ESTIMATION OF PARAMETERS IN THE GENERALIZED BETA DISTRIBUTION

BY

OWEN WHITBY

TECHNICAL REPORT NO. 29
NOVEMBER 3, 1971

PREPARED UNDER THE AUSPICES OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-14

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATION OF PARAMETERS IN THE GENERALIZED BETA DISTRIBUTION

BY

OWEN WHITBY

TECHNICAL REPORT NO. 29
November 3, 1971

PREPARED UNDER THE AUSPICES
OF
PUBLIC HEALTH SERVICE GRANT USPHS-5TI GM 25-14

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Notation and Preliminary Results</td>
<td>3</td>
</tr>
<tr>
<td>3. General Existence and Distribution Theorems</td>
<td>8</td>
</tr>
<tr>
<td>4. Regular Case (&gt; 2)</td>
<td>14</td>
</tr>
<tr>
<td>5. Boundary Non-regular Case (= 2)</td>
<td>23</td>
</tr>
<tr>
<td>6. Non-regular Case (&lt; 2)</td>
<td>34</td>
</tr>
<tr>
<td>7. Linearization of Likelihood Equations</td>
<td>48</td>
</tr>
<tr>
<td>8. Initial Estimators</td>
<td>59</td>
</tr>
<tr>
<td>9. Final Remarks</td>
<td>67</td>
</tr>
<tr>
<td>References</td>
<td>68</td>
</tr>
</tbody>
</table>
1. **Introduction**

The specific problem considered is that of estimating the parameters of the generalized Beta distribution. Many of the results, however, are stated and proven for a general distribution. The method of maximum likelihood is chiefly considered because the order statistic is minimal sufficient whenever at least one of the end points is unknown, and, therefore, no pretty estimators can be expected.

The generalized Beta distribution with parameters $a$, $b$, $\gamma$, $\delta$ ($-\infty < a < b < +\infty$, $\gamma > 0$, $\delta > 0$) is denoted by $B_{\gamma,\delta}(a,b)$ and has density

$$f(y; \theta) = \begin{cases} \frac{1}{B(\gamma,\delta)} \frac{1}{(b-a)^{\gamma+\delta-1}} (y-a)^{\gamma-1} (b-y)^{\delta-1}, & \text{for } a < y < b \\ 0, & \text{otherwise} \end{cases}$$

where $\theta = (a,b,\gamma,\delta)'$ and $B(\gamma,\delta)$ is the beta function. The parameters $a$ and $b$ are the left and right end points respectively, and $\gamma$ and $\delta$ are the governing shape parameters for $a$ and $b$ respectively.

The existence, consistency, and asymptotic normality and efficiency of a root of the likelihood equations are usually proven under conditions similar to those given by Cramér [1946, 500-501] or Kulldorff [1957], which, among other things, allow Taylor expansion of the derivative of the log-likelihood function in a fixed neighborhood of the true parameter value. For the generalized Beta distribution such
conditions cover the case of maximum likelihood estimation of the shape parameters when the end points are known. Gnanadesikan et al. [1967] discuss the numerical solution of the likelihood equations in this case.

When estimating at least one end point of the generalized Beta distribution no such fixed neighborhood of Taylor expansion validity exists, but if the governing shape parameters are large enough (> 2, regular case) the conditions can be weakened to allow Taylor expansion in a sequence of shrinking neighborhoods, and the usual asymptotic results, with \( n^{1/2} \) normalization, can be proven. Section 4 contains these results.

Section 5 presents similar results, with \( (n \ln n)^{1/2} \) normalization, when the governing shape parameters are equal to two (boundary non-regular case). A partial answer to the asymptotic distribution when the governing shape parameter is between one and two (non-regular case) is given in Section 6.

Solution of the likelihood equations, including linearization (for the regular and boundary non-regular cases) and selection of initial estimators, is discussed in Sections 7 and 8.
2. **Notation and Preliminary Results**

The following notation will be used as required. All vectors are to be interpreted as column vectors unless otherwise noted. The Euclidean norm, or any equivalent norm, on \( \mathbb{R}^p \) (the \( p \)-fold product space of the reals, \( \mathbb{R} \)) is denoted by \( \| \cdot \| \). \( S_{\eta}(x) \) and \( \bar{S}_{\eta}(x) \) denote, respectively, the open and closed spherical neighborhoods of \( x \) with radius \( \eta \). Matrices, \( p \times p \), will be considered as elements of \( \mathbb{R}^{p^2} \) with norm \( \| \cdot \| \) when required. \( \mathcal{J} > 0 \) means that \( \mathcal{J} \) is a symmetric, positive definite matrix.

The \( p \)-dimensional multivariate normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \) will be denoted by \( \mathcal{N}_p(\mu, \Sigma) \). \( \mathcal{Z} \) or \( \mathcal{Z}_p \) will denote a \( \mathcal{N}_p(0, I) \) random vector. Occasionally \( \mathcal{N}_p(\mu, \Sigma) \) itself will be used to denote a random vector with this law. \( N(\mu, \nu) \) is used instead of \( \mathcal{N}_1(\mu, \nu) \). \( Y_{1n} < Y_{2n} < \cdots < Y_{nn} \) will denote the ordered observations from a simple random sample \( Y_1, Y_2, \ldots, Y_n \).

If \( \{E_n\} \) is a sequence of measurable events, then for a pre-selected \( \epsilon > 0 \), the phrase "\( E_n \) occurs with all but small probability" \( (E_n \text{ wabsp}) \) means that \( \exists \ n_{\epsilon} \) such that \( P(E_n) > 1 - \epsilon \ \ \forall \ n \geq n_{\epsilon} \). "In probability" (i.p.) is used in the sense of Pratt [1959]. That is, an event \( E \) (for which measurability need not even be defined) is said to occur i.p. iff \( \forall \ \epsilon > 0, \ \exists \ \text{measurable events} \ E_n(\epsilon) \ \text{such that} \)

\[ (i) \ \ P(E_n(\epsilon)) > 1 - \epsilon \ \ \forall \ n, \ \text{and} \ \ (ii) \ \ \bigcap_{n=1}^{\infty} E_n(\epsilon) \subseteq E. \]

For a sequence of measurable events \( \{E_n\} \), "\( E_n \) i.p." is used to mean that
\[ E = \{ x : \exists n_x \text{ such that } x_n \in E_n \forall n \geq n_x \} \text{ occurs i.p.; or equivalently,} \]
\[ \forall \epsilon > 0, E_n \text{ w.p.} \]  

The Mann-Wald symbols, \( O_p \) and \( o_p \), have the usual interpretation. Thus, if \( \langle g(n) \rangle \) is a sequence of positive numbers, \( Z_n = O_p(g(n)) \) and \( Z_n = o_p(g(n)) \) mean that the sequence of random variables \( \{ z_n \} \) is bounded in probability and converges to zero in probability, respectively.

To complete this section a result is noted and two useful lemmata concerning \( B_\gamma, \delta(a, b) \) are proven. Results for the asymptotic distribution of extremes (e.g., Gnedenko [1945], Gumbel [1960]) assert, for \( B_\gamma, \delta(a, b) \), that, as \( n \rightarrow +\infty \), the distribution of the minimum of \( n \) observations is given by

\[
P( y \leq \frac{1}{n} (Y_{1n} - a) > y ) = \begin{cases} 
\exp \left[ - \frac{1}{\gamma B(\gamma, \delta)} \left( \frac{y}{b-a} \right)^\gamma \right] + o(1), & \text{for } y > 0 \\
1, & \text{for } y \leq 0
\end{cases}
\]

and the analogous result holds for \( Y_{nn} \).

**Lemma 2.1:** If \( Y_1, Y_2, \ldots \) are i.i.d. \( B_\gamma, \delta(a, b) \), and if
\[ a_n = a + o(n^{-1/\gamma}) \], then

(a) for each \( \epsilon \geq 0 \),
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{1}{(Y_i - a_n)^\epsilon} = \begin{cases} 
o_p(n^{\epsilon/\gamma - 1}), & \forall \gamma \in (0, \epsilon) \text{ if } \gamma \leq \epsilon \\
o_p(1), & \text{if } \gamma > \epsilon,
\end{cases}
\]
(b) \[ \frac{1}{\ln n} - a_n = O_p(n^{-1/\gamma}). \]

Moreover, if \( a_n \) is a random variable such that \( a_n = a + o_p(n^{-1/\gamma}) \), the results still hold.

Proof: Consider \( X_1, X_2, \ldots \) i.i.d. \( \mathcal{B}(0,1) \). For any \( \epsilon > 0 \), let \( W_i = X_i^{-\epsilon} \). Then \( \mathbb{E}[W_i^{\gamma'/\epsilon}] = \mathbb{E}[X_i^{-\gamma}] < +\infty \) iff \( \gamma' < \gamma \). For \( 0 < \gamma' < \gamma \leq \epsilon \), a theorem of Marcinkiewicz (Loève [1963], 242-3) shows that \( (n^{\epsilon'/\gamma'})^{-1} \sum_{i=1}^{n} W_i = (n^{\epsilon'/\gamma'})^{-1} \sum_{i=1}^{n} (X_i^{-\epsilon})^{-1} \) a.s. \( \rightarrow 0 \). This easily yields \( n^{-1} \sum_{i=1}^{n} (X_i^{-\epsilon})^{-1} = o_p(n^{\epsilon'/\gamma'-1}) \) for \( 0 < \gamma' < \gamma \leq \epsilon \). The W.L.L.N. gives \( n^{-1} \sum_{i=1}^{n} (X_i^{-\epsilon})^{-1} = o_p(1) \) for \( \gamma > \epsilon \), since \( \mathbb{E}[X_i^{-\epsilon}] < +\infty \) in this case.

Next it is shown that \( \forall \epsilon \in [0,1) \), and \( \forall \epsilon > 0 \),

\[ [Y_i - a - o(n^{-1/\gamma})]^{-\epsilon} \geq c(Y_i - a)^{-\epsilon}, \quad \forall i = 1, 2, \ldots, n, \text{ i.p.} \]

That is

\[ (1 - c^{1/\epsilon})(Y_i - a) \geq o(n^{-1/\gamma}), \quad \forall i = 1, 2, \ldots, n, \text{ i.p.} \]

But
\[ P((1-c^{-1/\varepsilon})(Y_1-a) \geq o(n^{-1/\gamma}), \quad \forall 1 \leq 1, 2, \ldots, n) \]

\[ = P \left\{ Y_1 - a \geq \frac{o(n^{-1/\gamma})}{1 - c^{-1/\varepsilon}} \right\} \]

\[ = \exp \left\{ - \frac{1}{\gamma B(\gamma, \delta)} \left[ \frac{o(1)}{b-a} \right] \right\} + o(1) \longrightarrow 1, \text{ as } n \rightarrow +\infty, \]

as is required.

Use of these two preliminary results in reverse order yields

(a) by showing that, i.p.,

\[ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(Y_i - a)^{\varepsilon}} \leq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{c(Y_i - a)^{\varepsilon}} = \frac{1}{c(b-a)^{\varepsilon}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{X_i^{\varepsilon}} \]

\[ = \begin{cases} 
  o_p(n^{\varepsilon/\gamma - 1}), & \text{if } 0 < \gamma' < \gamma \leq \varepsilon \\
  o_p(1), & \text{if } \varepsilon < \gamma .
\end{cases} \]

Now, for any $\lambda > 0$,

\[ P \left\{ \left| \frac{1}{Y_{1n} - a} \right| \leq \lambda n^{1/\gamma} \right\} = P \left\{ \left| Y_{1n} - a \right| \geq \frac{1}{\lambda n^{1/\gamma}} \right\} \]

\[ \geq P \left\{ Y_{1n} - a \geq \frac{1}{\lambda n^{1/\gamma}} \right\} \]

\[ = P(Y_{1n} - a \geq o(n^{-1/\gamma}) + \lambda^{-1} n^{-1/\gamma}) \]

\[ = \exp \left\{ - \frac{1}{\gamma B(\gamma, \delta)} \left[ \frac{o(1) + \lambda^{-1}}{b-a} \right] \right\} + o(1), \text{ from (2.1)} \]

\[ \longrightarrow \exp \left\{ - \frac{1}{\gamma B(\gamma, \delta)} \left[ \frac{1}{\lambda(b-a)} \right] \right\}, \quad \text{as } n \rightarrow +\infty \]

\[ \longrightarrow 1, \quad \text{as } \lambda \rightarrow +\infty. \]
Hence, \((Y_{\ln a_n})^{-1} = o_p(n^{1/\gamma})\).

When \(a_n\) is a random variable, the results follow directly from the "in probability" theorems of Pratt [1959] which state roughly that the algebra of 0 and \(o\) extends to \(o_p\) and \(o_p\). \(\|\)
3. **General Existence and Distribution Theorems**

The following lemma is an existence theorem for a root of a system of equations. \( g(\cdot), z, x, x_0 \) and \( \zeta(\cdot) \) are elements of \( \mathbb{R}^p \), \( p \geq 1 \), and \( A \) is a \( p \times p \) matrix. \( J_G(x) \) is the matrix of first partial derivatives of \( g \) wrt \( x \) at \( x \).

**Lemma 3.1:** If \( \exists \ z, A, \zeta(\cdot), \eta > 0 \) such that

(i) \( g(x) = z + A(x - x_0) + \zeta(x) \), \( \forall x \in S_{2\eta}(x_0) \)

(ii) \( |A| \neq 0 \)

(iii) \( \|A^{-1}z\| < \eta \)

(iv) \( \|A^{-1}\zeta(x)\| < \eta/2 \), \( \forall x \in S_{3\eta}(x_0) \)

(v) \( g(x) \) is differentiable, wrt each coordinate of \( x \), in \( S_{3\eta}(x_0) \)

(vi) \( |J_G(x)| \neq 0 \), \( \forall x \in S_{3\eta}(x_0) \),

then \( \exists \) a solution \( \hat{x} \) of \( g(x) = 0 \) such that \( \hat{x} \in S_{3\eta}(x_0) \).

**Proof:** This lemma is actually a slightly stronger version of the inverse function theorem (the size of the neighborhood of invertibility is also estimated), and reference is made to the proof of this theorem as it appears on pp. 177-9 of Rudin [1953].

First, define \( x_L = x_0 - A^{-1}z \), and note that it is the root of \( g + A(x - x_0) = 0 \). Observe that \( S_{2\eta}(x_L) \subset S_{3\eta}(x_0) \), since
\[ \| \mathbf{x}_L - x_o \| = \| A^{-1} \mathbf{z} \| < \eta \] by (iii). Next, define \( G_A(\cdot) \) by
\[ G_A(z) = A^{-1} \Gamma(z) = A^{-1} \mathbf{z} + (z - x_o) + A^{-1} \mathbf{z}(x_L), \]
and define \( x_L = G_A(x_L) \), noting in passing that \( \| x_L \| = \| A^{-1} \mathbf{z}(x_L) \| < \eta/2 \) by (iv). By (vi)
\[ |J_{G_A}(\hat{z})| = |A^{-1}| |J_{\mathbf{z}}(\hat{z})| \neq 0 \quad \forall \hat{z} \in S_{2\eta}(x_L), \]
and, therefore, in Rudin's proof the closed sphere on which \( \mathbf{f} = G_A \) is one-to-one can be taken as \( S = \tilde{S}_{2\eta}(x_L) \). The boundary of \( S \) is then \( T = \{ \hat{z} : \| \hat{z} - x_L \| = 2\eta \}. \)
Now, for \( \hat{z} \in T,
\[
\| G_A(\hat{z}) - x_L \| = \| A^{-1} \mathbf{z} + \hat{z} - x_o + A^{-1} \mathbf{z}(\hat{z}) - A^{-1} \mathbf{z}(x_L) \|
= \| \mathbf{z} - x_L + A^{-1} \mathbf{z}(\hat{z}) - A^{-1} \mathbf{z}(x_L) \|
\geq \| \mathbf{z} - x_L \| - \| A^{-1} \mathbf{z}(\hat{z}) - A^{-1} \mathbf{z}(x_L) \|
> 2\eta - (\frac{\eta}{2} + \frac{\eta}{2}) = \eta > 2\| x_L \|. \]

Hence, in Rudin's proof \( \lambda \) can be chosen so that \( \lambda > \| x_L \|. \) Therefore,
\( Q \in V = S_{\lambda}(x_L), \) which is an open sphere with the property that
\( \forall \mathbf{z} \in V, \exists \hat{z} \in S_{2\eta}(x_L) \) such that \( G_A(\hat{z}) = \mathbf{z}. \) In particular,
\( \exists \hat{x}_A \in S_{2\eta}(x_L) \) such that \( G_A(\hat{x}_A) = Q. \) But \( G(\hat{x}_A) = AG_A(\hat{x}_A) = Q. \)
Therefore, \( \hat{x} = \hat{x}_A \) is a solution of \( G(\hat{x}) = Q \) and \( \hat{x} \in S_{2\eta}(x_L) \subset S_{3\eta}(x_o). \)

Let \( P \) be the probability measure for the random sequence
\( \mathbf{x} = (x_1, x_2, x_3, \ldots) \). When the probability measures are indexed by
a parameter \( \theta \), the absence of a specific indexing value shall indicate
that the "null" or "true" value \( \theta_0 \) is intended. Thus \( P = P_{\theta_0}. \)

9
First to be considered is the problem of estimating a parameter \( \tilde{\theta} \in \Theta \subset \mathbb{R}^p \) by the method of maximum likelihood (in the loose sense); that is, by finding an appropriate root of the system of likelihood equations. Unless the contrary is explicitly stated, this loose sense will always be implied. The following general theorem concerns the existence and asymptotic distribution of a root of a system of stochastic equations with specified structure. The results for the regular \( (c_n \propto n^{1/2}) \) and boundary non-regular \( (c_n \propto n^{1/2} \ln^{1/2} n) \) cases of the generalized Beta distribution will follow directly from this theorem.

It will always be assumed that the true value \( \tilde{\theta}_o \) is an interior point of \( \Theta \).

The quantities \( g, g_n(\cdot, \cdot), A_{on}(\cdot), \) and \( R_n(\cdot, \cdot) \) are column p-vectors, while \( A_{ln}(\cdot) \) and \( J \) are \( p \times p \) matrices. \( J_n(\tilde{\theta}, \tilde{\gamma}) \) is the matrix of first partial derivatives of \( g_n(\tilde{\theta}, \tilde{\gamma}) \) wrt \( \tilde{\theta} \) at \( \tilde{\theta} \).

**Theorem 3.2:** If \( \exists A_{on}(\cdot), A_{ln}(\cdot), R_n(\cdot, \cdot), J > 0, \) and a sequence of positive constants \( (c_n) \) such that

\[
\begin{align*}
(1) & \quad g_n(\tilde{\theta}, \tilde{\gamma}) = A_{on}(\tilde{\gamma}) + A_{ln}(\tilde{\gamma})(\tilde{\theta} - \tilde{\theta}_o) + R_n(\tilde{\theta}, \tilde{\gamma}) \\
(2) & \quad c_nA_{on}(\tilde{\gamma}) = \frac{1}{2} J_n + o_p(1) \\
(3) & \quad A_{ln}(\tilde{\gamma}) = -J + o_p(1) \\
(4) & \quad \forall m > 0, R_n(\tilde{\theta}, \tilde{\gamma}) = o_p(c_n^{-1}), \text{ unif.} \quad \forall \tilde{\theta} \sim m/c_n(\tilde{\theta}_o)
\end{align*}
\]

and, in addition, when \( p = 1, \)
(v') \( \forall m > 0, \mathcal{Q}_n(\theta, \mathbf{x}) \) is continuous in \( \theta \) on \( \hat{\mathcal{S}}_{m/c_n}(\theta_0) \), i.p.

or, when \( p > 2 \).

(v'') \( \forall m > 0, \mathcal{Q}_n(\theta, \mathbf{x}) \) is differentiable wrt \( \theta \) on \( \mathcal{S}_{m/c_n}(\theta_0) \), i.p.

and

(viii') \( \forall m > 0, |\mathcal{J}_n(\theta, \mathbf{x})| \neq 0 \ \forall \theta \in \mathcal{S}_{m/c_n}(\theta_0), \) i.p.,

then, i.p., \( \exists \) a root \( \hat{\theta}_n \) of \( \mathcal{Q}_n(\theta, \mathbf{x}) = 0 \) with the property that

\[
\mathcal{C}_n(\hat{\theta}_n - \theta_0) = \mathcal{J}(\hat{\theta}_n, \mathbf{J}^{-1}) + o_p(1).
\]

Remarks.

1. A trivial modification of the proof shows this theorem to be true

if \( A_{1n} \) is also a function of \( \theta \), and (iii) is replaced by

(iii bis) \( \forall m > 0, A_{1n}(\theta, \mathbf{x}) = -J + o_p(1) \), unif. \( \forall \theta \in \hat{\mathcal{S}}_{m/c_n}(\theta_0) \).

2. The more natural, and stronger, condition to verify for (vi') will often be

(vi' bis) \( \forall m > 0, \mathcal{J}_n(\theta, \mathbf{x}) = -J + o_p(1) \), unif. \( \forall \theta \in \mathcal{S}_{m/c_n}(\theta_0) \).

3. In order to make clear the meaning of conditions such as those given above, (iv) is here written in more extensive form.

(iv) \( \forall m > 0 \) and \( \forall \varepsilon > 0, \)

\[
P(\mathcal{H}_n(\theta, \mathbf{x}) \geq \varepsilon_{n}^{-1}, \ \forall \theta \in \hat{\mathcal{S}}_{m/c_n}(\theta_0)) \rightarrow 0, \text{as } n \rightarrow +\infty.
\]
Proof: Since $\mathcal{J} > 0$, (iii) implies that $A_{\ln}^{-1}(\mathcal{X}) = -j^{\mathcal{J}^{-1}} + o_p(1)$. By the multivariate Slutsky theorem (e.g., Rao [1965], (x), 102)

\[
\|A_{\ln}^{-1}(\mathcal{X}) A_{on}(\mathcal{Y})\| = \|[j^{\mathcal{J}^{-1}} + o_p(1)] c_n^{-1}[j^{\mathcal{J}^{-1}/2} + o_p(1)]\|
\]

\[= c_n^{-1}\|j^{-1/2}\| + o_p(1) = c_n^{-1}o_p(1)\]

and, \(\forall m > 0,\)

\[
\|A_{\ln}^{-1}(\mathcal{X}) R_n(\theta, \mathcal{X})\| = \|[j^{\mathcal{J}^{-1}} + o_p(1)] o_p(c_n^{-1})\|
\]

\[= c_n^{-1}o_p(1), \text{ unif. } \forall \theta \in S_{m/c_n}(\theta_0).\]

The existence section of the proof is separated into two cases.

Single Parameter \((p = 1)\): Given any \(\epsilon > 0\), \(\exists m\) such that, by using (i) and the above results, and defining \(\delta_n = \theta_0 + m c_n^{-1}\), it can be seen that

\[
A_{\ln}^{-1}(\mathcal{X}) G_n(\theta_n^+, \mathcal{X}) = A_{\ln}^{-1}(\mathcal{X}) A_{on}(\mathcal{Y}) + m c_n^{-1} + A_{\ln}^{-1}(\mathcal{X}) R_n(\theta_n^+, \mathcal{X})
\]

is dominated wabsb by the middle term, and hence

\[
A_{\ln}^{-1}(\mathcal{X}) G_n(\theta_n^+, \mathcal{X}) \geq 0 \text{ wabsb.}
\]

Therefore, by \((v')\), wabsb \(\exists\) a root \(\hat{\theta}_n\) of \(G_n(\theta, \mathcal{X}) = 0\) such that \(\hat{\theta}_n \in S_{m/c_n}(\theta_0)\). That is, \(\exists \hat{\theta}_n = \theta_0 + o_p(c_n^{-1})\) such that \(G_n(\hat{\theta}_n, \mathcal{X}) = 0, \text{ i.p.}\).
Multiparameter ($p \geq 2$): Given any $\varepsilon > 0$, the first results in this proof, plus (v") and (vi"), show easily that $\exists m$ such that for $\eta_n = m c_n^{-1}$ the hypotheses of Lemma 3.1 are satisfied w.a.s.p for $Q_n(\theta, \chi)$ as a function of $\theta$. The lemma, therefore, asserts the existence w.a.s.p of a solution $\hat{\theta}_n$ of $G_n(\theta, \chi) = 0$ such that $\hat{\theta}_n \in S_{m/c_n}(\theta_0)$. That is, $\exists \hat{\theta}_n = \theta_0 + o_p(c_n^{-1})$ such that $G_n(\hat{\theta}_n, \chi) = 0$.

The proof is easily completed by noting that (iv) and $\hat{\theta}_n = \theta_0 + o_p(c_n^{-1})$ yield $R_n(\hat{\theta}_n, \chi_n) = o_p(c_n^{-1})$. Then manipulation of $G_n(\hat{\theta}_n, \chi) = 0$ and another application of the multivariate Slutsky theorem give

$$c_n(\hat{\theta}_n - \theta_0) = c_n^{-1}(\chi)[c_n A_n(\chi) + c_n R_n(\hat{\theta}_n, \chi)]$$

$$= [\chi^{-1} + o_p(1)] [\chi^{-1/2} \chi + o_p(1)] = \chi^{-1/2} \chi + o_p(1).$$

Corollary 3.3: If in Theorem 3.2, instead of $R_n(\theta, \chi)$ the remainder term is written as $((\frac{1}{2} (\theta - \theta_0) r_{n j}(\theta, \chi)(\theta - \theta_0))^{\phi}_{j=1})$, and 3.2(iv) is replaced by

3.3(iv) $\forall m > 0$, and $\forall j = 1, 2, \ldots, p$, $r_{n j}(\theta, \chi) = o_p(c_n)$, unif. $\forall \theta \in S_{m/c_n}(\theta_0),$

then the conclusion of Theorem 3.2 is still true.

Proof: 3.2(iv) follows trivially from 3.3(iv).
4. Regular Case (> 2)

The results of the preceding section will now be applied to the so-called regular case of maximum likelihood estimation of the parameter \( \theta \in \Theta \subset \mathbb{R}^p \) when a random sample of size \( n \) is available from a fixed distribution with density \( f(y; \theta) \).

**Theorem 4.1.** If \( g(\theta, y) \) is a column vector valued function satisfying:

1. \( g(\theta, y) = \frac{\partial \ln f(y; \theta)}{\partial \theta} \) wherever the right hand side is defined, so that \( g_n(\theta, y) = \frac{1}{n} \sum_{i=1}^{n} g(\theta, y_i) = 0 \) is an equivalent form of the system of likelihood equations for \( \theta \),

2. \( E_g(\theta, Y_i) = 0 \)

3. \( E[g(\theta, Y_i)]g(\theta, Y_i)' = -E \frac{\partial g(\theta, Y_i)}{\partial \theta} = \mathbf{J} > 0 \)

4. \( \forall m > 0, P \left\{ \frac{\partial^2 g(\theta, Y_i)}{\partial \theta_j \partial \theta_j} \right\} \) is continuous in \( \tilde{S}_{m/n}^{1/2}(\theta_0) \)

\[ \forall j, j', j'' = 1, 2, \ldots, p, \forall i = 1, 2, \ldots, n \]

\[ \rightarrow 1, \text{ as } n \rightarrow +\infty \]
(v) \( \forall m > 0, \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{1} (1-t) \frac{\partial^2 g_j(\omega, Y_i)}{\partial \omega_j \partial \omega_j} \bigg|_{\omega=\bar{\omega}_t} \) \( dt = o_p(n^{-1/2}) \),

uniformly \( \forall \bar{\omega} \in S \), \( \forall j, j', j'' = 1, 2, \ldots, p \) for \( \epsilon = 0 \) and \( 1, \)

where \( \omega_t = \theta + t(\bar{\theta} - \theta) \)

or

(v') \( \forall m > 0, \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 g_j(\theta, Y_i)}{\partial \theta_j \partial \theta_j} = o_p(n^{1/2}) \),

uniformly \( \forall \theta \in S \), \( \forall j, j', j'' = 1, 2, \ldots, p \)

then, i.p., \( \exists \) a root \( \hat{\theta}_n \) of \( q_n(\theta, Y) = 0 \) with the property that

\[ n^{1/2}(\hat{\theta}_n - \theta_0) = J^{-1/2} g_j + o_p(1). \]

**Remark:** When \( p = 1 \), (iv) can be replaced by

(iv') \( \forall m > 0, P[g_j(\theta, Y_i) \text{ is continuous in } S^{1/2}(\theta, \bar{\theta})_i, \forall i = 1, 2, \ldots, n} \)

\[ \rightarrow 1, \text{ as } n \rightarrow + \infty. \]

**Proof:** (iv) implies that \( \forall m > 0 \), the Taylor expansion of \( q_n(\theta, Y) \)
holds \( \forall \theta \in S \), i.p. This expansion may be written

\[ q_n(\theta, Y) = \frac{1}{n} \sum_{i=1}^{n} q(\theta_0, Y_i) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial q(\theta_0, Y_i)}{\partial \theta_0} (\theta - \theta_0) + R_n(\theta, Y) \]

where \( R_n(\theta, Y) = (R_{n1}(\theta, Y), \ldots, R_{np}(\theta, Y))' \), and

\[ R_{nj}(\theta, Y) = \frac{1}{n} \sum_{i=1}^{n} \left( \int_{0}^{1} (1-t) \sum_{j', j''=1}^{p} (\theta_j - \theta_{0j})(\theta_{j''} - \theta_{0j''}) \frac{\partial^2 g_j(\omega, Y_i)}{\partial \omega_j \partial \omega_j} \bigg|_{\omega=\bar{\omega}_t} \right) dt \]
or
\[
R_{n,j}(\theta, \chi) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\omega_i} \sum_{j', j''} (\theta_{j'} - \theta_{o,j'})(\theta_{j''} - \theta_{o,j''}) \frac{\partial^2 g_i(\omega, Y_i)}{\partial \omega_j \partial \omega_{j'}} \bigg|_{\omega = \omega_j}
\]
\[
= \frac{1}{2} (\theta - \theta_o)' \mathcal{R}_{n,j}(\theta, \chi) (\theta - \theta_o)
\]
where \(\omega_j\) is some point on the line segment between \(\theta\) and \(\theta_o\). Thus 3.2(i) holds.

The multivariate analogue of the Lindeberg-Lévy C.L.T. (e.g., Rao [1965], 108), together with (ii) and (iii), yields
\[
n^{-\frac{1}{2}} \mathcal{A}_{on}(\chi) = n^{-\frac{1}{2}} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(\chi, Y_i) = \mathcal{L}(\chi, \chi) + o_p(1).
\]
Thus 3.2(ii) holds.

Khintchine's W.L.L.N. and (iii) show that
\[
\mathcal{A}_{ln}(\chi) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(\theta_o, Y_i)}{\partial \theta_o} = -\lambda + o_p(1),
\]
which is 3.2(iii).

For the integral form of the remainder, it follows from (v) with \(\epsilon = 1\), that \(\forall m > 0\) and for each \(j = 1, 2, \ldots, p\),
\[
|R_{n,j}(\theta, \chi)|
\leq \sum_{j', j''} \left| \frac{\partial^2 g_i(\omega, Y_i)}{\partial \omega_j \partial \omega_{j'}} \bigg|_{\omega = \omega_j} \right| \frac{1}{n} \sum_{i=1}^{n} \int_0^1 (1-t) \left| \frac{\partial^2 g_i(\omega, Y_i)}{\partial \omega_j \partial \omega_{j'}} \bigg|_{\omega = \omega_t} \right| dt
\]
\[
\leq \sum_{j', j''} m \frac{2 \omega}{n} \left( \frac{1}{2} \right) = o_p(n^{-1/2}), \text{ unif. } \forall \theta \in \Sigma_{m/n}^{1/2} (\theta_o).
\]
Therefore, \(\forall m > 0\), \(R_n(\theta, \chi) = o_p(n^{-1/2})\), unif. \(\forall \theta \in \Sigma_{m/n}^{1/2} (\theta_o)\), and 3.2(iv) holds. 3.3(iv) holds trivially from (v').
By (iii), (iv), (v) with $\epsilon = 0$ (or $(v')$) and the W.L.L.N., the following Taylor expansion can be written $\forall m > 0$:

$$[\tilde{G}_n(\theta, X)]_{jj},$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i(\theta, X_i)}{\partial \theta_j},$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i(\theta_0, X_i)}{\partial \theta_j} + \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \sum_{j'=1}^{p} (\theta_j'' - \theta_0 j'') \frac{\partial^2 g_i(\theta, X_i)}{\partial \theta_j \partial \theta_j''} \bigg|_{\theta = 0} dt$$

$$= -[\tilde{G}]_{jj} + o_p(1) + mn^{-1/2} o_p(n^{1/2})$$

$$= -[\tilde{G}]_{jj} + o_p(1), \ \text{unif.} \ \forall \theta \in \mathbb{S}_{m/n}^{1/2}$$

Therefore, 3.2(vi" bis) and, hence, 3.2(vi") hold.

3.2(v') or 3.2(v"), whichever is appropriate, is true by virtue of (iv) or (iv').

Hence, Theorem 3.2 or Corollary 3.3 applies with $c_n = n^{1/2}$ to yield the required result. ||

Theorem 4.1 can now be applied to maximum likelihood estimation of the parameters of $B_{\gamma, \delta}(a, b)$ in the regular case ($\gamma, \delta > 2$). The well-known psi or digamma function, $\psi(z) = \frac{d}{dz} \ln \Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}$, and the trigamma function, $\psi'(z) = \frac{d}{dz} \psi(z)$, are used in the following work.

Theorem 4.2: For the generalized Beta distribution with unknown parameter $\theta = (a, b, \gamma, \delta)'$ and $\gamma, \delta > 2$, i.p. $\exists$ a root $\hat{\theta}_n$ of the system of likelihood equations such that $n^{1/2}(\hat{\theta}_n - \theta_0) = \mathcal{L}(\theta_0, \delta^{-1}) + o_p(1)$,
where the information matrix is

\[
\gamma = \begin{bmatrix}
\frac{\delta(y+\delta-1)}{(b-a)^2(y-2)} & \frac{y+\delta-1}{(b-a)^2} & \frac{\delta}{(b-a)(\delta-1)} & -\frac{\gamma}{b-a} \\
\frac{y+\delta-1}{(b-a)^2} & \frac{y(y+\delta-1)}{(b-a)^2(\delta-2)} & \frac{1}{b-a} & -\frac{y}{(b-a)(\delta-1)} \\
\frac{\delta}{(b-a)(\delta-1)} & \frac{1}{b-a} & \psi'(y)-\psi'(y+\delta) & -\psi'(y+\delta) \\
-\frac{\gamma}{b-a} & -\frac{y}{(b-a)(\delta-1)} & -\psi'(y+\delta) & \psi'(\delta)-\psi'(y+\delta)
\end{bmatrix}.
\]

Proof:

4.1(i): \[ g_1(\theta, y) = \frac{Y+\delta-1}{b-a} - \frac{y-1}{y-a} \]

\[ g_2(\theta, y) = -\frac{Y+\delta-1}{b-a} + \frac{\delta-1}{b-y} \]

\[ g_3(\theta, y) = \psi(y+\delta) - \psi(y) + \ln \left( \frac{Y-a}{b-a} \right) \]

\[ g_4(\theta, y) = \psi(y+\delta) - \psi(\delta) + \ln \left( \frac{b-y}{b-a} \right) \]

The system of likelihood equations is \[ g_n(\theta, Y) = \frac{1}{n} \sum_{i=1}^{n} g(\theta, Y_i) = 0, \]

where \[ g = (g_1, g_2, g_3, g_4)' \].

4.1(ii): It is easily verified that \[ E_g(\theta_o, Y_i) = 0. \]

4.1(iii): It is tedious work, but a straightforward task, to verify that

\[ E[g(\theta_o, Y_i)] [g(\theta_o, Y_i)']' = -E \frac{\partial g(\theta_o, Y_i)}{\partial \theta_o} = g > 0. \]
4.1(i): \( \frac{\partial^2 g_i(\hat{\theta}, Y_i)}{\partial \hat{\theta}_j \partial \hat{\theta}_l} \) is continuous on \( \tilde{S}_{m/n^{1/2}(\hat{\theta}_0)} \) \( \forall i = 1, 2, \ldots, n \)

if \((Y_i - a)^\epsilon\) and \((b - Y_i)^\epsilon\) are continuous on \( \tilde{S}_{m/n^{1/2}(\hat{\theta}_0)} \) \( \forall i = 1, 2, \ldots, n \) for \( \epsilon = 1, 2, 3 \). The result (2.1) and its analogue

insure that \( Y_{1n} - a \_m \_m^{1/2} > 0 \) and \( b_0 - mn^{1/2} - Y_{mn} > 0 \) i.p.

since \( \gamma, \delta > 2 \). Therefore, i.p., \( Y_i - a \neq 0 \) and \( b - Y_i \neq 0 \)

\( \forall i = 1, 2, \ldots, n \), \( \forall \hat{\theta} \in \tilde{S}_{m/n^{1/2}(\hat{\theta}_0)} \), and the desired result follows.

4.1(i'): This condition will be satisfied if \( \forall m > 0 \), and for \( \epsilon = 1, 2, 3 \), it is true that \( n^{-1} \sum_{i=1}^{n} (Y_i - a)^{-\epsilon} = o_p(n^{1/2}) \) and

\( n^{-1} \sum_{i=1}^{n} (b - Y_i)^{-\epsilon} = o_p(n^{1/2}) \) for \( \forall \hat{\theta} \in \tilde{S}_{m/n^{1/2}(\hat{\theta}_0)} \).

By Lemma 2.1,

\( \forall m > 0, n^{-1} \sum_{i=1}^{n} (Y_i - a)^{-\epsilon} = o_p(1) \) for \( \forall \hat{\theta} \in \tilde{S}_{m/n^{1/2}(\hat{\theta}_0)} \) when \( \epsilon = 1, 2 \) since \( \gamma > 2 \geq \epsilon \), and the same result holds for \( \epsilon = 3 \) when \( \gamma > 3 \).

When \( 2 < \gamma \leq 3 \), take \( \gamma' = 2 \), then \( \forall m > 0 \) Lemma 2.1 gives

\( n^{-1} \sum_{i=1}^{n} (Y_i - a)^{-\epsilon} = o_p(n^{1/2}) \) for \( \forall \hat{\theta} \in \tilde{S}_{m/n^{1/2}(\hat{\theta}_0)} \). The analogous results hold for \( n^{-1} \sum_{i=1}^{n} (b - Y_i)^{-\epsilon} \).

Thus Theorem 4.1, its conditions having been verified, asserts

the required result with regard to a root \( \hat{\theta}_n \).

Corollary 4.3: (a) When only subsets of the parameters \( a, b, \gamma, \delta \) are unknown (and still \( \gamma, \delta > 2 \)), the properties of the maximum likelihood estimator (mle) of the reduced parameter \( \hat{\theta} \) are the same as in the unreduced case, but now the information matrix \( J \) is found as
usual by deleting from the original full matrix the rows and columns corresponding to the known parameters. Provided that the corresponding end point is known, the result also holds for the full range \( \gamma \) and/or \( \delta \) greater than zero.

(b) When \( a, b, \gamma = \delta > 2 \) are unknown, all of the preceding results hold for \( \theta = (a, b, \gamma)' \), but the information matrix is now

\[
J_e = \begin{bmatrix}
\frac{\gamma(2\gamma-1)}{(b-a)^2(\gamma-2)} & \frac{2\gamma-1}{(b-a)^2} & \frac{1}{(b-a)(\gamma-1)} \\
\frac{2\gamma-1}{(b-a)^2} & \frac{\gamma(2\gamma-1)}{(b-a)^2(\gamma-2)} & \frac{-\gamma}{(b-a)(\gamma-1)} \\
\frac{1}{(b-a)(\gamma-1)} & \frac{-\gamma}{(b-a)(\gamma-1)} & 2[\psi'(\gamma)-2\psi'(2\gamma)]
\end{bmatrix}
\]

Proof: All parts of the proof follow directly from, or are essentially the same as, the corresponding parts in the proof of Theorem 4.2. ||

Corollary 4.4: In Theorem 4.2 and Corollary 4.3 each mle \( \hat{\theta}_n \) achieves (asymptotically) its respective Cramér-Rao "lower bound" for the covariance matrix of unbiased estimators for \( \theta_0 \).

Proof: The conditions of Rao [1965, p. 265] and/or of Kendall and Stuart [1967, Ex. 17.21, p. 33] are easily verified for each particular parameter set. ||
Theorem 3.2 and Corollary 4.4 assert the existence of a consistent, asymptotically normal, efficient (CANE) sequence of roots \( \langle \hat{\theta}_n \rangle \) of \( \varphi_n(\hat{\theta}, \chi) = 0 \). That is, for each \( n \) there exists an appropriate selection from among the roots of \( \varphi_n(\hat{\theta}, \chi) = 0 \) (if no roots exist, any arbitrary value may be chosen) so that the sequence \( \langle \hat{\theta}_n \rangle \) is CANE for \( \theta_0 \). In any particular problem it may or may not be possible to ascertain a correct selection procedure without resorting to an actual evaluation of the likelihood function at each contending root.

A problem which will enjoy special attention from time to time is that of estimating the end point \( a \) in \( B_{r, \delta}(a, b) \) when the other three parameters are known. In this situation the existence, for each \( n \), of a specific root \( \hat{\theta}_n \) (which is CANE for \( a_0 \) when \( r > 2 \)) can be asserted. (The properties of \( \hat{\theta}_n \) when \( 1 < r \leq 2 \) will be discussed in Sections 5 and 6.) First, the graph of

\[
G_n(a, \chi) = \frac{1}{n} \sum_{i=1}^{\infty} \left[ \frac{r+\delta-1}{b-a} - \frac{Y_i-1}{Y_i-a} \right] = \frac{r+\delta-1}{b-a} - \frac{1}{n} \sum_{i=1}^{\infty} \frac{Y_i-1}{Y_1-a}
\]

will be sketched. See Figure 4.1.

For every \( n \geq 1 \) there is one real root between each pair of adjacent observations, and one root below the smallest observation. It is this smallest root which is the correct choice for \( \hat{\theta}_n \) because, for \( r > 2 \), \( Y_{\infty} > a^+ \) i.p. and the smallest root is, therefore, the only root which can possibly fall in \([a^-_n, a^+_n]\) i.p. Hence, it must be the root whose existence the theorems assert. In practice it is
actually better to find the smallest root of \((b-a) G_n(a, \bar{X}) = 0\), since the expression on the left-hand side is monotone increasing for \(a < Y_{1n}\).

\[ G_n(a, \bar{X}) \]

\[ \hat{a}_n = R_{1n} R_{2n} R_{3n} R_{4n} \cdots R_{nn} \]

\[ 0 \quad 1n \quad 2n \quad 3n \quad \cdots \quad nn \quad b \quad a-axis \]

**Figure 4.1.** Sketch of \(G_n(a, \bar{X})\) for \(\gamma > 1\) showing representative locations of the \(n\) real roots of \(G_n(a, \bar{X}) = 0\).

As a final comment in this section, it should be noted that it is easy to show that for \(\gamma > 2, (\hat{a}_n - a_o)/(Y_{1n} - a_o) = O_p(n^{1/\gamma - 1/2})\), and thus \(\hat{a}_n\) is a much better estimator for \(a_o\) than in \(Y_{1n}\).
5. **Boundary Non-regular Case** \((= 2)\)

The case in which at least one end point is unknown, and its governing shape parameter has the value \(2\), will be treated in this section. This case is not regular in the sense of having CANE estimators with \(n^{1/2}\) normalization, nor is it non-regular in the sense of yielding no CAN estimators under any normalization as is the case when the shape parameter is between 1 and 2. There are, however, CAN (even CANE) estimators with \((n \ln n)^{1/2}\) normalization when the shape parameter has the boundary value 2. Whence the name "boundary non-regular" case.

LeCam [1960] states on page 178 that the mle for the shift parameter of the triangular density \([1 - |x|]^+\) is asymptotically normal (and asymptotically sufficient) when the appropriate normalizing sequence \((n \ln n)^{1/2}\) is used. This statement suggested that the boundary non-regular case of estimating \(\alpha\) in \(B_{\gamma,\delta}(a, b)\) when \(b, \gamma = 2, \delta\) are known might yield a similar result since the density vanishes with finite, non-zero slope at the point of contact. Before the result is stated in the form of a theorem, the following useful lemma is presented.

**Lemma 5.1:** If \(X_1, X_2, \ldots\) are i.i.d. \(B_{2,\delta}(0,1)\), then

\[
(a) \quad \left[ \frac{\delta(\delta+1)}{2} \frac{n \ln n}{n} \right]^{-1/2} \sum_{i=1}^{n} \left( \frac{1}{X_i} - \delta - 1 \right) = N(0,1) + o_p(1)
\]

and

\[
(b) \quad \left[ \frac{\delta(\delta+1)}{2} \frac{n \ln n}{n} \right]^{-1} \sum_{i=1}^{n} \frac{1}{X_i^2} = 1 + o_p(1).
\]
Furthermore, if $\delta = 2$, then

\[
\frac{1}{(3n \ln n)^{1/2}} \sum_{i=1}^{n} \left[ 3 - \frac{1}{X_i} \right] = \mathcal{N}(Q, I) + o_p(1).
\]

Proof: (a): Let $W_i = X_i^{-1}$. Then $W_i \in (1, \infty)$ w.p. 1. It follows, using the notation of Feller [1966, IX.8], that, for $w \geq 1$,

\[
U_w(w) = \int_{-w}^{w} y^2 dF_w(y) = \int_{-w}^{w} \frac{1}{1/w} \delta(\delta+1) x(1-x)^{\delta-1} dx
\]

\[
= \delta(\delta+1) \ln w + O(1), \text{ as } w \to +\infty.
\]

Therefore, $U_w(w)$ varies slowly at $+\infty$, and, hence, the distribution of $X_i^{-1}$ belongs to the domain of attraction of the normal distribution (Feller [1966], Theorem IX.8.1a(i)). The correct normalizing sequence satisfies $n^{-1/2} U_w(c_n) \to 1$ as $n \to +\infty$, and LeCam's comment suggests that $c_n = (cn \ln n)^{1/2}$, from which it is easily found that

\[
c_n = \left[ \frac{\delta(\delta+1)}{2} n \ln n \right]^{1/2}.
\]

Thus, with centering at $E[X_i^{-1}] = \delta+1$, Feller [1966, Theorem IX.2(ii)] gives

\[
\frac{1}{c_n} \sum_{i=1}^{n} \left( \frac{1}{X_i} - \delta - 1 \right) = N(0,1) + o_p(1).
\]

(b): Let $V_{in} = \frac{1}{n \ln n} \frac{1}{X_i^2}$. The $V_{in}$, $i = 1, 2, \ldots, n$, are clearly uniformly asymptotically negligible. Since, $\forall s > 0$ (and letting $s_n = (sn \ln n)^{-1/2}$)
\[ \int_0^s x(1-x)^{q-1} \, dx = \left[ \frac{x^2}{2} + o(x^2) \right]_0^s, \text{ as } x \to 0, \]
\[ = o\left(\frac{1}{n \ln n}\right), \]

and
\[ \int_{s_n}^1 \frac{1}{x} (1-x)^{q-1} \, dx = \left[ \ln x + o(\ln x) \right]_{s_n}^1, \text{ as } x \to 0, \]
\[ = \frac{1}{2} \ln n + o(\ln \ln n), \]

and
\[ \int_{s_n}^1 \frac{1}{x^2} (1-x)^{q-1} \, dx = \left[ -\frac{1}{2x^2} + o(x^{-2}) \right]_{s_n}^1, \text{ as } x \to 0, \]
\[ = o(n \ln n), \]

it is immediate that, \( \forall s > 0, \)
\[ \sum_{i=1}^n P(|V_{in}^i| > s) = o\left(\frac{1}{\ln n}\right) = o(1), \]

and
\[ n \text{ var}(V'_{in}) \]
\[ = nE(V'_{in})^2 - nE^2(V'_{in}) \]
\[ = \left[ n0\left(\frac{1}{n \ln n}\right) + \frac{n}{2} \frac{n}{2} 0(n \ln n) \right] - n\left[ 0\left(\frac{1}{n \ln n}\right) + 0\left(\frac{1}{n}\right) \right]^2 \]
\[ = o\left(\frac{1}{\ln n}\right) - n0\left(\frac{1}{2n}\right) = o\left(\frac{1}{\ln n}\right), \]

where \( V'_{in} \) is \( V_{in} \) truncated at \( \pm s \). Therefore, by a W.L.L.N. for
triangular arrays (Feller [1966], Theorem IX.9.1) \exists constants $d_n$ such that 
\[ \frac{1}{n \ln n} \sum_{i=1}^{n} \frac{1}{X_i^2} - d_n = o_p(1) \]. The constants can be taken as 
\[ d_n = nE[V_{1n}'] = nO(\frac{1}{\ln n}) + \frac{n\delta(\delta+1)}{n \ln n} \frac{1}{2} \ln n + \frac{n}{\ln n} O(\ln \ln n) \]
\[ = \frac{\delta(\delta+1)}{2} + o(1), \]
which in fact shows that convergence holds if all $d_n$ are taken as 
\[ d_n = \frac{\delta(\delta+1)}{2}. \]
(c): Let $W_1 = -\frac{u_1}{x_1} + \frac{u_2}{1-x_1}$. Consider the case $u_1, u_2 \geq 0$, with strict inequality holding for at least one $u_1$. Then, for $w \to +\infty$, with 
\[ r_w = \frac{u_1}{w} + o(\frac{1}{w}) \] and 
\[ s_w = 1 - \frac{u_2}{w} + o(\frac{1}{w}), \]
\[ U_w(w) = \int_{-w}^{w} x^2 \text{d}F(w) = \int_{r_w}^{s_w} \left( -\frac{u_1}{x} + \frac{u_2}{1-x} \right) 6x(1-x) \text{d}x \]
\[ = 6[u_1^2 \ln x - u_2^2 \ln(1-x) - (u_1 + u_2)^2 x]_{r_w}^{s_w} \]
\[ = 6[u_1^2 \ln(w-u_2) + u_2^2 \ln(w-u_1) + o(1)] \]
which varies slowly at $+\infty$. Therefore, the distribution of $W_1$ is attracted to the normal law, as in the above proof for part (a). This time \[ \frac{n}{c_n} U_w(c_n) \to l \] as $n \to +\infty$ for $c_n = \lfloor 3(u_1^2 + u_2^2) n \ln n \rfloor^{1/2}$, and, since $E[W_1] = -3u_1 + 3u_2$, it is easily found that
\[
\frac{1}{(3n \ln n)^{1/2}} \sum_{i=1}^{n} \left[ u_1(3 - \frac{1}{X_i}) + u_2(\frac{1}{1-X_i} - 3) \right]
= \frac{1}{(3n \ln n)^{1/2}} \sum_{i=1}^{n} \left[ -\frac{u_1}{X_i} + \frac{u_2}{1-X_i} - 3(u_2-u_1) \right]
= (u_1^2 + u_2^2)Z + o_p(1) = N(0, u_1^2 + u_2^2) + o_p(1).
\]

For the cases \( u_1 < 0 \) and/or \( u_2 < 0 \) the limits on the integral form of \( U_w(w) \) merely become \( |u_1|/w + o(1/w) \) and \( 1 - |u_2|/w + o(1/w) \) even though \(-u_1/x + u_2/(1-x)\) may be bounded away from zero. The resulting asymptotic distribution is the same as above. The required result is a direct consequence of the fact that \( \gamma \sim \mathcal{N}(\mu, \Sigma) \) iff \( \forall \, \mu, \, u'y \sim N(\mu'y, \, u'y) \) \( \) (Rao [1965], 436 ff.). ||

Now the theorem.

**Theorem 5.2:** If \( \gamma = 8 = 2 \) are known for \( B_y, (a,b) \), then, i.p. \( \exists \) a root \( (\hat{a}_n, \hat{b}_n)' \) of the likelihood equations with the property that
\[
(n \ln n)^{1/2} \left( \begin{array}{c}
\hat{a}_n - a_o \\
\hat{b}_n - b_o
\end{array} \right) = \mathcal{N} \left( 0, \frac{(b_o - a_o)^2}{3} I_2 \right) + o_p(1).
\]

**Proof:** The likelihood equations are
\[
0 = \sum_{i=1}^{n} \left[ \frac{3}{b-a} - \frac{1}{y_i-a} \right]
\]
and

\[ 0 = \sum_{i=1}^{n} \left( -\frac{3}{b_{o}-a_{o}} + \frac{1}{Y_{i}-a_{o}} \right), \]

which, by addition and subtraction of the same terms using \( a_{o} \) and \( b_{o} \), may also be written as

\[ 0 = \sum_{i=1}^{n} \left( \frac{3}{b_{o}-a_{o}} - \frac{1}{Y_{i}-a_{o}} \right) + (a-a_{o}) \sum_{i=1}^{n} \left[ \frac{3}{(b-a)(b_{o}-a_{o})} - \frac{1}{(Y_{i}-a)(Y_{i}-a_{o})} \right] \]

\[-\frac{3(b-b_{o})}{n(b-a)(b_{o}-a_{o})} \]

and

\[ 0 = \sum_{i=1}^{n} \left( \frac{3}{b_{o}-a_{o}} + \frac{1}{b_{o}-Y_{i}} \right) + (b-b_{o}) \sum_{i=1}^{n} \left[ \frac{3}{(b-a)(b_{o}-a_{o})} - \frac{1}{(b-Y_{i})(b_{o}-Y_{i})} \right] \]

\[-\frac{3(a-a_{o})}{n(b-a)(b_{o}-a_{o})} \]

After division by \( c_{n}^{2} = 3n \ln n \) these equations may be written in matrix form as

\[ Q = c_{n}(a, b; Y) \]

\[ = \sum_{i=1}^{n} \left( \frac{3}{b_{o}-a_{o}} - \frac{1}{Y_{i}-a_{o}} \right) + \sum_{i=1}^{n} \left[ \frac{1}{(Y_{i}-a_{o})^{2}} - \frac{0}{(b_{o}-Y_{i})^{2}} \right] \]

\[ + \frac{1}{c_{n}^{2}} \sum_{i=1}^{n} \left( \frac{(a-a_{o})}{(Y_{i}-a)(Y_{i}-a_{o})^{2}} + \frac{(a-a_{o})}{(b_{o}-a_{o})^{2}} - \frac{(a-a_{o})}{(b-a)(b_{o}-a)} \right) \]

\[ + \frac{1}{c_{n}^{2}} \sum_{i=1}^{n} \left( \frac{(b-b_{o})}{(b_{o}-Y_{i})^{2}} - \frac{(a-a_{o})}{(b-a)(b_{o}-a)} \right) \]
The conditions of Theorem 3.2 can be verified for this representation of \( G_n \).

If \( X_1, X_2, \ldots \) are i.i.d. \( B_{2,2}(0,1) \), the distribution of the first term of \( G_n \) is the same as the distribution of

\[
\frac{1}{b - a} \sum_{i=1}^{n} \left( \frac{3 - 1}{X_i} \right) = \frac{1}{b - a} \frac{1}{c_n} \sum_{i=1}^{n} \frac{1}{1 - X_i} \quad + \quad o_p(1),
\]

where the final step is the result of Lemma 5.1(c). Therefore, 3.2(ii) is satisfied with \( J = 1/(b - a)^2 \).

In the second term the upper left hand element of the matrix has the same distribution as

\[
\frac{-1}{(b - a)^2} \frac{1}{2} \sum_{i=1}^{n} \frac{1}{X_i} = \frac{-1}{(b - a)^2} + o_p(1), \quad \text{by Lemma 5.1(b)},
\]

and the lower right element is \(-1/(b - a)^2 + o_p(1)\) by the analogous result for \( \sum_{i=1}^{n} \frac{1}{1 - X_i^2} \). Hence, 3.2(iii) is satisfied.

For any \( m > 0 \), and \( \forall a, b \) such that \(|a - a_o| \leq mc^{-1}_n\), \(|b - b_o| \leq mc^{-1}_n\),

\[
\left| \frac{1}{c_n} \sum_{i=1}^{n} \frac{(a - a_o) - (b - b_o)}{(b - a)(b - a_o)} \right| \leq \frac{2mc^{-1}_n}{c_n} \frac{1}{(b - a_o - 2mc^{-1}_n)(b - a_o)} \leq \frac{1}{c_n} o(\frac{n}{n \ln n}) = c^{-1}_n o(1),
\]

and
\[
\frac{1}{c_n} \sum_{i=1}^{n} \frac{(a-a_o)^2}{(Y_i-a)(Y_i-a_o)} \leq \frac{\min_{i=1,...,n} |Y_i-a|}{\min_{i=1,...,n} |Y_i-a_o|} \frac{1}{c_n} \sum_{i=1}^{n} \frac{1}{(Y_i-a_o)^2}
\]

\[
= \frac{2c_n^{-2}}{Y_n-a} \frac{Y_{n-1}}{\min_{i=1,...,n} |Y_i-a|} O_p(1)
\]

\[
= c_n^{-1} O_p(n^{1/2})[1 + o_p(1)] O_p(1)
\]

\[
= c_n^{-1} O_p(1)
\]

because

\[
\frac{1}{c_n} \sum_{i=1}^{n} \frac{1}{(Y_i-a_o)^2} = O_p(1)
\]

by Lemma 5.1(b), \(1/(Y_{n-1}-a) = O_p(n^{1/2})\) by Lemma 2.1(b), and

\[P(Y_{n-1} > a) > P(Y_{n-1} > a_o + mc_n^{-1}) \rightarrow 1\] as \(n \rightarrow +\infty\), by (2.1).

The symmetric nature of \(B_{2,2}(a,b)\) shows that the same result holds for the term involving \(b\) and \(b_o\), and, therefore, 3.2(iv) is satisfied.

\(G_n\) is certainly differentiable wrt \(a\) and \(b\) whenever \(a < Y_{n-1}\) and \(b > Y_{n-1}\), and these events occur i.p. By minor modification of the arguments used above, it is readily found that

\[
J_{G_n}(a,b;Y) = \frac{1}{c_n} \sum_{i=1}^{n} \left( \begin{array}{ccc}
\frac{3}{(b-a)^2} - \frac{1}{(Y_i-a)^2} \\
\frac{-3}{(b-a)^2} & \frac{3}{(b-a)^2} & -\frac{1}{(b-Y_i)^2}
\end{array} \right)
\]

\[
= -J + o_p(1), \text{ unif. } \forall a,b \text{ such that } |a-a_o| \leq mc_n^{-1}, \quad |b-b_o| \leq mc_n^{-1}.
\]
This completes the verification of the conditions of Theorem 3.2, and guarantees the desired result. \[ \|

The diagonal nature of the covariance matrix of the asymptotic normal distribution shows that \( \hat{a}_n \) and \( \hat{b}_n \) are asymptotically independent. This is not the case when \( \gamma, \delta > 2 \).

When only one end point is unknown, and its governing shape parameter is 2, the results are essentially the same as those above, regardless of the value of the other shape parameter. The following theorem states the results.

**Theorem 5.3:** If \( b, \gamma = 2, \delta \) are known in \( R_{\gamma, \delta}(a, b) \), then for every \( n \), \( \exists \) a unique root, \( \hat{r}_n \), of the likelihood equation, which is less than \( Y_{1n} \); and this root has the property that

\[
(n \ln n)^{1/2} (\hat{a} - a_0) = N(0, \frac{2}{\delta(\delta+1)} (b-a_0)^2) + o_p(1)
\]

**Proof:** Once again Theorem 3.2, in conjunction with Lemma 5.1, asserts the existence i.p. of some root with the specified asymptotic properties. The discussion at the end of Section 4 demonstrates the unicity and location of the root. \[ \|

**Corollary 5.4:** The estimator, \( \hat{a}_n \), in Theorem 5.3 asymptotically attains the Kiefer lower bound for unbiased estimators of \( a_0 \).
Proof: One of Polfeldt's theorems (Polfeldt [1970], Theorem 3.4) gives a computational form of Kiefer's variance bound for unbiased estimators of the location parameter of a distribution which is one-sided to the right, whose density varies regularly with exponent 1 at the end point, and which satisfies several other technical assumptions. \( B_{2,0}(a,b) \) satisfies the correct assumptions when only \( a \) is unknown, except that the support of the density is finite rather than semi-infinite. Polfeldt's proof can, however, be modified to cover this case, and the Kiefer variance bound can again be written as \( A_n^2 (1 + o(1)) \) where \( A_n = G^{-1}(n^{-1}) \), \( G(v) = v^2 T(v) \), and \( T(v) = \int_{v}^{b-a} y^{-2} f_2, o(y; 0, b-a) \, dy \).

But,

\[
T(v) = \frac{\delta (\delta + 1)}{b-a} \int_{v}^{b-a} \frac{1}{y} (b-a-y)^{\delta-1} \, dy = \frac{\delta (\delta + 1)}{(b-a)^2} \int_{v}^{b-a} \frac{1}{y} \left(1 - \frac{y}{b-a}\right)^{\delta-1} \, dy
\]

\[
= \frac{\delta (\delta + 1)}{(b-a)^2} \left[ \ln y + o(y) \right]_{v}^{b-a} = \frac{\delta (\delta + 1)}{(b-a)^2} \ln \frac{1}{v} + o(1),
\]

and, therefore, if \( A_n = (b-a) \left[ \frac{\delta (\delta + 1)}{2} \right] ^{\frac{1}{2}} n \ln n \right]^{-1/2} \), then \( G(A_n) = A_n^2 T(A_n) = \frac{2}{n \ln n} \frac{1}{2} (\ln n + \ln \ln n) = \frac{1}{n} + o\left(\frac{1}{n}\right) \). Therefore, the Kiefer variance bound is \( (b-a)^2 \frac{2}{\delta (\delta + 1)} \frac{1}{n \ln n} (1 + o(1)) \) and \( \hat{A}_n \) attains it asymptotically as claimed. ||

When estimating both \( a \) and \( b \) in \( B_{2,0}(a,b) \), with the shape parameters known, the result of Corollary 5.4, coupled with the asymptotic independence of \( \hat{A}_n \) and \( \hat{B}_n \), suggests that \( (\hat{A}_n, \hat{B}_n)' \) asymptotically achieves the joint "Kiefer lower bound" for the covariance matrix of unbiased estimators of \( (a,b)' \).
Finally, it is easy to show that for $\gamma = 2, \ldots (\hat{a}_n - a_0)/(Y_{ln} - a_0) = o_p(\ln^{-1/2} n)$, and thus $\hat{a}_n$ is a much better estimator for $a_0$ than is $Y_{ln}$. 
6. **Non-regular Case (< 2)**

In this section we shall examine the case in which just one end point is unknown, and its governing shape parameter is less than two.

The case in which the governing shape parameter is known and is less than or equal to one is rapidly dispensed with since direct examination of the likelihood function shows that the mle (in the strong sense) of a single unknown end point is the corresponding sample extreme. Thus \( \hat{\alpha}_n = Y_{1n} \) and \( \hat{\beta}_n = Y_{nn} \), and together these are the sensible mle choices when both end points are unknown and both shape parameters are known and less than or equal to one. Note that they are not the only possible choices when, for example, \( \gamma < 1 \) and \( \delta < 1 \) are known, since in this case the likelihood function achieves its supremum value of \(+\infty\) at any point \((a, b)\) in \( (a, b) : a = Y_{1n}, Y_{nn} \leq b < +\infty \) or \(-\infty < a \leq Y_{1n}, b = Y_{nn} \). The properties of \( \hat{\alpha}_n = Y_{1n} \) and \( \hat{\beta}_n = Y_{nn} \), both singly and jointly, are immediate consequences of the well-known distribution theory for order statistics.

In order to facilitate investigation of the appropriate root when \( 1 < \gamma < 2 \) we shall rewrite the likelihood equation for \( a \) as

\[
\frac{a-b}{\gamma-1} n G(a) = \sum_{i=1}^{n} \left[ \frac{b-a}{Y_i-a} - \zeta \right] = 0
\]

where \( \zeta = (\gamma+\delta-1)/(\gamma-1) \). Because of the affine invariance of the estimation problem, it is sufficient to study the roots of the likelihood equation
\[ \Sigma_n(\alpha) = \sum_{i=1}^{n} \left[ \frac{1-\alpha}{X_i-\alpha} - \xi \right] = 0 \]

where \( X_1, X_2, \ldots \) are i.i.d. \( \mathcal{E}(0,1) \), and \( a = (b_o - a_o)\alpha + a_o \).

The following lemma merely establishes the shape of \( \Sigma_n(\alpha) \).

See Figure 6.1.

**Lemma 6.1:** For \( \gamma > 1 \),

(i) \( \Sigma_n(\alpha) \) is monotone strictly increasing for \( \alpha \in (-\infty, X_{1n}) \),

\( (X_{1n}, X_{2n}), \ldots, (X_{n-1,n}, X_{nn}), (X_{nn}, +\infty) \)

(ii) \( \Sigma_n(\alpha) \) has monotone strictly increasing second derivative for \( \alpha \in \) above intervals

(iii) \( \Sigma_n(\alpha) \) is strictly convex for \( \alpha \in (-\infty, X_{1n}) \), strictly concave for \( \alpha \in (X_{nn}, +\infty) \), and has exactly one inflection point (passing from concave to convex) in each of the other intervals

(iv) \[ \lim_{\alpha \to -\infty} \Sigma_n(\alpha) = \lim_{\alpha \to +\infty} \Sigma_n(\alpha) = n(1-\xi) = -n \frac{\delta}{\gamma-1} < 0, \]

\[ \lim_{\alpha \uparrow X_{1n}} \Sigma_n(\alpha) = +\infty, \text{ and } \lim_{\alpha \downarrow X_{1n}} \Sigma_n(\alpha) = -\infty. \]

**Proof:** The proof is straightforward. ||
Figure 6.1. Sketch of $\Sigma_n(\alpha)$ for $\gamma > 1$.

We shall endeavor to exploit the convex, strictly increasing nature of $\Sigma_n(\alpha)$ in $(-\infty, X_{1n})$ in order to gain knowledge of the (finite sample size and) asymptotic properties of the rule $\hat{\alpha}_n$.

We shall write $S_n(\alpha)$ for $\sum_{i=1}^{n} \frac{1}{X_i - \alpha}$ so that

$$\Sigma_n(\alpha) = (1-\alpha) S_n(\alpha) - n\xi.$$

**Theorem 6.2:** If $b, L < \gamma < 2, b$ are known in $B_{\gamma, \delta}(a, b)$, then there exists a unique root $\hat{\alpha}_n < Y_{1n}$ of the likelihood equation, and $\hat{\alpha}_n$ is consistent for $a_o$.  

36
Proof: The existence of a unique root $\hat{\alpha}_n < X_{1n}$ of $E_n(\alpha) = 0$ is immediate from Lemma 6.1. It remains only to show that $\hat{\alpha}_n$ is consistent for 0.

For $\alpha < 0$, $\xi(\alpha) = E[(X_1 - \alpha)^{-1}]$ exists. Therefore, $\frac{1}{n} S_n(\alpha) = \xi(\alpha) + o_p(1)$ by Khintchine's W.I.L.N. That is,

$$\frac{F_{S_n}(\alpha)(x)}{n} \rightarrow \begin{cases} 0, & \forall x < \xi(\alpha) \\ 1, & \forall x > \xi(\alpha) \end{cases}, \text{ as } n \rightarrow +\infty.$$

For $\alpha < 0$, $(1-\alpha) \xi(\alpha) = E[(1-\alpha)/(X_1 - \alpha)]$ is increasing in $\alpha$, and, for $\gamma > 1$, is continuous on the left at $\alpha = 0$ because

$$\left| \frac{1-\alpha}{X_1 - \alpha} \right| < \frac{2}{X_1} \quad \text{for} \quad |\alpha| < 1,$$

and because $\xi = E[X_1^{-1}]$ exists. It follows that, $\forall \alpha < 0$, $\xi = \xi(0^-) > (1-\alpha) \xi(\alpha)$, and hence

$$\frac{F_{S_n}(\alpha)(\frac{\xi}{1-\alpha})}{n} \rightarrow 1 \text{ as } n \rightarrow +\infty$$

since $\xi/(1-\alpha) > \xi(\alpha)$. By the shape of $E_n(\alpha)$, and because $E_n(\alpha)$ is a continuous random variable for $\alpha < 0$, we find that, $\forall \alpha < 0$,

$$F_{\hat{\alpha}_n}(\alpha) = P(\hat{\alpha}_n \leq \alpha) = P(S_n(\alpha) > 0) = P\left(\frac{S_n(\alpha)}{n} > \frac{\xi}{1-\alpha}\right)$$

$$= 1 - \frac{F_{S_n}(\alpha)(\frac{\xi}{1-\alpha})}{n} \rightarrow 0, \quad \text{as } n \rightarrow +\infty.$$

37
Since we also have $\hat{\alpha}_n < X_{ln}$ and $X_{ln}$ is consistent for 0, we can conclude that $\hat{\alpha}_n \rightarrow P > 0$. ||

For large samples, and in large part for small samples also, we must clearly rely upon asymptotic results, and it is this aspect of the problem which we shall investigate next. Unfortunately, the only asymptotic results obtained so far are for $F_{\alpha_n}$, and it is these results which we now present.

For any $x < 0$, $F_{\alpha_n}(x) = F_{\alpha_n}(x/c_n) = P(\sum_{n}(x/c_n) > 0) = P((1/d_n)\sum_{n}(x/c_n) > 0)$, and we shall see that, for the correct choice of $c_n$ and $d_n$, $(1/d_n)\sum_{n}(x/c_n)$ has a proper asymptotic distribution whose characteristic function will be found. In theory at least,

$\lim_{n \to \infty} F_{\alpha_n}(x)$ can then be evaluated.

Theorem 6.3: If $1 < \gamma < 2$, $\delta > 0$, $c_n = cn^{1/\gamma}$, and $d_n = \rho c_n = \rho cn^{1/\gamma}$, where $c$, $\rho$ are any positive constants, then, for any $x < 0$, the limit law of $(1/d_n)\sum_{n}(x/c_n)$ exists and in the Kolmogorov representation of the characteristic function of this infinitely divisible (i.d) law the mean is

$$-\frac{(-x)^{\gamma-1}}{\rho c^\gamma} \frac{B(\gamma-1, 2-\gamma)}{B(\gamma, \delta)},$$

and the canonical measure is
\[ K(u) = \begin{cases} 
0 & \text{for } u \leq 0 \\
\frac{(-x)^{\gamma-2}}{\rho^2} \frac{B(-\rho u; 2-\gamma, \gamma)}{B(\gamma, \delta)} & \text{for } 0 < u < \frac{1}{-x\rho} \\
\frac{(-x)^{\gamma-2}}{\rho^2} \frac{B(\gamma, 2-\gamma)}{B(\gamma, \delta)} & \text{for } \frac{1}{-x\rho} \leq u,
\end{cases} \]

where

\[ B(x; 2-\gamma, \gamma) = \int_0^x s^{1-\gamma}(1-s)^{\gamma-1} ds = \int \frac{1}{1-x} s^{\gamma-1}(1-s)^{\gamma-1} ds = \int \frac{z^{\gamma-1}}{1 - (z+1)^2} dz. \]

**Proof:** First,

\[ E\left( \frac{1}{d_n} \sum \frac{X_i}{C_n} \right) \]

\[ = \frac{n}{d_n} E\left( \frac{1-x/C_n}{X-x/C_n} - \zeta \right) = \frac{n}{d_n} E\left( \frac{1}{X-x/C_n} - \zeta \right) - \frac{x}{C_n} \frac{1}{d_n} E\left( \frac{1}{X-x/C_n} \right) \]

\[ = \frac{n}{d_n} \frac{1}{B(\gamma, \delta)} \int_0^1 \left[ \frac{1}{y-x/C_n} - \frac{1}{y} \right] y^{\gamma-1}(1-y)^{\delta-1} dy + O(n^{1-2/\gamma}), \]

since \( 0 < E((X-x/C_n)^{-1}) < E[X^{-1}] < +\infty, \)

\[ = -\frac{n}{d_n} \frac{-x}{C_n} \frac{1}{B(\gamma, \delta)} \int_0^1 \frac{y^{\gamma-2}(1-y)^{\delta-1}}{y-x/C_n} dy + O(n^{1-2/\gamma}) \]

\[ = -\frac{n}{d_n} \frac{-x}{C_n} \frac{1}{B(\gamma, \delta)} \int_0^1 \frac{z^{\gamma-2}(1-(-x/C_n)z)^{\delta-1}}{z+1} dz + O(n^{1-2/\gamma}) \]

by setting \( z = \frac{C_n}{-x} y \),
\[
\begin{align*}
= & \ - \frac{n}{d_n} \left( \frac{c_n}{x} \right)^{\gamma - 1} \frac{1}{B(\gamma, \delta)} \int_0^{\infty} \frac{z^{\gamma - 2}}{z+1} \ dz + O\left(n^{1-2/\gamma}\right) \\
& \text{by Lemma 6.4 with } \delta = 1, \ s = \gamma - 2, \ t = 1, \ \\
= & \ - \frac{(-x)^{\gamma - 1}}{\rho c^\gamma} \frac{B(\gamma - 1, 2 - \gamma)}{B(\gamma, \delta)} + O\left(n^{1-2/\gamma}\right) \quad \text{by setting } u = \frac{z}{z+1}, \\
= & \ \beta + O\left(n^{1-2/\gamma}\right), \quad \text{for convenience of future reference.}
\end{align*}
\]

Also,

\[
\begin{align*}
\text{var} \left[ \frac{1}{d_n} \sum_n \left( \frac{x}{c_n} \right) \right] \\
= & \ \frac{n}{d_n^2} \text{var} \left( \frac{1 - x/c_n}{x - x/c_n} \right) = \frac{n}{d_n^2} \left( 1 - \frac{x}{c_n} \right)^2 \ E\left[ \frac{1}{(X - x/c_n)^2} \right] - \frac{1}{n} \left[ \frac{n}{d_n} \frac{1 - x/c_n}{X - x/c_n} - \xi \right] \frac{n}{d_n} \xi^2 \\
= & \ \frac{n}{d_n} \left[ 1 + O\left(n^{-1/\gamma}\right) \right] \frac{1}{B(\gamma, \delta)} \int_0^{\infty} \frac{y^{\gamma - 1}(1-y)^{\delta - 1}}{(y-x/c_n)^2} \ dy \\
& \ - \frac{1}{n} \left[ \beta + O\left(n^{1-2/\gamma}\right) \right]^2 - \frac{2\gamma}{d_n} \left[ \beta + O\left(n^{1-2/\gamma}\right) \right] - \frac{n}{d_n^2} \xi^2 \\
= & \ \frac{n}{d_n^2} \left( \frac{c_n}{x} \right)^{\gamma - 2} \left[ 1 + O\left(n^{-1/\gamma}\right) \right] \frac{1}{B(\gamma, \delta)} \int_0^{\infty} \frac{z^{\gamma - 1}(1-x/c_n)^{\delta - 1}}{(z+1)^2} \ dz + O\left(n^{1-2/\gamma}\right), \\
& \text{by setting } z = (c_n/x)y \text{ and because } \min(1, \frac{2}{\gamma - 1}, \frac{1}{\gamma}) = \frac{2}{\gamma - 1} \text{ for } 1 < \gamma < 2,
\end{align*}
\]
\[
\frac{n}{d_n} \left( \frac{-x}{c_n} \right)^{y-2} \frac{1}{B(y, \delta)} \int_0^\infty \frac{z^{y-1}}{(z+1)^2} \, dz + o(n^{1-2/y}),
\]

by Lemma 6.4 with \( s = y-1, \ t = 2 \),

\[
\frac{(-x)^{y-2}}{\rho c \gamma} \frac{B(y, 2-y)}{B(y, \delta)} + o(n^{1-2/y}).
\]

Thus the conditions of the Extended Convergence Criterion (Loève [1963], 294-5) are satisfied, and it remains only to verify that the canonical measure is as stated, since the mean of the i.d. law will be \( \beta \) as found above.

Let

\[
x'_n = \frac{1}{d_n} \left[ \frac{1-x/c_n}{X-x/c_n} - \xi \right], \quad \beta_n = \mathbb{E}(X'_n), \quad F_n(y) = P(X'_n \leq y),
\]

and

\[
K_n(u) = n \int_{-\infty}^u \frac{2}{y} dF_n(y + \beta_n).
\]

Since

\[
f_x(v) = \frac{1}{B(y, \delta)} v^{y-1}(1-v)^{\delta-1}, \quad 0 < v < 1,
\]

and

\[
X = \frac{x}{c_n} + \frac{1-x/c_n}{d_n X'_n + \xi},
\]

we find that \( X'_n \) has density
\[ f_n(z) = \frac{1}{B(\gamma, \delta)} \left[ \frac{1-x/c_n}{d_n z + \xi} \right]^{\gamma-1} \left[ 1 - \frac{x}{c_n} \right]^{\delta-1} \frac{d_n (1-x/c_n)}{(d_n z + \xi)^2} \]

with support
\[ \left( \frac{1}{d_n} [1-\xi], \frac{1}{d_n} \left[ \frac{1-x/c_n}{-x/c_n} - \xi \right] \right) . \]

Now, letting \( y = z - \beta_n \), we have, for
\[ u \in \left( \frac{1}{d_n} [1-\xi] - \beta_n, \frac{1}{d_n} \left[ \frac{1-x/c_n}{-x/c_n} - \xi \right] - \beta_n \right) \]
\[ = (O(n^{-1/\gamma}), \frac{1}{-x_0} + O(n^{-1/\gamma})) , \]

(6.1) \( K_n(u) \)
\[ = n \int_{-\infty}^{u} y^2 f_n(y + \beta_n) \, dy = n \int_{[1-\xi]/d_n}^{u+\beta_n} (z-\beta_n)^2 f_n(z) \, dz \]
\[ = \frac{n}{B(\gamma, \delta)} \left[ \frac{c_n}{-x} \right]^{\gamma-1} \lambda_n \left[ \frac{c_n}{d_n} \frac{1-x/c_n}{-x} \frac{1}{r+1} - \frac{\xi}{d_n} - \beta_n \right]^{\delta-1} \left( 1 - \frac{x}{c_n} r \right)^{\delta-1} \, dr , \]

where
\[ \lambda_n = (1 - \frac{x}{c_n}) \frac{c_n}{x} \frac{1}{d_n} u + \frac{1}{d_n} \beta_n + \frac{\xi}{d_n} - 1 = \frac{1}{-x_0 u} - 1 + o(1) \]

and we have let \( r = \frac{c_n}{-x} \left[ \frac{x}{c_n} + \frac{1-x/c_n}{d_n z + \xi} \right] \)
\[ = \frac{(-x)^{\gamma-2}}{\rho^2 c^{\gamma}} \frac{1}{B(\gamma, \delta)} \int_{-x_0 u}^{1} \frac{r^{\gamma-1}}{(r+1)^2} \, dr + O(n^{1-2/\gamma}) , \]

42
by expanding the squared term in the integrand, noting that the integrals with lower limit \( \lambda_n \) differ from those with lower limit \((-xpu)^{-1} - 1\) by at most \( O(n^{-2/\gamma}) \), and then using Lemma 6.4 on each integral.

Therefore \( K(u) = \lim_{n \to \infty} K_n(u) \) for \( 0 < u < 1/xp \), as required. From the first expression, (6.1), for \( K_n(u) \), it is immediately seen that

\[ K(u) = \lim_{n \to \infty} K_n(u) = 0 \text{ for } u < 0. \text{ Also, } K(0) = \lim_{u \downarrow 0} K(u) = 0, \]

and

\[ K(u) = \lim_{v \uparrow 1/(-xp)} K(v) = \frac{(-x)^{\gamma-2}}{\rho c_\gamma} \frac{B(\gamma, 2-\gamma)}{B(\gamma, \delta)} \text{ for } u \geq \frac{1}{-xp}, \]

since

\[ \int_0^\infty \frac{r^{\gamma-1}}{(r+1)^\delta} \, dr \text{ exists. } \]

By setting \( v = -xpu \), the logarithm of the characteristic function of the limit law of

\[ \frac{1}{\rho c_\gamma} \sum_{i=1}^n \left[ \frac{1 - x/(cn_1^{1/\gamma})}{x_i - x/(cn_1^{1/\gamma}) - \xi} \right] \]

may be written, for \( x < 0 \), as

\[ (6.2) \quad \psi(t) = \frac{(-x)^{\gamma}}{\rho c_\gamma} \frac{1}{B(\gamma, \delta)} \int_0^1 \left( e^{itv/(-xp)} - 1 - \frac{itv}{-xp} \right) \frac{(1-v)^{-1}}{v^{\gamma+1}} \, dv + i\beta t. \]

Let \( V_x \) be a random variable whose characteristic function is \( \exp[\psi(t)] \), where \( \psi(t) \) is given by (6.2). Then, for \( x < 0 \),

\[ \lim_{n \to \infty} \sum_{n=1}^\infty \sum_{n=1}^\infty \sum_{n=1}^\infty P(V_x > 0) \text{ can theoretically be found by inversion of the characteristic function; or, since, all moments of } V_x \text{ exist,} \]
methods à la Shohat and Tamarkin [1943] or Royden [1953] can be used to construct bounds on \( P(V_x > 0) \). The semi-invariants (cumulants) of \( V_x \) are easily found to be

\[
k_1 = -\frac{(-x)^{-1}}{\rho c \gamma} \frac{B(\gamma-1, 2-\gamma)}{B(\gamma, \delta)}
\]

and

\[
k_k = \frac{(-x)^{-k}}{\rho c \gamma^k} \frac{B(\gamma, k-\gamma)}{B(\gamma, \delta)} \quad \text{for} \quad k \geq 2.
\]

Since \( \rho > 0 \) can be chosen arbitrarily, the choice \( \rho = -1/x \) might prove of particular use in evaluating \( P(V_x > 0) \). This choice gives

\[
\psi(t) = \left(\frac{-x}{c}\right)^{\gamma} \frac{1}{B(\gamma, \delta)} \int_0^1 (e^{itv} - 1 - itv) \frac{(1-v)^{\gamma-1}}{v^{\gamma+1}} dv - it \left(\frac{-x}{c}\right)^{\gamma} \frac{B(\gamma-1, 2-\gamma)}{B(\gamma, \delta)}.
\]

We are now in a position to improve the consistency result for \( \hat{\alpha}_n \) even if we do not wish to evaluate \( \lim_{n \to \infty} F_{c_n, n_n} \). Using the well-known inequality (Feller [1966], 150)

\[
P(X - E(X) > t) = \frac{\var(X)}{\text{var}(X) + t^2} \quad \forall t > 0,
\]

we find that, for \( x < 0 \),

\[
\lim_{n \to \infty} F_{c_n, n_n} (x) = P(V_x > 0) = P(V_x - \mu > -\mu) \leq \frac{\sigma^2}{\sigma^2 + \mu^2}
\]

\[
= \left[ 1 + \left(\frac{-x}{c}\right)^{\gamma} \frac{\Gamma(2-\gamma) \Gamma(\gamma+\delta)}{(\gamma-1)^2} \right]^{-1} \to 0 \quad \text{as} \quad x \to -\infty.
\]

Since it is also true that \( \hat{\alpha}_n < X_{ln} = \alpha(p_{n-1/\gamma}) \), we can conclude that \( \hat{\alpha}_n = o_p(n^{-1/\gamma}) \), the same order as \( X_{ln} \).
For \( x > 0 \), no results are yet known for \( \lim_{n \to \infty} F_{\frac{X}{c_n}}(x) \). The best that can be said at present is that

\[
F_{\frac{X}{c_n}}(x) = P\left( \frac{X}{c_n} < X_{ln}, \sum_{n} \frac{X}{c_n} > 0 \right) + F_{\frac{X}{c_n}}(x).
\]

Gnedenko's [1943] result, quoted at (2.1), gives the limit of the second term, but the limit of the first term is not known.

We end this section with the lemma whose results were used in the proof of Theorem 6.3.

**Lemma 6.4:** If \( \epsilon \geq 0, \delta > 0 \), and \( c_n = cn^{1/\gamma} > 0 \), then

\[
I_n = \int_{\epsilon}^{c_n} \frac{z^{s(1-z/c_n)^{\delta-1}}}{(z+1)^t} \, dz = \begin{cases} 
O(n^{(s-t+1)/\gamma}) & \text{for } 0 \leq t < s+1 \\
\int_{\epsilon}^{c_n} \frac{z^s}{(z+1)^t} \, dz + O(n^{(s-t+1)/\gamma}) & \text{for } 0 < s+1 < t < s+2.
\end{cases}
\]

**Proof:** Set \( y = z/c_n \). Then, for \( 0 \leq t \),

\[
I_n = \int_{\epsilon/c_n}^{c_n} \frac{y^{s(1-y)^{\delta-1}}}{(y + 1/c_n)^t} \, dy \leq \int_{0}^{1} y^{s-t}(1-y)^{\delta-1} \, dy < +\infty,
\]

provided \( s-t > -1 \). This completes the proof for \( 0 \leq t < s+1 \).

For the second part, write

\[
I_n = \int_{\epsilon}^{c_n} \frac{z^s}{(z+1)^t} \, dz - \int_{\epsilon/c_n}^{c_n} \frac{z^s}{(z+1)^t} \, dz + \int_{\epsilon/c_n}^{c_n} \frac{z^s}{(z+1)^t} \left[ (1-z/c_n)^{\delta-1} - 1 \right] \, dz
\]

\[
= I - J_n + R_n,
\]

and note that integral \( I \) exists for \( 0 < s+1 < t \). The second
integral is easily taken care of since

\[ 0 < J_n < \int_0^\infty z^{s-t} \, dz = \frac{c_n^{s-t+1}}{s-t+1} = O(n^{(s-t+1)/\gamma}) \text{ for } s-t+1 \neq 0. \]

For \( R_n \), again set \( y = z/c_n \), then, when \( 0 \leq t \),

\[
\frac{|R_n|}{c_n^{s-t+1}} = \int_0^{\infty} \frac{y^s}{(y+1/c_n)^t} \left| (1-y)^{\delta-1} - 1 \right| dy \\
\leq \int_0^1 y^{s-t} \left| (1-y)^{\delta-1} - 1 \right| dy < +\infty \quad \text{for } s-t+1 > -1,
\]

since the integrand is \( O(y^{s-t+1}) \) as \( y \downarrow 0 \), and there is no problem as \( y \uparrow 1 \) because \( \delta > 0 \). Therefore, \( R_n = O(n^{(s-t+1)/\gamma}) \) for \( 0 \leq t < s+2 \), and putting everything together we find \( I_n = I + O(n^{(s-t+1)/\gamma}) \), for \( 0 < s+1 < t < s+2 \). ||

What can be said about the asymptotic efficiency of the mle for \( a \) when \( 0 < \gamma < 2 \), and \( a \) is the only unknown parameter? For this parameter range the proofs of the Polfeldt [1970] results can undoubtedly be modified to cover the generalized Beta case in addition to the one-sided situation for which they were originally designed. Our Corollary 5.4 is an example of how Polfeldt's proofs can be so modified.

For \( 1 < \gamma < 2 \) Polfeldt's Theorem 3.3 shows that the Kiefer lower bound for unbiased estimators of \( a \) is \( O(n^{-2}/\gamma) \). Thus our mle \( \hat{a}_n \) has the optimal rate of consistency, \( O_p(n^{-1}/\gamma) \), but we do not yet know enough about its distribution to compare the constants.
associated with its rate to those given by Polfeldt for the Kiefer lower bound. In this case, then, the asymptotic efficiency of $\hat{\theta}_n$ is neither ruled out nor confirmed.

When $0 < \gamma \leq 1$ the mle is $Y_{ln}$. Polfeldt's Theorem 3.3 shows $Y_{ln}$ to be asymptotically efficient when $\gamma = 1$. For $0 < \gamma < 1$ Polfeldt's Section 4 results for linear combinations of order statistics show that, while $Y_{ln}$ has the correct order of consistency, it is not asymptotically efficient. Polfeldt's Table 4 shows that there are linear combinations of several extreme order statistics which are asymptotically more efficient than $Y_{ln}$. For a range of $\gamma$ values and various numbers of extreme order statistics, the same table gives the asymptotically optimal coefficients for these linear combinations. Polfeldt's Section 5 demonstrates, furthermore, that for $0 < \gamma < 1$ even these linear combinations are asymptotically less efficient than some best quadratic estimators, which in turn still do not attain the Kiefer lower bound provided by Polfeldt's Theorem 3.3 and his Sections 3.7 and 3.8.

As a final comment we note that, except for $\gamma = 1$, no estimator is known which asymptotically achieves either the Kiefer or Polfeldt lower bound when $0 < \gamma < 2$, nor is it known whether they can indeed be attained.
7. **Linearization of Likelihood Equations**

Recall that for $B_{\gamma,\delta}(a,b)$ the system of likelihood equations can be written

\[(7.1) \quad g_n(\theta, y) = \frac{1}{n} \sum_{i=1}^{n} g(\theta, Y_i) = 0,\]

where the components of $g(\theta, y)$ are

\[(7.2) \quad g_1(\theta, y) = \frac{Y+\delta-1}{b-a} - \frac{Y-1}{y-a}\]

\[(7.3) \quad g_2(\theta, y) = -\frac{Y+\delta-1}{b-a} + \frac{\delta-1}{b-y}\]

\[(7.4) \quad g_3(\theta, y) = \psi(Y+\delta) - \psi(Y) + \ln\left(\frac{Y-a}{b-a}\right)\]

\[(7.5) \quad g_4(\theta, y) = \psi(Y+\delta) - \psi(\delta) + \ln\left(\frac{b-y}{b-a}\right).\]

If (7.2) is multiplied by $(b-a)/(\gamma-1)$, and (7.3) by $(b-a)/(\delta-1)$, then each of the four components will be monotone in each of the four parameters individually, and numerical solution of the system should be facilitated. While the following discussion will deal with the system as given in (7.1)-(7.5), the results will apply also to the monotone system.
Solution of the likelihood system by iteration on the linearized equations has been suggested in the literature by many authors. The coefficients used in the linearization may consist of the elements of the information matrix, or of various estimates of these elements, but each method is essentially the Newton-Raphson method for solving a system of equations iteratively.

Kale [1961; 1962] proves, under usual (Cramér type) regularity conditions, that if the initial estimator is consistent, then any of its iterates (i.e., first, second, third, ...) is also consistent. Kale says nothing about the "rate of consistency". It is well-known, however (Ferguson [1958, pp. 1049-50]; Chernoff [1962, p. 405]; Dubey [1966, p. 230]), that if the initial estimator is $O_p(n^{-1/2})$-consistent, and if usual regularity conditions are satisfied, then a single Newton-Raphson iteration is sufficient to produce a CANE estimator. Furthermore, it is part of the (apparently unpublished, but easily proven) lore of statistics that it is sufficient for the initial estimator to be $O_p(n^{-1/4})$-consistent for a single iteration to yield a CANE estimator. Published proofs for the $O_p(n^{-1/2})$ case (and the easy proof referred to above for the $O_p(n^{-1/4})$ case) use the Cramér regularity conditions which, as we have noted before, do not hold when estimating the end points of the generalized Beta distribution.

Consideration of the graph of $G_n(s, Y)$ in Figure 4.1 makes it clear that the $O_p(n^{-1/4})$ result cannot possibly be true for the generalized Beta distribution when estimating the end points unless the governing shape parameters have a value of at least 4. For instance,
if \( r < 4 \) and we are estimating \( a \), then we can find \( o_p(n^{-1/4}) \)-consistent estimators which will i.p. fall above \( Y_{1n} \), and which will yield first iterates also falling above \( Y_{1n} \); or consider the initial estimator \( Y_{1n} \) which is itself \( o_p(n^{-1/4}) \)-consistent, but for which the Newton-Raphson iteration is not even defined since the required derivative does not exist at \( Y_{1n} \).

We shall prove that, when dealing with \( B_r,(a,b) \), for shape parameters greater than or equal to 4 the \( o_p(n^{-1/4}) \) result holds, and for shape parameters between 2 and 4 the CANE result holds for initial estimators of higher orders of consistency. The result is still better than the \( O(n^{-1/2}) \) case, which follows as a corollary.

**Theorem 7.1**: For the generalized Beta distribution, let

\[
A_{\ln}(\theta, \bar{Y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g(\theta, Y_i)}{\partial \theta} \bigg|_{\theta = \bar{Y}}, \quad \text{and} \quad A_{\on}(\theta, \bar{Y}) = \frac{1}{n} \sum_{i=1}^{n} g(\theta, Y_i),
\]

and w.l.o.g. if both end points are unknown let \( r_0 \leq \delta_0 \), and if only one end point is unknown let it be \( a \). If

\[
\tilde{\theta} \sim n = \begin{cases} 
\tilde{\theta}_0 + o_p(n^{-1/r_0}) & \text{for } 2 < r_0 < 4 \\
\tilde{\theta}_0 + o_p(n^{-1/4}) & \text{for } 4 \leq r_0,
\end{cases}
\]

then

\[
\tilde{\theta} \sim n = \tilde{\theta} \sim n - A_{\ln}^{-1}(\tilde{\theta} \sim n, \bar{Y}) A_{\on}(\tilde{\theta} \sim n, \bar{Y})
\]

exists i.p. and has the same asymptotic distribution as \( \tilde{\theta} \sim n \).
Proof: For the generalized Beta distribution the Taylor expansion of 
\[ g_n(\theta, Y) = \frac{1}{n} \sum_{i=1}^{n} g(\theta, Y_i) \] 
about \( \bar{\theta}_n \) exists and is valid at \( \hat{\theta}_n \) i.p., 
because \( 2 \leq \gamma_o \leq \delta_o, \bar{\theta}_n = \theta_o + o_p(n^{-1/2}) \), and 
\( \hat{\theta}_n = \theta_o + n^{1/2} \gamma_o^{-1/2} + o_p(n^{-1/2}) \). The expansion is

\[ Q = Q_n(\bar{\theta}_n, \bar{Y}) = A_{on}(\bar{\theta}_n, \bar{Y}) + A_{ln}(\bar{\theta}_n, \bar{Y})(\bar{\theta}_n - \bar{\theta}_n) + R_n(\hat{\theta}_n, \bar{\theta}_n, \bar{Y}) \]  
(7.9)

where \( R_n(\hat{\theta}_n, \bar{\theta}_n, \bar{Y}) = \frac{1}{2} (\bar{\theta}_n - \bar{\theta}_n)' \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 g_i(\theta, Y_i)}{\partial \theta \partial \theta} \right]_{\bar{\theta}_n} (\bar{\theta}_n - \bar{\theta}_n) \)

for some \( \omega_{nj} \) on the line segment joining \( \bar{\theta}_n \) and \( \bar{\theta}_n \). By Lemma 7.2(a), \( A_{ln}(\bar{\theta}_n, \bar{Y}) = -J + o_p(1) \), and since \( J > 0 \) we have 
\( A_{ln}(\bar{\theta}_n, \bar{Y}) = -J^{-1} + o_p(1) \). Hence, \( \bar{\theta}_n \) exists i.p., and by straightforward 
manipulation the Taylor expansion at (7.9) yields

\[ \hat{\theta}_n = \bar{\theta}_n - A_{ln}(\bar{\theta}_n, \bar{Y}) A_{on}(\bar{\theta}_n, \bar{Y}) - A_{ln}(\bar{\theta}_n, \bar{Y}) R_n(\hat{\theta}_n, \bar{\theta}_n, \bar{Y}) \]

\[ = \bar{\theta}_n + [J^{-1} + o_p(1)] R_n(\hat{\theta}_n, \bar{\theta}_n, \bar{Y}) \].

Lemma 7.2(b) shows that \( R_n(\hat{\theta}_n, \bar{\theta}_n, \bar{Y}) = o_p(n^{-1/2}) \), and this gives 
\( \hat{\theta}_n = \bar{\theta}_n + o_p(n^{-1/2}) \). Therefore,

\[ n^{1/2}(\hat{\theta}_n - \theta_o) = n^{1/2}(\bar{\theta}_n - \theta_o) + n^{1/2}(\bar{\theta}_n - \hat{\theta}_n) \]

\[ = n^{1/2}(\bar{\theta}_n - \theta_o) + o_p(1), \]

51
and by Slutsky's theorem we see that \( n^{1/2}(\tilde{\theta}_n - \theta_0) \) and \( n^{1/2}(\hat{\theta}_n - \theta_0) \) have the same limit law, namely \( \gamma(\theta, \gamma^{-1}) \).

**Remarks:**

1. Careful scrutiny of the proof of the following lemma reveals that the above result will still hold if each component of \( \tilde{\theta}_n \) is of an order of consistency appropriate for the result to hold in the case that it is the single unknown parameter in the problem. That is, \( \tilde{y}_n \) and \( \tilde{\sigma}_n \) need be only \( o_p(n^{-1/4}) \)-consistent, while \( \tilde{a}_n \) and \( \tilde{b}_n \) need be individually consistent only to the order required by the value of their respective governing shape parameters.

2. The result clearly holds when some subset of the parameters is known and the remaining likelihood equations are linearized analogously.

3. That the first iterate is CANE when the initial estimator is \( \hat{\theta}_n = \theta_0 + O_p(n^{-1/2}) \) is clearly a consequence of the much stronger Theorem 7.1.

Now we shall establish the lemma used in the proof of the theorem.

**Lemma 7.2:** Under the conditions of Theorem 7.1,

(a) \( A_{1n}(\tilde{\theta}_n, \tilde{y}) = J + o_p(1) \)

(b) \( R_n(\tilde{\theta}_n, \tilde{\theta}_n, \tilde{y}) = o_p(n^{-1/2}) \).
Proof: Theorems 4.1 and 4.2 showed that \( A_{\ln}(\theta, \bar{\gamma}) = -J + o_p(1) \). For (a) it remains only to show that \( A_{\ln}(\bar{\theta}, \bar{\gamma}) - A_{\ln}(\theta, \bar{\gamma}) = o_p(1) \). Recall that if a function \( \phi \) is differentiable at \( \theta \) and has derivative matrix \( \dot{\phi}(\theta) \) there, we can write

\[
\phi(\bar{\theta}) - \phi(\theta) = \dot{\phi}(\theta)(\bar{\theta} - \theta) + o(\|\bar{\theta} - \theta\|) = o(\|\bar{\theta} - \theta\|),
\]

for \( \|\bar{\theta} - \theta\| = o(1) \). It follows immediately that all terms which involve only \( \bar{\theta} \) and \( \theta \) in the difference matrix, \( A_{\ln}(\bar{\theta}, \bar{\gamma}) - A_{\ln}(\theta, \bar{\gamma}) \), are \( o_p(\text{const}) = o_p(1) \) since their derivatives exist at \( \theta \) by direct check. The remaining terms are of the form

\[
\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{(Y_i - a)} - \frac{1}{(Y_i - \bar{a})} \right) = -\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{(Y_i - a)} - \frac{1}{(Y_i - \bar{a})} \right)
\]

for \( \epsilon = 1 \) and 2,

or have factors \( \bar{\gamma} \) and \( \gamma \) in the numerators; but these latter factors are easily seen not to alter the order statements of the result we shall prove. Similar symmetric terms in \( b_n, b, \bar{b}, \bar{b} \) also occur.

By simple algebra and a trivial modification of Lemma 2.1 the above difference is, for \( \epsilon = 1 \),

\[
\left( a_n - a \right) \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{(Y_i - a_n)(Y_i - a)} \right) = \begin{cases} o_p(n^{-1/2}) & \text{for } 2 < \gamma < 4 \\ o_p(n^{-1/4}) & \text{for } 4 \leq \gamma \end{cases}
\]

and for \( \epsilon = 2 \),

\[
53
\]
\[
\left(\bar{a}_n - a_0\right) \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{1}{(y_i - \bar{a}_n)^2 (y_i - a_0)} + \frac{1}{(y_i - \bar{a}_n)(y_i - a_0)^2} \right]
= \begin{cases}
\mathcal{O}_p\left(n^{\frac{-1}{2} / \gamma_0}\right) \mathcal{O}_p\left(n^{2 / \gamma_0 - 1 / 2}\right) & \text{for } 2 < \gamma_0 \leq 3 \\
\mathcal{O}_p\left(n^{\frac{-1}{2} / \gamma_0}\right) \mathcal{O}_p\left(1\right) & \text{for } 3 < \gamma_0 < 4 \\
\mathcal{O}_p\left(n^{\frac{-1}{4} / \gamma_0}\right) \mathcal{O}_p\left(1\right) & \text{for } 4 \leq \gamma_0
\end{cases} = \mathcal{O}_p\left(1\right),
\]

where we have chosen \( \gamma' = 6\gamma_0/(4+\gamma_0) \) in Lemma 2.1 for \( 2 < \gamma_0 < 3 \).

Thus the difference of the matrices is \( \mathcal{O}_p\left(1\right) \), and the proof of (a) is complete.

For (b) note first that \( \mathcal{W}_{nj} = \bar{\theta}_n + \mathcal{O}_p\left(n^{-1 / \gamma_0}\right) \) because \( \bar{\theta}_n = \bar{\theta}_0 + \mathcal{O}_p\left(n^{-1 / \gamma_0}\right) \). Troublesome terms in this part are of the type \( \frac{1}{n} \frac{1}{\sum_{i=1}^{n} (y_i - \bar{a}_n)^3} \), but with the same \( \gamma' \), Lemma 2.1 again gives the result since

\[
\mathbb{R}_n(\bar{\theta}_n, \bar{\theta}_n, \bar{\gamma}) = \begin{cases}
\mathcal{O}_p\left(n^{-\frac{1}{2}} / \gamma_0\right) \mathcal{O}_p\left(n^{\frac{2}{\gamma_0} - \frac{1}{2}}\right) & \text{for } 2 < \gamma_0 \leq 3 \\
\mathcal{O}_p\left(n^{-\frac{1}{2}} / \gamma_0\right) \mathcal{O}_p\left(1\right) & \text{for } 3 < \gamma_0 < 4 \\
\mathcal{O}_p\left(n^{-\frac{1}{4}} / \gamma_0\right) \mathcal{O}_p\left(1\right) & \text{for } 4 \leq \gamma_0
\end{cases} = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right).
\]

54
Further Remark: By using the type of argument employed in Lemma 7.2, it is relatively easy to check that Theorem 7.1 holds also when $A_{in}(\tilde{\theta}_n, x)$ is replaced, in the definition of $\tilde{\theta}_n$, by

$$-\frac{1}{n} \sum_{i=1}^{n} [g(\tilde{\theta}_n, Y_i)][g(\tilde{\theta}_n, Y_i)'] \quad \text{or by} \quad -\tilde{J}^*(\tilde{\theta}_n),$$

where $\tilde{J}^*(\tilde{\theta}_n)$ has the functional form of $\tilde{J}$, but has been evaluated at $\tilde{\theta} = \tilde{\theta}_n$.

In order to prove a similar linearization result when the shape parameters have the value two, we shall need the following lemma.

**Lemma 7.3:** If $X_1, X_2, \ldots$ i.i.d. $B_2, c(0,1)$, then

$$\sum_{i=1}^{n} X_i^{-3} = o_p(n^{3/2} v_n) \iff v_n \longrightarrow +\infty.$$

**Proof:** Let $c_n = n^{3/2} v_n$ and $V_{2n}^{-3} = c_n X_1^{-3}$, let $V_{in}$ be $V_{in}$ truncated at $\pm s$, and let $s_n = (sc_n)^{-1/3}$. Then

$$(7.10) \quad \sum_{i=1}^{n} P(\mid V_{in} \mid > s) = n \int_{0}^{s} x(1-x)^{\delta-1} dx = O(n^{2/3} s_n^2) = o(v_n^{2/3}) = o(1) \quad \text{iff} \quad v_n \longrightarrow +\infty,$$

and

$$E(V'_{in}) = s \int_{0}^{s} x(1-x)^{\delta-1} dx + \frac{1}{c_n} \int_{s_n}^{1} \frac{1}{2} (1-x)^{\delta-1} dx$$

$$= o(c_n^{-2/3}) = o(n^{-1} v_n^{-2/3}).$$
and

\[ E((V_{in}^{'})^2) = s^2 \int_0^1 x(1-x)^8 \, dx + \frac{1}{c_n^2} \int_0^1 \frac{1}{x} (1-x)^8 \, dx \]

\[ = O(c_n^{-2/3}) = O(n^{-\frac{1}{3}}v_n^{-2/3}), \]

from which we find

\[ (7.11) \quad n \operatorname{var}(V_{in}^{'}) = nO(n^{-1}v_n^{-2/3}) = nO(n^{-1}v_n^{-2/3}) = O(v_n^{-2/3}) = o(1) \]

iff \( v_n \to + \infty \).

The W.L.L.N. for triangular arrays (Feller [1966], Theorem IX.9.1), again asserts that

\[ c_n^{-1} \sum_{i=1}^n X_i^{-3} - nE(V_{in}^{'}) = o_p(1) \quad \text{iff (7.10) and} \]

\[ (7.11) \quad \text{hold. That is,} \quad (n^{3/2}v_n^{-1}) \sum_{i=1}^n X_i^{-3} = o_p(1) \quad \text{iff} \]

\[ v_n \to + \infty \].

Now with \( A_{on}, A_{ln} \) and \( B_n \) defined as in (7.6) and (7.9) except that the divisors \( n \) are replaced by the divisors \( 3n \ln n \), and \( \sim = (a,b)' \), the following theorem can be proven.

**Theorem 7.4**: For \( B_n, (a,b) \) when \( \gamma = \delta = 2 \) are known, if

\[ \tilde{\theta}_o = \theta_o + o_p(n^{-1/2}), \]

then

\[ \tilde{\theta}_n = \tilde{\theta}_n - A_{ln}^{-1}(\tilde{\theta}_n, \tilde{\gamma}) A_{on}(\tilde{\theta}_n, \tilde{\gamma}) \]

exists i.p. and has the same asymptotic distribution as \( \tilde{\theta}_n \).
Proof: The Taylor expansion of \( G_n(\tilde{\theta}, \tilde{X}) \) about \( \tilde{\theta} \) exists and is valid at \( \tilde{\theta} \) i.p. By Lemma 5.1(b), \( A_{\ln}(\tilde{\theta}, \tilde{X}) = 1/(b_o - a_o)^2 + o_p(1) \). The only terms requiring special attention in \( A_{\ln}(\tilde{\theta}, \tilde{X}) - A_{\ln}(\tilde{\theta}, \tilde{X}) \) are of the type

\[
(7.12) \quad \frac{1}{3n \ln n} \sum_{i=1}^{n} \left( \frac{1}{(Y_i - \tilde{a}_n)^2} - \frac{1}{(Y_i - a_o)^2} \right)
\]

\[
= (a_n - a_o) \frac{1}{3n \ln n} \sum_{i=1}^{n} \left[ \frac{1}{(Y_i - \tilde{a}_n)^2} + \frac{1}{(Y_i - a_o)^2} \right].
\]

Choosing \( v_n = \ln^{1/2} n \) in Lemma 7.3, and using the fact that for any \( c, 0 < c < 1 \), and any \( k > 0, \)

\[
P(c(\tilde{Y}_i - a_o)^k \leq [\tilde{Y}_i - a_o - o_p(n^{-1/2})]^k, \forall i = 1, 2, \ldots, n) \rightarrow 1
\]

as \( n \rightarrow \infty \), we find that (7.12) is \( o_p(n^{-1/2})(n \ln n)^{-1} o_p(n^{3/2} \ln^{1/2} n) = o_p(\ln^{1/2} n) = o_p(1) \). Therefore, \( A_{\ln}(\tilde{\theta}, \tilde{X}) - A_{\ln}(\tilde{\theta}, \tilde{X}) = o_p(1) \), and \( A_{\ln}(\tilde{\theta}, \tilde{X}) \) is invertible i.p., and \( \tilde{\theta}_n \) exists i.p. Terms in \( R_n \) which do not involve the \( Y_i \)'s pose no problem, and terms which do involve the \( Y_i \)'s are of the type

\[
(7.13) \quad \frac{1}{2} (\tilde{\theta}_n - \tilde{a}_n) \frac{1}{3n \ln n} \sum_{i=1}^{n} \frac{1}{(Y_i - \tilde{a}_n)^2} (\tilde{\theta}_n - \tilde{a}_n).
\]

Since \( a_{n1} = a_o + o_p(n^{-1/2}) \), the above result also shows that (7.13) is \( o_p(n^{-1/2})(n \ln n)^{-1} o_p(n^{3/2} \ln^{1/2} n) = o_p(n^{-1/2} \ln^{-1/2} n) \), and hence that \( A_{\ln}(\tilde{\theta}, \tilde{X}) R_n(\tilde{\theta}, \tilde{\theta}, \tilde{X}) = o_p(n^{-1/2} \ln^{-1/2} n) \). By a method analogous
to the one used in the proof of Theorem 7.1 these results show that
\[ \hat{\theta} = \tilde{\theta} + o_p(n^{-1/2} \ln^{-1/2} n), \]
and therefore that \( \hat{\theta} \) and \( \tilde{\theta} \) have
the same asymptotic distribution. ||
8. Initial Estimators

If the preceding results concerning linearization of the likelihood equations are to be of any use, we must be able to find suitable initial estimators for $a$, $b$, $\gamma$, $\delta$.

The first natural method to consider is the method of moments. The first moment, and the second, third and fourth central moments of $B_{\gamma, \delta}(a,b)$ are

\begin{align*}
(8.1) \quad \mu &= a + (b-a) \frac{\gamma}{\gamma + \delta} \\
(8.2) \quad \mu_2 &= (b-a)^2 \frac{\gamma \delta}{(\gamma + \delta)^2 (\gamma + \delta + 1)} \\
(8.3) \quad \mu_3 &= (b-a)^3 \frac{2\gamma \delta (\delta - \gamma)}{(\gamma + \delta)^3 (\gamma + \delta + 1)(\gamma + \delta + 2)} \\
(8.4) \quad \mu_4 &= (b-a)^4 \frac{2\gamma \delta ((\gamma + \delta)^2 + \gamma \delta (\gamma + \delta - 2) + (\gamma - \delta)^2)}{(\gamma + \delta)^4 (\gamma + \delta + 1)(\gamma + \delta + 2)(\gamma + \delta + 3)}
\end{align*}

These equations can be solved explicitly for $a$, $b$, $\gamma$, $\delta$. The solutions can be more easily expressed if the coefficients of skewness and excess,

\begin{align*}
\gamma_1 &= \frac{\mu_3}{\mu_2^{3/2}} \quad \text{and} \quad \gamma_2 = \frac{\mu_4}{\mu_2^2} - 3,
\end{align*}

are also introduced. From the definitions of these two coefficients, which are independent of $a$ and $b$, the relationships

59
\[ \tau \equiv \gamma + \delta = -2 + \frac{2\gamma_2 + 12}{(3\gamma_1^2 - 2\gamma_2)} \]

(8.6)

\[ \pi \equiv \gamma \delta = 4\tau^2(\tau+1)/[\gamma_1^2(\tau+2)^2 + 16(\tau+1)] \]

can be derived, where \( \tau \) and \( \pi \) are defined as the sum and product, respectively, of the two shape parameters. The following expressions for \( \gamma \) and \( \delta \) are easily found by simple manipulation:

\[ \gamma = \frac{1}{2} \left[ \tau - \text{sgn}(\gamma_1)(\tau^2 - 4\pi)^{1/2} \right] \]

(8.7)

\[ = \frac{1}{2} \tau - \frac{1}{2} \gamma_1 \tau(\tau+2)/[\gamma_1^2(\tau+2)^2 + 16(\tau+1)]^{1/2} \]

\[ \delta = \tau - \gamma. \]

The solutions

\[ a = \mu - \tau[\mu_2(\tau+1)/\pi]^{1/2} \]

(8.8)

\[ b = \mu + \delta[\mu_2(\tau+1)/\pi]^{1/2} \]

result from (8.1) and (8.2) by straightforward calculation.

By substituting the corresponding sample quantities in (8.7) and (8.8), via the intermediary computations of (8.5) and (8.6), the moment estimators for \( a, b, \gamma, \delta \) are easily computed. For all values of the parameters the moment estimators with normalizing sequence \( n^{1/2} \), have asymptotically a proper normal distribution. Thus, for \( \gamma > 2 \),

60
When $\delta > 2$, the moment estimators can serve as initial estimators to yield CANE first iterates; but the moment estimators are not suitable as initial estimators when $\gamma = 2$ and/or $\delta = 2$, if CANE first iterates are desired for $a$ and $b$.

There are of course many other statistics, in addition to the sample skewness and excess, which are invariant under location and scale transformations and from which other initial estimators for $\gamma$ and $\delta$ (and consequently for $a$ and $b$ via (8.8)) can be found. Judicious choice of such statistics might allow us to find initial estimators which are more efficient or easier to compute than the moment estimators. We shall briefly consider a class of such invariant statistics constructed by forming the ratio of two appropriate linear combinations of the order statistics. We shall restrict the class by considering only linear combinations of several sample quantiles.

The initial estimators $\bar{\gamma}_n$, $\bar{\delta}_n$, $\bar{a}_n$, $\bar{b}_n$ found by this method will be $O_p(n^{-1/2})$-consistent, and will therefore yield CANE first iterates when $\gamma, \delta > 2$.

Let $U_{in} = \sum_{j=1}^{r} a_{ij} Y_{[np_j]}n$ and $V_{in} = \sum_{j=1}^{r} b_{ij} Y_{[np_j]}n$, where $0 < p_1 < p_2 < \cdots < p_r < 1$; $a_{ij}$ and $b_{ij}$, $j = 1, 2, \ldots, r$, are constants, and $Y_{[np_j]}n$ is the $[np_j]$-th ordered observation from a sample of size $n$. The case $[np_j] = 0$ causes no trouble since we shall be concerned only with asymptotic results. In practice the $Y_{[np_j]}n$ could be replaced by $Y_{[np_j]+1,n}$. The asymptotic theory will be unaffected by this change.
The ratio \( S_{in} = U_{in}/V_{in} \) is invariant under scale transformations on the distribution of \( Y \), and if \( \sum_{j=1}^{r} a_{ij} = 0 = \sum_{j=1}^{r} b_{ij} \), the ratio is also invariant under location transformations. If the zero sum condition is imposed for all \( i \), then it is sufficient, when studying the distribution of \( S_n = (S_{1n}, S_{2n}, \ldots, S_{pn})' \), to consider the \( Y_{kn} \) as the ordered observations from the standard distribution with location parameter 0, and scale parameter 1.

Now let \( \lambda_j = F^{-1}(p_j) \) be the \( p_j \) quantile of the distribution of \( Y \), and let \( \mu_i = \sum_{j=1}^{r} a_{ij} \lambda_j \), \( \nu_i = \sum_{j=1}^{r} b_{ij} \lambda_j \), \( s_i = \mu_i/\nu_i \), \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)' \) and \( Y[\cdot,n] = (Y[np_1], n', Y[np_2], n', \ldots, Y[np_r], n') \).

Provided that \( \nu_1 \neq 0 \), the Taylor expansion of \( S_{in} \) about \( \lambda \) is

\[
S_{in} = s_i + \sum_{j=1}^{r} d_{ij} (Y[np_j], n - \lambda_j) + R_{in}
\]

\[
= s_i + D_i (Y[\cdot,n] - \lambda) + R_{in},
\]

where \( d_{ij} = (a_{ij} \nu_i - b_{ij} \mu_i)/\nu_i^2 \) and \( D_i = (d_{i1}, d_{i2}, \ldots, d_{ir})' \).

When the density \( f \) is non-zero and differentiable with a bounded derivative in some neighborhood of each \( \lambda_j \), then

\[
n^{-1/2}(Y[\cdot,n] - \lambda) = \sqrt{n} (Q, \Lambda) + o_p(1)
\]

where the elements of the covariance matrix are \( \lambda_{kj} = \lambda_{jk} = \frac{\lambda_j (1-\lambda_k)}{f(\lambda_j) f(\lambda_k)} \) for \( j \leq k \). From this result it also follows easily that the remainder term \( R_{in} \) is \( o_p(n^{-1/2}) \), in fact, \( o_p(n^{-1}) \), provided that
\( \nu_i \neq 0 \). It is immediate that

\[
n^{1/2}(s_i - \theta_i) = N(0, D_i'AD_i) + o_p(1),
\]

and with \( D = (D_1, D_2, \ldots, D_p) \) that

\[
n^{1/2}(\theta_n - \theta) = \mathcal{N}_p(0, \tilde{\Sigma}) + o_p(1),
\]

where \( \tilde{\Sigma} = DAD \).

Each \( s_i \) is a function of the parameter \( \theta \) which we seek to estimate. Suppose \( s_i = \varphi_i(\theta) \). Consider the case in which \( \dim(\varphi) = \dim(\theta) = p \), and the \( p \) elements of the system \( \varphi = \varphi(\theta) \) are functionally independent. It is proposed to estimate \( \theta \) as follows:

1. From a sample of size \( n \) compute \( \tilde{\theta}_n \).
2. Solve \( \tilde{\theta}_n = \varphi(\theta) \) for \( \theta \), calling the solution \( \hat{\theta}_n \).

When the second partial derivatives of \( \varphi \) exist and are continuous in a neighborhood of \( \theta_0 \), the existence and distribution proofs for mle's can be mimicked to show first that, \( i.p. \) \( \exists \) a solution \( \hat{\theta}_n \) of \( \tilde{\theta}_n = \varphi(\theta) \) such that \( \hat{\theta}_n = \theta_0 + o_p(n^{-1/2}) \), and then from

\[
\tilde{\theta}_n = \varphi(\theta_0) + J_\varphi(\tilde{\theta}_n - \theta_0) + R_n(\tilde{\theta}_n, \theta_0),
\]

that

\[
n^{1/2}(\tilde{\theta}_n - \theta_0) = \mathcal{N}_p(0, \varphi) + o_p(1),
\]

63
where
\[ J_\varphi = \left. \frac{\partial \varphi(\theta)}{\partial \theta} \right|_{\theta_0} \quad \text{and} \quad \varphi = (J_\varphi^{-1}) \frac{\partial}{\partial (J_\varphi^{-1})}. \]

Note that \( J_\varphi \) is non-singular because the elements of \( \varphi(\theta) \) are functionally independent.

Note also that, since \( \theta_0 \) is unknown, the \( b_{ij} \)'s must be chosen so that all \( v_i \neq 0 \) for all \( \theta \) in the parameter space. One sufficient condition for \( v_i \neq 0 \) for all \( \theta \) is that
\[
\exists \, J_1 \ni b_{ij} \begin{cases} 
\leq (\geq) 0 & \text{for } j \leq J_1 \\
\geq (\leq) 0 & \text{for } j > J_1,
\end{cases}
\]

where either both inequalities without or both inequalities within the parentheses are satisfied. This is not a necessary condition as the sequence of \( b_{ij} \)'s, \((-2, 1, -1, 2)\), clearly demonstrates. It is, however, relatively easy to show that if \( \sum_{j=1}^{r} b_{ij} = 0 \) where not all \( b_{ij} = 0 \), then a necessary and sufficient condition for \( \sum_{j=1}^{r} b_{ij} \lambda_j \neq 0 \) for any continuous random variable, is
\[
\sum_{j=1}^{k} b_{ij} \geq 0 \quad \forall \, k = 1, 2, \ldots, r
\]
or
\[
\sum_{j=k}^{r} b_{ij} \geq 0 \quad \forall \, k = 1, 2, \ldots, r.
\]

When finding initial estimators for \( \gamma, \delta \) in the Beta case, we use \( \theta = (\gamma, \delta)' \), \( p = 2 \), and in computing \( J_\varphi \) we need
\[ \frac{\partial u_i}{\partial \gamma} = \sum_{j=1}^{r} a_{ij} \frac{p_j I_1(y, \delta) - I_1(\lambda_j; y, \delta)}{f(\lambda_j)} \]

and

\[ \frac{\partial u_i}{\partial \delta} = \sum_{j=1}^{r} a_{ij} \frac{p_j I_2(y, \delta) - I_2(\lambda_j; y, \delta)}{f(\lambda_j)}, \]

where

\[ I_1(y; r, \delta) = [B(r, \delta)]^{-1} \frac{\partial}{\partial y} B(y; r, \delta), \]

\[ I_2(y; r, \delta) = [B(r, \delta)]^{-1} \frac{\partial}{\partial \delta} B(y; r, \delta), \]

\( B(y; r, \delta) \) is the incomplete beta integral, and \( I_k(y, \delta) = I_k(1, y, \delta) \).

Replacing the \( a_{ij} \) by \( b_{ij} \) yields \( \partial \nu_i / \partial \gamma \) and \( \partial \nu_i / \partial \delta \).

Ideally we would like to choose the \( p_j, a_{ij} \) and \( b_{ij} \) so as to optimize, in an appropriate sense, the covariance matrix \( \varphi \).

Unfortunately, there seems to be no way of analytically selecting optimal values for these parameters, though it would be relatively simple to compute \( \varphi \) for various sets of \( p_j, a_{ij} \) and \( b_{ij} \) at various values of \( (r, \delta) \), and then to determine which sets, if any, are "better" than others.

Preliminary work on the numerical aspects of this method of estimation indicates that, for the Beta case, graphical solution of \( \varphi_n = \varphi(\theta) \) is quite easy, and that relatively simple algebraic approximations to the curves involved may allow similar algebraic approximations to the roots \( \hat{r}, \hat{\delta} \).

Finally, the \( U_{in} \) and \( V_{in} \) may be replaced by statistics of the type
\[ T_n^{(1)} = \frac{1}{n} \sum_{j=1}^{n} J_i \left( \frac{j}{n+1} \right) h_i(Y_{jn}) + \sum_{j=1}^{r} a_{ij} h_i(Y_{jn}, n), \]

for which the results of Chernoff et al. [1967] also yield asymptotic normality of \( n^{1/2}(\tilde{\theta}_n - \theta) \), and the consequent asymptotic normality of \( n^{1/2}(\tilde{\theta}_n - \theta_0) \), under appropriate conditions on \( a_{ij}, J_i \) and \( h_i \).
9. Final Remarks

When an alternate parametrization in terms of the shape parameters \( \gamma \) and \( \delta \), and the midrange \( \tau \) and range \( \lambda \) is used, the properties of the mle's, moment estimators, and ratio estimators are easily found from the foregoing results and the fact that \( \tau = (a+b)/2 \) and \( \lambda = b-a \). That is \( (\tau, \lambda, \gamma, \delta)' \) is a linear transformation of \( (a, b, \gamma, \delta)' \).

The joint asymptotic distribution of estimators which require different normalizing sequences should be deducible from a knowledge of the asymptotic distributions of all linear combinations of the normalized estimators. These latter distributions should not be difficult to derive. The method, though without differing normalizations, has essentially been used in finding the asymptotic distribution of \( (\hat{\alpha}_n, \hat{\delta}_n)' \) when \( \gamma = \delta = 2 \) are known.
REFERENCES

Chernoff, Herman [1962]. Optimal accelerated life designs for estimation. Technometrics 4, 381-408.


