The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 894–906

CONSISTENCY A POSTERIORI[1]

BY ROBERT H. BERK[2]

Hebrew University

0. Summary. Conditions are given under which a sequence of posterior distributions converges weakly to a degenerate distribution. It is not assumed that the model being used actually governs the data. Exponential models are studied in particular and are shown to be well behaved under mild assumptions. Large-deviation results for U-statistics and posterior odds are also given.

1. Introduction. Let $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \cdots$ be a sample sequence, independent and identically distributed random variables, having common distribution F on the (measurable) sample space $(\mathscr{X}, \mathscr{B})$. F will refer also to the joint distribution of the sequence. Let $p(x \mid \theta)$ be a family of probability densities on $(\mathscr{X}, \mathscr{B})$, with respect to the σ-finite measure μ. θ ranges in the measurable parameter space $(\Theta, \mathscr{A})$; we assume $p(\cdot \mid \cdot)$ is $\mathscr{A} \times \mathscr{B}$ measurable. P denotes a (possibly unnormed) prior measure on $(\Theta, \mathscr{A})$. We assume there is a probability measure Q on $(\Theta, \mathscr{A})$ equivalent to P. (This is so if, e.g., $(\Theta, \mathscr{A}, P)$ is σ-finite.) We consider the behavior of the sequence of posterior distributions $\mathbf{P}_n$ where, for $A \in \mathscr{A}$

$$\mathbf{P}_n A = \int_A \prod_1^n p(\mathbf{X}_i \mid \theta)\, dP / \int_\Theta \prod_1^n p(\mathbf{X}_i \mid \theta)\, dP. \tag{1.1}$$

Below we detail assumptions that guarantee that $\mathbf{P}_n$ is a.s. well defined. We consider the question of consistency a posteriori: the a.s. weak convergence of $\mathbf{P}_n$ to a degenerate distribution. The following heuristic discussion gives an indication of the approach to the problem and the results obtained. Here and throughout, expectations are under F.

Let $\boldsymbol{l}_n(\theta) = n^{-1} \sum_1^n \ln [p(\mathbf{X}_i \mid \theta) p^*(\mathbf{X}_i)]$, where p^* is some positive function. If p^* can be chosen so that $\lambda(\theta) = E \ln [p(\mathbf{X} \mid \theta) p^*(\mathbf{X})]$ exists, $F(\boldsymbol{l}_n(\theta) \to \lambda(\theta)) = 1$ and the exceptional set of data sequences can depend on θ. In Section 2 we show that this entails $F(\boldsymbol{l}_n \to \lambda[P]) = 1$, where the exceptional θ-set can depend on the observed data sequence. We may rewrite (1.1) as

$$\mathbf{P}_n A = \int_A \exp\{n \boldsymbol{l}_n\}\, dP / \int_\Theta \exp\{n \boldsymbol{l}_n\}\, dP. \tag{1.2}$$

$(\mathbf{P}_n A)^{1/n}$ is then seen to be a ratio of n-norms (of random functions). In the following heuristic argument, we replace $\boldsymbol{l}_n$ by λ in (1.2) to obtain

$$(P_n A)^{1/n} = [\int_A \exp\{n\lambda\}\, dP / \int_\Theta \exp\{n\lambda\}\, dP]^{1/n} = \|e^\lambda 1_A\|_n / \|e^\lambda\|_n. \tag{1.3}$$

Received August 20, 1968; revised October 14, 1969.

[1] Work supported by NSF grants GP-6008 and GP-7350. Revision carried out while the author was an NSF Postdoctoral Fellow.

[2] On leave from the University of Michigan.

P_n is a probability distribution and (to the extent that $\boldsymbol{l}_n$ resembles λ) resembles $\mathbf{P}_n$. From (1.3) we see that

$$(P_n A)^{1/n} \to \|e^{\lambda} 1_A\|_\infty / \|e^{\lambda}\|_\infty < 1 \text{ iff ess sup } \{\lambda(\theta): \theta \in A\} < \text{ess sup } \{\lambda(\theta): \theta \in \Theta\} = \lambda^*.$$

Hence if A is a set on which λ does not (essentially) achieve its essential supremum, $P_n A \to 0$. So if λ achieves its essential supremum λ^* at some point θ^* and is essentially bounded below λ^* off (open) neighborhoods of θ^*, then for any such neighborhood, U, $P_n U \to 1$. This says that P_n converges weakly to the distribution degenerate at θ^*.

This analysis of P_n entails the behavior of λ at its essential supremum. The analysis of $\mathbf{P}_n$ involves something more: Accounting for the extra complexity of (1.2), as compared with (1.3), entails the following: Given a sequence $\{v_n\}$ of measurable functions on Θ that converge $[P]$ to v, what can be said about the behavior of $\|v_n\|_n$? In particular, does $\|v_n\|_n \to \|v\|_\infty$? We consider this question and other preliminaries in Section 2. In Section 3 we give general results concerning the consistency of $\mathbf{P}_n$. Section 4 deals with the special cases of exponential and continuous models. There it is (implicitly) shown that under reasonable conditions, the answer to the last question above is affirmative for the functions that concern us. In Section 5 we give related large-deviation results for $\mathbf{P}_n$. Section 6 illustrates the theory with two examples: an exponential model, the univariate normal distribution with unknown mean and variance; a continuous but non-exponential model, the Cauchy distribution with unknown location. In the former case, the condition $0 < \operatorname{Var} \mathbf{X} < \infty$ is seen to assure for a large class of prior distributions that $\mathbf{P}_n$ converges weakly $[F]$ to the distribution degenerate at (ξ_F, σ_F), where $\xi_F = E\mathbf{X}$, $\sigma_F{}^2 = \operatorname{Var} \mathbf{X}$. In the latter case, $E(\log|\mathbf{X}|)^+ < \infty$ assures weak convergence $[F]$ of $\mathbf{P}_n$.

We introduce the following notation:

1.1 *Definitions and assumptions.* (a) p is jointly measurable.

(b) Let $C = (p > 0) \in \mathscr{A} \times \mathscr{B}$, C_θ is the cross-section at θ (C_θ is the carrier of $p(\cdot \mid \theta)$) and $\Theta_F = (FC_\theta = 1)$. Then $P\Theta_F > 0$. (The measurability of Θ_F is an easy consequence of Fubini's theorem.)

(c) $l(x \mid \theta) = \ln[p(x \mid \theta) p^*(x)]$, where $p^* > 0$. The choice of p^* is discussed below.

(d) $\boldsymbol{l}_n(\theta) = \sum_1^n l(\mathbf{X}_k \mid \theta)/n$.

(e) When p is an exponential model, i.e., $l(x \mid \theta) = \lambda_0(\theta) + \sum_1^m a_i(x)\lambda_i(\theta) = \lambda_0(\theta) + a'(x)\Lambda(\theta)$ ($'$ denotes transpose), we write $\mathbf{a}_n = \sum_1^n a(\mathbf{X}_k)/n$ and $\alpha_F = Ea(\mathbf{X})$. Thus $\boldsymbol{l}_n = \mathbf{a}_n{}'\Lambda + \lambda_0$.

(f) For any random variable $\boldsymbol{l}$, $E\boldsymbol{l}$ exists if $E\boldsymbol{l}^+$ and $E\boldsymbol{l}^-$ are not both ∞. We assume p^* in (c) may be chosen so that $P(El(\mathbf{X} \mid \theta)$ does not exist$) = 0$. Further conditions on p^* appear in Section 3.

(g) $\lambda(\theta) = El(\mathbf{X} \mid \theta)$. For an exponential model, $\lambda = \alpha_F{}'\Lambda + \lambda_0$.

(h) If v is a measurable real-valued function on Θ, for $A \in \mathscr{A}$ $P\sup_A v = P\operatorname{ess\,sup} \{v(\theta): \theta \in A)$, $P\|v\|_{An} = (\int_A |v|^n dP)^{1/n}$ and $P\|v\|_{A\infty} = P\sup_A |v|$. The subscript A is deleted when $A = \Theta$. $A' = \Theta - A$.

(i) $\lambda^* = P\sup\lambda$ and $A_\delta = (\lambda^* - \delta \leqq \lambda \leqq \lambda^*)$.

With the above notation, we may write (1.2) as

$$\mathbf{P}_n A = (P\|\exp \boldsymbol{l}_n\|_{An}/P\|\exp \boldsymbol{l}_n\|_n)^n. \tag{1.4}$$

2. Preliminaries. We consider first the convergence of $\boldsymbol{l}_n$ to λ.

LEMMA 2.1. *Assume Condition* 1.1(f). *Then* $F(\boldsymbol{l}_n \to \lambda[P]) = 1$, *where the exceptional* θ-*set can depend on the observed data sequence.*

PROOF. Let $(\Omega, \mathscr{F}) = (\Theta \times \mathscr{X}^\infty, \mathscr{A} \times \mathscr{B}^\infty)$. Here only, x is a point in $\mathscr{X}^\infty$. Let $C = (\boldsymbol{l}_n \to \lambda) \in \mathscr{F}$. The strong law of large numbers implies that for all $\theta \in \Theta$, $FC_\theta = 1 \Rightarrow (Q \times F)C = 1 \Rightarrow F(QC_x = 1) = 1$; i.e., $F(\boldsymbol{l}_n \to \lambda[P]) = 1$.

If p is an exponential model, an even stronger conclusion is possible. For by hypothesis, $Ea(\mathbf{X})$ exists, hence $F(\mathbf{a}_n \to Ea(\mathbf{X})) = 1 \Rightarrow F(\boldsymbol{l}_n \to \lambda$ everywhere$) = 1$. When Θ is a topological space, various continuity assumptions on $p(x \mid \cdot)$ can insure everywhere convergence w.p.1. The details are omitted.

A basic fact about the behavior of $P\|v_n\|_n$ is given by

LEMMA 2.2. $v_n \to v[P] \Rightarrow \liminf P\|v_n\|_n \geqq P\|v\|_\infty$.

PROOF. Since $v_n \to v[P]$, $v_n \to v[Q]$ and since Q is finite, $v_n \to_Q v$. Choose $\varepsilon > 0$. Let $Q(|v| > P\|v\|_\infty - \varepsilon/2) = 2\delta > 0$. For n sufficiently large, $Q(|v_n - v| > \varepsilon/2) < \delta$ and then $Q(|v_n| > P\|v\|_\infty - \varepsilon) > \delta$. It then follows that for n sufficiently large, $P(|v_n| > P\|v\|_\infty - \varepsilon)$ is bounded away from zero, by $\gamma < \infty$, say. (For suppose $QC_n > \delta$ for all n while $PC_n \to 0$. By passing to a subsequence, we may suppose $\sum PC_n < \infty$. Letting $D_n = \cup_{k>n} C_n$, $PD_n \downarrow 0$ while $QD_n > \delta$. Letting $D = \cap D_n$, $PD = 0$ while $QD \geqq \delta$, contradicting the equivalence of P and Q.) Finally,

$$P\|v_n\|_n \geqq (\textstyle\int_{(|v_n| > P\|v\|_\infty - \varepsilon)} |v_n|^n \, dP)^{1/n} > (P\|v\|_\infty - \varepsilon)\gamma^{1/n} \to P\|v\|_\infty - \varepsilon. \quad \square$$

We consider next an extension of the notion "$P\sup$" to random functions. As the conclusions depend only on the null sets of P, we can suppose $P\Theta = 1$. Let $0 \leqq v(x \mid \theta)$ be jointly measurable. For $x \in \mathscr{X}$, let $h(x) = P\sup v(x \mid \theta)$. We note that h is measurable since $h = \lim h_n$, where $h_n(x) = (\int v^n(x \mid \theta)\, dP)^{1/n}$. Thus we may define $P\sup v(\mathbf{X} \mid \theta) = h(\mathbf{X})$. Let $N = (v(x \mid \theta) \leqq h(x)) \in \mathscr{A} \times \mathscr{B}$. Letting $N_x \in \mathscr{A}$ be the cross-section at x, $\forall x \in \mathscr{X}$, $PN_x = 1$. Hence $(P \times F)N = 1$, implying $P(FN_\theta = 1) = 1$. I.e.,

$$P(v(\mathbf{X} \mid \theta) \leqq h(\mathbf{X}) \text{ w.p.1}) = 1. \tag{2.1}$$

Parenthetically, we also note the following:

If $\{\mathbf{Y}_\theta : \theta \in \Theta\}$ is a collection of random variables defined on $(\mathscr{X}, \mathscr{B}, F)$, it is well known that there is a random variable $\mathbf{Y} = F\sup \mathbf{Y}_\theta$ satisfying: $\forall\theta$, $\mathbf{Y}_\theta \leqq \mathbf{Y}[F]$ and if $\forall\theta$, $\mathbf{Y}_\theta \leqq \mathbf{Y}'[F]$, then $\mathbf{Y} \leqq \mathbf{Y}'[F]$. Moreover, $\mathbf{Y}$ is the actual supremum of a denumerable subfamily of $\{\mathbf{Y}_\theta\}$. Let $F\sup v(\mathbf{X} \mid \theta) = k(\mathbf{X})$ and $N = (v(x \mid \theta) \leqq k(x)) \in \mathscr{A} \times \mathscr{B}$. Then $\forall\theta \in \Theta, FN_\theta = 1$, hence $F(PN_x = 1) = 1$ or $F(v(x \mid \theta) \leqq$

$k(x)[P]) = 1$, which implies $F((\int v^n(x \mid \theta)\, dP)^{1/n} \leqq k(x)) = 1$. I.e., $h_n(\mathbf{X}) \leqq k(\mathbf{X})[F]$, hence

$$P \sup v(\mathbf{X} \mid \theta) \leqq F \sup v(\mathbf{X} \mid \theta)[F]. \tag{2.2}$$

A parallel development holds for $P \sup_A$, for $A \in \mathscr{A}$.

An immediate consequence of Lemma 2.2 is

COROLLARY 2.3. *If* $\mathbf{v}, \mathbf{v}_1, \cdots$ *are random functions* (*i.e.* $\mathrm{v} = v(\mathbf{X}, \theta)$, *etc.*) *and* $\mathrm{F}(\mathbf{v}_n \to \mathbf{v}[P]) = 1$, *then* $F(\liminf P\|\mathbf{v}_n\|_n \geqq P\|\mathbf{v}\|_\infty) = 1$.

We also require the following notion.

DEFINITION 2.4. A sequence of events $\{B_n\}$ is said to be exponentially bounded if for some $c > 0$, $\rho < 1$, $FB_n < c\rho^n$. A random variable $\mathbf{Y}$ is exponentially bounded if the events $(|\mathbf{Y}| > n)$ are. Equivalently, for some $r > 0$; $Ee^{r|\mathbf{Y}|} < \infty$. We note that if $\mathbf{Y}$ is exponentially bounded, so is $r\mathbf{Y}$ for every real r. If $\{B_n\}$ is exponentially bounded, then $F(B_n \text{ i.o.}) = 0$. Hence if $\mathbf{s} =$ last time B_n occurs, $F(\mathbf{s} < \infty) = 1$. From the relation $FB_n \leqq F(\mathbf{s} \geqq n) \leqq \sum_n^\infty FB_k$, we see that $\{B_n\}$ is exponentially bounded $\Leftrightarrow \mathbf{s}$ is exponentially bounded. See Section 3 of Wijsman (1968) for a more detailed discussion of this notion.

3. Weak convergence of $\mathbf{P}_n$. We consider first a criterion that assures $\mathbf{P}_n$ is eventually well defined. We note first that 1.1(b) assures that a.s., the denominator in (1.1) does not vanish. We may argue this as follows: Assume first $\Theta_F = \Theta$. Then 1.1(b) implies $(F \times Q)C = 1$; hence $F(\int p(x \mid \theta)\, dP > 0) = 1$. A similar argument shows that $\int \prod_1^n p(\mathbf{X}_i \mid \theta)\, dP > 0$ w.p.1. If $\Theta_F \subset \Theta$, replace Θ by Θ_F, C by $C \cap (\mathscr{X} \times \Theta_F)$ and renormalize Q. Then $\int_{\Theta_F} \prod_1^n p(\mathbf{X}_i \mid \theta)\, dP > 0$ w.p.1.

To avoid pathological behavior, the denominator in (1.1) must eventually become finite. To this end, we consider the following. For $n = 0, 1, \cdots$,

$$\mathbf{Z}_n = \int \exp\{n\boldsymbol{l}_n\}\, dP. \tag{3.1}$$

Note that $\mathbf{Z}_0 = P\Theta$. Let $\mathbf{S}$ be the first time $\mathbf{Z}_n < \infty$; $\mathbf{S} = +\infty$ if no such n occurs. Let $\boldsymbol{\sigma}$ be the last time $\mathbf{Z}_n = +\infty$; we take $\boldsymbol{\sigma}$ to be zero if all the $\mathbf{Z}_n$ are finite and to be $+\infty$ if all are infinite. $\mathbf{S}$ is a stopping time on the sequence $\mathbf{X}_1, \mathbf{X}_2, \cdots$ and $\boldsymbol{\sigma}$ is a reverse stopping time. $\mathbf{S} \leqq \boldsymbol{\sigma} + 1$.

DEFINITION 3.1. P becomes proper if $F(\mathbf{S} < \infty) = 1$. P remains proper if $F(\boldsymbol{\sigma} < \infty) = 1$. The following result, although it does not cover all cases of interest, is basic.

THEOREM 3.2. *Suppose* $F \ll \mu$. *Then the following are equivalent.* (i) *P remains proper.* (ii) *P becomes proper.* (iii) *For some* $s \geqq 0$, $F(\mathbf{Z}_s < \infty) > 0$. (iv) $\boldsymbol{\sigma}$ *is exponentially bounded.*

PROOF. Clearly (i)$\Rightarrow$(ii). Also, (ii) clearly entails (iii). Suppose then that (iii) holds. We note first that if $P\Theta < \infty$,

$$\int \prod_1^n p(x_i \mid \theta)\, d\mu^n = 1 \Rightarrow \mathbf{Z}_n = \int \prod_1^n p(\mathbf{X}_i \mid \theta) p^*(\mathbf{X}_i)\, dP < \infty[\mu],$$

hence a.s. $[F]$. Thus the event $(\boldsymbol{\sigma} > ns)$ is seen to entail the event

$$\left(\int \exp\{\textstyle\sum_{ks+1}^{ks+s} l(\mathbf{X}_i \mid \theta)\}\, dP = +\infty,\ k = 0, 1, \cdots, n-1\right).$$

Letting $F(\mathbf{Z}_s < \infty) = \beta > 0$, we see that $F(\boldsymbol{\sigma} > ns) \leqq (1-\beta)^n$, showing that $\boldsymbol{\sigma}$ is exponentially bounded. Hence (iii)⇒(iv). Finally, if $\boldsymbol{\sigma}$ is exponentially bounded, $F(\boldsymbol{\sigma} < \infty) = 1$, so that P remains proper. □

We will want to consider situations in which the condition $F \ll \mu$ is not necessarily satisfied. For that reason, further results on P becoming proper are given below (Propositions 4.3 and 4.4).

The following theorem is our basic convergence result for $\mathbf{P}_n$. Under suitable conditions (discussed below), it entails the weak convergence of $\mathbf{P}_n$.

THEOREM 3.3. *Choose* $A \in \mathscr{A}$. *If* (i) 1.1(f) *holds and* $\lambda^* > -\infty$, (ii) $F(\limsup(P \sup_A \boldsymbol{l}_n) < \lambda^* - \gamma)$ *for some* $\gamma > 0$ *and* (iii) *P becomes proper, then* $F(\mathbf{P}_n A \to 0) = 1$.

REMARK. If condition 1.1(b) fails, $\lambda \equiv -\infty[P]$, so that $\lambda^* = -\infty$. We also note that $\limsup(P \sup_A \boldsymbol{l}_n)$ is symmetric in $\mathbf{X}_1, \mathbf{X}_2, \cdots$ and, by the Hewitt–Savage 0–1 law, is a.s. constant. It is thus no more general to assume $F(\limsup(P \sup_A \boldsymbol{l}_n) < \lambda^*) = 1$. Before proving the theorem, we establish

LEMMA 3.4. *If P becomes proper and for* $A \in \mathscr{A}$, $F(\limsup(P \sup_A \boldsymbol{l}_n) = a) = 1$, *then* $F(\limsup P\|\exp \boldsymbol{l}_n\|_{An} \leqq e^a) = 1$.

PROOF. Let $\mathbf{P} = \mathbf{P}_S$; $\mathbf{P} \ll P$. ($\mathbf{P}$ is the first $\mathbf{P}_n$ that is well defined.) Let $\mathbf{Z} = \mathbf{Z}_S$ and on $(\mathbf{S} < n)$, $\boldsymbol{l}_n' = \sum_{S+1}^n l(\mathbf{X}_i \mid \theta)/n$. $F(0 < \mathbf{Z} < \infty) = 1$. Then on $(\mathbf{S} < n)$,

$$P\|\exp \boldsymbol{l}_n\|_{An} = \mathbf{Z}^{1/n}\mathbf{P}\|\exp \boldsymbol{l}_n'\|_{An} \leqq \mathbf{Z}^{1/n} \exp\{P \sup_A \boldsymbol{l}_n'\}[F]. \tag{3.2}$$

Conditional on $\mathbf{S}$, the behavior of $\boldsymbol{l}_n'$ is like that of $\boldsymbol{l}_n$, so by the hypothesis, since $F(\mathbf{S} < \infty) = 1$, $F(\limsup(P \sup_A \boldsymbol{l}_n') = a \mid \mathbf{S}) = 1[F]$.

This, together with (3.2), implies

$$F(\limsup P\|\exp \boldsymbol{l}_n\|_{An} \leqq e^a) = 1. \quad □ \tag{3.3}$$

PROOF OF THEOREM. By Theorem 2.1, $F(\boldsymbol{l}_n \to \lambda[P]) = 1$. We then see from Corollary 2.3 that

$$F(\liminf P\|\exp \boldsymbol{l}_n\|_n \geqq P\|\exp \lambda\|_\infty = \exp \lambda^*) = 1. \tag{3.4}$$

From Lemma 3.4 we see that $F(\limsup P\|\exp \boldsymbol{l}_n\|_{An} < \exp\{\lambda^* - \gamma\}) = 1$. Together, these imply $F(\limsup(\mathbf{P}_n A)^{1/n} < e^{-\gamma}) = 1$; hence $F(\lim \mathbf{P}_n A = 0) = 1$. □

When $A = A_\delta'$, the following stronger condition often holds:

$$F(\limsup(P \sup_A \boldsymbol{l}_n) \leqq P \sup_A \lambda \leqq \lambda^* - \delta) = 1. \tag{3.5}$$

Provided $\lambda^* < \infty$, we then have $F(\mathbf{P}_n A_\delta \to 1) = 1$. Sufficient conditions for this are given in Section 4. An important consequence appears in Corollary 3.6.

Suppose now that Θ is a Hausdorff topological space and that the open sets are in $\mathscr{A}$. A sequence $\{P_n\}$ of measures on $(\Theta, \mathscr{A})$ is said to converge weakly to P^* if for all bounded continuous $v: \Theta \to R$, $\int v\, dP_n \to \int v\, dP^*$. We further say

DEFINITION 3.5. A collection $\mathscr{A}^*$ of subsets of Θ is a weak base at $\theta^* \in \Theta$ if for every neighborhood U of θ^*, there is a set $A \in \mathscr{A}^*$ so that $A \subset U$.

REMARK. Since Θ is Hausdorff, if $\mathscr{A}^*$ is a weak base at θ^*, $\cap\{A : A \in \mathscr{A}^*\} \subset \{\theta^*\}$.

COROLLARY 3.6. *If* $F(\lim \mathbf{P}_n A_\delta = 1) = 1$ *for all* $\delta > 0$ *and if* $\{A_\delta : \delta > 0\}$ *is a weak base at* $\theta^* \in \Theta$, *then* w.p.1, $\mathbf{P}_n$ *converges weakly to* P^*, *the probability distribution degenerate at* θ^*.

PROOF. Let $\mathscr{U}$ be the open neighborhoods of θ^*. Since $\{A_\delta : \delta > 0\}$ is a weak base, for every $U \in \mathscr{U}$, there is an $A_{1/k} < U$; k a positive integer. $F(\mathbf{P}_n A_{1/k} \to 1, k = 1, 2, \cdots) = 1$, hence $F(\mathbf{P}_n U \to 1, U \in \mathscr{U}) = 1$. This last condition entails the a.s. weak convergence of $\mathbf{P}_n$ to P^*. □

To be assured of weak convergence, two things must be checked. First, whether the condition of Theorem 3.3 holds for A_δ' (for sufficiently small δ) and second, whether the A_δ form a weak based at some θ^*. The second condition is perhaps easier to verify, as it depends only on the nature of λ. For many models, λ is continuous, so that the A_δ are closed sets. If they are also compact for sufficiently small δ and decrease to $\{\theta^*\}$, they form a weak base at θ^*. If Θ is locally compact, a necessary and sufficient condition that the A_δ form a weak base at θ^* is that for sufficiently small δ, $\{cl\, A_\delta\}$ form a nested system of compacts decreasing to $\{\theta^*\}$. Or, if Θ is metric, then diam $A_\delta \downarrow 0$ and $cl\, A_\delta \downarrow \{\theta^*\}$ is necessary and sufficient. If the A_δ form a weak base at θ^*, then $\lim_{\theta \to \theta^*} \lambda(\theta) = \lambda^*$; if λ is continuous too, $\lambda^* = \lambda(\theta^*)$.

Verifying the condition of Theorem 3.3 for A_δ' is usually more difficult. It is insured by Wald's (1949) classical conditions and variants (Berk (1966), Huber 1967)) for the consistency of the maximum likelihood estimator. At the same time, these conditions insure that the A_δ form a weak base at some θ^*. (See the discussion below on continuous models).

4. Special structures. In this section we discuss how the foregoing applies to certain exponential and continuous models. In the former case, it is relatively easy to give simple conditions that insure consistency a posteriori.

Suppose p is an exponential model: $l(x \mid \theta) = a'(x)\Lambda(\theta) + \lambda_0(\theta)$. Associated with p are certain convex functions: For $\alpha = (\alpha_1, \cdots, \alpha_m) \in R^m$, let

$$L_A(\alpha) = P \sup_A \{\alpha'\Lambda(\theta) + \lambda_0(\theta)\}.$$

We write $L_\Theta = L$. The theory in Section 2 shows that L_A is measurable. It is readily seen that L_A is subadditive and homogeneous of order one and is hence convex. Then $D = (L < \infty)$ is a convex cone. Theorem 24 of Eggleston (1968) shows that L is continuous on D°, the relative interior of D (considered as a subset of $\mathscr{L}(D)$, its linear span, with the relative topology). We note that

$$P \sup_A \boldsymbol{l}_n = L_A(\mathbf{a}_n). \tag{4.1}$$

PROPOSITION 4.1. *Suppose* (i) $F(a(\mathbf{X}) \in \mathscr{L}(D)) = 1$, (ii) $E|a_i(\mathbf{X})| < \infty$, $i = 1, \cdots, m$ *and* (iii), $\alpha_F = Ea(\mathbf{X}) \in D^\circ$. *Then for any* $A \in \mathscr{A}$, $F(P \sup_A \boldsymbol{l}_n \to P \sup_A \lambda) = 1$.

PROOF. Since $a(\mathbf{X}) \in \mathscr{L}(D)$ w.p.1., the same is true of $\mathbf{a}_n$, for all n. Furthermore, since w.p.1. $\mathbf{a}_n \to \alpha_F \in D^\circ$, w.p.1. $\mathbf{a}_n$ is eventually in D°. Choose $A \in \mathscr{A}$. Since $L_A \leqq L$, $D_A = (L_A < \infty) \supset D$. Since L_A is continuous on $(D_A)^\circ \supset D^\circ$ and $\mathbf{a}_n$ is eventually in D°, w.p.1. $L_A(\mathbf{a}_n) \to L_A(\alpha_F)$. I.e. (see (4.1)), $F(P \sup_A \mathbf{l}_n \to P \sup_A \lambda) = 1$. □

COROLLARY 4.2. *Under the hypothesis of Proposition* 4.1, *if P becomes proper*, $\forall \delta > 0$, $F(\mathbf{P}_n A_\delta \to 1) = 1$.

PROOF. For an exponential model, $\Theta_F = \Theta$, while under Proposition 4.1(ii), $El(\mathbf{X} \mid \theta)$ exists for all $\theta \in \Theta$ and $\lambda^* > -\infty$. Writing $A_\delta' = A$, Proposition 4.1 implies that $F(\lim (P \sup_A \mathbf{l}_n) = P \sup_A \lambda \leqq \lambda^* - \delta) = 1$. Since $\alpha_F \in D^\circ$, $\lambda^* = L(\alpha_F) < \infty$, hence the conditions of Theorem 3.2 hold. □

REMARK. It is usually easy to verify directly whether the A_δ form a weak base at some $\theta^* \in \Theta$. If so, the conclusion of the corollary can be strengthened to $F(\mathbf{P}_n \to_w P^*) = 1$. Such a verification is illustrated in Section 6.

For exponential models, it is often easy to determine that P becomes proper via the following. Let

$$h_n(\alpha) = \int \exp\{n(\lambda_0 + \alpha'\Lambda)\}\, dP \tag{4.2}$$

and $H_n = (h_n < \infty)$. h_n is easily seen to be convex, hence so is H_n. Let $H = \liminf H_n = \bigcup_n I_n$, where $I_n = \bigcap_{k>n} H_k$. Since H is the union of increasing convex sets, H is also convex.

PROPOSITION 4.3. *If* (i) $F(a(\mathbf{X}) \in \mathscr{L}(H)) = 1$, (ii) $E|a_i(\mathbf{X})| < \infty$, $i = 1, \cdots, m$ *and* $\alpha_F \in H^\circ$, *then P remains proper.*

PROOF. As in the proof of Proposition 4.1, we may conclude that w.p.1., $\mathbf{a}_n$ eventually remains in H°. Let $\mathbf{s}-1$ be the last time $\mathbf{a}_n \notin H$. Then $F(\mathbf{s} < \infty) = 1$. Since $H = \bigcup I_n$, there is a random positive integer $\mathbf{t}$ so that $\mathbf{a}_s \in I_t$ w.p.1. $F(\mathbf{t} < \infty) = 1$. Hence on $(\mathbf{s} \vee \mathbf{t} < n)$, by the definition of $\mathbf{s}$, $\mathbf{a}_n \in I_t \subset I_n$. Thus $\mathbf{Z}_n = h_n(\mathbf{a}_n) < \infty$. We thus see that $\boldsymbol{\sigma} \leqq \mathbf{s} \vee \mathbf{t}$; hence $F(\boldsymbol{\sigma} < \infty) = 1$. □

REMARK. The combined hypotheses of Propositions 4.1 and 4.3 require that w.p.1, $a(\mathbf{X}) \in \mathscr{L}(H) \cap \mathscr{L}(D) = \mathscr{L}(H \cap D)$. We note that $H_n \cap D = I_n \cap D$, since if $\exp\{\lambda_0 + \alpha'\Lambda\}$ is essentially bounded and its nth power is integrable, so is any higher power. Thus we have the simpler representation: $H \cap D = \bigcup (H_n \cap D)$.

Although not used in the sequel, we include here another result similar to Proposition 4.3. By strengthening Assumption 4.3(i), it is possible to drop assumption 4.3(ii) and still conclude that $\boldsymbol{\sigma}$ is finite a.s. In fact, that it is exponentially bounded. Let $\bar{H}$ denote the closure of H in $\mathscr{L}(H)$.

PROPOSITION 4.4. *If for some positive integer s*, $F(\mathbf{a}_s \in \bar{H}) = 1$ *and* $F(\mathbf{a}_s \in H^\circ) > 0$, *then* $\boldsymbol{\sigma}$ *is exponentially bounded.*

PROOF. Let $\mathbf{t}$ be the first $n \geqq s$ for which $\mathbf{a}_n \in H^\circ$, or be $+\infty$ if no such n occurs. For $n \geqq s$, $\mathbf{a}_n$ is an average of $\binom{n}{s}$ points, all a.s. in $\bar{H}$ and hence is itself a.s. in $\bar{H}$.

To wit: $\mathbf{a}_n$ is the mean of the averages of the $\binom{n}{s}$ possible selections of s from $\{a(\mathbf{X}_1), \cdots, a(\mathbf{X}_n)\}$. Moreover, if any one of these $\binom{n}{s}$ averages is in H°, $\mathbf{a}_k \in H^\circ$ a.s. for all $k \geqq n$. Thus the event $(\mathbf{t} > ns)$ is seen to entail the event

$$\left(\textstyle\sum_{ks+1}^{ks+s} a(\mathbf{X}_i)/s \notin H^\circ,\ k = 0, 1, \cdots, n-1\right).$$

Letting $F(\mathbf{a}_s \in H^\circ) = \beta > 0$, $F(\mathbf{t} > ns) \leqq (1-\beta)^n$, showing that $\mathbf{t}$ is exponentially bounded. We also see from the above that $\boldsymbol{\sigma} \leqq \mathbf{t}$. □

Under the following conditions we may also verify Condition (ii) of Theorem 3.3, for $A = A_\delta'$.

ASSUMPTIONS 4.5. (a) Θ is a separable metric space and $\mathscr{A}$ contains the open sets. (b) $p(\mathbf{X} \mid \theta)$ is a.s. continuous in θ. (c) For all $\theta \in \Theta$ there is a neighborhood U of θ so that $EP\sup_U |l(\mathbf{X}|\cdot)| < \infty$. (d) $\lambda^* > -\infty$ and for $\delta > 0$ there is an integer $s \geqq 1$ and a compact $K \subset \Theta$ so that $EP\sup_{K'} \boldsymbol{l}_s < \lambda^* - \delta$.

PROPOSITION 4.6. *Conditions* (i) *and* (ii) *of Theorem* 3.3 *are met if, in addition to* Assumptions 4.5, 1.1(f) *holds.*

PROOF. (A fuller treatment of this demonstration may be found in Berk (1966).) Condition (i) is subsumed by the hypotheses. Write $A_\delta' = A$. Choose $\delta > 0$ and then s and K as in Assumption 4.5(d). For $\theta \in K'$, $\lambda(\theta) = E\boldsymbol{l}_s(\theta) < \lambda^* - \delta$; hence $A \supset K'$, or $A = (A \cap K) \cup K'$. We note now that the set of real-valued continuous functions on K is a separable Banach space under $P\|\cdot\|_{K\infty}$. (This norm is the ordinary sup-norm over the compact subset of K that supports P.) By the strong law of large numbers for Banach-valued random variables,

$$F(P\|\boldsymbol{l}_n - \lambda\|_{K\infty} \to 0) = 1.$$

(It must be checked that $EP\|l(\mathbf{X}|\cdot)\|_{K\infty} < \infty$. Since K is compact, this follows from Assumption 4.5(c).) For $\theta \in K'$, $\boldsymbol{l}_n(\theta) \leqq \mathbf{u}_n$, where $\mathbf{u}_n$ is the U-statistic based on $\mathbf{X}_1, \cdots, \mathbf{X}_n$ formed from the kernel $u(x_1, \cdots, x_s) = P\sup_{K'} \sum_1^s l(x_i \mid \theta)/s$. As shown in Berk (1966), $F(\mathbf{u}_n \to E\mathbf{u}_s < \lambda^* - \delta) = 1$. Since $P\sup_{K'} \boldsymbol{l}_n \leqq \mathbf{u}_n$,

$$F(\limsup (P\sup_{K'} \boldsymbol{l}_n) < \lambda^* - \delta) = 1. \tag{4.3}$$

Also, by (4.2),

$$F(P\sup_{A \cap K} \boldsymbol{l}_n \to P\sup_{A \cap K} \lambda \leqq \lambda^* - \delta) = 1. \tag{4.4}$$

Together, (4.3) and (4.4) imply $F(\limsup (P\sup_A \boldsymbol{l}_n) \leqq \lambda^* - \delta) = 1$. It follows from Assumptions 4.5(c, d) that $\lambda^* < \infty$, hence Condition (ii) of Theorem 3.3 holds. □

5. Large deviations. Under further conditions, Theorem 3.3 can be strengthened to give a large deviation result for $n^{-1} \log \mathbf{P}_n A$. We consider first exponential models.

Taking D and H_n as in Section 4, let $\Delta_n = D \cap H_n$ and $\Delta = D \cap H$. Δ_n and Δ are convex and, as remarked after Proposition 4.3, $\Delta_n \uparrow \Delta$. For any $A \in \mathscr{A}$, L_A is continuous on Δ°.

THEOREM 5.1. *Suppose* (i) $F(a(\mathbf{X})\in\mathscr{L}(\Delta))=1$, (ii) *for* $i=1,\cdots,m$, $a_i(\mathbf{X})$ *is exponentially bounded and* (iii) $\alpha_F\in\Delta^\circ$. Then for all $\gamma<1$, $\{(\mathbf{P}_nA_\delta'>e^{-n\gamma\delta})\}$ *is exponentially bounded.*

PROOF. We first argue that there is a positive integer t so that $\alpha_F\in\Delta_t^\circ$ and $\mathscr{L}(\Delta_t)=\mathscr{L}(\Delta)$. For since $\alpha_F\in\Delta^\circ$, there is a neighborhood, U, of α_F, open in $\mathscr{L}(\Delta)$, so that $U\subset\Delta$. Because Δ is convex, we may suppose that U is the interior of the convex hull of k points, $\alpha_1,\cdots,\alpha_k$, say. Here $k-1\leqq m$ is the dimensionality of Δ. Since $\alpha_1,\cdots,\alpha_k\in\Delta$ and $\Delta_n\uparrow\Delta$, there is a positive integer t so that $\alpha_1,\cdots,\alpha_k\in\Delta_t$. It follows that $U\subset\Delta_t$. Thus $\alpha_F\in\Delta_t^\circ$ and $\mathscr{L}(\Delta_t)=\mathscr{L}(U)=\mathscr{L}(\Delta)$.

Write $A_\delta'=A$ and let $N=\{\alpha:|\alpha_i-\alpha_{Fi}|<\varepsilon,\ i=1,\cdots,m\}\cap\mathscr{L}(\Delta_t)$, where ε is chosen to satisfy three conditions: $N\subset\Delta_t^\circ$ (possible, since $\alpha_F\in\Delta_t^\circ$), $\varepsilon<\beta\delta/Mm$ ($\beta>0$ and $M>0$ are specified below) and for $\alpha\in N$, $L_A(\alpha)<L_A(\alpha_F)+\beta\delta\leqq\lambda^*-\delta+\beta\delta$.

From (1.4) we see that

$$(\mathbf{P}_nA)^{1/n}=P\|\exp \boldsymbol{l}_n\|_{An}/P\|\exp \boldsymbol{l}_n\|_n. \tag{5.1}$$

Considering first the denominator, let $U=(|\lambda_i|<M,\ i=1,\cdots,m)$. Since $PA_{\beta\delta}>0$, by choosing M large enough, we can assure that $P(U\cap A_{\beta\delta})>0$. Let $V=U\cap A_{\beta\delta}$ and $v=PV$. Then on V, $|\boldsymbol{l}_n-\lambda|\leqq\sum_1^m|\mathbf{a}_{ni}-\alpha_{Fi}|\ |\lambda_i|\leqq Mm\sup_i|\mathbf{a}_{ni}-\alpha_{Fi}|$. Hence $a_n\in N$ implies $|\boldsymbol{l}_n-\lambda|\leqq\beta\delta$ on V, which implies $P\|\exp\boldsymbol{l}_n\|_n\geqq(\int_V\exp n\boldsymbol{l}_n\}\,dP)^{1/n}\geqq\exp\{\lambda^*-2\beta\delta\}v^{1/n}$. Hence for n sufficiently large, $\mathbf{a}_n\in N$ implies

$$P\|\exp \boldsymbol{l}_n\|_n\geqq\exp\{\lambda^*-3\beta\delta\}. \tag{5.2}$$

Considering now the numerator of (5.1), let $\eta=\sup_N h_t(\alpha)<\infty$. Note also that by adjusting p^* if necessary, we may suppose $\lambda^*>\delta$. Then $\mathbf{a}_n\in N$ implies

$$\begin{aligned}\int_A\exp\{n\boldsymbol{l}_n\}\,dP&=\int_A\exp\{n\mathbf{a}_n'\Lambda+\lambda_0\}\,dP\\&\leqq\exp\{(n-t)L_A(\mathbf{a}_n)\}\int_A\exp t\{\mathbf{a}_n'\Lambda+\lambda_0\}\,dP\\&\leqq\exp\{(n-t)L_A(\mathbf{a}_n)\}\eta<\exp\{n(\lambda^*-\delta+2\beta\delta)\}\end{aligned}\tag{5.3}$$

for n sufficiently large, since on N, $L_A<\lambda^*-\delta+\beta\delta$. Thus (5.1), (5.2) and (5.3) show that if $\mathbf{a}_n\in N$, $\mathbf{P}_nA\leqq e^{-n\delta(1-5\beta)}$.

Since $a(\mathbf{X})\in\mathscr{L}(\Delta)=\mathscr{L}(\Delta_t)$ w.p.1, $\mathbf{a}_n\in\mathscr{L}(\Delta_t)$ w.p.1 for all n. The usual large deviation result for sample means then shows that $\{(\mathbf{a}_n\notin N)\}$ is exponentially bounded. Thus, choosing β so that $1-5\beta>\gamma$, $\{(\mathbf{P}_nA>e^{-n\gamma\delta})\}$ is exponentially bounded. □

REMARK. The first inequality in (5.3) shows that as soon as $\mathbf{a}_n$ enters Δ_t, P becomes proper. Since $N\subset\Delta_t$, we see, in fact, that $\boldsymbol{\sigma}$ is exponentially bounded here.

As a prelude to the next theorem, we establish a large deviation result for U-statistics, a fact that may be of independent interest.

Let $u:\mathscr{X}^s\to R$ be measurable and symmetric in its arguments and for $n\geqq s$, let

$$\mathbf{u}_n=\sum u(\mathbf{X}_{i_1},\cdots,\mathbf{X}_{i_s})/\binom{n}{s},$$

where the sum is over all $i_1<\cdots<i_s$ selected from $\{1,\cdots,n\}$.

THEOREM 5.2. *If* $\mathbf{u}_s$ *is exponentially bounded and* $E\mathbf{u}_s = 0$, *then for all* $\varepsilon > 0$, $\{(|\mathbf{u}_n| > \varepsilon)\}$ *is exponentially bounded.*

PROOF. Following Hoeffding (1963), for $n \geqq s$, let

$$v_n(x_1, \cdots, x_n) = u(x_1, \cdots, x_s) + u(x_{s+1}, \cdots, x_{2s}) + \cdots + u(x_{ks-s+1}, \cdots, x_{ks}),$$

where $k = [n/s]$, the largest integer contained in n/s. Then

$$\mathbf{u}_n = \sum v_n(\mathbf{X}_{i_1}, \cdots, \mathbf{X}_{i_n})/kn!,$$

the sum being taken over all $n!$ permutations of $\{1, \cdots, n\}$. Since $k\mathbf{u}_n$ is an average of $(n!)$ identically distributed statistics, inequality (5.2) of Hoeffding (1963) shows that for $t > 0$, $F(k\mathbf{u}_n > k\varepsilon) \leqq E\exp\{t(\mathbf{v} - k\varepsilon)\}$, where $\mathbf{v} = v_n(\mathbf{X}_1, \cdots, \mathbf{X}_n)$. Since $\mathbf{v}$ is a sum of k independent identically distributed random variables, $E\exp\{t\mathbf{v}\} = (E\exp\{t\mathbf{u}_s\})^k = \phi^k(t)$, say. Thus $P(\mathbf{u}_n > \varepsilon) \leqq \exp\{-tk\varepsilon\}\phi^k(t)$. Since $\mathbf{u}_s$ is exponentially bounded and $E\mathbf{u}_s = 0$, $\phi(t) = 1 + O(t^2)$ for t near 0; thus for some $c, c' > 0$ and $\gamma, \rho < 1$ $P(\mathbf{u}_n > \varepsilon) < c\gamma^k < c'\rho^n$. The same argument applied to $-\mathbf{u}_n$ shows that $\{(\mathbf{u}_n < -\varepsilon)\}$ is also exponentially bounded. □

We now give a version of Theorem 5.1 for certain continuous models. We take $\mathbf{S}$ as in Section 3, s, U and K as in Assumptions 4.5. Let $\mathbf{Z} = \mathbf{Z}_S$ and $\mathbf{P} = \mathbf{P}_S$.

THEOREM 5.3. *If Assumptions* 4.5 *hold, if* P *assigns finite measure to every compact subset of* Θ *and if* $P\sup_U|l(\mathbf{X}|\cdot)|$, $\mathbf{S}$ *and* $(\ln \mathbf{Z})^+$ *are exponentially bounded, then for all* $\gamma < 1$, $\{(\mathbf{P}_n A_\delta' > e^{-n\gamma\delta})\}$ *is exponentially bounded.*

PROOF. Let $A_\delta' = A$.

$$\mathbf{P}_n A = \textstyle\int_A \exp\{n\boldsymbol{l}_n\}\,dP/\int \exp\{n\boldsymbol{l}_n\}\,dP. \tag{5.4}$$

Considering first the denominator, choose K as in Assumption 4.5(d); $A_\delta \subset K$. We establish first that for $\varepsilon > 0$, $\{(P\|\boldsymbol{l}_n - \lambda\|_{K\infty} > \varepsilon)\}$ is exponentially bounded. (The only property of K we require is that it is compact.) Since $EP\sup_U|l(\mathbf{X}|\cdot)| < \infty$ and l is continuous at θ, as $U \supset V \downarrow \{\theta\}$, $EP\sup_V l(\mathbf{X}|\cdot) \downarrow El(\mathbf{X}(\theta) = \lambda(\theta)$. Replacing l by $-l$ shows also that $EP\inf_V l(\mathbf{X}|\cdot) \uparrow \lambda(\theta)$. For each θ we may thus choose an open set V so that $\theta \in V \subset U$ and so that

$$\max\{|EP\sup_V l(\mathbf{X}|\cdot) - \lambda(\theta)|, |EP\inf_V l(\mathbf{X}|\cdot) - \lambda(\theta)|\} < \varepsilon/2. \tag{5.5}$$

There is a finite number of the $\{V\}$, say $V_1, \cdots, V_m$, that cover K. From (5.5), it follows that

$$\begin{aligned} P\sup_K|\boldsymbol{l}_n - \lambda| \leqq \max_{1 \leqq j \leqq m} \max\{&|\textstyle\sum_{i=1}^n P\sup_{V_j} l(\mathbf{X}_i|\cdot)/n - EP\sup_{V_j} l(X|\cdot)|, \\ &|\textstyle\sum_{i=1}^n P\inf_{V_j} l(\mathbf{X}_i|\cdot)/n - EP\inf_{V_j} l(\mathbf{X}|\cdot)|\} + \varepsilon/2. \end{aligned} \tag{5.6}$$

Let $\mathbf{B}_n$ denote the RHS of (5.6). The usual exponential bounds for sample means (a special case of Theorem 5.2, e.g.) implies that $\{(\mathbf{B}_n > \varepsilon)\}$ is exponentially bounded. (Note that $P\sup_V l(\mathbf{X}|\cdot)$ is exponentially bounded since $P\sup_U|l(\mathbf{X}|\cdot)|$

is, by assumption, and $V \subset U$. From (5.6) follows the exponential boundedness of

(5.7) $$\{(P \sup_K |\boldsymbol{l}_n - \lambda| > \varepsilon)\}.$$

We use this result to obtain bounds on the numerator and denominator in (5.4). Treating first the denominator, for $\beta > 0$ we have

(5.8) $$\int \exp\{n\boldsymbol{l}_n\}\, dP \geqq \int_{A_{\beta\delta}} \exp\{n\boldsymbol{l}_n\}\, dP.$$

For $\beta < 1$, $A_{\beta\delta} \subset K$ and hence is compact (see Berk (1966), Lemma 2). From (5.8) we see that $P\sup_{A_{\beta\delta}}|\boldsymbol{l}_n - \lambda| < \beta\delta$ implies that $\int \exp\{n\boldsymbol{l}_n\}\, dP \geqq \exp\{n(\lambda^* - 2\beta\delta)\} PA_{\beta\delta}$. Since $\{(P\sup_{A_{\beta\delta}}|\boldsymbol{l}_n - \lambda| > \beta\delta)\}$ is exponentially bounded, the same is true of

(5.9) $$\{(\int \exp\{n\boldsymbol{l}_n\}\, dP < \exp\{n(\lambda^* - 3\beta\delta)\})\}.$$

We consider next the numerator of (5.4). We note that $A = (K - A_\delta) \cup K'$, while

(5.10) $$\int_{K - A_\delta} \exp\{n\boldsymbol{l}_n\}\, dP \leqq \exp\{n(\lambda^* - \delta + \beta\delta)\} PK \leqq \exp\{n(\lambda^* - \delta + 2\beta\delta)\}$$

if $P\sup_K |\boldsymbol{l}_n - \lambda| < \beta\delta$ and n is sufficiently large. We thus have that

(5.11) $$\{(\int_{K - A_\delta} \exp\{n\boldsymbol{l}_n\}\, dP > \exp\{n(\lambda^* - \delta + 2\beta\delta)\})\}$$

is exponentially bounded.

For $n > s$ and on $(\mathbf{S} < n - s)$, let $\boldsymbol{l}_n' = \sum_{\mathbf{S}+1}^{n} l(\mathbf{X}_i \mid \theta)/n$ and let $\mathbf{u}_n'$ be the U-statistic based on $\mathbf{X}_{\mathbf{S}+1}, \cdots, \mathbf{X}_n$, formed from the kernel $u(\mathbf{X}_1, \cdots, \mathbf{X}_s) = P\sup_{K'} \boldsymbol{l}_s$. Then on $(\mathbf{S} < n - s)$,

(5.12) $$\int_{K'} \exp\{n\boldsymbol{l}_n\}\, dP = \mathbf{Z}\int_{K'} \exp\{n\boldsymbol{l}_n'\}\, d\mathbf{P} \leqq \mathbf{Z}\exp\{(n - \mathbf{S})\mathbf{u}_n'\}$$

From Theorem 5.2 and the independence of $\mathbf{S}$ and $\mathbf{u}_n'$, we may conclude that for some $c > 0$ and $\rho < 1$, $F(\mathbf{u}_n' > \lambda^* - \delta + \beta\delta \mid \mathbf{S}) < c\rho^{n-\mathbf{S}}$. Again, by adjusting $p^*(x)$, we may suppose $\lambda^* > \delta$. Since $\mathbf{S}$ is exponentially bounded, by increasing ρ if necessary, we have that $E\rho^{-\mathbf{S}} < \infty$, hence that $\{((n - \mathbf{S})\mathbf{u}_n' > n(\lambda^* - \delta + \beta\delta))\}$ is exponentially bounded. (We take $\mathbf{u}_n' = 0$ on $(\mathbf{S} \geqq n - s)$.) Since also $(\ln \mathbf{Z})^+$ is exponentially bounded, the same is true of

(5.13) $$\{(\int_{K'} \exp\{n\boldsymbol{l}_n\}\, dP > n(\lambda^* - \delta + 2\beta\delta))\}.$$

The exponential boundedness of (5.9), (5.11) and (5.13) then imply that $\{(\mathbf{P}_n A > \exp\{-n\delta(1 - 6\beta)\})\}$ is exponentially bounded. Choosing β so that $1 - 6\beta > \gamma$ concludes the proof. □

An immediate consequence of the foregoing is

COROLLARY 5.4. *If the A_δ form a weak base at θ^* and the conditions of either Theorem* 5.1 *or* 5.3 *hold, then for every neighborhood U, of θ^*, there is a $\beta > 0$ so that $\{(\mathbf{P}_n U' > e^{-n\beta})\}$ is exponentially bounded.*

6. Examples. We illustrate the foregoing with two examples.

I. *Exponential model.* Let $p(x \mid \theta) = \exp\{-(x - \xi)^2/2\sigma^2\}/\sigma(2\pi)^{\frac{1}{2}}$, $-\infty < x < \infty$, $\theta = (\xi, \sigma^2)$. Choosing $p^*(x) = (2\pi)^{\frac{1}{2}}$,

(6.1) $$l(x \mid \theta) = -x^2/2\sigma^2 + \xi x/\sigma^2 - \xi^2/2\sigma^2 - \ln \sigma.$$

Since the carrier of $p(\cdot \mid \theta)$ is $(-\infty, \infty)$ for every θ, 1.1(b) is automatically satisfied. The relevant expectations in 1.1(f) certainly exist if $E\mathbf{X}^2 < \infty$. Of course then $\lambda^* > -\infty$.

We verify consistency a posteriori via Propositions 4.1–4.3. We assume P dominates Lebesgue measure on Θ. Replacing x by α_1 and x^2 by α_2 in (6.1), we have

$$\lambda_0 + \alpha'\Lambda = -\alpha_2/2\sigma^2 + \xi\alpha_1/\sigma^2 - \xi^2/2\sigma^2 - \ln\sigma = -(\alpha_1 - \xi)^2/2\sigma^2 - (\alpha_2 - \alpha_1{}^2)/2\sigma^2 - \ln\sigma. \tag{6.2}$$

It can be readily seen from (6.2) that $D = (\alpha_1{}^2 < \alpha_2)$ (and that the maximizing values in (6.2) are $\xi = \alpha_1$ and $\sigma^2 = \alpha_2 - \alpha_1{}^2$). Letting $a_1(x) = x$, $a_2(x) = x^2$, we see that $F(a(\mathbf{X}) \in \mathscr{L}(D)) = 1$, while $\alpha_F \in D^\circ$ unless $\mathbf{X}$ is degenerate. $E\mathbf{X}^2 < \infty$ also assures that Proposition 4.1(ii) holds. Referring to Corollary 4.2, we see that if P becomes proper, $F(\mathbf{P}_n A_\delta \to 1) = 1$. We note that, if, e.g., $dP = \sigma^a\, d\xi\, d\sigma$, P becomes proper if $\mathbf{X}$ is not degenerate (in fact, as soon as $n > a+2$ and at least two observations differ). This may be established directly, but we do so using Proposition 4.3: From (6.2), one sees that for $n > a+2$, $H_n \supset D$; hence for $n > a+2$, $I_n \supset D$. Thus $F(a(\mathbf{X}) \in \mathscr{L}(H)) = 1$ and if $\mathbf{X}$ is not degenerate, $\alpha_F \in D^\circ \subset H^\circ$. Thus P remains proper.

We examine the A_δ. Let $\xi_F = E\mathbf{X}$ and $v_F = E\mathbf{X}^2$.

$$\lambda(\theta) = -v_F/2\sigma^2 + \xi\xi_F/\sigma^2 - \xi^2/2\sigma^2 - \ln\sigma.$$

Since P dominates Lebesgue measure on Θ, it is easily verified that $\lambda^* = \lambda(\theta^*)$, where $\theta^* = (\xi_F, \sigma_F{}^2)$, $\sigma_F{}^2 = v_F - \xi_F{}^2$. Thus for the univariate normal model, we have, asymptotically, for a large class of priors, consistency to the true mean and variance, even though the data are not normal, provided $0 < \sigma_F < \infty$.

We turn to Theorem 5.1, working with $dP = \sigma^a\, d\xi\, d\sigma$ as an illustration. Since for $n > a+2$, $\Delta_n = D \cap H_n = D$, $\Delta = D$. Moreover, the hypotheses of Theorem 5.1 hold if $\mathbf{X}^2$ is exponentially bounded and not degenerate. Thus (see Corollary 5.4) for any neighborhood U, of $\theta^* = (\xi_F, \sigma_F{}^2)$, for some $\beta > 0$, $\{(\mathbf{P}_n U' > e^{-n\beta})\}$ is exponentially bounded.

II. *Continuous model.* $p(x \mid \theta) = f(x-\theta)$, $-\infty < \theta < \infty$, where $f(x) = 1/\pi(1+x^2)$, $-\infty < x < \infty$. We choose $p^*(x) = \pi(1+x^2)$, so that $l(x \mid \theta) = \ln[(1+x^2)/(1+(x-\theta)^2)]$. Since $|l(\cdot \mid \theta)|$ is bounded, $El(\mathbf{X} \mid \theta)$ always exists and the requirement in Proposition 4.6 that $\lambda^* \neq -\infty$ holds. To verify Assumption 4.5(c), let $U = (-\varepsilon, \varepsilon)$, $\varepsilon > 0$. For $\theta \in U$, $|l(x \mid \theta)| \leqq \ln(1+\varepsilon^2)$ if $|x| \leqq \varepsilon$, while $\ln[(1+x^2)/(1+(|x|+\varepsilon)^2)] \leqq l(x \mid \theta) \leqq \ln[(1+x^2)/(1+(|x|-\varepsilon)^2)]$ if $|x| > \varepsilon$. Thus $EP\sup_U |l(\mathbf{X} \mid \theta)| < \infty$. By translation, the same is true for $U = (\theta-\varepsilon, \theta+\varepsilon)$, for all $\theta \in \Theta$. Taking $s = 1$ and $K = [-M, M]$, $M > 0$, in Assumption 4.5(d), for $\theta \in K'$, $l(x \mid \theta) \leqq \ln[(1+x^2)/(1+(M-|x|)^2)]$ if $|x| \leqq M$, while $\sup_{K'} l(x \mid \theta) = \ln(1+x^2)$ if $|x| > M$. Thus if we assume $E(\ln|\mathbf{X}|)^+ < \infty$, by dominated convergence we see that $E\sup_{K'} l(\mathbf{X} \mid \theta) \downarrow -\infty$ as $M \uparrow \infty$. Hence the required K exists. Proposition 4.6 and Theorem 3.3 then apply if P becomes proper. (E.g., if P is Lebesgue measure on $\Theta = R$, P becomes proper after one observation. This is generally true for a location family.)

Assumptions 4.5(b), (c) insure that $\lambda(\theta)$ is continuous in θ and the above analysis shows that $\lambda(\theta) \to -\infty$ as $|\theta| \to \infty$. Hence if λ has a unique maximum, at θ^*, say, the A_δ form a weak base there. We note in particular that if the model holds ($F \sim p(\cdot \mid \theta^*)$, say), then $E(\ln|\mathbf{X}|)^+ < \infty$ and λ has a unique maximum at θ^*.

Turning to Theorem 5.3, we note first that if P is Lebesgue measure, $\mathbf{Z} = \mathbf{Z}_1 = \pi(1+\mathbf{X}_1{}^2)$. Hence $(\ln \mathbf{Z})^+$ as well as the remaining quantities in the hypothesis are exponentially bounded provided $E|\mathbf{X}|^r < \infty$ for some $r > 0$.

Unhappily, it is not known whether Assumptions 4.5 can be verified for the Cauchy location/scale family. The trouble spot is Assumption 4.5(d). Choosing $s = 1$ gives $\sup_{K'} l(x \mid \theta) \equiv +\infty$ for all K. Choosing $s = 2$ gives $\sup_{K'} \mathbf{l}_2 = n[(1+\mathbf{X}_1{}^2)(1+\mathbf{X}_2{}^2)/(\mathbf{X}_1 - \mathbf{X}_2)^2]$ for all K which cannot be given an arbitrarily small expectation. It is not known if choosing $s \geqq 3$ will do.

We note too that the normal model could be analyzed as a continuous model. One then finds that more stringent assumptions about $\mathbf{X}$ are required.

Acknowledgment. The author is indebted to Professor R. A. Wijsman for suggesting many improvements that are incorporated into the exposition.

REFERENCES

[1] BERK, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

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

[3] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–20.

[4] HUBER, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 221–233. Univ. of California Press, Berkeley.

[5] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

[6] WIJSMAN, R. A. (1968). Bounds on the sample size distribution for a class of invariant sequential probability ratio tests. *Ann. Math. Statist.* **39** 1048–1056.