid observations is asymptotically normal and the sample minimum is asymptotically Weibull. This is exactly the case considered throughout this paper.

The proof of the main result depends on the following result that is implicitly assumed in Chow & Teugels (1978).

**Proposition 1** Two random variables, X and Y, are independent if there are functions \( g(t) \) and \( h(u) \) such that

\[
E (\exp(itX)1_{Y \leq u}) = g(t)h(u)
\]
for all $t \in \mathbb{R}$ and $u \in \mathbb{R}$.

An outline of a proof of this proposition is as follows. By assumption we have

$$g(t)h(u) = E \left( e^{itX} 1_{[Y \leq u]} \right)$$

$$= EE \left[ e^{itX} 1_{[Y \leq u]} | X \right]$$

$$= E e^{itX} E \left[ 1_{[Y \leq u]} | X \right]$$

$$= E e^{itX} \Pr(Y \leq u | X).$$

Consider the function $\psi(t, s) = g(t) \int e^{isu} dh(u)$, which is clearly a product of a function of $t$ and a function of $s$. We can write

$$\psi(t, s) = \int \exp(isu) g(t) \, dh(u)$$

$$= \int \exp(isu) \, d \left( E \exp(itX) \Pr(Y \leq u | X) \right)$$

$$= E \exp(itX) \int \exp(isu) \, d \Pr(Y \leq u | X)$$

$$= E \exp(itX) E \left[ \exp(isY) | X \right]$$

$$= E \left( \exp(itX + isY) \right).$$

The last expression is the joint characteristic function of $X$ and $Y$. Since this can be expressed as a product of functions of $t$ and $s$, $X$ and $Y$ are independent. \hfill \Box

**Theorem 4** Asymptotic Joint Distribution of the Sample Minimum and the Sample Mean

Suppose $x_1, x_2, \ldots, x_n$ are independent and identically distributed as in the hypotheses of Theorem 3. Let $\alpha_n = \min\{x_1, x_2, \ldots, x_n\}$. Assume that $E x_1^4 < \infty$. If $a_n = \sqrt{n}$, $b_n = a_n^{-1} E x_1$, and $c_n = n^{-1}$, then $c_n^{-1}(\alpha_n - d)$ and $(a_n^{-1} \sum x_i - nb_n)$ are asymptotically independent (Chow & Teugels (1978)).
PROOF of Theorem 4

Following Chow and Teugels (1978) we prove this using Proposition 1. Let $S_n \equiv \sum_{i=1}^{n} x_i$ and consider

$$E \{ \exp[it(a_n^{-1}S_n - nb_n)1[\sigma_n \geq c_nv + d]] \}.$$  \hspace{1cm} (2.18)

This is

$$E \{ \exp[it(a_n^{-1}x_1 - b_n) + \cdots + it(a_n^{-1}x_n - b_n)]1[\sigma_n \geq c_nv + d] \}$$

$$= E \prod_{i=1}^{n} \{ \exp[it(a_n^{-1}x_i - b_n)]1[x_i \geq c_nv + d] \}$$

$$= \left\{ \int_{-\infty}^{\infty} \exp[it(a_n^{-1}x_1 - b_n)]1[x_1 \geq c_nv + d] dF(x_1) \right\}^n$$

$$= \left\{ \int_{-\infty}^{\infty} \exp[ity]1[y \geq c_n^{-1}(c_nv + d) - b_n] dF(a_ny + a_nb_n) \right\}^n. \hspace{1cm} (2.19)$$

(Here we let $y = a_n^{-1}x_1 - b_n$ and so $x_1 = a_n y + a_nb_n$. If $x_1 \geq c_nv + d$ then $a_n y + a_nb_n \geq c_nv + d$ implying $y \geq a_n^{-1}(c_nv + d) - b_n$.) So,

r.h.s. (2.19) = \left\{ \int_{a_n^{-1}(c_nv + d) - b_n}^{\infty} \exp[ity] dF(a_ny + a_nb_n) \right\}^n

$$= \left\{ \int_{-\infty}^{\infty} \exp[ity] dF(a_ny + a_nb_n)
\quad \quad - \int_{-\infty}^{a_n^{-1}(c_nv + d) - b_n} \exp[ity] dF(a_ny + a_nb_n) \right\}^n
$$

$$= \left\{ 1 + \int_{-\infty}^{\infty} (\exp[ity] - 1) dF(a_ny + a_nb_n)
\quad \quad - \int_{-\infty}^{a_n^{-1}(c_nv + d) - b_n} \exp[ity] dF(a_ny + a_nb_n) \right\}^n. \hspace{1cm} (2.20)$$

Then let $h(z) = c_n^{-1}(a_n(z + b_n) - d)$. $y = h^{-1}(z)$, and note that $h(a_n^{-1}(c_nv + d) - b_n) = v$
so that

\[
\text{r.h.s. (2.21)} = \left\{ 1 + \int_{-\infty}^{\infty} (\exp[i\pi y] - 1) dF(a_n y + a_n b_n) - \int_{-\infty}^{\infty} \exp[i\theta^{-1}(z)] dF(a_n h^{-1}(z) + a_n b_n) \right\}^n
\]

\[
= \left\{ 1 + \frac{\psi_1}{n} - \frac{\psi_2}{n} \right\}^n,
\]

for

\[
\psi_1 = \int_{-\infty}^{\infty} n(\exp[i\pi y] - 1) dF(a_n y + a_n b_n)
\]

and \( \psi_2 = \frac{1}{n} \int_{-\infty}^{\infty} n \exp\left\{ i\theta^{-1}(c_n z + d) - b_n \right\} dF(d + c_n z) \).

We will analyze \( \psi_1 \) and \( \psi_2 \) separately. Using the facts that \( \cos(\pi y) \) and \( \sin(\pi y) \) are bounded functions, and that \( y = a_n^{-1} x_1 - b_n \), we have

\[
\psi_1 = \int_{-\infty}^{\infty} n[\cos(\pi y) + i \sin(\pi y) - 1] dF(a_n y + a_n b_n)
\]

\[
= n \int_{-\infty}^{\infty} [\cos(\pi y) - 1] dF(a_n y + a_n b_n) + \int_{-\infty}^{\infty} \sin(\pi y) dF(a_n y + a_n b_n)
\]

\[
= \int_{-\infty}^{\infty} \left\{ -\frac{\left(1 + \cos(t x_1^*)\right)^2}{2} + (-1) \cos(t x_1^*) \right\} \frac{[(t x_1^* - E x_1^*)]^4}{4! n} dF(x_1)
\]

\[
+ i \int_{-\infty}^{\infty} \left\{ \frac{1}{n} \frac{1/2 (x_1^* - E x_1^*)) + (-1) \cos(t x_1^{**}) \right\} \frac{[(t x_1^{**} - E x_1^{**})]^3}{3! n^{3/2}} dF(x_1)
\]

for some \( x_1^* \in (-|x_1 - E x_1|, |x_1 - E x_1|) \) and \( x_1^{**} \in (-|x_1 - E x_1|, |x_1 - E x_1|) \). Now,

\[
- \int_{-\infty}^{\infty} \frac{[t x_1^* (x_1 - E x_1)]^2}{2} dF(x_1) = -\frac{t^2 \kappa^2 \sigma^2}{2}
\]

and

\[
\int_{-\infty}^{\infty} \frac{n^{1/2} t \kappa (x_1 - E x_1)}{dF(x_1)} = 0.
\]

The Dominated Convergence Theorem and the finiteness of \( E x_1^* \) imply

\[
\lim_{n \to \infty} \int_{-\infty}^{\infty} \cos(t x_1^*) \frac{[(t x_1^* - E x_1^*)]^4}{4! n} dF(x_1) = 0
\]

and

\[
\lim_{n \to \infty} \int_{-\infty}^{\infty} \cos(t x_1^{**}) \frac{[(t x_1^{**} - E x_1^{**})]^3}{3! n^{3/2}} dF(x_1) = 0.
\]
So

\[ \psi_1 \rightarrow \frac{\sigma^2 \kappa^2 t^2}{2}. \]

As a preliminary to the analysis of \( \psi_2 \), note that the measure defined by \( ndF(d + c_n z) \) and that defined by \( d[nF(d + c_n z)] \) are the same. The factor \( n \) can belong to either the measure or the integrand without changing value of an integral because of the linearity of the derivative.

Now, \( d = \sup \{ x | F(x) = 0 \} \), so it is clear that for all \( z < 0 \), \( F(d + c_n z) = 0 \). Hence, the signed measure \( nF(d + c_n z) \) places zero mass to the left of 0, implying the lower limit of integration for \( \psi_2 \) can be replaced by zero. Thus we have:

\[
\psi_2 = \int_{-\infty}^{\nu} \exp[-ib_n t + it a_n^{-1} c_n z + it a_n^{-1} d] \, nF(d + c_n z) \\
= \int_0^{\nu} \exp[-ib_n t + it a_n^{-1} c_n z + it a_n^{-1} d] \, [nF(d + c_n z)] \\
\rightarrow \int_0^{\nu} \exp[0 + 0 + 0] \frac{d}{dz} [f(d) \times z] \, dz \\
= v f(d). 
\]

The limit in display (2.22) is a consequence of Proposition 18, page 270 of Royden (1988) since the measures converge to a uniform measure on \((0, \nu)\) and the integrand is bounded.

Using the limits of \( \psi_1 \) and \( \psi_2 \) we have

\[ E \{ \exp[it(a_n^{-1} S_n - nb_n)1_{[a_n \geq c_n \nu + d]}] \} \rightarrow \exp \left( -\frac{\sigma^2 \kappa^2 t^2}{2} + v f(d) \right) \]

which factors and allows application of Proposition 1. \( \Box \)
3 INFERENCE IN TRUNCATED EXPONENTIAL FAMILIES (WITH UNKNOWN TRUNCATION PARAMETER)

In this chapter we consider (as a first preliminary to analysis of our motivating problem) inference based on independent and identically distributed observations from a left-truncated exponential family of distributions with unknown truncation parameter. Versions of the results presented here follow equally well for right-truncated families of distributions and most will hold for doubly-truncated distributions, although we will explicitly state only the left-truncated results.

Inference based on independent and identically distributed data is typically directed at finding consistent and efficient point estimators for the parameters. Maximum likelihood is typically a good approach. Maximum likelihood estimators are often solutions to the likelihood equations, the set of equations derived by setting the first partial derivatives of the log-likelihood equal to zero. At least in a situation where the regularity conditions of Theorem 1 are satisfied, MLEs are consistent and asymptotically normal. This chapter considers likelihood-based estimation in a non-regular family. It demonstrates that the particular kind of non-regularity introduced by an unknown truncation parameter poses no great theoretical difficulty, because of the kind of asymptotic independence of means and minima established in Theorem 4 of the previous chapter. Through simulation we further demonstrate that our asymptotic results can be relevant for samples of practical size.
3.1 Exponential Families

Many common families of distributions are (untruncated) exponential families. In fact, the most common continuous probability densities and discrete probability mass functions are members of exponential families. A regular exponential family consists of those distributions whose densities with respect to some measure can be written in the form

$$f_\eta(x) = g(\eta)h(x)\exp\left[\sum_{j=1}^{k}\eta_j t_j(x)\right]. \quad (3.1)$$

A couple of examples are in order.

The Weibull family is a popular family of distributions in the detection limits literature as well as in reliability. The two-parameter Weibull density with respect to Lebesgue measure on $[0, \infty)$ is of the form

$$f_{\theta, \alpha}(x) = \frac{\beta}{\theta^\alpha}x^{\alpha-1}\exp\left[-\frac{x^\alpha}{\theta}\right] \quad x \geq 0.$$

We can re-express this as

$$f_{\theta, \alpha}(x) = \left(\frac{\beta}{\theta^\alpha}\right)\exp\left[(\alpha - 1)\log x - \left(x^\alpha/\theta\right)^\alpha\right]
= g(\eta)\exp[\eta_1t_1(x) + \eta_2t_2(x)],$$

where we have defined

$$\eta_1 = \beta - 1, \quad t_1(x) = \log(x), \quad g(\eta) = -(\eta_1 + 1)\eta_2,$$

$$\eta_2 = -\theta^{-\alpha}, \quad t_2(x) = x^\alpha, \quad \text{and } h(x) = 1.$$

The normal distributions also form a regular exponential family. They can be put into the standard form by defining

$$\eta_1 = -\frac{1}{2\sigma^2}, \quad t_1(x) = x^2, \quad g(\eta) = \frac{\sqrt{-\eta_1}}{\sqrt{\pi}}\exp\left[\frac{\eta_2^2}{4\eta_1}\right],$$

$$\eta_2 = \frac{\mu}{\sigma^2}, \quad t_2(x) = x, \quad \text{and } h(x) = 1.$$
With these conventions we see that, for $x \in \mathbb{R}$, the normal probability density (with respect to Lebesgue measure) can be written as

$$f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$$

$$= \left[ \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{\mu^2}{2\sigma^2} \right) \right] \exp \left[ \left( \frac{-1}{2\sigma^2} \right) x^2 + \left( \frac{\mu}{\sigma^2} \right) x \right]$$

$$= \left[ \left( \frac{\sqrt{-(-1/2\sigma^2)}}{\sqrt{\pi}} \right) \exp \left( \frac{\mu^2/\sigma^4}{-4/2\sigma^2} \right) \right] \exp \left[ \left( \frac{-1}{2\sigma^2} \right) x^2 + \left( \frac{\mu}{\sigma^2} \right) x \right]$$

$$= h(x)g(\eta) \exp [\eta_1 t_1(x) + \eta_2 t_2(x)].$$

### 3.2 Notation

We establish notation that will be used for the remainder of this chapter. Define

- $\eta_j$: the $j^{th}$ of $k$ parameters whose estimator is found by solving the likelihood equations,
- $\hat{\eta}_j$: the solution of the likelihood equation for $\eta_j$,
- $d$: a truncation parameter,
- $\hat{d}$: a maximum likelihood estimator of $d$,
- $f_{\eta, d}(x)$: the density (with respect to Lebesgue measure) of a member of the truncated exponential family with parameters $(\eta, d)$,
- $L(\eta, d)$: the log-likelihood based on a random sample from the truncated exponential family with parameters $(\eta, d)$,
- $L_{\eta_j}(\eta, d)$: the first partial of $L(\eta, d)$ with respect to $\eta_j$, $\partial L(\eta, d)/\partial \eta_j$,
- $L_d(\eta, d)$: the first partial of $L(\eta, d)$ with respect to $d$, $\partial L(\eta, d)/\partial d$,
- $t$: the vector $\{t_1(X), t_2(X), \ldots, t_k(X)\}'$,.
the sample average of \( t_j(x_i) \), \( \overline{t}_j = n^{-1} \sum_{i=1}^{n} t_j(x_i) \),

\( \mathbf{t} \) the vector \( \{\overline{t}_1, \overline{t}_2, \ldots, \overline{t}_k\} \),

\( \eta \cdot \mathbf{t}(x) \) shorthand for \( \sum_{j=1}^{k} \eta_j t_j(x) \),

\( \mathcal{K}(\eta, d) \) the inverse of the normalizing constant, \( \left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] dx \right)^{-1} \),

\( \mathcal{K}'(\eta, d) \) the first partial of \( \mathcal{K} \) with respect to \( \eta_j \), \( \mathcal{K}'(\eta, d) = \frac{\partial}{\partial \eta_j} \mathcal{K}(\eta, d) \),

\( \mathcal{K}''(\eta, d) \) the second partial of \( \mathcal{K} \) with respect to \( \eta_j \), \( \mathcal{K}''(\eta, d) = \frac{\partial^2}{\partial \eta_j^2} \mathcal{K}(\eta, d) \),

\( \mathcal{K}''_{jm}(\eta, d) \) the mixed partial of \( \mathcal{K}(\eta, d) \), \( \mathcal{K}''_{jm}(\eta, d) = \frac{\partial^2}{\partial \eta_j \partial \eta_m} \mathcal{K}(\eta, d) \),

\( r_j(\eta, d) \) the expected value of \( t_j(X) \), \( \text{Et}_j(X) = \int_{d}^{\infty} t_j(x) \mathcal{K}(\eta, d) h(x) \exp[\eta \cdot \mathbf{t}(x)] dx \),

\( \mathbf{r}(\eta, d) \) the vector \( \{r_1(\eta, d), r_2(\eta, d), \ldots, r_k(\eta, d)\} \),

\( r_{jm}^i(\eta, d) \) the first partial of \( r_j(\eta, d) \) with respect to \( \eta_m \), \( \partial r_j(\eta, d)/\partial \eta_m \),

\( r_j^d(\eta, d) \) the first partial of \( r_j(\eta, d) \) with respect to \( d \), \( \partial r_j(\eta, d)/\partial d \).

### 3.3 Truncated Exponential Families

We turn our attention to families of distributions with densities with respect to Lebesgue measure on \( \mathbb{R} \) of the form

\[
\int_{\eta, d} f_{\eta, d}(x) = \frac{g(\eta)h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] 1_{[d \leq x]}}{\int_{d}^{\infty} g(\eta)h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] dx} = \frac{h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] 1_{[d \leq x]}}{\int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] dx} \quad (3.2)
\]
The joint density of iid data from the truncated exponential family (3.2) is

\[
f_{\eta, d}(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left\{ \frac{h(x_i) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x_i) \right] 1_{[x_i \geq d]}}{\left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] \, dx \right)} \right\}
\]

\[
= \frac{\left( \prod_{i=1}^{n} h(x_i) \exp \left[ \sum_{j=1}^{k} \sum_{i=1}^{n} \eta_j t_j(x_i) \right] 1_{[\min\{x_1, x_2, \ldots, x_n\} \geq d]} \right)}{\left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] \, dx \right)^n}.
\]

It is often unpleasant to work with the joint density (or likelihood) directly. For purposes of maximizing (or minimizing) we may work with the logarithm of the likelihood since \(\log(\cdot)\) is a monotone increasing function. The logarithm has the nice property that products become sums and exponents become multipliers. The log-likelihood for iid data from the truncated exponential family (3.2) is

\[
L(\eta, d) = \log \left\{ \frac{\left( \prod_{i=1}^{n} h(x_i) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x_i) \right] 1_{[\min\{x_1, x_2, \ldots, x_n\} \geq d]} \right)}{\left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] \, dx \right)^n} \right\}
\]

\[
= \sum_{i=1}^{n} \log h(x_i) + \sum_{i=1}^{n} \sum_{j=1}^{k} \eta_j t_j(x_i) - n \log \left[ \int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx \right] \quad (3.3)
\]

\[
= \sum_{i=1}^{n} \log h(x_i) + n \sum_{j=1}^{k} \eta_j t_j + n \log [\mathcal{K}(\eta, d)]. \quad (3.4)
\]

### 3.4 Moments in a Truncated Exponential Family

Here we note some properties of the family (3.2) that will prove helpful in the analysis of likelihood-based estimation. The necessary results concern the expected value and variance of the various \(t_j(X)\) terms from (3.2).
All the asymptotic results that follow ultimately depend on smoothness properties of \( \mathbf{r}(\mathbf{\eta}, d) \). We may explicitly demonstrate smoothness by computing the various partial derivatives of \( \mathbf{r}(\mathbf{\eta}, d) \). By definition

\[
    r_j(\mathbf{\eta}, d) = \frac{\int_d^\infty t_j(x)h(x)\exp\left[\sum_{j=1}^{k} \eta_j t_j(x)\right] \, dx}{\int_d^\infty h(x)\exp\left[\sum_{j=1}^{k} \eta_j t_j(x)\right] \, dx}
\]

Partial derivatives of \( r_j(\mathbf{\eta}, d) \) exist and are easily obtained. First

\[
    r_j^{\eta_j}(\mathbf{\eta}, d) = \left\{ \frac{\left(\int_d^\infty h(x)\exp[\mathbf{\eta} \cdot \mathbf{t}(x)] \, dx\right) \left(\int_d^\infty t_j(x)^2 h(x)\exp[\mathbf{\eta} \cdot \mathbf{t}(x)] \, dx\right)}{\left(\int_d^\infty h(x)\exp[\mathbf{\eta} \cdot \mathbf{t}(x)] \, dx\right)^2} \right\} \\
    - \left\{ \frac{\left(\int_d^\infty t_j(x)h(x)\exp[\mathbf{\eta} \cdot \mathbf{t}(x)] \, dx\right)^2}{\left(\int_d^\infty h(x)\exp[\mathbf{\eta} \cdot \mathbf{t}(x)] \, dx\right)^2} \right\} \\
    = \text{Et}_j^2(X) - [\text{Et}_j(X)]^2 \\
    = \text{Var}_j(X).
\]  

(3.5)

Similarly

\[
    r_j^{m}(\mathbf{\eta}, d) = \text{E}[t_j(X)t_m(X)] - \text{Et}_j(X)\text{Et}_m(X) \\
    = \text{Cov}(t_j(X), t_m(X)).
\]

The interesting fact that the various partials of \( r_j(\mathbf{\eta}, d) \) with respect to the entries of \( \mathbf{\eta} \) are the variance and covariances of \( t_j(X) \) has far reaching implications (not the least of which is that the log-likelihood has a unique maximum). We unify these results in Lemma 1, ultimately showing the relation between the constant of integration in the density (3.2) and the variance-covariance matrix of \( \mathbf{t}(X) \). But first we display the partial
of $r_j(\eta, d)$ with respect to $d$,

$$
\begin{align*}
\frac{\partial^2}{\partial \eta^2} r_j(\eta, d) &= -\left\{ \left( \int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx \right) t_i(d) h(d) \exp [\eta \cdot t(x)] \right\} \\
&\quad \left\{ \left( \int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx \right)^2 \right\} \\
&= -t_j(d) f_{\eta,d}(d) + E t_j(X) f_{\eta,d}(x) \\
&= [E t_j(X) - t_j(d)] f_{\eta,d}(d). \tag{3.6}
\end{align*}
$$

**Lemma 1** Let $X$ have density (3.2). The expected value of $t_j(X)$ is

$$
r_j(\eta, d) = E t_j(X) = \frac{\partial}{\partial \eta_j} \left[ -\log \left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] \, dx \right)^{-1} \right] = \frac{\partial}{\partial \eta_j} [ -\log \mathcal{K}(\eta, d)].
$$

The variance of $t_j(X)$ is given by

$$
r_j^{\eta j}(\eta, d) = \text{Var} t_j(X) = \frac{\partial^2}{\partial \eta^2} \left[ -\log \left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] \, dx \right)^{-1} \right] = \frac{\partial^2}{\partial \eta_j^2} [ -\log \mathcal{K}(\eta, d)],
$$

and the covariance between $t_j(X)$ and $t_m(X)$ is

$$
r_j^{\eta m}(\eta, d) = r_j^{\eta m}(\eta, d) = \text{Cov}(t_j(X), t_m(X)) = \frac{\partial^2}{\partial \eta_j \partial \eta_m} \left[ -\log \left( \int_{d}^{\infty} h(x) \exp \left[ \sum_{j=1}^{k} \eta_j t_j(x) \right] \, dx \right)^{-1} \right] = \frac{\partial^2}{\partial \eta_j \partial \eta_m} [ -\log \mathcal{K}(\eta, d)]
$$

for all $j \in \{1, 2, \cdots, k\}$ and $m \in \{1, 2, \cdots, k\}$. 
PROOF of Lemma 1

From here forward express $\sum_{j=1}^{k} \eta_j t_j(x)$ as $\eta \cdot t(x)$. Then,

$$\frac{\partial}{\partial \eta_j} f_{\eta,d}(x) = \frac{\partial}{\partial \eta_j} \{ \mathcal{K}(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \}$$

$$= \mathcal{K}'(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} + \mathcal{K}(\eta, d) t_j(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}}.$$

Integrating,

$$\int \frac{\partial}{\partial \eta_j} f_{\eta,d}(x) dx = \int \mathcal{K}'(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} dx + \int \mathcal{K}(\eta, d) t_j(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} dx.$$

So

$$\frac{\partial}{\partial \eta_j} (1) = \left( \frac{\mathcal{K}'(\eta, d)}{\mathcal{K}(\eta, d)} \right) \int \mathcal{K}(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} dx$$

$$+ \text{Et}_j(X),$$

that is

$$0 = \left( \frac{\mathcal{K}'(\eta, d)}{\mathcal{K}(\eta, d)} \right) + \text{Et}_j(X),$$

so

$$\text{Et}_j(X) = \frac{\partial}{\partial \eta_j} [- \log \mathcal{K}(\eta, d)].$$

The second partial derivative of the density with respect to $\eta_j$ can, similarly, be used to find the second moment of $t_j(X)$.

$$\frac{\partial^2}{\partial \eta_j^2} f_{\eta,d}(x) = \frac{\partial}{\partial \eta_j} \{ \mathcal{K}'(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \}$$

$$+ \mathcal{K}(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}}$$

$$= \mathcal{K}'(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}}$$

$$+ 2 \mathcal{K}(\eta, d) h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}}$$

$$+ \mathcal{K}(\eta, d) h(x) [t_j(x)]^2 \exp [\eta \cdot t(x)] 1_{\{x \geq d\}}$$

$$= \psi_1''(x) + \psi_2''(x) + \psi_3''(x),$$
where
\[ \psi_1^m(x) = \mathcal{K}_{jj}''(\eta, d)h(x)\exp[\eta \cdot t(x)]1_{\{x \geq d\}}, \]
\[ \psi_2^m(x) = 2\mathcal{K}_j'(\eta, d)h(x)t_j(x)\exp[\eta \cdot t(x)]1_{\{x \geq d\}}, \]
and
\[ \psi_3^m(x) = \mathcal{K}(\eta, d)h(x)[t_j(x)]^2\exp[\eta \cdot t(x)]1_{\{x \geq d\}}. \]

Then integrating,
\[ \int \psi_1^m(x)dx = \left( \frac{\mathcal{K}_{jj}''(\eta, d)}{\mathcal{K}(\eta, d)} \right) \int \mathcal{K}(\eta, d)h(x)\exp[\eta \cdot t(x)]1_{\{x \geq d\}}dx \]
\[ = \left( \frac{\mathcal{K}_{jj}''(\eta, d)}{\mathcal{K}(\eta, d)} \right). \]

\[ \int \psi_2^m(x)dx = 2\left( \frac{\mathcal{K}_j'(\eta, d)}{\mathcal{K}(\eta, d)} \right) \int \mathcal{K}(\eta, d)h(x)t_j(x)\exp[\eta \cdot t(x)]1_{\{x \geq d\}}dx \]
\[ = 2\left( \frac{\mathcal{K}_j'(\eta, d)}{\mathcal{K}(\eta, d)} \right) E t_j(X) \]
\[ = 2\left( \frac{\mathcal{K}_j'(\eta, d)}{\mathcal{K}(\eta, d)} \right) \left[ \frac{K_j'(\eta, d)}{\mathcal{K}(\eta, d)} \right], \]
and
\[ \int \psi_3^m(x)dx = \int \mathcal{K}(\eta, d)h(x)[t_j(x)]^2\exp[\eta \cdot t(x)]1_{\{x \geq d\}}dx \]
\[ = E t_j^2(X). \]

We arrive at the second moment of \( t_j(X) \) through
\[ \int \frac{\partial^2}{\partial \eta^2} f_{\eta, d}(x)dx = \left( \frac{\mathcal{K}_{jj}''(\eta, d)}{\mathcal{K}(\eta, d)} \right) - 2\left( \frac{\mathcal{K}_j'(\eta, d)}{\mathcal{K}(\eta, d)} \right)^2 + E t_j^2(X). \]

So,
\[ \frac{\partial^2}{\partial \eta^2} \int f_{\eta, d}(x)dx = - \left\{ 2\left( \frac{\mathcal{K}_j'(\eta, d)}{\mathcal{K}(\eta, d)} \right)^2 - \left( \frac{\mathcal{K}_{jj}''(\eta, d)}{\mathcal{K}(\eta, d)} \right) \right\} + E t_j^2(X) \]
and
\[ E t_j^2(X) - \frac{\partial^2}{\partial \eta^2}(1) = \left\{ \frac{2\left[ \mathcal{K}_j'(\eta, d) \right]^2 - \mathcal{K}_{jj}''(\eta, d)\mathcal{K}(\eta, d)}{[\mathcal{K}(\eta, d)]^2} \right\}. \]
Thus,

\[ \text{Vart}_j(X) = E t_j^2(X) - [E t_j(X)]^2 \]

\[ = \left( \frac{2 [K''_j(\eta, d)]^2 - K''_j(\eta, d)K(\eta, d)}{[K(\eta, d)]^2} \right) - \left( \frac{K'_j(\eta, d)}{K(\eta, d)} \right)^2 \]

\[ = - \left( \frac{K''_j(\eta, d)K(\eta, d) - [K'_j(\eta, d)]^2}{[K(\eta, d)]^2} \right) \]

\[ = \frac{\partial}{\partial \eta_j} \left[ - \left( \frac{K'_j(\eta, d)}{K(\eta, d)} \right) \right] \]

\[ = \frac{\partial^2}{\partial \eta_j^2} \left[ - \log K(\eta, d) \right]. \]

To find the covariance between \( t_j(X) \) and \( t_m(X) \) we look at the mixed partial

\[ \frac{\partial^2}{\partial \eta_j \partial \eta_m} f_{\eta, u}(x) = \frac{\partial}{\partial \eta_m} \left\{ K'_j(\eta, d)h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} + K(\eta, d)t_j(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \right\} \]

\[ = K''_jm(\eta, d)h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \]

\[ + K'_j(\eta, d)t_m(x)h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \]

\[ + K'_m(\eta)t_j(x)h(x) \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \]

\[ + K(\eta, d)[t_j(x)t_m(x)] \exp [\eta \cdot t(x)] 1_{\{x \geq d\}} \]

\[ = \psi_1^*(x) + \psi_2^*(x) + \psi_3^*(x) + \psi_4^*(x), \quad (3.7) \]

where

\[ \int \psi_1^*(x)dx = \left( \frac{K''_jm(\eta, d)}{K(\eta, d)} \right), \]

\[ \int \psi_2^*(x)dx = \left( \frac{K'_j(\eta, d)}{K(\eta, d)} \right) Et_m(X) = - \left( \frac{K'_j(\eta, d)K'_m(\eta, d)}{K(\eta, d)^2} \right), \]

\[ \int \psi_3^*(x)dx = \left( \frac{K'_m(\eta, d)}{K(\eta, d)} \right) Et_j(X) = - \left( \frac{K'_j(\eta, d)K'_m(\eta, d)}{K(\eta, d)^2} \right), \]

\[ \int \psi_4^*(x)dx = \left( \frac{K''_m(\eta, d)}{K(\eta, d)} \right). \]
and \( \int \psi_j^*(x) dx = E_t_j(X) t_m(X) \). Integrating both sides of (3.7), reversing the order of integration and differentiating with respect to \( \eta_j \) and then \( \eta_m \), we have that the l.h.s. is zero. This allows us to complete the proof by writing

\[
\text{Cov}(t_j(X), t_m(X)) = E(t_j(X) t_m(X)) - E_t_j(X) E_t_m(X)
\]

\[
= - \left[ \frac{\kappa_{jm}'(\eta, d)}{\kappa(\eta, d)} \right] - 2 \left( \frac{\kappa_j''(\eta) \kappa_m'(\eta)}{[\kappa(\eta, d)]^2} \right) - \left( \frac{\kappa_j'(\eta, d)}{\kappa(\eta, d)} \right) \left( \frac{\kappa_m'(\eta, d)}{\kappa(\eta, d)} \right)
\]

\[
= - \left[ \frac{\kappa_{jm}'(\eta, d)}{\kappa(\eta, d)} \right] - \left( \frac{\kappa_m'(\eta, d) \kappa_j'(\eta, d)}{[\kappa(\eta, d)]^2} \right)
\]

\[
= \frac{\partial^2}{\partial \eta_j \partial \eta_m} \{- \log [\kappa(\eta, d)]\}.
\]

With the lemma established, we may use matrix notation to compactly express the results.

\[
E^*(A') = E^*(A') = \begin{bmatrix} E_t_1(X) \\ E_t_2(X) \\ \vdots \\ E_t_k(X) \end{bmatrix} = \begin{bmatrix} r_1(\eta, d) \\ r_2(\eta, d) \\ \vdots \\ r_k(\eta, d) \end{bmatrix} = r(\eta, d).
\]

The covariance matrix of \( \mathbf{t}(X) \) is then given by

\[
\text{Var}[\mathbf{t}(X)] = \left[ \frac{\partial r(\eta, d)}{\partial \eta} \right] = \begin{bmatrix} r_1(\eta, d) & r_2(\eta, d) & \cdots & r_k(\eta, d) \\ r_2(\eta, d) & r_2(\eta, d) & \cdots & r_k(\eta, d) \\ \vdots & \vdots & \ddots & \vdots \\ r_k(\eta, d) & r_k(\eta, d) & \cdots & r_k(\eta, d) \end{bmatrix}.
\]
3.5 Likelihood-Based Estimation

Simple inspection of equation (3.3) reveals that a likelihood-based estimator of $d$ is

$$\hat{d} = \min\{x_1, \ldots, x_n\}.$$ 

The integral in display (3.4) must be positive, as the integrand is the unscaled density which, by definition, is non-negative. Since a positive multiple of this integral is being subtracted in the log-likelihood, we choose $d$ to give the smallest possible value of the integral. This occurs by making $d$ as large as permissible. And, of course, $d$ cannot be larger than the smallest observation.

We compute a local maximum of $L(\eta, \hat{d})$ by solving the "likelihood equations." The necessary partial derivatives are

$$\frac{\partial L(\eta, \hat{d})}{\partial \eta_j} = \sum_{i=1}^{n} t_j(x_i) - \left[ \frac{\int_{\hat{d}}^{\infty} t_j(x) h(x) \exp \left( \sum_{j=1}^{k} \eta_j t_j(x) \right) \, dx}{\int_{\hat{d}}^{\infty} h(x) \exp \left( \sum_{j=1}^{k} \eta_j t_j(x) \right) \, dx} \right]$$

$$= n \tilde{r}_j - n r_j(\eta, \hat{d}).$$

Setting

$$\frac{\partial L(\eta, \hat{d})}{\partial \eta_j} = 0$$

produces the equation

$$\tilde{r}_j = r_j(\eta, \hat{d}). \quad (3.8)$$

A solution to the set of equations defined by the $k$ versions of (3.8) for $j \in \{1, 2, \ldots, k\}$ is a likelihood-based estimator of $\eta$.

If one can find such a solution and wishes to conclude that a local maximum of the likelihood has been found, additional argument is required. The literature often restricts itself to discussing only consistent solutions to the likelihood equations. For truncated members of the exponential family it is possible to demonstrate that such an estimator produces at least a local maximum of the likelihood. Kennedy and Gentle (1980, p. 429)
state that for continuously differentiable partials, negative definiteness of the Hessian matrix is all that is needed to show that a solution to the likelihood equations is a local maximum. By examining the relationship between the Hessian matrix and the covariance matrix of \( t(X) \), we can hope to infer negative definiteness. That is, we have shown that \( \text{Var}(X) \) is simply

\[
\frac{\partial^2 L(\eta, d)}{\partial \eta^2} = -\frac{1}{n} \left[ \frac{\partial^2 L(\eta, d)}{\partial \eta^2} \right].
\]

Since this is a covariance matrix, it must be non-negative definite. But for fixed \( d \), members of the exponential family are such that the natural parameter space contains an open ball in \( \mathbb{R}^k \), which in turn implies that the matrix is nonsingular, positive definite. It is clear, therefore, that this matrix of second partials of the log-likelihood is negative definite and \( \hat{\eta} \) is a local maximizer of \( L(\eta, \hat{d}) \).

### 3.6 Inference in a Truncated Exponential Family

Likelihood-based inference in regular exponential families has been the subject of much research. Our purpose here is to study the impact of truncation on inference in these families.

**Theorem 5** Asymptotic Independence of the Standardized “Data Vector” and the Sample Minimum

Assume \( x_1, x_2, \ldots, x_n \) are independent and identically distributed from the truncated exponential family, (3.2), where \( h(x) \) is a continuous positive function of \( x \). Let \( \alpha_n = \min(x_1, \ldots, x_n) \), \( \tilde{t}_j = n^{-1} \sum_{i=1}^n t_j(x_i) \), \( T_j = \sqrt{n}(\tilde{t}_j - E_t(x_1)) \), and \( T = (T_1, \cdots, T_k)' \). The vector \( T \) is asymptotically independent of \( n(\hat{d} - d) \).

**PROOF of Theorem 5**

We follow the example of Theorem 4. If we can demonstrate that

\[
E \left\{ \exp \left( i \sum_{j=1}^k s_j T_j \right) 1_{[\alpha_n \geq d + \frac{\alpha_n}{n}]} \right\} = (3.9)
\]
has a limit that factors into a product of functions of the vector \((s_1, s_2, \ldots, s_k)\) and \(u\), then we will have demonstrated asymptotic independence. If \(F\) is the marginal distribution function corresponding to the family (3.2), then quantity (3.9) is

\[
\int_{d+\frac{n}{n}}^{\infty} \cdots \int_{d+\frac{n}{n}}^{\infty} \exp \left\{ i \sqrt{n} \sum_{j=1}^{k} s_j \sum_{i=1}^{n} \left[ \frac{t_j(x_i) - E t_j(x_i)}{n} \right] \right\} dF(x_1) \cdots dF(x_n)
\]

\[
= \left( \int_{d+\frac{n}{n}}^{\infty} \exp \left\{ i \sum_{j=1}^{k} s_j \left[ \frac{t_j(x_1) - E t_j(x_1)}{\sqrt{n}} \right] \right\} dF(x_1) \right)^n
\]

\[
= \left( 1 + \frac{\psi_1}{n} - \frac{\psi_2}{n} \right)^n \tag{3.10}
\]

where

\[
\psi_1 = \int_{d}^{\infty} n \left\{ \exp \left[ i \sum_{j=1}^{k} s_j n^{-1/2} (t_j(x_1) - E t_j(x_1)) \right] - 1 \right\} dF(x_1)
\]

and \(\psi_2 = \int_{d}^{d+\frac{n}{n}} n \left\{ \exp \left[ i \sum_{j=1}^{k} s_j n^{-1/2} (t_j(x_1) - E t_j(x_1)) \right] \right\} dF(x_1)\).

Now

\[
\psi_2 = \int_{d}^{d+\frac{n}{n}} n \left\{ \exp \left[ \sum_{j=1}^{k} \frac{i s_j (t_j(x_1) - E t_j(x_1))}{\sqrt{n}} \right] - 1 \right\} dF(x_1) + n \int_{d}^{d+\frac{n}{n}} dF(x_1).
\]

So

\[
\left| \psi_2 - n \int_{d}^{d+\frac{n}{n}} dF(x_1) \right| \leq \int_{d}^{d+\frac{n}{n}} n \left| \exp \left[ \sum_{j=1}^{k} \frac{i s_j (t_j(x_1) - E t_j(x_1))}{\sqrt{n}} \right] - 1 \right| dF(x_1). \tag{3.11}
\]

Then using the fact that \(|\exp(i\theta) - 1| \leq |\theta|\),

\[
\left| \psi_2 - n \int_{d}^{d+\frac{n}{n}} dF(x_1) \right| \leq \int_{d}^{d+\frac{n}{n}} n \left| \sqrt{n} \sum_{j=1}^{k} s_j (t_j(x_1) - E t_j(x_1)) \right| f_{\eta,d}(x_1) dF(x_1)
\]

\[
\leq \sqrt{n} \sum_{j=1}^{k} \int_{d}^{d+\frac{n}{n}} \left| s_j (t_j(x_1) - E t_j(x_1)) \right| f_{\eta,d}(x_1) dF(x_1)
\]

\[
= \frac{u}{\sqrt{n}} \sum_{j=1}^{k} |s_j (t_j(\xi_j) - E t_j(x_1))| f_{\eta,d}(\xi_j)
\]
for some \( \xi_1, \xi_2, \cdots, \xi_k \in (d, d + \frac{u}{n}) \). So \( \left| \psi_2 - n \int_d^{d + \frac{u}{n}} dF(x_1) \right| \to 0 \) as \( n \to \infty \). Thus

\[
\psi_2 \to u f_{\eta, d}(d).
\]

With

\[
\sigma_j^2 = \int_d^{\infty} [t_j(x_1) - Et_j(x_1)]^2 f_{\eta, d}(x_1) \, dx_1,
\]

\( \psi_1 \) converges to \( -\sum_{j=1}^{k} s_j^2 \sigma_j^2 / 2 \) by an argument identical to that used in Theorem 4. Thus we see that expression (3.10) converges to

\[
\exp \left[ -\sum_{j=1}^{k} s_j^2 \sigma_j^2 / 2 + u f_{\eta, d}(d) \right]
\]

which, by Proposition 1, finishes the argument.

We next demonstrate that the likelihood-based estimator of \( \hat{d} \) in display (3.2) is asymptotically independent of a consistent solution of the likelihood equations (3.8).

**Theorem 6 Asymptotic Joint Distribution of \( \hat{\eta} \) and \( \hat{d} \)**

Suppose \( x_1, x_2, \ldots, x_n \) are independent and identically distributed from distribution (3.2). Further assume that \( \hat{\eta} \) solves the matrix equation given by (3.8). If \( \hat{\eta} \) is consistent for \( \eta \) then \( \sqrt{n}(\hat{\eta} - \eta) \) is asymptotically normal with mean 0, and covariance matrix

\[
\left[ \frac{\partial r(\eta, d)}{\partial \eta} \right]^{-1} = -n \left[ \frac{\partial^2 L(\eta, d)}{\partial^2 \eta^2} \right]^{-1}.
\]

Further \( \sqrt{n}(\hat{\eta} - \eta) \) is asymptotically independent of \( n(\hat{d} - d) \).

**PROOF of Theorem 6**

Suppose \( \hat{\eta} \) solves the version of the equations in (3.8) where \( d \) replaces \( \hat{d} \). Then

\[
r(\hat{\eta}, d) = \bar{\ell} \text{ and } r(\hat{\eta}, \hat{d}) = \bar{\ell} \text{ together imply } r(\hat{\eta}, d) = r(\hat{\eta}, \hat{d}).
\]

For \( j = 1, 2, \cdots, k \)

\[
r_j(\hat{\eta}, \hat{d}) - r_j(\eta, d) = r_j(\hat{\eta}, \hat{d}) - r_j(\hat{\eta}, d) + r_j(\hat{\eta}, d) - r_j(\eta, d)
\]

\[
= r_j(\hat{\eta}, d) - r_j(\hat{\eta}, d) + r_j(\hat{\eta}, d) - r_j(\eta, d).
\]
So

$$\tilde{t}_j - E\hat{t}_j(X) = \sum_{m=1}^{k} (\tilde{\eta}_j - \hat{\eta}_j) \left[ \frac{\partial r_j(\eta_j^*, \eta)}{\partial \eta_j} \right] + r_j(\hat{\eta}, \eta) - r_j(\tilde{\eta}, \eta)$$

for some $\eta_j^*$ on the segment between $\hat{\eta}$ and $\tilde{\eta}$.

Now write $r_j(\hat{\eta}, \eta)$ two ways

$$r_j(\hat{\eta}, \eta) = r_j(\tilde{\eta}, \eta) + \sum_{m=1}^{k} (\tilde{\eta}_m - \hat{\eta}_m) \left[ \frac{\partial r_j(\eta^*, \eta)}{\partial \eta_m} \right]$$

(3.12)

and

$$r_j(\hat{\eta}, \eta) = r_j(\tilde{\eta}, \hat{\eta}) + (\eta - \hat{\eta}) \left[ \frac{\partial r_j(\eta^*, \hat{\eta})}{\partial \eta} \right]$$

(3.13)

for some $d^* \in (\eta, \hat{\eta})$. (See Figure 3.1 for a $k = 1$ illustration.)

So, recalling that

Figure 3.1 Illustration of Equations (3.12) and (3.13)
\[ r_j(\hat{\eta}, \hat{d}) = r_j(\hat{\eta}, d) = \hat{t}_j, \] we have
\[
(d - \hat{d}) \left[ \frac{\partial r_j(\hat{\eta}, d^*_j)}{\partial d} \right] = \sum_{m=1}^{k} (\hat{\eta}_m - \hat{\eta}_m') \left[ \frac{\partial r_j(\eta^*_j, d)}{\partial \eta_m} \right].
\]

Thus
\[
\hat{t}_j - E \hat{t}_j(X) = (d - \hat{d}) \frac{\partial r_j(\hat{\eta}, d^*_j)}{\partial d} + r_j(\hat{\eta}, d) - r_j(\eta, d).
\]

Collecting these equalities for \( j = 1, 2, \ldots, k \) into matrix form
\[
\begin{bmatrix}
\frac{\partial r_1(\hat{\eta}, d^*_1)}{\partial d} \\
\vdots \\
\frac{\partial r_k(\hat{\eta}, d^*_k)}{\partial d}
\end{bmatrix}
\]

Hence
\[
\sqrt{n}(\hat{r}(\hat{\eta}, d) - r(\eta, d)) = \sqrt{n}(\hat{t} - r(\eta, d)) + \sqrt{n}(\hat{d} - d).
\]

Now then, the assumed consistency of \( \hat{\eta} \), the fact that each \( d^*_j \) is between \( d \) and \( \hat{d} \) and the smoothness of the \( r_j \) together imply that the vector of partials of \( r \) with respect to \( d \) converges in probability to a constant vector. Then, the fact that \( \sqrt{n}(\hat{d} - d) \) converges to 0 in probability implies that
\[
\sqrt{n}(\hat{r}(\hat{\eta}, d) - r(\eta, d))
\]
has the same limiting distribution as
\[
\sqrt{n}(\hat{t} - r(\eta, d)),
\]

namely multivariate normal with mean 0 and covariance matrix \( \text{Var}\hat{t} \).

But then for some \( \eta^*_j \) on the segment between \( \hat{\eta} \) and \( \eta \)
\[
\begin{align*}
r_j(\hat{\eta}, d) - r_j(\eta, d) &= \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m') \left[ \frac{\partial r_j(\eta^*_j, d)}{\partial \eta_m} \right] \\
&= \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m') \left[ \frac{\partial r_j(\eta^*_j, d)}{\partial \eta_m} \right].
\end{align*}
\]
In matrix notation

\[
\mathbf{r}(\hat{\eta}, d) - \mathbf{r}(\eta, d) = \begin{bmatrix}
\frac{\partial r_1(\eta_1^{**}, d)}{\partial \eta_1} & \frac{\partial r_1(\eta_1^{**}, d)}{\partial \eta_2} & \cdots & \frac{\partial r_1(\eta_1^{**}, d)}{\partial \eta_k} \\
\frac{\partial r_2(\eta_2^{**}, d)}{\partial \eta_1} & \frac{\partial r_2(\eta_2^{**}, d)}{\partial \eta_2} & \cdots & \frac{\partial r_2(\eta_2^{**}, d)}{\partial \eta_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial r_k(\eta_k^{**}, d)}{\partial \eta_1} & \frac{\partial r_k(\eta_k^{**}, d)}{\partial \eta_2} & \cdots & \frac{\partial r_k(\eta_k^{**}, d)}{\partial \eta_k}
\end{bmatrix} (\hat{\eta} - \eta).
\]

Note that the matrix of partials multiplying \((\hat{\eta} - \eta)\) in this expression converges in probability (because of the consistency of \(\hat{\eta}\) and the smoothness of the \(r_j\)'s) to the matrix

\[
\frac{\partial \mathbf{r}(\eta, d)}{\partial \eta} = \text{Var} t.
\]

Thus

\[
\sqrt{n}(\hat{\eta} - \eta) \overset{d}{=} \sqrt{n} \begin{bmatrix}
\frac{\partial r_1(\eta_1^{**}, d)}{\partial \eta_1} & \frac{\partial r_1(\eta_1^{**}, d)}{\partial \eta_2} & \cdots & \frac{\partial r_1(\eta_1^{**}, d)}{\partial \eta_k} \\
\frac{\partial r_2(\eta_2^{**}, d)}{\partial \eta_1} & \frac{\partial r_2(\eta_2^{**}, d)}{\partial \eta_2} & \cdots & \frac{\partial r_2(\eta_2^{**}, d)}{\partial \eta_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial r_k(\eta_k^{**}, d)}{\partial \eta_1} & \frac{\partial r_k(\eta_k^{**}, d)}{\partial \eta_2} & \cdots & \frac{\partial r_k(\eta_k^{**}, d)}{\partial \eta_k}
\end{bmatrix}^{-1} (\mathbf{r}(\hat{\eta}, d) - \mathbf{r}(\eta, d))
\]

has a multivariate normal limiting distribution with mean \(0\) and covariance matrix \([\text{Var} t]^{-1}\).

Regarding the asymptotic independence of \(\sqrt{n}(\hat{\eta} - \eta)\) and \(\sqrt{n}(\hat{d} - d)\), note that

\[
\begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix} = \begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix} + \begin{pmatrix}
\sqrt{n}(\hat{\eta} - \hat{\eta}) \\
0
\end{pmatrix}.
\]

The proof shows that \(\sqrt{n}(\hat{\eta} - \eta)\) converges to \(0\) in probability. So

\[
\begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix} \text{ and } \begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix}
\]
have the same limiting distribution. But \( \hat{\eta} \) is a function of \( \tilde{\theta} \) through \( \mathbf{r}^{-1}(\cdot, d) \) (the inverse of \( \mathbf{r}(\cdot, d) \) mapping \( \mathbb{R}^k \) to \( \mathbb{R}^k \)). The smoothness of \( \mathbf{r}(\cdot, d) \) implies that \( \sqrt{n}(\hat{\eta} - \eta) \) is essentially a linear function of \( \sqrt{n}(\tilde{\theta} - \mathbb{E}(\tilde{\theta})) \) which by Theorem 5 is asymptotically independent of \( n(\hat{d} - d) \).

\[ \square \]

### 3.7 Likelihood Ratio Tests

The theory we developed for the asymptotic distribution of likelihood-based estimators in model (3.2) has implications for likelihood ratio testing in that family. To begin with, consider the hypotheses

\[ \begin{align*}
\mathcal{H}_0 : \quad & (\eta, d) = (\eta_0, d_0) \\
\mathcal{H}_a : \quad & \text{not } \mathcal{H}_0.
\end{align*} \]

The likelihood ratio approach to testing these is to look at the statistic

\[ \Lambda(\eta_0, d_0) = \sup_{(\eta, d)} \frac{\prod_{i=1}^{n} f_{\eta,d}(x_i)}{\prod_{i=1}^{n} f_{\eta_0,d_0}(x_i)} \]  

and reject \( \mathcal{H}_0 \) for large values of this statistic. The asymptotic null distribution of \( \Lambda \) is, thus, of considerable interest. Classical results imply that for likelihood ratio testing of point null hypotheses in \( k \)-parameter families, \( \chi_k^2 \) limiting distributions are relevant. Here we need to assess the impact of the non-regularity introduced into the exponential family by the truncation parameter.

Consider a second order multivariate Taylor expansion of

\[ \log \left( \prod_{i=1}^{n} f_{\eta_0,d_0}(x_i) \right) \]

at the point \((\hat{\eta}, \hat{d})\). This is
\[
\sum_{i=1}^{n} \log f_{\eta_0, d_0}(x_i) = \sum_{i=1}^{n} \log f_{\hat{\eta}, \hat{d}}(x_i) + (d_0 - \hat{d}) \sum_{i=1}^{n} \frac{\partial}{\partial \hat{d}} \log f_{\hat{\eta}, \hat{d}}(x_i) \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j) \frac{\partial}{\partial \eta_j} \log f_{\eta, d}(x_i) + \left[ \frac{1}{2} \right] (d_0 - \hat{d})^2 \sum_{i=1}^{n} \frac{\partial^2}{\partial d^2} \log f_{\eta, d}(x_i) \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)(d_0 - \hat{d}) \frac{\partial^2}{\partial \eta_j \partial d} \log f_{\eta, d}(x_i) \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)(\eta_{0j_2} - \hat{\eta}_{j_2}) \frac{\partial^3}{\partial \eta_{j_1} \partial \eta_{j_2}} \log f_{\eta, d}(x_i) \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)(d_0 - \hat{d})^2 \frac{\partial^3}{\partial \eta_j^2 \partial d} \log f_{\eta, d}(x_i) \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)(\eta_{0j_2} - \hat{\eta}_{j_2})(d_0 - \hat{d}) \frac{\partial^3}{\partial \eta_{j_1} \partial \eta_{j_2} \partial d} \log f_{\eta, d}(x_i) \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)(\eta_{0j_2} - \hat{\eta}_{j_2})(\eta_{0j_3} - \hat{\eta}_{j_3}) \frac{\partial^3}{\partial \eta_{j_1} \partial \eta_{j_2} \partial \eta_{j_3}} \log f_{\eta, d}(x_i) \\
(3.16)
\]

for some point \((\eta^*, d^*) = \alpha(\eta_0, d_0) + (1 - \alpha)(\hat{\eta}, \hat{d})\) for \(\alpha \in (0, 1)\). With \((\eta_0, d_0)\) probability one, \(\min(x_1, \ldots, x_n) > d_0\), and this expansion makes sense. Classical results (those for families that satisfy the conditions of Theorem 1) follow, in the simplest case, from showing that the second order terms in such an expansion are asymptotically independent chi-square random variables with a single degree of freedom. (With \(k\) such variables, the sum is asymptotically \(\chi^2_k\).) More generally, you often demonstrate that the
second order terms form, collectively, a $k$-dimensional quadratic form whose asymptotic limiting distribution is chi-square. To get the classical results in regular families, one also needs to know that the remainder terms vanish in the limit. This follows from the regularity conditions of Theorem 1.

We will show that similar results hold in our particular non-regular situation. The classical assumptions are not satisfied in our problem.

The most obvious departure from the classical situation in our model is the lack of smoothness of the likelihood in the parameter $d$. In the $k = 1$ illustration in Figure 3.2, every value to the right of $\hat{d}$ has 0 likelihood. In particular the likelihood is discontinuous on every $\epsilon$-ball centered at $(\hat{\eta}, \hat{d})$. This produces an apparent technical difficulty in making an expansion like (3.16) since Taylor’s Theorem is usually stated in terms of an open neighborhood about the point of expansion. But this difficulty is more apparent
than real, since with \((\mathbf{\eta}_0, d_0)\) probability one, \((\mathbf{\eta}_0, d_0)\) has positive likelihood and the likelihood is smooth along the line segment connecting \((\mathbf{\eta}_0, d_0)\) and \((\hat{\mathbf{\eta}}, \hat{d})\).

The general form in display (3.16) simplifies because all second and higher order derivatives of \(\log f_{\eta, d}(x)\) with respect to the \(\eta_j\)'s are constant in \(x\), as are first and higher order derivatives with respect to \(d\). That is, rewriting (3.16) for our family (3.2)

\[
\sum_{i=1}^{n} \log f_{\eta_0, d_0}(x_i) = \sum_{i=1}^{n} \log f_{\eta, d}(x_i) + n(d_0 - \hat{d}) \frac{\partial}{\partial d} \log f_{\eta, d} \bigg|_{\eta, d} \\
+ \sum_{i=1}^{n} \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j) \frac{\partial}{\partial \eta_j} \log f_{\eta, d}(x_i) \bigg|_{\eta, d} + \left[ \frac{n}{2} \right] \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)^2 \frac{\partial^2}{\partial \eta_j^2} \log f_{\eta, d} \bigg|_{\eta, d} \\
+ n \sum_{j_1 \neq j_2} (\eta_{0j_1} - \hat{\eta}_{j_1})(\eta_{0j_2} - \hat{\eta}_{j_2}) \frac{\partial^2}{\partial \eta_{j_1} \partial \eta_{j_2}} \log f_{\eta, d} \bigg|_{\eta, d} \\
+ \left[ \frac{n}{2} \right] (d_0 - \hat{d})^2 \frac{\partial^2}{\partial \eta_j^2} \log f_{\eta, d} \bigg|_{\eta, d}
\]

From display (3.15)

\[
\lambda(\mathbf{\eta}_0, d_0) = 2 \log \Lambda(\mathbf{\eta}_0, d_0) = -2 \left[ \sum_{i=1}^{n} \log f_{\eta_0, d_0}(x_i) - \sum_{i=1}^{n} \log f_{\hat{\eta}, \hat{d}}(x_i) \right].
\]

We can approximate this random variable using the Taylor expansion (3.17) and, by doing so, hope to understand the asymptotic null behavior of \(\Lambda\) in terms of the behavior of a polynomial in \((d_0 - \hat{d})\) and the \((\eta_{0j} - \hat{\eta}_j)\) for \(j \in \{1, 2, \ldots, k\}\).
It is immediately clear that

\[(\eta_{0j} - \hat{\eta}_j) \sum_{i=1}^{n} \frac{\partial}{\partial \eta_j} \log f_{\eta,d}(x_i) \bigg|_{\hat{\eta}_d} = 0 \quad \text{for all } j\]

since the value of \( \hat{\eta} \) is chosen so that the first partials are equal to zero. Examining the second order term in \( \eta \) requires a little care. It is clear

\[-\left( n \sum_{j=1}^{k} (\eta_{0j} - \hat{\eta}_j)^2 \frac{\partial^2}{\partial \eta_j^2} \log f_{\eta,d} \bigg|_{\hat{\eta}_d} \right.

\left. + 2n \sum_{j_1 \neq j_2} (\eta_{0j_1} - \hat{\eta}_{j_1})(\eta_{0j_2} - \hat{\eta}_{j_2}) \frac{\partial^2}{\partial \eta_{j_1} \partial \eta_{j_2}} \log f_{\eta,d} \bigg|_{\hat{\eta}_d} \right) \quad (3.18)

is the quadratic form

\[-[\sqrt{n}(\eta_0 - \hat{\eta})] \left[ \left( \frac{1}{n} \right) \frac{\partial^2 L(\eta, d)}{\partial^2 \eta^2} \right]_{\eta_d} \left[ \sqrt{n}(\eta_0 - \hat{\eta}) \right]. \quad (3.19)\]

By Theorem 6 we know that, under \( H_0 \), \( \sqrt{n}(\eta_0 - \hat{\eta}) \) converges to a normal random variable with mean 0 and covariance matrix \([\text{Var}(t(X))]^{-1}\), where \( t \) is distributed according to the parameters \((\eta_0, d_0)\). Since \( (\partial^2 / \partial^2 \eta_j^2) \log f_{\eta,d} \) and \( (\partial^2 / \partial \eta_j \partial \eta_m) \log f_{\eta,d} \) are continuous in \( \eta \) and \( d \) within the region of interest, and since \( (\hat{\eta}, \hat{d}) \) is consistent, it is immediate that

\[
\lim_{n \to \infty} \frac{\partial^2}{\partial \eta_{j_1} \partial \eta_{j_2}} \log f_{\eta,d} \bigg|_{\hat{\eta}_d} = \frac{\partial^2}{\partial \eta_{j_1} \partial \eta_{j_2}} \log f_{\eta,d} \bigg|_{(\eta_0, d_0)} \quad \text{for all } j_1, j_2 \in \{1, 2, \ldots, k\}. \quad (3.20)
\]

Lemma 1 establishes that \( \left( \frac{1}{n} \right) \left[ \frac{\partial^2 L(\eta,d)}{\partial \eta^2} \right] = -\text{Var}(t(X)) \). Thus, we may appeal to standard results (see, for example, Result 4.7, p. 140, Johnson and Wichern, 1992) to conclude that the quantity in display (3.19) is asymptotically \( \chi^2 \).

Continuity of the partial derivatives and consistency of \( (\hat{\eta}, \hat{d}) \) enable us to conclude that all second order mixed-partial terms (that contain \( \hat{d} \)) as well as all third and higher order terms converge to zero in probability. The cross-product second order term

\[-2n(\eta_{0j} - \hat{\eta}_j)(d_0 - \hat{d}) \frac{\partial^2}{\partial \eta \partial d} \log f_{\eta,d} \bigg|_{\hat{\eta}_d} \quad (3.21)\]

provides an example of the typical argument for demonstrating a term to be asymptotically negligible. We have shown \( n(d - d_0) \) converges to an \( \text{Exp}(1) \) variable while \( (\eta_0 - \hat{\eta}_j) \) converges to zero and the derivative term converges in probability to \( (\partial^2 / \partial \eta \partial d) \log f_{\eta,d} \bigg|_{\eta_0,d_0} \).

We next observe that the linear term in \( d \) in display (3.17) is asymptotically exponentially distributed.

**Theorem 7 (Sampling Distribution of the Linear Term for the Truncation Parameter)** Assume \( x_1, \ldots, x_n \) are independent and identically distributed from the distribution with density (3.2). Then

\[
(d - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log f_{\eta,d}(x_i) \bigg|_{\hat{\eta},d}
\]

converges in distribution to an exponential random variable with mean one.

**PROOF of Theorem 7**

It is clear that

\[
(d - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log f_{\eta,d}(x_i) \bigg|_{\hat{\eta},d} = (d - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log \left\{ \frac{h(x) \exp [\eta \cdot t(x)] 1_{(x \geq d)}}{\int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx} \right\} \bigg|_{\eta,d}
\]

\[
= -n(d - d) \frac{\partial}{\partial d} \log \left( \int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx \right) \bigg|_{\eta,d}
\]

\[
= n(d - d) \times \frac{h(d) \exp [\eta \cdot t(d)]}{\left( \int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx \right)}.
\]

Theorem 3 tells us that \( n(d - d) \) converges in distribution to an exponential random variable with mean \( f_{\eta,d}(d)^{-1} \). Using formula (3.2) we see that

\[
f_{\eta,d}(d)^{-1} = \frac{\left( \int_{d}^{\infty} h(x) \exp [\eta \cdot t(x)] \, dx \right)}{h(d) \exp [\eta \cdot t(d)]}.
\]
Continuity of the density and consistency of \((\hat{\eta}, \hat{d})\) implies that

\[
\frac{h(\hat{d}) \exp \left[ \hat{\eta} \cdot t(\hat{d}) \right]}{\left( \int_{\hat{d}} h(x) \exp \left[ \hat{\eta} \cdot t(x) \right] \, dx \right)} \rightarrow \frac{h(d) \exp \left[ \eta \cdot t(d) \right]}{\left( \int_{d} h(x) \exp \left[ \eta \cdot t(x) \right] \, dx \right)}.
\]

The product in (3.22) is, therefore, asymptotically exponential with mean one.

By Theorem 7 and the previous discussion, we see that, in the limit, the likelihood ratio statistic reduces to the sum of two independent random variables, one of which is \(\chi_k^2\) and the other of which is twice an exponential random variable with mean one. But an exponential variable with mean two is \(\chi_2^2\). It is well known that the sum of independent chi-square random variables is chi-square with degrees of freedom equal to the sum of the component degrees of freedom. So the limiting null distribution of \(\lambda(\eta_0, d_0) = 2 \log \Lambda(\eta_0, d_0)\) is \(\chi_{k+2}^2\).

### 3.7.1 Confidence Regions

A large \(n\) confidence region for \((\eta, d)\) can be had by inverting likelihood ratio tests of \(H_0 : (\eta, d) = (\eta_0, d_0)\). That is, an approximate \(\gamma\)-level confidence region for \((\eta, d)\) is given by

\[
\left\{ (\eta, d) \bigg| -2 \left[ \sum_{i=1}^{n} \log f_{\eta, d}(x_i) \bigg|_{\eta, d} - \sum_{i=1}^{n} \log f_{\eta, d}(x_i) \bigg|_{\eta_0, d} \right] \leq \chi_{k+2}^2(\gamma) \right\}. \tag{3.23}
\]

### 3.7.2 Profile Likelihood

It makes sense to consider likelihood ratio tests for parts of the parameter vector \((\eta, d)\). The most obvious version is probably the case \(H_0 : \eta = \eta_0\). The likelihood ratio statistic in this case involves maximizing the likelihood over \(d\) where \(\eta = \eta_0\). That is, we consider

\[
L(\hat{\eta}, \hat{d}) - \sup_d L(\eta_0, d).
\tag{3.24}
\]
It is obvious (from casual inspection) that \( \sup_d \sum \log f(x_i|\eta_0, d) \) is achieved when \( d = \min\{x_i\}_{i=1}^n = \hat{d} \). So we consider a Taylor expansion of (3.24) in \( \eta \) only.

In the case where \( k = 1 \), for some \( \eta^* \) on the interval between \( \eta_0 \) and \( \hat{\eta} \),

\[
L(\eta_0, \hat{d}) - L(\hat{\eta}, \hat{d}) = (\eta_0 - \hat{\eta}) \sum_{i=1}^n \frac{\partial}{\partial \eta} \log f_{\eta,d}(x_i) \bigg|_{\eta} + \left[ \frac{n}{2} \right] (\eta_0 - \hat{\eta})^2 \frac{\partial^2}{\partial \eta^2} \log f_{\eta,d} \bigg|_{\eta} \]

\[
+ \left[ \frac{n}{6} \right] (\eta_0 - \hat{\eta})^3 \frac{\partial^3}{\partial \eta^3} \log f_{\eta,d} \bigg|_{\eta^*}.
\]

The same sort of arguments that were used in the analysis of likelihood ratio testing for \( H_0 : (\eta, d) = (\eta_0, d_0) \) apply here as well. Continuity of the log-likelihood and its derivatives in \( \eta \) on the interval between \( \eta_0 \) and \( \hat{\eta} \) ensures that the third order term is asymptotically zero under the null hypothesis, since \( n(\eta_0 - \hat{\eta})^3 \to 0 \). The first order term is zero by the definition of \( \eta \). The second order term is asymptotically \( \chi_1^2 \) since, under \( H_0 \), the expression \( \sqrt{n}(\hat{\eta} - \eta) \) is asymptotically normal and the continuity of derivatives of the function (3.2) in \( (\eta, d) \) implies

\[
\frac{\partial^2}{\partial \eta^2} \log f_{\eta,d} \bigg|_{\hat{\eta}} \to \frac{\partial^2}{\partial \eta^2} \log f_{\eta_0,d}
\]

(where \( d \) is the value of the truncation parameter).

The extension of this result to \( k > 1 \) is similar to (but easier than) our analysis of the null behavior of the likelihood ratio test statistic for \( H_0 : (\eta, d) = (\eta_0, d_0) \). We need to express the second order partials as quadratic forms and use standard results concerning quadratic forms of normal random vectors. The resultant test statistic is asymptotically \( \chi_k^2 \) under \( H_0 : \eta = \eta_0 \). Thus, in terms of handling data from a truncated distribution with unknown truncation point, one may approximate the threshold parameter with the sample minimum and proceed as if the parameter is known. Asymptotically the likelihood ratio test statistic has the same distribution as if the threshold parameter were known.
3.8 Numerics for the Truncated Normal Problem

Having developed asymptotic theory for likelihood-based inference in general continuous truncated exponential families we now present a specific application. We study the truncated normal problem since the normal distribution is part of our motivating problem and has many other applications.

As discussed in the last chapter, Cohen (1959) devised a method for obtaining maximum likelihood estimates in the usual parameterization for the truncated normal distribution with known truncation point. His method involves an auxiliary function of a single parameter. Since the sample minimum is the maximum likelihood estimator of an unknown truncation point in the normal family, we can make use of Cohen's work for numerical purposes by treating the sample minimum as a known truncation point. In what follows an explicit procedure is presented for getting MLEs numerically by taking advantage of this auxiliary function. (Added to Cohen's effort is the invertability of the auxiliary function and a prescription for estimating a value of that inverse both directly and through the use of Base SAS® software.) Define

\[ Z(\xi) \equiv \frac{\phi(\xi)}{1 - \Phi(\xi)} \quad \text{and} \quad g(\hat{d}, \bar{x}, s^2) \equiv \frac{s^2}{(\bar{x} - \hat{d})^2} \quad (3.25) \]

(where \( s^2 = [(n - 1)/n]s^2 \)) and further define

\[ \lambda(\xi) \equiv \{1 - Z(\xi)[Z(\xi) - \xi]\} \{Z(\xi) - \xi\}^{-2}. \quad (3.26) \]

Then

\[ \frac{\partial \lambda}{\partial \xi} = \frac{(Z - \xi)^2(Z + \xi Z' - 2Z Z') + 2(1 - Z')(Z - \xi)(1 - Z^2 + Z\xi)}{(Z - \xi)^4} \]

where

\[ Z' = \frac{\partial Z}{\partial \xi} = \frac{\xi\Phi(\xi)\phi(\xi) - \xi\phi(\xi) + [\phi(\xi)]^2}{[1 - \Phi(\xi)]^2}. \]
Figure 3.3 is a plot of $\lambda(\xi)$ and its first derivative.

Since the derivative of $\lambda(\xi)$ exists and is positive (at least on $[-5, 5]$) we know that a smooth inverse for $\lambda, \lambda^{-1}$, exists. We may then obtain our estimates of $\mu$ and $\sigma^2$ as

$$\hat{\sigma}^2 = s^2 + \theta\{\lambda^{-1}[g(\hat{d}, \bar{x}, s^2)]\}(\bar{x} - \hat{d})^2$$

(3.27)

and

$$\hat{\mu} = \bar{x} - \theta\{\lambda^{-1}[g(\hat{d}, \bar{x}, s^2)]\}(\bar{x} - \hat{d}).$$

(3.28)

where the function $\theta(\cdot)$ is defined in display (2.8). Figure 3.4 is a plot of the function $\theta(g)$ (or $\theta(\lambda^{-1}(g))$, to be precise).
We now have reached an explicit method for obtaining the MLEs for \( \mu \) and \( \sigma \) in, more or less, closed form. The only difficulty occurs with computing the function \( \lambda^{-1}(g) \). There are several reasonable methods.

The Stone-Weierstrass Theorem guarantees that there is a sequence of polynomials that converge to \( \lambda^{-1} \). By selecting a mesh of values for \( \xi \), we might generate a collection of ordered pairs, \((\xi, \lambda(\xi))\), that can be used to develop an approximating polynomial (through regression). (This procedure is unambiguous in the sense that, for a given \( \xi \), one can explicitly compute \( \lambda(\xi) \). Hence, although one wishes to approximate the
Figure 3.5  The inverse of $\lambda$ and the fifth degree polynomial generated by fitting points $(g, \lambda^{-1}(g))$ for $g \in \{0.01, 0.02, \ldots, 0.99\}$

function $\lambda^{-1}$, none of points used for generating the approximating polynomial have error.) We used this method to produce the polynomial

$$p(g) = -6.474948 + 49.875654g - 215.036450g^2 + 484.408232g^3 - 515.706547g^4 + 210.885646g^5.$$  

$p(g)$ is within 0.1 of $\lambda^{-1}(g)$ as long as $\lambda^{-1}$ is in the interval $(-3.5, 3.5)$. (Since, in our application, $\lambda^{-1}$ produces the $z$-score of the truncation point, this is presumably the common situation in practice.) Figure 3.5 shows $p$ overlayed with $\lambda^{-1}$. Further
tinkering reveals that a seventeenth degree polynomial yields two places of accuracy for approximating $\lambda^{-1}$ as long as $\lambda^{-1}$ is in $(-3.0, 3.0)$.

For accuracy beyond the second decimal place one may use an iterative routine, such as Newton-Raphson, or a binary search where the starting value for $\xi = \lambda^{-1}(x)$ is derived from a polynomial approximation. (The author’s experience with the problem suggests that, typically, using a binary search results in faster convergence.)

Another possibility for improving the approximation of $\lambda^{-1}$ is to use a polynomial fit over a smaller part of the possible domain. The distribution of $s^2/(\alpha - d)^2$ is complicated. We can, however, get an idea of what are likely values of $s^2/(\alpha - d)^2$ through simulation. This will depend upon the part of the normal tail that is truncated. Table 3.1 was constructed by simulating 500 samples of size one hundred each using $\mu = 0$ and $\sigma = 1$. The minimum, maximum, average, and standard deviation of $s^2/(\alpha - d)^2$ are given.

Once the maximum likelihood estimates of $\mu$ and $\sigma$ are had, we may immediately compute the maximum likelihood estimates of $\eta_1$ and $\eta_2$. Theorem 9.2.1 of Bain and Englehart (1987) ensures that transformations of MLEs are MLEs of the transformed parameters. Since $\eta_1 = -1/2\sigma^2$ and $\eta_2 = \mu/\sigma^2$ we know that $\hat{\eta}_1 = -1/2\hat{\sigma}^2$ and $\hat{\eta}_2 = \hat{\mu}/\hat{\sigma}^2$ are MLEs of $\eta_1$ and $\eta_2$ respectively.

3.9 Exploring Inference for the Usual Parameterization of the Truncated Normal Distribution

Natural first extensions for Theorems 5 and 6 would be to state results for maximum likelihood estimation for parameterizations other than the “natural” one. Considering that one can usually switch between parameterizations through the application of a smooth transformation, we should expect that the nature of asymptotics for likelihood-based estimators would not fundamentally change. Estimators for the two types of pa-
rameters would still be asymptotically independent, those for the "regular" parameters would still be asymptotically normal, and that for the truncation parameter, asymptotically exponential. Further, since point null hypotheses about the vector \( \eta \) or the vector \((\eta, d)\) correspond to point null hypotheses after transformation, the discussion of likelihood ratio testing carries over directly to "other" parameterizations.

The fact that one can switch between parameterizations through the use of a smooth function lends itself to an explicit method for obtaining the asymptotic variances of the estimators of "usual" parameters in terms of the elements of \( \text{Var}(\tilde{X}) \). We may use the approximate covariance matrix of the estimators of natural parameters and the matrix of first partials of the transforming function to produce the approximate covariance matrix of MLEs in the usual parameterization. In the normal problem let \( \sigma_{\hat{\eta}_1}, \sigma_{\hat{\eta}_2} \), and \( \rho \) be the asymptotic standard deviations and correlation for \( \hat{\eta}_1 \) and \( \hat{\eta}_2 \). The delta method implies

---

Table 3.1 The behavior of \( s^2 / (\bar{x} - d)^2 \) based on 500 samples of size 100

<table>
<thead>
<tr>
<th>( d )</th>
<th>( \text{min} )</th>
<th>( \text{max} )</th>
<th>( \text{mean} )</th>
<th>( \text{std dev} )</th>
<th>( d )</th>
<th>( \text{min} )</th>
<th>( \text{max} )</th>
<th>( \text{mean} )</th>
<th>( \text{std dev} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.75</td>
<td>0.0970</td>
<td>0.3404</td>
<td>0.1852</td>
<td>0.04045</td>
<td>0.00</td>
<td>0.3918</td>
<td>1.1029</td>
<td>0.5740</td>
<td>0.0858</td>
</tr>
<tr>
<td>-2.50</td>
<td>0.1183</td>
<td>0.3268</td>
<td>0.1949</td>
<td>0.03914</td>
<td>0.25</td>
<td>0.4201</td>
<td>0.9912</td>
<td>0.6206</td>
<td>0.0941</td>
</tr>
<tr>
<td>-2.25</td>
<td>0.1154</td>
<td>0.3401</td>
<td>0.2151</td>
<td>0.03617</td>
<td>0.50</td>
<td>0.4011</td>
<td>1.0988</td>
<td>0.6634</td>
<td>0.1071</td>
</tr>
<tr>
<td>-2.00</td>
<td>0.1475</td>
<td>0.4081</td>
<td>0.2398</td>
<td>0.03904</td>
<td>0.75</td>
<td>0.4611</td>
<td>1.0805</td>
<td>0.6900</td>
<td>0.0996</td>
</tr>
<tr>
<td>-1.75</td>
<td>0.1742</td>
<td>0.4098</td>
<td>0.2736</td>
<td>0.04093</td>
<td>1.00</td>
<td>0.4464</td>
<td>1.1158</td>
<td>0.7250</td>
<td>0.1065</td>
</tr>
<tr>
<td>-1.50</td>
<td>0.1827</td>
<td>0.4598</td>
<td>0.3073</td>
<td>0.04347</td>
<td>1.25</td>
<td>0.4914</td>
<td>1.2307</td>
<td>0.7423</td>
<td>0.1149</td>
</tr>
<tr>
<td>-1.25</td>
<td>0.2188</td>
<td>0.4811</td>
<td>0.3478</td>
<td>0.04879</td>
<td>1.55</td>
<td>0.5139</td>
<td>1.2490</td>
<td>0.7736</td>
<td>0.1203</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.2356</td>
<td>0.6252</td>
<td>0.3974</td>
<td>0.05916</td>
<td>1.75</td>
<td>0.5195</td>
<td>1.3273</td>
<td>0.8089</td>
<td>0.1239</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.2710</td>
<td>0.6459</td>
<td>0.4421</td>
<td>0.06218</td>
<td>2.00</td>
<td>0.5520</td>
<td>1.3198</td>
<td>0.8206</td>
<td>0.1286</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.2902</td>
<td>0.7012</td>
<td>0.4909</td>
<td>0.06742</td>
<td>2.25</td>
<td>0.5318</td>
<td>1.4878</td>
<td>0.8412</td>
<td>0.1465</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.2726</td>
<td>0.7968</td>
<td>0.5323</td>
<td>0.07894</td>
<td>2.50</td>
<td>0.4769</td>
<td>1.7825</td>
<td>0.8548</td>
<td>0.1511</td>
</tr>
</tbody>
</table>
that the limiting distribution of $\sqrt{n}(\hat{\mu} - \mu, \hat{\sigma} - \sigma)$ has covariance matrix

$$
\begin{bmatrix}
\frac{\partial \mu}{\partial \eta_1} & \frac{\partial \mu}{\partial \eta_2} \\
\frac{\partial \sigma}{\partial \eta_1} & \frac{\partial \sigma}{\partial \eta_2}
\end{bmatrix}
\begin{bmatrix}
\sigma_{\eta_1}^2 & \rho \sigma_{\eta_1} \sigma_{\eta_2} \\
\rho \sigma_{\eta_1} \sigma_{\eta_2} & \sigma_{\eta_2}^2
\end{bmatrix}
\begin{bmatrix}
\frac{\partial \mu}{\partial \eta_1} & \frac{\partial \sigma}{\partial \eta_1} \\
\frac{\partial \sigma}{\partial \eta_2} & \frac{\partial \sigma}{\partial \eta_2}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\frac{\sigma_{\eta_1}^2}{2n} & -\frac{1}{2n} \\
\frac{1}{2n} & 0
\end{bmatrix}
\begin{bmatrix}
\sigma_{\eta_1}^2 & \rho \sigma_{\eta_1} \sigma_{\eta_2} \\
\rho \sigma_{\eta_1} \sigma_{\eta_2} & \sigma_{\eta_2}^2
\end{bmatrix}
\begin{bmatrix}
\frac{\sigma_{\eta_1}^2}{2n} & -\frac{1}{2n} \\
\frac{1}{2n} & 0
\end{bmatrix}
$$

By substituting $\sigma_{\eta_1}^2$, $\sigma_{\eta_2}^2$, and $\rho$ from the inverse of the covariance matrix of $t(X)$ evaluated at $\hat{\eta}_1$, $\hat{\eta}_2$, and $\hat{d}$ from our estimation techniques developed last section, we can obtain estimates of the variances of $\hat{\mu}$ and $\hat{\sigma}$.

Here we explore the small sample behavior of the maximum likelihood estimators, $\hat{\mu}$ and $\hat{\sigma}$, and the null behavior of the likelihood ratio test statistic for testing $H_0 : (\mu, \sigma, d) = (\mu_0, \sigma_0, d_0)$ using simulation. We note that testing $H_0 : (\mu, \sigma, d) = (\mu_0, \sigma_0, d_0)$ is the same as testing $H_0 : (\eta_1, \eta_2, d) = (-1/2\sigma_0^2, \mu_0/\sigma_0^2, d_0)$ and so, in principle, we can use either form of the density when conducting the test and presumably arrive at the same inferences.

To begin, consider the nature of confidence regions for the vector $(\mu, \sigma, d)$ obtained by inverting likelihood ratio tests. Such a region will not be the Cartesian product of a confidence interval for $d$ and a confidence region for $\mu$ and $\sigma$. Nor will it have the elliptical shape that one normally associates with inverting likelihood ratio tests when all the parameters are of the natural variety. A typical likelihood-based region is shown in Figure 3.6 (the variance is held constant to make a two-dimensional plot). The confidence region is the triangular region on the plot. There are horizontal reference lines marking the true and estimated values of $\mu$ and vertical reference lines for the true and estimated values of $d$.

The joint density for a sample of size $n$ from the truncated normal distribution is

$$
f(x_1, \ldots, x_n | \mu, \sigma, d) = \left[1 - \Phi \left( \frac{d - \mu}{\sigma} \right) \right]^{-n} (\sigma \sqrt{2\pi})^{-n} \exp \left[ -\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2} \right] 1_{\{\min(x_i) \geq d\}}.$$
The log-likelihood is

\[ L(\mu, \sigma, d) = -n \log \left[ 1 - \Phi \left( \frac{d - \mu}{\sigma} \right) \right] - n \log(\sigma) - \frac{n}{2\sigma^2} (s^2 + (\bar{x} - \mu)^2). \quad (3.29) \]

We expand \( L(\mu_0, \sigma_0, d_0) \) about the point \( L(\hat{\mu}, \hat{\sigma}, \hat{d}) \) using a Taylor series

\[
\sum_{i=1}^{n} \log f_{\mu_0, \sigma_0, d_0}(x_i)
\]

\[
= \sum_{i=1}^{n} \log f_{\mu_0, \sigma_0, d_0}(x_i) + (\mu_0 - \hat{\mu}) \sum_{i=1}^{n} \frac{\partial}{\partial \mu} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ (\sigma_0 - \hat{\sigma}) \sum_{i=1}^{n} \frac{\partial}{\partial \sigma} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ \left( \frac{1}{2} \right) (\sigma_0 - \hat{\sigma})^2 \sum_{i=1}^{n} \frac{\partial^2}{\partial \mu^2} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ \left( \frac{1}{2} \right) (\sigma_0 - \hat{\sigma})^2 \sum_{i=1}^{n} \frac{\partial^2}{\partial \sigma^2} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ (\mu_0 - \hat{\mu})(d_0 - \hat{d}) \sum_{i=1}^{n} \frac{\partial^2}{\partial \mu \partial d} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ (\mu_0 - \hat{\mu})(\sigma_0 - \hat{\sigma}) \sum_{i=1}^{n} \frac{\partial^2}{\partial \mu \partial \sigma} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ (\sigma_0 - \hat{\sigma})(d_0 - \hat{d}) \sum_{i=1}^{n} \frac{\partial^2}{\partial \sigma \partial d} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\hat{\mu}, \hat{\sigma}, \hat{d}}
\]

\[
+ \sum_{j=0}^{3} \sum_{k=0}^{j-1} (\mu_0 - \hat{\mu})^j (\sigma_0 - \hat{\sigma})^k (d_0 - \hat{d})^{3-j-k} \sum_{i=1}^{n} \frac{\partial^3}{\partial \mu^j \partial \sigma^k \partial d^{3-j-k}} \log f_{\mu_0, \sigma_0, d_0}(x_i) \bigg|_{\mu^*, \sigma^*, d^*} 
\]

where \((\mu^*, \sigma^*, d^*)\) is on the segment connecting \((\mu_0, \sigma_0, d_0)\) and \((\hat{\mu}, \hat{\sigma}, \hat{d})\).

We need the first and second partial derivatives of the log-likelihood in order to explicitly compute the various terms of the expansion. Presumably (as we showed to be the case for the natural parameterization) the remainder terms converge to zero and so we will not produce explicit formulations for the third derivatives. (We shall, in fact, demonstrate that this apparently happens by showing that the left-hand side of (3.30) is approximately equal to the sum of select terms on the right-hand side. And this approximation improves with the sample size.)
A Typical 95% Confidence Set for $(d, \alpha)$

The various partials needed to compute the linear and quadratic terms in expression (3.30) are:

\[
\frac{\partial L}{\partial \mu} = -n \phi \left( \frac{d-\mu}{\sigma} \right) \left( \frac{1}{\sigma} \right) + \frac{n(\bar{x} - \mu)}{\sigma^2},
\]

\[
\frac{\partial L}{\partial \sigma} = \frac{n \phi \left( \frac{d-\mu}{\sigma^2} \right)}{\left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]} - \frac{n}{\sigma} + \frac{n}{\sigma^3} (\sigma^2 + (\bar{x} - \mu)),
\]

\[
\frac{\partial L}{\partial d} = \frac{n \phi \left( \frac{d-\mu}{\sigma} \right) \left( \frac{1}{\sigma^2} \right)}{\left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]},
\]

\[
\frac{\partial L^2}{\partial \mu^2} = n \left\{ \frac{\left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right] \left[ n(d-\mu) \phi \left( \frac{d-\mu}{\sigma} \right) + \frac{n}{\sigma^2} \phi \left( \frac{d-\mu}{\sigma} \right) \right] \left[ \phi \left( \frac{d-\mu}{\sigma} \right) \right]^2}{\left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]^2} \right\} + \frac{n(\bar{x} - \mu)}{\sigma^2},
\]
\[ \frac{\partial L^2}{\partial \sigma^2} = n \left\{ \frac{ \Phi \left( \frac{d-\mu}{\sigma} \right) \left( \frac{2d-2\mu}{\sigma^3} + \left( \frac{\mu-d}{\sigma^2} \right)^2 \right) \Phi \left( \frac{d-\mu}{\sigma} \right) + \left[ \frac{\phi \left( \frac{d-\mu}{\sigma} \right)}{\sigma^2} \right]^2 \frac{\mu-d}{\sigma^2} } { \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]^2 } \right\} + \frac{n}{\sigma^2} - \frac{n}{\sigma^4} [\bar{x}^2 + (\bar{x} - \mu)] , \]

\[ \frac{\partial L^2}{\partial \sigma \mu} = \frac{n}{\sigma} \left\{ \frac{ \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right] \frac{\phi \left( \frac{d-\mu}{\sigma} \right)}{\sigma^2} - \left[ \frac{\phi \left( \frac{d-\mu}{\sigma} \right)}{\sigma^2} \right]^2 } { \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]^2 } \right\} - \frac{2n}{\sigma^3} (\bar{x} - \mu) , \]

\[ \frac{\partial L^2}{\partial d \mu} = \frac{n}{\sigma} \left\{ \frac{ \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right] \left( \frac{\mu-d}{\sigma^2} \right) - \left[ \frac{\phi \left( \frac{d-\mu}{\sigma} \right)}{\sigma^2} \right]^2 } { \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]^2 } \right\} , \]

and

\[ \frac{\partial L^2}{\partial d \sigma} = \frac{n}{\sigma^2} \left\{ \frac{ \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right] \left[ \frac{\phi \left( \frac{d-\mu}{\sigma} \right)}{\sigma^2} \right] } { \left[ 1 - \Phi \left( \frac{d-\mu}{\sigma} \right) \right]^2 } \right\} . \]

(3.31)

It is interesting that in this parameterization the second and higher order derivatives with respect to \( \mu \) and \( \sigma \) depend on the data (unlike in the natural parameterization).

The principal practical difficulty when applying asymptotic theory to a real problem is knowing for what sample sizes one may trust the limiting results. For asymptotic results to be useful, it is necessary for the approximations they provide to be adequate at moderate sample sizes.

Presented here are the results of several small simulations for truncated normal data. The generating distribution was, in fact, standard normal with truncation point \(-1\). Five hundred replications were made with sample sizes of 30, 100, 1000 and 10000. Several plots are also presented to illustrate the behavior of the MLE vector and the Taylor approximation to the likelihood ratio statistic for \( H_0 : (\mu, \sigma, d) = (0, 1, -1) \). The key features in Table 3.2 are as follows. The last column is twice the sum of the asymptotically non-negligible terms (the first four columns) in the Taylor expansion (twice
Table 3.2 Summary statistics for the null distribution of terms in the Taylor expansion of the LRT statistic for $H_0 : (\mu, \sigma, d) = (0, 1, -1)$

<table>
<thead>
<tr>
<th>n</th>
<th>linear d</th>
<th>quad $\mu$</th>
<th>quad $\sigma$</th>
<th>cross $\mu-\sigma$</th>
<th>cross $\mu-d$</th>
<th>2xsum</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>MIN</td>
<td>0.010</td>
<td>0.000</td>
<td>0.000</td>
<td>-28.634</td>
<td>-.769</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>5.517</td>
<td>10.172</td>
<td>20.620</td>
<td>5.078</td>
<td>5.298</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>1.207</td>
<td>1.179</td>
<td>1.560</td>
<td>-1.585</td>
<td>0.217</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>1.412</td>
<td>1.494</td>
<td>2.502</td>
<td>3.235</td>
<td>0.506</td>
</tr>
<tr>
<td>100</td>
<td>MIN</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>-26.533</td>
<td>-.043</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>10.51</td>
<td>12.66</td>
<td>19.300</td>
<td>1.701</td>
<td>1.833</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>1.115</td>
<td>1.103</td>
<td>1.179</td>
<td>-1.181</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>1.307</td>
<td>1.577</td>
<td>1.998</td>
<td>2.442</td>
<td>0.143</td>
</tr>
<tr>
<td>1000</td>
<td>MIN</td>
<td>0.005</td>
<td>0.000</td>
<td>0.000</td>
<td>-18.498</td>
<td>-.036</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>6.285</td>
<td>11.480</td>
<td>16.304</td>
<td>1.337</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>0.947</td>
<td>1.055</td>
<td>1.048</td>
<td>-1.096</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>0.943</td>
<td>1.509</td>
<td>1.610</td>
<td>2.010</td>
<td>0.016</td>
</tr>
<tr>
<td>10000</td>
<td>MIN</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>-14.732</td>
<td>-.016</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>9.185</td>
<td>12.450</td>
<td>12.474</td>
<td>1.702</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>0.933</td>
<td>1.080</td>
<td>1.010</td>
<td>-1.025</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>0.918</td>
<td>1.605</td>
<td>1.469</td>
<td>1.863</td>
<td>0.004</td>
</tr>
</tbody>
</table>

because asymptotically it is twice the sum that is $\chi^2$. The limiting null distribution of the test statistic is $\chi^2_4$. The $\chi^2_4$ distribution has mean four and standard deviation 2.83. The sum column suggests that the second order Taylor approximation to the likelihood ratio statistic has converged to its limit (by sample size 1000). Figure 3.7 shows four panels comparing values of the LRT statistic with what one would compute using only terms from the Taylor expansion that are asymptotically not zero. In the limit all points lie on the line.

Another feature of the table is the decreasing trend (in absolute value) of the terms corresponding to the mixed partials of $d$ with $\mu$ and $d$ with $\sigma$. Theory suggests that these terms converge to zero as the sample size increases. This, in fact, is what happens.

We expect that the maximum likelihood estimators of $\mu$ and $\sigma$ to be approximately normal as $n$ increases. Figures 3.8 and 3.9 display estimated sampling distributions for $\hat{\mu}$ and $\hat{\sigma}$. The histograms and normal plots are consistent with our expectations. In addition, Table 3.3 seems to suggest that somewhere between 100 and 1,000 observations,
Table 3.3 Statistics for checking normality of maximum likelihood estimators for $\mu$ and $\sigma$

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Sample Mean</th>
<th>Standard Deviation</th>
<th>Shapiro Wilk</th>
<th>p-value</th>
<th>Sample Mean</th>
<th>Standard Deviation</th>
<th>Shapiro Wilk</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>-0.28622</td>
<td>0.68145</td>
<td>0.75343</td>
<td>0</td>
<td>1.084293</td>
<td>0.30567</td>
<td>0.90071</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>-0.05444</td>
<td>0.21748</td>
<td>0.93413</td>
<td>0</td>
<td>1.01673</td>
<td>0.13184</td>
<td>0.97728</td>
<td>0.01172</td>
</tr>
<tr>
<td>1000</td>
<td>-0.00714</td>
<td>0.05899</td>
<td>0.98504</td>
<td>0.50360</td>
<td>1.00487</td>
<td>0.03973</td>
<td>0.98341</td>
<td>0.31409</td>
</tr>
<tr>
<td>10000</td>
<td>-0.00156</td>
<td>0.01856</td>
<td>0.98962</td>
<td>0.92364</td>
<td>1.000342</td>
<td>0.01284</td>
<td>0.98870</td>
<td>0.87246</td>
</tr>
</tbody>
</table>

the distributions of $\hat{\mu}$ and $\hat{\sigma}$ are indistinguishable from normal.

The maximum likelihood estimator of $d$ is expected to be exponential while the first order term of the expansion (3.17) is expected to be $\chi^2_2$. The side by side plots of Figure 3.10 demonstrate this. It is well known the sample mean and sample variance are independent unbiased estimators of $\mu$ and $\sigma^2$ based on data from an untruncated normal distribution. In Table 3.2 we see clearly that the second order cross product term for $\mu$ and $\sigma$ is consistently close to negative one. Were independence between the estimators in the problem with no truncation to translate to the truncated version of the problem, we would expect this term to be very nearly zero. From Table 3.4 we see that truncation of sixteen percent ($\Phi(-1)$) of distribution introduces a strong negative correlation between $\hat{\mu}$ and $\hat{\sigma}$. Equations (3.27) and (3.28) are in accord with this fact since one systematically adds $\theta(\bar{x} - d)^2$ to the sample variance and subtracts $\theta(\bar{x} - d)$ from the sample mean to produce $\hat{\mu}$ and $\hat{\sigma}^2$.

Perhaps the most interesting aspect of the simulations is revealed in Figure 3.11. The

Table 3.4 Estimated correlations between sample moments and between maximum likelihood estimators of $\mu$ and $\sigma$

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Correlation $\bar{x}$ and $s^2$</th>
<th>Correlation $\hat{\mu}$ and $\hat{\sigma}$</th>
<th>Sample Size</th>
<th>Correlation $\bar{x}$ and $s^2$</th>
<th>Correlation $\hat{\mu}$ and $\hat{\sigma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.390267</td>
<td>-0.84886</td>
<td>1000</td>
<td>0.421604</td>
<td>-0.72072</td>
</tr>
<tr>
<td>100</td>
<td>0.408653</td>
<td>-0.76025</td>
<td>10000</td>
<td>0.452866</td>
<td>-0.68601</td>
</tr>
</tbody>
</table>
four panels are plots of the $\chi^2_4$ distribution expected order statistics against simulated ordered values of the likelihood ratio test statistic. Even for sample sizes as small as 30, the $\chi^2_4$ distribution is a remarkably good fit. This is, perhaps, surprising given that Figure 3.7 seems to indicate there can be quite large effects from terms that, only in the limit, are zero. Apparently the extra noise from these terms is more likely to mitigate departures from the asymptotic null distribution than exacerbate them.
Figure 3.7  The likelihood ratio test statistic vs. the sum of the asymptotically non-negligible terms of the Taylor series for sample sizes $n=30, 100, 1000, \text{ and } 10000$
Figure 3.8 The sampling distribution of $\hat{\mu}$ for $n=30, 100, 1000, 10000$
Figure 3.9  The sampling distribution of $\bar{x}$ for $n=30, 100, 1000, 10000$
Figure 3.10 Side-by-side $\chi^2$ plots of $\hat{d}$ and the first order Taylor term for $d$,

$$(\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log f_{n,d} \bigg|_{(\hat{d},d)}$$

for $n=30, 100, 1000, 10000$
Figure 3.11 $\chi^2$ plots, each showing the likelihood ratio test statistic from 500 simulations where the data producing the statistic is generated from a truncated standard normal with truncation parameter $-1$ and for sample sizes $n=30, 100, 1000, \text{ and } 10000$
4 INFERENCE IN CENSORED EXPONENTIAL FAMILIES
(WITH UNKNOWN CENSORING PARAMETER)

In this chapter we consider (as a second preliminary to analysis of our motivating problem) inference based on independent and identically distributed observations from a left-censored exponential family of distributions with unknown censoring point. Inference for left censored members of the regular exponential family is similar to that for their truncated counter-parts. The technical details are more complicated, owing to the nature of the observation vector. Under censoring, a data vector typically contains observations that are "realized" values and observations that you only know to be below the censoring point. This leads, in theoretical analysis, to dealing with distributions that are, in part, absolutely continuous with respect to Lebesgue measure and, in part, absolutely continuous with respect to a counting measure on \( -\infty \).

4.1 Notation

Before proceeding, we establish notation for dealing with censored exponential families. Of particular mention is the notation for expectation, since the expected value of complete observations inherently depends on the censoring point. We resolve the issue by defining expectation for only the uncensored version of the problem. Define, for this chapter,

\[ f_n(x) \]

the density (with respect to Lebesgue measure) of a member of the
exponential family with canonical parameter vector $\eta$, the density of which is $f_{\eta}(x) = h(x)K(\eta)\exp(\eta \cdot t(x))$,

$F_{\eta}(x)$ the cumulative distribution of $f_{\eta}(x)$,

$d$ the censoring point (smallest value that can be detected),

$f_{\eta,d}(x)$ the density $f_{\eta}(x)$ with left-censoring parameter $d$,

$\eta_j$ one of $k$ "regular" parameters,

$\eta$ the vector $(\eta_1, \eta_2, \cdots, \eta_k)'$,

$\hat{\eta}_j$ the $\eta_j$-solution of the "likelihood equations",

$\hat{d}$ the minimum observation of those greater than $d$, $\hat{d} = \min\{x_i|x_i \geq d\}$,

$n_{-\infty}$ the number of censored observations in a sample of size $n$,

$n_\infty$ the number of "realized" observations in the sample of size $n$,

$p$ the probability of a non-detect, $\int_{-\infty}^{d} h(x)K(\eta)\exp(\eta \cdot t(x)) dx$,

$\hat{p} = \frac{n_{-\infty}}{n}$ the sample proportion of non-detects,

$p$ an estimate of $p$ derived from $\hat{d}$ and $\hat{\eta}$ and the form of the density,

$\int_{-\infty}^{d} h(x)K(\hat{\eta})\exp(\hat{\eta} \cdot t(x)) dx$,

$E_tj(X)$ the expected value of $t_j(X)$ for the density $f_{\eta}(x)$,

$E_{tr}t_j(X)$ the expected value of $t_j(X)$ for the truncated density,

$E_{tr}t_j(X) = (1 - p)^{-1}\int_{d}^{\infty} t_j(x)h(x)K(\eta)\exp(\eta \cdot t(x)) dx$,

$\text{Var}_{tr}t_j(X)$ the variance of $t_j(X)$ for the truncated density,

$\text{Var}_{tr}t_j(X) = (1 - p)^{-1}\int_{d}^{\infty} (t_j(x) - E_{tr}t_j(X))^2h(x)K(\eta)\exp(\eta \cdot t(x)) dx$,

$\text{Cov}_{tr}(t_j, t_m)$ the covariance of $(t_j(X), t_m(X))$ for the truncated density,

$\text{Cov}_{tr}(t_j, t_m) = (1 - p)^{-1}\int_{d}^{\infty} (t_j - E_{tr}t_j)(t_m - E_{tr}t_m)f_{\eta}(x)dx$,

$\text{Var}_{tr}t(X)$ the variance matrix of $t(X)$ for the truncated density,

$\bar{t}_j$ the sample average value of $t_j$, $\bar{t}_j = \frac{1}{n_\infty}\sum_{\{x_i|\exists x_i \geq d\}} t_j(x_i)$,
\[ \mathcal{K}(\eta) = \left( \int h(x) \exp(\eta \cdot t(x)) \, dx \right)^{-1}, \]

\[ \mathcal{K}'(\eta) = \frac{\partial \mathcal{K}(\eta)}{\partial \eta}, \]

\[ \mathcal{K}''(\eta) = \frac{\partial^2 \mathcal{K}(\eta)}{\partial \eta^2}, \]

\[ \mathcal{K}'''(\eta) = \frac{\partial^3 \mathcal{K}(\eta)}{\partial \eta^3}. \]

Exp[\beta] the exponential distribution with mean \( \beta \), the density of which is given by \( f_\beta(x) = \frac{1}{\beta} \exp \left[ \frac{-x}{\beta} \right] 1_{\{x \geq 0\}} \).

### 4.2 Censored Exponential Families

An observation from the regular exponential family with unknown threshold parameter \( d \) may take the value \( -\infty \) if it is censored (is a "non-detect") or a value \( x \in [d, \infty) \). Such an observation has density with respect to Lebesgue measure on \( \mathbb{R} \) plus a unit point mass on \( -\infty \) given by

\[
    f_{\eta,d}(x) = \left\{ \int_{-\infty}^{d} h(x) \mathcal{K}(\eta) \exp \left( \sum_{j=1}^{k} \eta_j t_j(x) \right) \, dx \right\} 1_{\{x = -\infty\}} \\
    + h(x) \mathcal{K}(\eta) \exp \left( \sum_{j=1}^{k} \eta_j t_j(x) \right) 1_{\{x \geq d\}} \tag{4.1}
\]

for \( \eta \in \mathbb{R}^k \) and \( d \in \mathbb{R} \). The log likelihood for \( n \) iid observations is then

\[
    L(\eta, d) = \begin{cases} 
        n_{-\infty} \log \left( \int_{-\infty}^{d} h(x) \mathcal{K}(\eta) \exp(\eta \cdot t(x)) \, dx \right) \\
        \quad + \sum_{\{i \in x_i \geq d\}} \sum_{j=1}^{k} \eta_j t_j(x_i) \\
        \quad + n_{\mathbb{R}} \log \mathcal{K}(\eta) + \sum_{\{i \in x_i \geq d\}} \log h(x_i) & d \leq \min x_i \\
        -\infty & d > \min x_i. \tag{4.2}
    \end{cases}
\]
4.3 Maximum Likelihood Estimators

Simple inspection of display (4.2) reveals that, provided \( n_{-\infty} < n \) (the situation we will consider hence forth), the likelihood-based estimator of \( d \) is \( \hat{d} = \min\{x; x \geq -\infty\} \). This is because the integrand appearing there is necessarily nonnegative. The greater the upper limit of integration, therefore, the larger the integral. But the largest \( d \) producing \( L(\eta, d) > -\infty \) is the smallest observed real data value.

We compute a local maximum of \( L(\eta, \hat{d}) \) by setting the \( k \) partial derivatives with respect to the entries of \( \eta \) equal to zero and solving. The solution to this system of equations is a likelihood-based estimator of \( \eta \). We hope the solution takes on an appealing intuitive form that, asymptotically, has a distribution free of the uncertainty caused by the random \( \hat{d} \).

As a preliminary, we record three results that will be needed to simplify future expressions. First, exactly as in Lemma 1

\[
E_{t_j}(X) = -\frac{\partial}{\partial \eta_j} \log \mathcal{K}(\eta)
\]

for all \( j \). Second,

\[
\text{Vart}_j(X) = -\frac{\partial^2}{\partial \eta_j^2} \log \mathcal{K}(\eta) = \frac{\partial}{\partial \eta_j} E_{t_j}(X)
\]

and

\[
\text{Cov}(t_j(X), t_m(X)) = -\frac{\partial^2}{\partial \eta_j \partial \eta_m} \log \mathcal{K}(\eta) = \frac{\partial}{\partial \eta_j} E_{t_j}(x) = \frac{\partial}{\partial \eta_m} E_{t_m}(x).
\]
Third, the conditional expectation of $t_j(X)$ given $X < d$ is

$$E[t_j(X)|X < d] = \frac{\int_{-\infty}^{d} h(x)t_j(x)\mathcal{K}(\eta) \exp[\eta \cdot t(x)] \, dx}{\Pr(X < d)}$$

$$= \frac{\int_{-\infty}^{d} h(x)t_j(x)\mathcal{K}(\eta) \exp[\eta \cdot t(x)] \, dx}{\int_{-\infty}^{d} h(x)\mathcal{K}(\eta) \exp[\eta \cdot t(x)] \, dx}$$

$$= \frac{\int_{-\infty}^{d} h(x)t_j(x) \exp[\eta \cdot t(x)] \, dx}{\int_{-\infty}^{d} h(x) \exp[\eta \cdot t(x)] \, dx}.$$

This last "result" is, essentially, by definition.

The first partial derivative of the log-likelihood (4.2) with respect to $\eta_j$ is

$$\frac{\partial L(\eta, d)}{\partial \eta_j} = n_{-\infty} \left( \int_{-\infty}^{d} h(x)\mathcal{K}'(\eta) \exp[\eta \cdot t(x)] + h(x)t_j(x)\mathcal{K}(\eta) \exp[\eta \cdot t(x)] \, dx \right)$$

$$+ \frac{n_{\infty}}{\mathcal{K}(\eta)} + \sum_{\{x_1, x_2, \ldots\}} t_j(x_i)$$

$$= n \left( \frac{n_{\infty}}{n} \right) \bar{t}_j + n \left( \frac{n_{-\infty}}{n} \right) \left[ \left[ \frac{\mathcal{K}'(\eta)}{\mathcal{K}(\eta)} \int_{-\infty}^{d} h(x)\mathcal{K}(\eta) \exp[\eta \cdot t(x)] \, dx \right] \right.$$  

$$\left. + \left[ \int_{-\infty}^{d} h(x)\mathcal{K}(\eta) t_j(x) \exp[\eta \cdot t(x)] \, dx \right] \right] - n \left( \frac{n_{\infty}}{n} \right) \left( - \frac{\partial}{\partial \eta_j} \log \mathcal{K}(\eta) \right)$$

$$= n(1 - \hat{p})\bar{t}_j - n\hat{p} \left( - \frac{\partial}{\partial \eta_j} \log \mathcal{K}(\eta) \right) + n\hat{p}E[t_j(X)|X < d]$$

$$- n(1 - \hat{p}) \left( - \frac{\partial}{\partial \eta_j} \log \mathcal{K}(\eta) \right)$$

$$= n(1 - \hat{p})\bar{t}_j - n\hat{p}E[t_j(X) + n\hat{p}E[t_j(X)|X < d] - n(1 - \hat{p})Et_j(X)$$

$$= n(1 - \hat{p})\bar{t}_j + n\hat{p}E[t_j(X)|X < d] - nEt_j(X).$$
Setting this to zero we produce the equation

\[ \hat{p}E[t_j(X)|X < d] + (1 - \hat{p})\bar{t}_j = E_{t_j}(X). \]  

(4.3)

So an \( \eta \)-solution to the system of \( k \) equations (4.3) assumes an intuitively appealing form. That is, for any random variable \( Y \) and any function \( s \)

\[ E_s(Y) = \Pr(Y < d)E[s(Y)|Y < d] + \Pr(Y \geq d)E[s(Y)|Y \geq d]. \]  

(4.4)

Now, \( \bar{t}_j \) should approximate \( E_{t_j}(X) \), and \( \hat{p} \) should approximate \( p \). So equation (4.3) says that \( \eta \) should be chosen so that for each \( j \) an empirical approximation to \( E_{t_j}(X) \) should equal \( E_{t_j}(X) \).

To demonstrate that \( \eta \)-solutions to (4.3) are local maximizers of \( L(\eta, d) \), we again appeal to convexity. If the matrix of second partials of \( L(\eta, d) \) with respect to \( \eta \) can be shown to be negative definite, then the log-likelihood is convex in \( \eta \) and so any solution of the likelihood equations must maximize \( L(\eta, d) \). (The fact that \( d \) is unknown is unimportant because it does not affect the basic shape of the likelihood surface. Since \( d \) may be any real value, there is, for any given \( \hat{d} \) a fixed \( d \) problem for which the “known” censoring point is the particular value of \( \hat{d} \).)

It is a simple matter to demonstrate that second partials of the log-likelihood exist and are continuous, for we can easily compute them.

\[ \frac{\partial^2}{\partial \eta_j^2}L(\eta, d) = \frac{\partial}{\partial \eta_j} \left[ n(1 - \hat{p})\bar{t} + npE[t_j(X)|X < d] - nE_{t_j}(X) \right] \]
\[ = n \left( -\frac{\partial}{\partial \eta_j}E_{t_j}(X) + \frac{\partial}{\partial \eta_j}\hat{p}E[t_j(X)|X < d] \right) \]
\[ = n \left( -\text{Var}_{t_j}(X) + \hat{p}\text{Var}[t_j(X)|X < d] \right) \]  

(4.5)

The final expression follows as in Lemma 1. Likewise

\[ \frac{\partial^2}{\partial \eta_j \partial \eta_m}L(\eta, d) = n \left( -\frac{\partial}{\partial \eta_m}E_{t_j}(X) + \frac{\partial}{\partial \eta_m}\hat{p}E[t_j(X)|X < d] \right) \]
\[ = n \left( -\text{Cov}(t_j(X), t_m(X)) + \hat{p}\text{Cov}[t_j(X), t_m(X)|X < d] \right) \]  

(4.6)
and, for completeness,

\[
\frac{\partial}{\partial \eta_j} \frac{\partial}{\partial d} L(\eta, d) = n \hat{p} \frac{\partial}{\partial d} E [t_j(X) | X < d]
\]

\[
= n \hat{p} \left\{ \left( \int_{-\infty}^{d} \mathcal{K}(\eta) h(x) \exp[\eta \cdot t(x)] \, dx \right) t_j(d) h(d) \mathcal{K}(\eta) \exp[\eta \cdot t(d)] \right\}
\]

\[
- \left\{ \left( \int_{-\infty}^{d} \mathcal{K}(\eta) h(x) \exp[\eta \cdot t(x)] \, dx \right)^2 \right\}
\]

\[
= n \left( \frac{\hat{p}}{p} \right) t_j(d) f_{n,d}(d) - n \left( \frac{\hat{p}}{p} \right) f_{n,d}(d) E [t_j(X) | X < d]
\]

\[
= n \left( \frac{\hat{p}}{p} \right) f_{n,d}(d) \{ t_j(d) - E [t_j(X) | X < d] \}.
\]  

Combining (4.5) and (4.6), the matrix of second partials of the log-likelihood is

\[
\left[ \frac{\partial^2 L(\eta, d)}{\partial \eta^2} \right] = -n \left( \text{Var}(X) - \hat{p} \text{Var}[t(X) | X < d] \right)
\]  

where \( \text{Var}[t(X) | X < d] \) is

\[
\begin{bmatrix}
\text{Var}[t_1 | X < d] & \text{Cov}(t_1, t_2 | X < d) & \cdots & \text{Cov}(t_1, t_k | X < d) \\
\text{Cov}(t_2, t_1 | X < d) & \text{Var}[t_2 | X < d] & \cdots & \text{Cov}(t_2, t_k | X < d) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(t_k, t_1 | X < d) & \text{Cov}(t_k, t_2 | X < d) & \cdots & \text{Var}[t_k | X < d]
\end{bmatrix}
\]

and \( \text{Var}(X) \) is the unconditional covariance matrix.

To demonstrate that the matrix (4.8) is asymptotically negative definite we show that

\[
\text{Var}(X) - \hat{p} \text{Var}[t(X) | X < d]
\]  

is positive definite. To show this we recall (see Bain and Englehart (1987, p. 182)) that

\[
\text{Var}(X) = E \text{Var}[t(X) | 1_{\{X < d\}}] + \text{Var} E[t(X) | 1_{\{X < d\}}]
\]
and
\[
\mathbb{E}\hat{t}(X) = \mathbb{E}[\hat{t}(X)|1_{\{X < d\}}].
\]

From these and (4.4) we can write
\[
\text{Var}\hat{t}(X) = p\text{Var}[\hat{t}(X)|1_{\{X < d\}}]|X < d] + (1 - p)\text{Var}[\hat{t}(X)|1_{\{X < d\}}]|X \geq d]
\]
\[
+ p\mathbb{E}[(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))'(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))|X < d]
\]
\[
+ (1 - p)\mathbb{E}[(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))'(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))|X \geq d]
\]
\[
= p\text{Var}[\hat{t}(X)|X < d] + (1 - p)\text{Var}[\hat{t}(X)|X \geq d]
\]
\[
+ p\mathbb{E}[(E[\hat{t}(X)|X < d] - E\hat{t}(X))'(E[\hat{t}(X)|X < d] - E\hat{t}(X))]
\]
\[
+ (1 - p)\mathbb{E}[(E[\hat{t}(X)|X \geq d] - E\hat{t}(X))'(E[\hat{t}(X)|X \geq d] - E\hat{t}(X))].
\] (4.10)

It is clear that
\[
\mathbb{E} [\text{Var}[\hat{t}(X)|1_{\{X < d\}}]|X < d] = \text{Var}[\hat{t}(X)|X < d]
\]
and
\[
\mathbb{E} [\text{Var}[\hat{t}(X)|1_{\{X < d\}}]|X \geq d] = \text{Var}[\hat{t}(X)|X \geq d]
\]
since the evaluation of the indicator function based on the conditioning criteria is always one in this first case and zero in the second case. From this point we can write (4.10) as
\[
\text{Var}\hat{t}(X) = p\text{Var}[\hat{t}(X)|X < d] + (1 - p)\text{Var}[\hat{t}(X)|X \geq d]
\]
\[
+ p\mathbb{E}[(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))'(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))|X < d]
\]
\[
+ (1 - p)\mathbb{E}[(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))'(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))|X \geq d]
\]
\[
= p\text{Var}[\hat{t}(X)|X < d] + (1 - p)\text{Var}[\hat{t}(X)|X \geq d]
\]
\[
+ p\mathbb{E}[(E[\hat{t}(X)|X < d] - E\hat{t}(X))'(E[\hat{t}(X)|X < d] - E\hat{t}(X))]
\]
\[
+ (1 - p)\mathbb{E}[(E[\hat{t}(X)|X \geq d] - E\hat{t}(X))'(E[\hat{t}(X)|X \geq d] - E\hat{t}(X))].
\] (4.11)

By realizing that
\[
\mathbb{E} [(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))'(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))|X < d]
\]
\[
= (E[\hat{t}(X)|X < d] - E\hat{t}(X))'(E[\hat{t}(X)|X < d] - E\hat{t}(X))
\]
and
\[
\mathbb{E} [(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))'(E[\hat{t}(X)|1_{\{X < d\}}] - E\hat{t}(X))|X \geq d]
\]
\[
= (E[\hat{t}(X)|X \geq d] - E\hat{t}(X))'(E[\hat{t}(X)|X \geq d] - E\hat{t}(X))
\]
we eventually arrive at the expression

\[ \text{Var}(X) = p \text{Var}[\ell(X)|X < d] + (1 - p) \text{Var}[\ell(X)|X \geq d] \]
\[+ p(1 - p)(E[\ell(X)|X < d] - E[\ell(X)|X \geq d])' \]
\[= (E[\ell(X)|X < d] - E[\ell(X)|X \geq d]). \quad (4.12) \]

Hence, \( \text{Var}(X) - pE \text{Var}[\ell(X)|X < d] \) is simply the sum of \( (1 - p) \text{Var}[\ell(X)|X \geq d] \) and

\[ p(1 - p)(E[\ell(X)|X < d] - E[\ell(X)|X \geq d])' (E[\ell(X)|X < d] - E[\ell(X)|X \geq d]) \]

which is necessarily positive since \( p \in (0, 1) \).

### 4.4 Unknown Censoring Point Theorems

Our purpose is to study the effect of censoring on inference in the exponential family. We first need to establish probability results that guarantee that under left censoring, an average of uncensored observations, the count of the number of terms averaged and the minimum uncensored observation are asymptotically mutually independent. This being accomplished, we will demonstrate that the likelihood-based estimators in the censored exponential family are, in the limit, smooth functions of the average or the minimum, but not both. Thus, asymptotic independence of estimators of \( \eta \) and \( d \) will be demonstrated, leaving us only to show how to use the information from our analysis for the purposes of testing and constructing confidence regions.

**Theorem 8** Asymptotic Independence of a Random Index and the Corresponding Element of a Sequence Converging in Distribution Suppose \( X_n \) is \( l \)-dimensional. Assume \( X_n \) converges in distribution to \( X \). Assume further that \( B_n \) takes nonnegative integer values, \( B_n \xrightarrow{\text{w.p.}} \infty \), and that \( Z_n \) is a one-to-one transformation of \( B_n \) such that \( Z_n \xrightarrow{\text{D}} Z \) for some random variable \( Z \). Then if \( Y_n \) is such
that conditional on $B_n = b$, $Y_n$ has the same distribution as $X_b$

$$(Z_n, Y_n) \xrightarrow{D} (Z, X)$$

where $Z$ and $X$ are independent.

Proof of Theorem 8

Let $\psi_n(t) \equiv E \exp(it'X_n)$ be the characteristic function of $X_n$ and let $\psi(t) \equiv E \exp(it'X)$ be the characteristic function of $X$. It is standard (for reference see Billingsley (1986)), that that $\psi_n(t) \rightharpoonup \psi(t)$. Let $\Psi_n(s, t) = E \exp(isZ_n + it'Y_n)$. Then

$$\Psi_n(s, t) = E \left[ E \exp(isZ_n + it'Y_n) \right| B_n]$$

$$= E \exp(isZ_n)E \exp(it'Y_n|B_n)$$

$$= E \exp(isZ_n)\psi_{B_n}(t)$$

$$= E \exp(isZ_n)\psi(t) + E \exp(isZ_n)[\psi_{B_n}(t) - \psi(t)].$$

Now,

$$|E_{B_n}(\exp(isZ_n)\psi_{B_n}(t) - \psi(t))| \leq E_{B_n} |\exp(isZ_n)\{\psi_{B_n}(t) - \psi(t)\}|$$

$$= E_{B_n} |\psi_{B_n}(t) - \psi(t)|$$

$$\to 0.$$ 

This convergence follows since $B_n \to \infty$ w.p. 1. This implies that $|\psi_{B_n}(t) - \psi(t)| \to 0$ almost surely. Therefore, by the Lebesgue Dominated Convergence Theorem, we have $E|\psi_{B_n}(t) - \psi(t)| \to 0$. Thus, $\Psi_n(s, t) \rightharpoonup E \exp(isZ)\psi(t)$ which implies asymptotic independence. \qed
Theorem 9 Asymptotic Independence of the Number Censored, the Minimum and the Standardized "Data Vector" for the Uncensored Observations

Assume \(x_1, \ldots, x_n\) is an iid sample from family (4.1) where \(h(x)\) is positive and continuous. Let

\[ T_j = \sqrt{n_{\alpha}}(\hat{t}_j - E_{\alpha}(\hat{t}_j(X))) \quad \text{and} \quad T = (T_1, T_2, \cdots, T_k). \]

Then

\[ \frac{n_{\rightarrow \infty} - np}{\sqrt{np(1-p)}}, \quad T, \quad \text{and} \quad n_{\alpha}(\hat{d} - d) \]

are asymptotically independent,

\[ \frac{n_{\rightarrow \infty} - np}{\sqrt{np(1-p)}} \xrightarrow{D} N(0, 1), \]

\[ T \xrightarrow{D} N(0, \text{Var}_{\alpha}(\hat{t}(X))), \]

and \(n_{\alpha}(\hat{d} - d) \xrightarrow{D} \text{Exp}(1)\).

**PROOF of Theorem 9**

Conditional on \(x_i > -\infty, \quad x_i\) is from the truncated exponential family considered in Chapter 3. Therefore, conditioned on \(n_{\alpha}\), \((T, n_{\alpha}(\hat{d} - d))\) has the joint distribution of the standardized "data vector" and standardized minimum of a sample of size \(n_{\alpha}\) from the truncated exponential family (3.2). By the analysis of Chapter 3, these conditional joint distributions converge to the product of a \(N(0, \text{Var}_{\alpha}(\hat{t}(X)))\) distribution and an \text{Exp}(1) distribution.

Now \(n_{\alpha} \to \infty\) w.p. 1 and, since \(n_{\rightarrow \infty} = n - n_{\alpha}\),

\[ Z_n = \frac{n_{\rightarrow \infty} - np}{\sqrt{np(1-p)}} \]

is a one-to-one transformation of \(n_{\alpha}\). \(Z_n\) converges in distribution to \(Z\), the standard normal. So applying Theorem 8 with \(B_n = n_{\alpha}\), the result follows. \(\Box\)
Theorem 10 Asymptotic Independence of \( \hat{\eta} \) and \( \hat{d} \) for Members of the Censored Exponential Families

Suppose \( x_1, x_2, \ldots, x_n \) are iid from distribution (4.1). Further assume that \( \hat{\eta} \) solves the system of equations for \( j \in \{1, 2, \cdots, k\} \) given by

\[
\hat{p} E_\eta [t_j(X) | X < d] + (1 - \hat{p}) \hat{t}_j = E_\eta t_j(X).
\]

Assuming that \( \hat{\eta} \) is consistent, \( \hat{\eta} \) and \( \hat{d} \) are asymptotically independent. In addition, \( \sqrt{n}(\hat{\eta} - \eta) \) are asymptotically normal with mean 0 and covariance matrix the inverse of the Fisher information matrix for fixed \( d \).

**PROOF of Theorem 10**

For \( j = 1, 2, \cdots, k \) define the function \( \Gamma_j(\eta, d, p, \tilde{t}_j) : \mathbb{R}^{k+3} \to \mathbb{R} \) so that

\[
\Gamma_j(\eta, d, p, \tilde{t}_j) = p h_j(\eta, d) - g_j(\eta) + (1 - p) \tilde{t}_j
\]

where

\[
h_j(\eta, d) = E_\eta [t_j(X) | X < d] \quad \text{and} \quad g_j(\eta) = E_\eta t_j(X).
\]

Further, let

\[
h(\eta, d) = E_\eta [t(X) | X < d] \quad \text{and} \quad g(\eta) = E_\eta t(X).
\]

Letting \( \Gamma : \mathbb{R}^{2k+2} \to \mathbb{R}^k \) be the vector function defined by the \( k \) different \( \Gamma_j \)'s, we have by (4.3) that \( \Gamma(\hat{\eta}, \hat{d}, \hat{\rho}, \tilde{t}) = 0 \). Suppose that \( \hat{\eta} \) is such that \( \Gamma(\hat{\eta}, \hat{d}, \hat{\rho}, \tilde{t}) = 0 \). For \( j \in \{1, 2, \cdots, k\} \)

\[
\Gamma_j(\hat{\eta}, \hat{d}, \hat{\rho}, \tilde{t}_j) - \Gamma_j(\eta, d, \hat{\rho}, \tilde{t}_j) = \Gamma_j(\hat{\eta}, \hat{d}, \hat{\rho}, \tilde{t}_j) - \Gamma_j(\hat{\eta}, d, \hat{\rho}, \tilde{t}_j)
\]

\[
+ \Gamma_j(\hat{\eta}, \hat{d}, \hat{\rho}, \tilde{t}_j) - \Gamma_j(\eta, d, \hat{\rho}, \tilde{t}_j).
\]

(4.13)

By the Mean Value Theorem and substitution, equation (4.13) may be written as

\[
0 - (\hat{p} h_j(\eta, d) - g_j(\eta)) + (1 - \hat{p}) \tilde{t}_j = \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m) \left[ \frac{\partial \Gamma_j(\eta^*_m, d, \hat{\rho}, \tilde{t}_j)}{\partial \eta_m} \right]
\]

\[
+ \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m) \left[ \frac{\partial \Gamma_j(\eta^*_m, d, \hat{\rho}, \tilde{t}_j)}{\partial \eta_m} \right].
\]
Or,

\[-\sqrt{n}\{((1 - \hat{p})\bar{\xi}_j - [g_j(\eta) - \hat{p} h_j(\eta_d)]\}

= \sqrt{n} \left( \sum_{m=1}^{k} (\hat{\eta}_m - \hat{\eta}_m) \left[ \frac{\partial \Gamma_j(\eta, d, \hat{p}, \bar{\xi}_j)}{\partial \eta_m} \right] \right)

+ \sqrt{n} \left( \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m) \left[ \frac{\partial \Gamma_j(\eta, d, \hat{p}, \bar{\xi}_j)}{\partial \eta_m} \right] \right) \tag{4.14}

where \(\eta_j^*\) is on the line segment between \(\hat{\eta}\) and \(\hat{\eta}\) and \(\eta_j^{**}\) is on the segment between \(\eta\) and \(\hat{\eta}\). We may express \(\Gamma_j(\hat{\eta}, d, \hat{p}, \bar{\xi}_j)\) two different ways,

\[\Gamma_j(\hat{\eta}, d, \hat{p}, \bar{\xi}_j) = \Gamma_j(\hat{\eta}, d, \hat{p}, \bar{\xi}_j) + \sum_{m=1}^{k} (\hat{\eta}_m - \hat{\eta}_m) \left[ \frac{\partial \Gamma_j(\eta_j^*, d, \hat{p}, \bar{\xi}_j)}{\partial \eta_m} \right] \tag{4.15}\]

and

\[\Gamma_j(\hat{\eta}, d, \hat{p}, \bar{\xi}_j) = \Gamma_j(\hat{\eta}, \hat{d}, \hat{p}, \bar{\xi}_j) + (d - \hat{d}) \left[ \frac{\partial \Gamma_j(\eta, \hat{d}^*, \hat{p}, \bar{\xi}_j)}{\partial d} \right] \tag{4.16}\]

for some \(d^* \in (d, \hat{d})\). Recalling \(\Gamma_j(\hat{\eta}, \hat{d}, \hat{p}, \bar{\xi}_j) = \Gamma_j(\hat{\eta}, d, \hat{p}, \bar{\xi}_j)\), we combine (4.15) and (4.16) and (multiplying each side by \(\sqrt{n}\)) produce the equation

\[\sqrt{n} \sum_{m=1}^{k} (\hat{\eta}_m - \hat{\eta}_m) \left[ \frac{\partial \Gamma_j(\eta_j^*, d, \hat{p}, \bar{\xi}_j)}{\partial \eta_m} \right] = \sqrt{n}(d - \hat{d}) \left[ \frac{\partial \Gamma_j(\eta, \hat{d}^*, \hat{p}, \bar{\xi}_j)}{\partial d} \right]. \tag{4.17}\]

Taken together equations (4.14) and (4.17) give us

\[-\sqrt{n}\{((1 - \hat{p})\bar{\xi}_j - [g_j(\eta) - \hat{p} h_j(\eta, d)]\}

= \sqrt{n}(d - \hat{d}) \left[ \frac{\partial \Gamma_j(\eta, \hat{d}^*, \hat{p}, \bar{\xi}_j)}{\partial d} \right]

+ \sqrt{n} \left( \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m) \left[ \frac{\partial \Gamma_j(\eta_j^{**}, d, \hat{p}, \bar{\xi}_j)}{\partial \eta_m} \right] \right). \tag{4.18}\]

Now \(\sqrt{n}(d - \hat{d}) = \frac{\sqrt{n}}{\sqrt{n}\sqrt{n}}n \hat{a}(d - \hat{d})\) converges in probability to 0. As we have shown in equation (4.7)

\[\frac{\partial}{\partial d} E[t_j(X)|X < d] = \frac{1}{p} \int_{n,d}(d) \{t_j(d) - E[t_j(X)|X < d]\}.\]
So the consistency of $\hat{\theta}$, $\hat{p}$, and $\hat{\eta}$ imply $\partial \Gamma_j(\hat{\eta}_j, d^*, \hat{p}, \tilde{t}_j)/\partial d$ converges to

$$f_{\eta,d}(d) \{t_j(d) - E \{t_j(X)|X < d\}\},$$

a constant. Hence the first term on the right-hand side of (4.18) converges to zero in probability for all $j \in \{1, 2, \cdots, k\}$ and the large sample behavior of

$$\sqrt{n} \left( \sum_{m=1}^{k} (\hat{\eta}_m - \eta_m) \left[ \frac{\partial \Gamma_j(\eta_j^*, d, \hat{p}, \tilde{t}_j)}{\partial \eta_m} \right] \right)$$

is that of

$$-\sqrt{n} \{(1 - \hat{p})\tilde{t}_j - [g_j(\eta) - \hat{p} h_j(\eta, d)]\}. \quad (4.19)$$

Collecting these into matrix form, we have that the asymptotic behavior of

$$\sqrt{n} \begin{pmatrix} \frac{\partial \Gamma_1(\eta_1^*, d, \hat{p}, \tilde{t}_1)}{\partial \eta_1} & \cdots & \frac{\partial \Gamma_1(\eta_1^*, d, \hat{p}, \tilde{t}_1)}{\partial \eta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Gamma_k(\eta_k^*, d, \hat{p}, \tilde{t}_k)}{\partial \eta_1} & \cdots & \frac{\partial \Gamma_k(\eta_k^*, d, \hat{p}, \tilde{t}_k)}{\partial \eta_k} \end{pmatrix} (\hat{\eta} - \eta) \quad (4.20)$$

is that of

$$-\sqrt{n}((1 - \hat{p})\tilde{t} - g(\eta) + \hat{p}h(\eta, d)).$$

Now

$$\frac{\partial \Gamma_j(\eta_j^*, d, \hat{p}, \tilde{t}_j)}{\partial \eta_m} \rightarrow \frac{\partial \Gamma_j(\eta_j, d, p, E_{tr} t_j(X))}{\partial \eta_m}$$

in probability. Also,

$$\frac{\partial \Gamma_j(\eta_j, d, p, E_{tr} t_j(X))}{\partial \eta_m} = p \frac{\partial}{\partial \eta_m} E\{t_j(X)|X < d\} - \frac{\partial}{\partial \eta_m} E t_j(X)$$

$$= p \text{Cov}[t_j(X), t_m(X)|X < d] - \text{Cov}[t_j(X), t_m(X)].$$

So the matrix in (4.20) converges in probability to

$$-(\text{Var}(X) - p\text{Var}[t(X)|X < d]).$$
Therefore the asymptotic behavior of $\sqrt{n}(\hat{\eta} - \eta)$ is that of

$$(\text{Var}_t(X) - p\text{Var}_t(X)|X < d)^{-1} \sqrt{n}((1 - \hat{p})\bar{t} - g(\eta) + \hat{p}h(\eta, d)).$$

Now,

$$\sqrt{n}((1 - \hat{p})\bar{t} - g(\eta) + \hat{p}h(\eta, d))$$

$$= \sqrt{n}(1 - \hat{p})\left(\bar{t} - \frac{1}{1 - p}(g(\eta) - ph(\eta, d))\right)$$

$$+ \sqrt{n}(1 - \hat{p})\left[\frac{1}{1 - p} - \frac{1}{1 - \hat{p}}\right]g(\eta) + \sqrt{n}(1 - \hat{p})\left[-\frac{p}{1 - p} + \frac{\hat{p}}{1 - \hat{p}}\right]h(\eta, d)$$

$$= \sqrt{n}\left[\frac{n - n_{\infty}}{n^2}\right]^2\left(\bar{t} - \frac{1}{1 - p}(g(\eta) - ph(\eta, d))\right)$$

$$+ \sqrt{n}\left(\left[\frac{1}{1 - \hat{p}}\right] - 1\right)g(\eta) + \sqrt{n}\left(\hat{p} - p\left[\frac{1}{1 - p}\right]\right)h(\eta, d)$$

$$= \frac{n_1}{n}\sqrt{n_2}(\bar{t} - E_t(X)) + \frac{1}{1 - p}\sqrt{n}(\hat{p} - p)[h(\eta, d) - g(\eta)]. \quad (4.21)$$

Thus,

$$\sqrt{\frac{n_1}{n}} \to \sqrt{1 - p}$$

in probability. Hence the asymptotic behavior of (4.21) is that of

$$\sqrt{1 - p}\sqrt{n_1}(\bar{t} - E_t(X)) + \sqrt{\frac{1}{1 - p}\sqrt{n}(\hat{p} - p)}(h(\eta, d) - g(\eta)). \quad (4.22)$$

The first and second of these terms are asymptotically independent by virtue of Theorem 9, the first being asymptotically $N(0, (1 - p)\text{Var}_t(X))$ and the second being asymptotically $k$-dimensional normal with mean $0$ and covariance matrix

$$\frac{p}{1 - p}\begin{pmatrix}
(h_1(\eta, d) - g_1(\eta))^2 & \cdots & (h_1(\eta, d) - g_1(\eta))(h_k(\eta, d) - g_k(\eta)) \\
\vdots & \ddots & \vdots \\
(h_k(\eta, d) - g_k(\eta))(h_1(\eta, d) - g_1(\eta)) & \cdots & (h_k(\eta, d) - g_k(\eta))^2
\end{pmatrix}.$$
we have

\[(h_j(\eta, d) - g_j(\eta)) (h_m(\eta, d) - g_m(\eta)) = (1 - p)^2 (E[t_j(X)|X < d] - E[t_j(X)|X ≥ d]) \times (E[t_m(X)|X < d] - E[t_m(X)|X ≥ d]).\]

That is, the second term in (4.22) is asymptotically \(k\)-dimensional normal with mean 0 and covariance matrix the \(k \times k\) matrix

\[p(1 - p) \left[ (E[t(X)|X < d] - E[t(X)|X ≥ d])' (E[t(X)|X < d] - E[t(X)|X ≥ d]) \right].\]

(4.23)

So the sum (4.22) is asymptotically \(k\)-dimensional normal with mean 0 and covariance matrix the sum of \((1 - p)\text{Var}_r t(X)\) and (4.23). By (4.12) we know that this sum is \(\text{Var} t(X) - p \text{Var}[t(X)|X < d]\).

So finally, the limiting distribution of \(\sqrt{n}(\hat{\eta} - \eta)\) is \(k\)-variate normal, with mean 0 and covariance matrix

\[(\text{Var} t(X) - p \text{Var}[t(X)|X < d])^{-1}.
\]

Independence of \(\sqrt{n}(\hat{\eta} - \eta)\) and \(\sqrt{n}(\hat{d} - d)\) follows along the same lines as in Chapter 3. We can write

\[
\begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix}
= \begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix} + \begin{pmatrix}
\sqrt{n}(\hat{\eta} - \hat{\eta}) \\
0
\end{pmatrix}
\]

with the proof showing that \(\sqrt{n}(\hat{\eta} - \hat{\eta})\) converges to 0 in probability. So

\[
\begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix}
\text{ and } \begin{pmatrix}
\sqrt{n}(\hat{\eta} - \eta) \\
n(\hat{d} - d)
\end{pmatrix}
\]

have the same limiting distribution. But \(\hat{\eta}\) is a function of \(\hat{\eta}\) and \(\hat{d}\) through solving \(\Gamma(\cdot, d, \hat{d}, \hat{\eta}) = 0\) (an \(\mathbb{R}^{k+1}\) to \(\mathbb{R}^k\) function defined by setting \(\Gamma(\eta, d, \hat{d}, \hat{\eta}) = 0\) and letting \(\eta\)

depend on $\hat{p}$ and $\hat{t}$). The smoothness of $\Gamma$ implies that $\sqrt{n}(\hat{\eta} - \eta)$ is essentially a linear function of
\[
\sqrt{n}\begin{pmatrix} \hat{t} - E_t \hat{t} \\ \hat{p} - p \end{pmatrix}
\]
which, by Theorem 9, is asymptotically independent of $n(d - d)$. □

### 4.5 Likelihood Ratio Tests

The asymptotics of likelihood ratio testing for distributions with an unknown threshold parameter are (virtually) identical to those with an unknown truncation point. In both cases the asymptotic null distribution of the LR test statistic for a point null hypothesis concerning the entire parameter vector $(\eta, d)$ is $\chi^2_{k+2}$. This is because the asymptotic independence between the estimators of the threshold and the "regular" parameters allows us to tackle the two parts of a Taylor expansion separately and combine the results additively. Also, the asymptotic independence between the estimators relies (at the core) on the rate of convergence of the sample minimum being $O(n^{-1})$ while that of the estimators of "regular" parameters is $O(n^{-\frac{1}{2}})$.

For concreteness, we consider the hypotheses
\[
H_0 : (\eta, d) = (\eta_0, d_0)
\]
\[
H_\alpha : \text{not } H_0.
\]
and the corresponding likelihood ratio test statistic
\[
\lambda(\eta_0, d_0) = 2 \log \left[ \Lambda(\eta_0, d_0) \right] = 2 \log \left\{ \sup_{(\eta, d) \in \Theta_0} \prod_{i=1}^{n} f_{\eta, d}(x_i) \right\} = 2 \log \left\{ \sup_{(\eta, d) \in \Theta_0} \prod_{i=1}^{n} f_{\eta, d}(x_i) \right\}.
\]

We may, as before, express (4.24) using the Taylor expansion given in (3.16).
The regular portion of the MLE behaves identically in the truncation and censoring problems. That is to say $\sqrt{n}(\hat{\eta} - \eta)$ is asymptotically normal with mean 0 and covariance matrix equal to the inverse of the Fisher information

$$\left\{ \frac{1}{n} \left[ \frac{\partial^2 L(\eta, d)}{\partial \eta^2} \right] \right\}^{-1}. \quad (4.25)$$

This property makes for the necessary cancellation in expression (3.19) to cause the corresponding part of the Taylor expansion to have an asymptotic $\chi^2_k$ distribution along the lines outlined by Johnson and Wichern (1992, p. 140).

Regardless if the observations are censored or truncated, the first order threshold term in Taylor expansion (3.16) is distributed as a $\frac{1}{2} \chi^2_2$ (or Exp(1)) random variable in the limit.

**Theorem 11 (Distribution of the linear term for the censoring parameter in a Taylor expansion of $L(\eta, d)$)** Assume $x_1, \ldots, x_n$ are independent and identically distributed from a distribution with density (4.1). If $\hat{\eta}$ is consistent, then

$$(\hat{d} - d) \sum_{i=1}^{n} \left. \frac{\partial}{\partial \eta} \log f_{\hat{\eta}, d}(x_i) \right|_d$$

converges in distribution to an exponential random variable with mean one.

**PROOF of Theorem 11**

We write the variable (4.26) as

$$\left. (\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial \eta} \log f_{\hat{\eta}, d}(x_i) \right|_d = \left. (\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial \eta} \log \left\{ F_{\hat{\eta}}(d) 1_{\{x_i = -\infty\}} + f_{\hat{\eta}}(x) 1_{\{x_i \geq d\}} \right\} \right|_d$$

$$= n_{-\infty} (\hat{d} - d) \left[ \frac{f_{\hat{\eta}}(d)}{F_{\hat{\eta}}(d)} \right]. \quad (4.27)$$

Now

$$\frac{f_{\hat{\eta}}(\hat{d})}{F_{\hat{\eta}}(\hat{d})} \xrightarrow{p} \frac{f_{\eta}(d)}{F_{\eta}(d)}$$
by the right continuity of \( f_\eta(x) \) in \( x \), the continuity of \( f_\eta(x) \) in \( \eta \), the continuity of \( F_\eta(d) \) in \((\eta, d)\) and the consistency of both \( \hat{\eta} \) and \( \hat{d} \). Further,

\[
n_{-\infty} (\hat{d} - d) = \frac{n_{-\infty}}{n_\mathbb{R}} n_\mathbb{R} (\hat{d} - d)
\]

and

\[
\frac{n_{-\infty}}{n_\mathbb{R}} \xrightarrow{p} \frac{p}{1 - p}.
\]

It is a consequence of Theorems 3 and 8 that

\[
n_\mathbb{R} (\hat{d} - d) \xrightarrow{D} Y
\]

where \( Y \) is exponential with mean \( \frac{1 - p}{f_\eta(d)} \). So

\[
n_\mathbb{R} (\hat{d} - d) \frac{f_\hat{\eta}(\hat{d})}{F_\hat{\eta}(d)} \xrightarrow{D} W
\]

where \( W \) is exponential with mean

\[
\left( \frac{f_\eta(d)}{F_\eta(d)} \right) \left( \frac{p}{1 - p} \right) \left( \frac{1 - p}{f_\eta} \right) = 1.
\]

\( \square \)

As in the analysis of the truncated situation, the first order \( \eta \)-terms in a Taylor expansion of (4.24) are zero because \( \hat{\eta} \) is the solution to the likelihood equation. Also, all third order and higher terms along with mixed second order terms are asymptotically zero because their estimators converge at a rate (when you multiply them together) faster than \( O(n^{-1}) \). We have already addressed the first order censoring and second order “regular” terms above.

Procedurally speaking, “testing” for the censored problem is the same as “testing” for the truncated problem. One simply computes the expression in (4.24) and compares the result to a percentile of a \( \chi^2_{k+2} \). The tests are not, however, identical in the sense that taking the “realized” data (the \( x_i > -\infty \)) from a censored distribution and treating
them as if they were a sample of size \( n \) from a truncated distribution (and ignoring the count \( n_{-\infty} \)) is not, generally, the same as taking into account the censored observations by using representation (4.2) in (4.24). This is obvious if you consider that information would be lost by such a procedure. Asymptotically the information from a random sample of size \( n \) distributed according to (4.1) is \( n(\text{Var}(t(X)) - pE \text{Var}[t(X)|X < d]) \).

Treating that sample as if it were from density (3.2) we know, asymptotically, that \( n(1 - p) \) observations will be uncensored and the information contained in these is \( n(1 - p)\text{Var}[t(X)|X \geq d] \). We have shown in (4.12) that the difference of these two quantities is \( np(1 - p)(E[t(X)|X < d] - E[t(X)|X \geq d])^2 \), a positive result. Since the variance of the MLEs is inversely related to the Fisher information, the variance of the MLEs resulting from ignoring the count of censored observations will be larger.

Since \( \hat{d} \) is the sample minimum without regard to conditioning on \( \eta \), it will follow that the profile likelihood ratio test for \( H_0 : \eta = \eta_0 \) is asymptotically \( \chi^2_k \).

### 4.6 Numerics for the Censored Normal Problem

With the theory of likelihood-based inference for censored members of the exponential family in place, we now present an application. Due to its relationship with the motivating problem and its wide variety of applications, we consider the censored normal distribution.

As we discussed in Chapter 2, Cohen (1959) developed a technique for computing MLEs for the usual parameterization of the censored normal distribution with a known censoring point. We may directly adapt this for use when the censoring parameter is unknown by substituting the estimate, \( \hat{d} \), for the censoring point. Using Cohen's technique we need only evaluate a one-dimensional auxiliary function iteratively. We then use fixed transformations to convert the result into separate estimates for the population mean and population standard deviation.
The auxiliary function is the inverse of the left-hand side of expression (2.13) for a fixed value of $h$. It would be ideal to demonstrate that, indeed, an inverse exists. However, the details are so tedious that to do is impractical. Instead, we make a heuristic argument employing graphical techniques.

We need to invert the function

$$
\lambda(\xi|\hat{p}) = \left\{1 - \left(\frac{\hat{p}}{1-\hat{p}}\right) \left(\frac{\phi(-\xi)}{1-\Phi(-\xi)}\right) \left[\left(\frac{\hat{p}}{1-\hat{p}}\right) \left(\frac{\phi(-\xi)}{1-\Phi(-\xi)}\right) - \xi\right]\right\} \left(\frac{\hat{p}}{1-\hat{p}}\right) \left(\frac{\phi(-\xi)}{1-\Phi(-\xi)}\right) - \xi}^2
$$

(4.28)

at the point $g$ (where $g$ is set to $s^2/(\bar{x} - d)^2$) to arrive at a value, $\lambda^{-1}(g|\hat{p}) = \xi_g$. Then we get our adjusting factor, $\theta$, by computing

$$
\theta(\xi_g) = \frac{\hat{p}\phi(\xi_g)}{\hat{p}\phi(\xi_g) - \Phi(\xi_g)\xi_g(1 - \hat{p})}.
$$

(4.29)

With $\theta$ at hand we have the usual estimates of $\mu$ and $\sigma^2$,

$$
\hat{\sigma}^2 = \hat{s}^2 + \theta\{\lambda^{-1}[g(\hat{d}, \bar{x}, \hat{s}^2)]|\hat{p}\}\} (\bar{x} - \hat{d})^2
$$

(4.30)

and

$$
\hat{\mu} = \bar{x} - \theta\{\lambda^{-1}[g(\hat{d}, \bar{x}, \hat{s}^2)]|\hat{p}\}\} (\bar{x} - \hat{d}).
$$

(4.31)

Figure 4.1 shows various values of $\theta$ plotted against $\hat{p}$ and $g$.

As in the truncated case, it is possible to approximate the function $\lambda^{-1}$ using curve-fitting techniques or some iterative routine (incorporating Newton Raphson or a binary search) or a combination of both. A macro program written in SAS® is included in the appendix to assist in these computations.

4.7 Simulation for the Normal Distribution Parameterized in the Usual Manner

Asymptotic properties associated with estimators in the usual parameterization of the normal distribution with an unknown censoring point are similar to those where
there is an unknown truncation parameter. Since we can extend the asymptotic results for the censored case to include smooth reparameterizations as outlined in Chapter 3, the estimators that are solutions to the likelihood equations are asymptotically normal with variances related to those of the natural functions (the $t_j(x)$ in representation (4.1)). Limiting normality of the estimators is the key to ensuring likelihood ratio test statistics are asymptotically chi-square.

The joint density for a sample of size $n$ from the censored normal distribution in the usual parameterization has the form

$$f(x_1, \ldots, x_n | \mu, \sigma) = \frac{(n-\infty + n_{\infty})!}{n_{-\infty}! n_{\infty}!} \left[ \frac{d - \mu}{\sigma} \right]^{-n-\infty} (\sigma \sqrt{2\pi})^{-n_{\infty}} \exp \left[ - \sum_{x_i > -\infty} \frac{(x_i - \mu)^2}{2\sigma^2} \right].$$
We have shown that the maximizing vector of parameter estimates (for the natural parameterization) is asymptotically independent of \( \hat{d} \). Mirroring Chapter 3, we therefore can conclude, since the usual parameters are a smooth transformation of the natural ones, that \((\hat{\mu}, \hat{\sigma})\) is asymptotically bivariate normal while \( \hat{d} \) is asymptotically exponential. Since the marginals of a bivariate normal distribution are univariate normal, we can easily use graphical techniques to qualitatively lend credence to our assertion. Figures 4.2 and 4.3 display histograms and normal plots of \( \hat{\mu} \) and \( \hat{\sigma} \) for 500 simulations of iid censored normal data. (The parameters of the distribution from which the data are sampled are \((\mu, \sigma, d) = (0, 2, -1)\)). Table 4.1 gives statistics corresponding to the plots. Figure 4.4 shows the various exponential plots for the estimator, \( \hat{d} \), of the threshold parameter and for the first order Taylor term \((\hat{d} - d) \sum \frac{\partial}{\partial d} \log f_{\hat{\mu}, \hat{\sigma}, d}(x_i) |_{\hat{d}}\). In general we see that for large \( n \), while the asymptotics prevail, the standard deviations of the estimators are larger than those for the truncated situation. This follows from the fact that observations that are censored contribute less information than those which are not. This is in perfect agreement with Theorem 2.86 Schervish (1995) which states that for a random variable \( X \) and a function \( f \), the Fisher Information of \( f(X) \) is less than or equal to that of \( X \) (in our censored case \( f(X) = X 1_{x \geq d} + (-\infty) 1_{x < d} \)).

The log likelihood, \( L(\mu, \sigma, d) \), is

\[
-n_{-\infty} \log \left[ \Phi \left( \frac{d - \mu}{\sigma} \right) \right] - \frac{n_\sigma}{2} \log(2\pi) - n_\sigma \log(\sigma) - \frac{n_\sigma}{2\sigma^2} (s^2 + (x - \mu)^2). \tag{4.32}
\]

Table 4.1 Statistics for checking normality of maximum likelihood estimators for \( \mu \) and \( \sigma \)

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Sample Mean</th>
<th>Standard Deviation</th>
<th>Shapiro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wilk ( p )-value</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.05855</td>
<td>0.37314</td>
<td>0.97678</td>
</tr>
<tr>
<td>100</td>
<td>0.02295</td>
<td>0.21953</td>
<td>0.98081</td>
</tr>
<tr>
<td>1000</td>
<td>0.00083</td>
<td>0.06562</td>
<td>0.98787</td>
</tr>
<tr>
<td>10000</td>
<td>-0.00049</td>
<td>0.02282</td>
<td>0.98199</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Mean</th>
<th>Sample Standard Deviation</th>
<th>Shapiro</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Wilk ( p )-value</td>
</tr>
<tr>
<td>1.87903</td>
<td>0.31970</td>
<td>0.98174</td>
</tr>
<tr>
<td>1.96595</td>
<td>0.17921</td>
<td>0.98653</td>
</tr>
<tr>
<td>1.99772</td>
<td>0.05572</td>
<td>0.98587</td>
</tr>
<tr>
<td>2.00102</td>
<td>0.01832</td>
<td>0.98809</td>
</tr>
</tbody>
</table>
We may expand $L(\mu, \sigma, d)$ about the point $(\hat{\mu}, \hat{\sigma}, \hat{d})$ using a Taylor series like that given in equation (3.30). Use of such an expansion requires explicit evaluation of the various partial derivatives.

\[
\frac{\partial L}{\partial \mu} = \frac{-n_{-\infty} \phi \left( \frac{d - \mu}{\sigma} \right)}{\sigma \left( \Phi \left( \frac{d - \mu}{\sigma} \right) \right)} + \frac{n_{\infty} (\bar{x} - \mu)}{\sigma^2},
\]

\[
\frac{\partial L}{\partial \sigma} = \frac{n_{-\infty} \phi \left( \frac{d - \mu}{\sigma} \right) (\mu - d)}{\sigma^2 \left( \Phi \left( \frac{d - \mu}{\sigma} \right) \right)} - \frac{n_{\infty}}{\sigma^3} (s^2 + (\bar{x} - \mu)),
\]

\[
\frac{\partial L}{\partial d} = \frac{n_{-\infty} \phi \left( \frac{d - \mu}{\sigma} \right)}{\sigma \left( \Phi \left( \frac{d - \mu}{\sigma} \right) \right)}.
\]

\[
\frac{\partial L^2}{\partial \mu^2} = n_{-\infty} \left\{ \frac{\Phi \left( \frac{d - \mu}{\sigma} \right) \phi \left( \frac{d - \mu}{\sigma} \right) (d - \mu) + \sigma \left[ \phi \left( \frac{d - \mu}{\sigma} \right) \right]^2}{\sigma^3 \left( \Phi \left( \frac{d - \mu}{\sigma} \right) \right)^2} \right\} - \frac{n_{\infty}}{\sigma^2},
\]

\[
\frac{\partial L^2}{\partial \sigma^2} = n_{-\infty} \left\{ \frac{\Phi \left( \frac{d - \mu}{\sigma} \right) \phi \left( \frac{d - \mu}{\sigma} \right) \left[ 2 \sigma^2 (d - \mu) - (\mu - d)^2 \right] - \sigma \left[ \phi \left( \frac{d - \mu}{\sigma} \right) \right]^2 (\mu - d)^2}{\sigma^5 \left( \Phi \left( \frac{d - \mu}{\sigma} \right) \right)^2} \right\}
\]

\[
+ \frac{n_{\infty}}{\sigma^4} (s^2 + (\bar{x} - \mu)),
\]

\[
\frac{\partial L^2}{\partial \sigma \partial \mu} = n_{-\infty} \left\{ \frac{\Phi \left( \frac{d - \mu}{\sigma} \right) \phi \left( \frac{d - \mu}{\sigma} \right) \left[ \sigma^2 (\mu - d)^2 \right] - \sigma \left[ \phi \left( \frac{d - \mu}{\sigma} \right) \right]^2 (\mu - d)}{\sigma^4 \left( \Phi \left( \frac{d - \mu}{\sigma} \right) \right)^2} \right\}
\]

\[
- \frac{2n_{\infty}}{\sigma^3} (\bar{x} - \mu),
\]

\[
\frac{\partial L^2}{\partial d \partial \mu} = \frac{n_{-\infty}}{\sigma} \left\{ \frac{\Phi \left( \frac{d - \mu}{\sigma} \right) \phi \left( \frac{d - \mu}{\sigma} \right) \left[ (\mu - d) \right] - \phi \left( \frac{d - \mu}{\sigma} \right)^2 \left( \frac{d - \mu}{\sigma} \right)}{\left[ \Phi \left( \frac{d - \mu}{\sigma} \right) \right]^2} \right\},
\]

and

\[
\frac{\partial L^2}{\partial d \partial \sigma} = \frac{n_{-\infty}}{\sigma^2} \left\{ \frac{\Phi \left( \frac{d - \mu}{\sigma} \right) \left[ \phi \left( \frac{d - \mu}{\sigma} \right) \left( \frac{\mu - d}{\sigma} \right)^2 - \phi \left( \frac{d - \mu}{\sigma} \right) \right] - \phi \left( \frac{d - \mu}{\sigma} \right)^2 \left( \frac{d - \mu}{\sigma} \right)}{\left[ \Phi \left( \frac{d - \mu}{\sigma} \right) \right]^2} \right\}.
\]

Table 4.2 displays summaries of the simulated values of the various terms of a second order Taylor expansion of the LRT statistic for $H_0: (\mu, \sigma, d) = (0, 1, -1)$. The table
is to be compared to Table 3.2. We again observe that the mean and standard deviation of the sum of the leading terms in the Taylor expansion of the LRT statistic are not much different (as \( n \) gets larger than 100) from the values of 4 and 2.83, the mean and standard deviation of a \( \chi^2_4 \) random variable. The \( d \) cross product terms seem to converge to zero as \( n \) increases. From Figure 4.5 we also see a clear pattern that as \( n \) increases, the sum of first order \( d \) term and second order mean and variance terms is a good approximation to the LRT statistic. Looking at Figure 4.6 we see that a \( \chi^2_4 \) null distribution appears appropriate for the test statistic even for small values of \( n \).

Table 4.2 Summary statistics for the null distribution of terms in the Taylor expansion of the LRT statistic for \( H_0 : (\mu, \sigma, d) = (0, 1, -1) \)

<table>
<thead>
<tr>
<th></th>
<th>linear</th>
<th>quad</th>
<th>quad</th>
<th>cross</th>
<th>cross</th>
<th>cross</th>
<th>2( \times )sum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( d )</td>
<td>( \mu )</td>
<td>( \sigma )</td>
<td>( \mu - \sigma )</td>
<td>( \sigma - d )</td>
<td>( \mu - d )</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>MIN</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>-1.232</td>
<td>-3.537</td>
<td>-1.453</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>6.208</td>
<td>5.021</td>
<td>25.950</td>
<td>2.044</td>
<td>1.236</td>
<td>0.509</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>1.008</td>
<td>0.586</td>
<td>0.922</td>
<td>-0.037</td>
<td>-0.187</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>0.965</td>
<td>0.812</td>
<td>1.911</td>
<td>0.296</td>
<td>0.453</td>
<td>0.197</td>
</tr>
<tr>
<td>100</td>
<td>MIN</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>-1.926</td>
<td>-1.682</td>
<td>-0.734</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>5.310</td>
<td>5.521</td>
<td>11.077</td>
<td>1.531</td>
<td>0.540</td>
<td>0.511</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>0.965</td>
<td>0.579</td>
<td>0.633</td>
<td>-0.034</td>
<td>-0.051</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>0.967</td>
<td>0.837</td>
<td>1.057</td>
<td>0.271</td>
<td>0.191</td>
<td>0.112</td>
</tr>
<tr>
<td>1000</td>
<td>MIN</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.868</td>
<td>-0.231</td>
<td>-0.236</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>7.124</td>
<td>6.241</td>
<td>5.726</td>
<td>0.766</td>
<td>0.236</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>0.981</td>
<td>0.491</td>
<td>0.502</td>
<td>-0.049</td>
<td>-0.003</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>0.963</td>
<td>0.710</td>
<td>0.705</td>
<td>0.212</td>
<td>0.048</td>
<td>0.033</td>
</tr>
<tr>
<td>10000</td>
<td>MIN</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>-1.371</td>
<td>-0.147</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>MAX</td>
<td>8.557</td>
<td>4.841</td>
<td>6.982</td>
<td>0.737</td>
<td>0.131</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>MEAN</td>
<td>1.020</td>
<td>0.595</td>
<td>0.529</td>
<td>-0.059</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>STD DEV</td>
<td>1.068</td>
<td>0.815</td>
<td>0.747</td>
<td>0.264</td>
<td>0.017</td>
<td>0.011</td>
</tr>
</tbody>
</table>
Figure 4.2  The sampling distribution of $\hat{\mu}$ for $n=30, 100, 1000, 10000$
Figure 4.3  The sampling distribution of $\hat{\sigma}$ for $n=30, 100, 1000, 10000$
Figure 4.4 Side-by-side $\chi^2$ plots of $\hat{d}$ and the first order Taylor term for $d$,

$$(\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log f_{\hat{n},\hat{d}}(x_i) \bigg|_{\hat{d}}$$

for $n=30, 100, 1000, 10000$
Figure 4.5 The likelihood ratio test statistic vs. the sum of the asymptotically non-negligible terms of the Taylor series for sample sizes $n=30, 100, 1000,$ and $10000$.
Figure 4.6 $\chi^2$ Plots, each showing the likelihood ratio test statistic from 500 simulations where the data producing the statistic is randomly generated from a censored standard normal distribution with threshold parameter $-1$, for sample sizes $n=30, 100, 1000, \text{and } 10000$
5 INFERENCE IN CENSORED DISTRIBUTIONS WITH
(UNKNOWN CENSORING PARAMETER AND
UNKNOWN) "EXTRA MASS" AT $-\infty$

The theory developed thus far does not directly address the motivating problem introduced in Chapter 1. We have proven results addressing the relationship between the maximum likelihood estimator of a threshold parameter (whether it is an unknown censoring or unknown truncation parameter) and the maximum likelihood estimators of the natural parameters for a continuous exponential family. We have further argued that such results are essentially unchanged when we reparameterize and the new parameters are some smooth transformation of the natural parameters. The mixed point-mass/normal distribution in display (1.1) is not, however, covered by these results. The proofs developed for Theorems 6 and 10 make use of the structure of the exponential family and, to date, we have not found easy extensions outside this family. And so it is unclear how to formally proceed to the "extra mass" problem.

Although many formal technical details are currently missing in the "extra-mass" development, there are obvious conjectures concerning likelihood-based inference in this case. It seems clear that the limiting distribution of the estimator of the threshold is exponential and, further, that the contribution of such an estimator to the null distribution of the likelihood ratio statistic for testing a point null hypothesis for the entire parameter vector will be to add two degrees of freedom to a $\chi^2$ limit. The minimum uncensored observation probably converges to the censoring point at a rate much faster
than the other estimators converge to their parameters. Since this difference in rates was key to showing independence in the other two cases, it seems likely that the minimum uncensored observation and estimators derived as solutions to likelihood equations will be asymptotically independent. If this is the case then we can expect to treat the estimators of the regular parameters as being asymptotically normal with an asymptotic variance related to the Fisher information matrix for fixed $d$.

5.1 Censored Exponential Families with Extra Mass

The general “censored with extra mass distribution” may be defined by the density

$$F_{\theta,p,d}(x) = \{p + (1 - p)F_\theta(d)\} 1_{x = -\infty} + (1 - p)f_\theta(x)1_{x \geq d}$$  \hspace{1cm} (5.1)

with respect the sum of a unit point mass at $-\infty$ and Lebesgue measure on $\mathbb{R}$ where $d$ is the lower censoring point, $p$ is a mixing parameter that allows the additional probability of $x = -\infty$ (in excess of left-tail probability corresponding to the density), and $\theta$ is the parameter vector of the density $f_\theta(x)$. (We avoid the use of $\eta$ here so as to not leave the impression that the density necessarily specifies a regular exponential family.)

5.2 Discussion of the Asymptotics

It is possible to say something definite about the marginal behavior of the first order Taylor term for $\hat{d} = \min(x_i| x_i \geq d)$ when testing $H_0 : (p, \theta, d) = (p_0, \theta_0, d_0)$.

Theorem 12 (Distribution of the linear term for the censoring parameter)

Assume $x_1, \ldots, x_n$ are independent and identically distributed from distribution (5.1). If $(\hat{\theta}, \hat{p})$ is consistent, then

$$(\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log F_{\hat{\theta}, \hat{p}, d}(x_i) \bigg|_{\hat{d}}$$

converges in distribution to an exponential random variable with mean one.
PROOF of Theorem 12

We write

\[
(\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log F_{\hat{\theta}, \hat{\theta}, d}(x_i) \\
\bigg|_{\hat{d}} = (\hat{d} - d) \sum_{i=1}^{n} \frac{\partial}{\partial d} \log \left\{ (\hat{p} + (1 - \hat{p})) F_{\hat{\theta}}(d) 1_{\{x_i < \infty\}} + (1 - \hat{p}) f_{\theta}(x_i) 1_{\{x_i \geq d\}} \right\} \\
= n_{-\infty}(\hat{d} - d) \left[ \frac{(1 - \hat{p}) f_{\theta}(\hat{d})}{\hat{p} + (1 - \hat{p}) F_{\theta}(\hat{d})} \right] |_{\hat{d}}. \tag{5.2}
\]

Now

\[
\frac{(1 - \hat{p}) f_{\theta}(\hat{d})}{\hat{p} + (1 - \hat{p}) F_{\theta}(\hat{d})} \xrightarrow{p} \frac{(1 - p) f_{\theta}(d)}{p + (1 - p) F_{\theta}(d)}
\]

by the right continuity of \( f_{\theta}(x) \) in \( x \), the continuity of \( f_{\theta}(x) \) in \( \theta \), the continuity of \( F_{\theta}(d) \) in \( (\theta, d) \) and the consistency of \( \hat{\theta}, \hat{p} \) and \( \hat{d} \). Further,

\[
n_{-\infty}(\hat{d} - d) = n_{-\infty} n_{\hat{\theta}}(\hat{d} - d)
\]

and

\[
n_{-\infty} \xrightarrow{p} n_{\hat{\theta}} \xrightarrow{p} \frac{p + (1 - p) F_{\theta}(d)}{1 - p + (p - 1) F_{\theta}(d)}.
\]

It is a consequence of Theorems 3 and 8 that

\[
n_{\hat{\theta}}(\hat{d} - d) \xrightarrow{D} Y
\]

where \( Y \) is exponential with mean \( \frac{1 - p + (p - 1) F_{\theta}(d)}{(1 - p) F_{\theta}(d)} \). So

\[
n_{\hat{\theta}}(\hat{d} - d) \xrightarrow{D} \frac{(1 - \hat{p}) f_{\theta}(\hat{d})}{\hat{p} + (1 - \hat{p}) F_{\theta}(\hat{d})} \xrightarrow{D} W
\]

where \( W \) is exponential with mean

\[
\left( \frac{(1 - p) f_{\theta}(d)}{p + (1 - p) F_{\theta}(d)} \right) \left( \frac{p + (1 - p) F_{\theta}(d)}{1 - p + (p - 1) F_{\theta}(d)} \right) \left( \frac{1 - p + (p - 1) F_{\theta}(d)}{(1 - p) F_{\theta}(d)} \right) = 1
\]

\( \square \)
5.3 Numerics for the Normal with Extra Mass Problem

In the introduction we saw that in the usual parameterization the log-likelihood for a normal distribution with an unknown censoring point and extra-mass is given by (1.1). We also noted that maximizing (1.1) is equivalent to maximizing

\[
\log f(x|p, \hat{d}, \mu, \sigma) = n_{-\infty} \log \left[ p + (1 - p)\Phi \left( \frac{\hat{d} - \mu}{\sigma} \right) \right] + n_\infty (\log(1 - p) - \log \sigma)
\]

\[\quad - \frac{n_\infty}{2} \log(2\pi) - \frac{n_\infty}{2\sigma^2} (n_\infty (\bar{x} - \mu)^2 + s^2) \tag{5.3}\]

where \( \hat{d} = \min\{x_i| x_i > d\} \).

The author's experience with the problem has been that (5.3) cannot be maximized with a naive application of the Newton Raphson (or Fisher Scoring) algorithms with any degree of reliability. Two remedies to this situation have been found. First, the problem is readily solvable if one incorporates Newton Raphson into an iterative two-step routine that maximizes first over \((\mu, \sigma)\) holding \(p\) fixed, then solves for \(p\) using the current \((\mu, \sigma)\). This procedure is repeated until the changes in the values of the estimates are sufficiently small. The second approach is to extend Cohen's work and develop a two-step iterative procedure. The step whereby we maximize over \((\mu, \sigma)\) is accomplished through the use of an auxiliary function of one variable.

We demonstrate the latter procedure in what follows. To do so we need to find the first partial derivatives of the log likelihood with respect to \(\mu\) and \(\sigma\).

For \(\mu\)

\[
\frac{\partial L}{\partial \mu} = \frac{-n_{-\infty}(1 - p)\phi \left( \frac{\hat{d} - \mu}{\sigma} \right)}{\sigma [p + (1 - p)\Phi \left( \frac{\hat{d} - \mu}{\sigma} \right)]} + \frac{n_\infty (\bar{x} - \mu)}{\sigma^2}, \tag{5.4}\]

which implies

\[
(\bar{x} - \mu) = \sigma \left( \frac{h}{1 - h} \right) \left( \frac{(1 - p)\phi \left( \frac{\hat{d} - \mu}{\sigma} \right)}{p + (1 - p)\Phi \left( \frac{\hat{d} - \mu}{\sigma} \right)} \right) \tag{5.5}\]

where \(h = n_{-\infty}/n\).
For $\sigma^2$ we first note

$$\frac{\partial}{\partial \sigma^2} \left( \frac{d-\mu}{\sigma} \right) = \frac{\partial}{\partial \sigma^2} \left( \frac{d-\mu}{\sqrt{\sigma^2}} \right) = \left( \frac{-1}{2} \right) \left( \frac{d-\mu}{(\sigma^2)^{-3/2}} \right) = -\left( \frac{d-\mu}{2\sigma^3} \right)$$

and

$$\frac{\partial}{\partial \sigma^2} n^2 \log(\sigma) = \frac{\partial}{\partial \sigma^2} \frac{n^2}{2} \log(\sigma^2) = \frac{n^2}{2\sigma^2}.$$ 

So

$$\frac{\partial L}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2} \left\{ n_{-\infty} \log \left[ \frac{p + (1-p)\Phi \left( \frac{d-\mu}{\sigma} \right)}{\sigma} \right] \right\}$$

$$+ \frac{\partial}{\partial \sigma^2} \left\{ n_{\infty} \left( \log(1-p) - \log(\sigma) - \frac{1}{2} \log(2\pi) \right) \right\}$$

$$- \frac{\partial}{\partial \sigma^2} \left[ \frac{1}{2\sigma^2} (n_{\infty}(-\mu)^2 + n_{\infty} s^2_\sigma) \right]$$

$$= \frac{n_{-\infty}(1-p)\phi \left( \frac{d-\mu}{\sigma} \right)}{\left[ p + (1-p)\Phi \left( \frac{d-\mu}{\sigma} \right) \right]} - \frac{\partial}{\partial \sigma^2} \left[ \log(\sigma^2) \right] - \frac{n_{\infty}}{\sigma^4} (s^2_\sigma + (\bar{x} - \mu)^2)$$

$$= \frac{\left[ -n_{-\infty} (d - \mu) \right]}{2\sigma^3} \frac{(1-p)\phi \left( \frac{d-\mu}{\sigma} \right)}{\left[ p + (1-p)\Phi \left( \frac{d-\mu}{\sigma} \right) \right]} - \frac{n_{\infty}}{2\sigma^2} + \frac{n_{\infty}}{2\sigma^4} (s^2_\sigma + (\bar{x} - \mu)^2). \quad (5.6)$$

Setting (5.6) to zero, multiplying each side by $\sigma^4$ and dividing by $n_{\infty} + n_{-\infty}$ we arrive at

$$h\sigma^2 \left[ \frac{d-\mu}{\sigma} \right] \left[ \frac{(1-p)\phi \left( \frac{d-\mu}{\sigma} \right)}{\left[ p + (1-p)\Phi \left( \frac{d-\mu}{\sigma} \right) \right]} \right] + (1-h)\sigma^2 = (1-h)(s^2_\sigma + (\bar{x} - \mu)^2)$$

and

$$\sigma^2 \left\{ \left( \frac{h}{1-h} \right) \left[ \frac{d-\mu}{\sigma} \right] \left[ \frac{(1-p)\phi \left( \frac{d-\mu}{\sigma} \right)}{\left[ p + (1-p)\Phi \left( \frac{d-\mu}{\sigma} \right) \right]} \right] + 1 \right\} = (s^2_\sigma + (\bar{x} - \mu)^2).$$

Thus, if we let

$$\xi = \frac{d-\mu}{\sigma} \quad \text{and} \quad W(p, h, \xi) = \left( \frac{h}{1-h} \right) \left[ \frac{(1-p)\phi \left( \frac{d-\mu}{\sigma} \right)}{\left[ p + (1-p)\Phi \left( \frac{d-\mu}{\sigma} \right) \right]} \right]$$

we can rewrite the estimating equations as

$$d - \mu = \sigma \xi,$$

$$\bar{x} - \mu = \sigma W(p, h, \xi),$$

and $$s^2_\sigma + (\bar{x} - \mu)^2 = \sigma^2 [1 + \xi W(p, h, \xi)].$$
At this point we appeal to the same type of argument as in Sections 2.1.1 and 2.1.2 to produce the equation

\[ \left[ 1 - \frac{W(p, h, \xi)(W(p, h, \xi) - \xi)}{W(p, h, \xi) - \xi} \right] = \frac{s_\xi^2}{(\bar{x} - d)^2} \]  

(5.7)

with the auxiliary function

\[ \lambda^*(p, h, \xi) = \frac{W(p, h, \xi)}{W(p, h, \xi) - \xi} \]

where \( \xi \) is the value that solves (5.7). We derive the conditional maximum likelihood estimates for \( \mu \) and \( \sigma \) as

\[ \hat{\sigma}^2 = s_\xi^2 + \lambda^*(p, h, \xi)(\bar{x} - d)^2 \]  

(5.8)

and

\[ \hat{\mu} = \bar{x} - \lambda^*(p, h, \xi)(\bar{x} - d). \]  

(5.9)

The complete procedure, then, for finding the maximum likelihood estimators for \((p, \mu, \sigma)\) is as follows. The MLE of \( \hat{d} \) is \( \hat{d} = \min\{x_i|x_i > -\infty\} \). For any pair \((\hat{\mu}_{(i)}, \hat{\sigma}_{(i)})\) find the conditional MLE of \( p \) by solving

\[ \frac{\partial L(p, \hat{d}, \mu, \sigma)}{\partial p} = \frac{1 - \Phi \left( \frac{\hat{d} - \mu}{\sigma} \right)}{p + (1 - p)\Phi \left( \frac{\hat{d} - \mu}{\sigma} \right)} - \frac{n_\xi}{1 - p} = 0. \]  

(5.10)

That is

\[ \hat{p}_{(i+1)} = \max \left\{ 0, 1 - \frac{n_\xi}{n \left[ 1 - \Phi \left( \frac{\hat{d} - \mu_{(i)}}{\sigma_{(i)}} \right) \right]} \right\}. \]

Now compute \( W(p_{(i+1)}, h, \hat{\xi}) \) by first inverting (5.7) to get \( \hat{\xi} \) and recalling that \( h, p_{(i+1)}, \bar{x}, s_\xi^2 \) and \( \hat{d} \) are fixed quantities. From this we get \( \lambda_{(i+1)}^* \) and ultimately, through (5.8) and (5.9), \((\hat{\mu}_{(i+1)}, \hat{\sigma}_{(i+1)})\). This brings us to the beginning of another iteration.

There is still the matter of starting values. By choosing \( \hat{p} \) to be the sample proportion of zero counts minus the expected fraction of zero counts from a normal distribution with
mean \( \bar{x} \) and variance \( s^2 \) and censoring point \( \hat{d} \), we produce a starting value of \( p \) that results in an algorithm that we have empirically found to converge reliably.

If we employ the Newton Raphson approach then for starting values of \((\hat{\mu}, \hat{\sigma})\) given \( \hat{p} \), we may choose the \( \hat{\mu} \) and \( \hat{\sigma} \) from the previous iteration or \( \bar{x} \) and \( s^2 \) for the first iteration.

5.4 Simulation Results for the Normal with Extra Mass

As in the previous two chapters we present a number of plots and tables summarizing simulations using from density (5.3). We used parameters \((\mu, \sigma, p, d) = (0, 1, 0.2, -1)\). Four sample sizes were chosen, and for each, 500 runs were made. The sample sizes were \( n = 30 \), \( n = 100 \), \( n = 1000 \), and \( n = 10000 \).

Through the histograms and normal plots in Figures 5.1, 5.2, and 5.3 and the information given in the table and matrices below, we examine the asymptotic distribution of the “regular” parameter estimators. From the evidence available it seems reasonable that \((\hat{\mu}, \hat{\sigma}, \hat{p})\) are asymptotically normal. However, it seems likely that the asymptotics do not take over at sample sizes below one thousand. Of particular notice are the plots for \( \hat{p} \). From the histograms it is evident that for smaller sample sizes there is a large point mass at the value 0 (41.2% for \( n=30 \) and 11.2% for \( n=100 \)) indicating, at least for \( p \leq 0.2 \), it is often difficult to distinguish the extra-mass model from the ordinary censored model.

Table 5.1 contains statistics corresponding to the plots. Of interest is that at \( n = 10000 \) the Shapiro Wilk statistic for testing if \( \hat{\sigma} \) is normal suggests that normality is not likely while the reverse is true at \( n = 1000 \).

To determine if the negative inverse of the Fisher information matrix is a good approximation for the covariance matrix of the MLEs, we compute the inverse of the (expected) Fisher information at the true parameter values. Using this matrix we can easily derive the vector of approximate standard errors for \((\hat{\mu}, \hat{\sigma}, \hat{p})\) and an approximate
Figure 5.1  The sampling distribution of $\hat{\mu}$ for $n=30, 100, 1000, 10000$

correlation matrix,

\[
\begin{bmatrix}
0.02201 \\
0.01507 \\
0.00969
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
1 & -0.7160 & 0.7788 \\
-0.7160 & 1 & -0.7320 \\
0.7788 & -0.7320 & 1
\end{bmatrix}.
\]

The standard errors can be compared with the values in Table 5.1. The sample correlation matrix computed from the simulation of 500 samples of size $n = 10000$ is

\[
\begin{bmatrix}
1 & -0.661 & 0.743 \\
-0.661 & 1 & -0.686 \\
0.743 & -0.686 & 1
\end{bmatrix}.
\]

As is evident, the two sets of standard errors and correlations differ by no more than about 8% in value. Further, for $n \geq 1000$ these agree with the observed Fisher informa-
Figure 5.2 The sampling distribution of $\hat{\sigma}$ for $n=30, 100, 1000, 10000$

tion. We take this as evidence that the expected Fisher information provides reasonable approximate variances and covariances for large $n$, and that we may approximate the Fisher information using the observed Fisher information (saving the labor of integrating complicated second derivatives). For $n = 100$ or $n = 30$, however, the estimates provided by the observed Fisher information vary by as much as a factor of 2. The top portion of Figure 5.4 gives a 3-dimensional representation of the correlation structure among the "regular" MLEs.

We can use simulation to lend plausibility to speculation that the MLE for $(\mu, \sigma, p)$ is independent of that for $d$. The bottom portion of Figure 5.4 shows a number of scatter plots of $\hat{d}$ plotted against the other estimators. The rows are, from top to bottom, for $n = 30, 100, 1000, \text{and } n = 10000$. The order in each row is $\hat{d}-\hat{\mu}$, $\hat{d}-\hat{\sigma}$, and $\hat{d}-\hat{p}$. The
The sampling distribution of \( \hat{\rho} \) for \( n = 30, 100, 1000, 10000 \)
correlations are given in Table 5.2 with \( p \)-values for testing if the correlations are zero.
The analysis suggests that independence is appropriate at large sample sizes.

Figure 5.6 shows the likelihood ratio test statistic (for testing the point-null hypothesis \( H_0 : (\mu, \sigma, p, d) = (0, 1, 0.2, -1) \)) plotted against the sum of the first order Taylor term for \( d \) and the second order terms for the “regular” parameters. If our speculations regarding independence and rates of convergence are correct, asymptotically the sum should equal the LRT statistic. This seems to be the case. The partial derivatives necessary for the Taylor expansion are as follows.

\[
\frac{\partial L}{\partial \mu} = \frac{n_{-\infty}(1-p)\phi \left( \frac{\hat{d} - \mu}{\sigma} \right) \left( \frac{-1}{\sigma} \right)}{[p + (1-p)\Phi \left( \frac{\hat{d} - \mu}{\sigma} \right)]} + \frac{n_{\sigma}(\hat{d} - \mu)}{\sigma^2},
\]
Table 5.1 Statistics for checking normality of the maximum likelihood estimators for $\mu$, $\sigma$, and $\dot{p}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\dot{\mu}$</th>
<th>$\dot{\sigma}$</th>
<th>$\dot{\dot{p}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>Shap</td>
</tr>
<tr>
<td>30</td>
<td>-0.116</td>
<td>0.297</td>
<td>0.986</td>
</tr>
<tr>
<td>$10^5$</td>
<td>-0.044</td>
<td>0.218</td>
<td>0.970</td>
</tr>
<tr>
<td>$10^6$</td>
<td>-0.006</td>
<td>0.070</td>
<td>0.983</td>
</tr>
<tr>
<td>$10^7$</td>
<td>0.002</td>
<td>0.022</td>
<td>0.983</td>
</tr>
</tbody>
</table>

Table 5.2 Correlations (with $p$-values) between parameter estimators

<table>
<thead>
<tr>
<th>$n$</th>
<th>$d-\dot{\mu}$</th>
<th>$d-\dot{\sigma}$</th>
<th>$d-\dot{\dot{p}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.149129 (0.0008)</td>
<td>-0.063509 (0.1562)</td>
<td>-0.084915 (0.0578)</td>
</tr>
<tr>
<td>100</td>
<td>-0.031659 (0.4800)</td>
<td>-0.001481 (0.9736)</td>
<td>-0.095333 (0.0331)</td>
</tr>
<tr>
<td>1000</td>
<td>-0.038625 (0.3888)</td>
<td>-0.025091 (0.5757)</td>
<td>-0.026596 (0.5530)</td>
</tr>
<tr>
<td>10000</td>
<td>-0.012233 (0.7850)</td>
<td>0.011101 (0.8044)</td>
<td>0.054424 (0.2244)</td>
</tr>
</tbody>
</table>

\[
\frac{\partial L}{\partial \mu} = \frac{n_\infty (1 - p) \phi \left( \frac{\dot{\mu}}{\sigma} \right)}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} - \frac{n_\infty}{\sigma} \left[ p - \frac{1}{2} \frac{(\dot{\mu} - \mu)^2}{\sigma^2} \right],
\]

\[
\frac{\partial L}{\partial \sigma} = \frac{n_\infty (1 - p) \phi \left( \frac{\dot{\mu}}{\sigma} \right)}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} - \frac{n_\infty}{\sigma} \left[ p - \frac{1}{2} \frac{(\dot{\mu} - \mu)^2}{\sigma^2} \right],
\]

\[
\frac{\partial^2 L}{\partial \mu^2} = \frac{n_\infty (1 - p) \left[ \frac{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} \right]^2}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} - \frac{n_\infty}{\sigma^2},
\]

\[
\frac{\partial^2 L}{\partial \sigma^2} = \frac{n_\infty (1 - p) \left[ \frac{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} \right]^2}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} - \frac{n_\infty}{\sigma^2} \frac{3(n_\infty (\dot{\mu} - \mu)^2 + (n_\infty - 1) \sigma^4)}{\sigma^4},
\]

\[
\frac{\partial^2 L}{\partial \dot{\mu} \partial \sigma} = \frac{n_\infty (1 - p) \left[ \frac{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} \right]^2}{p + (1 - p) \Phi \left( \frac{\dot{\mu}}{\sigma} \right)} - \frac{n_\infty}{\sigma^2} \frac{3(n_\infty (\dot{\mu} - \mu)^2 + (n_\infty - 1) \sigma^4)}{\sigma^4}.
\]
\[
\frac{\partial^2 L}{\partial p^2} = \frac{-n_{\infty} \left(1 - \phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right)^2}{p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)} - \frac{n_{\infty}}{(1 - p)^2}
\]

\[
\frac{\partial^2 L}{\partial \mu \partial \sigma} = \frac{n_{\infty} (1 - p)}{\left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]^2} \left[\phi \left(\frac{\bar{d} - \mu}{\sigma}\right) \left(\frac{-(\bar{d} - \mu)^2}{\sigma^4}\right) + \phi \left(\frac{\bar{d} - \mu}{\sigma}\right) \left(\frac{1}{2}\right)\right]
\frac{n_{\infty} (1 - p)}{\left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]^2} - \frac{2n_{\infty}(\bar{x} - \mu)}{\sigma^3},
\]

\[
\frac{\partial^2 L}{\partial p \partial \mu} = \frac{n_{\infty} \left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right] \left[\phi \left(\frac{\bar{d} - \mu}{\sigma}\right) \left(\frac{1}{2}\right)\right]}{\left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]^2}
\frac{n_{\infty} \left[1 - \Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right] \left[(1 - p)\phi \left(\frac{\bar{d} - \mu}{\sigma}\right) \left(\frac{\mu}{\sigma^2}\right)\right]}{\left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]^2},
\]

and

\[
\frac{\partial^2 L}{\partial p \partial \sigma} = \frac{n_{\infty} \left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right] \left[\phi \left(\frac{\bar{d} - \mu}{\sigma}\right) \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]}{\left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]^2}
\frac{n_{\infty} \left[1 - \Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right] \left[(1 - p)\phi \left(\frac{\bar{d} - \mu}{\sigma}\right) \left(\frac{\mu - \bar{d}}{\sigma^2}\right)\right]}{\left[p + (1 - p)\Phi \left(\frac{\bar{d} - \mu}{\sigma}\right)\right]^2}.
\]

As was the case in the other two normal models considered in this thesis, a \(\chi^2\) distribution with two degrees of freedom more than the number of "regular" parameters seems to fit the simulation very well at moderate sample sizes. This is shown by Figure 5.5. In addition, it appears from Figure 5.6 that as \(n\) increases the sum of first order threshold and second order "regular" terms is a good approximation to the LRT statistic.
Figure 5.4 3D plots of $\hat{p}$ versus $(\hat{\mu}, \hat{\sigma})$ and scatter plots of $\hat{d}$ versus $\hat{\mu}, \hat{\sigma}$, and $\hat{p}$ for $n=30,100,1000,10000$
Figure 5.5 \( \chi^2 \) Plots, each showing the likelihood ratio test Statistic from 500 simulations where the data producing the statistic is randomly generated from a censored standard normal with extra mass and with threshold parameter \(-1\) and for sample sizes \(n=30, 100, 1000, \) and 10000.
Figure 5.6  The likelihood ratio test statistic vs. the sum of the asymptotically non-negligible terms of the Taylor series for sample sizes $n=30, 100, 1000, $ and $10000$. 

Sample Size for Simulation 30

Sample Size for Simulation 100

Sample Size for Simulation 1000

Sample Size for Simulation 10000
APPENDIX A SAS® Programs

Presented here are a number of programs needed to implement the various computations displayed throughout the thesis. We employ several conventions in discussing the programs. All nouns with a specific meaning when used in the context of discussing a SAS® program are displayed in the Sans Serif font. We present objects contained within the programs (macros, procedures, data steps, options) in the Courier font; all other text, in the standard \texttt{\LaTeX} Roman font. Wherever possible intuitive variable names are chosen to represent the names for variables appearing in the algorithms (\texttt{xbar, std}) for example. Variable names beginning with an under-score store temporary values that have no meaning outside the program itself.

A.1 Chapter 3 programs

To implement the procedures outlined in Section 3.8, we include a number of programs and macros.

A.1.1 Data Step Functions needed to Implement Cohen’s Algorithm

The first set of SAS® macros are simply functions. These functions must be invoked within the SAS Data Step environment.

\begin{align*}
\%\phi(xi) & \quad \phi(\xi) = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{\xi^2}{2} \right] \\
\%d\phi(xi) & \quad \frac{\partial}{\partial \xi} \phi(\xi) = -\xi \phi(\xi) \\
\%Z(xi) & \quad Z(\xi) \equiv \phi(\xi)/(1 - \Phi(\xi)) \\
\%Zprime(xi) & \quad Z'(\xi) = \frac{\partial Z}{\partial \xi} = \frac{\xi \Phi(\xi) \phi(\xi) - \xi \phi(\xi) + [\phi(\xi)]^2}{[1 - \Phi(\xi)]^2} \\
\%lambda(xi) & \quad \lambda(\xi) = \{1 - Z(\xi) [Z(\xi) - \xi]\} \{Z(\xi) - \xi\}^{-2}
\end{align*}
\[ \frac{\partial \lambda}{\partial \xi} = \frac{(Z - \xi)^2(Z + \xi Z' - 2ZZ') + 2(1 - Z')(Z - \xi)(1 - Z^2 + ZZ')}{(Z - \xi)^4} \]

\[ \theta(\xi) = \frac{Z(\xi)}{Z(\xi) - \xi} \]

\[ \frac{d}{d\xi} \theta(\xi) = \frac{Z(\xi) - \xi Z'(\xi)}{(Z(\xi) - \xi)^2} \]

\begin{verbatim}
%macro phi(xi);
    (1/(sqrt(2)*gamma(.5)))*exp(-(&xi*&xi)/2)
%mend phi;

%macro dphi(xi);
    -&xi*(1/(sqrt(2)*gamma(.5)))*exp(-(&xi*&xi)/2)
%mend dphi;

%macro Z(xi);
    %smallphi(&xi)/(1-probnorm(&xi))
%mend Z;

%macro Zprime(xi);
    ((1-probnorm(&xi))*%smphipr(&xi)+%smallphi(&xi)*%smallphi(&xi))/((1-probnorm(&xi))*(1-probnorm(&xi)))
%mend Zprime;

%macro lambda(xi);
    (1-%Z(&xi)*(%Z(&xi)-&xi))/((%Z(&xi)-&xi)*(%Z(&xi)-&xi))
%mend lambda;

%macro lambdapr(xi);
    (((%Z(&xi)-&xi)*(%Z(&xi)-&xi)*(-2*%Z(&xi)*Zprime(&xi)+%Z(&xi)+&xi*Zprime(&xi))+
      (1-%Z(&xi)*(%Z(&xi)-&xi))*2*%Z(&xi)*(Zprime(&xi)-1))/
     ((%Z(&xi)-&xi)*(%Z(&xi)-&xi)*(%Z(&xi)-&xi)*(%Z(&xi)-&xi))
%mend lambdapr;

%macro theta(xi);
    %Z(&xi)/(%Z(&xi)-&xi)
%mend theta;

%macro thetapr(xi);
    ((%Z(&xi)-&xi*Zprime(&xi))/((%Z(&xi)-&xi)*(%Z(&xi)-&xi))
%mend thetapr;
\end{verbatim}
A.1.2 Deriving the MLEs for a variable contained in a SAS Data Set

The macro \texttt{Jitrmle}, takes the SAS Data Set specified by the argument supplied to the named parameter \texttt{DATA} with the variable specified by the named parameter \texttt{RAW} and computes the MLE of $\mu$ and $\sigma$ for a normal distribution. The algorithm used is the one discussed in Section 8 of Chapter 3. The optional parameter $D$ will be used for the truncation point.

The input you need is as follows:

<table>
<thead>
<tr>
<th>Named Input</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{DATA}</td>
<td>the (2-level) SAS name for the dataset,</td>
</tr>
<tr>
<td>\texttt{RAW}</td>
<td>the variable containing the data,</td>
</tr>
<tr>
<td>\texttt{CONVERGE}</td>
<td>the accuracy of $\lambda^{-1}$ (default is $10^{-6}$),</td>
</tr>
<tr>
<td>\texttt{MAXITER}</td>
<td>the maximum number of iterations (default is 200),</td>
</tr>
<tr>
<td>\texttt{D}</td>
<td>the specified truncation point (default is minimum of \texttt{RAW}),</td>
</tr>
<tr>
<td>\texttt{OUT}</td>
<td>the name of a SAS Data Set to write the results (default is output).</td>
</tr>
</tbody>
</table>

To generate a truncated normal data set (here, named "\texttt{data}" of size $n = 1000$) you may use the following SAS® data step.

```sas
data data(keep=raw);
  do i=1 to 1000;
    do until(b>-1);
      b=rannor(0);
    end;
    output raw;
  end;
run;
```

The program produces six output macro variables and writes their values to the SAS log.

<table>
<thead>
<tr>
<th>Output Macro Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{muhat}</td>
<td>$\hat{\mu} = \bar{x} - \theta{\lambda^{-1}[g(d, \bar{x}, s^2)]}(\bar{x} - \hat{d})$.</td>
</tr>
<tr>
<td>\texttt{sigmahat}</td>
<td>$\hat{\sigma} = \sqrt{s^2 + \theta{\lambda^{-1}[g(d, \bar{x}, s^2)]}(\bar{x} - \hat{d})^2}$.</td>
</tr>
<tr>
<td>\texttt{dhat}</td>
<td>$\hat{d} = \min{x_i; i \in 1, 2, \ldots, n}$ or $D$ if specified</td>
</tr>
<tr>
<td>\texttt{s_sigma}</td>
<td>$\sqrt{-\left(\frac{\partial^2 L}{\partial \mu^2} \left( \frac{\partial^2 L}{\partial \mu^2} \frac{\partial^2 L}{\partial \sigma^2} - \frac{\partial^2 L}{\partial \mu \partial \sigma} \right)^{-1} \right)}$</td>
</tr>
<tr>
<td>\texttt{s_mu}</td>
<td>$\sqrt{-\left(\frac{\partial^2 L}{\partial \sigma^2} \left( \frac{\partial^2 L}{\partial \mu^2} \frac{\partial^2 L}{\partial \sigma^2} - \frac{\partial^2 L}{\partial \mu \partial \sigma} \right)^{-1} \right)}$</td>
</tr>
<tr>
<td>\texttt{corr}</td>
<td>$\sqrt{\frac{\partial^2 L}{\partial \mu \partial \sigma} \frac{\partial^2 L}{\partial \mu \partial \sigma} - \frac{\partial^2 L}{\partial \mu \partial \sigma} \left( \frac{\partial^2 L}{\partial \mu^2} \frac{\partial^2 L}{\partial \sigma^2} \right)}$</td>
</tr>
</tbody>
</table>
%macro trmle(data=,raw=,converge=, maxiter=, d=. , out=);
  %global muhat sigmahat;
  %let _a=%index(&data,);
  %if _a=0 %then %do;
    %let libname=work;
    %let memname=&data;
  %end;
  %else %do;
    %let libname=%trim(%left(%substr(&data,1,%eval(&_a-1))));
    %let memname=%trim(%left(%substr(&data,
      %eval(&_a+1),%eval(%length(&data)-&_a))));
  %end;
proc sql;
  select nobs into :n from sashelp.vtable
  where libname=upcase("libname")
  and memname=upcase("memname");
  select min(&raw) into y from libname..memname;
quit;
%let n=%left(%trim(&n));
%put n=&n;
%if &converge= %then %let converge=0.000001;
%if &maxiter= %then %let maxiter=200;
%if %sysevalf(4d,boolean)=0 %then %do;
  %if &d ne 0 %then %let d=minimum; %put missing d, d=minimum=&d;
  %end;
%let min=%sysfunc(min(&d,&minimum));
data _null_;
  array xx(4n);
  do _i_= 1 to 4n;
    set libname..memname;
    xx(_i_)=raw;
  end;
  n=4n;
  xbar=mean(of xx1-xx4n);
  std= sqrt( ((4n-l)/4n)*var(of xx1-xx4n) );
  min=min(min(of xx1-xx4n),&d);
  g=std**2/((xbar-min)**(xbar-min));
  start=-6.474948 + (49.875654 )*g - (215.036450 )*g**2 
    + (484.408232 )*g**3 -(515.706547 )*g**4 +(210.885646 )*g**5;
  prstart=-100;
  do while(abs(start-prstart)>&converge);
    iter+1;
    prstart=start;
    lambda=%lambda(start);
    if lambda> g then start=start-(.5**iter)*(.2);
    else start=start+(.5**iter)*(.2);
    diff=lambda-g;
    theta=%theta(start);
    if iter>&maxiter then stop;
  end;
  muhat=xbar-theta*(xbar-min);
  varhat=(std**2)+theta*(xbar-min)*(xbar-min);
\[ \text{sigmahat} = \sqrt{\text{varhat}}; \]

\[ \text{put / muhat= / varhat= / sigmahat= / min=;} \]

\[ \text{call symput('muhat',muhat); call symput('sigmahat',sigmahat);} \]

\[ / \text{*** Comput the Estimated Variance of the MLEs ***} / \]

\[ \text{ds2=}%d1ls2\text{(min,muhat,sigmat)}\; \text{put ds2=} ; \]

\[ \text{dm2=}%d1lmu2\text{(min,muhat,sigmat)}; \text{put dm2=} ; \]

\[ \text{dms=}%d1lms\text{(min,muhat,sigmat)}; \text{put dms=} ; \]

\[ \text{varmu=}-(\text{ds2}/(\text{dm2*ds2-dms*dms})); \text{put s_musqrt(varmu);} \]

\[ \text{vars=}-(\text{dm2}/(\text{dm2*ds2-dms*dms})); \text{put s_sigmarsqrt(vars);} \]

\[ \text{cov=}-(\text{dms}/(\text{dm2*ds2-dms*dms})); \text{put s_musqrt(cov)}; \]

\[ \text{put s_sigma=} / \text{ s_musqrt=} / \text{ corr=} ; \]

\[ \text{call symput('muhat',muhat);} \]

\[ \text{call symput('sigmahat',sigmahat);} \]

\[ \text{call symput('dhat',min);} \]

\[ \text{call symput('s_mu',s_mu);} \]

\[ \text{call symput('s_sigma',s_sigma);} \]

\[ \text{call symput('corr',corr);} \]

\[ \text{run;} \]

\%if &out= \%then \%put NO DATA SET CREATED;

\%else \%do;

\%data &out; muhat=&muhat; dhat=&dhat; sigmahat=&sigmahat; run;

\%end;

\%mend trmle;

Here is some output generated by \%trmle using 2 samples of 1000 observations each taken from standard normals with truncation points of -1 and 3 respectively.

\%trmle(data=raw, raw=b, converge=0.000000001, \%out=result);

missing d, d=minimum=-0.9985

\begin{verbatim}
MUHAT=0.069350517
VARHAT=0.9166547215
SIGMAHAT=0.9575305329
MIN=-0.9985
DS2=4424.858995
DM2=-515.0877596
DMS=678.3923545
S_MU=0.0499410938
S_SIGMA=0.0355963626
CORR=-0.667920905
\end{verbatim}

missing d, d=minimum=3.000053

\begin{verbatim}
MUHAT=-0.630070972
VARHAT=1.18288698537
SIGMAHAT=1.0875982042
MIN=3 000053
DS2=3000 965334
DM2=-51.50877596
DMS=392.5923545
S_MU=2.4573574371
S_SIGMA=0.3219426606
CORR=-0.998391202
\end{verbatim}
A.1.3 Likelihood Ratio Test functions

These macros allow one to calculate partial derivatives of the log-likelihood of a left-truncated normal distribution. They can be used for doing Newton Raphson or for computing the Fisher information.

These macros must be called within a SAS data step. Each macro has three positional parameters (dhat, muhat, and sigmahat). These must be supplied. In addition, the statistics n, xbar and std are needed to be defined in the data step calling the macro. The following example shows 3 of the macros being called to create the SAS® data set "Graph".

```
data graph;
  retain xbar 0.3278736565 std 0.770257147 n 100
  min -0.983346899 muhat 0.1181753755;
  do sigmahat=0.5 to 2.5 by .1;
    loglike=%loglike(niin,muhat,sigmahat);
    dlls =%dlls(min,muhat,sigmahat);
    diff=(dlls-lag1(dlls))*10;
    dlls2 =%dlls2(min,muhat,sigmahat);
  end;
run;
```

Macros with Definitions

\[
\frac{d}{\sigma} \log \left[ 1 - \Phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right) \right] - n \log(\sigma) - \frac{n}{2\sigma^2} (s^2 + (\bar{x} - \hat{\mu})^2)
\]

\[
\frac{n\phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right)}{\sigma \left( 1 - \Phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right) \right)}
\]

\[
\frac{-n\phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right) \left( \frac{1}{\sigma^2} \right) + \frac{n(\bar{x} - \hat{\mu})}{\sigma^2}}{1 - \Phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right)}
\]

\[
\frac{n\phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right) \left( \frac{1}{\sigma^2} \right) + \frac{n(\bar{x} - \hat{\mu})}{\sigma^2}}{1 - \Phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right)}
\]

\[
\frac{n\phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right) \left( \frac{1}{\sigma^2} \right) + \frac{n(\bar{x} - \hat{\mu})}{\sigma^2}}{1 - \Phi \left( \frac{\mu - \hat{\mu}}{\sigma} \right)}
\]

\[
v = \frac{n(\bar{x} - \hat{\mu})}{\sigma^2} + \frac{n}{\sigma^2} \left( s^2 + (\bar{x} - \hat{\mu})^2 \right)
\]

\[
v = \frac{n(\bar{x} - \hat{\mu})}{\sigma^2} + \frac{n}{\sigma^2} \left( s^2 + (\bar{x} - \hat{\mu})^2 \right)
\]
\[
%\text{dllms} \quad n \left\{ \frac{1 - \Phi \left( \frac{\bar{X} - \mu}{\sigma} \right)}{\left( \frac{\sigma}{\sqrt{n}} \right)^2} \phi \left( \frac{\bar{X} - \mu}{\sigma} \right) \left( \frac{\sigma}{\sqrt{n}} \right)^2 \right\} - \frac{2n}{\sigma^2} (\bar{x} - \mu)
\]

\[
%\text{dllmd} \quad \frac{n}{\sigma} \left\{ \frac{1 - \Phi \left( \frac{\bar{X} - \mu}{\sigma} \right)}{\left( \frac{\sigma}{\sqrt{n}} \right)^2} \phi \left( \frac{\bar{X} - \mu}{\sigma} \right) \left( \frac{\sigma}{\sqrt{n}} \right)^2 \right\}
\]

\[
%\text{dllsd} \quad \frac{n}{\sigma^2} \left\{ \frac{1 - \Phi \left( \frac{\bar{X} - \mu}{\sigma} \right)}{\left( \frac{\sigma}{\sqrt{n}} \right)^2} \phi \left( \frac{\bar{X} - \mu}{\sigma} \right) \left( \frac{\sigma}{\sqrt{n}} \right)^2 \right\}
\]

%macro arg(dhat,muhat,sigmahat);
\end{verbatim}

*(&dhat-&muhat)/&sigmahat
%mend arg;

%macro dargs(dhat,muhat,sigmahat);
\end{verbatim}

(&muhat-&dhat)/&sigmahat**2
%mend dargs;

%macro loglike(dhat,muhat,sigmahat);
\end{verbatim}

-(&n*log(1-probnorm( %arg(&dhat,&muhat,&sigmahat)) )
-(&n*log(&sigmahat)-(5/2)*(&sigmahat**2)
*(&n*std*std+&n*(xbar-&muhat)*(xbar-&muhat))
%mend loglike;

%macro dllld(dhat,muhat,sigmahat);
\end{verbatim}

(&n*/phi( %arg(&dhat,&muhat,&sigmahat))/
\end{verbatim}

(&sigmahat*(1-probnorm( %arg(&dhat,&muhat,&sigmahat))))
%mend dllld;

%macro dllmu(dhat,muhat,sigmahat);
\end{verbatim}

(-&n*/%phi( %arg(&dhat,&muhat,&sigmahat)))/
\end{verbatim}

(&sigmahat*(1-probnorm( %arg(&dhat,&muhat,&sigmahat))))
+%n*(xbar-&muhat)/&sigmahat**2
%mend dllmu;

%macro dllmu2(dhat,muhat,sigmahat);
\end{verbatim}

(-&n/*%phi( %arg(&dhat,&muhat,&sigmahat)))*(-1/2)*%phi( %arg(&dhat,&muhat,&sigmahat))+(1/2)*%phi( %arg(&dhat,&muhat,&sigmahat))/
\end{verbatim}

(1-probnorm( %arg(&dhat,&muhat,&sigmahat)))**2-%n/&sigmahat**2
%mend;

%macro dlls(dhat,muhat,sigmahat);
\end{verbatim}

-%n*(-%phi( %arg(&dhat,&muhat,&sigmahat))*
\end{verbatim}

%dargs(&dhat,&muhat,&sigmahat))/
\end{verbatim}

(1-probnorm( %arg(&dhat,&muhat,&sigmahat)))-%n/&sigmahat +
\end{verbatim}

(1/2)*%phi( %arg(&dhat,&muhat,&sigmahat))*(%n*std**2+%n*(xbar-&muhat)**2)
%mend dlls;

%macro dlls2(dhat,muhat,sigmahat);
\end{verbatim}

-%n*(1-probnorm( %arg(&dhat,&muhat,&sigmahat)))*
((%phi(%arg(&dhat,&muhat,&sigmahat))))*
((-2*(&muhat-&dhat))/&sigmahat**3)*
(%dargs(&dhat,&muhat,&sigmahat))*
(%phi(%arg(&dhat,&muhat,&sigmahat)))*
(%dargs(&dhat,&muhat,&sigmahat))*
(-%phi(%arg(&dhat,&muhat,&sigmahat)))*
(%dargs(&dhat,&muhat,&sigmahat))*
((1-probnorm(%arg(&dhat,&muhat,&sigmahat))))**2+&n/&sigmahat**2
-(3/&sigmahat**4)*(&n*std**2+&n*(xbar-&muhat)**2)
%end dlls2;
%
%macro dllms(dhat,muhat,sigmahat);
&n*( (1-probnorm(%arg(&dhat,&muhat,&sigmahat))))*
((%phi(%arg(&dhat,&muhat,&sigmahat)))*(1/&sigmahat**2)
+(%dargs(&dhat,&muhat,&sigmahat))*
(%phi(%arg(&dhat,&muhat,&sigmahat)))*
(-1/&sigmahat))-%phi( %arg(&dhat,&muhat,&sigmahat )))*
(%dargs(&dhat,&muhat,&sigmahat))*
(-%phi(%arg(&dhat,&muhat,&sigmahat)))*(-1/&sigmahat)/
(1-probnorm(%arg(&dhat,&muhat,&sigmahat)))**2
+(1/&sigmahat**3)*(-2*%n*(xbar-&muhat))
%end dllms;
%
%macro dllmd(dhat,muhat,sigmahat);
( (1-probnorm(%arg(&dhat,&muhat,&sigmahat))))*
%phi(%arg(&dhat,&muhat,&sigmahat))*
(%phi(%arg(&dhat,&muhat,&sigmahat)))*
-%arg(&dhat,&muhat,&sigmahat)*(-&n/&sigmahat**2) -
(&n/&sigmahat**2)*(%phi(%arg(&dhat,&muhat,&sigmahat)))*
(%phi(%arg(&dhat,&muhat,&sigmahat)))/
(1-probnorm(%arg(&dhat,&muhat,&sigmahat)))**2
%end dllmd;
%
%macro dllsd(dhat,muhat,sigmahat);
&/(/&sigmahat*%sigmahat))*
((1-probnorm(%arg(&dhat,&muhat,&sigmahat))))*
((&muhat-&dhat)*%phi(%arg(&dhat,&muhat,&sigmahat)))*
((&muhat-&dhat)/(&sigmahat*%sigmahat))-
(%phi(%arg(&dhat,&muhat,&sigmahat)))/
-(%phi(%arg(&dhat,&muhat,&sigmahat)))*
(%phi(%arg(&dhat,&muhat,&sigmahat)))*
(&dhat-&muhat)/&sigmahat)))/
(1-probnorm(%arg(&dhat,&muhat,&sigmahat)))**2
%end dllsd;
A.1.4 Sample Code for the Plots

All plots in the dissertation were created using SAS® Graph software. The code is often tedious. Here is the program for generating Figure 3.3. Setting options for the various portions of the graph is the bulk of the work once the macro functions are in place.

```sas
%include '<macro storage locations>';
data plots;
  retain step .1;
  do xi=-5 to 5 by step;
    lambda=%lambda(xi);
    lambdapr=%lambdapr(xi);
    label lamda='Function' lambdapr='Derivative';
    output;
  end;
run;
filename graph1 '<storage location> ';
goptions t2a'get=pscolor device=psrectll hsize=0 vsize=0
  horigin=0 vorigin=0 border
gsfnaine=graph1 gsfmode=replace;
title f=swiss 'Plot of ' f=greek '1(c) ' f=swiss 'and ' f=math 'k'
f=greek '1(c)/' f=math 'k' f=greek 'c';
axis1 order=(-5 to 5 by 2) value=(height=.85 tick=4)
  label=(f=greek 'c' j=c);
axis2 label=(a=90 f=greek '1(c)' f=swiss 'and '
  f=math 'k' f=greek '1(c)/'
  f=math 'k' f=greek 'c' j=center);
legend1 offset=(5cm,7cm) shape=line(1cm) across=1
  label=(position=top justify=center 'Key') frame mode=share;
symbol1 interpol=join value=point line=1;
symbol2 interpol=join value=point line=2;
proc gplot data=plots;
  plot (F Fprime)*xi / overlay
    haxis=axis1 vaxis=axis2 legend=legend1;
run;
```
A.2 Chapter 4 Programs

A.2.1 Functions needed to Implement Cohen's Algorithm

Here are a number of SAS® macros needed for computing MLEs by implementing the procedure in Section 6 of Chapter 4. In addition, many of the same macros that were needed for computing MLEs (for the truncated normal likelihood) are also needed here.

The macros that begin dll- are the partial derivatives of the log-likelihood with respect to the various parameters. Their forms are the same as given in Section 4.7 (except that the following are evaluated at the MLE).

Macros with Definitions

The first 5 functions require the positional parameters \( h \) and \( \xi \) (in that order)

\[
\%Y \quad \left( \frac{h}{1-h} \right) Z(-\xi) = \left( \frac{h \phi(\xi)}{(1-h)\Phi(\xi)} \right)
\]

\[
\%Yprime \quad -\left( \frac{h}{1-h} \right) Z'(-\xi) = \frac{h \xi [1 - \Phi(\xi)] \phi(\xi) - h \xi \phi(\xi) + h \phi(\xi)^2}{(1-h)\Phi(\xi)^2}
\]

\[
\%cohen18 \quad \frac{1 - (\frac{h}{1-h}) Z(-\xi) \left[ \left( \frac{h}{1-h} \right) Z(-\xi) - \xi \right]}{\left[ \left( \frac{h}{1-h} \right) Z(-\xi) - \xi \right]^2}
\]

\[
\%lambda \quad \frac{Y(h, \xi)}{1 - Y(h, \xi)} = \frac{\left( \frac{h}{1-h} \right) \frac{\phi(\xi)}{(1-\Phi(\xi))}}{1 - \left[ \left( \frac{h}{1-h} \right) \frac{\phi(\xi)}{(1-\Phi(\xi))} \right]} = \frac{1}{\frac{(1-h)\Phi(\xi)}{h \phi(x)} - 1}
\]

\[
\%censd \quad \frac{\partial}{\partial \xi} \left( \frac{1 - (\frac{h}{1-h}) Z(-\xi) \left[ \left( \frac{h}{1-h} \right) Z(-\xi) - \xi \right]}{\left[ \left( \frac{h}{1-h} \right) Z(-\xi) - \xi \right]^2} \right)
\]

These functions require the positional parameters nzero, nplus, dhat, mhat and sigmahat

\[
\%loglike \quad -n_{-\infty} \log \left[ \frac{\phi \left( \frac{d - \hat{\mu}}{\hat{\sigma}} \right)}{\Phi \left( \frac{d - \hat{\mu}}{\hat{\sigma}} \right)} \right] - \frac{n_{\mu}}{2} \log(2\pi) - n_{\mu} \log(\hat{\sigma}) - \frac{n_{\mu}}{2\hat{\sigma}^2} (\hat{x} - \hat{\mu})^2
\]

\[
\%dlld \quad \frac{n_{-\infty} \phi \left( \frac{d - \hat{\mu}}{\hat{\sigma}} \right)}{\hat{\sigma} \Phi \left( \frac{d - \hat{\mu}}{\hat{\sigma}} \right)}
\]

\[
\%dllmu \quad \frac{-n_{-\infty} \phi \left( \frac{d - \hat{\mu}}{\hat{\sigma}} \right)}{\hat{\sigma} \Phi \left( \frac{d - \hat{\mu}}{\hat{\sigma}} \right)} + \frac{n_{\mu} (\hat{x} - \hat{\mu})}{\hat{\sigma}^2}
\]
\[ \frac{n_{-\infty} \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) (\hat{d} - \hat{\mu})}{\hat{\sigma}^2 \left[ \Phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \right]} - \frac{n_\alpha}{\hat{\sigma}^2} + \frac{n_\alpha}{\hat{\sigma}^3} (\hat{d}^2 + (\bar{x} - \hat{\mu})) \]

\[ n_{-\infty} \left\{ \frac{\phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) (\hat{d} - \hat{\mu}) + \hat{\sigma} \left[ \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \right]^2}{\hat{\sigma}^3 \left[ \Phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \right]^2} \right\} - \frac{n_\alpha}{\hat{\sigma}^2} \]

\[ n_{-\infty} \left\{ \frac{\phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) [2 \hat{\sigma}^2 (\hat{d} - \hat{\mu}) - (\hat{d} - \hat{\mu})^2]}{\hat{\sigma}^3 \left[ \Phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \right]^2} \right\} + \frac{n_\alpha}{\hat{\sigma}^2} \left( \hat{d}^2 + (\bar{x} - \hat{\mu}) \right) \]

\[ n_{-\infty} \left\{ \frac{\phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) [\hat{d}^2 - (\hat{\mu} - \hat{d})^2] - \hat{\sigma} \left[ \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \right]^2 (\hat{d} - \hat{\mu})}{\hat{\sigma}^4 \left[ \Phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \right]^2} \right\} - \frac{2n_\alpha}{\hat{\sigma}^2} (\bar{x} - \hat{\mu}) \]

\[ \frac{n_{-\infty}}{\hat{\sigma}} \left\{ \frac{\phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) (\hat{d} - \hat{\mu}) - \phi \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \left( \frac{1}{\hat{\sigma}} \right)}{\Phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right)} \right\} \]

\[ \frac{n_{-\infty}}{\hat{\sigma}^2} \left\{ \frac{\phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) (\hat{d} - \hat{\mu}) - \phi \longleftarrow \phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right) \left( \frac{1}{\hat{\sigma}} \right)}{\Phi \left( \frac{\hat{d} - \hat{\mu}}{\hat{\sigma}} \right)} \right\} \]

%macro Y(h, xi);
(\&h/(i-\&h))%Z(-&xi)
%mend;
%macro Yprime(h, xi);
(\&h/(\&h-i))%Zprime(-&xi)
%mend;
%macro cohen18(h, xi);
(i-%Y(&h, &xi)*)((%Y(&h, &xi)-&xi))/((%Y(&h, &xi)-&xi)*(%Y(&h, &xi)-&xi))
%'%mend;
%'%macro censd(h, xi);
(\(\frac{(C_{XY}(fth,4xi) - ftxi)}{Yprime(h, &xi)}\) \* \(\frac{(C_{XY}(h, ftxi) - ftxi)}{Yprime(h, &xi)}\) * (-2 * \(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\)) - (1 - \(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\)) * 2 * \(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\) * (-2 * \(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\)) - (1 - \(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\)) * \(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\))
%'%mend;
%'%macro lambda(h, xi);
\(\frac{1}{Yprime(h, 4xi)}\) * \(\frac{1}{Yprime(h, ftxi)}\)
%'%mend;
%'%macro loglike(nzero,nplus,dhat,muhat,sigmahat);
&nzero*log(probnorm( \(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\)) ) -&nplus*log(&sigmahat) -(.5)*(1/(&sigmahat**2))
(&nplus*varmle+nplus*(xbar-&muhat))*(xbar-&muhat)
%'%mend loglike;
%'%macro dlld(nzero,nplus,dhat,muhat,sigmahat);
(\(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\))
%'%mend dlld;
%'%macro dllmu(nzero,nplus,dhat,muhat,sigmahat);
(-\(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\))
%'%mend dllmu;
%'%macro dlls(nzero,nplus,dhat,muhat,sigmahat);
-%nzero*(-\(\frac{phi( \(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\))}{\(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\)}\)) / (probnorm( \(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\)) - &nplus/&sigmahat + (1/&sigmahat**3)(*(&nplus*varmle+nplus*(xbar-&muhat))*(xbar-&muhat))
%'%mend dlls;
%'%macro dllmu2(nzero,nplus,dhat,muhat,sigmahat);
(-&nzero/&sigmahat)*
( \(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\))
%'%mend dllmu2;
%'%macro dlls2(nzero,nplus,dhat,muhat,sigmahat);
&nzero*((probnorm( \(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\)) * \(\frac{phi( \(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\))}{\(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\)}\)) * (-2 * &muhat-\(\frac{arg(dhat, &muhat, &sigmahat)}{\frac{arg(dhat, &muhat, &sigmahat)}}\)) / &sigmahat**3) + (%dargs&dhat&muhat&sigmahat)*
%'%mend dlls2;
\[
\begin{align*}
(\%dargs(\&dhat, \&muhat, \&sigmahat)) * \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))* \\
(\%dargs(\&dhat, \&muhat, \&sigmahat)) / \\
(\text{probnorm}(\%arg(\&dhat, \&muhat, \&sigmahat)))**2 \\
+ \&nplus/\&sigmahat**2 \\
- (3/\&sigmahat**4)*(\&nplus*\text{varml}e+\&nplus*(xbar-\&muhat)*(xbar-\&muhat))
\end{align*}
\]

\%mend dlls2;

\%macro dllms(nzero,nplus,dhat,muhat,sigmahat);
\&nzero*((\text{probnorm}(\%arg(\&dhat, \&muhat, \&sigmahat)))* \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))*(1/\&sigmahat**2) \\
+((\%dargs(\&dhat, \&muhat, \&sigmahat))*) \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))*(-1/\&sigmahat)) \\
-(\%phi(\%arg(\&dhat, \&muhat, &sigmahat)))* \\
(\%dargs(\&dhat, \&muhat, \&sigmahat))**2 \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))*(1/\&sigmahat))/ \\
(\text{probnorm}(\%arg(\&dhat, \&muhat, \&sigmahat)))**2 \\
+ (1/\&sigmahat**3)*(-2*\&nplus*(xbar-\&muhat))
\%mend dllms;

\%macro dllsd(nzero,nplus,dhat,muhat,sigmahat);
\&nzero*((\text{probnorm}(\%arg(\&dhat, \&muhat, \&sigmahat)))* \\
(\%dargs(\&dhat, \&muhat, \&sigmahat)))* \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))*(1/\&sigmahat)) \\
+(\%dargs(\&dhat, \&muhat, \&sigmahat))**2 \\
-((\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))* \\
(\%dargs(\&dhat, \&muhat, \&sigmahat)))* \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))).
\%mend dllsd;

\%macro dllmd(nzero,nplus,dhat,muhat,sigmahat);
\&nzero/\&sigmahat)*((\text{probnorm}(\%arg(\&dhat, \&muhat, \&sigmahat)))* \\
\%phi(\%arg(\&dhat, \&muhat, &sigmahat))**2 \\
(1/\&sigmahat)) + (1/\&sigmahat) \\
(\%phi(\%arg(\&dhat, \&muhat, \&sigmahat)))* \\
(\%dargs(\&dhat, \&muhat, \&sigmahat)))/ \\
(\text{probnorm}(\%arg(\&dhat, \&muhat, \&sigmahat)))**2 \\
\%mend dllmd;
A.2.2 Computing MLEs for the censored normal distribution

The macro %cnmle, takes the SAS Data Set specified by the argument supplied to the named parameter DATA with the variable specified by the named parameter RAW and computes the MLEs of μ and σ for a censored normal distribution. The program uses the algorithm discussed in Section 6 of Chapter 4. This algorithm converges reliably at well over 99% censoring.

There are a number of optional parameters that may be specified. Optional parameters are those with “defaults” listed after them. The input you need is as follows:

<table>
<thead>
<tr>
<th>Named Input</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA</td>
<td>the (2-level) SAS name for the dataset,</td>
</tr>
<tr>
<td>RAW</td>
<td>the variable containing the data,</td>
</tr>
<tr>
<td>CONVERGE</td>
<td>the accuracy of $\lambda^{-1}$ (default is $10^{-6}$),</td>
</tr>
<tr>
<td>MAXITER</td>
<td>the maximum number of iterations (default is 200),</td>
</tr>
<tr>
<td>D</td>
<td>the specified truncation point (default is minimum of RAW),</td>
</tr>
<tr>
<td>OUT</td>
<td>the name of a SAS Data Set to write the results (default is output).</td>
</tr>
</tbody>
</table>

The program produces six output macro variables and writes their values to the SAS log. These output values correspond to the parameter estimates and their standard errors along with the correlation between the estimates (as computed from the observed Fisher information).

<table>
<thead>
<tr>
<th>Output Macro Variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>muhat</td>
<td>$\hat{\mu} = \bar{x} - \theta{\lambda^{-1}[g(\hat{d}, \bar{x}, s^2)]}(\bar{x} - \hat{d})$.</td>
</tr>
<tr>
<td>sigmahat</td>
<td>$\hat{\sigma} = \sqrt{s^2 + \theta{\lambda^{-1}[g(\hat{d}, \bar{x}, s^2)]}(\bar{x} - \hat{d})^2}$.</td>
</tr>
<tr>
<td>dhat</td>
<td>$\hat{d} = \min{x_i</td>
</tr>
<tr>
<td>s_sigma</td>
<td>$\sqrt{-\left(\frac{\partial^2 L}{\partial \mu^2} \left[ \frac{\partial^2 L}{\partial \mu^2} \frac{\partial^2 L}{\partial \sigma^2} - \left[ \frac{\partial^2 L}{\partial \mu \sigma} \right]^{-1} \right] \right)}$</td>
</tr>
<tr>
<td>s_mu</td>
<td>$\sqrt{-\left(\frac{\partial^2 L}{\partial \sigma^2} \left[ \frac{\partial^2 L}{\partial \mu^2} \frac{\partial^2 L}{\partial \sigma^2} - \left[ \frac{\partial^2 L}{\partial \mu \sigma} \right]^{-1} \right] \right)}$</td>
</tr>
<tr>
<td>corr</td>
<td>$\sqrt{\frac{\partial^2 L}{\partial \mu \partial \sigma} \left[ \frac{\partial^2 L}{\partial \mu^2} \frac{\partial^2 L}{\partial \sigma^2} - \left[ \frac{\partial^2 L}{\partial \mu \sigma} \right]^{-1} \right]}$</td>
</tr>
</tbody>
</table>

Here is the sample output for submitting

```
%cnmle(data=temp, raw=raw, converge=.000000001, out=result);
```

where temp is a data set with $(\mu, \sigma, d) = (0, 2, 6)$ and $n = 10000$. As you can see only 14 data points were available, yet the program does a reasonable job at generating parameter estimates.

<table>
<thead>
<tr>
<th>Output Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUHAT</td>
<td>0.0823620288</td>
</tr>
<tr>
<td>VARHAT</td>
<td>3.9983563607</td>
</tr>
<tr>
<td>SIGMAHAT</td>
<td>1.999589048</td>
</tr>
<tr>
<td>D</td>
<td>6.059533</td>
</tr>
<tr>
<td>S_MU</td>
<td>1.4026922451</td>
</tr>
<tr>
<td>S_SIGMA</td>
<td>0.4620675</td>
</tr>
<tr>
<td>CORR</td>
<td>-0.9932500146</td>
</tr>
<tr>
<td>percent censored</td>
<td>0.9986</td>
</tr>
</tbody>
</table>
%macro cnmle(data=, raw=, converge=, maxiter=, d=., out=output);
%global muhat sigmahat;
%let _a=%index(&data, .);
%if &_a=0 %then %do;
%let libname=work;
%let memname=&data;
%end;
%else %do;
%let libname=%trim(%left(%substr(&data, 1, %eval(&_a-1))));
%let memname=%trim(%left(%substr(&data, %eval(&_a+1),
%eval(%length(&data)-&_a))));
%end;

proc sql;
select nobs into %trim(%left(:nobs))
from sashelp.vtable
where libnsun=upcase("&libname")
and memneuae=upcase("&memname");
quit;
%if &converge= %then %let converge=0.000001;
%if &maxiter= %then %let maxiter=200;
%if %sysevalf(&d, boolean)=0 %then %do;
%if &d ne 0 %then %let d=&minimum;
%end;
%let min=%sysfunc(min(&d, &minimum));
data _null_
/******* Compute summary statistics of the data *******
/*****************************/
array xx(4nobs);
do i = 1 to inobs;
set &libname..&memname;
xx(i)=&raw;
if xx(i)=. then nzero+1;
else do;
   sum+xx(i); sumsq+xx(i)*xx(i); n+1;
end;
xbar=sum/n;
std= sqrt((1/n) *(sumsq-sum*sum/n)); varmle=std;
min=&min; dhat=min;
p=nzero/n;
target=std*std/((xbar-min)*(xbar-min));
/*****************************/
/******* Compute the MLEs for mu and sigma *******
/*****************************/
left=5.5;
/*****************************/
/**** Compute the Right End Point for Search ******/
lr=2; rr=5; deltar=1;
iter=0;
do while(abs(deltar)&converge);
cr=(rr+lr)/2;
deltar=(%Y(p,cr)-cr);
if deltar<=0 then rr=cr;
else lr=cr;
iter+1;
if iter>100 then stop;
end;
right=rr-.001;
*put right=;

/******************************
***** Implement Binary Search
******************************
delta=1; iter=0;
do while(abs(delta)>&converge);
   center=(right+left)/2;
   delta=(%cohen18(p,center)-target);
   if delta<=0 then left=center;
   else right=center;
   iter+1;
   if iter>1000 then do; put 'oh oh!'; stop; end;
end;
lambda=%lambda(p,center);
muhat=xbar-lambda*(xbar-min);
varhat=(std**2)+lambda*(xbar-min)*(xbar-min);
sigmahat=sqrt(varhat);
put / 03 muhat= / 03 varhat= / 03 sigmahat= /
   03 "dhat"=min / 03 "percent censored"=p ;

/******************************
***** Compute Estimated Variance of the MLEs
******************************
ds2=%dlls2(nzero,nplus,dhat,muhat,sigmahat); put ds2=;
dm2=%dllmu2(nzero,nplus,dhat,muhat,sigmahat); put dm2=;
dms=%dllms(nzero,nplus,dhat,muhat,sigmahat); put dms=;
varmu=-(ds2/(dm2*d2-dms*dms)); s_mu=sqrt(varmu);
vars=-(dm2/(dm2*d2-dms*dms)); s_sigma=sqrt(vars);
cov=-(dms/(dm2*d2-dms*dms)); corr=cov/(s_mu*s_sigma);
put s_mu= / s_sigma= / corr=;
call symput('muhat',muhat); call symput('s_mu',s_mu);
call symput('sigmahat',sigmahat); call symput('s_sigma',s_sigma);
call symput('dhat',min); call symput('corr',corr);
run;
%if &out= %then %put NO DATA SET CREATED;
%else %do;
data &out; muhat=&muhat; dhat=&dhat; sigmahat=&sigmahat;
s_mu=&s_mu; s_sigma=&s_sigma; corr=&corr; run;
%end;
%mend cnmle
A.2.3 Table Generation

We can use the macro %mle to form a look-up table using the program below. By using SAS® to write SAS® code, we can generate a sequence of macro calls for the macro %mle and write those to the external file _temp1. By submitting a %include _temp1 we execute %mle and write its results to the external file _temp2. We then use a data step to read that file. (This allows us to recycle the macro %mle without needing to rewrite its contents within a data-step and reformat its output.)

```sas
%macro mle(percent, goal);
  data _null_;
  target=&goal; p=&percent; left=-5.5; lr=-2; rr=5; deltar=1; iter=0;
  do while(abs(deltar)>&converge),*
    cr=(rr+lr)/2; deltar=(%Y(&percent,cr)-cr);
    if deltar<=0 then rr=cr; else lr=cr;
  end;
  right=rr-.001;
  delta=1; iter=0;
  do while(abs(delta)>ftconverge); 
    center=(right+left)/2;
    delta=(%cohenl8(p,center)-target) ;
    if delta<=0 then left=center; 
    else right=center;
  end;
  lainbda=*yilambda(p, center);
  put "lhs= " target fll5 "top " p <035 "table " Isimbda <055 "center " center;
  run;
%mend;

filename .tempi '/home/tomp/_junk_';
filename _temp2 '/home/tomp/ junk ';

data _null_; 
  file _temp1;
  do percent=.01 to .51 by .05;
    do goal=0 to 1.5 by .1;
      put '*/,mle( ' percent ' , ' goal ' );';
    end;
  end;
  run;
  %include .tempi;
  proc printto;
  data plot(drop=key);
    length key $3;
    infile _temp2;
    input @1 key $char3. @;
    if key='lhs';
    input @06 lhs @19 top @41 table @62 center;
  run;
```

This code uses the `mle` macro to generate a sequence of calls to the `mle` function, which is then written to an external file. The `mle` function is used to find the maximum likelihood estimate (MLE) of a parameter, in this case `p`, given a target and a goal. The resulting values are then written to a data set, which can be printed to another file for further analysis.
A.1 Utility Macros

There are a number of utility macros that you need to \%include to use the “censored with extra mass macro.”

The first SAS Macro is %process, which operates on the variable “raw” from the SAS data set “data” and generates a number of sample statistics. The stored value (in raw) corresponding to observations for which you know only that they fell below the censoring threshold must be the period, “.” (it is trivial to switch another value to a period if raw identifies censored observation in some other way).

\%macro process(data = ,raw = );
%let _a=%index(&data, );
%if _a=0 %then do; %let libname=work; %let memname=&data; %end;
%else %do; %let libname=%trim(%left(%substr(&data,1,%eval(&_a-1))));
%let memname=%trim(%left(%substr(&data,%eval(&_a+1),
%eval(length(&data)-&_a)))); %end;
proc sql;
select nobs into %left(mobs) from sashelp.vtable
where libname=upcase(&libname) and memname=upcase("&memname");
quit;
%let nobs=%trim(%left(mobs));
data sumdata(keep=sum sumsq nzero nplus dhat xbar
 varmle smle phat goal);
array xx(&nobs);
do i = 1 to &nobs;
set &libname..&memname(rename=(ftraw=tl));
if tl=0 then do; xx(i)=.; nzero+1; end;
else do; xx(i) = tl; sum+xx(i); sumsq+xx(i)*xx(i); end;
end;
nplus=&nobs-nzero; dhat=min(of xx1-xx&nobs);
xbar=sum/nplus; varmle = (1/nplus)*(sumsq-sum*sum/nplus);
smle=sqrt (v2unnle); phat = nzero/nobs;
goal = varmle/((xbar-dhat)*(xbar-dhat)); output;
run;
%mend;

The next SAS macro is %cohenls. This macro computes the conditional MLEs of \( \mu \) and \( \sigma \) for a given \((\hat{p}, \hat{d})\). This macro is not as sophisticated as the %mle (meaning that there are a few (very rare) cases that %mle will correctly resolve, but %cohenls will not). However, those instances occur only when \( s/(\hat{x} - d) \) is very small (implying that the censoring threshold is several standard deviations from the mean and casting doubt on the usefulness of treating the data as “censored” in the first place). In cases where the macro does not resolve, it is probably more sensible to treat the problem as being a mixed “normal” and point mass at zero (rather than censored normal with point mass at zero).

\%macro cohenls(p, h, cohen, converge);
data mle;
  keep center delta lambda p h;
p=&p;h=&h;target=&cohen;
converge=&&converge;
left=-5.5; right=0; delta=1;
do while(abs(delta)>&converge);
  center=(right+left)/2; cohen18=%cohen18(p,h,center);
  delta=(%cohen18(p, h, center)-target);
  if delta<=0 then left=center; else right=center; end;
  xi=center;
  lambda=%lambda(p, h, xi);
end;
%mend;

A.3.2 Function Macros

Macros with Definitions

Here are some additional functions that you need to %include to use the censored with extra mass macro.

The first 5 functions require (p, h, xi) be specified

\[
\%W \left( \frac{h}{1-h} \right) \left( \frac{(1-p)\phi\left( \frac{d-\mu}{\sigma} \right)}{p+(1-p)\Phi\left( \frac{d-\mu}{\sigma} \right)} \right)
\]

\[
\%cohen18 \left( \frac{h}{1-h} \right) \left( \frac{(1-p)\phi(\xi)}{(\hat{\beta}+(1-\hat{\beta})\Phi(\xi))} \right)
\]

\[
\%lambda \left( W(\hat{\beta}, h, \xi)/(W(\hat{\beta}, h, \xi) - \xi) \right)
\]

\[
\%loglike \left( n_{\infty} \log \left[ \hat{\beta} + (1 - \hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right) \right] + n_2 \left( \log(1 - \hat{\beta}) - \log \sigma \right) - \frac{n_2}{2} \log(2\pi) \right)
\]

\[
\%dllmu \left( \frac{n_{\infty}(1-\hat{\beta})\phi\left( \frac{d-\mu}{\sigma} \right)(\frac{d-\mu}{\sigma})}{\hat{\beta}+(1-\hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right)} + \frac{n_2(\hat{x}-\hat{\mu})}{\sigma^2} \right)
\]

\[
\%dlls \left( \frac{n_{\infty}(1-\hat{\beta})\phi\left( \frac{d-\mu}{\sigma} \right)(\frac{-d-\mu}{\sigma^2})}{\hat{\beta}+(1-\hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right)} - \frac{n_2}{\sigma} + \frac{n_2(\hat{x}-\hat{\mu})^2+(n_2-1)x^2}{\sigma^4} \right)
\]

\[
\%dllmu2 \left( \frac{n_{\infty}(1-\hat{\beta})\left[ \phi(\hat{\beta}+(1-\hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right)) \right]\left[ \phi(\hat{\beta}+(1-\hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right)) \right]-\left[ \frac{n_2\phi\left( \frac{d-\mu}{\sigma} \right)^2}{\hat{\beta}+(1-\hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right)} \right]^2 }{\hat{\beta}+(1-\hat{\beta})\Phi\left( \frac{d-\mu}{\sigma} \right)^2} - \frac{n_2}{\sigma^2} \right)
\]
$%dlls2 \quad n_{\infty}(1-p) \left\{ \left[ \hat{p} + (1-p) \Phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \right] \phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \left( \frac{2(\hat{d}-\hat{\mu})}{\sigma^2} \right) - \phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \left( \frac{-\hat{d}-\hat{\mu}}{\sigma^2} \right) \right\} $ \\

$\quad - \frac{n_{\infty}(1-p) \left[ \hat{p} + (1-p) \Phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \right]^2}{\left[ \hat{p} + (1-p) \Phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \right]^2} \quad + \frac{m}{\sigma^2} - 3(n_{\infty}(\bar{d}-\hat{\mu})^2 + (n_{\infty}-1)s^2) $ \\

$%dllms \quad n_{\infty}(1-p) \left\{ \left[ \hat{p} + (1-p) \Phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \right] \phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \left( \frac{-\hat{d}-\hat{\mu}}{\sigma^2} \right) + \phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \left( \frac{1}{\sigma^2} \right) \right\} $ \\

$\quad - \frac{n_{\infty}(1-p) \left[ \hat{p} + (1-p) \Phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \right]^2}{\left[ \hat{p} + (1-p) \Phi \left( \frac{d-\hat{\mu}}{\sigma} \right) \right]^2} \quad - \frac{2m(\bar{d}-\hat{\mu})}{\sigma^3} $ \\

* h=sample fraction of zero-counts;  
* p=extra mass probability; 

%macro W(p, h, xi);  
(\&h/(1-\&h))*( (1-\&p)*%smallphi(\&xi)/(\&p+(1-\&p)*probnorm(\&xi)) )  
%end; 

%macro cohen18(p, h, xi);  
((W(\&p,\&h,\&xi)-\&xi))/ 
(((W(\&p,\&h,\&xi)-\&xi)*(W(\&p,\&h,\&xi)-\&xi))  
%end; 

%macro lambda(p, h, xi);  
W(\&p,\&h,\&xi)/(W(\&p,\&h,\&xi)-\&xi)  
%end; 

%macro loglike(nzero,nplus,phat,dhat,muhat,sigmahat);  
&nzero*\log(\&phat+(1-\&phat)*probnorm( %arg(\&dhat,\&muhat,\&sigmahat) ))  
+&nplus*( \log(1-\&phat)-\log(\&sigmahat) )  
-(.5)*(/(\&sigmahat*\&sigmahat)*  
(\&nplus*\varmle+\&nplus*(\xbar-\&muhat)\xbar-\&muhat))  
%end loglike; 

%macro dild(nzero,nplus,phat,dhat,muhat,sigmahat);  
(&nzero*(1-\&phat)*\&phii(\&arg(\&dhat,\&muhat,\&sigmahat))/ 
(\&sigmahat+(\&phat*(1-\&phat)*probnorm( (\&dhat-\&muhat)/\&sigmahat ))))  
%end dild; 

%macro dllmu(nzero,nplus,phat,dhat,muhat,sigmahat);  
(-&nzero*(1-\&phat)*\&phii(\&arg(\&dhat,\&muhat,\&sigmahat))/ 
(\&sigmahat+(\&phat*(1-\&phat)*probnorm( (\&dhat-\&muhat)/\&sigmahat )))*  
(1/(\&sigmahat*\&sigmahat))*\&nplus*(\xbar-\&muhat))  
%end dllmu; 

%macro dlls(nzero,nplus,phat,dhat,muhat,sigmahat);  
(-&nzero*(1-\&phat)*\&phii(\&arg(\&dhat,\&muhat,\&sigmahat))/ 
(\&sigmahat+(\&phat*(1-\&phat)*probnorm( (\&dhat-\&muhat)/\&sigmahat )))-  
&nplus/\&sigmahat  
+(\&nplus*\varmle+\&nplus*(\xbar-\&muhat)\xbar-\&muhat))/\&sigmahat**3
%mend dlls;
%macro dllmu2(nzero,nplus,phat,dhat,muhat,sigmahat);
&nzero*((( ( &phat-1)/&sigmahat )*((&dhat-&muhat)/&sigmahat**2)*%phi(%arg(&dhat,&muhat,&sigmahat))*)
&phat+(1-&phat)*probnorm( (%dhat-&muhat)/&sigmahat ))
-(((1-&phat)/&sigmahat)*%phi(%arg(%dhat,&muhat,&sigmahat))**2)/
(&phat+(1-&phat)*probnorm( (%dhat-&muhat)/&sigmahat ))**2
-&nplus/&sigmahat**2
%mend dllmu2;

%macro dlls2(nzero,nplus,phat,dhat,muhat,sigmahat);
&nzero*(&phat+(1-&phat)*probnorm( (%dhat-&muhat)/&sigmahat ))*
( (1-&phat)*%phi(%arg(&dhat,&muhat,&sigmahat))*
-(&dhat-&muhat)**3/&sigmahat**5+2*(&dhat-&muhat)/&sigmahat**3 ))-
((1-&phat)*%phi(%arg(&dhat,&muhat,&sigmahat))*
((&dhat-&muhat)/&sigmahat**2))**2)/
(&phat+(1-&phat)*probnorm( (%dhat-&muhat)/&sigmahat ))**2
+nplus/&sigmahat**2
-3*(&nplus*varmle+nplus*(xbar-&muhat)*(xbar-&muhat))/&sigmahat**4
%mend dlls2;

%macro dllms(nzero,nplus,phat,dhat,muhat,sigmahat);
&nzero*(1-&phat)*
((( &phat+(1-&phat)*probnorm( (%dhat-&muhat)/&sigmahat ))*
%phi(%arg(&dhat,&muhat,&sigmahat))*
-(&dhat-&muhat)**2/&sigmahat**4)+
%phi(%arg(&dhat,&muhat,&sigmahat))
/&sigmahat**2 ) -
(1-&phat)*
%phi(%arg(&dhat,&muhat,&sigmahat))**2*
(&dhat-&muhat)**2/&sigmahat**2)/
(&phat+(1-&phat)*probnorm( (%dhat-&muhat)/&sigmahat ))**2
-2*&nplus*(xbar-&muhat)/&sigmahat**3
%mend dllms;
A.3.2 Extra Mass Program

By specifying the data and raw positional parameters \%exmass will compute the MLEs for the "normal distribution with extra mass". \%exmass stores the output in the SAS data set "work.em". Among the output variables are:

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZERO</td>
<td>number of missing counts,</td>
</tr>
<tr>
<td>SUM</td>
<td>of the non-missing values,</td>
</tr>
<tr>
<td>SUMSQ</td>
<td>uncorrected sum of squares of non-missing values,</td>
</tr>
<tr>
<td>NPLUS</td>
<td>number of positive counts,</td>
</tr>
<tr>
<td>DHAT</td>
<td>sample minimum,</td>
</tr>
<tr>
<td>XBAR</td>
<td>sum/nplus,</td>
</tr>
<tr>
<td>VARMLE</td>
<td>sample variance of nonmissing observations (1/NPLUS version),</td>
</tr>
<tr>
<td>SMLE</td>
<td>square root of VARMLE,</td>
</tr>
<tr>
<td>H</td>
<td>nzero/n,</td>
</tr>
<tr>
<td>GOAL</td>
<td>varmle/((xbar-dhat)*(xbar-dhat)),</td>
</tr>
<tr>
<td>NEW_P</td>
<td>the MLE estimate of ( p ),</td>
</tr>
<tr>
<td>MUHAT</td>
<td>the MLE estimate of ( \mu ),</td>
</tr>
<tr>
<td>SIGMAHAT</td>
<td>the MLE estimate of ( \sigma ),</td>
</tr>
<tr>
<td>DHAT</td>
<td>the MLE estimate of ( d ),</td>
</tr>
<tr>
<td>LOGLIKE</td>
<td>( L(d, p, \mu, \sigma) ).</td>
</tr>
</tbody>
</table>

%macro exmass(data, raw);
%let stop=run;
%let old=1;
%process(data=&data, raw=&raw);
data _null_; set sumdata;
call symput('h', phat);
call symput('p', phat);
call symput('new_p', phat);
call symput('cohen', goal);
run;
%do %while(&stop ne stop);
%cohenls(&new_p, &h, &cohen, .000001);
data em;
merge mle sumdata;
sigmahat=sqrt(smle*smle + lambda*(xbar-dhat)*(xbar-dhat));
muhat=xbar - lambda*(xbar-dhat);
new_p = max(0,1-(nplus/(nzero+nplus))/
(1-probnorm((dhat-muhat)/sigmahat)) );
call symput('new_p',new_p);
loglike=%loglike(nzero,nplus,new_p,dhat,muhat,sigmahat);
old=symget('old');
if abs(old-loglike)<.0001 then do;
call symput('stop','stop');
end;
call symput('old',loglike);
run;
%end;
%mend macro;
A.3.2 Newton Raphson approach

To solve for MLEs in the "normal with extra mass" model where you have "moderate" censoring, Newton's approach works quickly. Presented here is a program, %simemass, that incorporates Newton Raphson (built into a simulation) programmed in SAS® IML.

To run the program you must first %include the macros %simdata and, during execution, have a filename reference pointing to program %newton. For example

```
filename newton '/home/tomp/St699/NR.Extra.Mass/Macro/newton.sas';
```

This Macro Simulates SAMPLES number of data sets (where SAMPLES is an integer greater than or equal to one). For each data set simulated you need to supply the following parameters:

-nob - number of observations per data set sampled

- p - the extra percent to be censored

- mu - the mean of the normal

- sigma - the std dev of the normal

- d - the left threshold of detection

SAS returns the SAS data set EM.MLE<SAMPLES> where, for each data set sampled, there is a single observation. The variables in the new data set are the following:

- P "MLE for parameter p"

- MU "MLE for parameter mu"

- SIGMA "MLE for parameter sigma"

- NPLUS "Number of detects in sample"

- DHA "MLE for lower threshold of detection"

- ITER "Number of Iterations to Convergence"

- LOGLIKE "Loglikelihood value at MLE"

- OLDLOGLK "Loglikelihood at parms generating data"

- XBAR "Sample mean of the detects"

- STD "Sample standard deviation of detects";

```
%macro simemass(samples=,nob=,p=,d=,mu=,sigma=);
%global nplus dhat xbar std;
%include simulate;
proc sql;
create table mle<nob>
(P num label="MLE for parameter p",
 M num label="MLE for parameter mu",
 S num label="MLE for parameter sigma",
 D num label="MLE for lower threshold of detection",
 LL num label="Loglikelihood value at MLE",
 OLDLOGLK num label="Loglikelihood at parms generating data",
 D2M num label="2nd Partial of Log Likelihood wrt mu",
 D2S num label="2nd Partial of Log Likelihood wrt sigma",
 D2P num label="2nd Partial of Log Likelihood wrt p",
 DSM num label="Mixed Partial of LogLike wrt mu sigma",
```
DPM     num label="Mixed Partial of LogLike wrt mu and p",
DPS     num label="Mixed Partial of LogLike wrt p and sigma",
NPLUS   num label="Number of detects in sample",
NZERO   num label="Number of non-detects in sample",
XBAR    num label="Sample mean of the detects",
STD     num label="Sample standard deviation of detects",
ITER   num label="Number of Iterations to Convergence";
quit;
do ii=1 %to &samples;
  %simdata(n=&nobs,p=&p,d=&d,mu=&mu,sigma=&sigma);
%include newton;
proc datasets lib=Hork nolist;
  append base=aletoobs neH=sortset force;
quit;
%end;
%mend emass;

%macro simdata(n=,p=,d=,mu=,sigma=);
data_null;
n=&n;p=&p;d=&d;m=mu;sigma=&sigma;
array xx(4n);
do i=1 to &n;
  tl=ranbin(0,1,1-&p);
  if tl=0 then do;
    xx(i)=.;
    nzero+1;
  end;
  else do;
    temp=&sigma*rannor(0)+μ
    if temp < &d then do;
      xx(i) = .;
      nzero+1;
    end;
    else do;
      xx(i) = temp; siim+xx(i); sumsq+xx(i)*xx(i);
    end;
  end;
  end;
nplus=&n-nzero;
dhat=min(of xx1-xx&n);
xbar=sum/nplus;
std=sqrt((1/(nplus-1))*(sumsq-sum*sum/nplus));
call symput('nplus',nplus);
call symput('dhat',dhat);
call symput('xbar',xbar);
call symput('std',std);
run;
%mend;
This is the program "newton".

PROC IML;

*------------------------- Define Functions Needed (Log Likelihood and Derivatives) ---------------------;

start ll;
   ll= nzero#log(p+(1-p)#probnorm((dhat-mu)/s))+ nplus#(log(1-p)-log(s))-.5#(s)##(-2)#(nplus#(xbar-mu)##2+(nplus-1)#std##2);
finish ll;

start oldll;
   oldll=nzero#log( po+(1-po)#probnorm( (dhato-muo)/so ))+
   nplU3#( log(l-po)-log(3o) )-.5#(so)##(-2)#( nplus#(xb£u:-muo)##2+(nplus-l)#std##2 )
finish oldll;

start newp;
   old = p;
   p=max(1-(nplus/(nzero+nplus)) / (1-probnorm((dhat-mu)/s)),0);
finish newp;

start normpdf;
   arg=(dhat-mu)/s;
   smallphi=((2#gamma(.5)##2)##(-.5)#exp(-.5#(arg#arg)));
   denom=(p+(1-p)#probnorm( (dhat-mu)/s ));
   bigphi=probnorm((dhat-mu)/s);
finish normpdf;

start dm;
   dm=nzero#((p)->smallphi)/(s#(p+(1-p)#probnorm( (dhat-mu)/s )))+
   nplus#(xbar-mu)/(s##s);
finish dm;

start d2m;
   d2m=nzero#((p)-#smallphi#(dhat-mu)/s##2)/denom##2
   -nplus/s##2+(nplus#(xbar-mu)##2+(nplus-1)#std##2)/s#3;
finish d2m;

start ds;
   ds=(nzero#(((p)-#smallphi#(-arg))/#(denom))-nplus/s
   +nplus#(xbar-mu)##2+(nplus-1)#std##2)/s##3;
finish ds;

start d2s;
   d2s=nzero#((denom#((p)-#smallphi#(-dhat-mu)/s##2)##2)/denom##2
   -nplus/s##2-3#(nplus#(xbar-mu)##2+(nplus-1)#std##2)/s##4;
finish d2s;

start dsm;
   dsm=nzero#(((p)-#smallphi#((-dhat-mu)/s##2)+#smallphi#(dhat-mu)/s##2)/denom##2
   -2#nplus#(xbar-mu)/s##3;
finish dsm;

start d2p;
   suml=-nzero#(1-p)#(denom#smallphi#(dhat-mu)/s##2)/denom##2;
   sum2=-nplus/(1-p)##2;
   d2p=suml-sum2;
finish d2p;
start dpm;
  dpm=nnzro#(denom#(smallphi/s) -
          (1-bigphi)#(1-p#smallphi/s))/denom##2;
finish dpm;

start dps;
  dps=nnzro#(denom#(smallphi#arg/s) -
           (1-bigphi)#(1-p#smallphi#arg/s))/denom##2;
finish dps;

*--------- Define Constants Needed From the Sample------------------*
 nplus=nnplus; /* Number of Positive Counts from Sample */
nzero=&nobs-nplus; /* Number of Non-Detects from Sample */
dhat =&dhat; /* Sample Minimum (MLE of lower threshold) */
xbar =&xbar; /* Sample Mean of Detects */
std =&std; /* Sample (Unbiased) Std Dev of detects */

*--------- Specify Starting Values for Parameter Estimators----------*
  s =&std; /* Use Unb Std Dev as initial Est. of sigma */
  mu =xbar; /* Use sample mean as initial Est. of mu */
  p=2; run newp; /* Initial estimate of P */
  x=(mu/s); /* vector of initial estimators of mu and sigma */

*--------- Specify Program Control Constants-----------------------*
 converge=.00000001; /* Conv. crit. for Newton Rhapson */
 converg2=.00001; /* Conv. crit. for updating p */
 maxiter=1000; /* maximum number of NR iterations and */
                /* max num p-update iterations */

*--------- Transfer Simulation Parameters--------------------------*
  po=&p; muo=&mu; so=&sigma; dhato=&d;

start x; /*--- Newton's Method for Solving for a Maximum ----------*/
  mu=x[1,1]; s=x[2,1]; /* is set to 0 since you maximize a log- */
  run normpdf; /* likelihood by solving the likelihood eqn */
  run dm; /* which are the equations where the first */
  run ds; /* partials are set to zero */
  f=(dm/ds);
finish x;

start J; /* J is the jacobian, the matrix of 2nd */
  run d2m; /* partials of the log-likelihood with */
  run d2s; /* respect to mu and sigma */
  run dsm;
  j=(d2m|d2s)/(dsm|d2s);
finish J;
start newton;
    run F;
    do iter=1 to maxiter while(max(abs(f))>converge);
        run J;
            x=x-solve(j,f);
        run F;
    end;
    mu=x[1]; s=x[2];
finish newton;

start info;
    run d2p;
    run dps;
    run dpm;
    info=(d2m ||| dsm ||| dpm )//
        (dsm ||| d2s ||| dps )//
        (dpm ||| dps ||| d2p );
    cov=-inv(info);
    diag=sqrt(diag(coT));
    corr=(cov[1,2]/sqrt(cov[1,1]*cov[2,2]))//
         (cov[1,3]/sqrt(cov[1,1]*cov[3,3])) //
         (cov[2,1]/sqrt(cov[2,2]*cov[3,3]))//
         (cov[3,1]/sqrt(cov[1,1]*cov[3,3]))//
         (cov[3,2]/sqrt(cov[2,2]*cov[3,3])) //
         (cov[3,3]/sqrt(cov[3,3]));
    print info cov;
    print diag corr;
finish info;

start newmle;
    outline=(p ||| mu ||| s ||| dhat ||| ll ||| oldll ||| d2m ||| d2s ||| d2p ||| dsm ||| dpm ||| dps ||| nplus ||| nzero ||| xbar ||| std ||| click);
    create sortset from outline[colname='p','m','s','d','ll','oldll','d2m','d2s','d2p','dsm','dpm','dps','nplus','nzero','xbar','std','click'];
    append from outline;
    close sortset;
finish newmle;

start Main;
    do click=1 to maxiter while(abs(p-old)>converg2);
        run newton;
        run newp;
    end;
run ll;
run oldll;
run info;
run newmle;
finish main;

run main;
quit;
BIBLIOGRAPHY


