ON THE ASYMPTOTIC BEHAVIOR OF BAYES' ESTIMATES IN THE DISCRETE CASE[1]

BY DAVID A. FREEDMAN

University of California, Berkeley

1. Summary. Doob (1949) obtained a very general result on the consistency of Bayes' estimates. Loosely, if any consistent estimates are available, then the Bayes' estimates are consistent for almost all values of the parameter under the prior measure. If the parameter is thought of as being selected by nature through a random mechanism whose probability law is known, Doob's result is completely satisfactory. On the other hand, in some circumstances it is necessary to identify the exceptional null set. For example, if the parameter is thought of as fixed but unknown, and the prior measure is chosen as a convenient way to calculate estimates, it is important to know for which null set the method fails. In particular, it is desirable to choose the prior so that the null set is in fact empty.

The problem is very delicate; considerable work [8], [9], [12] has been done on it recently, in quite general contexts and under severe regularity assumptions. It might therefore be of interest to discuss the simplest possible case, that of independent, identically distributed, *discrete* observations, in some detail. This will be done in Sections 3 and 4 when the observations take a *finite* set of possible values. Under this assumption, Section 3 shows that the posterior probability converges to point mass at the true parameter value among almost all sample sequences (for short, the posterior is consistent; see Definition 1) exactly for parameter values in the topological carrier of the prior. In Section 4, the asymptotic normality of the posterior is shown to follow from a local smoothness assumption about the prior.

In both sections, results are obtained for priors which admit the possibility of an infinite number of states. The results of these sections are not entirely new; see pp. 333 ff. of [7], pp. 224 ff. of [10], [11]. They have not appeared in the literature, to the best of our knowledge, in a form as precise as Theorems 1, 3, 4. Theorem 2 is essentially the relevant special case of Theorem 7.4 of Schwartz (1961).

In Sections 5 and 6, the case of a countable number of possible values is treated. We believe the results to be new. Here the general problem appears, because priors which assign positive mass near the true parameter value may lead to ridiculous estimates. The results of Section 3 (let alone 4) are false. In fact, Theorem 5 of Section 5 gives the following construction. Suppose that under the true parameter value the observations take an infinite number of values with positive probability. Then given any spurious (sub-)stochastic probability

Received September 18, 1962.

[1] Prepared with the partial support of the National Science Foundation, Grant G-14648.

distribution, it is possible to find a prior assigning positive mass to any neighborhood of the true parameter value, but leading to a posterior probability which converges for almost all sample sequences to point mass at the spurious distribution. Indeed, there is a prior assigning positive mass to every open set of parameters, for which the posterior is consistent only at a set of parameters of the first category.

To some extent, this happens because at any stage information about a finite number of states only is available, but on the basis of this evidence, conclusions must be drawn about all states. If the prior measure has a serious prejudice about the shape of the tails, disaster ensues. In Section 6, it is shown that a simple condition on the prior measure (which serves to limit this prejudice) ensures the consistency of the posterior.

Prior probabilities leading to posterior distributions consistent at all and asymptotically normal at essentially all (see Remark 3, Section 3) parameter values are constructed. Section 5 is independent of Sections 3 and 4; Section 6 is not. Section 6 overlaps to some extent with unpublished work of Kiefer and Wolfowitz; it has been extended in certain directions by Fabius (1963).

The results of this paper were announced in [5]; some related work for continuous state space is described in [3]. It is a pleasure to thank two very helpful referees: whatever expository merit Section 5 has is due to them and to L. J. Savage.

2. Notations and definitions. We use the following notation throughout the paper. The letter Ω denotes an abstract space and $\mathfrak{F}$ a σ-field of subsets of Ω. The letter I denotes a countable set and $\{X_n : n = 1, 2, \cdots\}$ is a sequence of I-valued random variables ("the observations") on $(\Omega, \mathfrak{F})$. Let S be the space of functions from I to $[0, 1]$, in the product topology; S is compact and metrizable. Let $L = \{\lambda \mid \lambda \varepsilon S \text{ and } \sum_{i \varepsilon I} \lambda(i) \leqq 1\}$, in the relative topology. Then L is compact. Let "the parameter space" $\Lambda = \{\lambda \mid \lambda \varepsilon L \text{ and } \sum_{i \varepsilon I} \lambda(i) = 1\}$. If I is finite then Λ is closed and nowhere dense in L; but if I is infinite then Λ is a dense G_δ in L, so $L - \Lambda$ is of the first category. Suppose that for each $\lambda \varepsilon \Lambda$ there is a probability P_λ on $\mathfrak{F}$ under which the $\{X_n\}$ are independent with common distribution $P_\lambda\{\omega \mid \omega \varepsilon \Omega, X_n(\omega) = i\} = \lambda(i)$, $i \varepsilon I$. We define P_λ only for $\lambda \varepsilon \Lambda$. Let μ be a probability ("the prior") on the Borel σ-field $\mathcal{B}$ of L. For technical reasons, it is convenient to allow $\mu(\Lambda) < 1$. If $\theta \varepsilon \Lambda$ is the "true parameter value," we will often project μ onto the subspace spanned by $\lambda \rightarrow \{\lambda(i) : \theta(i) > 0\}$. Even if $\mu(\Lambda) = 1$, the projection of μ may assign positive mass to sub-stochastic vectors. The reader should draw a picture for an I of three points and a θ vanishing at one of them. Let $C(\mu)$, the topological carrier of μ, be the smallest compact subset of L of μ-measure 1, so that $\lambda \varepsilon C(\mu)$ if and only if $\lambda \varepsilon L$ and every L-neighborhood of λ has positive μ-measure.

The "posterior distribution" $\mu_{n,\omega}$ of λ given $X_1(\omega) \cdots X_n(\omega)$ is defined by

$$(1) \qquad \mu_{n,\omega}(A) = \left[\int_A \prod_{j=1}^n \lambda[X_j(\omega)]\mu(d\lambda) \bigg/ \int_L \prod_{j=1}^n \lambda[X_j(\omega)]\mu(d\lambda) \right]$$

for $A \varepsilon \mathcal{B}$ and nonzero denominator. When defined, $\mu_{n,\omega}$ is a probability on $\mathcal{B}$ dominated by μ. If $\theta \varepsilon C(\mu)$, or more generally if $\theta(i) > 0$ implies $\int_L \lambda(i)\mu(d\lambda) > 0$, then $\mu_{n,\omega}$ is defined a.s. $[P_\theta]$. If $\theta(i) > 0$ and $\int_L \lambda(i)\mu(d\lambda) = 0$ for some $i \varepsilon I$, then for P_θ-almost all ω, the denominator in (1) is eventually 0.

If the loss $L(\lambda, \theta)$ caused by estimating λ when the true parameter value is θ satisfies $L(\lambda, \theta) = \sum_{i \varepsilon I} a(i)[\lambda(i) - \theta(i)]^2$ where $a \varepsilon \Lambda$ is fixed, then the Bayes' estimate $\beta_{n,\omega}(\cdot)$ of θ is

$$\beta_{n,\omega}(i) = \int_L \lambda(i)\mu_{n,\omega}(d\lambda). \tag{2}$$

The weak $*$ topology is assigned to the space of probabilities on $\mathcal{B}$, meaning $\mu_n \to \mu$ if and only if $\int_L f\, d\mu_n \to \int_L f\, d\mu$ for each continuous f on L. Point mass at λ is written δ_λ. The μ_n-measure of each L-neighborhood of λ converges to 1 if and only if $\mu_n \to \delta_\lambda$.

Let $n_i(\omega)$ be the number of i's among $[X_j(\omega), 1 \leqq j \leqq n]$, so that

$$\prod_{j=1}^{n} \lambda[X_j(\omega)] = \prod_{i \varepsilon I} \lambda(i)^{n_i(\omega)}.$$

Here and throughout we understand that $0^0 = 1$; an empty sum is 0; an empty product is 1.

DEFINITION 1. *If μ is a probability on $\mathcal{B}$ and $\theta \varepsilon \Lambda$, we say (θ, μ) is consistent if and only if $\mu_{n,\omega} \to \delta_\theta$ for P_θ-almost all ω.*

If (θ, μ) is consistent, plainly $\lim_{n\to\infty}\beta_{n,\omega}(i) = \theta(i)$ for P_θ-almost all ω.

We will summarize here the main facts known about consistency, for *finite* I and priors μ *concentrated in* Λ. Theorem 2 of Section 3 implies: (θ, μ) is consistent exactly for θ in the topological carrier of μ. Doob (1949) and a standard argument gives the weaker result: (θ, μ) is consistent for μ-almost all $\theta \varepsilon \Lambda$. Blackwell and Dubins (1962) imply, via a different standard argument, a result which is not comparable: if μ and $\hat{\mu}$ are equivalent probabilities on $\mathcal{B}$ concentrated in Λ, then for μ-almost all $\lambda \varepsilon \Lambda$, the variation norm of $\mu_{n,\omega} - \hat{\mu}_{n,\omega}$ converges to 0 as n tends to ∞ for P_λ-almost all ω.

For *countable* I, the results of Doob and of Blackwell-Dubins still hold. According to Remark 6 of Section 5, Doob's exceptional null set may be much larger than the largest open set of prior probability 0.

Miscellaneous conventions. Unless noted, $\lambda \varepsilon L$ and $\lambda(i)$ is its ith coordinate, $i \varepsilon I$. The letter j indexes the random variable X_j, and ranges from 1 to n, the number of observations on hand. The letter θ is reserved for points of Λ. Logarithms to base e are written log; logarithms to base b are written $\log_b$. If $\{Z_m : 1 \leqq m < \infty\}$ are random variables on $(\Omega, \mathcal{F}, P)$ and $\{a_m : 1 \leqq m < \infty\}$ are real numbers, phrases like "Z_m is bounded by a_m eventually a.s. $[P]$" mean: there is a P-null set $E \varepsilon \mathcal{F}$ and a natural number valued random variable τ defined on $\Omega - E$ such that if $\omega \varepsilon \Omega - E$ and $m \geqq \tau(\omega)$ then $Z_m(\omega) \leqq a_m$.

If μ is a measure on $\mathcal{B}$ and ϕ is a measurable vector function on $(L, \mathcal{B})$, phrases like "the distribution of ϕ under μ" signify the measure $\mu\phi^{-1}$.

3. Bayes' estimates are consistent. The principal result of this section is

THEOREM 1. *Suppose $\theta \varepsilon \Lambda$ and $\{i \mid \theta(i) > 0\}$ is finite. Let μ be a probability on $\mathcal{B}$. Then (θ, μ) is consistent if and only if θ is in the topological carrier of μ.*

PROOF. The "only if" part is clear from discussion of (1). If the $\{\mu_{n,\omega}\}$ are defined a.s. $[P_\theta]$ they and all their subsequential limits concentrate on $C(\mu)$. If $\theta \not\varepsilon C(\mu)$ either $\{\mu_{n,\omega}\}$ is eventually undefined a.s. $[P_\theta]$ or $\{\mu_{n,\omega}\}$ is always defined a.s. $[P_\theta]$ but δ_θ is not a limit point of the sequence. The other implication is nontrivial, and will be deduced from Theorem 2.

Let λ and p be in L. The entropy of λ relative to p is $H(\lambda \mid p) = -\sum_{i \varepsilon I} p(i) \log \lambda(i)$, with $\log 0 = -\infty$ and $0 \cdot -\infty = 0$. Thus $0 \leqq H(\lambda \mid p) \leqq \infty$, and

$$H(\alpha_1\lambda_1 + \alpha_2\lambda_2 \mid p) \leqq \alpha_1 H(\lambda_1 \mid p) + \alpha_2 H(\lambda_2 \mid p) \tag{3}$$

for α_1 and α_2 nonnegative, $\alpha_1 + \alpha_2 = 1$; λ_1, λ_2, p in L. Moreover, if $p \varepsilon L$ and $p(i) > 0$ for some i, there is a unique positive c with $cp \varepsilon \Lambda$; then

$$H(\lambda \mid p) \geqq H(cp \mid p). \tag{4}$$

These facts are well known and follow easily from Jensen's inequality and the concavity of the logarithm function. We write $H(\theta \mid \theta) = H(\theta)$.

THEOREM 2. *Let $\theta \varepsilon \Lambda$ have $H(\theta) < \infty$. Let μ be a probability on $\mathcal{B}$ with the property that for any neighborhood U of θ in L and any $\delta > 0$, $\mu\{\lambda \mid \lambda \varepsilon U$ and $H(\lambda \mid \theta) < H(\theta) + \delta\} > 0$. Then (θ, μ) is consistent.*

REMARK. By example, the condition is not necessary.

PROOF. Let $I_+ = \{i \mid i \varepsilon I$ and $\theta(i) > 0\}$. Let $S_+, L_+, \Lambda_+, \mathcal{B}_+$ be defined as before with I_+ in place of I. Let ν and $\nu_{n,\omega}$ be the projections of μ and $\mu_{n,\omega}$ on $\mathcal{B}_+$. We still have

$$\nu_{n,\omega}(A) = \left[\int_A \prod_{j=1}^n \lambda[X_j(\omega)]\nu(d\lambda) \Big/ \int_{L_+} \prod_{j=1}^n \lambda[X_j(\omega)]\nu(d\lambda)\right]$$

for $A \varepsilon \mathcal{B}_+$. Enumerate I so that if I_+ is finite then $I_+ = \{1, 2, \cdots, N\}$, and if I_+ is infinite then $I_+ = \{1, 2, \cdots\}$. In the second case, we choose N large. Precisely, let δ be positive but less than 1. Write $r(k) = \sum_{i=k+1}^\infty \theta(i)$. Choose N so large that

(i) $r(N) < 4\delta$,

(ii) $\sum_{i=1}^N \theta(i) \log \theta(i) < -H(\theta) + \delta$.

Define $x^+ = x, x \geqq 0; = 0, x < 0$ and let

$$V = \left\{\lambda \mid \lambda \varepsilon L_+ \text{ and } \sum_{i=1}^N \frac{\theta(i)[\lambda(i) - \theta(i)]^2}{[\theta(i) + (\lambda(i) - \theta(i))^+]^2} < 16\delta\right\}. \tag{5}$$

The crucial fact to establish is

$$\lim_{n\to\infty} \nu_{n,\omega}(V) = 1 \quad \text{a.s. } [P_\theta], \tag{6}$$

which is equivalent to

$$\lim_{n\to\infty} \nu_{n,\omega}(L_+ - V)/\nu_{n,\omega}(V) = 0 \quad \text{a.s. } [P_\theta]. \tag{7}$$

By discarding a P_θ-null set, suppose $\lim_{n\to\infty} n^{-1}n_i(\omega) = \theta(i)$, $1 \leqq i \leqq N$ for all ω. We will begin the proof of (7) by establishing

$$\text{(8)} \qquad \sup_{\lambda \varepsilon L_+ - V} \sum_{i=1}^{\infty} n^{-1}n_i(\omega) \log \lambda(i) < -H(\theta) - 2\delta$$

for large enough n; compare (2.12) on page 892 of Kiefer and Wolfowitz (1956), an inequality they attribute to Wald. Let ∂V be the boundary of V in L_+. The left side of (8) is bounded by the left side of

$$\text{(9)} \quad \sup_{\lambda \varepsilon L_+ - V} \sum_{i=1}^{N} n^{-1}n_i(\omega) \log \lambda(i) = \max_{\lambda \varepsilon \partial V} \sum_{i=1}^{N} n^{-1}n_i(\omega) \log \lambda(i),$$

an equality holding for large n. To see this, consider the points θ_N and $\theta_{n,\omega}$ of Λ_+ defined as

$$\begin{aligned} \theta_N(i) &= [1 - r(N)]^{-1}\theta(i), && 1 \leqq i \leqq N, \\ &= 0, \quad \text{elsewhere;} \\ \theta_{n,\omega}(i) &= n^{-1}n_i(\omega) \Big/ \sum_{i=1}^{N} n^{-1}n_i(\omega), && 1 \leqq i \leqq N, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

Now $\theta_N \varepsilon V$ by (i), so for n sufficiently large, $\theta_{n,\omega} \varepsilon V$. But the function $\lambda \to \sum_{i=1}^{N} n^{-1}n_i(\omega) \log \lambda(i)$ is concave by (3) and has a maximum at $\theta_{n,\omega}$ by (4). This proves (9) by an easy argument. The right side of (9) may be estimated using the obvious inequality: $\log(1 + x) \leqq x - \frac{1}{2}(1 + x^+)^{-2}x^2$, valid for $x \geqq -1$. Indeed, if $\lambda \varepsilon \partial V$,

$$\begin{aligned} &\sum_{i=1}^{N} \theta(i) \log \lambda(i) - \sum_{i=1}^{N} \theta(i) \log \theta(i) \\ &\qquad = \sum_{i=1}^{N} \theta(i) \log [1 + \theta(i)^{-1}(\lambda(i) - \theta(i))] \\ &\qquad \leqq \sum_{i=1}^{N} \lambda(i) - \sum_{i=1}^{N} \theta(i) - \frac{1}{2}\sum_{i=1}^{N} \frac{\theta(i)[\lambda(i) - \theta(i)]^2}{[\theta(i) + (\lambda(i) - \theta(i))^+]^2} < -4\delta, \end{aligned}$$

the first two terms together being bounded by $r(N) < 4\delta$, and the last being exactly 8δ since $\lambda \varepsilon \partial V$. Using Condition (ii),

$$\text{(10)} \qquad \max_{\lambda \varepsilon \partial V} \sum_{i=1}^{N} \theta(i) \log \lambda(i) < -H(\theta) - 3\delta.$$

It is routine to deduce from (10) that eventually

$$\max_{\lambda \varepsilon \partial V} \sum_{i=1}^{N} n^{-1}n_i(\omega) \log \lambda(i) < -H(\theta) - 2\delta,$$

proving (8).

From (8) follows

$$\int_{L_+ - V} \prod_{j=1}^{n} \lambda(X_j)\nu(d\lambda) \leqq \exp\{-n[H(\theta) + 2\delta]\}. \tag{11}$$

Let $V_0 = \{\lambda \mid \lambda \varepsilon V \text{ and } H(\lambda \mid \theta) < H(\theta) + \delta\}$. Then $\nu(V_0) > 0$ by hypothesis. For any particular $\lambda \varepsilon V_0$, the strong law implies

$$\liminf_{n\to\infty} n^{-1} \sum_{j=1}^{n} \log \lambda[X_j(\omega)] > -H(\theta) - \delta \tag{12}$$

for P_θ-almost all ω. The exceptional P_θ-null subset of Ω depends, of course, on λ. By Fubini's theorem, there is a P_θ-null set $E \varepsilon \mathfrak{F}$ such that if $\omega \not\varepsilon E$ then (12) holds for ν-almost all $\lambda \varepsilon V_0$. Here the exceptional ν-null subset of L_+ depends on ω. By Fatou's lemma,

$$\liminf_{n\to\infty} \int_{V_0} \left[\prod_{j=1}^{n} \lambda(X_j)\right] \exp\{n[H(\theta) + \delta]\}\, \nu(d\lambda) > \nu(V_0) \quad \text{a.s.}[P_\theta], \tag{13}$$

so eventually a.s. $[P_\theta]$

$$\int_{V_0} \prod_{j=1}^{n} \lambda(X_j)\nu(d\lambda) \geqq \nu(V_0) \exp\{-n[H(\theta) + \delta]\}. \tag{14}$$

A fortiori, (14) holds with V_0 replaced by V on the left side; comparing this with (11) leads to

$$\nu_{n,\omega}(L_+ - V)/\nu_{n,\omega}(V) \leqq \nu(V_0)^{-1} \exp(-n\delta) \tag{15}$$

eventually a.s. $[P_\theta]$, which implies (7). This completes the proof of (6).

Now let $\delta_k \to 0$ as $k \to \infty$. Apply this argument with $\delta = \delta_k$ to secure an L_+-neighborhood V_k of θ^+ defined by (5) with $\delta = \delta_k$, and a P_θ-null set E_k such that $\omega \not\varepsilon E_k$ implies $\lim_{n\to\infty} \nu_{n,\omega}(V_k) = 1$. For $\omega \not\varepsilon \bigcup_{k=1}^{\infty} E_k$, $\lim_{n\to\infty} \nu_{n,\omega}(V_k) = 1$ for all k. If $\lambda \varepsilon L$, let λ^+ be its restriction to I_+. Since $\{V_k : 1 \leqq k < \infty\}$ is a basis for the L_+-neighborhoods of θ^+, therefore $\nu_{n,\omega} \to \delta_{\theta^+}$. If $I - I_+$ is empty, the proof terminates. Otherwise, this standard reasoning is needed.

Let J be a finite subset of I disjoint from I_+; V is an L_+-neighborhood of θ^+ and $\delta > 0$. Then $W = \{\lambda \mid \lambda \varepsilon L, \max_{i\varepsilon J} \lambda(i) < \delta, \lambda^+ \varepsilon V\}$ is an L-neighborhood of θ, and by varying J, V, δ we get a basis for the L-neighborhoods of θ. If $\nu_{n,\omega} \to \delta_{\theta^+}$ then $\mu_{n,\omega}(W) \geqq \nu_{n,\omega}\{\lambda : \lambda \varepsilon L_+; \lambda \varepsilon V; \sum_{i\varepsilon I_+} \lambda(i) > 1 - \frac{1}{2}\delta\} \to 1$, so $\mu_{n,\omega}(W) \to 1$, and $\mu_{n,\omega} \to \delta_\theta$. This completes the proof of Theorem 2.

Remark 1. For future use, if (15) holds for V, it holds for any superset of V—since expanding V decreases the numerator and increases the denominator of the left side. But (15) has been established for a countable basis of the L_+-neighborhood system at θ^+ (with δ depending on V). Hence, for ω outside a P_θ-null set, for any L_+-neighborhood V of θ^+ there is a $\delta > 0$ depending on V but not ω for which (15) holds eventually.

Proof of Theorem 1. If I_+ is finite then $\lambda \to H(\lambda \mid \theta)$ is a continuous function from L to $[0, \infty]$ and $H(\theta) < \infty$. Thus $\theta \varepsilon C(\mu)$ implies the condition of

Theorem 2. More careful arguing shows that $\mu_{n,\omega} \to \delta_\theta$ exactly for ω where $n^{-1}n_i(\omega) \to \theta(i)$, $i \varepsilon I$.

REMARK 2. If I_+ is infinite, then $\lambda \to H(\lambda \mid \theta)$ is identically $+\infty$ when $H(\theta) = \infty$. When $H(\theta) < \infty$, it is lower semi-continuous but not continuous, being $+\infty$ on a dense G_δ. This function $p \to H(p)$ on L is also lower semi-continuous but not continuous, being $+\infty$ on a dense G_δ.

4. The posterior is asymptotically normal. In this section, we assume (i) $\theta \varepsilon \Lambda$ and I_+ is finite, enumerated as $\{1 \cdots N\}$. We suppose (ii) θ is nondegenerate, implying $N \geqq 2$. We use the notation of Section 3, as well as the symbols Λ_k and L_k, which are Λ and L of Section 2 with $I = \{1 \cdots k\}$. Thus $L_+ = L_N$. If $\lambda \varepsilon L$ or L_N then $\lambda*$ is its restriction to $\{1 \cdots N-1\}$. Let R^k be k-dimensional Euclidean space; $\mathbf{y} \varepsilon R^k$ has components $y_1 \cdots y_k$. Define $s(\mathbf{y}) = \sum_{i=1}^{N-1} y_i$ and $h(\mathbf{y}) = \frac{1}{2}\sum_{i=1}^{N-1} y_i^2/\theta(i) + \frac{1}{2}s(\mathbf{y})^2/\theta(N)$ for $\mathbf{y} \varepsilon R^{N-1}$. Abbreviate the function $\lambda \to \{n^{\frac{1}{2}}[\lambda(i) - n^{-1}n_i(\omega)],\ 1 \leqq i \leqq N-1\}$ from L, L_{N-1} or L_N to R^{N-1} as $\phi_{n,\omega}$. We call $\nu(\Lambda_N) = 1$ Case A and $\nu(\Lambda_N) < 1$ Case B. In Case A, we assume (iii) the distribution of $\lambda \to \lambda*$ under μ has a continuous positive density f (i.e., with respect to Lebesgue measure) in an L_{N-1}-neighborhood T of $\theta*$. In Case B, we assume (iv) in an L_N-neighborhood T of θ^+ the distribution of $\lambda \to \lambda^+$ under μ has a density f which is positive and continuous at θ^+. The results of this section are:

THEOREM 3. *In Case* A, *for P_θ-almost all ω, the distribution of $\phi_{n,\omega}$ under $\mu_{n,\omega}$, when restricted to any fixed compact subset of R^{N-1}, is eventually absolutely continuous with continuous positive density converging uniformly to*

$$\mathbf{y} \to [\theta(1) \cdots \theta(N)]^{-1}(2\pi)^{-\frac{1}{2}(N-1)} \exp[-h(\mathbf{y})]. \tag{16}$$

In Case B, *for P_θ-almost all ω, the distribution of $\phi_{n,\omega}$ under $\mu_{n,\omega}$ converges in variation norm to the distribution with density* (16).

REMARK 3. If (ii) does not hold, meaning $\theta(1) = 1$, suppose the distribution of $\lambda \to \lambda(1)$ under μ has a continuous positive density at 1. If $n_1(\omega) = n$ for all n, the asymptotic distribution of $\lambda \to n[1 - \lambda(1)]$ under $\mu_{n,\omega}$ is exponential. See the proof of Corollary 2.

REMARK 4. In Case B, there is no regularity condition on the μ-distribution of $\lambda \to \lambda*$ guaranteeing the asymptotic normality of $\phi_{n,\omega}$ under $\mu_{n,\omega}$. For any probability distribution on L_{N-1} which has a continuous, positive density in a neighborhood of $\theta*$, there is a probability μ on $\mathcal{B}$ giving $\lambda \to \lambda*$ this distribution, and assigning positive mass to any L-neighborhood of θ; but for which $\phi_{n,\omega}$ has a limiting distribution under $\mu_{n,\omega}$ for P_θ-almost no ω.

The Bayes' estimates $\beta_{n,\omega}(i)$ were defined by Equation (2).

THEOREM 4. *For P_θ-almost all ω and $1 \leqq i \leqq N$, $\lim_{n\to\infty} n^{\frac{1}{2}}[\beta_{n,\omega}(i) - n^{-1}n_i(\omega)] =$ 0.*

COROLLARY 1. *Theorem 3 continues to hold with $\phi_{n,\omega}(\lambda)$ defined as $\{n^{\frac{1}{2}}[\lambda(i) - \beta_{n,\omega}(i)] : 1 \leqq i \leqq N-1\}$.*

COROLLARY 2. *Along almost all sample sequences, the posterior distribution,* *when centered at its mean and rescaled by* $n^{\frac{1}{2}}$, *converges to a limiting normal distribution; this is precisely the limiting joint distribution of the maximum likelihood estimates, when centered at its mean and rescaled by* $n^{\frac{1}{2}}$.

This striking and mysterious fact is true in a wide variety of situations; see LeCam (1957). Of course, the Bayes' estimates and the maximum likelihood estimates have the same asymptotic distribution under P_θ by Theorem 4.

The proofs of these results are computational; we hope that the notation to be introduced here will simplify the reading. Let M be the closed, solid sphere of radius m about the origin in R^{N-1}. Define

$$b_{n,\omega} = \int_{L_N} \prod_{i=1}^{N} \lambda(i)^{n_i(\omega)} \nu(d\lambda)$$

$$c_{n,\omega} = n^{-\frac{1}{2}(N-1)} \prod_{i=1}^{N} [n^{-1} n_i(\omega)]^{n_i(\omega)}$$

$$d_{n,\omega} = [(n + N - 1)!]^{-1} \prod_{i=1}^{N} [n_i(\omega)!]$$

$$M_{n,\omega} = \{\lambda \mid \lambda \varepsilon L_{N-1} ; \phi_{n,\omega}(\lambda) \varepsilon M\}$$

$$u_{n,\omega}(\mathbf{y}) = n^{-1} n_N(\omega) - n^{-\frac{1}{2}} s(\mathbf{y})$$

$$h_{n,\omega}(\mathbf{y}) = -\sum_{i=1}^{N-1} n_i(\omega) \log [1 + n^{\frac{1}{2}} n_i(\omega)^{-1} y_i] - n_N(\omega) \log [1 - n^{\frac{1}{2}} n_N(\omega)^{-1} s(\mathbf{y})]$$

$$f_{n,\omega}(\mathbf{y}) = f[n^{-1} n_i(\omega) + n^{-\frac{1}{2}} y_i, 1 \leqq i \leqq N - 1] \quad \text{Case A}$$

$$f_{n,\omega}(\mathbf{y}; x) = f[n^{-1} n_i(\omega) + n^{-\frac{1}{2}} y_i, 1 \leqq i \leqq N - 1; x] \quad \text{Case B,}$$

with $f(\lambda) = 0$ for $\lambda \varepsilon L_N - T$.

Let ϵ be a positive number so small that

(i) $|x| \leqq \epsilon$ implies $\log (1 + x) \leqq x - \frac{1}{4}x^2$,

(ii) if $V = \{\lambda \mid \lambda \varepsilon L_{N-1} ; |\lambda(i) - \theta(i)| < 2\epsilon, 1 \leqq i \leqq N - 1\}$ then $V \subset T \cap \{\lambda \mid \lambda \varepsilon L_{N-1} ; \frac{1}{2}f(\theta*) < f(\lambda) < 2f(\theta*)\}$ in Case A and $\{\lambda \mid \lambda \varepsilon L_N ; \lambda* \varepsilon V\} \subset T \cap \{\lambda \mid \lambda \varepsilon L_N ; \frac{1}{2}f(\theta^+) < f(\lambda) < 2f(\theta^+)\}$ in Case B.

Let

$$D_n = \{\mathbf{y} \mid \mathbf{y} \varepsilon R^{N-1}; |y_i| < \epsilon n^{\frac{1}{2}}, i \leqq i \leqq N - 1; |s(\mathbf{y})| < \epsilon n^{\frac{1}{2}}\}$$

$$F_{n,\omega}(\mathbf{y}) = n u_{n,\omega}(\mathbf{y})^{-n_N(\omega)} \exp [-h_{n,\omega}(\mathbf{y})] \int_{u_{n,\omega}(\mathbf{y})-\epsilon}^{u_{n,\omega}(\mathbf{y})} f_{n,\omega}(\mathbf{y}; x)\, x^{n_N(\omega)}\, dx$$

$$T_{n,\omega}(\mathbf{y}) = n u_{n,\omega}(\mathbf{y})^{-n_N(\omega)} \int_{u_{n,\omega}(\mathbf{y})-\epsilon}^{u_{n,\omega}(\mathbf{y})} x^{n_N(\omega)}\, dx.$$

By discarding a P_θ-null set, we suppose throughout the proofs: $n^{-1}n_i(\omega) \rightarrow \theta(i)$, $1 \leqq i \leqq N$; and (15)—so (7)—holds for all L_N-neighborhoods V of θ^+, with δ depending on V. See Remark 1, Section 2.

PROOF OF THEOREM 3, CASE A. For large n, the set $M_{n,\omega} \subset T$ of Condition (iii), so the $\mu_{n,\omega}$-distribution of $\phi_{n,\omega}$, restricted to M, has density

$$b_{n,\omega}^{-1} c_{n,\omega} f_{n,\omega}(\mathbf{y}) \exp[-h_{n,\omega}(\mathbf{y})]. \tag{17}$$

From (7), $b_{n,\omega}$ is asymptotically equivalent to $f(\theta*)$ times

$$\int_{L_{N-1}} \left[\prod_{i=1}^{N-1} \lambda(i)^{n_i(\omega)}\right] [1 - s(\lambda)]^{n_N(\omega)}\, d\lambda = d_{n,\omega}. \tag{18}$$

Finally,

$$\lim_{n\to\infty} f_{n,\omega}(\mathbf{y}) = f(\theta*); \tag{19}$$

$$\lim_{n\to\infty} h_{n,\omega}(\mathbf{y}) = h(\mathbf{y}); \tag{20}$$

$$\lim_{n\to\infty} c_{n,\omega}/d_{n,\omega} = [\theta(1) \cdots \theta(N)]^{-1} (2\pi)^{-\frac{1}{2}(N-1)}. \tag{21}$$

Here (19) and (20) hold uniformly over $\mathbf{y} \in M$; (20) may be verified by considering Taylor's expansion of $\log(1 + x)$; (19) follows trivially from assumption (iii); and (21) is a consequence of Stirling's formula. This completes the proof of Theorem 3, Case A.

PROOF OF THEOREM 3, CASE B. The $\mu_{n,\omega}$-distribution of $\phi_{n,\omega}$ is the sum of two measures: the first, $\rho_{n,\omega}$, is the distribution measure of $\phi_{n,\omega}$ under $\nu_{n,\omega}$ restricted to T; the second, $\sigma_{n,\omega}$, is the distribution measure of $\phi_{n,\omega}$ under $\nu_{n,\omega}$ restricted to $L_N - T$. Of course, $\sigma_{n,\omega}(R^{N-1}) \to 0$ by (7). The density of $\rho_{n,\omega}$ is $b_{n,\omega}^{-1} n^{-\frac{1}{2}(N-1)} \prod_{i=1}^{N-1} [n^{-1} n_i(\omega) + n^{-\frac{1}{2}} y_i]^{n_i(\omega)}$ times

$$\int_0^{u_{n,\omega}(\mathbf{y})} f_{n,\omega}(\mathbf{y}; x)\, x^{n_N(\omega)}\, dx. \tag{22}$$

If $0 < \alpha < \beta < 1$, then $\lim_{n\to\infty} \int_0^{\beta-\alpha} x^n\, dx / \int_{\beta-\alpha}^{\beta} x^n\, dx = 0$; from this and regularity Condition (iv) it follows easily that (22) is asymptotically equivalent to $f(\theta^+) \int_0^{u_{n,\omega}(\mathbf{y})} x^{n_N(\omega)}\, dx$, uniformly on M. As usual, $b_{n,\omega}$ is essentially $f(\theta^+) \int_{L_N} \prod_{i=1}^{N} \lambda(i)^{n_i(\omega)}\, d\lambda$. By a trival modification of the argument for Case A, the density of $\rho_{n,\omega}$ converges to (16) uniformly on compact subsets of R^{N-1}, which completes the proof of Theorem 3.

PROOF OF THEOREM 4, CASE A. Let V_1 and V_2 be L_N-neighborhoods of θ^+. Then (15) implies

$$\lim_{n\to\infty} n^{\frac{1}{2}} \left[\frac{\int_{V_1} \lambda(i) \prod_{j=1}^{n} \lambda(X_j)\nu(d\lambda)}{\int_{V_2} \prod_{j=1}^{n} \lambda(X_j)\nu(d\lambda)} - \beta_{n,\omega}(i) \right] = 0. \tag{23}$$

Hence (23) holds when V_1 and V_2 are replaced by $\{\lambda \mid \lambda \in L_N;\ |\lambda(i) - n^{-1}n_i(\omega)| < \epsilon,\ 1 \leqq i \leqq N-1;\ |\sum_{i=1}^{N-1} \lambda(i) - n^{-1}n_i(\omega)| < \epsilon\}$. With this substitution, the first ratio in (23) is eventually

$$n^{-1} n_i(\omega) + n^{-\frac{1}{2}} \frac{\int_{D_n} y_i f_{n,\omega}(\mathbf{y}) \exp[-h_{n,\omega}(\mathbf{y})]\, d\mathbf{y}}{\int_{D_n} f_{n,\omega}(\mathbf{y}) \exp[-h_{n,\omega}(\mathbf{y})]\, d\mathbf{y}};$$

to see this, use Condition (ii) on ϵ and change variables in the obvious way.

In view of (19) and (20), the denominator of the last fraction has a positive lim inf. The theorem follows, therefore, from

$$\lim_{n\to\infty} \int_{D_n} y_i f_{n,\omega}(\mathbf{y}) \exp[-h_{n,\omega}(\mathbf{y})]\, d\mathbf{y} = 0. \tag{24}$$

For large n, we have $M \subset D_n$. Write the integral in (24) as an integral over M plus an integral over $D_n - M$. By (19), (20) and symmetry, the first integral converges to 0. Eventually, $nn_i(\omega)^{-1} \geqq [2\theta(i)]^{-1}$; then by Condition (i) on ϵ,

$$h_{n,\omega}(\mathbf{y}) \geqq \tfrac{1}{4}h(\mathbf{y}). \tag{25}$$

Using Condition (ii) on ϵ, the integral over $D_n - M$ is eventually bounded in absolute value by $2f(\theta*) \int_{\mathbf{y}\notin M} |y_i| \exp[-\frac{1}{4}h(\mathbf{y})]\, d\mathbf{y}$, which can be made arbitrarily small by choosing m sufficiently large. This completes the proof of (24), and so of Theorem 4, Case A.

Proof of Theorem 4, Case B. Equation (23) holds with V_1 and V_2 replaced by

$$\Big\{\lambda \mid \lambda \varepsilon L_N ; | \lambda(i) - n^{-1} n_i(\omega) | < \epsilon, 1 \leqq i \leqq N - 1; \quad \left|\sum_{i=1}^{N-1} \lambda(i) - n^{-1}n_i(\omega)\right| < \epsilon; \sum_{i=1}^{N} \lambda(i) > 1 - \epsilon\Big\}.$$

Using regularity Condition (iv), Condition (ii) on ϵ, and changing variables, we see that the first ratio in (23), after the indicated substitution, is eventually $n^{-1}n_i(\omega) + n^{-\frac{1}{2}}[\int_{D_n} y_i F_{n,\omega}(\mathbf{y})\, d\mathbf{y} / \int_{D_n} F_{n,\omega}(\mathbf{y})\, d\mathbf{y}]$. We will prove that the numerator of the last fraction converges to zero, and omit the easier proof that its denominator has a positive lim inf. Split the numerator into an integral over M and an integral over $D_n - M$. As in estimating (22), $\lim_{n\to\infty} T_{n,\omega}(\mathbf{y}) = 1$ uniformly on M, so that $\lim_{n\to\infty} F_{n,\omega}(\mathbf{y}) = f(\theta^+) \exp[-h(\mathbf{y})]$ uniformly on M by (20). Hence the integral over M converges to 0 by symmetry. Using (25) and Condition (ii) on ϵ, eventually the integral over $D_n - M$ is bounded in absolute value by

$$2f(\theta^+) \int_{D_n - M} |y_i| \exp[-\tfrac{1}{4} h(\mathbf{y})]\, T_{n,\omega}(\mathbf{y})\, d\mathbf{y}. \tag{26}$$

Changing the lower limit of integration in its definition to 0 implies $T_{n,\omega}(\mathbf{y}) \leqq 1 - n^{\frac{1}{2}}(n_N(\omega) + 1)^{-1}s(\mathbf{y})$, so (26) is bounded by $2f(\theta^+)$ times

$$\int_{\mathbf{y}\notin M} |y_i| \exp[-\tfrac{1}{4} h(\mathbf{y})]\, d\mathbf{y} + n^{\frac{1}{2}}(n_N(\omega) + 1)^{-1} \cdot \int_{\mathbf{y}\notin M} |y_i|\, |s(\mathbf{y})| \exp[-\tfrac{1}{4} h(\mathbf{y})]\, d\mathbf{y}.$$

The lim $\sup_{n\to\infty}$ of (26) may therefore be made arbitrarily small by choosing m sufficiently large, completing the proof of Theorem 4.

PROOF OF COROLLARY 2. In Case A, this merely summarizes the previous work. In Case B, we must prove that for any $\delta > 0$: (i) $\mu_{n,\omega}\{\lambda \mid n^{\frac{1}{2}}|\lambda(i) - \beta_{n,\omega}(i)| \geqq \delta\} \to 0$ for $i \leqq 0$ and $i > N$; and (ii) $\mu_{n,\omega}\{\lambda : |n^{\frac{1}{2}} \sum_{i=1}^{N} (\lambda(i) - \beta_{n,\omega}(i))| \geqq \delta\} \to 0$. Because of the inequalities

$$\sum_{i \leqq 0, i > N} \beta_{n,\omega}(i) \leqq 1 - \sum_{i=1}^{N} \beta_{n,\omega}(i)$$

and

$$\sum_{i \leqq 0, i > N} \lambda(i) \leqq 1 - \sum_{i=1}^{N} \lambda(i),$$

it is enough to prove the sharper estimate $\lim_{n\to\infty} \mu_{n,\omega}\{\lambda \mid 1 - \sum_{i=1}^{N} \lambda(i) \geqq n^{-1}y\} = e^{-y}$, $0 < y < \infty$. In view of (15), the left side coincides with the limit of

$$\int_{\{\lambda \mid \lambda \varepsilon L_N; \Sigma_{i=1}^{N} \lambda(i) \leqq 1 - n^{-1}y\}} \prod_{i=1}^{N} \lambda(i)^{n_i(\omega)}\, d\lambda \Big/ \int_{L_N} \prod_{i=1}^{N} \lambda(i)^{n_i(\omega)}\, d\lambda.$$

Changing variables in the numerator, this fraction becomes

$$(1 - n^{-1}y)^N \prod_{i=1}^{N} (1 - n^{-1}y)^{n_i(\omega)},$$

which does converge to e^{-y}.

5. Bayes' estimates are inconsistent. Surprisingly, Theorem 1 is false without the assumption "$\{i \mid \theta(i) > 0\}$ is finite." A large class of counter-examples will be constructed in Theorem 5, but here is one relatively easy to visualize. Let $I = \{0, 1, \cdots\}$ and let M map $[\frac{1}{8}, \frac{7}{8}]$ into Λ by assigning to x the geometric distribution with parameter x, truncated at $f(x)$. More precisely, let $f(\frac{1}{4}) = f(\frac{3}{4}) = \infty$; elsewhere on $[\frac{1}{8}, \frac{7}{8}]$, the function f is natural number valued. It is nondecreasing in $[\frac{1}{8}, \frac{1}{4})$ and in $(\frac{1}{2}, \frac{3}{4})$; nonincreasing in $(\frac{1}{4}, \frac{1}{2})$ and in $(\frac{3}{4}, \frac{7}{8}]$. Then let $M(x)(i) = (1 - x)x^i$, $0 \leqq i < f(x)$; $= 0$, $i > f(x) + 1$. For x unequal to $\frac{1}{4}$ or $\frac{3}{4}$, choose $M(x)[f(x)]$ and $M(x)[f(x) + 1]$ so that $M(x) \varepsilon \Lambda$ and $x \to M(x)(i)$ is continuous (or C_∞, if it seems relevant) for $\frac{1}{8} \leqq x \leqq \frac{7}{8}$ and $0 \leqq i < \infty$. Since M is a homeomorphism of $[\frac{1}{8}, \frac{7}{8}]$ into Λ, its range is an arc containing $\theta = M(\frac{1}{4})$ and $q = M(\frac{3}{4})$ as interior points (i.e., in the relative topology). Let μ be the image by M of the uniform distribution on $[\frac{1}{8}, \frac{7}{8}]$, so that μ is a continuous probability on the arc and assigns positive mass to each neighborhood of θ. If f grows quickly at $\frac{3}{4}$ (e.g., $\lim_{x\to\frac{3}{4}} f(x)|x - \frac{3}{4}| = 1$) and slowly at $\frac{1}{4}$ (e.g., $\lim_{x\to\frac{1}{4}} f(x)/\{\log_4 \log_4 (|x - \frac{1}{4}|^{-1})\} = \frac{1}{2}$), then (θ, μ) is not consistent. Indeed, if $U_n = \max (X_1 \cdots X_n)$, then $\mu_{n,\omega}$ concentrates in the M-image of $\{x \mid \frac{1}{8} \leqq x \leqq \frac{7}{8}, f(x) \geqq U_n(\omega) - 1\}$. For large n, the last set consists of two intervals. The first contains $\frac{1}{4}$ and has width approximately 4^{-n}. The second contains $\frac{3}{4}$ and has width approximately $(\log_4 n)^{-1}$. The posterior mass in these intervals can be estimated, and indeed $\lim_{n\to\infty} \mu_{n,\omega} = \delta_q$ for P_θ-almost all ω.

We omit the details, since they parallel the proof of Theorem 5, and mention only that U_n is essentially $\log_4 n$ under P_θ. Of course, the example can be modified so that $x \to M(x)$ (i) is positive and analytic on $[\frac{1}{8}, \frac{7}{8}]$ for all i.

It is helpful to notice that if Λ is given the l_1-metric, then M distorts distance near $\frac{1}{4}$ and $\frac{3}{4}$. Indeed, if $|x - \frac{1}{4}|$ is small, both the distance and the arc length from $M(x)$ to $M(\frac{1}{4})$ are of the order $[\log_4 (|x - \frac{1}{4}|^{-1})]^{-1}$, rather than $|x - \frac{1}{4}|$. We know very little about arcs free from such distortion. Suppose the probability μ concentrates in a smooth arc A in Λ of finite length, and has a continuous positive density with respect to arc length. By Theorem 7.6 on page 48 of Schwartz (1961), $\lim_{n\to\infty} \beta_{n,\omega} = \theta$ in P_θ-probability for all $\theta \varepsilon A$. It is unknown whether (θ, μ) is consistent, or even whether $\lim_{n\to\infty} \beta_{n,\omega} = \theta$ a.s. $[P_\theta]$, for all $\theta \varepsilon A$.

THEOREM 5. *Let $\theta \varepsilon \Lambda$ and let $\{i \mid \theta(i) > 0\}$ be infinite. Let $q \varepsilon L - \{\theta\}$. There exists a probability μ on $\mathcal{B}$ with*

(i) $\mu(\Lambda) = 1$ *and* $\theta \varepsilon C(\mu)$;

(ii) *if* $\lambda \varepsilon C(\mu) - \{q\}$ *then* $\lambda \varepsilon \Lambda$ *and* $\theta \ll \lambda \ll \theta + q$;

(iii) $\lim_{n\to\infty} \mu_{n,\omega} = \delta_q$ *for P_θ-almost all* ω.

PROOF. We begin by outlining the program. The probability μ will assign mass α_m to $\theta_m \varepsilon \Lambda$ and β_m to $q_m \varepsilon \Lambda$, for $1 \leqq m < \infty$; with $\alpha_m > 0$ and $\beta_m > 0$ eventually, and $\sum_{m=1}^{\infty} (\alpha_m + \beta_m) = 1$. The vectors θ_m and q_m will be chosen so that $\theta \ll \theta_m \ll \theta + q$; $\theta_m \to \theta$; $\theta \ll q_m \ll \theta + q$; $q_m \to q$. Then μ will satisfy Conditions (i) and (ii). To secure (iii), q_m will assign small probabilities (by comparison with θ) to remote states. Each q_m will have one fewer small coordinate than q_{m+1}; as θ is sampled and the corresponding state observed, q_{m+1} will become progressively more plausible than q_m. The posterior mass in $\{q_m : 1 \leqq m < \infty\}$ will therefore shift toward the tail, that is, toward q. Each θ_m will also assign small probabilities to remote states, which, when observed, make θ_m unlikely; but α_m will decrease so fast that the tail of $\{\theta_m : 1 \leqq m < \infty\}$ will also become unlikely. Over-all, posterior mass will transfer from $\{\theta_m : 1 \leqq m < \infty\}$ to $\{q_m : 1 \leqq m < \infty\}$, making $\mu_{n,\omega} \to \delta_q$ a.s. $[P_\theta]$.

We will now give the details, beginning with the β_m and q_m. As usual, we enumerate I so that $\theta(i) > 0$ for $1 \leqq i < \infty$ while $\theta(i) = 0$ for $i < 1$. The latter index set may be empty or finite. We distinguish two cases: in Case A, $q \varepsilon \Lambda$; in Case B, $q \not\varepsilon \Lambda$.

In Case A, let m_0 be the least integer $m \geqq 1$ for which $\sum_{i \leqq m} q(i) > 0$. Write $K_0 = 2/\sum_{i \leqq m_0} q(i)$. If $m < m_0$ put $\beta_m = 0$; if $m \geqq m_0$, the β_m are arbitrary subject to

(c1) $\beta_m > 0, m \geqq m_0$,

(c2) $\sum_{m=m_0}^{\infty} \beta_m < 1$.

Write $\delta_m = 1/(m2^{m+1})$; then $\sup_{m \geqq m_0} (\delta_m/\delta_{m+1}) \leqq 4$ For $i \geqq m_0$ let $e(i)$ be arbitrary, subject to

(c3a) $0 < e(i) < 2^{-(i+1)}$;

(c4a) $[e(i)/\delta_i]^{\theta(i)} \leqq (K_0 + 4)^{-4}$.

For $m \geqq m_0$, equation (c3a) implies $m\delta_m + \sum_{i=m+1}^{\infty} e(i) < 2^{-m} < 1$; hence there is a unique φ_m satisfying

$$\varphi_m \sum_{i \leqq m} q(i) + m\delta_m + \sum_{i=m+1}^{\infty} e(i) = 1.$$

Then $\lim_{m\to\infty} \varphi_m = 1$; and for $m \geqq m_0$,

$$[2\sum_{i \leqq m} q(i)]^{-1} \leqq \varphi_m \leqq [\sum_{i \leqq m} q(i)]^{-1},$$

implying $\varphi_m/\varphi_{m+1} \leqq K_0$.

For $1 \leqq m < m_0$ let $q_m = q_{m_0}$. For $m \geqq m_0$, let

$$\begin{aligned} q_m(i) &= \varphi_m q(i), & i &\leqq 0,\\ &= \varphi_m q(i) + \delta_m, & 1 &\leqq i \leqq m,\\ &= e(i), & i &\geqq m+1. \end{aligned}$$

It is easy to verify that $q_m \varepsilon \Lambda$; $\theta \ll q_m \ll \theta + q$; $\lim_{m\to\infty} q_m = q$. For each $m \geqq m_0$, for P_θ-almost all ω, eventually $n_{m+1}(\omega) \geqq \frac{1}{2}n\theta(m+1)$; and then, by (c4a),

$$\begin{aligned} \frac{\mu_{n,\omega}\{q_m\}}{\mu_{n,\omega}\{q_{m+1}\}} &= \frac{\beta_m}{\beta_{m+1}} \left\{ \prod_{i=1}^{m} \left[\frac{\varphi_m q(i) + \delta_m}{\varphi_{m+1} q(i) + \delta_{m+1}} \right]^{n_i(\omega)} \right\} \cdot \left[\frac{e(m+1)}{\varphi_{m+1} q(m+1) + \delta_{m+1}} \right]^{n_{m+1}(\omega)} \\ &\leqq \frac{\beta_m}{\beta_{m+1}} \left\{ \prod_{i=1}^{m} \left[\frac{\varphi_m}{\varphi_{m+1}} + \frac{\delta_m}{\delta_{m+1}} \right]^{n_i(\omega)} \right\} \left[\frac{e(m+1)}{m+1} \right]^{n_{m+1}(\omega)} \\ &\leqq \frac{\beta_m}{\beta_{m+1}} (K_0 + 4)^n (K_0 + 4)^{-2n}, \end{aligned} \tag{27a}$$

which converges to 0 as $n \to \infty$. This completes the discussion of β_m and θ_m in Case A.

In Case B, let $m_0 = 1$. Let β_m be arbitrary, subject to (c1) and (c2). Let $\Delta = 1 - \sum_{i\varepsilon I} q(i) > 0$. For $i \geqq 1$ let $e(i)$ be arbitrary, subject to:

(c3b) $0 < e(i) < \Delta 2^{-i-1}$;

(c4b) $[2ie(i)/\Delta]^{\theta(i)} \leqq (1 + 4\Delta^{-1})^{-4}$.

By (c3b), $\sum_{i \leqq m} q(i) + \sum_{i=m+1}^{\infty} e(i) < 1$, and there is a unique δ_m satisfying

$$\sum_{i \leqq m} q(i) + m\delta_m + \sum_{i=m+1}^{\infty} e(i) = 1.$$

Again by (c3b), $2^{-1}\Delta m^{-1} \leqq \delta_m \leqq m^{-1}$, and $\delta_m/\delta_{m+1} \leqq 4\Delta^{-1}$. Let

$$\begin{aligned} q_m(i) &= q(i), & i &\leqq 0,\\ &= q(i) + \delta_m, & 1 &\leqq i \leqq m,\\ &= e(i), & i &\geqq m+1. \end{aligned}$$

Then $q_m \varepsilon \Lambda$; $\theta \ll q_m \ll \theta + q$; $q_m \to q$.

When $n_{m+1}(\omega) \geqq \frac{1}{2}n\theta(m+1)$, by (c4b)

$$\text{(27b)}\quad \begin{aligned}\frac{\mu_{n,\omega}\{q_m\}}{\mu_{n,\omega}\{q_{m+1}\}} &= \frac{\beta_m}{\beta_{m+1}}\left\{\prod_{i=1}^{m}\left[\frac{q(i)+\delta_m}{q(i)+\delta_{m+1}}\right]^{n_i(\omega)}\right\}\left[\frac{e(m+1)}{q(m+1)+\delta_{m+1}}\right]^{n_{m+1}(\omega)} \\ &\leqq \frac{\beta_m}{\beta_{m+1}}(1+4\Delta^{-1})^n(1+4\Delta^{-1})^{-2n},\end{aligned}$$

which converges to 0 as $n \to \infty$. This completes the discussion of β_m and q_m in Case B.

We turn now to the θ_m. For each n, the function $\sum_{m=1}^{\infty}\beta_m\prod_{j=1}^{n}q_m(X_j)$ is positive a.s. $[P_\theta]$; consequently, there is a real-valued function h on the natural numbers with $h(n)$ decreasing to 0 as n increases, and $\sum_{m=1}^{\infty}\beta_m\prod_{j=1}^{n}q_m(X_j) > h(n)$ eventually a.s. $[P_\theta]$. Moreover, if $U_n = \max[X_1 \cdots X_n]$, then $1 \leqq U_n \uparrow \infty$ a.s. $[P_\theta]$—*because* $\{i \mid \theta(i) > 0\}$ *is infinite*—and there is a nondecreasing, unbounded natural number valued function g on the natural numbers for which $U_n > g(n)$ eventually a.s. $[P_\theta]$. Then we construct a sequence $\{\epsilon(i) : 1 \leqq i < \infty\}$ arbitrary, subject to:

(c5) $\epsilon(i) \downarrow 0$ as $i \uparrow \infty$;
(c6) $\sum_{i=1}^{\infty}\epsilon(i) < \theta(i)$;
(c7) $\epsilon[g(n)] \leqq \frac{1}{2}h(n)^2$.

We determine c_m uniquely from the relation $c_m + \sum_{i=2}^{m}\theta(i) + \sum_{i=m+1}^{\infty}\epsilon(i) = 1$, so that c_m is positive and converges to $\theta(1)$ by (c6). We define

$$\begin{aligned}\theta_m(i) &= 0, & i &\leqq 0,\\ &= c_m, & i &= 1,\\ &= \theta(i), & 2 &\leqq i \leqq m,\\ &= \epsilon(i), & i &\geqq m+1.\end{aligned}$$

Then $\theta_m \varepsilon \Lambda$; $\theta \ll \theta_m \ll \theta + q$; $\theta_m \to \theta$. We take α_m positive and arbitrary, subject to:

(c8) $\sum_{m=1}^{\infty}(\alpha_m + \beta_m) = 1$;
(c9) $\sum_{m=g(n)}^{\infty}\alpha_m \leqq \frac{1}{2}h(n)^2$ eventually.

This completes the construction; (i) and (ii) of the theorem certainly hold, and we will now verify (iii). Let $C = \{\theta_m : 1 \leqq m < \infty; q_m : 1 \leqq m < \infty; q\}$ and $C_k = \{q_m : m \geqq k; q\}$. Then C is the topological carrier of μ and $\mu_{n,\omega}$ for P_θ-almost all ω. Because the C_k are compact, Condition (iii) holds provided $\lim_{n\to\infty}\mu_{n,\omega}(C_k) = 1$ a.s. $[P_\theta]$ for all k. But

$$\mu_{n,\omega}(C_{m_0}) \geqq \frac{\sum_{m=m_0}^{\infty}\beta_m\prod_{j=1}^{n}q_m(X_j)}{\sum_{m=1}^{\infty}\alpha_m\prod_{j=1}^{n}\theta_m(X_j) + \sum_{m=m_0}^{\infty}\beta_m\prod_{j=1}^{n}q_m(X_j)}.$$

Hence $\mu_{n,\omega}(C_{m_0}) \to 1$ a.s. $[P_\theta]$ provided

$$\text{(28)}\quad \lim_{n\to\infty}\sum_{m=1}^{\infty}\alpha_m\prod_{j=1}^{n}\theta_m(X_j)\Big/\sum_{m=m_0}^{\infty}\beta_m\prod_{j=1}^{n}q_m(X_j) = 0 \quad \text{a.s. } [P_\theta].$$

Now the denominator of this expression ultimately exceeds $h(n)$ a.s. $[P_\theta]$. On the other hand, we have chosen the parameters so that the numerator is eventually bounded by $h(n)^2$, a.s. $[P_\theta]$. To see this, write the numerator as $(\sum_{m=1}^{U_n-1} + \sum_{m=U_n}^{\infty}) \alpha_m \prod_{j=1}^{n} \theta_m(X_j)$. But $\sum_{m=U_n}^{\infty} \alpha_m \prod_{j=1}^{n} \theta_m(X_j)$ is eventually bounded by $\sum_{m=g(n)}^{\infty} \alpha_m$, a.s. $[P_\theta]$, and this sum is (c9) in turn eventually bounded by $\frac{1}{2}h(n)^2$. Finally, $\sum_{m=1}^{U_n-1} \alpha_m \prod_{j=1}^{n} \theta_m(X_j) \leqq \sum_{m=1}^{U_n-1} \alpha_m \epsilon(U_n) \leqq \epsilon[g(n)] \leqq \frac{1}{2}h(n)^2$ eventually a.s. $[P_\theta]$. Here the first bound comes from the definition of θ_m; the second from (c8) and the definition of g; and the last from (c7). This proves (28), and $\lim_{n\to\infty} \mu_{n,\omega}(C_{m_0}) = 1$ a.s. $[P_\theta]$. Estimates (27a, b) imply $\lim_{n\to\infty} \mu_{n,\omega}(C_{m_0} - C_m) = 0$ a.s. $[P_\theta]$ for $m \geqq m_0$, completing the proof.

A number of conclusions can be made by slight modification of the example. For one thing, there is a second probability $\hat{\mu}$ on $\mathcal{B}$ satisfying (i), (ii); equivalent to μ and agreeing with μ when restricted to a neighborhood of θ; but for which $(\theta, \hat{\mu})$ is consistent. It is obtained by making the β_m decrease much more quickly. We conclude:

REMARK 5. For infinite state space (as opposed to finite) consistency is not a local property, nor does it depend merely on the null sets of the prior.

Point masses were used to simplify the computation. Because Λ is convex, they can be eliminated without difficulty: join θ_m to θ_{m+1}, $1 \leqq m < \infty$ by a line segment, and let μ assign mass α_m to that segment, distributed uniformly over it. Join θ_1 to q_{m_0} by a line segment carrying mass β_{m_0} in a uniform way; join q_m to q_{m+1}, $m_0 \leqq m < \infty$ by a line segment of total uniform mass β_{m+1}. Thus θ has been joined to q by a polygonal arc (with an infinite number of corners) which carries μ. By constructing more θ_m's and q_m's, we can continue the arc past q and past θ, if desired. The theorem continues to hold, with substantially the same proof. Of course, the arc can be made smooth.

If $q \equiv \theta$ (i.e., $q(i) > 0$ if and only if $\theta(i) > 0$), by modifying the construction slightly it is possible to have the carrier of μ precisely $\{\lambda \mid \lambda \in L \text{ and } \lambda \ll \theta\}$, while (i) and (iii) hold. If also $q \in \Lambda$, then all the q_m may be replaced by q. If $q \in \Lambda$, is equivalent to θ, and $H(\theta) < \infty$, an easy argument shows that θ_m and q_m can be chosen to have finite entropy relative to θ. Of course (see Theorem 2), $H(\theta_m \mid \theta)$ does not converge to $H(\theta)$. If Condition (ii) is dropped, then $e(i)$ and $\epsilon(i)$ may be put equal to 0, and the estimates are simplified.

Using substantially the same ideas, we proved:

REMARK 6. There is a continuous probability μ on $\mathcal{B}$ concentrated in Λ and assigning positive mass to every open subset of Λ, for which $\{\theta \mid \theta \in \Lambda;\ (\theta, \mu) \text{ is consistent}\}$ is of the first category in Λ. Even the larger $\{\theta \mid \theta \in \Lambda;\ \beta_{n,\omega} \to \theta \text{ in } P_\theta \text{ probability}\}$ is first category.

6. Tail-free measures. In this section we develop a very simple condition on priors ensuring that the results of Sections 3 and 4 go over to the countable case. The effect of this condition is to insist the prior be open-minded about the tails of the $\{\lambda \mid \lambda \in \Lambda\}$. The condition depends on the ordering of I; let us suppose, therefore, that $I = \{1, 2, \cdots\}$. Let $S_k(\lambda) = \sum_{i=1}^{k} \lambda(i)$, $1 \leqq k < \infty$, $\lambda \in L$.

DEFINITION 2. The probability μ on $\mathcal{B}$ is called tail-free provided:
(i) $\mu\{\lambda \mid \lambda \varepsilon L; S_k(\lambda) < 1\} = 1$ for all k;
(ii) there is a natural number $N = [\mu]$ such that the function $\lambda \to \{\lambda(i): 1 \leqq i \leqq N\}$ and all the functions $\lambda \to [1 - S_{N+k-1}(\lambda)]^{-1}\lambda(N + k)$, $1 \leqq k < \infty$ are mutually independent under μ.

Condition (i) is not essential, and is imposed for expository convenience. The μ-distribution probability of $\lambda \to [\lambda(i): 1 \leqq i \leqq N]$ will be denoted $\mu^{(N)}$; the μ-distribution probability of $\lambda \to [1 - S_{N+k-1}(\lambda)]^{-1}\lambda(N + k)$ will be denoted $\mu^{(N+k)}$, $1 \leqq k < \infty$. Thus $\mu^{(N)}(L_N - \Lambda_N) = 1$; $\mu^{(N+k)}[0, 1) = 1$, $1 \leqq k < \infty$. Conversely, given an N and probabilities $m^{(N+k)}$, $0 \leqq k < \infty$ with $m^{(N)}(L_N - \Lambda_N) = 1$ and $m^{(N+k)}[0, 1) = 1$, $1 \leqq k < \infty$, there is a unique tail-free probability μ on $\mathcal{B}$ for which: $[\mu] = N$; $m^{(N+k)} = \mu^{(N+k)}$, $0 \leqq k < \infty$.

As an easy consequence of Condition (ii), for each $k \geqq 0$ the functions

$$\lambda \to \{\lambda(i) : 1 \leqq i \leqq N + k\}$$

$$\lambda \to \{[1 - S_{N+k}(\lambda)]^{-1}\lambda(N + k + i) : 1 \leqq i < \infty\}$$

are independent under μ. The second function maps L into L, so maps μ into a probability $\hat{\mu}$. Of course, $\hat{\mu}$ is tail-free with $[\hat{\mu}] = 1$ and $\hat{\mu}^{(i)} = \mu^{(N+k+i)}$, $1 \leqq i < \infty$.

Let $k \geqq 0$, $n_m \geqq 0$, $N + k + 1 \leqq m \leqq M$; $s_i = \sum_{m=N+i}^{M} n_m$, $k + 1 \leqq i \leqq M - N$. Let C_r denote the class of functions G on L of the form $G(\lambda) = g[\lambda(i) : 1 \leqq i \leqq r]$, where g is a real continuous function on L_r.

It is almost obvious from the previous remark that if $G \varepsilon C_{N+k}$ then

$$\begin{aligned}(29)\quad &\int_L G(\lambda) \prod_{m=N+k+1}^{M} \lambda(m)^{n_m}\mu(d\lambda) \\ &= \left[\prod_{i=k+1}^{M-N} \int_0^1 x^{n_{N+i}}(1 - x)^{s_{i+1}}\mu^{(N+i)}(dx)\right] \int_L G(\lambda)[1 - S_{N+k}(\lambda)]^{s_{k+1}}\mu(d\lambda).\end{aligned}$$

By trivial manipulations,

$$(30)\quad \int_L G(\lambda)\mu_{n,\omega}(d\lambda) = \frac{\int_L G(\lambda)\left[\prod_{i=1}^{N+k} \lambda(i)^{n_i(\omega)}\right][1 - S_{N+k}(\lambda)]^{n-\Sigma_{i=1}^{N+k}n_i(\omega)}\mu(d\lambda)}{\int_L \left[\prod_{i=1}^{N+k} \lambda(i)^{n_i(\omega)}\right][1 - S_{N+k}(\lambda)]^{n-\Sigma_{i=1}^{N+k}n_i(\omega)}\mu(d\lambda)}$$

in the sense that, if the left side exists so does the right, and they are equal. In particular, if μ is tail-free so is $\mu_{n,\omega}$, with $[\mu] = [\mu_{n,\omega}] = N$, and $\mu_{n,\omega}^{(N+k)}$ is absolutely continuous with respect to $\mu^{(N+k)}$; having density proportional to

$$x \to x^{n_{N+k}(\omega)}(1 - x)^{n-\Sigma_{i=1}^{N+k}n_i(\omega)}$$

on $[0, 1)$ when $k \geqq 1$; and density proportional to

$$\lambda \to \left[\prod_{i=1}^{N} \lambda(i)^{n_i(\omega)}\right]\left[1 - \sum_{i=1}^{N} \lambda(i)\right]^{n-\Sigma_{i=1}^{N}n_i(\omega)}$$

on L_N when $k = 0$.

For a further interpretation of (30), observe that $\mu(\Lambda) = 0$ or 1 according as $\sum_{j=1}^{\infty} \int_0^1 x\mu^{(N+j)}(dx)$ converges or diverges. In the latter case, define the usual joint distribution D_μ for $\{\lambda;\ X_j : 1 \leqq j < \infty\}$ by requiring λ to have D_μ-distribution probability μ, and $\{X_j\ ;\ 1 \leqq j < \infty\}$ to have the D_μ-conditional distribution given λ: the $\{X_j\}$ are independent with common distribution λ. Naturally, $\mu_{n,\omega}$ is the D_μ-conditional distribution of λ given $\{X_j(\omega) : 1 \leqq j \leqq n\}$. The validity of (30) for all $G \varepsilon C_{N+k}$ says: *If μ is tail-free, and $[\mu] = N$, the D_μ-conditional distribution of $\{\lambda(i) : 1 \leqq i \leqq N + k\}$ given $\{n_i(\omega) : 1 \leqq i \leqq N + k\}$ is equal to its D_μ-conditional distribution given $\{X_j(\omega) : 1 \leqq j \leqq n\}$.*

The converse is also true. If (30) holds for all k, all $G \varepsilon C_{N+k}$, all n, then μ is tail-free, $[\mu] = N$. We omit the proof.

THEOREM 6. *Let μ be a tail-free probability on $\mathcal{B}$ and $\theta \varepsilon \Lambda$. Then (θ, μ) is consistent if and only if θ is in the topological carrier of μ.*

PROOF. The necessity is clear. For sufficiency, remember that C_k is separable and $\bigcup_{k=1}^{\infty} C_k$ is dense in the set of continuous functions on L. Hence it is enough to prove that $\int_L G d\mu_{n,\omega} \to G(\theta)$ a.s. $[P_\theta]$ for each k and each $G \varepsilon C_{N+k}$, where $N = [\mu]$. Evaluate the integral by (30). Then express the numerator and denominator of (30) as integrals over Λ_{N+k+1} with respect to the μ-distribution probability of $\lambda \to \{\lambda(i) : 1 \leqq i \leqq N + k;\ 1 - S_{N+k}(\lambda)\}$. An application of Theorem 1 completes the proof.

Let μ be a tail-free probability on $\mathcal{B}$ with $[\mu] = N$; let $\mu^{(N)}$ have a continuous, positive density over L_N ; let $\mu^{(N+i)}$ have a continuous, positive density on [0, 1], for $1 \leqq i < \infty$. If $\theta \varepsilon \Lambda$ then (θ, μ) is consistent by Theorem 6. Much more is true. Suppose θ is nondegenerate, and let H be a proper subset of $\{i \mid \theta(i) > 0\}$, with a finite number h of elements. Let $\phi_{n,\omega}$ map L into R^h by the relation

$$\phi_{n,\omega}(\lambda) = \{n^{\frac{1}{2}}[\lambda(i) - n^{-1}n_i(\omega)] : i \varepsilon H\}.$$

For $\lambda \varepsilon L$ or R^h let $s(\lambda) = \sum_{i\varepsilon H} \lambda(i)$.

THEOREM 7. *For P_θ-almost all ω, the $\mu_{n,\omega}$-distribution of $\phi_{n,\omega}$, when restricted to any fixed compact subset of R^h, is eventually absolutely continuous with a continuous positive density converging uniformly to*

$$\mathbf{y} \to \{[1 - s(\theta)]\prod_{i\varepsilon H} \theta(i)\}^{-\frac{1}{2}}(2\pi)^{-\frac{1}{2}h} \exp\{-\tfrac{1}{2}\sum_{i\varepsilon H} y_i^2/\theta(i) - \tfrac{1}{2}s(\mathbf{y})^2/(1 - s(\theta))\}.$$

PROOF. If $H = \{1, 2, \cdots, N + k\}$ for some $k \geqq 0$, the argument of Theorem 6 reduces the proof to an application of Theorem 3, Case A. We omit the general proof as both routine and tedious.

REMARK 7. Of course, $n^{-1}n_i(\omega)$ could be replaced by $\beta_{n,\omega}(i)$ in the definition of $\phi_{n,\omega}$ without changing the conclusion of the theorem.

Tail-free probabilities are a natural generalization of Dirichlet measures to the infinite dimensional case; as may be seen by taking $[\mu] = 1$ and beta distributions for the $\mu^{(i)}$. If also $\mu^{(i)} = \mu^{(1)}$, $1 \leqq i < \infty$, Theorem 3.1 of Fabius (1963) extends Theorem 7 thus. Let $\{a_i : 1 \leqq i < \infty\}$ be a bounded sequence of real numbers. Then for P_θ-almost all ω, the $\mu_{n,\omega}$-distribution of $\lambda \to n^{\frac{1}{2}} \sum_{i=1}^{\infty} a_i[\lambda(i) -$

$\beta_{n,\omega}(i)]$ converges in the weak $*$ topology to normal with mean 0 and variance $\sum_{i=1}^{\infty} a_i^2\theta(i) - [\sum_{i=1}^{\infty} a_i\theta(i)]^2$.

REFERENCES

[1] Blackwell, David and Dubins, Lester (1962). Merging of opinions with increasing information. *Ann. Math. Statist.* **33** 882–886.

[2] Doob, J. L. (1949). Application of the theory of martingales. *Coll. Int. du CNRS* 22–28 Paris.

[3] Dubins, Lester E. and Freedman, David A. (1963). Random distribution functions. *Bull. Amer. Math. Soc.* **69** 548–557.

[4] Fabius, Jaap (1963). Asymptotic behavior of Bayes' estimates. Thesis, Department of Statistics, University of California at Berkeley.

[5] Freedman, David A. (1962). Les propriétés asymptotiques de solutions de Bayes; cas discret. *C. R. Acad. Sci., Paris* **255** 1851–1852.

[6] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

[7] Laplace, Pierre Simon (1820). *Théorie Analytique des Probabilités*, 3rd ed. Courcier, Paris.

[8] LeCam, Lucien (1957). Locally asymptotically normal families of distributions. *Univ. Calif. Publ. Statist.* **3** 37–98.

[9] LeCam, Lucien (1958). Les propriétés asymptotiques des solutions de Bayes. *Publ. de l'Inst. de Statist. de l'Univ. de Paris*, **7** 17–35.

[10] Mises, R. v. (1931) *Wahrscheinlichkeitsrechnung*, Springer, Berlin.

[11] Mises, R. v. (1938). A modification of Bayes' problem. *Ann. Math. Statist.* **9** 256–259.

[12] Schwartz, Lorraine (1961). Consistency of Bayes' procedures. Thesis, Department of Statistics, University of California at Berkeley.