*The Annals of Statistics*
1985, Vol. 13, No. 4, 1445–1464

# BAYESIAN NONPARAMETRIC ESTIMATION OF THE MEDIAN; PART II: ASYMPTOTIC PROPERTIES OF THE ESTIMATES

BY HANI DOSS

*Florida State University*

For data $\theta + \varepsilon_i$, $i = 1, \dots, n$ where $\varepsilon_i$ are i.i.d. $\sim F$ with the median of $F$ equal to 0 but $F$ otherwise unknown, it is desired to estimate $\theta$. In Doss (1985) priors are put on the pair $(F, \theta)$, the marginal posterior distribution of $\theta$ is computed, and the mean of the posterior is taken as the estimate of $\theta$. In the present paper a frequentist point of view is adopted. The consistency properties of the Bayes estimates computed in Doss (1985) are investigated when the prior on $F$ is of the "Dirichlet-type." Any $F$ whose median is 0 is in the support of these priors. It is shown that if the $\varepsilon_i$ are i.i.d. from a discrete distribution, then the Bayes estimates are consistent. However, if the distribution of the $\varepsilon_i$s is continuous, the Bayes estimates can be inconsistent.

**1. Introduction.** For data $X_i = \theta + \varepsilon_i$, $i = 1, \dots, n$ where the $\varepsilon_i$s are i.i.d. $\sim F$ with the median of $F$ equal to 0 but $F$ otherwise unknown, it is desired to estimate $\theta$. In Doss (1985) the prior $\mathscr{D}_\alpha^* \times \nu$ is put on the pair $(F, \theta)$, where $\mathscr{D}_\alpha^*$ is the Dirichlet prior with parameter $\alpha$ conditioned on the event $\{\text{med } F = 0\}$, and $\nu$ is an arbitrary prior on $\theta$. Here, $\alpha$ is any finite nonnull measure on $\mathscr{R}$ such that $\alpha(-\infty, 0) = \alpha(0, \infty)$, and $\alpha$ has no mass at 0. Under the additional condition that $\alpha/(\alpha\{\mathscr{R}\})$ has a continuous density, Theorem 1 of Doss (1985) gives $\nu(d\theta | X)$, the marginal posterior distribution of $\theta$ given $X_1, \dots, X_n$. The Bayes estimate under squared error as loss is $\hat{\theta} = \int \theta \nu(d\theta | X)$. For the general perspective, see the introduction to Doss (1985), which serves as an introduction to this paper as well.

In this paper the consistency properties of the posterior and of the Bayes estimate under squared error loss are analyzed. For a proper understanding of the results, it is necessary to first give a clear notion of what is meant by consistency in the Bayesian context.

Let $\{P_\psi;\ \psi \in \Pi\}$ be a parametric family of distributions, let $\pi$ be a prior on $\psi$, and denote by $\pi(d\psi | X_1, \dots, X_n)$ the posterior distribution of $\psi$ given $X_1, \dots, X_n$.

The posterior $\pi(d\psi | X_1, \dots, X_n)$ is called *consistent at* $\psi_0$ if for $X_1, X_2, \dots$ i.i.d. $\sim P_{\psi_0}$, $\pi(d\psi | X_1, \dots, X_n)$ converges in distribution to the point mass at $\psi_0$, a.s. $[P_{\psi_0}^\infty]$. Let $\mathscr{F}$ be a subfamily of $(P_\psi;\ \psi \in \Pi)$. The posterior is *consistent for the family* $\mathscr{F}$ if it is consistent for all $\psi_0$ such that $P_{\psi_0} \in \mathscr{F}$. These two notions of consistency refer to the posterior, and not to estimators.

Received September 1983; revised February 1985.

[1]Research supported by National Science Foundation Grant MCS 80-24649 and Air Force Office of Scientific Research Grant F49620-82-K-007.

*AMS* 1980 *subject classifications.* Primary 62A15; secondary 62G05.

*Key words and phrases.* Bayes estimator, Dirichlet process prior, posterior distribution, consistency, estimation of the median.

For a loss function $L$, the corresponding Bayes rule $\hat{\psi}$ is *consistent at* $\psi_0$ if $X_1, X_2, \ldots$ are i.i.d. $\sim P_{\psi_0}$ implies that $\hat{\psi}(X_1, \ldots, X_n)$ converges to $\psi_0$ a.s. $[P_{\psi_0}^{\infty}]$. $\hat{\psi}$ is *consistent for the family* $\mathscr{F}$ if $\hat{\psi}$ is consistent at $\psi_0$ for every $\psi_0$ such that $P_{\psi_0} \in \mathscr{F}$. The last two notions of consistency can of course be applied to any estimator $\tilde{\psi}$ of $\psi$.

A desirable property for an estimator $\hat{\psi}$ is obviously consistency for the entire parametric family $\{P_\psi; \psi \in \Pi\}$.

The main results of this paper can be summarized as follows. Suppose that the $\varepsilon_i$s are i.i.d. $\sim F$, where $F$ is a fixed c.d.f. with median 0 (as opposed to a c.d.f. randomly chosen from $\mathscr{D}_\alpha^*$). It is found that roughly speaking, if $F$ is discrete, then the posterior and the Bayes estimate are both consistent. However, it is found that when $F$ is continuous, the posterior and the Bayes estimate $\hat{\theta}$ can be inconsistent. This inconsistent behavior can occur in three ways. The description is easiest in terms of the estimator $\hat{\theta}$.

(a) $\hat{\theta}$ converges to a wrong value.

(b) $\hat{\theta}$ oscillates between two wrong values $a$ and $b$, with $a < \theta < b$: $[F]$ a.e., there exist subsequences $\{n_k\}$ and $\{n_j\}$ such that $\hat{\theta} \to a$ along $\{n_k\}$ and $\hat{\theta} \to b$ along $\{n_j\}$.

(c) $[F]$ a.e., $\{\hat{\theta}_n; n = 1, 2, \ldots\}$ is dense in $\mathscr{R}$: For all $a \in \mathscr{R}$, there exists a subsequence $\{n_k\}$ such that $\theta_{n_k} \to a$.

In all three cases, the posterior $\nu(d\theta | X)$ behaves in an analogous way.

It is shown that any c.d.f. $F$ whose median is 0 is in the support of $\mathscr{D}_\alpha^*$. Thus, $(a)$, $(b)$ and $(c)$ provide examples of inconsistent Bayes rules.

Doob (1949) has proved that under very general conditions, the posterior $\pi(d\psi | X)$ is consistent at $\psi_0$ for $[\pi]$ a.e. $\psi_0$. Doob's result raises the question of what is the $\pi$-null set, and more importantly, when is it empty. LeCam (1953, 1958), Freedman (1963) and Schwartz (1965) have shown that under strong regularity on $\Pi$ and $\pi$, the answer is that

$$\pi(d\psi | X) \quad \text{is consistent at } \psi_0$$

if and only if

$$\psi_0 \in \operatorname{supp}(\pi). \tag{1.1}$$

The assumptions required for the validity of (1.1) sometimes are severe enough to essentially restrict the result to finite dimensional $\Pi$. Indeed, Freedman (1963) presented a counterexample involving priors on the set of distributions on the natural numbers. This counterexample is, however, somewhat contrived. Diaconis and Freedman (1985a, b) considered the "symmetrized Dirichlet" priors introduced by Dalal (1979a, b) as the priors on $F$. They showed that inconsistent behavior can occur in that situation; see also Diaconis and Freedman (1982).

The studies in Diaconis and Freedman (1985a, b) and in the present paper thus provide natural examples of inconsistent Bayes rules.

Section 2 provides preliminaries and a heuristic explanation of how consistent or inconsistent behavior arises. Section 3 gives the results concerning consistency. In that section there is also a description of some consistency results outside of the Bayesian and decision-theoretic framework. Section 4 contains the results concerning inconsistency, and Section 5 a short summary of the results of the paper.

**2. Preliminaries and heuristics.** Consider the space $\mathscr{P}^{**}$ consisting of all c.d.f.s on $\mathscr{R}$ with unique median equal to 0, and let the topology on $\mathscr{P}^{**}$ be the topology of weak convergence. Let $\Pi^* = \mathscr{P}^{**} \times \mathscr{R}$ have the product topology and the $\sigma$ field that this topology induces. The measure on $\Pi^*$ is to be $\mathscr{D}_\alpha^* \times \nu$.

REMARK. If $F$ is a c.d.f. with nonunique median, the median of $F$ can still be defined as the midpoint of the interval of medians. The theory developed here can equally well be worked out for the space of c.d.f.s with possibly nonunique median equal to 0. However, $\mathscr{P}^{**}$ is used instead in order to avoid technical complications.

In Doss (1985) the posterior $\nu(d\theta|X)$ is given under the condition that $\alpha_0$ be absolutely continuous, with a continuous density $\alpha_0'$. This condition is assumed in the sequel. The following assumptions are introduced.

ASSUMPTIONS.

A1. $\operatorname{supp}(\alpha) = \mathscr{R}$.

A2. $\operatorname{supp}(\nu) = \mathscr{R}$.

A3. $\int|\theta|\nu(d\theta) < \infty$.

PROPOSITION. (1) *Under* A1, $(\mathscr{D}_\alpha^* \times \nu)(\Pi^*) = 1$. (*Thus, the prior* $\mathscr{D}_\alpha^* \times \nu$ *can be put on* $\Pi^*$.)

(2) *Under* A1 *and* A2, $\operatorname{supp}(\mathscr{D}_\alpha^* \times \nu) = \Pi^*$.

The proof of the proposition is straightforward and is omitted; see Doss (1983b).

In Doss (1985) the posterior distribution of $\theta$ given $X_1, \dots, X_n$ is found to be

$$\nu(d\theta|X) = c(X)[\Pi^*\alpha_0'(X_i - \theta)]M(X,\theta)\nu(d\theta), \tag{2.1}$$

where

$$[M(X,\theta)]^{-1} = \Gamma\left(\tfrac{1}{2}\alpha(\infty) + nF_n(\theta)\right)\Gamma\left(\tfrac{1}{2}\alpha(\infty) + n(1 - F_n(\theta))\right). \tag{2.2}$$

$F_n$ is the empirical distribution function of $X_1, \dots, X_n$, $c(X)$ is a normalizing constant, and the * indicates that the product is over distinct $X_i$s only.

The Bayes rule under squared error loss is

$$\hat{\theta}(X) = \int_{-\infty}^{\infty} \theta\nu(d\theta|X). \tag{2.3}$$

In order to study the asymptotic behavior of $\nu(d\theta|X)$ and of $\hat{\theta}(X)$, it is necessary to first understand the asymptotic behavior of the factor $M(X,\theta)$. This behavior involves the function $\psi$ defined by

$$\psi(t) = t \log t + (1 - t)\log(1 - t) \quad \text{for} \quad t \in (0,1). \tag{2.4}$$

Note that $\psi$ is symmetric about $\frac{1}{2}$ and has a unique minimum there.

The random variables $X_1, X_2, \dots$ are assumed i.i.d. $\sim F$, where $F$ is any distribution function on $\mathscr{R}$ with a unique median equal to 0. Let $B(n)$ be defined by

$$B(n) = (2\pi)^{-1}\exp[n + (1 - \alpha(\infty) - n)\log n]. \tag{2.5}$$

LEMMA. *For* $[F]$ *a.e.* $\{X_i\}_{i=1}^{\infty}$,

$$M(X, \theta) \sim B(n)\exp[-n\psi(F_n(\theta))][F_n(\theta)(1 - F_n(\theta))]^{(1/2)-[\alpha(\infty)]/2} \tag{2.6}$$

*uniformly for* $\theta$ *in any interval* $[a, b]$ *such that* $0 < F(a)$ *and* $F(b) < 1$.

PROOF. The proof follows computationally from Stirling's formula and is omitted. □

A heuristic explanation of the behavior of the posterior can now be given. A more complete and precise account is given in Sections 3 and 4.

Consider first the parametric model that corresponds to the prior $\delta_{\alpha_0} \times \nu$ on $\Pi^*$. In this model, the posterior may be written as

$$\nu(d\theta|X) = c(X)e^{nl_n(\theta)}\nu(d\theta), \tag{2.7}$$

where $l_n(\theta)$ is given by

$$l_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log \alpha_0'(X_i - \theta). \tag{2.8}$$

Let $\hat{\theta}^{\alpha_0}$ denote the maximum likelihood estimate of $\theta$ in the model where $X_1, \dots, X_n$ are i.i.d. with density $\alpha_0'(x - \theta)$. For example, if $\alpha_0$ is a normal distribution, $\hat{\theta}^{\alpha_0}$ is the mean of the observations. Assuming sufficient regularity, $l_n(\theta)$ may be expanded around $\hat{\theta}^{\alpha_0}$:

$$\begin{aligned}\exp[nl_n(\theta)] = \exp n\Big[&l_n(\hat{\theta}^{\alpha_0}) + (\theta - \hat{\theta}^{\alpha_0})\dot{l}_n(\hat{\theta}^{\alpha_0}) \\ &+ \tfrac{1}{2}(\theta - \hat{\theta}^{\alpha_0})^2\ddot{l}_n(\hat{\theta}^{\alpha_0}) + \text{terms of smaller order}\Big].\end{aligned} \tag{2.9}$$

In (2.9), $\dot{l}_n(\theta)$ and $\ddot{l}_n(\theta)$ denote the first and second derivatives, respectively, of $l_n(\theta)$ with respect to $\theta$. The term

$$\exp[nl_n(\hat{\theta}^{\alpha_0})] \tag{2.10}$$

is independent of $\theta$, and may thus be absorbed into the normalizing constant. Using the fact that $\dot{l}_n(\hat{\theta}^{\alpha_0}) = 0$, and ignoring the smaller order terms, (2.7) may be written as

$$\nu(d\theta|X) = c(X)\exp\left[\tfrac{1}{2}n(\theta - \hat{\theta}^{\alpha_0})^2\ddot{l}_n(\hat{\theta}^{\alpha_0})\right]\nu(d\theta). \tag{2.11}$$

[(2.11) is exact if $\alpha_0$ is a normal distribution.] Thus in the parametric model, under sufficient regularity, the posterior is sharply peaked around the maximum likelihood estimate. If $\hat{\theta}^{\alpha_0}$ is a consistent estimate of a quantity that is *not* the population median, which can happen if the data is generated by a different parametric model, then the posterior (2.11) is inconsistent. The reader should keep in mind the example where $\alpha_0$ is the normal distribution. In this example, the maximum likelihood estimate, and hence the Bayes estimate are consistent estimators of the population mean, which may differ from the population median.

Return now to the posterior (2.1) for the model where the prior $\mathscr{D}_\alpha^* \times \nu$ is put on $\Pi^*$. Assume temporarily that the asymptotic expression for $M(X, \theta)$ given by (2.6) is uniform for $\theta$ ranging over $\mathscr{R}$. Since $B(n)$ is a constant independent of $\theta$, it can be completely ignored. The factor

$$(2.12) \qquad [F_n(\theta)(1 - F_n(\theta))]^{1/2 - [\alpha(\infty)]/2}$$

is asymptotically negligible relative to

$$(2.13) \qquad \exp[-n\psi(F_n(\theta))],$$

and will also be ignored. Thus, heuristically, we can replace $M(X, \theta)$ by (2.13). Replacing the asymptotic equivalence by an equality, $\nu(d\theta | X)$ is written

$$(2.14) \qquad \nu(d\theta|X) = c(X)\exp[-n\psi(F_n(\theta))]\left[\prod_{i=1}^{n}{}^{*} \alpha_0'(X_i - \theta)\right]\nu(d\theta).$$

Since $\psi(t)$ has a unique minimum at $t = \frac{1}{2}$, $\psi(F_n(\theta))$ has a minimum when $F_n(\theta) = \frac{1}{2}$, i.e., when $\theta = \text{med}\{X_1, \dots, X_n\}$. Thus, asymptotically, (2.13) [and $M(X, \theta)$] is sharply peaked at the empirical median. Letting

$$(2.15) \qquad l_n^*(\theta) = \frac{1}{n}\sum_{i=1}^{n}{}^{*} \log \alpha_0'(X_i - \theta),$$

(the * indicates that the sum is over distinct $X_i$s only), (2.14) is rewritten

$$(2.16) \qquad \nu(d\theta|X) = c(X)\exp[-n\psi(F_n(\theta))]\exp[nl_n^*(\theta)]\nu(d\theta).$$

The factor (2.13) does not depend on the prior choice of $\alpha$ and is thus "nonparametric." The factor $\exp[nl_n^*(\theta)]$ may then be called the "parametric component" of the posterior.

If $F$ is discrete, the sum on the right-hand side of (2.15) contains a small number of terms, and it turns out that $[F]$ a.s., $l_n^*(\theta) \to 0$ in $\nu$ probability as $n \to \infty$. Thus, (2.13) dominates. If the median of $F$ is assumed unique, the empirical median is a consistent estimator of it, so that the posterior is consistent.

If $F$ is continuous, (2.14) is simply

$$(2.17) \qquad \nu(d\theta|X) = c(X)\exp[-n\psi(F_n(\theta))]\exp[nl_n(\theta)]\nu(d\theta).$$

As was the case for the parametric model, under regularity, (2.17) is approximately equal to

(2.18)

$$\nu(d\theta|X) = c(X)\exp[-n\psi(F_n(\theta))]\exp\left[\tfrac{1}{2}n(\theta - \hat{\theta}^{\alpha_0})^2\ddot{l}_n(\hat{\theta}^{\alpha_0})\right]\nu(d\theta).$$

Observe that

$$\exp[-n\psi(F_n(\theta))]$$

and

$$\exp\left[\tfrac{1}{2}n(\theta - \hat{\theta}^{\alpha_0})^2\ddot{l}_n(\hat{\theta}^{\alpha_0})\right]$$

are of the same order of magnitude. If $\hat{\theta}^{\alpha_0}$ is a consistent estimate of a functional $T(F)$ that is not the median of $F$, (2.18) converges weakly to a point mass at some point strictly between the median of $F$ and $T(F)$. Thus, the posterior is inconsistent.

The Bayes estimate under squared error as loss is

$$\hat{\theta}(X) = \int \theta\nu(d\theta|X). \tag{2.19}$$

The definition of consistency of the posterior $\nu(d\theta|X)$ given in Section 1 involves convergence of $\nu(d\theta|X)$ to a point mass. The definition of consistency of the estimator (2.19) is the usual one, i.e., $\hat{\theta}(X)$ is consistent if it converges a.s. to the true value of $\theta$. The results obtained concerning consistency and inconsistency of the posterior yield (under Assumption A3) corresponding results for the Bayes estimator. The Bayes estimator will be inconsistent, not because the prior $\nu$ has heavy tails, but because the posterior $\nu(d\theta|X)$ is asymptotically a delta function at an incorrect value.

It is interesting to note that when $F$ is continuous, the data do not "swamp the prior," which is unusual.

## 3. Consistency.

ASSUMPTIONS.

A4. $F$ has a unique median at $\theta_0$.

A5. $F\{\theta_0\} = 0$.

A6. $\alpha_0'$ is bounded above.

A7. For every $\theta \in \mathscr{R}$, $E_F|\log \alpha_0'(X - \theta)| < \infty$.

THEOREM 1. *Assume* A2 *and* A4–A7. *If $F$ is discrete, then the posterior given by* (2.1) *is consistent for all $\theta$.*

REMARK. Assumption A5 is necessary. The theorem is not true without it. A6 is not crucial and can be considerably weakened, but at the cost of a complication of the proof. A7 is roughly the condition that the tails of $F$ not be very much heavier than the tails of $\alpha_0$.

PROOF OF THEOREM 1. The posterior is

$$\nu(d\theta|X) = c(X)M(X,\theta)\exp[nl_n^*(\theta)]\nu(d\theta), \tag{3.1}$$

where $l_n^*(\theta)$ is given by (2.15). Our first goal is to show that the factor $\exp[nl_n^*(\theta)]$ is asymptotically negligible.

Let $F$ be written as $F = \sum_{j=1}^{\infty} p_j \delta_{a_j}$. Let $\theta \in \mathscr{R}$ be fixed, and let $\varepsilon > 0$. By A7, there exists $K$ such that

$$\sum_{j=K+1}^{\infty} p_j |\log \alpha_0'(a_j - \theta)| < \varepsilon. \tag{3.2}$$

We have

$$\begin{aligned} l_n^*(\theta) &= \frac{1}{n} \sum_{j=1}^{K} \log \alpha_0'(a_j - \theta) I(X_i = a_j \text{ for some } i = 1, \dots, n) \\ &\quad + \frac{1}{n} \sum_{i=1}^{n}{}^* \log \alpha_0'(X_i - \theta) I(X_i \notin \{a_1, \dots, a_K\}). \end{aligned} \tag{3.3}$$

The first sum clearly goes to 0 as $n \to \infty$. The absolute value of the second sum is bounded by

$$\frac{1}{n} \sum_{i=1}^{n} |\log \alpha_0'(X_i - \theta) I(X_i \notin \{a_1, \dots, a_K\})|. \tag{3.4}$$

By the strong law of large numbers, for $[F]$ a.e. $\{X_i\}_{i=1}^{\infty}$, (3.4) converges to the left side of (3.2), which is less than $\varepsilon$. Thus, for each fixed $\theta$, $l_n^*(\theta) \to 0$ a.s. $[F]$. In particular,

$$\nu\{\theta \in \mathscr{R};\ l_n^*(\theta) \to 0 \text{ a.s. } [F]\} = 1. \tag{3.5}$$

By Fubini's theorem,

$$F\{\{X_i\}_{i=1}^{\infty};\ l_n^*(\theta) \to 0 \text{ a.e. } [\nu]\} = 1. \tag{3.6}$$

In particular,

$$\text{for } [F] \text{ a.e. sequence } \{X_i\}_{i=1}^{\infty}, \quad l_n^*(\theta) \to 0 \quad \text{in } \nu \text{ probability}. \tag{3.7}$$

Without loss of generality, assume that the median of $F$ is 0. We want to show that $\nu(d\theta|X) \to \delta_0$, and it is sufficient (and necessary) to show that

$$\text{for every } \varepsilon > 0, \quad \nu\{(-\infty, -\varepsilon] \cup [\varepsilon, \infty)|X\} \to 0 \quad \text{as } n \to \infty. \tag{3.8}$$

We will show that

$$\text{for every } \varepsilon > 0, \quad \nu\{[\varepsilon, \infty)|X\} \to 0 \quad \text{as } n \to \infty; \tag{3.9}$$

the corresponding statement for the set $(-\infty, -\varepsilon]$ is proved in an identical way. Let

$$\begin{aligned} N_n &= \int_{[\varepsilon, \infty)} M(X, \theta) \exp[nl_n^*(\theta)] \nu(d\theta), \\ D_n &= \int_{-\infty}^{\infty} M(X, \theta) \exp[nl_n^*(\theta)] \nu(d\theta), \end{aligned} \tag{3.10}$$

so that $\nu([\varepsilon, \infty)|X) = N_n/D_n$. $N_n$ will be bounded above and $D_n$ bounded below.

By the strong law of large numbers,

$$\text{(3.11)} \qquad \text{a.s. } [F], \qquad F_n(\varepsilon) \to F(\varepsilon) > \tfrac{1}{2}.$$

The inequality in (3.11) is due to the fact that 0 is the unique median of $F$. It follows that for large $n$, $\operatorname{med}\{X_1, \dots, X_n\} < \varepsilon$, and so

$$\text{(3.12)} \qquad M(X, \theta) < M(X, \varepsilon) \quad \text{for all } \theta \in [\varepsilon, \infty).$$

Without loss of generality, we may assume that $F(\varepsilon) < 1$. By the lemma of Section 2 and (3.11),

$$\text{(3.13)} \quad M(X, \varepsilon) \sim B(n)\exp[-n\psi(F_n(\varepsilon))][F(\varepsilon)(1 - F(\varepsilon))]^{1/2 - [\alpha(\infty)]/2}.$$

Assumption A6 is that $\alpha_0'$ is bounded, and without loss of generality, the bound may be taken to be 1. Hence

$$\text{(3.14)} \qquad N_n \le C_1 B(n)\exp[-n\psi(F_n(\varepsilon))],$$

where $C_1$ is a constant.

Consider now $D_n$. Let $\tau > 0$. We have

$$\text{(3.15)} \qquad D_n \ge \int_{(-\tau, \tau)} M(X, \theta)\exp[nl_n^*(\theta)]\,\nu(d\theta).$$

For $[F]$ a.e. $\{X_i\}_{i=1}^{\infty}$, for large $n$, $\operatorname{med}\{X_1, \dots, X_n\} \in (-\tau, \tau)$. Hence, for $\theta \in (-\tau, \tau)$ and for large $n$,

$$\text{(3.16)} \qquad M(X, \theta) \ge \min\{M(X, -\tau), M(X, \tau)\}.$$

It is assumed that $M(X, \tau) \le M(X, -\tau)$. This is done largely for notational simplicity and is made without loss of generality. Thus, for large $n$,

$$\text{(3.17)} \qquad D_n \ge C_2 \int_{(-\tau, \tau)} B(n)\exp[-n\psi(F_n(\tau))]\exp[nl_n^*(\theta)]\,\nu(d\theta),$$

where $C_2$ is a constant. Combining this with (3.14) gives

$$\text{(3.18)} \qquad \frac{N_n}{D_n} \le \frac{C_3\exp[-n[\psi(F_n(\varepsilon)) - \psi(F_n(\tau))]]}{\int_{(-\tau, \tau)}\exp[nl_n^*(\theta)]\,\nu(d\theta)},$$

where $C_3$ is a constant.

Since $F$ is assumed to have no atom at 0 by A5, $\tau$ may be chosen so that $\frac{1}{2} < F(\tau) < F(\varepsilon)$. Let $\eta > 0$ be so small that

$$\text{(3.19)} \qquad \tfrac{1}{2} < F(\tau) + \eta < F(\varepsilon) - \eta.$$

By the strong law of large numbers and (3.18), for large $n$,

$$\text{(3.20)} \qquad \frac{N_n}{D_n} \le \frac{C_3\exp[-n[\psi(F(\varepsilon) - \eta) - \psi(F(\tau) + \eta)]]}{\int_{(-\tau, \tau)}\exp[nl_n^*(\theta)]\,\nu(d\theta)}.$$

Let

$$\text{(3.21)} \qquad \delta = \psi(F(\varepsilon) - \eta) - \psi(F(\tau) + \eta).$$

By (3.19), $\delta$ is positive. By (3.7),

$$\text{for } [F] \text{ a.e. } \{X_i\}_{i=1}^{\infty}, \quad \nu\left\{l_n^*(\theta) > \frac{\delta}{2}\right\} \to 0 \quad \text{as } n \to \infty. \tag{3.22}$$

Since $0 \in \operatorname{supp}(\nu)$, there exists a constant $C_4 > 0$ such that for all large $n$,

$$\nu\left\{\theta \in (-\tau, \tau); l_n^*(\theta) < \frac{\delta}{2}\right\} > C_4. \tag{3.23}$$

Hence, for large $n$,

$$\int_{(-\tau, \tau)} \exp[n l_n^*(\theta)] \nu(d\theta) \geq C_4 e^{-n(\delta/2)}. \tag{3.24}$$

(3.20), (3.21), and (3.24) combine to give

$$\text{for } [F] \text{ a.e. } \{X_i\}_{i=1}^{\infty}, \quad \frac{N_n}{D_n} \leq e^{-n(\delta/3)} \quad \text{for large } n. \tag{3.25}$$

This completes the proof of Theorem 1. □

It should be noted that Theorem 1 is a considerable strengthening of the result of Doob mentioned in the Introduction.

COROLLARY 1. *Under the assumptions of Theorem* 1 *and* A3, *the Bayes estimate*

$$\hat{\theta}(X) = \int_{-\infty}^{\infty} \theta \nu(d\theta|X) \tag{2.3}$$

*is strongly consistent.*

PROOF. Let $\varepsilon > 0$. We may write

$$\hat{\theta}(X) = \int_{[-\varepsilon, \varepsilon]} \theta \nu(d\theta|X) + \int_{[-\varepsilon, \varepsilon]^c} \theta \nu(d\theta|X). \tag{3.26}$$

The absolute value of the first integral is less than or equal to $\varepsilon$. From the proof of Theorem 1, it is immediate that for $[F]$ a.e. $\{X_i\}_{i=1}^{\infty}$, there exists $\delta > 0$ such that

$$\nu(d\theta|X) \leq e^{-n\delta} \nu(d\theta) \quad \text{on } [-\varepsilon, \varepsilon]^c. \tag{3.27}$$

Hence, the absolute value of the second integral in (3.26) is less than or equal to

$$e^{-n\delta} \int_{-\infty}^{\infty} |\theta| \nu(d\theta). \tag{3.28}$$

This is enough to prove the corollary. □

Outside the decision-theoretic Bayesian framework, some positive results can be obtained. Rigorous proofs will not be provided, however.

First, the estimator is consistent if $F$ is equal to $\alpha_0$, "the prior guess at $F$." This is true assuming the usual regularity conditions that ensure consistency of the maximum likelihood estimate. A proof follows fairly directly from (2.18).

Second, if $-\log \alpha_0'$ is convex, the estimator is consistent for all symmetric $F$ satisfying A5 and

$$\int -\log \alpha_0'(x-\theta)\,dF(x) < \infty \quad \text{for all } \theta. \tag{3.29}$$

A heuristic argument proceeds as follows. Assume that $F$ is continuous, so that $\nu(d\theta|X)$ is asymptotically given by (2.17). Let

$$l(\theta) = \int \log \alpha_0'(x-\theta)\,dF(x). \tag{3.30}$$

By convexity of $-\log \alpha_0'$, symmetry of $F$, and (3.29), $l(\theta)$ has a unique maximum at $\theta = 0$. By the strong law of large numbers, for fixed $\theta$, $l_n(\theta)$ is close to $l(\theta)$, and $F_n(\theta)$ is close to $F(\theta)$. Assume that this holds uniformly in $\theta$. Then by (2.17), $\nu(d\theta|X)$ resembles a delta function at 0, and the estimator is consistent. A rigorous version of this argument can be obtained from Theorem 2.1 of Freedman and Diaconis (1982). If $F$ has a discrete component, a combination of the proof of Theorem 1 and of the above argument yields the consistency.

Third, if $\alpha_0$ is the double-exponential distribution i.e., $\alpha_0'(X) = \frac{1}{2}e^{-|x|}$, then the estimator is consistent for all continuous $F$ with a unique median. This is because for the double-exponential density, roughly speaking, the maximum likelihood estimate is the sample median. Consistency is then clear from (2.17).

**4. Inconsistency.** In this section it is proved that the posterior given by (2.1) and the corresponding estimator $\hat{\theta}$ given by (2.3) can be inconsistent. This inconsistent behavior can occur in three ways. The description is easiest in terms of the estimator $\hat{\theta}$.

(a) $\hat{\theta}$ converges to a wrong value, a.e. $[F]$. Roughly speaking, this happens when the mle based on the parametric model $X_1, \dots, X_n$ i.i.d. with density $\alpha_0'(x-\theta)$ consistently estimates a functional $T(F)$ that is not equal to the median of $F$. Here $F$ is not symmetric about its median.

(b) $\hat{\theta}$ oscillates between two wrong values $a$ and $b$, with $a < \text{med}(F) < b$: $[F]$ a.e., there exist subsequences $\{n_k\}$ and $\{n_j\}$ such that $\hat{\theta} \to a$ along $\{n_k\}$ and $\hat{\theta} \to b$ along $\{n_j\}$. This happens when $F$ has an atom at its median. $F$ may be taken to have bounded support, and may be taken symmetric about its median.

(c) $[F]$ a.e., $\{\hat{\theta}_n;\ n = 1, 2, \dots\}$ is dense in $\mathscr{R}$; for all $a \in \mathscr{R}$, there exists a subsequence $\{n_k\}$ such that $\hat{\theta}_{n_k} \to a$. This may happen if the tails of $F$ are much bigger than the tails of $\alpha_0$. $F$ may be taken symmetric about its median.

In all three cases, the posterior $\nu(d\theta|X)$ behaves in an analogous way.

Rather than prove general results giving exact conditions under which the estimator is inconsistent, rigorous proofs will be provided only for three special cases. However, the special cases and the proofs give enough of an indication as to what sort of conditions yield general results.

*(a) Inconsistency: $\hat{\theta}$ converges to a wrong value.*

THEOREM 2. *Let $\alpha_0$ be the standard normal distribution, and assume* A2. *Let $F$ be a distribution with density $f$ and a unique median $m$. Assume that $F$ has* *a finite mean* $\mu$ *and that* $\mu \neq m$. *Let*

$$h(\theta) = \tfrac{1}{2}(\theta - \mu)^2 + \psi(F(\theta)) \quad \textit{for } \theta \in \mathscr{R}, \tag{4.1}$$

*and let*

$$S = \{\theta;\ h \textit{ assumes its minimum at } \theta\}. \tag{4.2}$$

*Then, S lies between m and* $\mu$, *and is bounded away from m. The asymptotic support of the posterior* $\nu(d\theta|X)$ *is S in the sense that for any open set O containing S,* $[F]$ *a.e.,* $\nu(O|X) \to 1$. *Thus, the posterior is inconsistent at F.*

PROOF. Assume without loss of generality that $m < \mu$. It is clear that $h$ is strictly decreasing on $(-\infty, m]$ and strictly increasing on $[\mu, \infty)$. Also, since $h$ is continuous, $\inf\{h(\theta);\ \theta \in [m, \mu]\}$ is achieved on $[m, \mu]$. Thus, $S$ is not empty, and is contained in $[m, \mu]$. Let us now see that $S$ is bounded away from $m$. A computation gives that $h'(m) = m - \mu$, which is negative. This fact, together with the definition of the derivative implies that there exists an $\varepsilon > 0$ such that for all $\theta \in (m, m + \varepsilon)$, $h(\theta) < h(m)$. Thus $S$ is bounded away from $m$.

Let $O$ be any open set containing $S$. To show that $[F]$ a.e., $\nu(O^c|X) \to 0$, we will show that

$$[F] \text{ a.e.,} \quad \nu(O^c \cap (-\infty, m)|X) \to 0, \tag{4.3}$$

$$[F] \text{ a.e.,} \quad \nu(O^c \cap [m, \mu + 1]|X) \to 0, \tag{4.4}$$

and

$$[F] \text{ a.e.,} \quad \nu(O^c \cap (\mu + 1, \infty)|X) \to 0. \tag{4.5}$$

Let $N_{1n}$, $N_{2n}$, $N_{3n}$, and $D_n$ be defined by

$$N_{1n} = \int_{O^c \cap (-\infty, m)} \exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta)\nu(d\theta), \tag{4.6}$$

$$N_{2n} = \int_{O^c \cap [m, \mu+1]} \exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta)\nu(d\theta), \tag{4.7}$$

$$N_{3n} = \int_{O^c \cap (\mu+1, \infty)} \exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta)\nu(d\theta), \tag{4.8}$$

and

$$D_n = \int_{-\infty}^{\infty} \exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta)\nu(d\theta). \tag{4.9}$$

Then, to show (4.3), (4.4), and (4.5), we need to show that $[F]$ a.e., $N_{in}/D_n \to 0$ for $i = 1, 2, 3$. As in the proof of Theorem 1, $D_n$ will be bounded from below and the $N_{in}$ will be bounded above.

Consider first $D_n$. Let $s \in S$. For $\tau > 0$,

$$D_n \geq \int_{s-\tau}^{s+\tau} \exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta)\nu(d\theta). \tag{4.10}$$

Take $\tau$ to be small enough so that $F(s - \tau)$ and $F(s + \tau)$ are both strictly between 0 and 1. Then, by the lemma, $[F]$ a.e., for large $n$, the integrand in (4.10) is greater than

$$K \exp\left[-n(\theta - \bar{X})^2/2\right] B(n) \exp[-n\psi(F_n(\theta))]\,\nu(d\theta) \tag{4.11}$$

for $\theta \in (s - \tau, s + \tau)$. In (4.11), $K$ is a constant. By the strong law of large numbers applied to $\bar{X}$, and by the Glivenko–Cantelli theorem, it follows that

$$D_n \geq e^{o(n)} \int_{s-\tau}^{s+\tau} B(n) e^{-nh(\theta)} \nu(d\theta). \tag{4.12}$$

Consider now $N_{1n}$. For $\theta \in (-\infty, m)$ and all large $n$,

$$\exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta) \leq \exp\left[-n(m - \bar{X})^2/2\right] M(X, M_n), \tag{4.13}$$

where $M_n = \text{med}\{X_1, \dots, X_n\}$. By the lemma, the right side of (4.13) is less than

$$2\left(\tfrac{1}{4}\right)^{1/2 - [\alpha(\infty)]/2} B(n) \exp\left[-n(m - \bar{X})^2/2\right] \exp[-n\psi(F_n(M_n))]. \tag{4.14}$$

By the strong law of large numbers, this is equal to

$$B(n) \exp[-n(h(m) + o(1))]. \tag{4.15}$$

Thus,

$$N_{1n} \leq e^{o(n)} B(n) e^{-nh(m)}. \tag{4.16}$$

Using the fact that $h(m) > h(s)$, together with Assumption A2, $\tau$ in (4.12) may be taken small enough to show that there exists $\delta > 0$ such that $N_{1n}/D_n \leq e^{-n\delta}$ for large $n$.

Consider now $N_{2n}$. By the strong law of large numbers, the lemma, and the Glivenko–Cantelli theorem,

$$\exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta) = \exp[-n(h(\theta) + o(1))] B(n) \tag{4.17}$$

uniformly for $\theta \in [m, \mu + 1]$. [Without loss of generality, we may assume that $F(\mu + 1) < 1$.] Observe that

$$\min\{h(\theta);\ \theta \in O^c \cap [m, \mu + 1]\} > h(s). \tag{4.18}$$

We can combine (4.18), (4.17), and (4.12) to show that there exists $\delta > 0$ such that $N_{2n}/D_n \leq e^{-n\delta}$ for large $n$.

Finally, consider $N_{3n}$. For $\theta \in (\mu + 1, \infty)$, we have

(4.19)

$$\exp\left[-n(\theta - \bar{X})^2/2\right] M(X, \theta) \leq \exp\left[-n(\mu + 1 - \bar{X})^2/2\right] M(X, \mu + 1).$$

By the lemma and the strong law of large numbers, the right side of (4.19) is equal to

$$B(n) \exp[-n(h(\mu + 1) + o(1))]. \tag{4.20}$$

Thus,

$$N_{3n} \leq B(n) \exp[-n(h(\mu + 1) + o(1))]. \tag{4.21}$$

Using the fact that $h(\mu + 1) > h(s)$, (4.21), (4.12), and Assumption A2 may be combined to show that there exists $\delta > 0$ such that $N_{3n}/D_n \leq e^{-n\delta}$ for large $n$. This completes the proof of Theorem 2. □

Note that if $S$ consists of one point $s \in (m, \mu]$, the posterior converges to the point mass at $s$. Examples where $S$ consists of only one point abound, and in fact, it is difficult to construct a distribution $F$ such that the corresponding $S$ contains more than one point.

COROLLARY 2. *Under the assumptions of Theorem* 2 *and under* A3, *the estimator* $\hat{\theta}$ *given by* (2.3) *is inconsistent*:

$$P\{\hat{\theta} \to m\} = 0.$$

*If the set S given by* (4.2) *consists of a unique point s, then* $\hat{\theta}$ *converges to* $s$ $[F]$ *a.e.*

PROOF. The proof is similar to the proof of Corollary 1, and is omitted. □

Recall that $\hat{\theta}^{\alpha_0}$ denotes the maximum likelihood estimator of $\theta$ under the model $X_1, \ldots, X_n$ i.i.d. with density $\alpha_0'(x - \theta)$. If $\alpha_0$ is a normal distribution, $\hat{\theta}^{\alpha_0}$ is $\bar{X}$, the mean of the observations, which is a consistent estimate of the population mean when the latter is finite.

Suppose $\alpha_0$ is another distribution that is symmetric about 0. The log likelihood of $\theta$ based on a sample of size $n$ is

$$nl(\theta) = \sum_{i=1}^{n} \log \alpha_0'(X_i - \theta). \tag{4.22}$$

The following discussion is informal. Suppose that $X_1, X_2, \ldots$ are i.i.d. $\sim F$, and let

$$k(\theta) = E_F \log \alpha_0'(X_1 - \theta). \tag{4.23}$$

Assume that $k$ has a unique minimum, which is achieved at $T(F)$. Under suitable regularity (see for example Huber, 1967), $\hat{\theta}^{\alpha_0}$ converges to $T(F)$, a.s. $[F]$. The results for the posterior $\nu(d\theta|X)$ given by (2.1) and the Bayes estimate $\hat{\theta}$ given by (2.3) are analogous to those given by Theorem 2, where $\alpha_0$ is the normal distribution: If $T(F) \neq m(F)$ [$m(F)$ is the median of $F$], then $\nu(d\theta|X)$ is asymptotically supported by a set $S$ lying between $m$ and $T(F)$, and bounded away from $m$. Both $\nu(d\theta|X)$ and $\hat{\theta}$ are inconsistent.

EXAMPLES. Let

$$\alpha_0'(x) = c\exp(-|x|^a) \qquad a > 1.$$

The convexity of $-\log \alpha_0'$ insures that $\hat{\theta}^{\alpha_0}$ is always unique. It is easy to show that there exist distributions $F$ with $m(F)$ unique, and such that $T(F)$ is unique and not equal to $m(F)$.

*(b) Inconsistency: $\hat{\theta}$ oscillates between two wrong values.* In the following theorem, the measure $\alpha$ that parameterizes $\mathscr{D}_\alpha^*$ plays no essential role.

THEOREM 3. *Assume that $\nu$ is continuous and satisfies* A2, *and assume that $\alpha_0$ has an everywhere positive density that is bounded above. Let $X_1, X_2, \ldots$ be i.i.d. $\sim F$, where $F$ is given by*

$$F = \tfrac{1}{3}(\delta_{-1} + \delta_0 + \delta_1). \tag{4.24}$$

*Let $\nu_-$ and $\nu_+$ be the two probability measures given by*

$$\begin{aligned} \nu_-(d\theta) &= c_-(X)\alpha_0'(1+\theta)\alpha_0'(\theta)\alpha_0'(1-\theta)I(\theta \in [-1,0])\nu(d\theta), \\ \nu_+(d\theta) &= c_+(X)\alpha_0'(1+\theta)\alpha_0'(\theta)\alpha_0'(1-\theta)I(\theta \in [0,1])\nu(d\theta). \end{aligned} \tag{4.25}$$

*Then,*

(1) *For $[F]$ a.e. $\{X_i\}_{i=1}^{\infty}$, there exist subsequences $\{n_k\}$ and $\{n_j\}$, such that $\nu(d\theta|X)$ converges in absolute deviation norm to $\nu_+$ along $\{n_k\}$ and to $\nu_-$ along $\{n_j\}$.*

(2) *For every $0 < \varepsilon < \frac{1}{2}$,*

$$\begin{aligned} &P\Big\{\sup_{A \in \mathscr{B}} |\nu(A|X) - \nu_-(A)| < \varepsilon\Big\} \to \tfrac{1}{2}, \\ &P\Big\{\sup_{A \in \mathscr{B}} |\nu(A|X) - \nu_+(A)| < \varepsilon\Big\} \to \tfrac{1}{2} \end{aligned} \tag{4.26}$$

*as $n \to \infty$.*

PROOF. We will show that $[F]$ a.e., there exists a subsequence $\{n_k\}$ such that $\nu(d\theta|X)$ converges to $\nu_+$ in absolute deviation norm along $\{n_k\}$. The corresponding statement concerning $\nu_-$ is proved in the same way.

Let the measure $\bar{\nu}$ be defined by

$$\bar{\nu}(d\theta) = \alpha_0'(1+\theta)\alpha_0'(\theta)\alpha_0'(1-\theta)\nu(d\theta). \tag{4.27}$$

Then $[F]$ a.e., for large $n$,

$$\nu(d\theta|X) = c(X)M(X,\theta)\bar{\nu}(d\theta). \tag{4.28}$$

Recall that $M(X,\theta)$ is constant between observations. A simple calculation shows that for $\theta_1 \in [-1,1]^c$, $\theta_2 \in (-1,1)$, and any $\varepsilon > 0$, for large $n$,

$$\frac{M(X,\theta_1)}{M(X,\theta_2)} \le \exp\big[-n\big(\psi\big(\tfrac{1}{3}\big) - \varepsilon\big)\big]. \tag{4.29}$$

It is then easy to see that $[F]$ a.e.,

$$\nu([-1,1]^c|X) \le \exp\big[-n\big(\psi\big(\tfrac{1}{3}\big) - \varepsilon\big)\big] \tag{4.30}$$

for large $n$.

By (4.30),

$$\nu([-1,0]|X) \sim \frac{\int_{[-1,0]} M(X,\theta)\bar{\nu}(d\theta)}{\int_{[-1,1]} M(X,\theta)\bar{\nu}(d\theta)}. \tag{4.31}$$

By (4.31) and the lemma,

$$\nu([-1,0]|X) \sim \frac{\int_{[-1,0]}\exp[-n\psi(F_n(\theta))]\bar{\nu}(d\theta)}{\int_{[-1,1]}\exp[-n\psi(F_n(\theta))]\bar{\nu}(d\theta)}. \tag{4.32}$$

Consider now (4.32). If we could replace $F_n(\theta)$ by $F(\theta)$, we would have that $\psi(F_n(\theta)) = \psi(\frac{1}{3})$ for all $\theta$ such that $0 < |\theta| < 1$. The result would be that $\nu([-1,0]|X) \to \bar{\nu}([-1,0])$. However, $F_n(\theta)$ fluctuates about $F(\theta)$ enough to make a big difference.

By the law of the iterated logarithm for multinomial random vectors (see Lemma 3 of Finkelstein, 1971) $[F]$ a.e., the set of limit points of the sequence of random vectors

$$\sqrt{\frac{n}{2\log\log n}}\,\left(F_n\{-1\} - \tfrac{1}{3}, F_n\{0\} - \tfrac{1}{3}, F_n\{1\} - \tfrac{1}{3}\right) \tag{4.33}$$

is

$$S = \left\{(x_1, x_2, x_3);\ \sum_{i=1}^{3} x_i = 0 \text{ and } \sum_{i=1}^{3} x_i^2 \le \tfrac{1}{3}\right\}. \tag{4.34}$$

The point $(-1/\sqrt{6}, 0, 1/\sqrt{6})$ is an element of $S$. Let $\{n_k\}$ be a subsequence along which (4.33) converges to $(-1/\sqrt{6}, 0, 1/\sqrt{6})$. Then, for all large $k$,

$$\begin{aligned} \sqrt{\frac{n}{2\log\log n}}\left(F_n(\theta_1) - \frac{1}{3}\right) &\le \frac{-1}{\sqrt{12}}, \\ \sqrt{\frac{n}{2\log\log n}}\left(F_n(\theta_2) - \frac{2}{3}\right) &\le -\frac{1}{\sqrt{12}} \end{aligned} \tag{4.35}$$

for all $\theta_1 \in (-1,0)$ and all $\theta_2 \in (0,1)$. (The subscript $k$ has been suppressed.) By (4.32) and (4.35),

$$\nu([-1,0]|X) \le \frac{\int_{[-1,0]}\exp\left\{-n\psi\left(\frac{1}{3} - \sqrt{\frac{\log\log n}{6n}}\right)\right\}\bar{\nu}(d\theta)}{\int_{[0,1]}\exp\left\{-n\psi\left(\frac{2}{3} - \sqrt{\frac{\log\log n}{6n}}\right)\right\}\bar{\nu}(d\theta)}. \tag{4.36}$$

By Taylor's theorem, the right side of (4.36) is less than

$$\frac{\int_{[-1,0]}\exp\left\{-n\psi\left(\frac{1}{3}\right) + \psi'\left(\frac{1}{3}\right)\sqrt{\frac{\log\log n}{12n}}\right\}\bar{\nu}(d\theta)}{\int_{[0,1]}\exp\left\{-n\psi\left(\frac{2}{3}\right) + \psi'\left(\frac{2}{3}\right)\sqrt{\frac{\log\log n}{12n}}\right\}\bar{\nu}(d\theta)}. \tag{4.37}$$

Since $\psi(\frac{1}{3}) = \psi(\frac{2}{3})$, $\psi'(\frac{1}{3}) < 0$, and $\psi'(\frac{2}{3}) > 0$, (4.37) goes to 0 as $k \to \infty$. This, together with (4.30) yields the first part of the theorem.

Let us now prove (4.26). By (4.32), for large $n$, with probability arbitrarily close to 1,

$$\nu([-1,0]|X) \sim \bar{\nu}([-1,0])\big[\bar{\nu}([-1,0]) + \bar{\nu}([0,1]) \cdot \exp n\{\psi(F_n(-1)) - \psi(F_n(0))\}\big]^{-1}. \tag{4.38}$$

By Taylor's theorem and a little algebra, the right side of (4.38) is equal to

$$\bar{\nu}([-1,0])\big[\bar{\nu}([-1,0]) + \bar{\nu}([0,1]) \cdot \exp\big(\sqrt{n}\,\psi'(\tfrac{1}{3})\{\sqrt{n}\,(F_n(-1) - \tfrac{1}{3}) + \sqrt{n}\,(F_n(0) - \tfrac{2}{3})\} + O_p(1)\big)\big]^{-1} \tag{4.39}$$

The random variables

$$\sqrt{n}\,(F_n(-1) - \tfrac{1}{3}) + \sqrt{n}\,(F_n(0) - \tfrac{2}{3}) \tag{4.40}$$

are asymptotically normally distributed with mean 0. Let the events $I_n$ be defined by

$$I_n = \big\{\sqrt{n}\,(F_n(-1) - \tfrac{1}{3}) + \sqrt{n}\,(F_n(0) - \tfrac{2}{3}) < -n^{-1/4}\big\}. \tag{4.41}$$

We have $P(I_n) \to \frac{1}{2}$ as $n \to \infty$. For the purpose of this discussion, the $O_p(1)$ term in (4.39) can be ignored. Thus, by (4.39), on $I_n$, $\nu([-1,0]|X) \to 0$ as $n \to \infty$. Hence,

$$\liminf_{n\to\infty} P\Big\{\sup_{A\in\mathscr{B}} |\nu(A|X) - \nu_+(A)| < \varepsilon\Big\} \geq \tfrac{1}{2} \quad \text{for every } \varepsilon > 0. \tag{4.42}$$

Similarly,

$$\liminf_{n\to\infty} P\Big\{\sup_{A\in\mathscr{B}} |\nu(A|X) - \nu_-(A)| < \varepsilon\Big\} \geq \tfrac{1}{2} \quad \text{for every } \varepsilon > 0. \tag{4.43}$$

By considering the sets $[0,1]$ and $[-1,0]$, we note that

$$\text{if} \quad \sup_{A\in\mathscr{B}} |\nu(A|X) - \nu_+(A)| < \varepsilon, \quad \text{then} \quad \sup_{A\in\mathscr{B}} |\nu(A|X) - \nu_-(A)| > 1 - \varepsilon.$$

This observation, together with (4.42), (4.43), and the assumption that $\varepsilon < \frac{1}{2}$, now gives (4.26). □

COROLLARY 3. *Assume* A3 *and the conditions of Theorem* 3. *Let* $\theta_-$ *and* $\theta_+$ *be given by*

$$\theta_- = \int_{[-1,0]} \theta\,\nu_-(d\theta), \qquad \theta_+ = \int_{[0,1]} \theta\,\nu_+(d\theta). \tag{4.44}$$

*Then,* $\theta_- < 0 < \theta_+$, *and*

(1) *For* $[F]$ *a.e.* $\{X_i\}_{i=1}^{\infty}$, *there exist subsequences* $\{n_k\}$ *and* $\{n_j\}$ *such that the estimate* $\hat{\theta}$ *converges to* $\theta_+$ *along* $\{n_k\}$ *and to* $\theta_-$ *along* $\{n_j\}$.

(2) *For every* $0 < \varepsilon < \frac{1}{2}$, *as* $n \to \infty$

$$P\{|\hat{\theta} - \theta_-| < \varepsilon\} \to \tfrac{1}{2},$$
$$P\{|\hat{\theta} - \theta_+| < \varepsilon\} \to \tfrac{1}{2}. \tag{4.45}$$

PROOF. The proof is similar to the proof of Corollary 1, and is omitted. □

*(c) Inconsistency: $\hat{\theta}$ is dense in $\mathscr{R}$.* Let $C_\sigma$ denote the Cauchy distribution with median 0 and scale parameter $\sigma$:

$$C_\sigma(x) = \int_{-\infty}^{x} \frac{\sigma}{\pi(\sigma^2 + t^2)}\, dt. \tag{4.46}$$

THEOREM 4. *Let $\alpha_0$ be the standard normal distribution and assume* A2. *For all $\sigma$ sufficiently large, if $X_1, X_2, \dots$ are i.i.d. $\sim C_\sigma$, then $[C_\sigma]$ a.e. $\{X_i\}_{i=1}^\infty$, for every $a \in \mathscr{R}$, there exists a subsequence $\{n_k\}$ such that along $\{n_k\}$ $\nu(d\theta|X)$ converges in distribution to the point mass at $a$.*

PROOF. Let $S_n = X_1 + X_2 + \cdots + X_n$. $[C_\sigma]$ a.e., the sequence $\{S_n/n;\ n = 1, 2, \dots\}$ is dense in $\mathscr{R}$. Intuitively, this is because the random variables $S_1, S_2/2, S_3/3, \dots$ all have the distribution $[C_\sigma]$. If they were in addition independent, the result would follow immediately. They are of course not independent. However, for sufficiently large $n_2$, $S_1$ and $S_{n_2}/n_2$ are "nearly independent"; for sufficiently large $n_3$, $S_1, S_{n_2}/n_2, S_{n_3}/n_3$ are "nearly independent", etc. Thus, a subsequence $\{n_k\}$ can be chosen so that $S_1, S_{n_2}/n_2, S_{n_3}/n_3$,. are nearly i.i.d. $\sim [C_\sigma]$. Consequently, $\{S_{n_k}/n_k;\ k = 1, 2, \dots\}$ is dense in $\mathscr{R}$. The above argument can be made rigorous. A less transparent but quicker proof may be obtained from the Hewitt–Savage 0–1 law.

Let $\mu \in \mathscr{R}$, and let $\{n_k\}$ be a subsequence such that $S_{n_k}/n_k \to \mu$. Attention is restricted to the subsequence $\{n_k\}$, and the subscript $k$ is henceforth suppressed. As was seen in the proof of Theorem 2, the asymptotic support of $\nu(d\theta|X)$ is the set $S$ defined through (4.1) and (4.2). Our goal is to show that

(a) $S$ consists of a unique point, call it $S(\mu)$, and

(b) by varying $\mu$, $S(\mu)$ can be made to equal any preassigned value $a$.

For $h$ defined by (4.1), we have

$$h'(\theta) = C_\sigma'(\theta)\log\left[\frac{C_\sigma(\theta)}{1 - C_\sigma(\theta)}\right] + (\theta - \mu) \tag{4.47}$$

and

$$h''(\theta) = 1 + \frac{[C_\sigma'(\theta)]^2}{C_\sigma(\theta)[1 - C_\sigma(\theta)]} + C_\sigma''(\theta)\log\left[\frac{C_\sigma(\theta)}{1 - C_\sigma(\theta)}\right]. \tag{4.48}$$

Consider the third term on the right side of (4.48). We have

$$C_\sigma''(\theta)\log\left[\frac{C_\sigma(\theta)}{1-C_\sigma(\theta)}\right]=\frac{1}{\sigma^2}C_1''\left(\frac{\theta}{\sigma}\right)\log\left[\frac{C_1\left(\frac{\theta}{\sigma}\right)}{1-C_1\left(\frac{\theta}{\sigma}\right)}\right]. \tag{4.49}$$

It is easy to see that

$$\sup_{t\in\mathscr{R}}\left|C_1''(t)\log\left[\frac{C_1(t)}{1-C_1(t)}\right]\right|<\infty. \tag{4.50}$$

Note that the second term in (4.48) is always positive. This observation, together with (4.49) and (4.50) shows that $\sigma$ can be chosen so that $h''(\theta)$ is always positive.

Assume $\mu > 0$. The set $S$ is contained in $(0, \mu]$. Since $h'(0) < 0 < h'(\mu)$, the equation $h'(\theta) = 0$ has a root in $(0, \mu)$, and since $h''(0)$ is always positive, this root is unique. Call it $S(\mu)$. $h$ has a unique minimum at $S(\mu)$, and as was seen in the proof of Theorem 2, $\nu(d\theta|X)$ converges in distribution to the point mass at $S(\mu)$. Let $a \in \mathscr{R}$ be given, and assume that $a > 0$. Let $\mu$ be given by

$$\mu = C_\sigma'(a)\log\left[\frac{C_\sigma(a)}{1-C_\sigma(a)}\right]+a. \tag{4.51}$$

Then, $a$ solves $h'(\theta) = 0$, and is the unique solution. This completes the proof of Theorem 4. □

COROLLARY 4. *Assume A3 and te conditions of Theorem 4. Then,* $[C_\sigma]$ *a.e.* $\{X_i\}_{i=1}^\infty$, *for every* $a \in \mathscr{R}$, *there exists a subsequence* $\{n_k\}$ *such that along* $\{n_k\}$, $\hat\theta \to a$.

PROOF. The proof is similar to the proof of Corollary 1, and is omitted. □

**5. Summary.** For the problem of estimating the location of a distribution function the shape of which is only partially known, the approach used by Dalal, Diaconis and Freedman and in the present work was to put a prior on the unknown c.d.f. Dalal and Diaconis and Freedman considered the "symmetrized Dirichlet priors," which give probability one to the set of symmetric c.d.f.s. Diaconis and Freedman (1985a, b) showed that the estimates obtained from these priors can be inconsistent, while Doss (1984) showed that they are extremely sensitive to the symmetry assumption.

In Doss (1985) the priors $\mathscr{D}_\alpha^*$ were considered for the more general problem of estimating the median of a distribution. If $\alpha_0$ was chosen symmetric, the prior $\mathscr{D}_\alpha^*$ was "centered" at $\alpha_0$, but put all its mass on asymmetric distributions. For these priors, the results are mixed. Only the basic question of consistency has been studied in detail.

On the positive side, the estimator is consistent if $F$ is equal to the prior guess $\alpha_0$, and remains consistent if $F$ deviates from $\alpha_0$ as long as $F$ is discrete. For almost any choice of $\alpha_0$, the mle $\hat{\theta}^{\alpha_0}$ does not have this property. Also, as was indicated by the comments at the end of Section 3, there are certain choices of $\alpha_0$ which will yield an estimator that is consistent for a subset of $\mathscr{P}^{**}$ that is large enough to include, for example, the continuous distributions.

On the negative side, Theorem 3 states the following. There exists a set $E$ of distributions, $E \subset \mathscr{P}^{**}$, such that for any $\alpha_0$ (including the double-exponential distribution), the estimator based on $\mathscr{D}_\alpha^*$ is inconsistent for all $F \in E$. Theorem 2 states that if $\alpha_0$ is a normal distribution, there is a set $E \subset \mathscr{P}^{**}$ such that $E$ is dense in $\mathscr{P}^{**}$, and the estimator based on $\mathscr{D}_\alpha^*$ is inconsistent for all $F \in E$. Thus, the estimator is not even consistent in a neighborhood of $\alpha_0$.

The results obtained by Diaconis and Freedman and in the present work give some information on the behavior of estimators of location obtained by putting a prior on an unknown c.d.f. However, the following problem still remains unsolved. Find a class $\mathscr{C}$ of priors such that

(i) For any symmetric c.d.f. $\alpha_0$ there is a member of $\mathscr{C}$ that is in some suitable sense "centered" at $\alpha_0$.

(ii) The class yields estimators $\hat{\theta}_{\alpha_0}$ of the median which are tractable.

(iii) The estimator $\hat{\theta}_{\alpha_0}$ is

(a) efficient at $\alpha_0$,

(b) robust under small, possibly asymmetric perturbations of $\alpha_0$,

(c) still a consistent estimator of the median if the true distribution is distant from $\alpha_0$.

**Acknowledgments.** I would like to thank Professor Persi Diaconis for his valuable guidance. I am also grateful to Fred Huffer, Satish Iyengar, and Tom Sellke for some interesting discussions, and to a referee for his careful reading of the paper.

## REFERENCES

DALAL, S. R. (1979a). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Process. Appl.* **9** 99–107.

DALAL, S. R. (1979b). Nonparametric and robust Bayes estimation of location. *Optimizing Methods in Statistics* (J. S. Rustagi, ed.) 141–166. Academic Press, New York.

DIACONIS, P. and FREEDMAN, D. (1982). Bayes rules for location problems. *Proceedings of the Third Purdue Symposium on Statistical Decision Theory and Related Topics* (S. S. Gupta and J. Berger (eds.). Academic Press, New York.

DIACONIS, P. and FREEDMAN, D. (1986a). On the consistency of Bayes estimates. To appear in *Ann. Statist.*

DIACONIS, P. and FREEDMAN, D. (1986b). On inconsistent Bayes estimates of location. To appear in *Ann. Statist.*

DOOB, J. L. (1949). Application of the theory of martingales. Colloques Internationaux du CNRS, 22–28, Paris.

DOSS, H. (1983a). Bayesian nonparametric estimation of location. Ph.D. thesis, Stanford University.

DOSS, H. (1983b). Bayesian nonparametric estimation of the median; Part II: Asymptotic properties of the estimates. Florida State University Technical Report No. M657.

DOSS, H. (1984). Bayesian estimation in the symmetric location problem. *Z. Wahrsch. verw. Gebiete* **68** 127–147.

DOSS, H. (1985). Bayesian nonparametric estimation of the median; Part I: Computation of the estimates. *Ann. Statist.* **13** 1432–1444.

FINKELSTEIN, H. (1971). The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.* **42** 607–615.

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

FREEDMAN, D. and DIACONIS, P. (1982). On inconsistent $M$-estimators. *Ann. Statist.* **10**, 454–461.

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

LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *University of California Publications Statistics* **1** 277–330 (1955).

LECAM, L. (1958). Les propriétés asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris* **7** 17–35.

SCHWARTZ, L. (1965). On Bayes procedures. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 4*, 10–16.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32306