The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 193-204

CONSISTENCY AND ASYMPTOTIC NORMALITY OF MLE'S FOR EXPONENTIAL MODELS[1]

BY ROBERT H. BERK

Rutgers University

0. Summary. Conditions are given for the strong consistency and asymptotic normality of the MLE (maximum likelihood estimator) for multiparameter exponential models. Because of the special structure assumed, the conditions are less restrictive than required by general theorems in this area. The technique involves certain convex functions on Euclidean spaces that arise naturally in the present context. Some examples are considered; among them, the multinomial distribution. Some convexity and continuity properties of multivariate cumulant generating functions are also discussed.

1. Introduction. Let $\mathbf{Y}, \mathbf{Y}_1, \mathbf{Y}_2, \cdots$ be a sequence of m-dimensional random variables. All distributions we consider render the sequence i.i.d. We work with a natural exponential family of pdf's for $\mathbf{Y}$ of the form

$$p(y \mid \omega) = \exp\{\omega' y - c(\omega)\} . \tag{1.1}$$

The pdf's in (1.1) are relative to some σ-finite measure $\nu \neq 0$ on R^m. We say ν generates the exponential family p. The parameter ω ranges in a subset Ω of the natural parameter set $\Omega(\nu) = \{\omega : \int \exp\{\omega' y\}\, d\nu(y) < \infty\}) \subset R^m$. Clearly $c(\omega) = \log \int \exp\{\omega' y\}\, d\nu(y)$ and $\Omega(\nu) = (c < \infty)$. Throughout, we assume that $\Omega(\nu)$ has non-empty interior $\Omega^0(\nu)$ and that the mapping $\omega \to p(\cdot \mid \omega)$ is $1 - 1$ on $\Omega(\nu)$. This entails no essential loss of generality, as indicated below in Section 2.

We consider consistency and asymptotic normality for the MLE of ω in Ω. Section 2 contains some preliminary results; consistency is discussed in Section 3. In Section 4, the consistency results are extended to general exponential families. Some examples are considered in Section 5 and asymptotic normality is discussed in Section 6. The multinomial distribution is treated in Section 5 too. Although not quite an exponential family, with slight modification our methods apply in this case as well. Besides illustrating a different approach to this model, we are able to improve slightly on the results previously given in the literature.

2. Convexity and continuity properties. We collect here some facts about the function c. Many of these may be viewed as properties of multivariate Laplace

Received May 11, 1970; revised June 14, 1971.

[1] This work was begun while the author held an NSF Postdoctoral Fellowship and was further supported by NSF Grant GP-7350.

transforms or cumulant generating functions. The reader is referred to Eggleston (1958) for a general discussion of the facts about convexity cited here. In the sequel, $C(\nu)$ denotes the convex hull of the support of ν: $C(\nu)$ is the smallest convex subset of R^m whose complement is a ν-null set.

A. We note that $\Omega(\nu)$ is a convex subset of R^m. This follows directly from the fact that $h(\omega) = \int \exp\{\omega' y\}\, d\nu(y)$ is convex. It follows from Fatou's lemma that h is lower semi-continuous (lsc) on R^m: if $\omega_n \to \omega$, $\liminf h(\omega_n) \geqq h(\omega)$. Thus $c = \log h$ is also lsc. Moreover, c is convex on R^m. For if u, $v \in R^m$, it follows from Hölder's inequality that for $0 < t < 1$, $h(tu + (1-t)v) \leqq (h(u))^t (h(v))^{1-t}$.

Upon taking logarithms, we see that

$$c(tu + (1-t)v) \leqq tc(u) + (1-t)c(v)\,. \tag{2.1}$$

Regarding strict inequality in (2.1), we may assert

LEMMA 2.1. *The following statements are equivalent.*

(i) ν *is not supported on a flat.*
(ii) c *is strictly convex on* $\Omega(\nu)$.
(iii) $\omega \to p(\cdot \mid \omega)$ *is* $1-1$ *on* $\Omega(\nu)$.

PROOF. (i) $\Leftrightarrow$ (ii). Choose u and v in $\Omega(\nu)$, $u \neq v$. It is enough to consider $t = \frac{1}{2}$ in (2.1). The necessary and sufficient condition for equality in Schwarz' inequality, applied to the preceding, shows that there is equality in (2.1) iff $(u-v)'y$ is a.s. constant $[\nu]$. This happens iff ν is supported on a flat.

(i) $\Leftrightarrow$ (iii). We have $p(\cdot \mid u) = p(\cdot \mid v)[\nu] \Leftrightarrow u'y - c(u) = v'y - c(v)[\nu] \Leftrightarrow (u-v)'y = c(u) - c(v)[\nu] \Leftrightarrow \nu$ is supported on a flat. The first statement is the negation of (iii); the last, the negation of (i). □

By transferring ν to a space of lower dimension, if necessary, we may always arrange that 2.1 (i) holds. Thus the assumption that 2.1 (iii) holds entails no essential loss of generality.

B. Let P_ω be the measure corresponding to $p(\cdot \mid \omega)\, d\nu$. Since p cannot vanish, for all $\omega \in \Omega(\nu)$, $P_\omega \equiv \nu$. In particular, $C(P_\omega) = C(\nu)$. Any P_ω generates the same exponential family as ν, except that $\Omega(P_\omega) = \Omega(\nu) - \omega = \{v - \omega : v \in \Omega(\nu)\}$. For many purposes, we may thus assume that ν is a probability measure and that $\Omega(\nu)$ contains the origin. A non-singular linear transformation H of R^m yields a corresponding transformation of ν and p: Corresponding to the transformed variable $H\mathbf{Y}$ is an isomorphic exponential family of pdfs, generated by νH^{-1} and with normalization function cH^{-1}. The common convex support of the family is $C(\nu H^{-1}) = HC(\nu)$ and the new index set is $\Omega(\nu H^{-1}) = H\Omega(\nu)$.

Assuming that $\Omega^0(\nu) \neq \varnothing$ also entails no essential loss of generality. For if $\Omega(\nu)$ spans a flat of dimension $r < m$, by translating and rotating with

an appropriate H, we may assume that $\Omega(\nu)$ spans the subspace $V = \{(\omega_1, \cdots, \omega_m): \omega_{r+1} = \cdots = \omega_m = 0\}$. Then as (y, ω) ranges through $R^m \times \Omega(\nu)$, p depends on y and ω only through their first r coordinates. The measure ν_r on R^r induced by the projection $(y_1, \cdots, y_m) \to (y_1, \cdots, y_r)$ then generates the same exponential family p, considered now as a function on $R^r \times R^r$. Moreover, $\Omega^0(\nu_r) \neq \emptyset$.

C. A direct consequence of convexity is that c is continuous on $\Omega^0(\nu)$. (See Theorem 24 of Eggleston, 1958.) In fact, c is infinitely differentiable on $\Omega^0(\nu)$. This may be seen from the fact that the moment generating function (mgf) of $\mathbf{Y}$ under P_ω is $E_\omega \exp\{u'\mathbf{Y}\} = \exp\{c(u + \omega) - c(\omega)\}$. If $\omega \in \Omega^0(\nu)$, the mgf is finite as u ranges in a neighborhood of zero. Equivalently, letting $|y|$ denote Euclidean length in R^m, $|\mathbf{Y}|$ has a non-trivial mgf under P_ω for $\omega \in \Omega^0(\nu)$. Then all coordinates of $\mathbf{Y}$ have moments of all orders. On $\Omega^0(\nu)$, we also have $E_\omega\mathbf{Y} = \dot{c}(\omega)$, where $\dot{c} = (\partial c/\partial\omega_1, \cdots, \partial c/\partial\omega_m)'$ and $E_\omega[(\mathbf{Y} - \dot{c}(\omega)) \times (\mathbf{Y} - \dot{c}(\omega))'] = \ddot{c}(\omega)$, where $\ddot{c} = (\partial^2 c/\partial\omega_i\partial\omega_j)$. Because $\mathbf{Y}$ does not lie in a flat $[P_\omega]$, $\ddot{c}$ is positive definite.

D. For $\omega \in \Omega(\nu)$, $E_\omega\mathbf{Y}$ exists if $E_\omega|\mathbf{Y}| < \infty$. Let $\Omega_1(\nu) = \{\omega \in \Omega(\nu): E_\omega|\mathbf{Y}| < \infty\}$. It is easily seen that $\Omega_1(\nu)$ is convex and $\Omega_1(\nu) \supset \Omega^0(\nu)$. The mapping $\omega \to E_\omega\mathbf{Y}$ of $\Omega_1(\nu)$ into R^m is an extension of $\dot{c}$ from $\Omega^0(\nu)$ to $\Omega_1(\nu)$. We denote this extended mapping by $\dot{c}$ also.

LEMMA 2.2. *The mapping $\dot{c}$ is $1 - 1$ on $\Omega_1(\nu)$.*

PROOF. For $v, \omega \in \Omega(\nu)$, we have always $E_\omega \log[p(\mathbf{Y} \mid v)/p(\mathbf{Y} \mid \omega)] \leqq 0$. Equality holds only if $P_v = P_\omega$, which, is view of 2.1 (iii), happens only if $v = \omega$. Thus by taking $v \neq \omega \in \Omega_1(\nu)$ and letting $\eta = \dot{c}(\omega)$, we obtain $0 > E_\omega(v - \omega)'\mathbf{Y} + c(\omega) - c(v)$ or

$$v'\eta - c(v) < \omega'\eta - c(\omega)\,. \tag{2.2}$$

Thus the LHS of (2.2) attains a unique maximum at ω, which implies that the correspondence $\omega \to \eta$ is unique. □

Then on $\Omega_1(\nu)$, one could alternatively parametrize the exponential family by the expectation $\eta = \dot{c}(\omega)$. The conceptual advantages in doing so are discussed below in Section 3.

E. It is convenient to introduce the likelihood function $q(y \mid \omega) = \omega'y - c(\omega)$. q represents the likelihood for the sufficient statistic $\mathbf{Y}_n = \sum_1^n \mathbf{Y}_i/n$ as well as for a single observation. The properties of c established above imply that $q(y \mid \cdot)$ is concave and usc in ω. In particular, $q(y \mid \cdot)$ attains a maximum on any compact subset of R^m.

In the sequel, the set of outcomes for which the likelihood is bounded plays an important role. Accordingly, for $y \in R^m$ and $V \subset \Omega(\nu)$ we define

$$q(y \mid V) = \sup\{\omega'y - c(\omega): \omega \in V\}\,. \tag{2.3}$$

Since q is linear in y, it follows that $q(\cdot \mid V)$ is convex and lsc. Let $B(V) = (q(\cdot \mid V) < \infty)$ and $B(\nu) = B(\Omega(\nu))$. Then for all $V \subset \Omega(\nu)$, $B(V)$ is convex and $B(V) \supset B(\nu)$. Theorem 24 of Eggleston (1958) shows that $q(\cdot \mid V)$ is continuous on $B^0(V)$ and *a fortiori*, on $B^0(\nu)$. The following lemma establishes that $B^0(\nu) \neq \emptyset$ and gives some relations among the sets we have considered. $\bar{C}$ denotes the closure of $C \subset R^m$.

LEMMA 2.3. (i) $\dot{c}(\Omega_1(\nu)) \subset B(\nu) \subset \overline{C(\nu)}$.

(ii) *If every support hyperplane intersects $C(\nu)$ in a ν-null set, then $B(\nu) \subset C^0(\nu)$.*

(iii) $\dot{c}(\Omega^0(\nu))$ *is open; thus* $B^0(\nu) \neq \emptyset$.

(iv) $\dot{c}(\Omega_1(\nu)) \subset C^0(\nu)$.

PROOF. The first inclusion in (i) follows from (2.2). We establish the second as follows. We note first that

$$\zeta \in B(\nu) \Leftrightarrow \inf \{\int \exp\{\omega'(y - \zeta)\}\, d\nu(y) : \omega \in \Omega(\nu)\} > 0 \,. \tag{2.4}$$

Since $\overline{C(\nu)}$ is closed and convex, if $\zeta \notin \overline{C(\nu)}$, there is a hyperplane through ζ, one of whose closed half-spaces does not meet $\overline{C(\nu)}$. I.e., there is an $\omega \neq 0$ in R^m so that $\omega'(y - \zeta) < 0$ for $y \in C(\nu)$. As discussed in (B), we may assume that ν is a probability measure. Then for $k > 0$, $\int \exp\{k\omega'(y - \zeta)\}\, d\nu(y) < 1$; thus $k\omega \in \Omega(\nu)$ for all $k > 0$. As $k \to \infty$, $\exp\{k\omega'(y - \zeta)\} \to 0$ pointwise on $C(\nu)$. By dominated convergence, $\int \exp\{k\omega'(y - \zeta)\}\, d\nu(y) \to 0$ as $k \to \infty$. Thus the second condition in (2.3) is violated and $\zeta \notin B(\nu)$.

(ii) We remark first that 2.1 (i) is a necessary condition for the hypothesis. Suppose then that $\zeta \notin C^0(\nu)$. Then there is a hyperplane through ζ, one of whose open half-spaces does not meet $C(\nu)$. I.e., for some $\omega \neq 0$, $\omega'(y - \zeta) \leqq 0$ for $y \in C(\nu)$. The hypothesis implies that $\nu\{y \in C(\nu) : \omega'(y - \zeta) = 0\} = 0$; hence $\omega'(y - \zeta) < 0$ a.e. $[\nu]$. The previous argument then shows that $\zeta \notin B(\nu)$.

(iii) We note that on $\Omega^0(\nu)$, the Jacobian matrix of $\dot{c}$ is $\ddot{c}$. Since $\det \ddot{c} > 0$ on $\Omega^0(\nu)$, it follows that $\dot{c}(\Omega^0(\nu))$ is open (see, e.g., Buck (1956), page 218, Theorem 24). Since $\Omega^0(\nu) \neq \emptyset$, it follows from (i) that $B^0(\nu) \neq \emptyset$.

(iv) If $\omega \in \Omega_1(\nu)$ and $\eta = E_\omega \mathbf{Y} \notin C^0(\nu)$, then for some $v \neq 0$, $v'(y - \eta) \geqq 0$ for $y \in C(\nu)$. I.e., $v'(\mathbf{Y} - \eta) \geqq 0$ $[P_\omega]$. But $E_\omega v'(\mathbf{Y} - \eta) = 0$; thus $v'(\mathbf{Y} - \eta) = 0$ $[P_\omega]$. That is P_ω is supported on a hyperplane, which contradicts 2.1 (i). □

3. Consistency for natural exponential families. Using the model (1.1), we treat consistency for the MLE of ω in $\Omega \subset \Omega(\nu)$ based on $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$. We do not suppose that P, the actual distribution of $\mathbf{Y}$ is in the model. P will also denote the distribution of the entire i.i.d. sequence $\mathbf{Y}, \mathbf{Y}_1, \cdots$. The normalized likelihood function for $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ is

$$n^{-1} \textstyle\sum_1^n \log p(\mathbf{Y}_i \mid \omega) = q(\bar{\mathbf{Y}}_n \mid \omega) \,. \tag{3.1}$$

If a conventional MLE is to exist, $q(\bar{\mathbf{Y}}_n \mid \cdot)$ must be bounded on Ω. That is,

we must have $\bar{\mathbf{Y}}_n \in B(\Omega)$. Assuming $\eta = E_P\mathbf{Y}$ exists, $\bar{\mathbf{Y}}_n \to \eta$ w.p.1. Thus if $\eta \in B^0(\Omega)$, then eventually $\bar{\mathbf{Y}}_n \in B(\Omega)$. This condition on η is an essential sufficient condition in the following theorem.

THEOREM 3.1. *Suppose* (i) Ω *is locally compact.* (ii) $E_P|\mathbf{Y}| < \infty$ *and* $\eta = E_P\mathbf{Y} \in B^0(\Omega)$. (iii) $q(\eta \mid \cdot)$ *attains a unique maximum on* Ω *at* ω *say, and is bounded below* $q(\eta \mid \omega)$ *off neighborhoods of* ω. *Then* w.p.1, $q(\bar{\mathbf{Y}}_n \mid \cdot)$ *eventually attains a maximum on* Ω. *If the maximizing point is unique, it is measurable. If* $\boldsymbol{\omega}_n$ *is a measurable maximizing point for* $q(\bar{\mathbf{Y}}_n \mid \cdot)$, *then* $P(\boldsymbol{\omega}_n \to \omega) = 1$. *If in addition,* Ω *is locally convex,* $q(\bar{\mathbf{Y}}_n \mid \cdot)$ *eventually attains a unique maximum on* Ω.

REMARK. Above, local compactness of Ω refers to its relative topology. This means that if U_ε is the closed ε-sphere about $\omega \in \Omega$, then $U_\varepsilon \cap \Omega$ is compact for ε sufficiently small. We take an estimator $\boldsymbol{\omega}$ in Ω to be measurable if for every relatively open $V \subset \Omega$, $(\boldsymbol{\omega} \in V)$ is measurable. If the maximizing point $\boldsymbol{\omega}_n$ is not assumed to be measurable, we can still assert that $\boldsymbol{\omega}_n \to \omega$ for a set of outcomes having P-measure one.

PROOF. Let $S = \{v \in \Omega : |v - \omega| \leqq \varepsilon\}$. Since Ω is locally compact, we may assume S is compact by taking ε sufficiently small. Let $V = \Omega - S$. Since w.p.1 $\bar{\mathbf{Y}}_n \to \eta \in B^0(\Omega)$, $\bar{\mathbf{Y}}_n$ is eventually in $B^0(\Omega)$. Then, by (iii) and the continuity of $q(\cdot \mid V)$, w.p.1 $q(\bar{\mathbf{Y}}_n \mid V) \to q(\eta \mid V) < q(\eta \mid \omega)$. Since also w.p.1 $q(\bar{\mathbf{Y}}_n \mid \omega) \to q(\eta \mid \omega)$, w.p.1 eventually $q(\bar{\mathbf{Y}}_n \mid \omega) > q(\bar{\mathbf{Y}}_n \mid V)$. Thus eventually, the global supremum of $q(\bar{\mathbf{Y}}_n \mid \cdot)$ occurs on S. Since S is compact, $q(\bar{\mathbf{Y}}_n \mid \cdot)$ eventually attains its supremum on Ω and all of its maximizing points are in S. Because ε is arbitrary, for any measurable MLE $\boldsymbol{\omega}_n$, $P(\boldsymbol{\omega}_n \to \omega) = 1$.

If Ω is locally convex, then for ε sufficiently small, S is convex as well as compact. Since $q(\bar{\mathbf{Y}}_n \mid \cdot)$ is strictly concave, it attains a unique maximum on S. Then as soon as $q(\bar{\mathbf{Y}}_n \mid \omega)$ becomes and remains larger than $q(\bar{\mathbf{Y}}_n \mid V)$, this unique maximizing point in S is the unique MLE on Ω.

Suppose finally that $q(\bar{\mathbf{Y}}_n \mid \cdot)$ attains a unique maximum on Ω at $\boldsymbol{\omega}_n$. We establish the measurability of $\boldsymbol{\omega}_n$ as follows. Let W be a compact subset of Ω. Then $(\boldsymbol{\omega}_n \in W) = (q(\bar{\mathbf{Y}}_n \mid W) = q(\bar{\mathbf{Y}}_n \mid \Omega))$. Since $q(\cdot \mid W)$ and $q(\cdot \mid \Omega)$ are lsc and hence measurable, we conclude that $\boldsymbol{\omega}_n$ is measurable. □

REMARK. The requirement that Ω be locally compact is satisfied if Ω is open or closed. The effect of this condition is to assure that $q(\bar{\mathbf{Y}}_n \mid \cdot)$ (eventually) does attain a maximum on Ω. If some such condition were not imposed, it could happen that although $q(\bar{\mathbf{Y}}_n \mid \cdot)$ is bounded on Ω (as soon as $\bar{\mathbf{Y}}_n \in B^0(\Omega)$), the supremum is never attained. One can also avoid this problem by considering almost-maximum likelihood estimators. I.e., choose $\boldsymbol{\omega}_{an}$ to satisfy $q(\bar{\mathbf{Y}}_n \mid \boldsymbol{\omega}_{an}) > q(\bar{\mathbf{Y}}_n \mid \Omega) - \delta_n$, where $0 < \delta_n \to 0$ as $n \to \infty$. Such a $\boldsymbol{\omega}_{an}$ always exists and can be chosen to be measurable. Eventually $\boldsymbol{\omega}_{an}$ is in S too, so

$\boldsymbol{\omega}_{an} \to \omega$ on a set having P-measure one. Condition 3.1 (i) is unnecessary with this approach.

If P is or resembles, in a certain sense, a distribution in $\Omega \cap \Omega_1(\nu)$, the conditions of Theorem 3.1 may be verified.

LEMMA 3.2. *Assume* (i) $E_P|\mathbf{Y}| < \infty$ *and for some* $\omega \in \Omega \cap \Omega_1(\nu)$, $E_P\mathbf{Y} = E_\omega\mathbf{Y}$. (ii) $\eta = E_P\mathbf{Y} \in B^0(\Omega)$. *Then* 3.1 (ii, iii) *hold.*

REMARK. It follows from Lemma 2.2 that ω must be unique.

PROOF. We discuss only 3.1 (iii). It follows from (2.1) that $q(\eta \mid \cdot)$ attains a unique maximum on $\Omega(\nu)$ at ω. If we replace Ω by $\Omega(\nu)$ in 3.1 (iii), the desired boundedness is a fact generally true about strictly concave functions on convex sets. *A fortiori*, the boundedness holds on $\Omega \subset \Omega(\nu)$. □

REMARK. If in 3.2 (i), $\omega \in \Omega \cap \Omega^0(\nu)$, 3.2 (ii) is redundant. This follows from Lemma 2.3 (i and iii). Then the following corollary, which applies to many common exponential families, is immediate.

COROLLARY 3.3. *Suppose* $\Omega \subset \Omega^0(\nu)$. *If for some* $\omega \in \Omega$, $E_P\mathbf{Y} = E_\omega\mathbf{Y}$, 3.1 (ii, iii) *hold.*

If $E_P\mathbf{Y} = E_\omega\mathbf{Y}$ for some ω in $\Omega^0(\nu)$, then $\eta = E_P\mathbf{Y} \in \dot{c}(\Omega^0(\nu))$. Since $\dot{c}(\Omega^0(\nu))$ is open (2.3(iii)), w.p.1 $\bar{\mathbf{Y}}_n$ is eventually in $\dot{c}(\Omega^0(\nu))$. When that happens, or more generally, if $\bar{\mathbf{Y}}_n \in \dot{c}(\Omega_1(\nu))$, $q(\bar{\mathbf{Y}}_n \mid \cdot)$ has a unique maximum on $\Omega(\nu)$ at the point $\boldsymbol{\omega}_n$ satisfying

$$\dot{c}(\boldsymbol{\omega}_n) = \bar{\mathbf{Y}}_n \,. \tag{3.2}$$

This follows from (2.2), although it may be heuristically seen by formally differentiating $q(\bar{\mathbf{Y}}_n \mid \omega) = \omega'\mathbf{Y}_n - c(\omega)$. (However, c is totally differentiable only on $\Omega^0(\nu)$ in general.) An advantage of the parameterization $\eta = \dot{c}(\omega)$ becomes evident here. The MLE for η in $\dot{c}(\Omega_1(\nu))$ is just $\bar{\mathbf{Y}}_n$, which is visibly consistent and asymptotically normal. (P. J. Bickel brought this fact to the author's attention.)

4. General exponential families. We extend the preceding results to general exponential families. Let $\mathbf{X}, \mathbf{X}_1, \cdots$ be a sequence of i.i.d. random variables with values in $\mathscr{X}$ and having common distribution F. F is efined on $\mathscr{A}$, a measurable structure for $\mathscr{X}$. A general exponential model for $\mathbf{X}$ is a family of pdf's of the form

$$f(x \mid \theta) = \exp\{\textstyle\sum_1^m \alpha_i(\theta)\beta_i(x) - \gamma(\theta)\}\,; \tag{4.1}$$

all pdf's are relative to some σ-finite measure $\mu \neq 0$ on $\mathscr{A}$. The parameter θ ranges in a parameter space Θ. We consider the MLE for θ based on the model f.

Let $\beta(x) = (\beta_1(x)), \cdots, \beta_m(x))'$ and $\mathbf{Y} = \beta(X)$. Clearly the MLE for θ depends only on $\mathbf{Y}, \mathbf{Y}_1, \cdots$. The mapping β induces on R^m an exponential family p (corresponding to f) of pdf's for $\mathbf{Y}$. By letting $\nu = \mu\beta^{-1}$ and $\alpha(\theta) = (\alpha_1(\theta), \cdots, \alpha_m(\theta))'$, the correspondence between f and p is given by

$$f(x \mid \theta) = p(\beta(x) \mid \alpha(\theta)) ,$$

where p is the natural exponential family induced by ν. If we write $\omega = \alpha(\theta)$, the corresponding range for ω is $\Omega = \alpha(\Theta)$. $P = F\beta^{-1}$ is the actual distribution of $\mathbf{Y}$.

In order that the basic condition 2.1 (i) hold for ν, β must not lie in a flat in $R_m[\mu]$. If the model f is identified (i.e., $\theta \to f(\cdot \mid \theta)$ is $1 - 1$), then $\alpha : \Theta \to \Omega$ must be $1 - 1$. We henceforth assume μ and α satisfy these conditions. Then $a = \alpha^{-1}$ exists and $a : \Omega \to \Theta$.

Under f, the normalized likelihood for $\mathbf{X}_1, \cdots, \mathbf{X}_n$ is

$$\begin{aligned} \mathbf{g}_n(\theta) &= n^{-1} \textstyle\sum_1^n \log f(\mathbf{X}_i \mid \theta) = \alpha'(\theta)\overline{\mathbf{Y}}_n - \gamma(\theta) \\ &= \alpha'(\theta)\overline{\mathbf{Y}}_n - c(\alpha(\theta)) = q(\overline{\mathbf{Y}}_n \mid \alpha(\theta)) . \end{aligned} \tag{4.2}$$

In order that convergence of the MLE in Θ have meaning, Θ must be a topological space. We take measurability of an estimator $\boldsymbol{\theta}$ ranging in Θ to mean Borel measurability. Theorem 3.1 carries over as follows.

THEOREM 4.1. *Suppose* (i) β *does not lie in a flat* $[\mu]$. (ii) *Conditions* 3.1 (i–iii) *hold for* (p, Ω, P). (iii) $a : \Omega \to \Theta$ *is continuous. Then* w.p.1, $\mathbf{g}_n$ *eventually attains a maximum on* Θ. *If the maximizing point is unique, it is measurable. If* $\boldsymbol{\theta}_n$ *is a measurable maximizing point for* $\mathbf{g}_n$, *then* $F(\boldsymbol{\theta}_n \to \theta) = 1$, *where* $\theta = a(\omega)$. *If in addition,* Ω *is locally convex,* $\mathbf{g}_n$ *eventually attains a unique maximum on* Θ.

PROOF. The theorem is a direct consequence of Theorem 3.1, using the identifications made above and the $1 - 1$ correspondence $\Omega \to \Theta$. In particular: (a) $\mathbf{g}_n$ attains a (unique) maximum on Θ iff $q(\overline{\mathbf{Y}}_n \mid \cdot)$ attains a (unique) maximum on Ω (see (4.2)). (b) If $\boldsymbol{\omega}_n$ is a measurable maximizing point for $q(\overline{\mathbf{Y}}_n \mid \cdot)$, then $\boldsymbol{\theta}_n = a(\boldsymbol{\omega}_n)$ is a maximizing point for $\mathbf{g}_n$ and is measurable (because a is continuous, hence measurable). (c) If $\boldsymbol{\omega}_n \to \omega$, by continuity $\boldsymbol{\theta}_n = a(\boldsymbol{\omega}_n) \to a(\omega) = \theta$. □

We interpret 3.1 (i–iii) in terms of (f, Θ, F) under the plausible assumption that α and a are continuous (and thus that Θ and Ω are homeomorphic). Regarding 3.1 (i), we note that local compactness is a topological invariant. Hence Ω is locally compact iff Θ is. Condition 3.1 (ii) requires that $\eta = E_F\beta(\mathbf{X})$ exist and belong to $B^0(\Omega)$. Note that $B(\Omega)$ may be defined intrinsically in terms of f:

$$B(\Omega) = \{y \in R^m : \sup\{\alpha'(\theta)y - \gamma(\theta) : \theta \in \Theta\} < \infty\} .$$

Similarly, we have the intrinsic representation

$$\Omega(\nu) = \{\omega \in R^m : \int \exp\{\omega'\beta(x)\}\, d\mu(x) < \infty\};$$

other analogous expressions may be given as well. Assuming the existence of η, we interpret 3.1 (iii) by defining

$$g_F(\theta) = E_F \log f(\mathbf{X} \mid \theta) = q(\eta \mid \alpha(\theta)).$$

Clearly $q(\eta \mid \cdot)$ attains a unique maximum on Ω iff g_F attains a unique maximum on Θ. Let ω and θ denote the respective maximizing points; $\omega = \alpha(\theta)$. Moreover, $q(\eta \mid \cdot)$ is bounded below $q(\eta \mid \omega)$ off neighborhoods of ω iff g_F is similarly bounded. E.g., if U is an open neighborhood of θ, $\sup\{g_F(t) : t \notin U\} = q(\eta \mid aU)$ and aU is open if α is continuous.

The import of the preceding discussion is that the whole development can be done for (f, Θ, F) directly, without explicitly calculating (p, Ω, P). Such an approach is taken in Berk (1970), in a different context. We remark that 3.2 and 3.3 can be reformulated to apply directly to (f, Ω, F).

5. Examples. The following indicates for some specific models the nature of the conditions discussed above. The last example deals with the multinomial distribution. Although it does not fall precisely within the above formulation, the same technique can be used to analyze it.

EXAMPLE 1. $N(\xi, \sigma^2)$. We obtain the model $f(x|\xi,\sigma) = \exp\{-(x-\xi)^2/2\sigma^2\}/\sigma(2\pi)^{\frac{1}{2}}$, $(\xi, \sigma) \in \Theta \subset R \times (0, \infty)$ upon taking $\mathscr{X} = R$, $d\mu(x) = dx/(2\pi)^{\frac{1}{2}}$, $\beta(x) = (x, -x^2/2)$, $\alpha(\xi, \sigma) = (\xi/\sigma^2, 1/\sigma^2)$ and $\gamma(\xi, \sigma) = -(\xi^2/2\sigma^2 + \log \sigma)$. For the natural parameterization $m = 2$ and ν is supported on the parabola $(2y_2 = -y_1^2)$. $\Omega(\nu) = \{(\omega_1, \omega_2) : \omega_2 > 0\} = R \times (0, \infty)$ and $c(\omega) = (\omega_1^2/\omega_2 - \log \omega_2)/2$. It is straightforward to calculate that $B(\nu) = (y_1^2 < -2y_2) = C^0(\nu)$. (See, e.g. (6.2) of Berk (1970) or note that $\dot{c}(\Omega(\nu)) = \{E_{(\xi,\sigma)}\beta(\mathbf{X}) : (\xi, \sigma) \in \Theta\} = C^0(\nu)$ and see 2.3 (i, ii).) Since α is a homeomorphism, existence and consistency of the MLE for locally compact Θ follows from 3.2. Of course, for $\Theta = R \times (0, \infty)$, the MLE is easily exhibited and studied ad hoc.

An example where $\Omega \subset \Omega(\nu)$ is the family $N(\delta\theta, \theta^2)$, where $\delta \neq 0$ is fixed and θ ranges in $\Theta = R - \{0\}$. The mapping α takes Θ into $\Omega = \{(\omega_1, \omega_2) : \omega_1^2 = \delta^2\omega_2, \omega_2 \neq 0\}$. Ω is locally compact. Thus, assuming the model holds, $\mathbf{g}_n$ eventually attains a maximum on Θ. (In fact, $\mathbf{g}_n$ attains a unique maximum w.p.1 for $n \geqq 1$.) When the model holds, consistency follows from 3.3.

The likelihood $\mathbf{g}_n$ has two relative maxima, at $(-\delta\bar{\mathbf{X}}_n \pm (\delta^2\bar{\mathbf{X}}_n + 4\mathbf{V}_n)^{\frac{1}{2}})/2$, where $\bar{\mathbf{X}}_n = \sum_1^n \mathbf{X}_i/n$ and $\mathbf{V}_n = \sum_1^n \mathbf{X}_i^2/n$. The absolute maximum is at $\boldsymbol{\theta}_n = (-\delta\bar{\mathbf{X}}_n + (\operatorname{sgn} \delta\bar{\mathbf{X}}_n)\cdot(\delta^2\bar{\mathbf{X}}_n^2 + 4\mathbf{V}_n)^{\frac{1}{2}})/2$ and consistency is easy to verify directly.

If $\mathbf{X} \sim F$ and $E_F\mathbf{X} = \xi \neq 0$ and $\operatorname{Var}_F \mathbf{X} = \sigma^2 < \infty$, $\boldsymbol{\theta}_n$ converges w.p.1 to the obvious limit, which is the θ of 4.1. Note that unless $\xi/\sigma = \delta$, F does not resemble any distribution in the model in the sense of 3.2 (ii). Thus one

must appeal directly to 3.1 (or 4.1). If $\xi = 0$, 3.1 (iii) fails, for although $\mathbf{g}_F$ attains a maximum on Θ, it is not unique. Examination of $\boldsymbol{\theta}_n$ shows that w.p.1, $\limsup \boldsymbol{\theta}_n = \sigma$ and $\liminf \boldsymbol{\theta}_n = -\sigma$, so that there is no consistent behavior in this case. One would not expect to have consistency if $\mathbf{g}_F$ does not attain a unique maximum. This point is further illustrated in the following example.

EXAMPLE 2. Let ν be Lebesgue measure on $[0, 1]$. Then $p(y|\omega) = \omega e^{\omega y}/(e^{\omega} - 1)$, $0 \leqq y \leqq 1$, $\omega \in R = \Omega(\nu)$. Here $C^0(\nu) = (0, 1) = \dot{c}(\Omega(\nu))$, so also $B^0(\nu) = (0, 1)$ (see 2.3 (i, ii)). Suppose $\mathbf{Y} \sim P$ and let $\eta = E_P\mathbf{Y}$. If $\eta \in (0, 1)$, then $q(\eta \mid \omega) = \omega\eta + \log[\omega/(e^{\omega} - 1)]$ has a unique maximum on $\Omega(\nu)$ and the MLE is consistent. If $\eta \notin (0, 1)$, so that 3.2 (i) is violated, $q(\eta \mid \cdot)$ is unbounded. If $\eta \notin [0, 1]$, then w.p.1, $q(\bar{\mathbf{Y}}_n \mid \omega)$ is eventually unbounded, so that no MLE exists. If $\eta = 1$ and $P(\mathbf{Y} = 1) < 1$, $\bar{\mathbf{Y}}_n$ oscillates about 1, so that a MLE both exists and does not exist infinitely often.

EXAMPLE 3. *Gamma distribution.* We consider the model $f(x \mid \omega) = \omega_2^{\omega_1} x^{\omega_1} e^{-\omega_2 x}/\Gamma(\omega_1)$ for $x > 0$, with $d\mu(x) = dx/x$ on $\mathscr{X} = (0, \infty)$. The model is already naturally parametrized, although it is more convenient to work with $(\mathscr{X}, \mu)$ than (R^2, ν). Here $\Omega(\nu) = \{(\omega_1, \omega_2) : \omega_1 > 0, \omega_2 > 0\} = \Omega^0(\nu)$. Taking $\Omega = \Omega(\nu)$, if $\omega \in \Omega$ obtains, it follows from 3.3 that the MLE eventually exists and is consistent. It does not seem possible to explicitly exhibit the MLE for this model. Corresponding results hold for suitable $\Omega \subset \Omega(\nu)$. The case $\Omega = (\omega_2 = 1)$ is treated in the classical way by Cramér (1946) page 504 *ff*. Similar remarks apply to the family of beta distributions.

EXAMPLE 4. *Multinomial distribution.* The theorems of this paper do not apply directly to a multinomial model unless all of the cell probabilities are positive. With zero cell probabilities, the multinomial can be viewed as an extended exponential model, with some components of ω being allowed to assume the value $-\infty$. We essentially adopt this point of view, but work explicitly with the usual parameterization, the cell probabilities. With a bit of circumlocution to allow for the value $-\infty$, we use the preceding methods to establish consistency in this case.

The sample space for one multinomial observation $\mathbf{Y}$ is $K = \{(1, 0, \cdots, 0), (0, 1, \cdots, 0), \cdots, (0, 0, \cdots, 0, 1)\} \subset R^m$ and the parameter set is $\Pi = \{\pi \in R^m : 0 \leqq \pi_i, \sum_1^m \pi_i = 1\}$. Π also serves as the sample space for the sufficient statistic $\bar{\mathbf{Y}}_n$. We choose as dominating measure ν the counting measure on K. Then the likelihood function for the outcome $y \in \Pi$ is

$$q(y \mid \pi) = \log p(y \mid \pi) = \sum y_i \log \pi_i , \tag{5.1}$$

where we define $-\infty \cdot 0 = 0$. We note that for $y \in \Pi$, $q(y \mid \cdot)$ is concave and use on Π (in fact, strictly concave where finite). Although $q(\cdot \mid \pi)$ is linear

in y, it may not be continuous on Π. The pathology occurs if the set $(q(\cdot \mid \pi) = -\infty)$ is not empty (which happens only if some components of π vanish). Nevertheless, $q(\cdot \mid \pi)$ is continuous on the set where it is finite. We denote this last set by $\Delta(\pi) = \{y \in \Pi : \sum y_i \log \pi_i > -\infty\}$. Thus $\Delta(\pi)$ contains those distributions dominated by π. Clearly $\Delta(\pi)$ is a closed convex subset of Π. (In fact, $\Delta(\pi)$ is the convex hull of the finite set $\Delta(\pi) \cap K$.)

If $V \subset \Pi$, $q(y \mid V) = \sup\{q(y \mid \pi) : \pi \in V\}$ is finite on the set $\Delta(V) = \cup \{\Delta(\pi) : \pi \in V\}$. As $q(\cdot \mid V)$ is clearly convex, it therefore has the usual continuity properties of convex functions: it is continuous on the relative interior of any convex subset of $\Delta(V)$. In particular, $q(\cdot \mid V)$ is continuous on $\Delta^0(\eta)$, the relative interior of $\Delta(\eta)$, for all $\eta \in \Delta(V)$. (Note that $\Delta^0(\eta)$ may be characterized as those points of $\Delta(\eta)$ having positive barycentric coordinates, or as those distributions in Π equivalent to η.) The sets $\{\Delta^0(\eta) : \eta \in \Delta(V)\}$ form a finite partition of $\Delta(V)$ (into vertices, edges, faces, etc. of Π). Since $q(\cdot \mid V)$ is piecewise continuous on this finite partition, it is measurable.

We establish consistency along the lines of 3.1. We take $\Omega \subset \Pi$ locally compact. Suppose P, the actual distribution of $\mathbf{Y}$, is indexed by $\eta \in \Pi$. Suppose further that $q(\eta \mid \cdot)$ attains a unique maximum on Ω at π and is bounded below $q(\eta \mid \pi)$ off neighborhoods of π. (In particular, if $\eta \in \Omega$, this follows as in 3.2, with $\pi = \eta$.) Referring to the proof of 3.1, we take $V = \{v \in \Omega : |v - \pi| > \varepsilon\}$. Then w.p.1, $q(\bar{\mathbf{Y}}_n \mid V) \to q(\eta \mid V)$, because w.p.1 $\bar{\mathbf{Y}}_n \to \eta \in \Delta^0(\eta)$ and $q(\cdot \mid V)$ is continuous on $\Delta^0(\eta)$. The proof then continues as in 3.1 and we conclude that w.p.1 a MLE $\boldsymbol{\pi}_n \in \Omega$ eventually exists and (assuming measurability) $P(\boldsymbol{\pi}_n \to \pi) = 1$. If Ω is locally convex, then we conclude that $\boldsymbol{\pi}_n$ is eventually unique and measurable. Note that the analog to assumption 3.1 (ii) is unnecessary here. The fact that $\bar{\mathbf{Y}}_n$ is eventually in $\Delta^0(\eta)$ plays the role of that condition.

If $\alpha : \Theta \to \Omega \subset \Pi$ is a reparametrization of Ω, the assumption that $a = \alpha^{-1}$ be continuous (as in 4.1) assures the consistency of the MLE $\boldsymbol{\theta}_n = a(\boldsymbol{\pi}_n)$.

A result much like the preceding is given by Rao (1957, 1965). His method is somewhat different and he restricts attention to the case $\eta \in \Omega$. His condition for the eventual existence and consistency of an MLE is that for all $\eta \in \Omega$, for some $\varepsilon > 0$, $A_\varepsilon = \{\pi \in \Pi : q(\eta \mid \eta) - q(\eta \mid \pi) \leqq \varepsilon\} \subset \Omega$. Since A_ε contains an open neighborhood (in Π) of η, such an Ω is open in Π and therefore locally compact and locally convex. Thus under Rao's conditions, the MLE is eventually unique. It is straightforward to verify that Rao's demonstration holds under the weaker condition that Ω is locally compact. For the case of reparametrization, Rao's condition 5e. 2.1 (Rao, 1965) is just another way of stating that α^{-1} be continuous.

6. Asymptotic normality. With some further restrictions, we establish the

asymptotic normality of the MLE. The conditions are somewhat weaker than those used in more general settings. (E.g., Cramér (1946), page 500 *ff.*) Basically, we require that the likelihood be continuously differentiable in the parameter.

We work with model (4.1), where we now suppose that Θ is an open subset of $R^k (k \leqq m)$ and that $\Omega = \alpha(\Theta) \subset \Omega^0(\nu)$. (We retain the identifications made in Section 4.) We suppose further that α is continuously differentiable on Θ and that the matrix $A(\theta) = (\partial\alpha_i/\partial\theta_j)$, $i = 1, \cdots, m$, $j = 1, \cdots, k$ is of rank k at every point of Θ. (Because c is infinitely differentiable on $\Omega^0(\nu)$, any differentiability condition for α is equivalent to one for $f(x \mid \cdot)$.) We assume that F resembles the model in that for some (necessarily unique) $\theta \in \Theta$, $E_F\beta(\mathbf{X}) = E_\theta\beta(\mathbf{X}) = \dot{c}(\alpha(\theta))$. Finally we suppose that (eventually) there is a measurable MLE, say $\boldsymbol{\theta}_n$, based on $\mathbf{g}_n$ and $F(\boldsymbol{\theta}_n \to \theta) = 1$. We define

$$I(\theta) = A'(\theta)\ddot{c}(\alpha(\theta))A(\theta) . \tag{6.1}$$

Because A is of full rank and $\ddot{c}$ is nonsingular, I is also nonsingular. $I(\theta)$ is the Fisher information matrix for $f(\cdot \mid \theta)$. (I.e., let $\mathbf{g}(\theta) = \log f(\mathbf{X} \mid \theta) = \alpha'(\theta)\mathbf{Y} - c(\alpha(\theta))$. Then $\dot{\mathbf{g}}(\theta) = (\partial\mathbf{g}/\partial\theta_1, \cdots, \partial\mathbf{g}/\partial\theta_k)' = A'(\theta)[\mathbf{Y} - \dot{c}(\alpha(\theta))]$ and $I(\theta) = \mathrm{Cov}_\theta\, \dot{\mathbf{g}}(\theta)$.)

THEOREM 6.1. *Under the preceding conditions,*

$$n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta) \to_{\mathscr{L}} N(0, I^{-1}(\theta)) .$$

PROOF. Since $\mathbf{g}_n$ is differentiable on the open set Θ, it follows that $\boldsymbol{\theta}_n$ satisfies the likelihood equation $\dot{\mathbf{g}}(\boldsymbol{\theta}_n) = 0$. From (4.1) we see that $\dot{\mathbf{g}}_n(\theta) = \bar{\mathbf{Y}}_n A(\theta) - \gamma(\theta) = [\bar{\mathbf{Y}}_n - \dot{c}(\alpha(\theta))]A(\theta)$. Hence the likelihood equation becomes

$$[\bar{\mathbf{Y}}_n - \dot{c}(\alpha(\theta_n))]A(\boldsymbol{\theta}_n) = 0 . \tag{6.2}$$

Since $\boldsymbol{\theta}_n \to \theta$ w.p. 1, by continuity of A,

$$A(\boldsymbol{\theta}_n) = A(\theta) + o_p(1) , \tag{6.3}$$

where $o_p(1)$ denotes a matrix, each of whose terms converges to zero in probability as $n \to \infty$. (The analogous use of order symbols below will be clear from the context.) Similarly,

$$\alpha(\boldsymbol{\theta}_n) - \alpha(\theta) = (\boldsymbol{\theta}_n - \theta)A'(\theta) + o(|\boldsymbol{\theta}_n - \theta|) ,$$

so that

$$\begin{aligned} \dot{c}(\alpha(\boldsymbol{\theta}_n)) - \dot{c}(\alpha(\theta)) &= [\alpha(\boldsymbol{\theta}_n) - \alpha(\theta)]\ddot{c}(\alpha(\theta)) + o(|\alpha(\boldsymbol{\theta}_n) - \alpha(\theta)|) \\ &= (\boldsymbol{\theta}_n - \theta)A'(\theta)\ddot{c}(\alpha(\theta)) + o(|\boldsymbol{\theta}_n - \theta|) . \end{aligned} \tag{6.4}$$

The likelihood equation (6.2) implies that

$$[\bar{\mathbf{Y}}_n - \dot{c}(\alpha(\theta))]A(\boldsymbol{\theta}_n) = [\dot{c}(\alpha(\boldsymbol{\theta}_n)) - \dot{c}(\alpha(\theta))]A(\boldsymbol{\theta}_n) .$$

Multiplying by $n^{\frac{1}{2}}$ and then using (6.3) and (6.4) gives

$$
\begin{aligned}
& n^{\frac{1}{2}}[\bar{\mathbf{Y}}_n - \dot{c}(\alpha(\theta))][A(\theta) + o_p(1)] \\
& \quad = n^{\frac{1}{2}}[\dot{c}(\alpha(\boldsymbol{\theta}_n)) - \dot{c}(\alpha(\theta))][A(\theta) + o_p(1)] \\
& \quad = n^{\frac{1}{2}}[(\boldsymbol{\theta}_n - \theta)A'(\theta)\ddot{c}(\alpha(\theta)) + o(|\boldsymbol{\theta}_n - \theta|)][A(\theta) + o_p(1)] \\
& \quad = n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta)I(\theta)[J + o_p(1)]\,,
\end{aligned} \tag{6.5}
$$

where J denotes the identity matrix of order k. The central limit theorem applied to $\bar{\mathbf{Y}}_n$, together with Slutsky's theorem, shows that the first term in (6.5) is $O_p(1)$. Since the last term in (6.5) is $O_p(n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta))$, it follows that $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta)$ is also $O_p(1)$. Hence the last term in (6.5) is $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta)I(\theta) + o_p(1)$. Since $n^{\frac{1}{2}}[\bar{\mathbf{Y}}_n - \dot{c}(\alpha(\theta))] \to_{\mathscr{L}} N(0, \ddot{c}(\alpha(\theta)))$, we conclude from (6.5) that $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta)I(\theta) \to_{\mathscr{L}} N(0, I(\theta))$. Because I is invertible, $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta) \to_{\mathscr{L}} N(0, I^{-1}(\theta))$. □

The foregoing theorem applies to Examples 1–3 of Section 5. For the multinomial distribution, Rao (*op cit.*) obtained normality under much the same conditions as those of this section. The following example shows what can happen if some of the conditions assumed in this section are weakened.

Suppose $f(x \mid \theta)$ is the $N(\theta, 1)$ pdf and $\Theta = [0, \infty)$. It is easily seen that $\boldsymbol{\theta}_n = \bar{\mathbf{X}}_n$ if $\bar{\mathbf{X}}_n \geqq 0$ and $\boldsymbol{\theta}_n = 0$ if $\bar{\mathbf{X}}_n < 0$. Thus if $E_F\mathbf{X} < 0$, $\boldsymbol{\theta}_n$ eventually remains zero, so that $\boldsymbol{\theta}_n$ is consistent (to $\theta = 0$) but not asymptotically normal (except in some degenerate sense, perhaps). Note that the conditions of Theorem 4.1 are satisfied with $\theta = 0$. If $E_F X = 0$ and $E_F\mathbf{X}^2 < \infty$, then $F(\boldsymbol{\theta}_n = 0) = F(\bar{\mathbf{X}}_n \leqq 0) \to \frac{1}{2}$ and again there is consistency but not asymptotic normality. Here F does resemble $\theta = 0 \in \Theta$, but Θ is not open. If one takes $\Theta = (0, \infty)$, then no bona fide MLE exists when $\bar{\mathbf{X}}_n \leqq 0$. Almost-MLE's do exist and when $E_F\mathbf{X} \leqq 0$, exhibit behavior similar to the above. This example indicates that the asymptotic behavior of $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \theta)$ can change markedly if one relaxes either the condition that Θ be open or that F resemble some θ in Θ.

REFERENCES

[1] BERK, R. H. (1970). Consistency a posteriori. *Ann. Math. Statist* **41** 894–906.

[2] BUCK, R. C. (1956). *Advanced Calculus*. McGraw-Hill, New York.

[3] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

[4] EGGLESTON, H. G. (1958). *Convexity*. University Press, Cambridge.

[5] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[6] RAO, C. R. (1957). Maximum likelihood estimation for the multinomial distribution. *Sankhyā* **18** 139–148.

[7] RAO, C. R. (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.