The Annals of Statistics
1990, Vol. 18, No. 3, 1317–1327

ON THE UNIFORM CONSISTENCY OF BAYES ESTIMATES FOR MULTINOMIAL PROBABILITIES

BY P. DIACONIS[1] AND D. FREEDMAN[2]

Harvard University and University of California

A k-sided die is thrown n times, to estimate the probabilities $\theta_1, \dots, \theta_k$ of landing on the various sides. The MLE of θ is the vector of empirical proportions $p = (p_1, \dots, p_k)$. Consider a set of Bayesians that put uniformly positive prior mass on all reasonable subsets of the parameter space. Their posterior distributions will be uniformly concentrated near p. Sharp bounds are given, using entropy. These bounds apply to all sample sequences: There are no exceptional null sets.

1. Introduction. This paper is about the consistency of Bayes estimates. The usual statement is that for almost all sample sequences, as the sample size goes to ∞ the posterior distribution piles up near the true value of the parameter. The objective is to reformulate this as a finite-sample result, without exceptional null sets or "true values" of parameters.

We begin with coin tossing, and develop an explicit inequality which shows that the posterior must concentrate near the observed fraction of heads. The inequality replaces the asymptotics and eliminates the null set; observed fraction stands in for the true parameter.

To be a little more specific, suppose there are j heads in n tosses of a coin. Consider the posterior odds ratio for a parameter interval of fixed length centered at j/n. The posterior odds are bounded below by ab^n, where $a > 0$ and $b > 1$ are computable constants. So the odds go to ∞ at an exponential rate.

If the prior assigns measure 0 to an interval, so will the posterior. Even if the prior assigns small positive mass to the interval, it may take a long time for the data to swamp the prior. The inequality must therefore take into account the degree to which the prior covers the parameter space.

The notion of "ϕ-positivity" is introduced, to measure coverage; ϕ is a positive function on $(0, 1)$. A prior μ is said to be ϕ-positive if μ assigns mass $\phi(h)$ or more to every closed interval of length h in $[0, 1]$. For example, if $\phi(h) = 0.1h$, then μ is ϕ-positive if and only if μ is bounded below by $0.1 \times$ Lebesgue measure, setwise. Priors with densities which have zeros—like betas—can be handled using more complicated ϕ's; so can singular priors.

The inequality on the posterior odds ratio holds uniformly in ϕ-positive priors μ, and uniformly in the fraction j/n of heads. Take any parameter

Received June 1988; revised August 1989.
[1]Research partially supported by NSF grant DMS 86-00235.
[2]Research partially supported by NSF grant DMS 86-01634.
AMS 1980 *subject classifications*. 62A15, 62E20.
Key words and phrases. Bayes estimates, consistency, Laplace's method, multinomial, Bernstein–von Mises theorem.

interval $[j/n - h, j/n + h]$: The posterior odds ratio for the inside versus the outside is bounded below by

$$\psi(h)e^{2nh^2}. \tag{1.1}$$

Here $\psi(h) > 0$ is computed from ϕ and does not otherwise depend on the prior. In effect, this is a weak law for the posterior, with an exponential error bound. Uniformity in the prior is relevant to arguments about intersubjective agreement; see Diaconis and Freedman (1986).

The rest of this paper is organized as follows: Section 2 gives a careful statement of the result for coin tossing; Section 3 has proofs. The extension to the multinomial is in Section 4, and the last section discusses the idea of ϕ-positivity. For more detailed arguments, see Diaconis and Freedman (1988).

History. In effect, we will estimate the posterior using the method of Laplace (1774); he showed that the posterior piles up near the MLE, but only for the uniform prior. (An easy modern proof uses Chebyshev's inequality, but that was not available to Laplace.) Some modern references on the consistency of Bayes estimates include Le Cam (1953), Le Cam and Schwartz (1960), Schwartz (1965), Freedman (1963) and Diaconis and Freedman (1986). Edwards, Lindman and Savage (1963) must be cited too; their idea was that the data eventually swamp a nondogmatic prior—the principle of stable estimation (pages 201–208).

If there is a smooth prior density with no zeros, better results are available. If j/n is bounded away from 0 or 1, the posterior is asymptotically normal; this result is often called the Bernstein–von Mises theorem—although Laplace got there first; references include Johnson (1967, 1970), Walker (1969), Ghosh, Sinha and Joshi (1982) and Le Cam (1986), Sections 12.3, 12.4 and 17.7. If j/n is close to 0 or 1, the posterior is asymptotically gamma. With some effort, the asymptotics can be converted to uniform estimates and stitched together. Under additional assumptions, higher order correction terms can be calculated as in deBruijn (1981).

2. The theorem for coin tossing. Let ϕ be a positive function on $(0, 1)$. A prior probability μ on $[0, 1]$ is "ϕ-positive" if $\mu[p, p + h] \geq \phi(h)$ for all p and h with $0 \leq p < p + h \leq 1$. For discussion and examples, see Section 5.

Let H be the relative entropy function:

$$H(p, \theta) = -p \log \theta - (1 - p)\log(1 - \theta). \tag{2.1}$$

Here $p = j/n$ is the relative frequency of heads, and θ is the parameter—the probability of heads. (The prior is a distribution over θ.) As is well known,

(2.2) $H(p, \cdot\,)$ is strictly convex, with a strict minimum at p.

For $0 < h < \frac{1}{2}$, let

$$g(h) = \inf_{p,\theta}\{H(p, \theta) - H(p, p): |\theta - p| \geq h\}. \tag{2.3}$$

As will be seen, g is convex, strictly increasing, and $g(h) > 2h^2$.

To state the main result, suppose a coin is tossed n times, and $p = j/n$ is the fraction of heads. Let $0 < h < \frac{1}{2}$. Let $R(n, p, h)$ be the posterior odds ratio for the inside of the parameter interval $[p - h, p + h]$ versus the outside, with respect to a ϕ-positive prior: The outside of the parameter interval is nonempty, because $h < \frac{1}{2}$. Let $0 < \varepsilon < 1$. There is a $\psi(h, \varepsilon) > 0$, which depends on ϕ, h and ε but not on n or p, such that the following inequality holds.

(2.4) THEOREM. $R(n, p, h) \geq \psi(h, \varepsilon)e^{n(1-\varepsilon)g(h)}$ *for* $0 < h < \frac{1}{4}$.

The first factor on the right does not depend on the data. It depends on the prior only through ϕ; it depends on h and ε. The second factor depends on h and ε too; but it depends on the data only through the sample size n. In particular, p is not involved on the right. The bound grows exponentially fast as $n \to \infty$. As it turns out, $\psi(h, \varepsilon)$ is the minimal prior mass in an interval of length about εh^2; more rigorously, $\psi(h, \varepsilon) = \phi(h^*)$, where $h^* = \min\{\frac{1}{2}\varepsilon g(h), h\}$.

The unattainable ideal version of the theorem has $\psi(h, \varepsilon)$ replaced by $\phi(h)$, and $\varepsilon = 0$ in the exponent. On the log scale, these blemishes vanish, as the corollary shows.

(2.5) COROLLARY. $\liminf_{n \to \infty} \inf_{p, \mu}(1/n)\log R(n, p, h) \geq g(h)$.

In (2.5), the prior μ is restricted to be ϕ-positive; $0 < h < \frac{1}{4}$; and $g(h)$ is best possible.

As will be seen, $g(h) > 2h^2$; so, for suitable $\psi(h) > 0$ depending only on ϕ,

(2.6) COROLLARY. $R(n, p, h) \geq \psi(h)e^{2nh^2}$ *for all* n, *all* $p \in [0, 1]$, *all* $h \in (0, \frac{1}{4})$ *and all* ϕ*-positive priors* μ.

To derive (2.6), take $\varepsilon = \varepsilon_h = g(h) - 2h^2$ in (2.4).

3. Proofs for the coin. Fix $0 < h < \frac{1}{4}$. Recall $g(h)$ from (2.3). Confirm that

$$
\begin{aligned}
g(h) &= \min_p \{H(p, p + h) - H(p, p) \colon 0 \leq p < 1 - h\} \\
&= \min_p \{H(p, p - h) - H(p, p) \colon h < p \leq 1\} > 0.
\end{aligned}
\tag{3.1}
$$

Indeed, $p \to H(p, p + h) - H(p, p)$ is continuous on $(0, 1 - h)$; positive by (2.2); tends to $\log 1/(1 - h) > 0$ as $p \to 0^+$; tends to ∞ as $p \to (1 - h)^-$. And $H(1 - p, 1 - \theta) = H(p, \theta)$.

Fairly sharp bounds on $g(h)$ are given in (3.5) and (3.6), but the proof of the main theorem only needs positivity.

PROOF OF THEOREM (2.4). The posterior odds ratio is

$$R(n,p,h) = \frac{\int_{[p-h,p+h]} e^{-nH(p,\theta)}\mu(d\theta)}{\int_{[0,p-h)\cup(p+h,1]} e^{-nH(p,\theta)}\mu(d\theta)}.$$

By (2.2), the denominator is at most $e^{-n[H(p,p\pm h)]}$. By (2.3), $H(p,p\pm h) \geq H(p,p) + g(h)$. An upper bound on the denominator is therefore $e^{-n[H(p,p)+g(h)]}$.

To complete the proof of Theorem (2.4), the numerator will be bounded from below by $e^{-n[H(p,p)+\varepsilon g(h)]}\psi(h,\varepsilon)$. In outline we choose a small, positive h^* and take the integral in the numerator over the subinterval $[p, p+h^*]$; for θ in this subinterval, by continuity, $H(p,\theta)$ is about $H(p,p)$.

To make this rigorous, let $0 < h < \frac{1}{4}$. Without real loss, suppose $0 \leq p \leq \frac{1}{2}$. Clearly,

$$\frac{\partial}{\partial\theta}H(p,\theta) = -\frac{p}{\theta} + \frac{1-p}{1-\theta}. \tag{3.2}$$

If $p \leq \theta \leq p+h$, the first term on the right in (3.2) is between -1 and 0. The second term is between 0 and $(1-p)/(1-p-h)$. The last expression increases with p to a maximum of $\frac{1}{2}/(\frac{1}{2}-h) < 2$, because $h < \frac{1}{4}$. Thus,

$$\left|\frac{\partial}{\partial\theta}H(p,\theta)\right| < 2, \quad \text{provided } p \leq \theta \leq p+h,\ 0 \leq p \leq \tfrac{1}{2},\ 0 < h < \tfrac{1}{4}. \tag{3.3}$$

Fix $\varepsilon > 0$. Let h^* be the smaller of h and $\frac{1}{2}\varepsilon g(h)$. Let $\psi(h,\varepsilon) = \phi(h^*)$, a positive lower bound on the prior μ-mass in $(p, p+h^*)$. By (3.3), $p \leq \theta \leq p + h^*$ entails $H(p,\theta) < H(p,p) + \varepsilon g(h)$. Since $h^* \leq h$, the numerator is bounded by

$$\begin{aligned}\int_{[p,p+h^*]} e^{-nH(p,\theta)}\mu(d\theta) &\geq e^{-n[H(p,p)+\varepsilon g(h)]}\mu[p,p+h^*] \\ &\geq e^{-n[H(p,p)+\varepsilon g(h)]}\psi(h,\varepsilon).\end{aligned}$$

□

PROOF OF COROLLARY (2.5). The inequality is immediate from (2.4). To see that $g(h)$ is best possible, fix h. For now, fix j and n too. Abbreviate $p = j/n$. We must bound $R(n,p,h)$ from above. As (2.2) shows, the numerator is bounded above by $e^{-nH(p,p)}$. The denominator is bounded below by the integral over $[p+h, p+h+\delta]$. For θ in that interval, $H(p,\theta)$ is at most $H(p + h + \delta)$, by (2.2). So the denominator is at least

$$\mu(p+h, p+h+\delta)\cdot e^{-n[H(p,p+h+\delta)]}.$$

If $p + h + \delta < 1$, then

$$\frac{1}{n}\log R(n,p,h) \leq O\left(\frac{1}{n}\right) + H(p,p+h+\delta) - H(p,p).$$

To complete the argument, let $n \to \infty$; let $p = j/n$ tend to a point where $H(p, p+h) - H(p,p)$ takes its minimum value $g(h)$; and let $\delta \to 0$. □

The function $g(h)$. We now look more closely at the function $g(h)$. Let $h \in (0, \frac{1}{2})$, so $h < 1 - h$. For $0 \le p < 1 - h$, let

$$D_{+h}(p) = H(p, p+h) - H(p,p).$$

For $h < p \le 1$, let

$$D_{-h}(p) = H(p, p-h) - H(p,p).$$

These are the "entropy differentials." Clearly, $D_{-h}(p) = D_{+h}(1-p)$.

(3.4) PROPOSITION. *The function* $p \to D_{+h}(p)$ *is strictly convex; the function* $h \to D_{+h}(p) - 2h^2$ *is strictly increasing and positive.*

PROOF. For the first claim, $x \to x\log[x/(x+\delta)]$ is strictly convex, provided $x > 0$ and $x + \delta > 0$; take $\delta = \pm h$ and $x = p$ or $1 - p$. For the second claim,

$$\frac{\partial}{\partial h} D_{+h}(p) = \frac{h}{(p+h)(1-p-h)} - 4h \ge 0$$

because $x(1-x) \le \frac{1}{4}$. □

REMARK. The convexity of the entropy differential can be used instead of (3.2) to make $H(p,\theta) \approx H(p,p)$ for $p < \theta < p + h^*$ in the numerator of the odds ratio; this alternative proof of (2.4) was suggested by associate editor.

(3.5) COROLLARY. $g(h) - 2h^2$ *is positive and increasing.*

(3.6) PROPOSITION. *Let* $0 < h < h_0 < \frac{1}{2}$. *Then* $g(h) < 2C_0h^2$, *where* $C_0 = -[\log(1-4h_0^2)]/4h_0^2$.

PROOF. Clearly, $g(h) < D_{+h}(\frac{1}{2})$, so $2g(h) < -\log(1-4h^2)$. But $u \to -[\log(1-u)]/u$ is strictly increasing. □

For example, take $h_0 = \frac{1}{10}$. Then $2h^2 < g(h) < 2.05h^2$ for $0 < h < \frac{1}{10}$. As the referee observes, $D_{+h}(p) > 2h^2$ is a special case of the inequality between the Kullback–Leibler number and variation distance:

$$\int |f - g| \le 2\sqrt{\int f \log\frac{f}{g}}\,.$$

4. The theorem for the multinomial. Let S_k be the simplex of all k-vectors θ with nonnegative coordinates θ_i adding to 1. Consider a die with k sides, labeled $1, \dots, k$. In n tosses, the relative frequencies with which these sides land form a vector $p = (p_1, \dots, p_k)$ in S_k.

(4.1) For $0 < h < 1/k$, Let $N_k(h, p)$ be the polyhedral neighborhood of p consisting of the $\theta \in S_k$ with $|\theta_i - p_i| \leq h$ for all i.

Plainly, $N_k(h, p)$ is the sphere around p of radius h—in the sup norm. Usually, this "sphere" is a cube.

To state the main result of this section, let ϕ be a positive function on (0, 1). Suppose a k-sided die is tossed n times. Let p be the vector of empirical frequencies. Let $0 < h < 1/k$. Let $R(n, p, h)$ be the posterior odds ratio for the inside of $N_k(h, p)$ versus the outside, with respect to a ϕ-positive prior: The outside is nonempty, because $h < 1/k$. Let $\varepsilon > 0$. Recall g from (2.3). There is a $\psi(h, \varepsilon) > 0$, which depends on ϕ, h and ε but not on n or p, such that the following inequality holds:

(4.2) THEOREM. $R(n, p, h) \geq \psi(h, \varepsilon) e^{n(1-\varepsilon)g(h)}$ *for* $0 < h < \frac{1}{2}k$.

(4.3) COROLLARY. $\liminf_{n \to \infty} \inf_{p, \mu} (1/n) \log R(n, p, h) \geq g(h)$.

In (4.2) and (4.3) the prior μ is restricted to be ϕ-positve, $0 < h < \frac{1}{2}k$ and $g(h)$ is best possible.

Informally, a prior μ on the simplex S_k is "ϕ-positive" if $\mu(S_{kh}) \geq \phi(h)$, where $S_{kh} \subset S_k$ has the same shape and orientation as S_k, but each edge of S_{kh} is h times the corresponding edge of S_k (in length). More formally, let l be an integer between 1 and k.

(4.4) DEFINITION. Let $T_k(l)$ be the $(k-1)$-dimensional simplex in R^k whose k extreme points $\{e^j: j = 1, \dots, k\}$ are as follows, with e_i^j being the ith coordinate of the vector e^j:

if $j = l$, then $e_i^j = 0$ for all i,

if $j \neq l$, then $e_l^j = -1$, $e_j^j = +1$ and $e_i^j = 0$ for $i \neq j, l$.

Plainly, if $x \in T_k(l)$ then $\sum_{i=1}^k x_i = 0$; furthermore, $-1 \leq x_l \leq 0$; and $0 \leq x_i \leq 1$ for $i \neq l$. For $p \in S_k$ and $0 < h < 1/k$ let

$$T_k(p, h, l) = p + hT_k(l) = \{p + hx: x \in T_k(l)\}. \tag{4.5}$$

To illustrate, let $h = 1$: If $p = (1, 0, 0, \dots, 0)$ and $l = 1$, then $T_k(p, h, l) = S_k$; likewise if $p = (0, 1, 0, \dots, 0)$ and $l = 2$, etc. With this notation, μ is "ϕ-positive" if $\mu\{T_k(p, h, l)\} \geq \phi(h)$ whenever $T_k(p, h, l) \subset S_k$. The definition does not really depend on l. Indeed, let C_l be the class of sets $T_k(p, h, l)$ which are wholly included in S_k, as p ranges over S_k: Then $C_1 = C_2 = \cdots = C_k$.

Let $H_k(p, \theta)$ be the relative entropy:

$$H_k(p, \theta) = -\sum_{i=1}^{k} p_i \log \theta_i. \tag{4.6}$$

This can be defined everywhere by the convention $0 \times \infty = 0$, but the limit of $H_k(p, \theta)$ is not well defined if, e.g., p_1 and θ_1 both tend to 0.

As the next result shows, the minimum entropy differentials do not depend on the dimension k; this reduces the general case to the case $k = 2$.

(4.7) PROPOSITION.

$$\inf_{p \in S_k,\, \theta \notin N_k(h, p)} [H_k(p, \theta) - H_k(p, p)] = g(h).$$

PROOF. Suppose $k \geq 3$. Since the entropy function (4.6) is convex in θ with its minimum at p, the infimum outside the convex polyhedron $N_k(h, p)$ is attained on the boundary. Consider the intersection of the boundary with

$$F = \{\theta\colon \theta \in S_k \text{ and } \theta_k = p_k + h\}.$$

Assume for the sake of argument that this face is nonempty, so $p_k + h \leq 1$. Consider

$$\inf_{\theta \in F} H_k(p, \theta) - H_k(p, p). \tag{4.8}$$

Now

$$\begin{aligned} H_k(p, \theta) &= -\sum_{i=1}^{k} p_i \log \theta_i \\ &= (1 - p_k) H_{k-1}(\tilde{p}, \tilde{\theta}) - p_k \log(p_k + h) - (1 - p_k)\log(1 - p_k - h), \end{aligned}$$

where $\tilde{p}_i = p_i/(1 - p_k)$ and $\tilde{\theta}_i = \theta_i/(1 - \theta_k)$ for $i = 1, \dots, k - 1$. So $\tilde{p}, \tilde{\theta} \in S_{k-1}$. Now $(1 - p_k)H_{k-1}(\tilde{p}, \tilde{\theta})$ is minimized in $\tilde{\theta}$ at $\tilde{\theta} = \tilde{p}$, and the value of the minimum is

$$-\sum_{i=1}^{k-1} p_i \log(p_i/1 - p_k) = -\sum_{i=1}^{k-1} p_i \log p_i + (1 - p_k)\log(1 - p_k).$$

As is easily verified, the location of the proposed minimum for $H_k(p, \cdot\,)$ is on the boundary of $N_k(h, p)$.

The infimum in (4.8) is seen to be

$$\begin{aligned} -p_k \log(p_k + h) - (1 - p_k)\log(1 - p_k - h) + p_k \log p_k \\ + (1 - p_k)\log(1 - p_k) = D_{+h}(p_k), \end{aligned}$$

whose minimum value is $g(h)$. This completes the proof of (4.7). □

PROOF OF THEOREM (4.2) AND COROLLARY (4.3). Suppose by renumbering that $p_1 \leq \cdots \leq p_k$. Let $l = k$. Recall the simplex $T_k(p, h, l)$ from (4.5). This simplex is wholly in the interior of S_k, because $p_k \geq 1/k > h$. It has k

extreme points:

$$\begin{array}{cccccc}
p_1 & p_2 & p_3 & \cdots & p_{k-1} & p_k \\
p_1 + h & p_2 & p_3 & \cdots & p_{k-1} & p_k - h \\
p_1 & p_2 + h & p_3 & \cdots & p_{k-1} & p_k - h \\
\cdot & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
p_1 & p_2 & p_3 & \cdots & p_{k-1} + h & p_k - h
\end{array}$$

And each extreme point is in S_k. The rest of the argument is as for the coin. □

5. Some facts about ϕ-positivity. This section has some remarks and examples on the idea of ϕ-positivity for the binomial case. Recall that ϕ is a positive function on $(0, 1)$; and the prior μ is ϕ-positive iff it assigns mass $\phi(h)$ or more to every closed interval of length h in $[0, 1]$.

(5.1) REMARK. Fix $a > 0$. If $\phi(h) > ah$ for all h, and μ is ϕ-positive, then μ is bounded setwise below by a times Lebesgue measure.

It is natural to conjecture that a ϕ-positive class of measures is bounded below (setwise) by a positive measure, but that turns out to be wrong; ϕ-positivity is a more general idea.

(5.2) EXAMPLE. There is a ϕ-positve class of probability measures $M = \{\mu\}$ on $[0, 1]$ such that if α is a measure and $\alpha \le \mu$ setwise for all $\mu \in M$, then $\alpha = 0$.

CONSTRUCTION. The class M will be countable. Let λ be Lebesgue measure on $[0, 1]$. Let λ_n assign mass $1/(n+1)$ to each of $0/n, 1/n, 2/n,, \ldots, n/n$. Let

$$\mu_n = \frac{n+1}{n+2}\lambda_n + \frac{1}{n+2}\lambda.$$

Let $Q = \{r\}$ be the rationals in $[0, 1]$, and I the irrationals. If $\alpha \le \mu_n$, then $\alpha\{r\} \le 1/(n+2)$ and $\alpha(I) \le 1/(n+2)$, so in the end $\alpha\{r\} = 0$, $\alpha(Q) = 0$, and $\alpha(I) = 0$.

We claim that $\{\mu_n\}$ is ϕ-positive, with $\phi(h) = h^2/4$. To verify this, consider the interval $[x, x+h]$. Suppose $(a-1)/n < x \le a/n$ and $b/n \le x + h < (b+1)/n$. Clearly, $(b-a)/n \ge h - 2/n$; so $b - a \ge nh - 2$. So there are at least $b - a + 1$ rationals of order n in $[x, x+h]$, and

$$\lambda_n[x, x+h] \ge \frac{nh-1}{n+1}.$$

Now

$$\begin{aligned}\mu_n[x, x+h] &\geq \frac{nh-1}{n+2}\\ &\geq \frac{1}{2}h - \frac{1}{n+2}\\ &\geq \frac{1}{4}h \quad \text{if } n+2 \geq \frac{4}{h}\\ &\geq \frac{1}{4}h^2.\end{aligned}$$

If $n + 2 < 4/h$, a lower bound on $\mu_n[x, x+h]$ is still $\frac{1}{4}h^2$, from the λ-term only. In fact, $\phi(h)$ is of order h^2, as one sees by taking n of order $1/h$. □

(5.3) REMARK. There is a connection with monotone rearrangements [Hardy, Littlewood and Pólya (1934)]. Let ϕ be convex, with derivative f, and $\phi(1) = 1$. So f is monotone nondecreasing, and its integral is 1. All rearrangements of f are ϕ-positive. Some rearrangements have bigger (and nonconvex) ϕ's; for such a ϕ, all rearrangements of its density will no longer be ϕ-positive. If $\phi(h) = ah^2$, the rearrangements can be bounded below only by a trivial measure.

(5.4) REMARK. Let M be a ϕ-positive class. Then the closed convex hull of M is ϕ-positve too. (The space of probabilities on $[0, 1]$ is endowed with the weak-star topology, which is compact and metrizable.)

If M consists of one prior, or finitely many priors, then there is a ϕ such that M is ϕ-positive; the next result is a small generalization.

(5.5) REMARK. Let M be a closed, convex class of probabilities on $[0, 1]$. Suppose that each element of M assigns positive mass to every open interval. Then there is a ϕ such that M is ϕ-positive.

PROOF. Fix h with $0 < h < 1$. Let $0 \leq x \leq 1 - h$. Let the continuous function f_x on $[0, 1]$ vanish to the left of x and to the right of $x + h$; let $f_x = 1$ at $x + \frac{1}{2}h$; complete f_x by linear interpolation. Now $\mu(f_x)$ is a continuous positive function of $\mu \in M$ and x; so it has a positive minimum: $\phi(h)$ can be defined as this minimum, over μ and x. □

Let M_ϕ be the class of ϕ-positive μ. When is M_ϕ nonempty? When is ϕ the exact inf, that is, $\phi(h) = \inf\{\mu[x, x+h]: \mu \in M_\phi \text{ and } 0 \leq x < x + h \leq 1\}$?

What are the extreme points of M_ϕ? At this point, we only have some scattered remarks as partial answers.

(5.6) EXAMPLE. Let $\phi(h) = h/10$, for $0 < h < 1$. One compact convex class M of ϕ-positive μ is the set of μ of the form

$$0.1 * \text{Lebesgue} + 0.9 * \nu,$$

where ν is any probability. The extreme points have $\nu = \delta_x$. This class is maximal, by a standard extension argument off intervals. There seem to be two other compact convex ϕ-positve classes M, which are minimal: take $\nu = \delta_0$ or δ_1. To get intermediate classes, mix over any compact set of δ_x's containing $x = 0$ or 1.

(5.7) EXAMPLE. Let $\phi(h) = h/2$ for $h < \frac{2}{3}$ and $\phi(h) = 2h$ for $\frac{2}{3} < h < 1$. The extreme points of the class of ϕ-positive μ seem to be as follows:

$$\tfrac{1}{2}\text{Lebesgue} + \tfrac{1}{2}\delta_a \quad \text{with } \tfrac{1}{3} \le a \le \tfrac{2}{3},$$

$$\tfrac{1}{2}\text{Lesbesgue} + \tfrac{1}{2}\left\{3a\delta_a + \text{density 3 on } \left(\tfrac{2}{3} + a, 1\right)\right\} \quad \text{for } a < \tfrac{1}{3}.$$

(5.8) REMARK. Let $M = \{\mu\}$ be ϕ-positive. Then $\phi(1/n) \le 1/n$, otherwise μ has mass greater than 1. Likewise, if ϕ is the exact inf of M, then $\phi(h) \ge n\theta(h/n)$.

On the other hand, as the next example shows, ϕ can decrease arbitrarily rapidly near 0.

(5.9) EXAMPLE. $a_n = 2^n\phi(1/2^n)$ can decrease arbitrarily rapidly.

CONSTRUCTION.. Let $a_1 < 1/2$, and $a_{n+1} < a_n$. Let μ_n have density equal to a_n on $[0, 1/2^n]$ and equal to b_n on $(1/2^n, 1]$. So b_n can be computed from a_n, and $b_n > 1$. Let $M = \{\mu_n\}$. We claim that M is ϕ-positive for suitable ϕ; and if ϕ is the exact inf, $\phi(1/2^n) = a_n/2^n$. Indeed, if $m \le n$, then

$$\mu_m[0, 1/2^n] = a_m/2^n.$$

On the other hand, if $m > n$,

$$\mu_m[0, 1/2^n] > a_n/2^n.$$

Indeed,

$$\mu_m[0, 1/2^n] > b_m(1/2^n - 1/2^m) > b_m/2^{n+1} > 1/2^{n+1} > a_n/2^n. \qquad \square$$

Acknowledgments. We thank the referee, the Associate Editor and the former Editor (Bill van Zwet) for close readings and very helpful technical suggestions.

REFERENCES

DE BRUIJN, N. G. (1981). *Asymptotic Methods in Analysis*. Dover, New York.

DIACONIS, P. and FREEDMAN, D. (1988). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

DIACONIS, P. and FREEDMAN, D. (1988). On the uniform consistency of Bayes estimates for multinomial probabilities. Technical Report 137, Dept. Statist., Univ. California, Berkeley.

EDWARDS, W., LINDMAN, H. and SAVAGE, L. J. (1963). Bayesian statistical interference for psychological research. *Psychol. Rev.* **70** 193–242.

FREEDMAN, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math. Statist.* **34** 1386–1403.

GHOSH, J. K., SINHA, B. K. and JOSHI, S. N. (1982). Expansions for posterior probability and integrated Bayes risk. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds.) **1** 403–456. Academic, New York.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934). *Inequalities*. Cambridge Univ. Press, Cambridge.

JOHNSON, R. (1967). An asymptotic expansion for posterior distributions. *Ann. Math. Statist.* **38** 1899–1906.

JOHNSON, R. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* **41** 851–864.

LAPLACE, P. S. (1774). Mémoire sur la probabilité des causes par les évènements. Memoires de mathématique et de physique presentés a l'académie royale des sciences, par divers savants, et lûs dans ses assemblées 6. Reprinted in *Laplace's Oeuvres Complètes* **8** 27–65. English translation by S. Stigler (1986) *Statist. Sci.* **1** 359–378.

LE CAM, L. (1953). On some asymptotic properties of the maximum likelihood estimates and related Bayes estimates. *Univ. Calif. Publ. Statist.* **1** 277–330.

LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LE CAM, L. and SCHWARTZ, L. (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* **31** 140–150.

SCHWARTZ, L. (1965). On Bayes' procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26.

WALKER, A. M. (1969). On the asymptotic behavior of posterior distributions. *J. Roy. Statist. Soc. Ser. B* **31** 80–88.

MATHEMATICS DEPARTMENT
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138

STATISTICS DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720