Weak Limit Theorems for Univariate k-Mean Clustering under a Nonregular Condition

REGIS J. SERINKO AND GUTTI JOGESH BABU*

Pennsylvania State University
Communicated by the Editors

A set of n points sampled from a common distribution F, is partitioned into k ≥ 2 groups that maximize the between group sum of squares. The asymptotic normality of the vector of probabilities of lying in each group and the vector of group means is known under the condition that a particular function, depending on F, has a nonsingular Hessian. This condition is not met by the double exponential distribution with k = 2. However, in this case it is shown that limiting distribution for the probability is \( b \, \text{sign}(W) \sqrt{|W|} \) and for the two means it is \( a_i \, \text{sign}(W) \sqrt{|W|} \), where \( W \sim N(0, 1) \) and \( b, a_1, \) and \( a_2 \) are constants. The rate of convergence is \( n^{1/4} \) and the joint asymptotic distribution for the two means is concentrated on the line \( x = y \).

A general theory is then developed for distributions with singular Hessians. It is shown that the projection of the probability vector onto some sequence of subspaces will have normal limiting distribution and that the rate of convergence is \( n^{1/2} \). Further, a sufficient condition is given to assure that the probability vector and vector of group means have limiting distributions, and the possible limiting distributions under this condition are characterized. The convergence is slower than \( n^{1/2} \).

1. Introduction

The clearest formulation of the k-mean clustering procedure is in terms of the minimization of the within group sum of squares. Let \( F \) be a probability distribution function on \( \mathbb{R} \) with a finite second moment. For \( k \geq 1 \) and \( a = (a_1, a_2, ..., a_k) \), let

\[
W(a) = \int \min_{1 \leq i \leq k} (x - a_i)^2 \, dF(x).
\]

Received November 28, 1990; revised September 27, 1991.
Key words and phrases: Bahadur's representation, singular Hessian, M-estimation, maximum likelihood, double exponential distribution.
* Research supported in part by NSA Grant MDA 904-90-H-1001.
Let $\mathbf{u} = (\mu_1, \mu_2, ..., \mu_k)$ be a vector, unique up to a permutation of indices, satisfying

$$W(\mathbf{u}) = \inf_{\mathbf{a}} W(\mathbf{a}).$$

For $k = 1$, $\mathbf{u} = \mu_1 = \int x \, dF(x)$ is the ordinary mean. For $k \geq 2$ one may use $\mathbf{u}$ to partition $\mathbb{R}$ into $k$ groups or clusters. The points in each group are those closer to some $\mu_i$ than to any other. In turn, $\mu_i$ is the conditional mean of the group. For $k > 2$, $\mathbf{u}$ is known as the cluster center vector. Denote the $i$th cluster by $C_i$ and let

$$p_i = \sum_{j=1}^{i} \int_{C_j} dF(x), \quad i = 1, 2, ..., k - 1.$$  

The split point vector is defined by $\mathbf{p} = (p_1, p_2, ..., p_{k-1})$.

Let $X_1, X_2, ..., X_n$ be a sample from $F$. The natural estimator of $\mathbf{u}$ is a vector $\mathbf{u}_n$ which minimizes

$$W_n(\mathbf{a}) = \frac{1}{n} \sum_{i=1}^{n} \min_{1 \leq j \leq k} (X_i - a_j)^2.$$  

Denote the points in the sample closer to $\mu_{in}$ than to any other component of $\mathbf{u}_n$ by $C_{in}$. One estimates $\mathbf{u}$ by $\mathbf{u}_n$ which has components

$$p_{in} = \frac{1}{n} \sum_{j=1}^{i} \sum_{m=1}^{n} I(\{X_m \in C_{jn}\}), \quad i = 1, 2, ..., k - 1,$$

where $I(A)$ is the indicator of the set $A$. This estimation procedure is called the method of $k$-mean clustering [4]. For the practical aspects of this estimation problem see Hartigan [5] and Jain and Dubes [7].

Hartigan [6] and Pollard [11] have given conditions on $F$ to assure the asymptotic normality of $n^{1/2}(\mathbf{p}_n - \mathbf{p})$ and $n^{1/2}(\mathbf{u}_n - \mathbf{u})$, respectively. (In addition, Hartigan [6] has shown that $\mathbf{p}_n$ is a weak consistent estimator of $\mathbf{p}$ and Pollard [9, 10] has established the strong consistency of $\mathbf{u}_n$. Pollard's [9–11] results hold for multivariate observations as well.) One of these conditions is that the Hessian (that is, the matrix of second-order partial derivatives) of $W(\mathbf{a})$ is nonsingular at $\mathbf{u}$.

The present study was motivated by the observation that, when $F$ is a double exponential distribution function and $k = 2$, the Hessian is singular while all other conditions of the weak limit theorems are met. This raises the natural question, do weak limits exist for the estimators in this case, and if so, what are they? The answer is yes and it is shown in Section 3 that the limiting distributions are non-Gaussian. Specifically,

$$n^{1/4}(\mathbf{p}_n - \mathbf{p}) \Rightarrow b \text{ sign}(W) \sqrt{|W|}$$
ASYMPTOTICS OF $k$-MEAN CLUSTERING

and

$$n^{1/4}(\mu_{jn} - \mu_j) \Rightarrow a_j \text{sign}(W) \sqrt{|W|}, \quad j = 1, 2,$$

where $W \sim N(0, 1)$ and $b, a_1,$ and $a_2$ are constants. Further, the joint asymptotic distribution of the standardized cluster centers is shown to concentrate on the line $x = y$ in the $x - y$ plane.

The next logical step is to take what is learned from the study of the double exponential distribution and develop a theory for distributions with singular Hessians and $k$ arbitrary but fixed. This is the second part of the work reported here. The asymptotic theory was found to be far richer in general than in the nonsingular case. If the Hessian is singular then at each point $t$, $\mathbb{R}^{k-1}$ splits into two orthogonal subspaces $W^{(s)}(t)$, which is one-dimensional for $t$ near $p$, and $W^{(f)}(t)$. It is found that the projection of $p_n - p$ into $W^{(f)}(p_*^*)$ (for some point $p_*^*$ near $p$) converges to a multivariate normal distribution at an $n^{1/2}$ rate. However, the weak limit of $p_n - p$ is governed by its projection into $W^{(s)}(p_*^*)$, which converges slower, if at all, to a limiting distribution. A sufficient condition is given to assure the existence of a real sequence $\{a_n\}$ and a random variable $X$ such that $a_n(p_n - p) \Rightarrow Xe^{(s)}(p)$ and $a_n(p_n - p) \Rightarrow XMe^{(f)}(p)$, where $e^{(s)}(p) \in W^{(s)}(p)$, $\|e^{(s)}(p)\| = 1$, and $M$ is a $k \times (k-1)$ matrix. In addition, all limiting distributions possible under the sufficient conditions are classified.

Like the $k$-mean procedure many estimation procedures including $M$-estimation are based on finding the extremum of some criterion function. The condition of a nonsingular Hessian of the population criterion function is used to assure asymptotic normality in the theory of these estimation procedures. For example, maximum likelihood estimation assumes that the information function is nonsingular. The present study provides a method of investigation in the absence of the nonsingularity condition and it illustrates the consequences of the failure of the condition.

The paper is organized as follows. Section 2 contains some preliminary results and assumptions used throughout the paper. The weak limit theorem for the double exponential with $k = 2$ is presented in Section 3. The limit theorem for arbitrary $k$ is given in Section 4. Appendix A contains the proofs of the preliminary lemmas presented in Section 2 and a proposition that is stated in Section 4. Finally, two variance-covariance matrices which appear in various result throughout the paper are given in Appendix B.

2. ASSUMPTIONS AND PRELIMINARY RESULTS

An alternate formulation based on the maximization of the between group sums of squares, rather than the minimization of the within group
sum of squares, is better suited to the problems considered here. To discuss
this formulation some notation is needed. For any distribution function \(G\),
define the quantile function by
\[
G^{-1}(p) = \inf\{x: G(x) \geq p\}, \quad 0 < p < 1.
\]  

Let \(V_k = \{t \in \mathbb{R}^{k-1} : 0 < t_1 < t_2 < \cdots < t_{k-2} < t_{k-1} < 1\}\), \(t_0 = 0\), and \(t_k = 1\).

Let \(F\) be a distribution function with finite first moment and let the
components of \(\mu(t, k) \in \mathbb{R}^k\) be given by
\[
\mu_j(t, k) = (t_j - t_{j-1})^{-1} \int_{t_{j-1}}^{t_j} F^{-1}(u) \, du, \quad t \in V_k, j = 1, 2, ..., k.
\]  

The function
\[
B(t, k) = \sum_{j=1}^{k} (t_j - t_{j-1})[\mu(t, k)]^2 - \mu^2
\]  
defined on \(V_k\) is called the split function, where \(\mu = \int_0^1 F^{-1}(u) \, du\). For \(t\) in
the boundary of \(V_k\), let \(B(t, k) = B(\nu, m)\), where \((m - 1)\) is the number of
distinct components of \(t\) which are different from zero or unity and \(\nu \in V_m\)
is the \((m - 1)\)-dimensional vector of these components. If none of the com-
oponents differ from zero or unity, then \(B(t, k)\) is taken as zero. If \(F\) has
finite second moment this definition of \(B\) on the boundary assures that it
is continuous on the closure of \(V_k\). Suppose that \(B(t, k)\) has unique maxi-
mum at \(p \in V_k\). If \(F\) has finite second moment, then it can be shown that
\(p\) and \(\mu(p, k)\) are the split point vector and cluster center vector,
which were defined in the Introduction. Let \(F_n\) be the empirical distribution function of
a sample of size \(n\) from \(F\). Define \(\mu_n(t, k) \in \mathbb{R}^k\) with components
\[
\mu_{jn}(t, k) = (t_j - t_{j-1})^{-1} \int_{t_{j-1}}^{t_j} F_n^{-1}(u) \, du, \quad t \in V_k, j = 1, 2, ..., k,
\]  
and set \(\mu_n = \int_0^1 F_n^{-1}(u) \, du\). The sample split point function is defined by
\[
B_n(t, k) = \sum_{j=1}^{k} (t_j - t_{j-1})[\mu_n(t, k)]^2 - \mu_n^2, \quad t \in V_k, n \geq k.
\]  
A point \(p_n \in V_k\) which maximizes \(B_n\) is taken as an estimator \(p\); it is called
the sample split point vector. The sample cluster center vector \(\mu_n(p_n, k)\) is
the estimator of \(\mu(p, k)\). These are the \(k\)-mean clustering estimators which
appear in the Introduction.

The notation is somewhat simplified by suppressing the dependence of
various quantities on \(k\). No confusion should result since \(k\) will always be
fixed.
Both $F$ and $F_n$ are left continuous functions; hence the directional derivatives of $\mu$ and $\mu_n$ exist. Consequently, the directional derivatives of both $B$ and $B_n$ exist. $B_j^{(+)}$ and $B_n^{(+)}$ will denote the directional derivatives in the direction of $\pm e_j$ of $B$ and $B_n$, respectively, where $e_j \in \mathbb{R}^{k-1}$ and it has all zero entries save the $j$th, which is unity. The directional derivatives are given by

$$B_j^{(\pm)}(t) = [\mu_{j+1}(t) - \mu_j(t)] \{\mu_{j+1}(t) + \mu_j(t) - 2F^{-1}(t_j^{\pm})\},$$

where $F^{-1}(t_j^{\pm}) = \lim_{\varepsilon \to 0} F^{-1}(t_j \pm \varepsilon)$. An expression for $B_n^{(\pm)}$ is obtained by affixing the subscript $n$ to all terms in (5). The corresponding vectors of directional derivatives are denoted by $B_n^{(\pm)}$ and $B_n^{(+)}$. (Hartigan [6] uses the notation $dB^{\pm}/dp$ for the vector of directional derivatives.) If the components of $t$ are continuity points of $F^{-1}$, then $B^{(1)}(t)$, the vector of first partial derivatives at $t$ exists. Further, the Hessian $B^{(2)}$ of $B$ exists at $t$ whenever $F^{-1}$ has derivatives at the components of $t$. In this formulation the condition that the Hessian of $W(a)$ is nonsingular at $\mu$ is replaced with the condition that $B^{(2)}(\mu)$ is nonsingular.

The assumptions needed in this paper are listed here for convenience:

(H1) $F$ is a distribution function with a finite second moment.

(H2) $F$ is a distribution function which gives rise to a split function (3) with a unique maximum at $p \in V_k$.

(H3) The derivative $f$ of $F$ exists and is continuous in some open neighborhood of $\bar{\mu}_j = F^{-1}(\rho_j)$ and $f(\bar{\mu}_j) > 0$, $j = 1, 2, ..., k-1$.

The following, two results which are used in the proofs of the main results, are proven in Appendix A.

**Lemma 1.** Suppose that $F$ is a distribution function which satisfies (H1)--(H3), then

$$(\mu_n(p_n) - \mu(p)) = (\mu_n(p) - \mu(p)) + o_p(n^{-1/2})$$

and, further,

$$n^{-1/2}(\mu_n(p) - \mu(p)) \Rightarrow MVN_k(0, \Sigma^{(0)}).$$

**Lemma 2.** Under the conditions of Lemma 1,

$$B^{(1)}(p_n) = -Z_n + o_p(n^{-1/2}),$$

where

$$n^{1/2}Z_n \Rightarrow MVN_{k-1}(0, \Sigma).$$
The form of the variance–covariance matrices, $\Sigma^{(0)}$ and $\Sigma$, are given in Appendix B. It should be noted that whenever $B^{(2)}(p)$ is nonsingular the mean value theorem and Lemma 2 lead to Hartigan's [6] central limit theorem (CLT) for $n^{1/2}(p_n - p)$.

3. The Double Exponential Distribution with $k = 2$

Throughout this section $k$ is assumed to be 2. The double exponential distribution is an example of a distribution which satisfies (H1)–(H3), but it does not satisfy the condition $B^{(2)}(p) \neq 0$ when $k = 2$. This distribution has p.d.f.,

$$f(x) = \frac{1}{2} \beta^{-1} \exp[-\beta |x|], \quad x \in \mathbb{R}, \beta > 0,$$

and quantile function

$$F^{-1}(t) = \begin{cases} \beta \log(2t), & 0 < t < \frac{1}{2} \\ -\beta \log(2(t - 1)), & \frac{1}{2} \leq t < 1, \end{cases}$$

where $\tau = 1 - t$. Without loss of generality take $\beta = 1$. Clearly, (H1) holds. Condition (H3) also holds, since $F^{-1}$ has continuous positive derivatives at all points of $(0, 1)$. A straightforward calculation gives

$$B^{(1)}(t) = \begin{cases} ([(\log(2t) - t)/t]^2 - 1], & 0 < t < \frac{1}{2} \\ ([(\log(2\tau - \tau)/t]^2 - 1], & \frac{1}{2} \leq t < 1, \end{cases}$$

which is seen to vanish if and only if $t = \frac{1}{2}$. Hence (H2) is seen to hold with $p = \frac{1}{2}$. Finally, simple algebra leads to

$$B^{(1)}(t) = -8[(t - \frac{1}{2})^2 \text{sign}(t - \frac{1}{2})](1 + o(1)),$$

as $t \to \frac{1}{2}$. From Eq. (9) one may conclude that $B^{(2)}(\frac{1}{2}) = 0$. Consequently, one of the main assumptions of Hartigan [6] and Pollard [10, 11] is violated. The following theorem gives the weak limit in this case.

**Theorem 1.** Suppose $X_1, X_2, ..., X_n$ are i.i.d. from a double exponential distribution with $\beta = 1$, then

$$n^{1/4}(p_n - p) \Rightarrow 2^{-1/2} \text{sign}(W) \sqrt{|W|}$$

$$n^{1/4}(\mu_{1n}(p_n) - \mu_1(p)) \Rightarrow \sqrt{2} \text{sign}(W) \sqrt{|W|}$$

$$n^{1/4}(\mu_{2n}(p_n) - \mu_2(p)) \Rightarrow \sqrt{2} \text{sign}(W) \sqrt{|W|},$$

where $W \sim N(0, 1)$. 
Remark 1. It is clear from the proof of this theorem (see Eq. (10)) that
\[ n^{1/4}[(\mu_{1,n}(p_n) - \mu_1(p)) - (\mu_{2,n}(p_n) - \mu_2(p))] \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty. \]
Hence the joint asymptotic distribution of \( n^{1/4}(\mu_{1,n}(p_n) - \mu_1(p)) \) and \( n^{1/4}(\mu_{2,n}(p_n) - \mu_2(p)) \) is concentrated on the line \( x = y \) in the \( x - y \) plane.

Proof. As noted earlier the double exponential distribution function satisfies (H1)--(H3). Hence Lemma 2 is applicable, which along with (9) yields
\[ -8[(p_n - \frac{1}{2})^2 \text{sign}(p_n - \frac{1}{2})](1 + o_p(1)) = -Z_n + o_p(n^{-1/2}), \]
where \( n^{1/2}Z_n \Rightarrow 4W \) with \( W \sim N(0, 1) \). The above expression is inverted and the mapping theorem (see Theorem 25.7 of [2, p. 343]) gives
\[ n^{1/4}(p_n - p) \Rightarrow 2^{-1/2} \text{sign}(W) \sqrt{|W|}. \]

It remains to prove the limit theorem for the cluster centers. The proof for \( j = 1 \) is presented. The proof for \( j = 2 \) is identical and is therefore omitted. By Lemma 1 one may write
\[ \mu_{1,n}(p_n) - \mu_1(p) = \mu_1(p_n) - \mu_1(p) + [\mu_{1,n}(p) - \mu_1(p)] + o_p(n^{-1/2}). \]
By the second part of Lemma 1, \([\mu_{1,n}(p) - \mu_1(p)] = O_p(n^{-1/2})\), while the mean value theorem and \( F^{-1}(p) = \frac{1}{2}(\mu_2(p) + \mu_1(p)) \) give
\[ \mu_1(p_n) - \mu_1(p) = \frac{\mu_2(p) - \mu_1(p)}{2p} (p_n - p)(1 + o_p(1)). \]
This yields
\[ n^{1/4}(\mu_{1,n}(p_n) - \mu_1(p)) \Rightarrow \frac{\mu_2(p) - \mu_1(p)}{2p} 2^{-1/2} \text{sign}(W) \sqrt{|W|}. \]
\( \mu_2(p) = -\mu_1(p) = 1 \) and \( p = \frac{1}{2} \) is substituted to complete the proof.

3. Weak Limit Theorems for \( k \geq 2 \)

In this section the problem of a singular Hessian for arbitrary but fixed \( k \geq 2 \) is taken up. In the last section it was seen that the limiting distribution of the estimators is determined by the limiting behavior of \( B^{(1)}(t) \) as \( t \to p \). When \( k > 2 \), if \( B^{(2)}(p) \) is singular, then different components of the vector \( B^{(1)}(t) \) can go to zero at different rates. Therefore the first point for study is the behavior of \( B^{(1)}(t) \) at \( p \).
Let $\mathbf{W}^{(s)}(t)$ denote the linear span of the set

$$\{ \mathbf{x} \in \mathbb{R}^{k-1} : \mathbf{x}^T \mathbf{B}^{(2)}(t) = \mathbf{x}^T \lambda_1(t) \},$$

where $\lambda_1(t)$ is the largest eigenvalue of $\mathbf{B}^{(2)}(t)$. $\mathbf{W}^{(f)}(t)$ will denote the orthogonal complement of $\mathbf{W}^{(s)}(t)$. Since $\mathbf{B}$ is a maximum at $\mathbf{p}$ the eigenvalues of $\mathbf{B}^{(2)}(\mathbf{p})$ will be nonpositive. Therefore, if $\text{Det} \mathbf{B}^{(2)}(\mathbf{p}) = 0$, the largest eigenvalue is zero. Roughly, as will be seen, whenever $\mathbf{B}^{(2)}(\mathbf{p})$ is singular the projection of $\mathbf{p}_n - \mathbf{p}$ onto $\mathbf{W}^{(s)}(t)$ will converge more slowly than its projection onto $\mathbf{W}^{(f)}(t)$. Hence, $\mathbf{W}^{(s)}(t)$ and $\mathbf{W}^{(f)}(t)$ will be called the slow and fast subspaces at $t$. It is this difference in convergence rates which produce a weak limit which is not multivariate normal. The projection matrices onto $\mathbf{W}^{(s)}(t)$ and $\mathbf{W}^{(f)}(t)$ will be denoted by $\mathbf{P}^{(s)}(t)$ and $\mathbf{P}^{(f)}(t)$, respectively. The Moore–Penrose inverse of a matrix $\mathbf{A}$ is denoted by $\mathbf{A}^+$ ([see [12, p. 253]. The following proposition, which is proven in Appendix A, gives the important properties of the projection matrices and subspaces.

**Proposition 1.** Suppose that $F$ satisfies (H1) and (H3) and that $\mathbf{B}^{(2)}(\mathbf{p})$ is singular; then the following hold:

(i) the dimension of $\mathbf{W}^{(s)}(\mathbf{p})$ is 1;

(ii) $\lim_{t \to \mathbf{p}} \mathbf{P}^{(s)}(t) = \mathbf{P}^{(s)}(\mathbf{p})$;

(iii) $\mathbf{W}^{(s)}(t)$ has dimension 1 for $t$ sufficiently near $\mathbf{p}$.

Further, for $t \neq \mathbf{p}$ in some neighborhood of $\mathbf{p}$, $[\mathbf{B}^{(2)}(t)]^{-1}$ exists and

(iv) $\lim_{t \to \mathbf{p}} \mathbf{P}^{(s)}(t)[\mathbf{B}^{(2)}(t)]^{-1} = [\mathbf{B}^{(2)}(\mathbf{p})]^{+}$.

Consequently, there is a single direction in $\mathbb{R}^{k-1}$ along which $\mathbf{B}^{(1)}(t)$ goes to zero at a different rate than in other directions, whenever $\mathbf{B}^{(2)}(\mathbf{p})$ is singular. This fact simplifies the analysis of the problem greatly.

Under condition (H3), the mean value theorem gives, for $t$ in some neighborhood of $\mathbf{p}$,

$$\mathbf{B}^{(1)}(t) = \mathbf{B}^{(2)}(t^*)(t - \mathbf{p}),$$

where each component of $t^*$ lies between the corresponding components of $t$ and $\mathbf{p}$. For convenience let $\pi_n = \mathbf{p}_n - \mathbf{p}$, $\pi^{(s)}_n = \mathbf{P}^{(s)}(\mathbf{p}_n - \mathbf{p})$, and $\pi^{(f)}_n = \mathbf{P}^{(f)}(\mathbf{p}_n^*)(\mathbf{p}_n - \mathbf{p})$. Further, let $\tau = t - \mathbf{p}$, $\tau^{(s)}(t') = \mathbf{P}^{(s)}(t')\tau$, and $\tau^{(f)}(t') = \mathbf{P}^{(f)}(t')$; for $t, t' \in \mathbb{R}^{k-1}$. Finally, by Proposition 1(iii), $\mathbf{W}^{(s)}(t)$ is one-dimensional for $t$ sufficiently near $\mathbf{p}$. Hence one may write for any $t' \in \mathbf{W}^{(s)}(t)$, $t' = t'e^{(s)}(t)$, where $e^{(s)}(t)$ is a unit vector in $\mathbf{W}^{(s)}(t)$, and $t' = e^{(s)}(t)^T t'$. The following is an intermediate weak limit result.
Lemma 3. Suppose that $F$ satisfies (H1) through (H3) and that $\text{Det } B^{(2)}(p) = 0$, then

$$n^{1/2}(p_n) \Rightarrow MVN_{k-1}(0, \Sigma^{(f)}),$$

where

$$\Sigma^{(f)} = [B^{(2)}(p)]^+ \Sigma [B^{(2)}(p)]^+;$$

$$n^{1/2}\pi_n^{(s)} \Rightarrow MVN_{k-1}(0, \Sigma^{(s)}),$$

where

$$\Sigma^{(s)} = P^{(s)}(p) B^{(2)}(p) P^{(s)}(p),$$

as $n \to \infty$.

Remark 2. The continuity of $B^{(2)}(t)$ in a neighborhood of $p$, the consistency of $p_n$, and the fact that $\lambda_1(p) = 0$, together with the second part of Lemma 3, imply that $\pi_n^{(s)}$ converges to zero at a rate which is slower than $n^{1/2}$. Since, $p_n - p = \pi_n^{(f)} + \pi_n^{(s)}$, the slower convergence of $\pi_n^{(s)}$ in turn implies the slower convergence of $p_n - p$. Hence, a necessary and sufficient condition for the existence of a sequence of real numbers $a_n$, converging to infinity, and a random vector $X$ such that $a_n(p_n - p) \Rightarrow X$ as $n \to \infty$, is that $a_n \pi_n^{(s)} \Rightarrow X$ as $n \to \infty$.

Proof of Lemma 3. The assumptions (H1) through (H3) allow one to invoke Lemma 2 and conclude

$$B^{(1)}(p_n) = -Z_n + o_p(n^{-1/2}), \quad (12)$$

where $n^{1/2}Z_n \Rightarrow MVN_{k-1}(0, \Sigma)$ as $n \to \infty$. For $n$ sufficiently large, assumption (H3), the weak consistency of $p_n$ (Theorem 3), and the mean value theorem (Eq. (11)) together yield

$$B^{(1)}(p_n) = B^{(2)}(p^*_n)(p_n - p) = B^{(2)}(p^*_n) \pi_n. \quad (13)$$

Equations (12) and (13) are combined to give

$$B^{(2)}(p^*_n) \pi_n = -Z_n + o_p(n^{-1/2}). \quad (14)$$

If $p^*_n \neq p$, then $P^{(f)}(p^*_n)[B^{(2)}(p_n^*)]^{-1}$ exists. On the other hand, by Proposition 1(iv) $P^{(f)}(t)[B^{(2)}(t)]^{-1}$ goes to $[B^{(2)}(p)]^+$ as $t$ goes to $p$. Therefore one may define $P^{(f)}(p)[B^{(2)}(p)]^{-1}$ to be $[B^{(2)}(p)]^+$. Both sides of (14) are multiplied by $P^{(f)}(p^*_n)[B^{(2)}(p^*_n)]$, to yield

$$\pi_n^{(f)} = -P^{(f)}(p^*_n)[B^{(2)}(p^*_n)]^{-1} Z_n + o_p(n^{-1/2}),$$
which along with Proposition 1(iv) and the consistency of \( p_n^{*} \) gives

\[
n^{1/2} \pi_n^{(f)} \Rightarrow MVN_{k-1}(0, \Sigma^{(f)}),
\]
as \( n \to \infty \). This completes the first part of the proof.

Note that \( P'(p,*) B_c^{(2)}(p,*) \pi_n = \lambda_1(p_n^{*}) \pi_n^{(s)} \) which combines with Eq. (14) to give

\[
\lambda_1(p_n^{*}) \pi_n^{(s)} = -P'(p,*) Z_n + o_p(n^{-1/2}).
\]

From this, Eq. (12), Proposition 1(ii), and consistency of \( p_n^{*} \), one concludes that

\[
n^{1/2} \lambda_1(p_n^{*}) \pi_n^{(s)} \Rightarrow MVN_{k-1}(0, \Sigma^{(s)}),
\]
as \( n \to \infty \). This completes the proof.

A consequence of Lemma 3 and Remark 2 is that the way in which \( P'(p) B^{(2)}(t) \) goes to the null vector at \( p \) will determine the existence of normalizing constants \( a_n \), such that \( a_n(p_n - p) \) has a weak limit. A new function is defined which aids in the study of the behavior of \( P'(p) B^{(1)}(t) \) at \( p \). Under the conditions of Proposition 1 there exists a neighborhood \( U \) of \( p \) such that \( W^{(s)}(t) \) is one-dimensional if \( t \in U \). Let \( W^{(s)} = \bigcup_{t \in U} W^{(s)}(t) \) and define a function \( g \) on \( W^{(s)} \) by

\[
g(x e^{(s)}(t)) = -e^{(s)}(p)^T B^{(1)}(p + x e^{(s)}(t)), \quad (15)
\]
where \( x \) is a sufficiently small real. Important properties of \( g \) are given in the next proposition.

**Proposition 2.** Suppose that \( F \) is a distribution function which satisfies (H1) through (H3) and that \( \det B^{(2)}(p) = 0 \) Then, for each \( t \in U \), the function \( g \) defined in (15) satisfies properties

(i) \( g(0) = 0 \)

(ii) \( \pm g(x e^{(s)}(t)) > 0 \) if \( \pm x > 0 \)

(iii) \( g(x e^{(s)}(t)) = o(x) \)

and, for \( t' \) sufficiently near \( p \),

\[
P'(p) B^{(1)}(p) B^{(1)}(t) + \tau^{(f)}(t) + p) = -g(\tau^{(s)}(t) e^{(s)}(t)) e^{(s)}(p) + P'(p) B^{(2)}(\tau^{(s)}(t) + \tau^{(f)}(t) + p) \tau^{(f)}(t).
\]

The proof is straightforward and therefore only outlined. Properties (i) and (ii) are consequences of the fact that, under (H1), \( B \) has a unique maximum at \( p \), and property (iii) follows from the singularity of \( B^{(2)}(p) \).
proof of Eq. (16) makes use of a Taylor series expansion of $\mathbf{B}^{(1)}(\tau^{(t)}(t) + \tau^{(t)}(t) + p)$ about $\tau^{(t)}(t) + p$. An immediate consequence of Proposition 2 is the following corollary.

**Corollary 1.** Under the conditions of Proposition 2,

$$n^{1/2}g(\pi_n^{(s)}e^{(s)}(p_n)) \Rightarrow N(0, \sigma^2)$$

as $n \to \infty$, where $\sigma^2 = e^{(s)}(p)^T \Sigma^{(s)}e^{(s)}(p)$.

The proof, which is once more straightforward and omitted, makes use of Proposition 2 and Lemmas 2 and 3.

From Corollary 1 it is seen that $g$ must be fairly well-behaved at the origin if $\pi_n^{(s)}$ is to be well-behaved. The following condition on $g$ is sufficient to assure the existence of normalizing constants $\{a_n\}$ such that $a_n \pi_n^{(s)}$ will have a weak limit.

**H4** $F$ is a distribution function, with $\text{Det} \mathbf{B}^{(2)}(p) = 0$, for which there exists an increasing real-valued function $r$, defined on a neighborhood $V$ of the origin of $\mathbb{R}$, with the following properties:

(a) $\lim_{x \to 0, t \to p = 0(x)} g(xe^{(s)}(t))/r(x) = 1$;

(b) $\rho_\pm(a) = \lim_{x \to 0} r(ax)/r(x)$ exists for $a$ in a dense set $A \subseteq [0, \infty)$.

**Remark 3.** If $B$ has a nonvanishing continuous fourth derivative in a neighborhood of $p$, then (H4) is satisfied with $r(x) = Cx^3$, where $C$ is a nonzero constant.

The following proposition contains two immediate consequences of (H4) which are used in the proof of the next weak limit theorem.

**Proposition 3.** (a) One has either

(i) $\rho_\pm(a) = a^{\lambda_\pm}$, $1 \leq \lambda_\pm < \infty$,

or

(ii) $\rho_\pm(a) = \begin{cases} 0 & a < 1 \\ 1 & a = 1 \\ \infty & a > 1. \end{cases}$

(b) If $\lim_{n \to \infty} b_n = b > 0$ and $x_n^\pm \to 0^\pm$ as $n \to \infty$ then

$$\lim_{n \to \infty} \frac{r(b_n x_n^\pm)}{r(x_n^\pm)} = \rho_\pm(b).$$
whenever \((a)(i)\) holds, and

\[
\lim_{n \to \infty} \frac{r(b_n x_n^+)}{r(x_n^+)} = \begin{cases} 
0 & b < 1 \\
1 & b > 1 
\end{cases}
\]

otherwise.

The proof of Proposition 3(a), which makes use of \((H4)(b)\), may be found in Feller [3, p. 275]). The proof of part (b) makes uses of a standard \(\text{lim inf-}\text{lim sup}\) argument. Both proofs are omitted.

For a random variable \(Y\), let \(Y^\pm = \max\{0, \pm Y\}\). The second and final major result can now be stated.

**Theorem 2.** Suppose that \(F\) is a distribution function which satisfies \((H1)\) through \((H4)\). Then there exists a real sequence \(\{a_n\}\) with \(\lim_{n \to \infty} a_n = \infty\) and a random variable \(X\) such that

\[
a_n(p_n - p) \Rightarrow X e^{(p)}(p) \quad \text{as} \quad n \to \infty.
\]

Further, \(X\) is given by one of

\[
X = (Z^n)^\beta - c(Z^-)^\beta, \\
X = (Z^+)\beta, \quad X = -(Z^-)^\beta, \\
X = \begin{cases} 
1 & \text{w.p. } \frac{1}{2} \\
0 & \text{w.p. } \frac{1}{2}
\end{cases}, \quad X = \begin{cases} 
-1 & \text{w.p. } \frac{1}{2} \\
0 & \text{w.p. } \frac{1}{2}
\end{cases},
\]

or

\[
X = \begin{cases} 
1 & \text{w.p. } \frac{1}{2} \\
-c & \text{w.p. } \frac{1}{2},
\end{cases}
\]

where \(0 < \beta \leq 1, 0 < c < \infty,\) and \(Z \sim N(0, \sigma^2)\) with \(\sigma^2 = e^{(p)}(p)^T \Sigma e^{(p)}(p)\).

**Proof.** As noted in Remark 2, it suffices to show the existence of a real sequence \(\{a_n\}\) and a random vector \(X\) such that \(a_n \pi_n^{(e)} \Rightarrow X\) as \(n \to \infty\), or equivalently, since the slow subspaces are one-dimensional, \(a_n \pi_n^{(e)} \Rightarrow X\) as \(n \to \infty\), where \(X = X e^{(p)}(p)\). Corollary 1 gives

\[
n^{1/2} g(\pi_n^{(e)}(p_n^*) \Rightarrow N(0, \sigma^2) \quad \text{as} \quad n \to \infty. \quad (17a)
\]

According to Remark 2, \(p_n^* - p = O(\pi_n^{(e)}).\) Therefore condition \((H4)(a)\) can be used to rewrite \((17a)\) as

\[
n^{1/2} r(\pi_n^{(e)}) \Rightarrow N(0, \sigma^2) \quad \text{as} \quad n \to \infty. \quad (17b)
\]
By Skorohod's theorem (see Theorem 25.6 of [2, p. 343]) there exists a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) on which there is defined a sequence of random variables \(\{V_n\}\), with the distribution of \(n^{1/2}r(\pi_n^{(s)})\), and a random variable \(V\), which is distributed as \(N(0, \sigma^2)\), such that \(\lim_{n \to \infty} V_n = V\) a.s. Since \(r\) is an increasing bounded function it has a left continuous inverse \(r^{-1}\). Let \(X_n = r^{-1}(V_n/n^{1/2})\), then \(X_n\) has the distribution of \(\pi_n^{(s)}\). It is shown below that there exist a real sequence \(\{a_n\}\) and a random variable \(X\) such that \(\lim_{n \to \infty} a_n X_n = X\) a.s. This implies that \(a_n \pi_n^{(s)} \Rightarrow X\).

The sequence \(\{a_n\}\) is constructed from the two sequences

\[
a_n^+ = [r^{-1}(n^{-1/2})]^{-1}
\]

and

\[
a_n^- = [-r^{-1}(-n^{-1/2})]^{-1}.
\]

The definitions of \(\{X_n\}\), \(V\), and \(a_n^\pm\), lead to

\[
\lim_{n \to \infty} r(a_n^+ X_n^+/a_n^+) / r(1/a_n) = V^+ \quad \text{a.s.} \tag{18a}
\]

and

\[
\lim_{n \to \infty} r(a_n^- X_n^- / -a_n^-) / r(-1/a_n^-) = V^- \quad \text{a.s.} \tag{18b}
\]

Assumption (H4)(b) and Eq. (18) are used to argue that both \(a_n^+ X_n^+\) and \(a_n^- X_n^-\) have an almost sure limit as \(n \to \infty\). To this end, let

\[
X^+ = \lim \inf_{n \to \infty} a_n^+ X_n^+,
\]

\[
X^- = \lim \inf_{n \to \infty} a_n^- X_n^-,
\]

\[
\bar{X}^+ = \lim \sup_{n \to \infty} a_n^+ X_n^+,
\]

\[
\bar{X}^- = \lim \sup_{n \to \infty} a_n^- X_n^-.
\]

(The reader should take care not to confuse \(X^\pm\) and \(\bar{X}^\pm\) with similar notation used for different objects by Hartigan [6].)

It is shown that \(X^\pm = \bar{X}^\pm\) a.s. For this purpose, one may find subsequences \(\{n_+(k)\}\), \(\{n_-(k)\}\), \(\{\bar{n}_+(k)\}\), and \(\{\bar{n}_-(k)\}\) such that

\[
\lim_{k \to \infty} a_{n_+(k)}^+ \bar{X}^+_{n_+(k)} = X^+; \tag{19a}
\]

\[
\lim_{k \to \infty} a_{n_-(k)}^- \bar{X}^-_{n_-(k)} = X^-; \tag{19b}
\]
To proceed from here two cases are distinguished.

**Case I.** Both $\rho_+(a)$ and $\rho_-(a)$ are finite for finite $a$. In this case, (18) and (19) give

$$V^{\pm} = \rho_{\pm}(X^{\pm}) = (X^{\pm})^{\beta_{\pm}} \text{ a.s.} \quad (20a)$$

and

$$V^{\pm} = \rho_{\pm}(\bar{X}^{\pm}) = (\bar{X}^{\pm})^{\beta_{\pm}} \text{ a.s.} \quad (20b)$$

This implies that $X^{\pm} = \bar{X}^{\pm} \text{ a.s. and}$

$$X^{\pm} = (V^{\pm})^{\beta_{\pm}} \text{ a.s.,}$$

where $0 < \beta_{\pm} \leq 1$.

**Case II.** Either $\rho_+(a) = \infty$ or $\rho_-(a) = \infty$ whenever $a > 1$. For definiteness, suppose that $\rho_+(a) = \infty$ whenever $a > 1$. In this case one cannot immediately conclude the equalities $V^{\pm} = \rho_{\pm}(X^{\pm})$ and $V^{\pm} = \rho_{\pm}(\bar{X}^{\pm})$ in (20). Define the sets

$$C_{+} = \{ \pm V > 0 \} \quad \text{and} \quad C_{\infty} = \{ |V| = \infty \},$$

and recall that $A^c$ is the set on which $X_n$ converges to 0 as $n \to \infty$. The limits (18) and (19) imply

$$0 < \lim_{k \to \infty} r(a_{\eta_{\pm}(k)}^+ X_{\eta_{\pm}(k)}^+(\omega)/a_{\eta_{\pm}(k)}^+) / r(1/a_{\eta_{\pm}(k)}^+) < \infty, \quad \omega \in A^c \cap C_+ \cap C_{\infty}^c,$$

and

$$0 < \lim_{k \to \infty} r(a_{\eta_{\pm}(k)}^+ X_{\eta_{\pm}(k)}^+(\omega)/a_{\eta_{\pm}(k)}^+) / r(1/a_{\eta_{\pm}(k)}^+) c < \infty, \quad \omega \in A^c \cap C_+ \cap C_{\infty}^c.$$
The same argument is used if \( p_+(a) \) is infinite for \( a > 1 \), and one may conclude that, \( \lim_{n \to \infty} a_n^- X_n^- = X \) with probability 1, where \( P(X = -1) = P(X = 0) = \frac{1}{2} \).

The final step in the proof is the construction of \( a_n \) and the classification of the limiting distributions. First suppose that \( \lim_{n \to \infty} a_n^+/a_n^- = c \) with \( 0 < c < \infty \); in this case take \( a_n = a_n^+ \). Under Case I, one has \( 0 < \beta_+ = \beta_- < \infty \), and \( \lim_{n \to \infty} a_n X_n = X \) with probability 1, where

\[
X = (V^+)^\beta - c(V^-)^\beta.
\]

On the other hand, under Case II, \( \lim_{n \to \infty} a_n X_n = X \) with probability 1, where \( P(X = 1) = P(X = -c) = \frac{1}{2} \).

Next suppose that \( a_n^\pm = o(a_n^\pm) \), where this notation means either \( a_n^+ = o(a_n^-) \) or \( a_n^- = o(a_n^+) \). In this case take \( a_n = a_n^\pm \). If \( p_\pm(a) \) is finite for finite \( a \), define \( \beta = \beta_\pm \), then \( 0 < \beta \leq 1 \) and \( \lim_{n \to \infty} a_n X_n = X \) with probability 1, where \( X = \pm (V^\pm)^\beta \). If \( p_\pm(a) \) is infinite for \( a > 1 \), then \( \lim_{n \to \infty} a_n X_n = X \) with probability 1, where \( P(X = \pm 1) = P(X = 0) = \frac{1}{2} \). This completes the proof.

The weak limit of the cluster center vector is given in the following corollary.

**Corollary 2.** Under the conditions of Theorem 2,

\[
a_n(\mu_n(p_n) - \mu(p)) \Rightarrow X\text{Me}^{(\epsilon)}(p),
\]

where \( \text{Me} \) is a \( k \times (k - 1) \) matrix, with elements

\[
(M)_{j} = \frac{\mu_{j+1}(p) - \mu_{j}(p)}{2(p_{j} - p_{j-1})}; \quad j = 1, 2, \ldots, k - 1;
\]

\[
(M)_{j-1} = \frac{\mu_{j}(p) - \mu_{j-1}(p)}{2(p_{j} - p_{j-1})}; \quad j = 2, 3, \ldots, k;
\]

and all the other elements vanish.

**Proof.** By Lemma 1,

\[
\mu_n(p_n) - \mu(p) = [\mu(p_n) - \mu(p)] + [\mu_n(p) - \mu(p)] + o_p(n^{-1/2}).
\]

By the second part of Lemma 1 and Remark 2,

\[
[\mu_n(p) - \mu(p)] = O_p(n^{-1/2}) - o_p(a_n^{-1}).
\]
On the other hand, the delta-method (see Problem 27.10 of [2, p. 380]) gives

\[ a_n[\mu(p_n) - \mu(p)] \Rightarrow X \text{Me}^{(s)}(p), \]

where \( M \) is the matrix of first partial derivatives of \( \mu(t) \) evaluated at \( p \). This leads to

\[ a_n(\mu_n(p_n) - \mu(p)) \Rightarrow X \text{Me}^{(s)}(p), \]

which completes the proof.

The proof of the limit theorem for \((p_n - p)\) makes intimate use of (H4) through Proposition 3 and it is difficult to see how to avoid this assumption. On the other hand, if one were able to deduce (H4) from the other assumptions, the theorem would be strengthened. One might hope that the properties of \( g \) given in Proposition 2 would be sufficient for (H4). The following is a counterexample to this hope.

**EXAMPLE 1.** By symmetry it suffices to construct a function defined only for positive \( x \). Let

\[ g_0(x) = \int_{0}^{x} h(v) \, dv, \quad x \in [0, 1], \]

where

\[ h(v) = \begin{cases} 
3v - (1/2)^{2k} & v \in ((1/2)^{2k+1}, (1/2)^{2k}], \\
(1/2)^{2k+1} & v \in ((1/2)^{2(k+1)}, (1/2)^{2k+1}],
\end{cases} \quad k = 0, 1, 2, ... 
\]

Define \( g \) by

\[ g(xe^{(t)}(v)) = g_0(xe^{(t)}(v) \cdot e^{(s)}(p)). \]

Clearly, \( g \) satisfies the three properties given in Proposition 2 and it is monotonic in \( x \) for fixed \( t \). If one takes \( t = p \), then (H4)(a) is satisfied with \( r(x) = g_0(x) \). However, as is now seen, it does not satisfy (H4)(b). By Proposition 3(i) it suffices to show that the limit in (H4)(b) does not hold for some \( a \). Consider the two sequences \( x_k = (1/2)^{2k} \) and \( x'_k = (1/2)^{2k+1}, k = 0, 1, 2, ... \). Let \( a = 2 \), then \( \lim_{k \to \infty} g_0(ax_k)/g_0(x_k) = \frac{3}{5} \) and \( \lim_{k \to \infty} g_0(ax'_k)/g_0(x'_k) = \frac{7}{2} \). Hence (H4)(b) does not hold for this \( g \).

This section is capped off by exhibiting a distribution function which satisfies the conditions of Theorem 3 with \( k = 3 \).
EXAMPLE 2. Define a distribution function $F$ by

$$
F(x) = \begin{cases}
0, & x < -a \\
p_1(p_2 - p_1)[a^2 - x^2], & -a \leq x < 0 \\
p_1(p_2 - p_1)[a^2 + x^2] + (p_2 - p_1)^2, & 0 \leq x \leq a \\
1, & a < x,
\end{cases}
$$

where $p_1 = \frac{1}{2}(1 - 1/\gamma)$, $p_2 = \frac{1}{2}(1 + 1/\gamma)$, $\gamma = a^2 - 1 = \sqrt{3} + \frac{3}{4}$. The quantile function for this distribution is

$$
F^{-1}(t) = \begin{cases}
\left( \frac{2p_1 p_2 - t}{p_1(p_2 - p_1)} \right)^{1/2}, & 0 \leq t \leq 2p_1 p_2 \\
0, & 2p_1 p_2 < t < p_1^2 + p_2^2 \\
\left( \frac{t - (p_1^2 + p_2^2)}{p_1(p_2 - p_1)} \right)^{1/2}, & p_1^2 + p_2^2 \leq t \leq 1.
\end{cases}
$$

Clearly, $F$ has finite second moment, i.e. it satisfies (H1). Further, the split function has a unique maximum whenever $k = 3$ at $p = (p_1, p_2)$; therefore (H2) is satisfied. The cluster center vector $\mu(p) = (-2, 0, 2)$. The distribution function $F$ has positive continuous derivatives everywhere on $(-a, a)$ except at the origin, hence (H3) is satisfied. The Hessian of the split function at $p$ is given by

$$
B^{(2)}(p) = \begin{pmatrix}
-2 & 2 \\
p_2 - p_1 & p_2 - p_1 \\
2 & -2 \\
p_2 - p_1 & p_2 - p_1
\end{pmatrix}.
$$

Clearly, $\text{Det} B^{(2)}(p) = 0$. Further, the split function has a continuous nonvanishing fourth derivative. Hence according to Remark 3, $r(x) = Cx^3$, where $C$ is a nonvanishing constant. Theorem 2 gives

$$
n^{1/6}(p_n - p) \Rightarrow d(W)^{1/3} e^{(3)}(p),
$$

where $W \sim N(0, 1)$, $d$ is a nonzero constant, and where $e^{(3)}(p)^T = (2^{-1/2}, 2^{-1/2})$. Corollary 2 gives

$$
n^{1/6}(\mu_n(p_n) - \mu(p)) \Rightarrow d(W)^{1/3} e,
$$

where $e^T = (p_1^{-1}, 2^{1/2}(p_2 - p_1), p_1^{-1})$. 

683/41/2-9
The results stated in Section 2 and Proposition 1 are proven here, but before this is done a consistency result is needed.

**Theorem 3 [6].** Suppose that $F$ is a distribution function which satisfies (H1) and (H2), then $\mathbf{p}_n \overset{p}{\to} \mathbf{p}$ as $n \to \infty$.

The proof of this theorem is omitted.

**Proof of Lemma 1.** To prove the first part of the lemma, it suffices to show the result for the components, i.e.,

$$\mu_{jn}(\mathbf{p}_n) - \mu_j(\mathbf{p}_n) = \mu_{jn}(\mathbf{p}) - \mu_j(\mathbf{p}) + o_p(n^{-1/2}); \quad j = 1, 2, \ldots, k. \quad (A1)$$

The left-hand side of (A1) is written as an integral,

$$\mu_{jn}(\mathbf{p}_n) - \mu_j(\mathbf{p}_n) = (p_{jn} - p_{j-1n}) \int_{p_{j-1n}}^{p_{jn}} (F_n^{-1}(u) - F^{-1}(u)) \, du$$

$$= (p_j - p_{j-1}) \int_{p_{j-1}}^{p_j} (F_n^{-1}(u) - F^{-1}(u)) \, du$$

$$+ \left[ (p_{jn} - p_{j-1n})^{-1} - (p_j - p_{j-1})^{-1} \right]$$

$$\times \int_{p_{j-1}}^{p_j} (F_n^{-1}(u) - F^{-1}(u)) \, du + R_{jn}, \quad j = 1, 2, \ldots, k, \quad (A2)$$

where

$$R_{jn} = \frac{\gamma_{j-1n}}{(p_{jn} - p_{j-1n})} \int_{p_{j-1n}}^{p_{jn}} (F_n^{-1}(u) - F^{-1}(u)) \, du$$

$$+ \frac{\gamma_{jn}}{(p_{jn} - p_{j-1n})} \int_{p_{j-1n}}^{p_{jn}} (F_n^{-1}(u) - F^{-1}(u)) \, du$$

with $\alpha_{in} = \min\{p_{in}, p_i\}$, $\beta_{in} = \max\{p_{in}, p_i\}$, and $\gamma_{in} = \text{sign}(p_{in} - p_i)$, $i = 1, 2, \ldots, k$.

Condition (H3) implies that $p_j$ and $p_{j-1}$ are continuity points of $F^{-1}$. Hence one may write

$$\int_{p_{j-1}}^{p_j} (F_n^{-1}(u) - F^{-1}(u)) \, du$$

$$= 1/n \sum_{r=1}^{n} \left[ X_r I\{F^{-1}(p_{j-1}) < X_r \leq F^{-1}(p_j)\} \right.$$

$$- \mu_j(\mathbf{p})(p_j - p_{j-1}) \left. + A_{jn} \right] \quad (A3)$$
where
\[
\Delta_{jn} = \frac{1}{n} \sum_{r=1}^{n} \left[ X_r \left( I\{F_n^{-1}(p_{j-1}) < X_r \leq F_n^{-1}(p_j)\} \right.ight.
\]
\[
- \left. I\{F_n^{-1}(p_{j-1}) < X_r \leq F_n^{-1}(p_j)\} \right].
\]

Condition (H1) and the CLT give that the first term in (A3) is $O_p(n^{-1/2})$. Condition (H3) leads to $\lim_{n \to \infty} |F_n^{-1}(p_i) - F^{-1}(p_i)| = 0$ a.s.; $i = 1, 2, \ldots, k$. This, in turn, along with Chebyshev's inequality yields $\Delta_{jn} = o_p(n^{-1/2})$. Therefore,
\[
\int_{p_{j-1}}^{p_j} (F_n^{-1}(u) - F^{-1}(u)) \, du = O_p(n^{-1/2}). \tag{A4}
\]

Consistency implies that
\[
\left[ (p_{jn} - p_{j-1n})^{-1} - (p_j - p_{j-1})^{-1} \right] = o_p(1).
\]

Hence the second term in (A2) is $o_p(n^{-1/2})$. To estimate $R_{jn}$ note that
\[
|R_{jn}| \leq \left( p_{jn} - p_{j-1n} \right)^{-1}
\times \left[ \int_{\alpha_{j-1n}}^{\beta_{jn}} |F_n^{-1}(u) - F^{-1}(u)| \, du + \int_{\alpha_{jn}}^{\beta_{jn}} |F_n^{-1}(u) - F^{-1}(u)| \, du \right],
\]

$j = 1, 2, \ldots, k$. A bound is obtained below for the second integral. An identical argument works for the first. Clearly,
\[
\int_{\alpha_{jn}}^{\beta_{jn}} |F_n^{-1}(u) - F^{-1}(u)| \, du 
\leq (\beta_{jn} - \alpha_{jn}) \sup_{u \in (\alpha_{jn}, \beta_{jn})} |F_n^{-1}(u) - F^{-1}(u)|, \quad j = 1, 2, \ldots, k.
\]

Consistency implies that $\alpha_{jn}$ and $\beta_{jn}$ both converge in probability to $p_j$, hence $(\beta_{jn} - \alpha_{jn}) = o_p(1)$. Further, by (H3), $F^{-1}$ has a continuous derivative in a neighborhood of $p_j$. Therefore, Bahadur's representation of quantiles (see [1]) and the convergence in probability of $\alpha_{jn}$ and $\beta_{jn}$ to $p_j$ give
\[
\sup_{u \in (\alpha_{jn}, \beta_{jn})} |F_n^{-1}(u) - F^{-1}(u)| = O_p(n^{-1/2}).
\]

This leads to $R_{jn} = o_p(n^{-1/2})$, $j = 1, 2, \ldots, k$. This establishes (A1), which completes the proof of the first part.
The second part of the lemma follows from

\[ \mu_j(p) - \mu_j(p) = (p_j - p_{j-1})^{-1} \int_{p_j}^{p_{j+1}} (F_n^{-1}(u) - F_n^{-1}(u)) \, du, \]

\( j = 1, 2, \ldots, k; \) (A3), the CLT, and the fact that \( A_j = o_p(n^{-1/2}). \) This completes the proof.

**Proof of Lemma 2.** Once again it suffices to show the results for the components, i.e.,

\[ B_j^{(1)}(p_n) = -Z_j + o_p(n^{-1/2}), \quad j = 1, 2, \ldots, k. \]

One only needs the right derivative of \( B_j, \) at \( p_j, j = 1, 2, \ldots, k, \) and stochastic equicontinuity in order to prove this. However, since both the right and left derivative are available a more direct and transparent proof is given. Recall that

\[ B_j^{(1)}(p_n) = (\mu_j + 1_n(p_n) - \mu_j(p_n)) \frac{1}{\sqrt{n}} [\mu_j + 1_n(p_n) - F_n^{-1}(p_j),], \quad (A5) \]

where \( F_n^{-1}(p_j, j = 1, 2, \ldots, k, \) It will be convenient to have some new notation. Let

\[ W_j = \mu_j(p) - \mu_j(p), \quad Y_j = W_j + 1_n - W_j, \]

\[ V_j = Y_j - 2 [F_n^{-1}(p_j) - F_n^{-1}(p_j)]. \]

Lemma 1, assumption (H3), and Bahadur's [1] representation for quantiles, consistency, and tightness give

\[ B_j^{(1)}(p_n) = B_j^{(1)}(p_n) + [\mu_j + 1_n(p_n) - \mu_j(p_n)] W_j, \]

\[ + Y_j [\mu_j + 1_n(p_n) + \mu_j(p_n) - 2F_n^{-1}(p_j)] + Y_n V_j + o_p(n^{-1/2}). \]

In terms of the new notation, the first part of Lemma 1 is

\[ \mu_j(p_n) = \mu_j(p) + W_j + o_p(n^{-1/2}). \]

This is substituted into (A5) to give

\[ B_j^{(1)}(p_n) = B_j^{(1)}(p_n) + [\mu_j + 1_n(p_n) - \mu_j(p_n)] V_j, \]

\[ + Y_j [\mu_j + 1_n(p_n) + \mu_j(p_n) - 2F_n^{-1}(p_j)] + Y_n V_j + o_p(n^{-1/2}). \]

Lemma 1 gives \( Y_j = O_p(n^{-1/2}). \) Assumption (H3), consistency, and Bahadur's [1] representation imply \( V_j = O_p(n^{-1/2}). \) Consistency of \( p_n, \)

(H2), and (H3) give \( \mu_j + 1_n(p_n) + \mu_j(p_n) - 2F_n^{-1}(p_j) \to 0 \) as \( n \to \infty. \) The
continuity of \( \mu(t) \) and consistency give \( \mu_j(p_n) - \mu_j(p) \rightarrow \mu_{j+1}(p) - \mu_j(p) \) as \( n \rightarrow \infty \). These observations are put together to yield

\[
B_j^{(\pm)}(p_n) = B_j^{(\pm)}(p_n') + [\mu_j(p) - \mu_j(p)] V_{jn} + o_p(n^{-1/2}). \tag{A6}
\]

Condition (H3) and the consistency of \( p_n \) imply that, for \( n \) sufficiently large, \( B_j^{(+)}(p_n) = B_j^{(-)}(p_n) \). Recall that \( p_n \) is the location of a maximum of \( B \), hence

\[
B_j^{(+)}(p_n) \leq 0 \leq B_j^{(-)}(p_n).
\]

Therefore,

\[
B_j^{(1)}(p_n) = -Z_{jn} + o_p(n^{-1/2}),
\]

where \( Z_{jn} = [\mu_j(p) - \mu_j(p)] V_{jn} \). Consistency, (H3), (A3), and Bahadur's [1] representation give

\[
V_{jn} = (1/n) \sum_{r=1}^{n} \eta_r + o_p(n^{-1/2}), \tag{A7}
\]

where

\[
\eta_r = (X, [j(p) - p_j])^{-1} I\{F^{-1}(p_j) < X, F^{-1}(p_{j+1}) \} + (p_j - p_{j-1})^{-1} I\{F^{-1}(p_{j-1}) < X, F^{-1}(p_j) \} + 2(f(\bar{\mu}_j)^{-1} I\{X, F^{-1}(p_j) \} - [\mu_j(p) + \mu_j(p) - 2f(\bar{\mu}_j)^{-1} p_j] \}; \quad r = 1, 2, \ldots, n. \tag{A8}
\]

The random variables (A8) have finite second moments and are centered. The proof is completed by invoking the CLT.

**Proof of Proposition 1.** To prove part (i), it suffices to show that \( k - 2 \) of the columns of \( B^{(2)}(p) \) form linearly independent vectors. The matrix \( B^{(2)}(p) \), which is symmetric, is given by

\[
[B^{(2)}(p)]_{ij} = \left( \mu_{j+1}(p) - \mu_j(p) \right) \times \left[ \frac{(\mu_{j+1}(p) - \mu_j(p)) p_{j+1} - p_{j-1}}{2(p_{j+1} - p_j)(p_j - p_{j-1})} - \frac{2}{f(\bar{\mu}_j)} \right],
\]

\[j = 1, 2, \ldots, k - 1;\]

\[
[B^{(2)}(p)]_{j+1} = \frac{(\mu_{j+2}(p) - \mu_{j+1}(p)) (\mu_{j+1}(p) - \mu_j(p))}{2(p_{j+1} - p_j)} = [B^{(2)}(p)]_{j+1},
\]

\[j = 1, 2, \ldots, k - 2. \] All the other elements are zero. Let \( v^{(r)} \in \mathbb{R}^{k-1} \), \( r = 1, 2, \ldots, k - 1 \), have components

\[
(v^{(r)})_j = [B^{(2)}(p)]_{ij}; \quad r = 1, 2, \ldots, k - 1; j = 1, 2, \ldots, k - 1.
\]
Define a vector \( \mathbf{v} \) by

\[
\mathbf{v} = \sum_{r=1}^{k-2} b^{(r)} \mathbf{v}^{(r)},
\]

where \( b^{(r)} \)'s are real numbers. It will be argued that \( \mathbf{v} = 0 \) implies that \( b^{(r)} = 0 \) for all \( r \).

Suppose that \( \mathbf{v} = 0 \). Then one has

\[
(v)_{k-1} = b^{(k-2)} \left[ B^{(2)}(p) \right]_{k-1k-2} = 0.
\]

(H2) implies that \( \left[ B^{(2)}(p) \right]_{k-1k-2} > 0 \); hence one has \( b^{(k-2)} = 0 \). Iteration of this argument leads to \( b^{(r)} = 0 \) for all \( r \). That is, the first \( (k-2) \) columns of \( B^{(2)}(p) \) are linearly independent. Since \( B^{(2)}(p) \) is singular, this implies that it has rank \( (k-2) \). In terms of the subspaces this is equivalent to the statement that \( W^{(v)}(p) \) is one-dimensional. This completes the proof of the first part of the proposition.

To prove part (ii) of the theorem it is noted that, under (H3), \( B^{(2)}(t) \) is continuous in some neighborhood of \( p \). Therefore, \( P^{(v)}(t) \) is continuous in this neighborhood and \( \lim_{t \to p} P^{(v)}(t) = P^{(v)}(p) \) (see [8, pp. 123–126]). This completes the proof of part (ii).

Next, part (iii) is considered. The continuity of \( B^{(2)}(t) \) in a neighborhood of \( p \), under (H3), along with its symmetry as a Hessian implies that its repeated eigenvalues \( \lambda_1(t) \geq \lambda_2(t) \geq \lambda_3(t) \geq \cdots \geq \lambda_{k-1}(t) \) are also continuous in this neighborhood (see [8, pp. 123–126]). This along with part (i) implies that \( \lambda_1(t) = \lambda_i(t) > \lambda_i(t), i \neq 1, \) for \( t \) sufficiently near \( p \). This immediately gives that \( W^{(v)}(t) \) is one-dimensional for \( t \) sufficiently near \( p \), and completes the proof of part (iii).

Finally, part (iv) is taken up. By assumption (H2), \( B \) has a unique maximum at \( p \) and by (H3), \( B^{(2)}(t) \) is continuous in a neighborhood of \( p \). Together these imply that if \( B^{(2)}(p) \) is singular, then \( [B^{(2)}(t)]^{-1} \) exists everywhere in some neighborhood of \( p \) with the exception of \( p \).

To complete the proof of the final part, let \( P_i(t) \) denote the projection onto the subspace spanned by the eigenvector corresponding to the repeated eigenvalue \( \lambda_i(t), i = 1, 2, \ldots, k-1 \). The following expansions are useful

\[
P^{(v)}(t) = \sum_{i=2}^{k-1} P_i(t),
\]

\[
[B^{(2)}(t)]^{-1} = \sum_{i=1}^{k-1} \lambda_i^{-1}(t) P_i(t),
\]
and

$$[B^{(2)}(p)]^+ = \sum_{i=2}^{k-1} \lambda_i^{-1}(p) P_i(p),$$

as is the following notation: For any two $r \times r$ matrices $A$ and $C$, the notation $A \preceq C$ and $|A|$ will denote $A_{ij} \preceq C_{ij}$ and $|A_{ij}|$, $i, j = 1, 2, \ldots, r$, respectively. With this convention one has

$$|P^{(f)}(t)[B^{(2)}(t)]^{-1} - [B^{(2)}(p)]^+|$$

$$= \left| \sum_{i=2}^{k-1} \left[ (\tilde{\lambda}_i^{-1}(t) - \tilde{\lambda}_i^{-1}(p) (P_i(t) - P_i(p))) \right] \right|$$

$$\leq |P^{(f)}(t)| \max_{i=2, \ldots, k-1} |\tilde{\lambda}_i^{-1}(t) - \tilde{\lambda}_i^{-1}(p)|$$

$$+ |P^{(s)}(t) - P^{(s)}(p)| \max_{i=2, \ldots, k-1} |\tilde{\lambda}_i^{-1}(p)|.$$

From part (i), $\tilde{\lambda}_i(p) \neq 0$, $i > 1$. Therefore, the second term on the far right is bounded away from infinity and, by part (ii), goes to zero as $t \to p$. On the other hand, the continuity of the repeated eigenvalues gives that the first term goes to zero as $t \to p$. This completes the proof.

**APPENDIX B**

The variance-covariance matrices which appear in Lemmas 1 and 2 are given here. Let

$$\sigma_j^2 = (p_j - p_{j-1})^{-1} \int_{p_{j-1}}^{p_j} (F_n^{-1}(u) - F^{-1}(u))^2 \, du,$$

$$\mu_j = \mu_j(p), \quad j = 1, 2, \ldots, k.$$

The first matrix (Lemma 1) has components

$$[\Sigma^{(0)}]_{jj} = \sigma_j^2 + \mu_j^2 \left[ \frac{1 - (p_j - p_{j-1})}{(p_j - p_{j-1})} \right]=$$

$$[\Sigma^{(0)}]_{jl} = \frac{-\mu_j \mu_l}{(p_j - p_{j-1})(p_l - p_{l-1})}, \quad l \neq j, l = 1, 2, \ldots, k.$$
The second matrix (Lemma 2) is given by

\[
\begin{align*}
\mathbf{[\Sigma]}_{jj} &= \mathbf{[\Sigma]}_{jj+1} + \mathbf{[\Sigma]}_{jj} + \left( \frac{2}{f(\mu_j)} \right)^2 p_j (1 - p_j) + 2 \mathbf{[\Sigma]}_{jj+1} \\
&+ 2 \left( \frac{2}{f(\mu_j)} \right) \mu_{j+1} p_j + 2 \left( \frac{2}{f(\mu_j)} \right) \mu_j (1 - p_j); \\
&\quad j = 1, 2, ..., k - 1,
\end{align*}
\]

\[
\mathbf{[\Sigma]}_{jl} = \mathbf{[\Sigma]}_{j+l+1} + \mathbf{[\Sigma]}_{j+l} + \mathbf{[\Sigma]}_{j+l+1} + \mathbf{[\Sigma]}_{j+l+1} \\
+ \left( \frac{2}{f(\mu_l)} \right) \left( \frac{2}{f(\mu_l)} \right) p_l (1 - p_l) \\
- \left( \frac{2}{f(\mu_l)} \right) p_l (\mu_{l+1} + \mu_l) + \left( \frac{2}{f(\mu_l)} \right) (1 - p_l) (\mu_{l+1} + \mu_l),
\]

\[l < j; j, l = 1, 2, ..., k - 1.\]

ACKNOWLEDGMENTS

We thank Steve Arnold for discussions about the Moore-Penrose inverse, and William Harkness, Gene Wayne, and an anonymous referee whose comments and criticisms led to a better presentation of the results.

REFERENCES