The Annals of Statistics
1986, Vol. 14, No. 1, 68–87

ON INCONSISTENT BAYES ESTIMATES OF LOCATION

By P. Diaconis[1] and D. Freedman[2]

Stanford University and University of California, Berkeley

In some relatively natural settings, Bayes estimates of location are shown to be inconsistent.

1. Introduction. Consider the problem of estimating a location parameter θ observed subject to errors ε_i. The data are modelled as

$$X_i = \theta + \varepsilon_i \tag{1.1}$$

where the ε_i are independent with unknown distribution function F. Bayes estimates are computed from a prior distribution for θ and F. One natural choice is to take θ and F independent, F having a Dirichlet distribution with parameter measure α. Ferguson (1974) contains a review of Dirichlet priors. The posterior distribution for such a prior will be given in Lemma 2.1.

With squared error as loss, the Bayes estimate of θ is the mean of the posterior distribution. One of our principal results is that for some prior distributions, the Bayes estimate is inconsistent: There are F's with a density symmetric about zero such that the Bayes estimate for θ oscillates between two nonzero numbers as data accumulate.

To be specific, suppose that the prior density f for θ is standard normal, while the parameter measure α for the Dirichlet is Cauchy, having density $\alpha' = g(x) = 1/\pi(1 + x^2)$. Let $\pi_n = \pi_n(X_1, \ldots, X_n)$ be the posterior distribution of θ and F given the data $X_1, \ldots, X_n$. We will construct a C_∞ density h with compact support, symmetric about 0 and with a strict maximum at 0, such that if $\theta = 0$ and the ε's are independent draws from h, the posterior distribution of θ oscillates between two false values $\pm\gamma$, and is therefore inconsistent. Here, γ is a positive number depending on h. If desired, h can be chosen strictly positive on the interior of its interval of support.

THEOREM 1. *Let X_i follow the model* (1.1), *where $\theta = 0$ and the ε_i have a compactly supported C_∞ density h, which is symmetric about* 0, *with a strict maximum at* 0. *For the prior, θ has the standard normal density, and F is independently drawn from the Dirichlet based on the standard Cauchy. For some h: as $n \to \infty$, almost surely, the posterior π_n given $X_1, \ldots, X_n$ concentrates near $\pm\gamma$, where γ is a positive number depending on h. For each large n, there is probability near $\frac{1}{2}$ that π_n concentrates close to γ, and probability near $\frac{1}{2}$ that π_n*

Received May 1983; revised August 1985.
[1] Research partially supported by National Science Foundation Grant MCS80-24649.
[2] Research partially supported by National Science Foundation Grant MCS83-01812.
AMS 1980 *subject classifications*. Primary 62A15; secondary 62E20.
Key words and phrases. Bayes estimates, consistency, Dirichlet priors.

concentrates close to $-\gamma$. *Moreover, for any* $\eta > 0$,

$$\limsup \pi_n\{(\theta, F): |\theta - \gamma| < \eta\} = 1 \quad a.e.,$$
$$\limsup \pi_n\{(\theta, F): |\theta + \gamma| < \eta\} = 1 \quad a.e.$$

Of course under the conditions of the theorem, the median is a consistent estimate of θ, as is any trimmed mean; so the Bayes estimates are worse than available objectivist procedures. Further motivation and philosophical discussion is in Diaconis and Freedman (1986). The proof of the theorem is deferred to Section 2. Estimates of the form considered in the theorem have actually been suggested in the Bayesian literature; see Dalal (1979a, 1979b, 1980). Dempster (1976) gives an extensive survey of the Bayesian approach to robust estimation, which is linked to estimates of the form considered in the theorem. Similar estimates have been suggested by Fraser (1976) and Johns (1979) from the frequentist viewpoint. The argument for Theorem 1 shows that for some underlying distributions these estimators are inconsistent.

The theorem remains valid if the normal prior density for θ is replaced by any smooth, everywhere positive density. Nor is the choice of Cauchy for α crucial. Any t-density works as well. However, if α has a density α' and $\log \alpha'$ is convex, then the posterior converges to point mass at θ, for essentially any choice of h. This may be shown using the arguments of Freedman and Diaconis (1982b).

The density h constructed for Theorem 1 has a single global maximum at 0; but h is not strongly unimodal—it has three local maxima (see Figure 1 in Section 2). Using the arguments of Freedman and Diaconis (1982b), it can be shown that if h is continuous, strongly unimodal, and symmetric about θ, the posterior converges to point mass at θ almost surely. (A density is strongly unimodal if it increases to its unique maximum and then decreases.)

In Diaconis and Freedman (1983b) we have an argument proving that any location mixture of Dirichlets gives a consistent estimate of the sampling distribution. So the marginal posterior distribution of the sampling distribution converges to point mass at $\mathscr{H}$, the distribution function with density h. This has a peculiar implication: Indeed, when π_n concentrates on θ's near γ, then π_n must concentrate on F's which are near $\mathscr{H}$ shifted to the left by γ; when π_n concentrates on θ's near $-\gamma$, then π_n must concentrate on F's which are near $\mathscr{H}$ shifted to the right by γ. Thus, π_n gets both θ and F badly wrong, but it gets the law of the data, namely F shifted by θ, nearly right.

One of the issues in this example is the identifiability of the parameters. In general, of course, the convolution $\delta_\theta * F$ cannot be decomposed into its components θ and F; here, δ_x is point mass at x. However, from the point of view of a Bayesian with the prior in the theorem, the two parameters θ and F are identifiable, as the next result shows.

THEOREM 2. *In the setting of Theorem* 1, *there is a measurable function* ϕ *such that* $\phi(\delta_\theta * F) = \theta$ *for almost all* θ *and* F.

In particular, a Bayesian will be convinced that his Bayes rule is consistent at almost all pairs (θ, F). It is therefore of mathematical interest, and we think also of interest from the point of view of the foundations of inference, to ask about consistency at particular (θ, F), especially for symmetric F where θ is objectively identifiable. That is why we present Theorem 1. We give an extensive discussion of the relation between consistency and Bayesian inference in Diaconis and Freedman (1986).

As noted above, one standard way to make θ identifiable is to require that the ε's be symmetric. Then, it seems reasonable to symmetrize the distribution function chosen from the Dirichlet. If G is the law of X, let G^- be the law of $-X$, and let $\overline{G} = \frac{1}{2}(G + G^-)$. So $\overline{G}$ is symmetric. Let $\overline{D}_\alpha$ be the law of $F = \overline{G}$, where G has law D_α; this is a "symmetrized" Dirichlet.

For the next theorem, let θ have a normal prior density, and let F be independent of θ, having the prior distribution $\overline{D}_\alpha$, where α is Cauchy. Let $\overline{\pi}_n$ be the posterior distribution for θ: This will be computed in Lemma 3.1. Again, as Theorem 3 below demonstrates, $\overline{\pi}_n$ can oscillate between two false values $\pm\gamma$ for θ. This time, the inconsistency spreads to the posterior opinion of the sampling distribution. Thus, a location mixture of symmetrized Dirichlets can be inconsistent for the sampling distribution—even when the latter is symmetric. So, a straightforward way of putting a prior on symmetric ε's does not cure the inconsistency of the Bayes procedures.

THEOREM 3. *Let X_i follow the model* (1.1), *where $\theta = 0$ and the ε_i have a compactly supported C_∞ density h, which is symmetric about* 0, *with a strict maximum at* 0. *For the prior, θ has the standard normal density, and F is independently drawn from the symmetrized Dirichlet based on the standard Cauchy. The posterior $\overline{\pi}_n$ given $X_1, \dots, X_n$ is computed from this prior. For suitable h: as $n \to \infty$, almost surely, $\overline{\pi}_n$ concentrates near $\pm\gamma$, where γ is a positive number depending on h. For each large n, there is probability near $\frac{1}{2}$ that $\overline{\pi}_n$ concentrates near γ, and probability near $\frac{1}{2}$ that $\overline{\pi}_n$ concentrates near $-\gamma$. Moreover, for any $\eta > 0$,*

$$\limsup_{n\to\infty} \overline{\pi}_n\{\theta\colon |\theta - \gamma| < \eta\} = 1 \quad a.e.,$$

$$\limsup_{n\to\infty} \overline{\pi}_n\{\theta\colon |\theta + \gamma| < \eta\} = 1 \quad a.e.$$

If desired, h can be chosen strictly positive on the interior of its interval of support.

The posteriors π_n and $\overline{\pi}_n$ are computed by using a theorem about the Dirichlet due to Korwar and Hollander (1973). Originally, we used a discretization argument. An abstract version of this is given in Section 4.

In Diaconis and Freedman (1982, 1983a) we discuss breakdown properties of the rules computed here. Related results appear in Huber (1984).

2. The first construction. The first step is to compute the posterior distribution of θ and F given the data; a similar result was conjectured by Dalal (1979a).

LEMMA 2.1. *With respect to the prior, let θ and F in the model* (1.1) *be independent, θ having density f and F being Dirichlet with parameter measure α which is absolutely continuous; let $g = \alpha'/\|\alpha\|$, where $\|\alpha\|$ is the mass of α. Let Δ_n be the set where $X_i \neq X_j$ for $1 \leqq i < j \leqq n$. On Δ_n, the posterior π_n can be characterized as follows*:

$$\pi_n(d\theta) = C_n^{-1} f(\theta) \prod_{i=1}^{n} g(X_i - \theta)\, d\theta,$$

where

$$C_n = \int_{-\infty}^{\infty} f(\theta) \prod_{i=1}^{n} g(X_i - \theta)\, d\theta,$$

$$\pi_n\{dF|\theta\} \quad \text{is} \quad D\left(\alpha + \sum_{i=1}^{n} \delta_{X_i - \theta}\right).$$

PROOF. Write P for the joint law of θ, F and $X_1, X_2, \ldots$. Thus, π_n is the law of θ and F given $X_1, \ldots, X_n$, computed according to P. On Δ_n, we may compute π_n by first conditioning on Δ_n, then on $X_1, \ldots, X_n$. But $\Delta_n = \{\varepsilon_i \neq \varepsilon_j$ for $1 \leq i < j \leq n\}$, so relative to P given Δ_n: The parameter θ still has density f; the ε_i are independent with common density g, and are independent of θ, by Theorem 2.5 of Korwar and Hollander (1973); and $X_i = \theta + \varepsilon_i$. □

REMARK 1. Lemma 2.1 gives the posterior when the observed values are all distinct. The argument also gives the posterior in general: $\pi_n(d\theta) = C_n^{-1}\pi_n(d\theta) = C_n^{-1} f(\theta)\Pi^* g(x_i - \theta)\, d\theta$, where $C_n = \int_{-\infty}^{\infty} f(\theta)\Pi^* g(x_i - \theta)\, d\theta$; the * signifies that the products are over distinct values only. Finally, $\pi_n(dF|\theta)$ remains the same.

REMARK 2. Under squared error, the Bayes estimate $\hat{\theta}(X_1, \ldots, X_n)$ of θ is the posterior mean. For the posterior computed in Lemma 2.1,

$$\hat{\theta}(X_1, \ldots, X_n) = C_n^{-1} \int_{-\infty}^{\infty} \theta f(\theta) \prod_{i=1}^{n} g(X_i - \theta)\, d\theta.$$

This coincides with the Bayes rule for the model (1.1) if ε_i has known density g; in the display, however, $g = \alpha'/\|\alpha\|$ is a feature of the prior for F.

Here is a sketch of the rest of the argument for Theorem 1; the rigor will follow. The computation is very similar to the one in Freedman and Diaconis (1982b). Let

$$M(x) = \log(1 + x^2), \tag{2.1}$$

so

$$g(x) = \frac{1}{\pi}\exp\{-M(x)\}.$$

Let

$$H(u) = \int M(x - u) h(x)\, dx. \tag{2.2}$$

The density h will be constructed so that $h(x)\,dx$ is essentially mass $\frac{1}{2}$ at each of $\pm a$, where $a > 1$. Then $H(u)$ has a local maximum at 0, and global minima at $\pm\gamma$, where $\gamma \doteq \sqrt{a^2 - 1}$. Now by Lemma 2.1, the posterior density of θ is

$$C_n^{-1}\pi^{-n}f(\theta)\exp\{-S_n(\theta)\}, \tag{2.3}$$

where

$$S_n(\theta) = \sum_{i=1}^{n} M(X_i - \theta) = nH(\theta) + \sqrt{n}\,G_n(\theta) \tag{2.4a}$$

and

$$G_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [M(X_i - \theta) - H(\theta)] \tag{2.4b}$$

is asymptotically a Gaussian process. In particular, the (unnormalized) posterior mass on a neighborhood of $\pm\gamma$ vanishes at the rate $\exp\{-nH(\gamma)\}$ while the posterior mass outside this neighborhood vanishes at a faster rate. Thus, the normalized posterior mass concentrates near $\pm\gamma$. The distribution of mass between the vicinity of γ and the vicinity of $-\gamma$ is controlled by the relative sizes of $G_n(\gamma)$ and $G_n(-\gamma)$; these two variables have a nonsingular Gaussian joint limiting distribution, so posterior mass shifts back and forth between the two neighborhoods.

LEMMA 2.2. *Fix $a > 1$. Recall M from* (2.1). *Let*

$$H_a(\theta) = \tfrac{1}{2}M(a - \theta) + \tfrac{1}{2}M(-a - \theta).$$

Then $H_a(\cdot)$ is symmetric, has a strict local maximum at 0 *where $H_a'' < 0$, and strict global minima at $\pm(a^2 - 1)^{1/2}$ where $H_a'' > 0$.*

PROOF. Calculus. □

LEMMA 2.3. *There is a compactly supported C_∞ probability density h which is symmetric about* 0 *with a strict maximum at* 0. *In addition,*

$$H(\theta) = \int M(x - \theta)h(x)\,dx$$

is symmetric, has a strict local maximum at 0 *where $H'' < 0$, and strict global minima at $\pm\gamma$ where $H'' > 0$. Here, $\gamma > 0$ depends on h, but is close to $(a^2 - 1)^{1/2}$. See Figure* 1.

PROOF. Choose a sequence h_n of densities which are C_∞, symmetric, supported on $[-2a, 2a]$, with strict maxima at 0, such that $h_n(\theta)\,d\theta \to \frac{1}{2}\delta_a + \frac{1}{2}\delta_{-a}$. Now look at the derivatives

$$\begin{aligned} H_n^{(j)}(\theta) &= (-1)^j \int M^{(j)}(x - \theta)h_n(x)\,dx \\ &\to (-1)^j \tfrac{1}{2}\left[M^{(j)}(a - \theta) + M^{(j)}(-a - \theta)\right] = H_a^{(j)}(\theta), \end{aligned}$$

because $M^{(j)}$ is bounded continuous. Now use Lemma 2.2. □

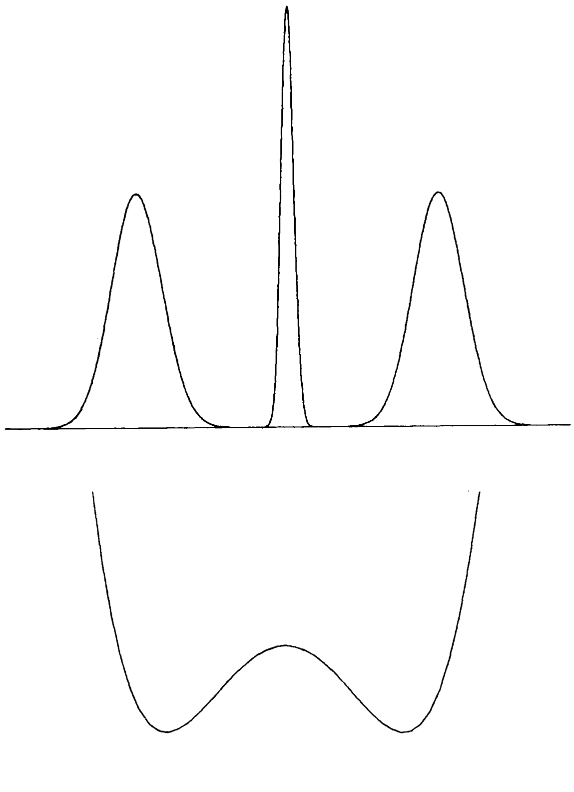

FIG. 1. *Graph of the density h and the function $H(\theta) = \int M(x - \theta)h(x)\,dx$.*

Let $X_1, X_2, \ldots,$ be independent, with the common density h constructed in Lemma 2.3. The asymptotic behavior of $S_n(\theta)$ in (2.4) will now be determined, with a view to proving the oscillatory behavior of the posterior; for inconsistency alone, see Berk (1966). Some results on the Brownian bridge will be helpful. To state them, let $\mathscr{H}$ be the distribution function of the density h constructed in Lemma 2.3, and let $\mathscr{H}_n$ be the empirical distribution function of $X_1, X_2, \ldots, X_n$. Let

$$B_n = \sqrt{n}\,(\mathscr{H}_n - \mathscr{H}).$$

Notice that B_n vanishes off $[-2a, 2a]$. Of course, $B_n = \tilde{B}_n \circ \mathscr{H}$, where $\tilde{B}_n$ is the approximate Brownian bridge on $[0,1]$ based on the empirical distribution of a sample of size n from the uniform. The law of the iterated logarithm (Chung, 1949) implies the following result.

LEMMA 2.4. *There is a finite constant A, and for almost all ω an $N = N_\omega < \infty$, such that $n > N$ entails $|B_n(t)| < A(\log\log n)^{1/2}$ for all t.*

Clearly, B_n converges weakly to the Brownian bridge $B_{\mathscr{H}} = \tilde{B} \circ \mathscr{H}$, where $\tilde{B}$ is the ordinary Brownian bridge on [0, 1].

LEMMA 2.5. *Let ϕ and ψ be two bounded continuous functions on the line. Then*

(a) $E\{\int \phi \, dB_{\mathscr{H}}\} = 0$;
(b) $\operatorname{cov}\{\int \phi \, dB_{\mathscr{H}}, \int \psi \, dB_{\mathscr{H}}\} = \operatorname{cov}[\phi(X_i), \psi(X_i)]$.

PROOF. This can be reduced to the corresponding result for the ordinary Brownian bridge $\tilde{B}$. Since $\tilde{B}(t) = W(t) - tW(1)$ for $0 \leqq t \leqq 1$, where W is a standard Weiner process, the result for $\tilde{B}$ is easily checked. □

LEMMA 2.6. *Let X have a symmetric, absolutely continuous distribution.*

(a) *The distribution of $M(X - \gamma) - M(X + \gamma)$ is absolutely continuous*;
(b) $P\{M(X - \gamma) = M(X + \gamma)\} = 0$;
(c) $E\{M(X - \gamma)^k\} = E\{M(X + \gamma)^k\}$ for $k = 1, 2, \dots$.

PROOF. (a) Let $\phi(t) = (t - \gamma) - M(t + \gamma)$. Then ϕ is smooth, and two-to-one except at 0.
(b) Immediate from (a).
(c) Use symmetry. □

The notation in the next lemma may seem perverse, but Z_+ is associated with behavior near γ, and Z_- with behavior near $-\gamma$.

LEMMA 2.7. *Let $Z_+ = \int M(u - \gamma)\, dB_{\mathscr{H}}(u)$ and $Z_- = \int M(u + \gamma)\, dB_{\mathscr{H}}(u)$. Then (Z_+, Z_-) has a nonsingular symmetric bivariate Gaussian distribution.*

PROOF. By Lemma 2.5(b),

$$\operatorname{cov}(Z_+, Z_-) = \operatorname{cov}\big[M(X_i - \gamma), M(X_i + \gamma)\big].$$

Symmetry follows from Lemma 2.6(c); if the distribution were singular, it would have to concentrate on the 45° line, contradicting Lemma 2.6(b). The mean is 0 by Lemma 2.5(a). □

Turn now to (2.4). Clearly,

$$G_n(\theta) = \int M(u - \theta)\, dB_n(u). \tag{2.5}$$

Then

$$\begin{aligned} G_n^{(j)}(\theta) &= (-1)^j \int M^{(j)}(u - \theta)\, dB_n(u) \\ &= (-1)^{j+1} \int B_n(u) M^{(j+1)}(u - \theta)\, du. \end{aligned} \tag{2.6}$$

Now Lemma 2.4 can be used.

LEMMA 2.8. *Let j be a fixed integer. There is a finite constant $A = A_j$, and for almost all ω an $N = N_\omega < \infty$, such that $n > N$ entails $|G_n^{(j)}(\theta)| < A(\log\log n)^{1/2}$ for all θ.*

LEMMA 2.9. *Under the conditions of Theorem 1, if $\eta > 0$, then on Δ_n, $\pi_n\{(\theta, F): |\theta - \gamma| < \eta$ or $|\theta + \gamma| < \eta\} \to 1$ a.e.*

PROOF. Let $D_n = C_n^{-1}\pi^{-n}$, the normalizing constant. Using (2.3) and Lemma 2.8, a.e. for large n,

$$\pi_n\{(\theta, F): |\theta - \gamma| < \eta/2 \quad \text{or} \quad |\theta + \gamma| < \eta/2\}$$
$$\geq 2\eta \cdot D_n \cdot \min_{|\theta - \gamma| < \eta/2} f(\theta) \cdot \exp\{-n\rho - A(n \log\log n)^{1/2}\},$$

where

$$\rho = \max_{|\theta-\gamma|<\eta/2} H(\theta).$$

On the other hand,

$$\pi_n\{(\theta, F): |\theta - \gamma| \geq \eta \quad \text{and} \quad |\theta + \gamma| \geq \eta\}$$
$$\leq D_n \cdot \int f(\theta)\, d\theta \cdot \exp\{-n\rho^* + A(n \log\log n)^{1/2}\},$$

where $\rho^* = \min_C \mathscr{H}(\theta)$, with $C = \{|\theta - \gamma| \geq \eta \text{ and } |\theta + \gamma| \geq \eta\}$. Clearly, $\rho^* > \rho$. □

LEMMA 2.10. *Let*

$$Z_{+n} = \int M(u - \gamma)\, dB_n(u), \qquad Z_{-n} = \int M(u + \gamma)\, dB_n(u),$$

$$Y_{+n} = \int M'(u - \gamma)\, dB_n(u), \qquad Y_{-n} = \int M(u + \gamma)\, dB_n(u).$$

Fix δ positive but small. Choose η positive but so small that for $|\theta - \gamma| \leq \eta$,

$$H''(\gamma) - \delta \leq H''(\theta) \leq H''(\gamma) + \delta.$$

Put $t = \sqrt{n}(\theta - \gamma)$. Almost surely, for all sufficiently large n, for all θ with $|\theta - \gamma| \leq \eta$, upper and lower bounds on $S_n(\theta)$ are, respectively,

$$nH(\gamma) + \sqrt{n} Z_{+n} - Y_{+n}t + \tfrac{1}{2}[H''(\gamma) + 2\delta]t^2$$

and

$$nH(\gamma) + \sqrt{n} Z_{+n} - Y_{+n}t + \tfrac{1}{2}[H''(\gamma) - 2\delta]t^2.$$

Likewise, putting $t = \sqrt{n}(\theta + \gamma)$, almost surely, for all sufficiently large n, for all θ with $|\theta + \gamma| \leq \eta$, upper and lower bounds on $S_n(\theta)$ are, respectively,

$$nH(\gamma) + \sqrt{n} Z_{-n} - Y_{-n}t + \tfrac{1}{2}[H''(\gamma) + 2\delta]t^2$$

and

$$nH(\gamma) + \sqrt{n} Z_{-n} - Y_{-n}t + \tfrac{1}{2}[H''(\gamma) - 2\delta]t^2.$$

PROOF. Only the result at γ needs to be proved, for the situation at $-\gamma$ is symmetric. Recall (2.4a). To estimate $H(\theta)$, use Taylor's theorem. Of course, $H'(\gamma) = 0$ because H has a minimum at γ; and $H''(\gamma) > 0$: see Lemma 2.3. Expanding around γ,

$$H(\gamma) + \tfrac{1}{2}[H''(\gamma) - \delta](\theta - \gamma)^2 \le H(\theta) \le H(\gamma) + \tfrac{1}{2}[H''(\gamma) + \delta](\theta - \gamma)^2.$$

Multiplying by n,

$$nH(\gamma) + \tfrac{1}{2}[H''(\gamma) - \delta]t^2 \le nH(\theta) \le nH(\gamma) + \tfrac{1}{2}[H''(\gamma) + \delta]t^2.$$

Now $G_n(\theta)$ must be estimated. Expanding M' around $u - \gamma$,

$$M'(u - \theta) = M'(u - \gamma) + M''(u - \gamma)(\gamma - \theta) + \tfrac{1}{2}M'''(\zeta_{u\theta})(\gamma - \theta)^2,$$

where $\zeta_{u\theta}$ is between $u - \theta$ and $u - \gamma$. Integrate both sides with respect to $dB_n(u)$, and then integrate by parts as in (2.6):

$$G_n(\theta) = Z_{+n} + (\gamma - \theta)Y_{+n} + \zeta_n(\theta),$$

so

$$\sqrt{n}\,G_n(\theta) = \sqrt{n}\,Z_{+n} - Y_{+n}t + \sqrt{n}\,\zeta_n(\theta),$$

where

$$\zeta_n(\theta) = -\tfrac{1}{2}(\gamma - \theta)^2 \int B_n(u) M'''(\zeta_{u\theta})\,du.$$

In view of Lemma 2.4, almost surely, for all sufficiently large n,

$$|\zeta_n(\theta)| \le A(\log\log n)^{1/2}(\gamma - \theta)^2 \quad \text{for all } \theta.$$

Then

$$\sqrt{n}\,|\zeta_n(\theta)| \le A(\log\log n/n)^{1/2}t^2 \le \tfrac{1}{2}\delta t^2$$

for n large. □

LEMMA 2.11. *Let $\sigma^2 = 1/[H''(\gamma) - 2\delta]$. Almost surely, for all sufficiently large n, the posterior θ-mass in $[\gamma - \eta, \gamma + \eta]$ is bounded above by*

$$C_n^{-1}\pi^{-n}\exp\{-nH(\gamma) - \sqrt{n}\,Z_{+n}\} \cdot \exp\left\{\frac{1}{2}\sigma^2 Y_{+n}^2\right\} \cdot \max_{[\gamma-\eta,\gamma+\eta]} f \cdot \sigma \cdot \sqrt{2\pi} \cdot \frac{1}{\sqrt{n}}.$$

Likewise near $-\gamma$, replacing Z_{+n} and Y_{+n} by Z_{-n} and Y_{-n}.

PROOF. In view of (2.3) and Lemma 2.10, the posterior density for θ is bounded above by

$$\begin{aligned} &C_n^{-1}\pi^{-n} \cdot \exp[-nH(\gamma) - \sqrt{n}\,Z_{+n}] \cdot \exp[\tfrac{1}{2}\sigma^2 Y_{+n}^2]\,\sigma\sqrt{2\pi} \\ &\quad\cdot \max_{[\gamma-\eta,\gamma+\eta]} f \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(t - \mu_n)^2}{2\sigma^2}\right], \end{aligned} \tag{2.7}$$

where $\mu_n = \sigma^2 Y_{+n}$. Integrate (2.7) over $\gamma - \eta \le \theta \le \gamma + \eta$ with respect to $d\theta$, by changing variables, to get

$$\frac{1}{\sqrt{n}}\frac{1}{\sigma\sqrt{2\pi}}\int_{-\eta\sqrt{n}}^{\eta\sqrt{n}}\exp\left[-\frac{(t-\mu_n)^2}{2\sigma^2}\right]dt \le \frac{1}{\sqrt{n}}. \qquad \square$$

LEMMA 2.12. *Let $\tau^2 = 1/[H''(\gamma)+2\delta]$. Let $\mathscr{N}$ be the standard normal distribution on the line. Almost surely, for all sufficiently large n, the posterior θ-mass in $[\gamma-\eta, \gamma+\eta]$ is bounded below by*

$$C_n^{-1}\pi^{-n}\exp\left[-nH(\gamma)-\sqrt{n}Z_{+n}\right]\cdot\exp\left\{\frac{1}{2}\tau^2Y_{+n}^2\right\}\cdot\min_{[\gamma-\eta,\gamma+\eta]} f\cdot\tau$$
$$\cdot\sqrt{2\pi}\cdot\frac{1}{\sqrt{n}}\cdot\mathscr{N}\left[\tau^2Y_{+n}-\eta\sqrt{n},\tau^2Y_{+n}+\eta\sqrt{n}\right].$$

Likewise near $-\gamma$, replacing Z_{+n} and Y_{+n} by Z_{-n} and Y_{-n}.

PROOF. As in Lemma 2.11. $\square$

PROOF OF THEOREM 1. Almost surely, the posterior concentrates near $\pm\gamma$: see Lemma 2.9. For each large n, it will now be argued, there is chance near $\frac{1}{2}$ that the posterior distribution of θ concentrates near $+\gamma$. This follows from Lemmas 2.11 and 2.12. Indeed, the 4-tuple

$$Z_{+n}, Z_{-n}, Y_{+n}, Y_{-n}$$

has a limiting distribution, where the first two coordinates are symmetric and jointly absolutely continuous: see Lemma 2.7. Thus, for K large but fixed and n large,

$$P\{|Y_{\pm n}|<K\} > 1-\delta,$$
$$P\left\{Z_{+n} < Z_{-n}-\frac{1}{K}\right\} > \frac{1}{2}-\delta.$$

If both events occur, the posterior mass near γ overwhelms that near $-\gamma$, in the ratio

$$\text{const.}\exp\left\{\sqrt{n}\left(Z_{-n}-Z_{+n}\right)\right\} > \text{const } e^{\sqrt{n}/K}.$$

Likewise, $-\gamma$ wins with probability near $\frac{1}{2}$.

The final assertions about the a.e. behavior of the lim sup follow from the Hewitt–Savage 0-1 law, because

$$P\{\limsup A_n\} \ge \limsup P\{A_n\} \ge \tfrac{1}{2},$$

where

$$A_n = \left\{\pi_n\left[|\theta-\gamma|<\eta\right] > 1-\delta\right\}$$

is a symmetric function of $X_1,\dots,X_n$. These are i.i.d., so $P\{\limsup A_n\} = 1$. $\square$

PROOF OF THEOREM 2. If $F \sim D(\alpha)$, then $\delta_\theta * F \sim D(\beta)$ where $\beta = \delta_\theta * \alpha$, and α is the Cauchy. Let $G \sim D(\beta)$ and

$$W_i = G\left(\frac{1}{i+1},\frac{1}{i}\right]\Big/G\left(0,\frac{1}{i}\right].$$

The W_i are independent β variables, with parameters p_i and q_i, where

$$p_i = \beta\left(\frac{1}{i+1}, \frac{1}{i}\right] \approx \beta'(0)\frac{1}{i^2},$$

$$q_i = \beta\left(0, \frac{1}{i+1}\right] \approx \beta'(0)\frac{1}{i}.$$

Next, $V_i = p_i(-\log W_i)$ is asymptotically exponential with parameter 1; especially, $c_i e^{-x} \le P\{V_i > x\} < c_i[1 - e^{-x/p_i}]^{-1} e^{-x}$, where $c_i = \Gamma(p_i + q_i)/p_i\Gamma(p_i)\Gamma(q_i) \to 1$ because $\Gamma(\delta) \approx 1/\delta$. Then

$$\frac{1}{n}\sum_{i=1}^{n} V_i \to 1$$

for $D(\beta)$-almost all G. Thus

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{i^2}(-\log W_i) \to \frac{1}{\beta'(0)} \quad \text{a.e.} \tag{2.8}$$

When α is Cauchy, $\beta = \delta_\theta * \alpha$,

$$\beta'(0) = \frac{1}{\pi}\frac{1}{1+\theta^2},$$

so

$$\theta^2 = \frac{1}{\pi\beta'(0)} - 1$$

can be recovered a.e. from $G = \delta_\theta * F$. Likewise, $(1-\theta)^2$ can be recovered, and then θ. The exceptional null set in (2.8) depends on β, i.e., on θ. By Fubini's theorem, (2.8) will hold for a.a. θ and F. □

3. The second construction. We begin by computing the posterior distribution $\bar{\pi}_n$ of θ given the data. The posterior distribution of F given θ and the data will not be needed. A result of the following form was conjectured by Dalal (1979a) who checked special cases for samples of size 2 and 3.

LEMMA 3.1. *With respect to the prior, let θ and F in the model* (1.1) *be independent, θ having a density f and F having the prior distribution $\bar{D}_\alpha$, a symmetrized Dirichlet with parameter measure α. Suppose α is symmetric and absolutely continuous; let $g = \alpha'/\|\alpha\|$, where $\|\alpha\|$ is the mass of α. The data are $X_1, \ldots, X_n$. Let Δ_n be the set where $X_i \ne X_j$ for $i \ne j$ and all θ_{ij} are distinct, with*

$$\theta_{ij} = \tfrac{1}{2}(X_i + X_j).$$

On Δ_n, the posterior $\bar{\pi}_n$ can be written as

$$\bar{\pi}_n = (\pi_{an} + \pi_{dn})/C_n.$$

Again, the constant C_n depends only on the data $X_1, \ldots, X_n$, and normalizes $\bar{\pi}_n$

to have mass 1. *The measure* π_{an} *is absolutely continuous, with density*

$$\|\alpha\| f(\theta) \prod_{k=1}^{n} g(X_k - \theta).$$

The measure π_{dn} *is discrete, with atoms at* θ_{ij} *for* $i \neq j$; *the mass at* θ_{ij} *is*

$$\tfrac{1}{2}[f(\theta_{ij})/g(\delta_{ij})] \prod_{k=1}^{n} g(X_k - \theta_{ij}),$$

where

$$\delta_{ij} = \tfrac{1}{2}(X_i - X_j).$$

PROOF. It is convenient to represent the joint distribution P of $\theta, F, X_1, X_2, \ldots$ as follows. Let β be the result of folding α onto $(0, \infty)$: Formally, β is a finite measure on $(0, \infty)$, and

$$\beta(0, x) = \alpha(-x, x).$$

Let $\theta \sim f$ and $G \sim D(\beta)$, independent. Given θ and G, let $\delta_1, \delta_2, \ldots$ be independent with common distribution G; let $\zeta_1, \zeta_2, \ldots$ be independent, each being ± 1 with probability $\frac{1}{2}$. We present F, ε_i, and X_i as

$$F = \tfrac{1}{2}(G + G^-), \qquad \varepsilon_i = \zeta_i \delta_i, \qquad X_i = \theta + \varepsilon_i.$$

We claim

(3.1) on Δ_n, no three δ's can be equal.

Indeed, suppose by way of contradiction that, e.g., $\delta_1 = \delta_2 = \delta_3$. Then $X_3 = X_1$ or X_2. Likewise,

(3.2) on Δ_n, at most one pair of δ's can be equal.

Indeed, suppose by way of contradiction that, e.g., $\delta_1 = \delta_2$ and $\delta_3 = \delta_4$. Then $\zeta_2 = -\zeta_1$, else $X_2 = X_1$; and $\zeta_4 = -\zeta_3$. So $\theta_{12} = \theta_{34}$, a contradiction.

Let Δ_0 be the event that $\delta_1, \ldots, \delta_n$ are all distinct. For $i \neq j$ among $1, \ldots, n$, let Δ_{ij} be the event that $\delta_i = \delta_j$, but all the other δ_k are distinct from each other and δ_i. Abbreviate $a = \|\alpha\| = \|\beta\|$. Then

$$P(\Delta_0) = a^{n-1}/(a+1)\cdots(a+n-1), \tag{3.3}$$

$$P(\Delta_{ij}) = a^{n-2}/(a+1)\cdots(a+n-1). \tag{3.4}$$

Formula (3.3) is immediate from Lemma 2.1 of Korwar and Hollander (1973); formula (3.4) is a small variation on that lemma, and can be derived from it by a symmetry argument. Or both formulas can be derived by discretization.

As (3.1)–(3.2) show, $\Delta_n \subset \Delta_0 \cup \bigcup_{1 \le i < j \le n} \Delta_{ij}$. Let $X = (X_1, \ldots, X_n)$. Bayes' rule shows that on Δ_n,

$$P\{d\theta|X\} = P\{\Delta_0|X\}P_0\{d\theta|X\} + \sum_{1 \le i < j \le n} P\{\Delta_{ij}|X\}P_{ij}\{d\theta|X\}, \tag{3.5}$$

where P_0 is the law of θ and $X = (X_1, \ldots, X_n)$ given Δ_0; likewise, P_{ij} is the law of θ and X given Δ_{ij}. More specifically, with $g = \alpha'/\|\alpha\|$:

$$\text{(3.6)}\qquad \begin{array}{l}\text{Relative to } P_0\text{: } \theta \text{ has density } f;\\ X_i - \theta \text{ has density } g\text{; all are independent.}\end{array}$$

$$\text{(3.7)}\qquad \begin{array}{l}\text{Relative to } P_{ij}\text{: } \theta \text{ has density } f;\ X_k - \theta \text{ has density } g;\\ \text{all are independent, except } (X_i - \theta) = -(X_j - \theta)\text{, i.e., } \theta = \theta_{ij}.\end{array}$$

This is because, given Δ_{ij}, the δ_k's for $k \neq j$ are independent with common distribution $\beta/\|\beta\|$, while $\delta_j = \delta_i$. This follows from Theorem 2.5 in Korwar and Hollander (1973) by a symmetry argument; or a discretization argument can be used. On Δ_n, $\zeta_j = -\zeta_i$ and $\varepsilon_j = -\varepsilon_i$. In particular, as will be seen in more detail below, P_0 and P_{ij} restricted to Δ_n, all put absolutely continuous distributions on $X_1, \ldots, X_n$, with densities f_0 and f_{ij}, respectively. The 0-term in (3.5) corresponds to π_{an}; the ij-term to the atom at θ_{ij} in π_{dn}.

Note that $P(\Delta_{ij})$ does not depend on i and j. Let

$$\text{(3.8)}\qquad D_n = P(\Delta_0) f_0(X_1, \ldots, X_n) + P(\Delta_{ij}) \sum_{1 \leq i < j \leq n} f_{ij}(X_1, \ldots, X_n).$$

In effect, D_n is the probability density of the data, computed a priori. It will develop that the normalizing C_n in the lemma is $D_n/P(\Delta_{ij}) = \|\alpha\| D_n/P(\Delta_0)$; see (3.3)–(3.4). By Bayes' rule,

$$P\{\Delta_0 | X\} = P(\Delta_0) f_0(X_1, \ldots, X_n)/D_n,$$
$$P\{\Delta_{ij} | X\} = P(\Delta_{ij}) f_{ij}(X_1, \ldots, X_n)/D_n.$$

The 0-term in (3.5) is easily dealt with: By (3.6),

$$f_0(X_1, \ldots, X_n) = \int_{-\infty}^{\infty} f(\theta) \prod_{k=1}^{n} g(X_k - \theta)\, d\theta$$

and

$$P_0\{d\theta | X\} = \frac{f(\theta) \prod_{k=1}^{n} g(X_k - \theta)\, d\theta}{f_0(X_1, \ldots, X_n)}.$$

So

$$P\{\Delta_0 | X\} P_0\{d\theta | X\} = [P(\Delta_0)/D_n]\, f(\theta) \prod_{k=1}^{n} g(X_k - \theta)\, d\theta,$$

as required.

The ij-term in (3.5) is a bit harder. Let $\varepsilon_k = X_k - \theta$ for $k \neq i, j$ and

$$\varepsilon = X_i - \theta = -(X_j - \theta).$$

Thus, $\theta \sim f$, $\varepsilon_k \sim g$, and $\varepsilon \sim g$, all are independent.

$$X_k = \theta + \varepsilon_k, \qquad k \neq i, j,$$
$$X_i = \theta + \varepsilon, \qquad X_j = \theta - \varepsilon.$$

Recall $\theta_{ij} = \frac{1}{2}(X_i + X_j)$ and $\delta_{ij} = \frac{1}{2}(X_i - X_j)$. Now f_{ij}, the density of $X_1, \ldots, X_n$, can be computed by the usual calculus, and is

$$f_{ij}(x_1,\dots,x_n) = \tfrac{1}{2} f(\theta_{ij}) g(\delta_{ij}) \prod_{k\neq i,j} g(x_k - \theta_{ij})$$

$$= \tfrac{1}{2}\left[f(\theta_{ij})/g(\delta_{ij})\right] \prod_{k=1}^{n} g(x_k - \theta_{ij}).$$

Of course,

$$P_{ij}\{d\theta | X\} \quad \text{is point mass at } \theta_{ij}.$$

This completes the proof, up to routine algebra. □

REMARK. Given θ and $X_1,\dots,X_n$, the law of F is $\overline{D}_\beta$, where

$$\beta = \alpha + \sum_{i=1}^{n} \delta_{X_i - \theta}.$$

See Lemma 4.3 of Diaconis and Freedman (1984). This and Lemma 3.1 determine the joint posterior distribution of θ and F.

Specialize now to the case where $g(x) = (1/\pi)1/(1 + x^2)$. Then

$$\pi_{an}(d\theta) = a\pi^{-n} f(\theta)\exp\left[-S_n(\theta)\right] d\theta,$$

where S_n was defined in (2.4), and $a = \|\alpha\|$, and

$$\pi_{dn}(d\theta) = \tfrac{1}{2}\pi^{-n} \sum_{1\leq i<j\leq n} \left[f(\theta_{ij})/g(\delta_{ij})\right]\exp\left[-S_n(\theta_{ij})\right]\delta_{\theta_{ij}}.$$

By construction, the X_j's will be independent with common density h, where h is as in Lemma 2.3, with modes at $-a$, 0, and a. If h vanishes except near its modes, it will be seen that π_{dn} is negligible by comparison with π_{an}, and the argument for Theorem 1 goes through unchanged. If h is strictly positive on the interior of its interval of support, however, π_{dn} dominates, and must be carefully estimated.

LEMMA 3.2. *Construct h as in Lemma 2.3, but require h to vanish except near 0 and $\pm a$. In particular, if X and X' are i.i.d. h, and $\overline{h}$ is the density of $\overline{X} = \tfrac{1}{2}(X + X')$, require $\overline{h}$ to vanish in a neighborhood of $\pm\gamma$. Under these circumstances, $\|\pi_{dn}\|/\|\pi_{an}\| \to 0$ a.e. as $n \to \infty$.*

PROOF. The argument is as in Lemma 2.9. Fix a small open interval around γ in which $\overline{h}$ vanishes. Find δ small but positive such that $H \geq H(\gamma) + \delta$ off that interval. Keep n so large (depending on ω, and possible except for a null set) that $|n^{-1}S_n(\theta) - H(\theta)| < \delta$ for all θ.

Since $f(\theta)\,d\theta$ assigns positive mass to neighborhoods of $\pm\gamma$, and $S_n(\theta) \doteq nH(\theta)$, the mass of π_{an} is at least

$$\text{const.}\,\pi^{-n}\exp\{-n[H(\gamma) + \delta]\}.$$

On the other hand, θ_{ij} is bounded away from $\pm\gamma$, so the mass of π_{dn} is at most

$$\text{const.}\,\pi^{-n}\exp\{-n[H(\gamma) + 2\delta]\}. \qquad \square$$

This completes the proof of Theorem 3 in case the density h of θ_{ij} vanishes near $\pm\gamma$; indeed, π_{an} was estimated in Lemma 2.10.

Turn now to the case where $\bar{h}(\gamma) = \bar{h}(-\gamma) > 0$. (For an existence proof, this case need not be considered—unless, for instance, we want h to be positive on the interior of its support.) By construction, the X_i are independent with a common density h, which is C_∞ and has compact support, as in Lemma 2.3. The factors $f(\theta_{ij})/g(\delta_{ij})$ in π_{dn} are therefore bounded above and below by positive constants, and can be ignored for the rest of the argument. Thus, let

$$\begin{aligned}\hat{\pi}_n(d\theta) &= \pi^{-n} \sum_{1 \le i < j \le n} \exp\left[-S_n(\theta_{ij})\right] \delta_{\theta_{ij}} \\ &= \pi^{-n}\binom{n}{2}\exp\left[-S_n(\theta)\right] q_n(d\theta),\end{aligned} \tag{3.9}$$

where q_n is the empirical distribution of the $\binom{n}{2}$ numbers $\theta_{ij} = 1 \le i < j \le n$, assigning mass $1/\binom{n}{2}$ to each. This $\hat{\pi}_n$ is a good-enough approximation to π_{dn}. Let q be the theoretical distribution of θ_{ij}, so q has density $\bar{h}$, and let

$$\tilde{\pi}_n(d\theta) = \pi^{-n}\binom{n}{2}\exp\left[-S_n(\theta)\right] q(d\theta). \tag{3.10}$$

As will be seen, $\tilde{\pi}_n$ and $\hat{\pi}_n$ are close.

LEMMA 3.3

(a) $\|\pi_{an}\|/\|\tilde{\pi}_n\| \to 0$ *a.e.*

(b) $\tilde{\pi}_n\{|\theta - \gamma| < \eta$ *or* $|\theta + \gamma| < \eta\}/\|\tilde{\pi}_n\| \to 1$ *a.e.*

(c) $\tilde{\pi}_n\{|\theta - \gamma| < \eta\}/\|\tilde{\pi}_n\| \to \frac{1}{2}$ *a.e. and* $\tilde{\pi}_n\{|\theta + \gamma| < \eta\}/\|\tilde{\pi}_n\| \to \frac{1}{2}$ *a.e.*

(d) $\limsup_{n\to\infty}\tilde{\pi}_n\{|\theta - \gamma| < \eta\}/\|\tilde{\pi}_n\| = \limsup_{n\to\infty}\tilde{\pi}_n\{|\theta + \gamma| < \eta\}/\|\tilde{\pi}_n\| = 1$ *a.e.*

PROOF. (a) Holds because of the factor $\binom{n}{2}$.

(b), (c), (d) Can be argued as in Section 2, because $\bar{h}(\pm\gamma) > 0$. □

The next part of the argument is designed to show that $\tilde{\pi}_n - \hat{\pi}_n$ is negligible by comparison with $\tilde{\pi}_n$, near $\pm\gamma$. Recall the notation of Lemma 2.10. Fix $\varepsilon > 0$. Define random variables $R_n^{\pm}$ as follows:

$$\tilde{\pi}_n[\gamma - \eta, \gamma + \eta] = \pi^{-n}\binom{n}{2}\frac{1}{\sqrt{n}}\exp\left[-nH(\gamma) - \sqrt{n}Z_{+n}\right]R_n^{+}, \tag{3.11}$$

$$\tilde{\pi}_n[-\gamma - \eta, -\gamma + \eta] = \pi^{-n}\binom{n}{2}\frac{1}{\sqrt{n}}\exp\left[-nH(\gamma) - \sqrt{n}Z_{-n}\right]R_n^{-}. \tag{3.12}$$

LEMMA 3.4. *The distributions of R_n^+ and R_n^- are tight.*

PROOF. This is argued as in Lemmas 2.10 and 2.11, because $\bar{h}(\pm\gamma) > 0$. □

Since the X_i have common density h, the H in Lemma 2.3 can be written as follows: $H(\theta) = E\{M(X_i - \theta)\}$. Recall that h is C_∞ and compactly supported; and $S_n(\theta) = \sum_{i=1}^n M(X_i - \theta)$.

LEMMA 3.5. *Almost surely, for each j, as $n \to \infty$, $S_n^{(j)}(\theta)/n \to H^{(j)}(\theta)$ uniformly in θ.*

PROOF. Starting with (2.4),

$$\frac{1}{n}S_n^{(j)}(\theta) = \frac{1}{n}\sum_{i=1}^{n} M^{(j)}(X_i - \theta)$$

$$= H^{(j)}(\theta) + \frac{1}{\sqrt{n}}G_n^{(j)}(\theta).$$

By Lemma 2.8,

$$G_n^{(j)}(\theta) = O(\log\log n)^{1/2}. \qquad \square$$

Recall that q_n is the empirical distribution of the θ_{ij}, and q is the common theoretical distribution of each θ_{ij}. Recall from Lemmas 2.3 and 2.3 that h is supported on $[-2a, 2a]$, where $a > 1$. Consider the sequence Q_n of processes $\{n^{1/2}[q_n(t) - q(t)]: \ -2a \le t \le 2a\}$. These processes are nearly uniformly bounded and equicontinuous, in the following sense.

LEMMA 3.6.

(a) *For any $\delta > 0$ there is a finite $B = B_\delta$ such that $P\{|Q_n(t)| \le B\} \ge 1 - \delta$ for all n.*

(b) *For any $\delta > 0$ there is a finite $n_0 = n_{0\delta}$ and a positive δ^* such that for all $n \ge n_0$,*

$$P\{|Q_n(t) - Q_n(s)| \le \delta \text{ for all } s, t \text{ with } |s - t| \le \delta^*\} \ge 1 - \delta.$$

PROOF. Let r_n be the empirical distribution of the $X_i + X_j$, and r the theoretical distribution of each $X_i + X_j$. It is enough to prove the assertions for $n^{1/2}(r_n - r)$. Let ρ_n and ρ be the empirical and theoretical of X_i. Thus, $r = \rho * \rho$. But $r_n \doteq \rho_n * \rho_n$. Indeed, as is easily verified,

$$\sqrt{n}\,\|r_n - \rho_n * \rho_n\|_\infty \to 0,$$

where $\|\ \|_\infty$ is sup norm. Thus it suffices to prove the assertions for

$$\sqrt{n}\,[\rho_n * \rho_n - \rho * \rho].$$

This last is

$$\sqrt{n}\,[\rho_n * \rho_n - \rho * \rho_n] + \sqrt{n}\,[\rho_n * \rho - \rho * \rho].$$

The first term at t is

$$\int \sqrt{n}\,[\rho_n(t-u) - \rho(t-u)]\,\rho_n(du)$$

and the second is

$$\int \sqrt{n}\,[\rho_n(t-u) - \rho(t-u)]\,\rho(du).$$

But with overwhelming probability, as a function of n,

$$\sqrt{n}\,[\rho_n(t-u) - \rho(t-u)]$$

is uniformly bounded and nearly equicontinuous. □

REMARK. $n^{1/2}[q_n(t) - q(t)]$ converges in the sense of the invariance principle to a Gaussian process with mean 0 and the same covariance structure as $\mathscr{H}[\frac{1}{2}(t - X)]$, where $X \sim \mathscr{H}$ the distribution function of h.

Recall $\hat{\pi}_n$ from (3.9) and $\tilde{\pi}_n$ from (3.10).

LEMMA 3.7. $(\hat{\pi}_n - \tilde{\pi}_n)[\gamma - \eta, \gamma + \eta] = o\{\tilde{\pi}_n[\gamma - \eta, \gamma + \eta]\}$ *in probability as* $n \to \infty$, *and likewise at* $-\gamma$.

PROOF. Multiply across by $n^{1/2}\pi^n / \binom{n}{2}$ to get rid of the extraneous normalizing constants. The assertion becomes

$$\text{(3.13)}\quad \sqrt{n}\int_{\gamma-\eta}^{\gamma+\eta} \exp[-S_n(\theta)](q_n - q)(d\theta) = o\{\exp[-nH(\gamma) - \sqrt{n}\,Z_{+n}]\}.$$

Integrate the left side of (3.13) by parts as $T_1 + T_2$, where

$$T_1 = \sqrt{n}\,\exp[-S_n(\theta)]\,\tilde{q}_n(\theta)\big|_{\gamma-\eta}^{\gamma+\eta},$$

$$T_2 = \int_{\gamma-\eta}^{\gamma+\eta} \sqrt{n}\,\tilde{q}_n(\theta)\exp[-S_n(\theta)]\,S_n'(\theta)\,d\theta,$$

$$\tilde{q}_n(\theta) = [q_n(\theta) - q(\theta)] - [q_n(\gamma) - q(\gamma)].$$

Take T_1 first: $\sqrt{n}\,\tilde{q}_n(\gamma \pm \eta)$ is small in probability if η is small, uniformly in n, by Lemma 3.6; and $\exp[-S_n(\gamma \pm \eta)]$ vanishes at a faster exponential rate than $\exp\{-nH(\gamma)\}$. Next take T_2. Again, $\sqrt{n}\,\tilde{q}_n(\theta)$ is small in probability for all θ with $|\theta - \gamma| \le \eta$ if η is small, uniformly in n, by Lemma 3.6. Thus, it suffices to prove

$$\text{(3.14)}\quad \int_{\gamma-\eta}^{\gamma+\eta} \exp[-S_n(\theta)]\,|S_n'(\theta)|\,d\theta = O_p\{\exp[-nH(\gamma) - \sqrt{n}\,Z_{+n}]\}.$$

Now a.e. for large n, S_n is strictly convex in $[\gamma - \eta, \gamma + \eta]$ and has a unique minimum on that interval, at say γ_n; see Lemmas 2.3 and 3.5: S_n''/n is uniformly close to H'' in a neighborhood of γ, where H'' is positive. Thus, S_n' is negative on $[\gamma - \eta, \gamma_n]$ and positive on $[\gamma_n, \gamma + \eta]$. The contribution from, e.g., the first interval is

$$\int_{\gamma-\eta}^{\gamma_n} \exp[-S_n(\theta)]\,[-S_n'(\theta)]\,d\theta = \exp[-S_n(\gamma_n)] - \exp[-S_n(\gamma - \eta)].$$

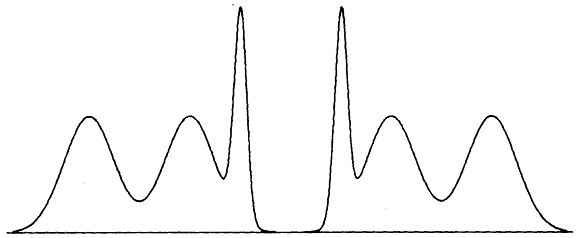

FIG. 2. *The density with six modes.*

Now $\exp[-S_n(\gamma - \eta)]$ vanishes at a faster exponential rate, and we are left with showing that

$$\exp[-S_n(\gamma_n)] = O_p\{\exp[-nH(\gamma) - \sqrt{n}\,Z_{+n}]\}. \tag{3.15}$$

Refer to Lemma 2.10: With $t = \sqrt{n}(\theta - \gamma)$ and $c = [H''(\gamma) - 2\delta]$,

$$\begin{aligned} S_n(t) &\geq nH(\gamma) + \sqrt{n}\,Z_{+n} - Y_{+n}t + \tfrac{1}{2}ct^2 \\ &= nH(\gamma) + \sqrt{n}\,Z_{+n} - \tfrac{1}{2}Y_{+n}^2/c + \tfrac{1}{2}c\left(t - c^{-1}Y_{+n}\right)^2. \end{aligned}$$

Thus

$$S_n(\gamma_n) \geq nH(\gamma) + \sqrt{n}\,Z_{+n} - \tfrac{1}{2}Y_{+n}^2/c$$

But Y_{+n} is tight. □

REMARK. Recall that h has modes at $-a$, 0, and a. But if, e.g., the posterior $\bar{\pi}_n$ concentrates on θ's near $-\gamma$, then it also concentrates on F's whose densities (Figure 2) have six modes, not three, at $-a-\gamma$, $-\gamma$, $\gamma - a$, $a - \gamma$, γ, $a + \gamma$.

4. An approximation theorem for conditional probabilities. At one time, we computed the posterior distributions in Theorems 1 and 3 by discretizing and passing to the limit. Proposition 4.1 gives a rigorous justification for this procedure, which may be useful in other contexts: e.g., see Ferguson (1973, 1974). To motivate the result, consider the following computation of a posterior. Let Λ be the set of probabilities λ on $\mathbb{R}$, and μ a prior probability on Λ. We wish to compute the posterior distribution of λ given a sample of size n from λ, by discretization. Let k be a large positive integer. Let f_k discretize $\mathbb{R}$ in the usual way: $f_k(x)$ is the least $j/k > x$. Let g_k lump Λ to match: $g_k(\lambda)$ assigns mass $\lambda((j-1)/k, j/k]$ to j/k.

To define the posterior carefully, let P_μ be the probability on $\Lambda \times \mathbb{R}^n$ for which

$$P_\mu(A \times B) = \int_A P_\lambda(B)\mu(d\lambda), \tag{4.1}$$

where A is a Borel subset of Λ and B a Borel subset of $\mathbb{R}^n$, while P_λ is the power probability λ^n on $\mathbb{R}^n$. This is the joint distribution of the parameter λ and the data $x \in \mathbb{R}^n$. Let m be the marginal probability on $\mathbb{R}^n$, namely, $m(B) = P_\mu(\Lambda \times B)$ for Borel $B \subset \mathbb{R}^n$. This is the law of the data. The posterior distribution is the "Markov kernel" $\pi(x, d\lambda)$ satisfying

$$(4.2) \qquad P_\mu(A \times B) = \int_B \pi(x, A) m(dx)$$

for Borel $A \subset \Lambda$ and $B \subset \mathbb{R}^n$. A Markov kernel $\pi(x, A)$ is measurable in x for each A, and a probability in A for each x. In sum, π is a regular conditional distribution for the parameter λ given the data x. In this case, posterior distributions are hard to compute, because the family P_λ is not dominated by a σ-finite measure; there is no likelihood function.

One way to compute the posterior is by discretization. Recall f_k and g_k above. Let ϕ_k discretize $\mathbb{R}^n$, by applying f_k separately to each coordinate. Let $\psi_k = g_k$. The law of ψ_k given ϕ_k is easy to compute, because all the probabilities are discrete; call the result $Q_k(x, d\lambda)$. Proposition 4.1 says that if $Q_k(x, d\lambda)$ converges to a Markov kernel $Q(x, d\lambda)$ as $k \to \infty$, then $Q(x, d\lambda)$ is the posterior.

Discretization arguments can be applied with more complex parameter spaces Λ. To prove Lemma 2.1 by discretization, let Λ consist of all pairs $\lambda = (\theta, F)$, where θ is real and F is a probability on the line. Let P_λ on $\mathbb{R}^n$ be F shifted to the right by θ, raised to the nth power. Let $\psi_k(\lambda) = (f_k(\theta), g_k(F))$. Lemma 3.1 can be handled the same way.

For Proposition 4.1, let $(\Omega, \mathscr{F}, P)$ be an arbitrary probability triple. Let $\mathscr{X}$ and $\mathscr{Y}$ be complete separable metric spaces. In the applications above, $\mathscr{X}$ is the data space and $\mathscr{Y}$ the parameter space; $\Omega = \mathscr{Y} \times \mathscr{X}$ and $P = P_\mu$. Let X and Y be Borel mappings from Ω to $\mathscr{X}$ and $\mathscr{Y}$, respectively: In the applications, these are just the projections. Let ϕ_k be a Borel mapping from $\mathscr{X}$ into itself such that $\phi_k(x) \to x$ pointwise as $k \to \infty$. Let ψ_k be a Borel mapping from $\mathscr{Y}$ into itself such that $\psi_k(y) \to y$ uniformly as $k \to \infty$. Let $Q_k(x, dy) = R_k(\phi_k(x), dy)$ be a regular conditional distribution for $\psi_k(Y)$ given $\phi_k(X) = \phi_k(x)$. Suppose $Q(x, dy)$ is a Markov kernel. Suppose $Q_k(x, dy) \to Q(x, dy)$ weak star for each x as $k \to \infty$. For certain discretizations, Pfanzagl (1979) shows this convergence is automatic.

Proposition 4.1. *Q is a regular conditional distribution for Y given X.*

Proof. Let g be bounded continuous on $\mathscr{X}$, while h is bounded and uniformly continuous on $\mathscr{Y}$. Then

$$(4.3) \qquad \begin{aligned} &\int_\Omega g[\phi_k(X)] Q_k(X, h \circ \psi_k)\, dP \\ &\quad = \int_\Omega g[\phi_k(X)] h[\psi_k(Y)]\, dP. \end{aligned}$$

As $k \to \infty$, the right side goes to $\int_\Omega g(X) h(Y)\, dP$; on the left side, $h \circ \psi_k$ can be replaced by h, with only a small error. Then the left side goes to $\int_\Omega g(X) Q(X, h)\, dP$. □

In the applications, it must be shown that $\psi_k(\lambda) \to \lambda$ as $k \to \infty$ uniformly in λ. The ψ_k of interest is the lumping function g_k defined above. The key fact is (4.4) below. Let d be the Lévy distance between probabilities on $\mathbb{R}$: so $d(F, G)$ is the inf of $\varepsilon > 0$ with $F(x) \leq G(x + \varepsilon) + \varepsilon$ and $G(x) \leq F(x + \varepsilon) + \varepsilon$. Then it is easy to show

$$d[F, g_k(F)] \leq 1/k \quad \text{for all } F. \tag{4.4}$$

That is, $g_k(F) \to F$ uniformly in F.

REFERENCES

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

CHUNG, K. L. (1949). An estimate concerning the Kolmogorov limiting distribution. *Trans. Amer. Math. Soc.* **67** 36–50.

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

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

DALAL, S. R. (1980). Bayesian non-parametric theory. In *Bayesian Statistics* (J. M. Bernardo et al., eds.) 521–534. University Press, Valencia, Spain.

DEMPSTER, A. (1976). A subjectivist looks at robustness. *Bull. Inst. Internat. Statist.* **46** (Book 1) 349–374.

DIACONIS, P and FREEDMAN, D. (1982). Bayes rules for location problems. *Statistical Decision Theory and Related Topics III* **1**. Academic, New York.

DIACONIS, P. and FREEDMAN, D. (1983a). Frequentist properties of Bayes rules. *Scientific Inference, Data Analysis and Robustness.* Academic, New York.

DIACONIS, P. and FREEDMAN, D. (1983b). On inconsistent Bayes rules in the discrete case. *Ann. Statist.* **11** 1109–1118.

DIACONIS, P. and FREEDMAN, D. (1984). Partial exchangeability and sufficiency. In *Statistics: Applications and New Directions.* (J. K. Ghosh and J. Roy, eds.) 205–236. Indian Statistical Institute, Calcutta.

DIACONIS, P. and FREEDMAN, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–26.

DOOB, J. L. (1949). Application of the theory of martingales. Colloque Internal du CNRS 22–28.

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.

FERGUSON, T. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2** 615–629.

FRASER, D. A. S. (1976). Necessary analysis and adaptive inferernce. *J. Amer. Statist. Assoc.* **71** 99–113.

FREEDMAN, D. and DIACONIS, P. (1982a). de Finetti's theorem for symmetric location families. *Ann. Statist.* **10** 184–189.

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

HUBER, P. (1984). Finite sample breakdown of *M* and *P* estimators. *Ann. Statist.* **12** 119–126.

JOHNS, M. V. (1979). Robust Pitman-like estimators. In *Robustness in Statistics* (R. Launer and G. Wilkensen, eds.) 49–60. Academic, New York.

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

PFANZAGL, J. (1979). Conditional distributions as derivatives. *Ann. Probab.* **7** 1046–1050.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD CALIFORNIA 94305

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720