The Annals of Statistics
1990, Vol. 18, No. 1, 443–453

CONSISTENT AND ROBUST BAYES PROCEDURES FOR LOCATION BASED ON PARTIAL INFORMATION

BY KJELL A. DOKSUM[1] AND ALBERT Y. LO[2]

University of California, Berkeley and SUNY, Buffalo

We consider Bayes procedures for a location parameter θ that are robust with respect to the shape of the distribution F of the data. The case where F is fixed (nonrandom) and the case where F has a Dirichlet distribution are both treated. The procedures are based on the posterior distributions of the location parameter given the partial information contained in a robust estimate of location. We show consistency and asymptotic normality of the procedures and give instances where the Bayes procedure based on the full sample diverges while the Bayes procedures based on partial information converges and is asymptotically normal. Finally, we show that robust confidence procedures can be given a Bayesian interpretation.

1. Introduction. In a frequentist setting, it has long been recognized that in semiparametric models it can be advantageous to use only part of the information contained in the sample. Thus partial likelihood methods, which in many instances corresponds to using only the information supplied by the ranks of the data, have been shown to be very useful for estimating the parameters in semiparametric models. See for instance Cox (1972, 1975), and Kalbfleisch and Prentice (1973, 1980).

In a Bayesian context, the use of partial information can be found in the work of Bernstein (1946), von Mises (1931), Pratt (1965), Savage and Saxena [see Savage (1969)] and Pettitt (1983) among others.

We consider Bayes procedures for location based on the partial information contained in robust estimates of location. We find that these procedures are consistent in some of the cases considered by Diaconis and Freedman (1986a, b) where the Bayes procedures based on the full sample diverges. Moreover, the posterior distribution of the location parameter given a robust estimate converges to a normal distribution and the Bayes procedure inherits the robustness properties of the robust estimate used in the conditioning. This result can be regarded as giving a Bayesian interpretation to robust estimation theory.

Section 2 treats the case where the shape F of the distribution is nonrandom, known or unknown. Here we obtain robust confidence intervals for location with a Bayesian interpretation in the spirit of Rubin (1984). In Section 3 the consistency and asymptotic normality results are established for the case where F is assumed to have a Dirichlet distribution. Section 4 contains a convergence lemma and proofs of two of the results in Sections 2.

Received November 1986; revised May 1989
[1]Research partially supported by NSF grant DMS-86-02083.
[2]Research partially supported by NSF grant MCS-81-02523-01.
AMS 1980 *subject classifications*. 62A15, 62E20.
Key words and phrases. Consistency, asymptotic normality, robustness, location problem, Bayes procedures, Dirichlet prior.

2. Consistent and robust Bayes procedures when the error distribution is nonrandom, known or unknown. We consider the location model where $X_1, \dots, X_n$ is assumed to satisfy

$X_i = \theta + \varepsilon_i$, $i = 1, \dots, n$.
The errors $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. F.
θ has density $\pi(\theta)$ and is independent of $\varepsilon_1, \dots, \varepsilon_n$.

Here θ is the location parameter of interest and the error distribution F is a nuisance parameter which, in this section, is assumed to be nonrandom.

The robustness literature is full of practical examples where this model is appropriate. Here is one from a newspaper headline.

EXAMPLE (*Washington Post*, February, 1986). Consider a rocket whose performance depends critically on the launch time temperature θ on its surface. Hand-held infrared measuring devices are used to read temperatures on the surface immediately before the launching of the rocket. The readings are subject to errors with unknown error distribution. Gross errors are suspected leading to a desire for a robust estimate of θ. The density $\pi(\theta)$ is known from readings by accurate instruments during days of nonlaunch conditions.

In terms of distributions, our model is

$$\text{(2.1)} \qquad \begin{aligned} &\theta \text{ has density } \pi(\theta). \\ &\text{Given } \theta,\ X_1, \dots, X_n \text{ are i.i.d. } F_\theta, \text{ where } F_\theta(x) = F(x - \theta). \end{aligned}$$

This is the usual Bayesian setup except F is not assumed to be $N(0, \sigma^2)$. We will call (2.1) the *Bayesian location model with distributions* π *and* F.

We first investigate the consistency, asymptotic normality and robustness of Bayes procedures when the distribution $H_\theta(x) = H(x - \theta)$ that actually generates the data is different from $F_\theta(x) = F(x - \theta)$. Thus we are considering a Bayesian version of Huber's (1964) robustness setup where he asked: "If we use the estimate appropriate for the model $F_\theta(x) = F(x - \theta)$, how does it perform if the true error distribution H is different from F?" Related Bayesian versions of this question have been considered by Freedman (1963, 1965), Fabius (1964) and Pratt (1965), as well as Rubin (1984), Doss (1984), Blackwell (1985) and Diaconis and Freedman (1986a, b).

2.1. *Consistency.* Using the arguments of Diaconis and Freedman (1986a, b), we immediately find that Bayes procedures can perform very badly when $H \neq F$. We compute the Bayes procedure using model (2.1) with F Cauchy, since a heavy-tailed F is a good candidate for coming up with a robust procedure. Here is what happens:

PROPOSITION 2.1. *Let Π_n be the posterior probability distribution of θ given $X_1, \dots, X_n$ computed according to model* (2.1) *with the prior π standard normal and F standard Cauchy. Suppose that $X_1, \dots, X_n$ is actually generated by a distribution $H \neq F$. It is possible to specify an H with an infinitely differentiable* *density h which is symmetric about zero (i.e., the true θ is zero), with a unique maximum at zero, such that Π_n is inconsistent. More precisely, as $n \to \infty$, almost surely (a.s. $[H]$), the posterior Π_n concentrates near $\pm\gamma$ for some positive number γ in the sense that, for each $\eta > 0$, as $n \to \infty$, $\Pi_n\{\theta: |\theta - \gamma| < \eta$ or $|\theta + \gamma| < \eta\} \to 1$ a.s. $[H]$. Moreover, for n large, the probability that Π_n concentrates close to γ is near $\frac{1}{2}$ and the probability that Π_n concentrates close to $-\gamma$ is near $\frac{1}{2}$.*

PROOF. Diaconis and Freedman (1986b) consider F random with a Dirichlet distribution $\mathbf{D}(\alpha)$, with $\alpha/\alpha(R)$ standard Cauchy. However, they point out that Korwar and Hollander (1973) have shown that Π_n for the model with $F \sim \mathbf{D}(\alpha)$, equals a.s. $[H]$ the Π_n for the model with F nonrandom and equal to $\alpha/\alpha(R)$. The result now follows from the arguments of Section 2 of Diaconis and Freedman (1986b). □

One of the surprising aspects of the above result is that a study of the likelihood function for the Cauchy model suggests the use of the sample median to estimate θ, and the sample median does quite well for the Diaconis–Freedman counterexample density h described in Proposition 2.1. In fact, in both these models, the sample median $\tilde{\theta}$ is strongly consistent and $\sqrt{n}(\tilde{\theta} - \theta)$ is nearly normal for moderate sample sizes n. This suggests a strategy for coming up with a consistent "Bayes" procedure: Since the posterior *given the sample* concentrates near the wrong values $\pm\gamma$, why not use the posterior *given the sample median* since then the posterior (by Bayes theorem) will be close to a normal distribution centered at the correct θ value? More generally, we would use the posterior distribution of θ given some consistent, well behaved estimate T_n. This idea can be found in the work of Bernstein (1946), von Mises (1931) and Pratt (1965), among others. We can think of it as a Bayes procedure based on the partial information contained in T_n. Or in other words, it uses a partial likelihood, i.e., the density (likelihood function) of T_n, rather than the full likelihood, in Bayes theorem. For instance, if T_n is the sample median, F has density f and n is odd, then the posterior density of θ given T_n is

$$\pi(\theta|T_n) \propto \pi(\theta)f(T_n - \theta)\{F(T_n - \theta)[1 - F(T_n - \theta)]\}^{(n-1)/2},$$

where $\propto$ denotes proportional to.

Besides the sample median, other good candidates for consistent robust estimates T_n would be the Hodges and Lehmann (1963) estimate, the trimmed mean or one of the Huber (1964) estimates.

Returning to the general case, we adopt Bernstein's (1946) condition on the estimate T_n of θ:

(2.2) The conditional distribution of $T_n - \theta$ given θ does not depend on θ.

Note that (2.2) is satisfied if T_n is a translation equivariant estimate of θ. If T_n satisfies (2.2), we will say that it is *translation equivariant in distribution*.

In what follows a sequence of random distribution functions $\tilde{G}_n$ will be said to *converge weakly in probability* to the distribution function G if $\tilde{G}_n(t)$ converge in probability to $G(t)$ at each continuity point t of G. This notion of convergence has also been used by Walker (1969) and Dawid (1970). Let δ_{θ_0} denote point mass at θ_0 and let $\Rightarrow$ denote weak convergence. Then our consistency result for the partial posterior is:

THEOREM 2.1. *Let $\Pi_n(\theta|T_n)$ be the posterior probability distribution of θ given T_n computed according to the Bayesian location model with error distribution F and with prior density π which is continuous and nonzero in a neighborhood of the true parameter value θ_0 and which is bounded on R. Assume that for $X_1, \dots, X_n$ a sample from F_θ, T_n is translation equivariant in distribution and that T_n converges in probability to θ. Finally, suppose that*

$$T_n \to \theta_0 \quad a.s. \left[H_{\theta_0}\right]. \tag{2.3}$$

We conclude that

(a) *$\Pi_n(\cdot|T_n)$ is consistent in the sense that*

$$\Pi_n(\cdot|T_n) \Rightarrow \delta_{\theta_0} \quad a.s. \left[H_{\theta_0}\right] \text{ as } n \to \infty.$$

(b) *If $\theta\pi(\theta)$ is also bounded on R, then the quadratic loss Bayes estimate $E(\theta|T_n)$ is consistent in the sense that $E(\theta|T_n) \to \theta_0$ a.s. $[H_{\theta_0}]$ as $n \to \infty$.*

(c) *If the convergence in* (2.3) *is in H_{θ_0} probability, then so is the convergence in* (a) *and* (b).

The proof, which is similar to the proof in Lo (1984), is given in Section 4.

The assumptions in the above result assume that T_n is consistent for samples from F_{θ_0} and for samples from H_{θ_0}. Without a condition of this type, identifiability is lost and no consistency result is possible. If F and H are symmetric about zero, then it is satisfied for (practically) all the T_n that have appeared in the literature. In particular, if T_n is the sample median and if π, F and H are as in Proposition 2.1, then $\Pi_n(\cdot|T_n)$ is consistent and we have an example where the posterior based on the entire sample is inconsistent, while the posterior based on partial information is consistent.

Note that if T_n is the sample median, $\Pi(\cdot|T_n)$ is consistent if F and H have medians zero and densities positive at zero.

2.2. *Asymptotic normality.* Next, we turn to the limit of the posterior distribution of $\sqrt{n}(\theta - T_n)$ given T_n computed according to the Bayesian location model with prior π and error distribution F. It turns out that if $X_1, \dots, X_n$ is generated by $H_\theta \neq F_\theta$, then this posterior does *not* converge to the asymptotic distribution that $-\sqrt{n}(T_n - \theta)$ has when $X_1, \dots, X_n$ is a sample from H_θ, but it converges to the asymptotic distribution that $-\sqrt{n}(T_n - \theta)$ has when $X_1, \dots, X_n$ is a sample from F_θ. Thus, in the limit, the distribution assumed in the model "dominates" the true distribution. Here is the result:

THEOREM 2.2. *Suppose that $\pi(\theta)$ and T_n satisfy the assumptions of Theorem* 2.1. *Assume that there exist a distribution function G and a sequence of* *constants* $\{a_n\}$ *such that for* $X_1, \ldots, X_n$ *a sample from* F_θ,

$$-a_n(T_n - \theta) \Rightarrow G. \tag{2.4}$$

Let $\tilde{\Pi}_n(\cdot \mid T_n)$ *denote the posterior probability distribution of* $a_n(\theta - T_n)$ *given* T_n *computed according to the Bayesian location model with error distribution* F. *Then*

(i) $\tilde{\Pi}_n(\cdot \mid T_n) \Rightarrow G$ *a.s.* $[H_{\theta_0}]$ *as* $n \to \infty$.

(ii) *If the convergence in* (2.3) *is in probability, so is the convergence in* (i).

The proof is given in Section 4.

Typically, $a_n = \sqrt{n}$ and G is the $N(0, \tau_F^2)$ distribution, where τ_F^2 is the asymptotic variance of $\sqrt{n}(T_n - \theta)$ when $X_1, \ldots, X_n$ is a sample from F.

In a much more general setting, Le Cam (1953, Theorem 7) considered the case corresponding to $F_\theta = H_\theta$ and established the convergence of the posterior density of $\sqrt{n}(\theta - M_n)$ given the sample $\underset{\sim}{X} = (X_1, \ldots, X_n)$ to a normal density, where M_n is the MLE (maximum likelihood estimate) of θ. Results of this type had been formally derived by Laplace (1820). Recently, Blackwell (1985), also in a more general setting, considered the case $F_\theta \neq H_\theta$ and established the convergence of the posterior density of $\sqrt{n}(\theta - M_n)$ given $\underset{\sim}{X}$ to a normal density determined by F_θ.

2.3. *Unknown error distribution.* Suppose that in model (2.1), F is taken to be H (which is unknown). Then, according to Theorem 2.2, the posterior distribution of θ given T_n can be approximated by the $N(T_n, \tau_H^2/n)$ distribution. Since τ_H^2 is unknown, we replace it by a consistent estimate $\hat{\tau}^2$ and propose the $N(T_n, \hat{\tau}^2/n)$ distribution as the approximate distribution of θ given T_n. Under certain conditions, this can be justified:

THEOREM 2.3. *Suppose that the prior density* π *is continuous and nonzero in a neighborhood of the true parameter value* θ_0 *and suppose that* π *is bounded on* R. *Assume that for* $X_1, \ldots, X_n$ *a sample from* H_θ, T_n *is translation equivariant in distribution and that* $-\sqrt{n}(T_n - \theta)$ *converge to an* $N(0, \tau_H^2)$ *distribution. If* $\hat{\tau}$ *is an estimate of* τ_H *which for each* $\varepsilon > 0$ *satisfies*

$$P\left[|\hat{\tau} - \tau_H| \geq \varepsilon \mid T_n\right] \to 0 \quad \text{as } n \to \infty \text{ a.s. } \left[H_{\theta_0}\right], \tag{2.5}$$

then for each $x \in R$,

$$P\left(\frac{\sqrt{n}(\theta - T_n)}{\hat{\tau}} \leq x \,\Big|\, T_n\right) \to \Phi(x) \quad \text{as } n \to \infty \text{ a.s. } \left[H_{\theta_0}\right],$$

where Φ *is the standard normal distribution function.*

PROOF. Let P^{T_n} denote the probability distribution of $(\theta, \hat{\tau})$ given T_n. On a set with probability 1, $\hat{\tau}$ converges in P^{T_n} probability to τ and $\sqrt{n}(\theta - T_n)$ converges in P^{T_n} law to $N(0, \tau^2)$. The result now follows from the Cramér–Slutsky theorem. □

REMARK 2.1. Theorem 2.3 implies that the confidence interval $T_n \pm z_\alpha \hat{\tau}$, where $z_\alpha = \Phi^{-1}(1 - \frac{1}{2}\alpha)$, has a Bayesian interpretation.

REMARK 2.2. When F is symmetric about zero, Theorem 2.2, in conjunction with the ideas of Stein (1956), can be used to construct an adaptive Bayes procedure: Let T_n be the adaptive estimate of location given by Stone (1975). Then, under the conditions of Theorem 2.2 and those of Stone, $\mathbf{L}(\sqrt{n}(\theta - T_n)|T_n)$ converges weakly in probability to the $N(0, 1/I(F))$ distribution, where $I(F)$ is the Fisher information of F. This $N(0, 1/I(F))$ limiting posterior distribution is the best possible for samples from $F(x - \theta)$. In fact, as shown by Le Cam (1953), the limiting posterior density of $\sqrt{n}(\theta - M_n)$ given $\underset{\sim}{X}$ is $N(0, 1/I(F))$. See also De Groot (1970).

REMARK 2.3. Lindley (personal communication) asks whether our partial posterior is the full sample posterior for some model. The answer is yes, approximately, in many interesting cases: Many robust estimates are maximum likelihood estimates (MLE's) for some model. Thus Huber's (1964) robust estimate is the MLE for Huber's least favorable distribution. Let $Q_\theta(x) = Q(x - \theta)$ denote the distribution for which T_n is the MLE. Then, when $H = F = Q$, the approximate posterior given in Theorem 2.2 is the approximate full sample posterior for samples from Q. This follows from Le Cam (1953). This remark corresponds to the idea that we obtain robust Bayes procedures by using a model distribution for $(X|\theta)$ with heavy tails.

REMARK 2.4. The results of this section apply not only in the location case. For instance, suppose we model $X_1, \dots, X_n$ to be a sample from a distribution F_θ with support (θ, ∞). Then if $T_n = X_{(1)}$ = smallest order statistic and if $X_1, \dots, X_n$ is a sample from a distribution H_θ with support (θ, ∞), then Theorems 2.1 and 2.2 apply.

3. Robust and consistent Bayes procedures when F has a Dirichlet distribution. Next we consider the case where F is treated as a nonparametric nuisance parameter with a Dirichlet [e.g. Ferguson (1973)] prior distribution. In particular we consider the model of Dalal (1979) and Diaconis and Freedman (1986a, b) where

(3.1)

θ has density $\pi(\theta)$.

F has the Dirichlet distribution $\mathbf{D}(\alpha)$ with absolutely continuous parameter measure α.

θ and F are independent.

Given (θ, F), $X_1, \dots, X_n$ are independent with distribution function $F_\theta(x) = F(x - \theta)$, all $x \in R$.

Again we consider the convergence of the Bayes procedure given as estimate T_n when the sample $X_1, \dots, X_n$ is generated by a continuous distribution H_θ not necessarily connected to the model (3.1). Let $\alpha(t) = \alpha((-\infty, t])$ and $\alpha_\theta(t) = \alpha(t - \theta)/\alpha(R)$. We find:

THEOREM 3.1. *Suppose that $\pi(\theta)$ is bounded on R and is continuous and nonzero in a neighborhood of the true parameter value θ_0, and suppose that the posterior $\Pi_n(\theta|T_n)$ of θ given T_n is computed assuming that $X_1, \dots, X_n$ is generated according to* (3.1). *In addition, assume that T_n is translation equivariant and that*

$$T_n \to \theta_0 \quad a.s. \left[H_{\theta_0}\right],\ H_{\theta_0} \text{ continuous}. \tag{3.2}$$

Then,

(a) *$\Pi_n(\cdot|T_n)$ is consistent, i.e., $\Pi_n(\cdot|T_n) \to \delta_{\theta_0}$ a.s. $[H_{\theta_0}]$ as $n \to \infty$.*

(b) *If $\theta\pi(\theta)$ is bounded on R, then $E(\theta|T_n)$ is consistent, i.e., $E(\theta|T_n) \to \theta_0$ a.s. $[H_{\theta_0}]$ as $n \to \infty$.*

(c) *If the convergence in* (3.2) *is in probability, so is the convergence in* (a) *and* (b).

(d) *If there is a distribution function G and a sequence of constants $\{a_n\}$ such that for $X_1, \dots, X_n$ a sample from α_θ,*

$$-a_n(T_n - \theta) \Rightarrow G, \tag{3.3}$$

then the posterior probability distribution $\tilde{\Pi}_n(\cdot|T_n)$ of $a_n(\theta - T_n)$ given T_n computed according to model (3.1) *converges in law a.s. $[H_{\theta_0}]$ to G, i.e.,*

$$\tilde{\Pi}_n(\cdot|T_n) \Rightarrow G \quad a.s. \left[H_{\theta_0}\right] \text{ as } n \to \infty. \tag{3.4}$$

(e) *If the convergence in* (3.3) *is in probability rather than a.s., so is the convergence in* (3.4).

PROOF. Let Δ_n be the part of the underlying probability space where $X_i \neq X_j$ for $1 \le i < j \le n$. Then Δ_n has H_{θ_0} probability 1. We will consider $\tilde{\Pi}_n(\cdot|T_n)$ on Δ_n. On Δ_n, we may find the conditional distribution of (θ, F) given T_n by first conditioning on Δ_n, then on T_n. By Theorem 2.5 of Korwar and Hollander (1973), given Δ_n, $X_1, \dots, X_n$ are i.i.d. with distribution α_θ and by the proof of Lemma 2.1 in Diaconis and Freedman (1986b), the joint distribution of θ and $X_1, \dots, X_n$ is the same as in model (2.1) with F_θ replaced by α_θ. Since T_n is a function of $X_1, \dots, X_n$, it follows that in Δ_n, $\tilde{\Pi}_n(\cdot|T_n)$ equals the posterior for model (2.1) with F_θ replaced by α_θ. The present result then follows from Theorems 2.1 and 2.2. □

If we apply this result to the Diaconis–Freedman (1986b) example where α is Cauchy and H is the Diaconis–Freedman distribution (see Proposition 2.1), we find that the posterior for model (3.1) based on the whole sample diverges, while the posterior of θ given the sample median converges a.s. to θ_0.

4. Proofs of the main results.

LEMMA 4.1. *Let $\{\mu_n\}$ be a sequence of probability measures on R, let $\{t_n\}$ be a sequence of real numbers and let $t_0 \in R$. If*

(i) *g is a bounded (by C) function on R and is continuous in a neighborhood of t_0,*
(ii) $\mu_n \Rightarrow \delta_0$,
(iii) $t_n \to t_0$,

then $\int |g(t_n - s) - g(t_n)|\mu_n(ds) \to 0$.

PROOF. Let $D_n(s, t_n) = |g(t_n - s) - g(t_n)|$. For any $\delta > 0$,

$$\int D_n(s, t_n)\mu_n(ds) = \int_{A_\delta} D_n(s, t_n)\mu_n(ds) + \int_{A_\delta^c} D_n(s, t_n)\mu_n(ds),$$

where $A_\delta = [|s| \leq \delta]$. Note that

$$\int_{A_\delta^c} D_n(s, t_n)\mu_n(ds) \leq 2C\mu_n(|s| > \delta),$$

which tends to zero by (ii).

It remains to consider the integral of D_n on A_δ. Let K be a closed and bounded interval containing t_0 in its interior. For n large enough, say $n \geq n_0$, $t_n \in K$. Let $I = \{t + s\colon\ t \in K,\ s \in [|s| \leq \delta]\}$ and note that I is a compact interval. Choose δ and K such that g is continuous on I. Hence g is uniformly continuous on I. That is, for all $\varepsilon > 0$, there is a $\delta > 0$ such that $|s| \leq \delta$ implies $\sup_{t \in K} |g(t - s) - g(t)| < \varepsilon$. Therefore,

$$\sup_{n \geq n_0} |g(t_n - s) - g(t_n)| < \varepsilon \quad \text{for } |s| \leq \delta.$$

It follows that

$$\int_{A_\delta} D_n(s, t_n)\mu_n(ds) \leq \varepsilon\mu_n[|s| \leq \delta] \leq \varepsilon$$

and the proof is complete. □

PROOF OF THEOREM 2.1. Let $\tilde{Q}_n$ denote the probability distribution of T_n for $X_1, \ldots, X_n$ a sample from F. Since T_n' is equivariant in distribution, then

$$P(T_n \in B|\theta) = \tilde{Q}_n(B - \theta)$$

for each Borel set B. It follows that

$$\Pi_n(\theta \in B|T_n) = \frac{\int_B \pi(\theta)Q_n(T_n - d\theta)}{\int \pi(\theta)Q_n(T_n - d\theta)},$$

where $Q_n(t) = \tilde{Q}_n((-\infty, t])$ is the distribution function of T_n when $\theta = 0$. By the change of variable $s = T_n - \theta$, we have

$$\Pi_n(\theta \in B|T_n) = \frac{\int_{T_n - B}\pi(T_n - s)Q_n(ds)}{\int\pi(T_n - s)Q_n(ds)}.$$

Similarly, the characteristic function corresponding to the posterior distribution $\Pi_n(\cdot|T_n)$ is

$$\phi_n(u|T_n) = \frac{\int e^{iu(T_n - s)}\pi(T_n - s)Q_n(ds)}{\int\pi(T_n - s)Q_n(ds)}. \tag{4.1}$$

Let $N_n(u)$ denote the numerator of this expression. Then

$$\left|N_n(u) - e^{iuT_n}\pi(T_n)\int e^{-ius}Q_n(ds)\right| \le \int\left|e^{-ius}\pi(T_n - s) - e^{-ius}\pi(T_n)\right|Q_n(ds)$$

$$= \int|\pi(T_n - s) - \pi(T_n)|Q_n(ds).$$

The last expression tends to zero a.s. $[H_{\theta_0}]$ by Lemma 4.1. Moreover, since $Q_n \Rightarrow \delta_0$, then $\int e^{-ius}Q_n(ds) \to 1$. Thus

$$\lim_{n\to\infty} N_n(u) = \lim_{n\to\infty} e^{iuT_n}\pi(T_n) = e^{iu\theta_0}\pi(\theta_0) \quad \text{a.s.} \left[H_{\theta_0}\right].$$

By a similar argument, the limit of the denominator in (4.1) is $\pi(\theta_0) > 0$. Thus

$$\lim_{n\to\infty} \phi_n(u|T_n) = e^{iu\theta_0} \quad \text{a.s.} \left[H_{\theta_0}\right] \tag{4.2}$$

and $\Pi_n(\theta|T_n) \Rightarrow \delta_{\theta_0}$ a.s. $[H_{\theta_0}]$. This completes the proof of (a).

To establish (b), note that

$$E(\theta|T_n) = \frac{\int(T_n - s)\pi(T_n - s)Q_n(ds)}{\int\pi(T_n - s)Q_n(ds)}.$$

As before, $\int\pi(T_n - s)Q_n(ds) \to \pi(\theta_0)$ a.s. $[H_{\theta_0}]$. Next note that

$$\left|\int(T_n - s)\pi(T_n - s)Q_n(ds) - T_n\pi(T_n)\int Q_n(ds)\right|$$

$$\le \int|(T_n - s)\pi(T_n - s) - T_n\pi(T_n)|Q_n(ds) \to 0 \quad \text{a.s.} \left[H_{\theta_0}\right]$$

by Lemma 4.1. Thus since $T_n\pi(T_n) \to \theta_0\pi(\theta_0)$ a.s. $[H_{\theta_0}]$, then $E(\theta|T_n) \to \theta_0$ a.s. $[H_{\theta_0}]$.

To establish (c), we use a Skorokhod representation and replace T_n with a sequence $Y_n(\theta_0)$ with the same conditional distribution given θ_0 as T_n, but with $Y_n(\theta_0) \to \theta_0$ a.s. $[H_{\theta_0}]$.

By construction, given θ_0, $Y_n(\theta_0)$ has the same probability distribution as T_n. Thus the proof of (a) and (b) leads to the conclusion that the characteristic function $\phi_n(u|Y_n(\theta_0))$ on the right-hand side of (4.1) with T_n replaced by $Y_n(\theta_0)$ converges a.s. $[H_{\theta_0}]$ to the appropriate limits. Since, $\phi_n(u|Y_n(\theta_0))$ has the same distributions as $\phi_n(u|T_n)$, then (c) follows for the posterior distribution. The proof for $E(\theta|T_n)$ is similar. □

PROOF OF THEOREM 2.2. Note that the posterior characteristic function of $a_n(\theta - T_n)$ given T_n is

$$\tilde{\phi}(u|T_n) = \frac{\int e^{ia_n(T_n - s - T_n)u}\pi(T_n - s)Q_n(ds)}{\int \pi(T_n - s)Q_n(ds)} \tag{4.3}$$
$$= \frac{\int e^{-iua_n s}\pi(T_n - s)Q_n(ds)}{\int \pi(T_n - s)Q_n(ds)}.$$

Let $\tilde{N}_n(u)$ denote the numerator of (4.3). Then

$$\left|\tilde{N}_n(u) - \pi(T_n)\int e^{-iua_n s}Q_n(ds)\right| \le \int |\pi(T_n - s) - \pi(T_n)|Q_n(ds) \to 0$$

a.s. $\left[H_{\theta_0}\right]$ by Lemma 4.1.

Next note that by (2.4),

$$\int e^{-iua_n s}Q_n(ds) \to \tilde{\phi}(u),$$

where $\tilde{\phi}(u)$ denotes the characteristic function of the conditional limiting distribution G. Thus $\tilde{N}_n(u) \to \pi(\theta_0)\tilde{\phi}(u)$ a.s. $[H_{\theta_0}]$. Similarly, the denominator of (4.3) converges a.s. $[H_{\theta_0}]$ to $\pi(\theta_0) > 0$. Thus $\tilde{\phi}_n(u|T_n) \to \tilde{\phi}(u)$ a.s. $[H_{\theta_0}]$ and the proof of (i) is completed. The (ii) part follows from a Skorokhod construction as in the proof of Theorem 2.1. □

Acknowledgment. The authors of this paper benefited greatly from discussions with L. Le Cam.

REFERENCES

BERNSTEIN, S. N. (1946). *Probability Theory*, 4th ed. Gostechizdat, Moscow. (In Russian. First edition 1917.)

BLACKWELL, D. (1985). Approximate normality of large products. Technical Report No. 54, Dept. Statist., Univ. of Calif., Berkeley.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–202.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

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

DAWID, A. P. (1970). On the limiting normality of posterior distributions. *Proc. Cambridge Phil. Soc.* **67** 625–633.

DE GROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

DIACONIS, P. and FREEDMAN, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–25.

DIACONIS, P. and FREEDMAN, D. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68–87.

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

FABIUS, J. (1964). Asymptotic behavior of Bayes estimates. *Ann. Math. Statist.* **35** 846–856.

FERGUSON, T. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

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

FREEDMAN, D. (1965). On the asymptotic behavior of Bayes estimates in the discrete case II. *Ann. Math. Statist.* **36** 454–456.

HODGES, J. L. JR. and LEHMANN, E. L. (1963). Estimates of location based or rank tests. *Ann. Math. Statist.* **34** 598–611.

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression model. *Biometrika* **60** 267–278.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Probab.* **1** 705–711.

LAPLACE, P. S. (1820). *Théorie Analytic des Probabilités* 338–353. Courcier, Paris (Third Edition).

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

LO, A. Y. (1984). Consistency in the location model: The undominated case. *Ann. Statist.* **12** 1584–1587.

PETTITT, A. N. (1983). Likelihood based inference using signed ranks for matched pairs. *J. Roy. Statist. Soc. Ser. B* **45** 287–296.

PRATT, J. W. (1965). Bayesian interpretation of standard inference statements. *J. Roy. Statist. Soc. Ser. B* **27** 169–203.

RUBIN, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.* **12** 1151–1172.

SAVAGE, I. R. (1969). Nonparametric statistics: A personal review. *Sankhyā Ser. A* **31** 107–144.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 187–196. Univ. California Press.

STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284.

VON MISES, R. (1931). *Wahrscheinlichkeitsrechung.* Springer, Berlin.

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

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1 OXFORD STREET
CAMBRIDGE, MASSACHUSETTS 02138

DEPARTMENT OF STATISTICS
SUNY AT BUFFALO
350 PORTER QUADRANGLE
BUILDING 4
BUFFALO, NEW YORK 14261