BAYESIAN ESTIMATION IN THE
SYMMETRIC LOCATION PROBLEM

by

Hani Doss

FSU Statistics Report M658

June, 1983

The Florida State University
Department of Statistics
Tallahassee, Florida 32306


Key words and phrases. Neutral to the right priors, Dirichlet priors, regular conditional distributions, estimation of location.
Bayesian Estimation in the Symmetric Location Problem

by

Hani Doss
Florida State University

Abstract

Let \( X_i = \theta + \epsilon_i \) for \( i = 1, \ldots, n \), where the \( \epsilon_i \)'s are i.i.d. \( \sim F \) and \( F \) is symmetric about \( 0 \). \( F \) is assumed unknown or only partially known, and the problem is to estimate \( \theta \). Priors are put on the pair \((F, \theta)\). The priors on \( F \) are obtained from Doksum's neutral to the right priors, and include "symmetrized Dirichlet" priors. The marginal posterior distribution of \( \theta \) given \( X_1, \ldots, X_n \) is computed and its general properties studied. It is found that for certain classes of distributions of the \( \epsilon_i \)'s, the posterior distribution of \( \theta \) is for all large \( n \) a point mass at the true value of \( \theta \). If the distribution of the \( \epsilon_i \)'s is not exactly symmetric, the Bayes estimates can behave very poorly.


Key words and phrases. Neutral to the right priors, Dirichlet priors, regular conditional distributions, estimation of location.
1. Introduction.

Let the data be \( X_i = \theta + \epsilon_i, \ i = 1, \ldots, n \), where the \( \epsilon_i \)'s are i.i.d. \( \sim F \) and \( F \) is symmetric about 0. The c.d.f. \( F \) is assumed to be equal or approximately equal to some fixed symmetric \( \alpha_0 \). The problem that motivated this research is described as follows. Find an estimator of \( \theta \) which is efficient if \( F \) is actually equal to \( \alpha_0 \), protects against small deviations from \( \alpha_0 \) and is still consistent if \( F \) is not equal to \( \alpha_0 \) but is symmetric about 0.

In a Bayesian approach one may proceed as follows. Let \( P^S \) equal the set of distributions on \( \mathbb{R} \) which are symmetric about 0, and let \( \Pi = P^S \times \mathbb{R} \). Put a prior on the generic point \((F,\theta) \in \Pi\), compute the marginal posterior distribution of \( \theta \) given \( X_1, \ldots, X_n \), and take the mean of the posterior as an estimate of \( \theta \).

Some justification for the Bayesian approach is given by a general result of Doob (1949). This result states, roughly, that if the prior \( \pi \) is put on the generic point \( \psi \) of the parameter space \( \Pi \), and if \( \{X_i\} \) are i.i.d. \( \sim P_{\psi_0} \) for \( \psi_0 \in \Pi \), then for \([\pi]\) a.e. \( \psi_0 \), the posterior distribution of \( \psi \) given \( X_1, \ldots, X_n \) converges weakly to the point mass at \( \psi_0 \) almost surely, \([P_{\psi_0}]\).

When a prior is put on \( P^S \), then for each \( t \in \mathbb{R} \), the quantity \( F(t) \) is a random variable, and the function \( G \) defined by \( G(t) = EF(t) \) is a distribution function on \( \mathbb{R} \). \( G \) can be viewed as the statistician's "prior guess" at the distribution \( F \). It is important for the statistician to give an indication of the strength of his faith in his prior guess. Confidence in the prior guess is expressed by a prior which concentrates most of its mass in small neighborhoods of \( G \). Having specified a topology on \( P^S \), it is also
important that a prior on \( P^s \) have large support with respect to this topology, so that, it is hoped, the data can change the prior if the initial guess at \( F \) is incorrect.

Let \( P \) denote the set of all probability measures on \( R \). The Dirichlet process priors (see Ferguson 1973, 1974) are probability measures on \( P \) which are parameterized by the set of all finite non-null measures on \( \alpha \). Let \( \alpha \) be a finite non-null measure on \( R \), let \( D_\alpha \) denote the corresponding Dirichlet process prior, and write \( \alpha = \alpha(R)\alpha_0 \) so that \( \alpha_0 \) is a probability measure. It turns out that if \( F(\cdot) \) is distributed according to \( D_\alpha \), then \( E_F(t) = \alpha_0(\{(-\infty, t]\}) \), while the quantity \( \alpha(R) \) expresses the degree of concentration of \( D_\alpha \) around its "center" \( \alpha_0 \). The Dirichlet process priors have the attractive feature that if the support of \( \alpha \) is \( R \) then \( D_\alpha \) is supported by all of \( P \). Another property is that they give mass 1 to the set of discrete distributions.

Dalal (1979a) used the Dirichlet priors to construct "symmetrized Dirichlet" priors on \( P^s \). A sample c.d.f. picked according to this prior is obtained as follows. Let \( \alpha \) be symmetric, and let \( F \sim D_\alpha \); then \( \frac{1}{2} + \frac{1}{2}F(t) - \frac{1}{2}F(-t) \) is a symmetric random c.d.f. that has the symmetrized Dirichlet distribution with parameter \( \alpha \), denoted \( D^s_\alpha \).

In independent work, Dalal (1979b) and Diaconis and Freedman (1983) took a prior \( \nu(d\theta) \) on \( R \) and put the product measure \( D^s_\alpha \times \nu \) on \( \Pi \). They computed the posterior distribution of \( \theta \) given a sample \( X_1, ..., X_n \) with the values \( X_i \neq X_j \) distinct, and computed the Bayes estimate of \( \theta \) under squared error as loss. Diaconis and Freedman were concerned mainly with the question of the consistency of the Bayes rules. They showed that for certain choices of \( \alpha \), the estimates obtained from the prior \( D^s_\alpha \) are inconsistent.
In this paper other classes of priors on \( \mathcal{P}_s \) are considered. These priors arise from the neutral to the right priors introduced by Doksum (1974) and include the priors \( \mathcal{P}_a^s \). The marginal posterior distribution of \( \theta \) given a sample \( X_1, \ldots, X_n \) is computed and the Bayes estimate of \( \theta \) under squared error as loss is obtained.

The qualitative features of the posterior distribution of \( \theta \) can be summarized as follows: assuming that the prior \( \nu \) on \( \theta \) is absolutely continuous, the posterior distribution of \( \theta \) is a convex combination of two probabilities, one discrete and the other absolutely continuous. The discrete component concentrates its mass on the points \( \frac{X_k + X_{k+1}}{2} \) for \( k \neq l \). If there is a point which is the midpoint of at least two distinct pairs of \( X \)'s, then the weight given to the absolutely continuous component is zero, and the posterior is concentrated on the midpoints \( \frac{X_i + X_{i+1}}{2} \) of highest multiplicity.

Excepting trivial cases, neutral to the right process priors concentrate their mass on discrete distributions, and the appearance of the atoms at the points \( \frac{X_k + X_{k+1}}{2} \), \( k \neq l \) is due to the fact that these priors give probability one to the set of symmetric discrete c.d.f.'s. The proposition in section 3 states that if \( F \) is picked according to a symmetrized Dirichlet prior, then for any \( \theta_0 \in \mathcal{R} \), for \( [F_{\theta_0}] \) almost every sample sequence \( X_1, X_2, \ldots \), the posterior distribution of \( \theta \) is for all large \( n \) a point mass at \( \theta_0 \) (\( F_{\theta} \) is defined by \( F_{\theta}(x) = F(x-\theta) \)). Other classes of symmetric c.d.f.'s \( F \) with this property are also identified. Thus, the resulting estimators can be extremely good if the data comes from an exactly symmetric discrete c.d.f.. It turns out that the slightest deviation from these assumptions can cause a disaster.
The theorem in section 3 gives the posterior distribution of $\theta$ in the general case, while the corollary in that section specializes this to the prior $\mathcal{D}_a^s$. The behavior of the posterior distribution is also discussed in that section. Section 4 gives the proofs of the theorems and corollary of section 3. In Doss (1983a,b,c) other priors on $F$ are considered. A random c.d.f. picked from one of these priors has median equal to zero, but need not be symmetric. The resulting Bayes estimates of $\theta$ protect against asymmetric contamination.

2. Preliminaries.

Let $F^s$ denote the $\sigma$-field on $p^s$ generated by the topology of weak convergence, and let $\mu$ be a probability measure on $(p^s, F^s)$. Let $\nu$ be a probability measure on $(\mathcal{R}, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel sets of $\mathcal{R}$. The product measure $\mu \times \nu$ on $\Pi$ with the product $\sigma$-field induces random variables $X_1, \ldots, X_n$ and defines a probability measure $P$ on $\mathcal{R}^n \times \Pi$ as follows:

$$P\{X_1 \leq x_1; \ldots; X_n \leq x_n; F \in C; \theta \in A\}$$

$$= \int \int \int_{\Pi} F(x_i - \theta) \mu(dF) \nu(d\theta)$$

(2.1)

where $x_1, \ldots, x_n \in \mathcal{R}$, $C \in F^s$ and $A \in \mathcal{B}$. It is desired to obtain a "regular" conditional distribution of $\theta$ given $X_1, \ldots, X_n$. This is done in section 3 for a certain class of priors $\mu$. The class of measures $\mu$ on $p^s$ is described below.

If $\mu$ is a measure on $p^s$, then the distribution function $F$ is a random function. It will be very helpful to view $\mu$ as a stochastic process.
\{F(t); t \in \mathbb{R}\}. Briefly, any stochastic process \{F(t); t \in \mathbb{R}\} which satisfies

(i) \( F \) is nondecreasing, a.s.

(ii) \( \lim_{t \to +\infty} F(t) = 1 \), a.s. \hspace{1cm} (2.2)

(iii) \( \lim_{t \to s^+} F(t) = F(s) \) for each \( s \in \mathbb{R} \), a.s.

(iv) \( F(-t) = 1 - F(t^-) \) for each \( t \geq 0 \), a.s.

induces a measure \( \mu \) on \( (\mathbb{P}^S, \mathbb{F}^S) \). Conversely, any measure \( \mu \) on \( (\mathbb{P}^S, \mathbb{F}^S) \)
induces a separable stochastic process \( \{F(t); t \in \mathbb{R}\} \) satisfying (i)-(iv) of
(2.2). Details can be extracted from Doksum (1974, pp. 189, 190).

Consider a probability measure concentrated on \([0, \infty)\), and let \( F \) denote
both the measure and its distribution function. (Throughout this work,
probability measures are identified with their cumulative distribution
functions, and the same symbol is used to denote both the measure and its
distribution function whenever convenient.) A symmetrized version \( F_\mathbb{S} \) of
\( F \) can be defined as follows:

\[
F_\mathbb{S}(t) = \frac{1}{2} + \frac{1}{2} F(t) - \frac{1}{2} F(-t^-). \hspace{1cm} (2.3)
\]

The correspondence above is 1-1. Thus, the construction of random
distribution functions on \( \mathbb{R} \) symmetric about 0 is equivalent to the construction
of random distribution functions on \([0, \infty)\). With this in mind, the class of
random distribution functions on \([0, \infty)\) to be considered is the class of
neutral to the right distribution functions introduced by Doksum (1974).
Only random distribution functions continuous in probability will be con-
sidered. (Recall that a stochastic process \( \{X(t)\} \) is continuous in proba-
bility if \( s \to t \) implies \( X(s) \to X(t) \) in probability.) Before proceeding,
a review of neutral to the right distributions will be given.
Let $F$ be a distribution function on $[0, \infty)$ with $F(0) = 0$. Define the function $Y(\cdot)$ on $[0, \infty]$ by

$$F(t) = 1 - e^{-Y(t)}.$$ \hfill (2.4)

Note that $Y(\cdot)$ is nondecreasing, right continuous, and that $Y(0) = 0$ and $Y(\infty) = \infty$. It is possible that $Y(t) = \infty$ for some finite $t$. Equation (2.4) gives a 1-1 correspondence between distribution functions $F$ on $[0, \infty)$ with $F(0) = 0$ and functions $Y(\cdot)$ on $[0, \infty]$ that are nondecreasing, right continuous and satisfy $Y(0) = 0$ and $Y(\infty) = \infty$. Thus, random distribution functions $F$ on $[0, \infty)$ with $F(0) = 0$ correspond to nondecreasing, right continuous stochastic processes such that $Y(0) = 0$ a.s., and $Y(\infty) = \infty$ a.s.. If the process \{$Y(t); t \in [0, \infty]$\} has independent increments, then the random distribution functions $F$ defined by (2.4) is called neutral to the right by Doksum (1974). Roughly speaking, a random distribution function $F$ on $[0, \infty)$ is neutral to the right if for every $t_1$ and $t_2$ with $0 \leq t_1 < t_2$,

$$\frac{1-F(t_2)}{1-F(t_1)} \quad \text{and} \quad \{F(t); t \leq t_1\}$$ \hfill (2.5)

are independent; that is, if the proportion of mass $F$ assigns to the sub-interval $(t_2, \infty)$ of the interval $(t_1, \infty)$ is independent of what $F$ does to the left of $t_1$.

A description of the process \{$Y(t); t \in [0, \infty]$\} will now be given. Attention will be restricted to the processes which are continuous in probability. The Lévy theory can be applied to such processes.

It is well-known that if \{$Y(t); t \in [0, \infty]$\} is an increasing independent increments process continuous in probability, then
\[ -\log E e^{-\lambda Y(t)} = m(t)\lambda + \int_0^\infty (1-e^{-\lambda x})dN_t(x) \quad \text{for} \ \lambda \geq 0, \quad (2.6) \]

where

(i) \( m(\cdot) \) is a continuous nondecreasing function with \( m(0) = 0 \).

(ii) \( \{N_t; t \geq 0\} \) is a nondecreasing, continuous family of measures on \((0, \infty)\), that is, for each Borel set \( A \subset (0, \infty) \), \( N_t(A) \) is a continuous, nondecreasing function of \( t \).

(iii) For each \( t \geq 0 \),

\[ \int_0^1 x \, dN_t(x) < \infty \quad \text{and} \quad N_t[1, \infty) < \infty. \]

The function \( m \) corresponds to the deterministic part of the process \( Y(t) \).

If \( m(\cdot) = 0 \), then \( Y(t) \) increases only by jumps, whose locations and sizes are random. For a rigorous account of the theory of nondecreasing processes with independent increments the reader is referred to Ito and McKean (1965, pp. 146-149); see also Ferguson and Klass (1972).

For \( F(t) \) defined by (2.4), the posterior distribution of the process \( \{F(t); t \in [0, \infty)\} \) given \( X_1, \ldots, X_n \) has been described by Doksum (1974). Ferguson (1974) in his review paper gives a nice summary of the results of Doksum. If \( F \) is neutral to the right, then the posterior distribution of \( F \) given a sample \( X_1, \ldots, X_n \) is again neutral to the right, and has jumps at the observed values \( X_1, \ldots, X_n \).

However, in the model used in this work, the observed random variables \( X_1, \ldots, X_n \) form a sample from a process \( F \) that is a mixture of neutral to the right processes. Moreover, \( F \) is viewed as a nuisance parameter. Consequently, the posterior distribution of \( F \) given a random sample will not be of direct concern.
Examples.

Two main classes of examples are now given.

Example 2.1. The Dirichlet process.

Let \( \alpha \) be a finite nonnull measure on \( R \). The random distribution \( F \) has the Dirichlet distribution with parameter \( \alpha \), denoted \( \mathcal{D}_\alpha \), if for every finite measurable partition \( \{ A_1, \ldots, A_k \} \) of [0, \( \infty \)) the random vector \((F(A_1), \ldots, F(A_k))\) has the Dirichlet distribution with parameter vector \((\alpha(A_1), \ldots, \alpha(A_k))\).

Write \( \alpha = \alpha(R)\alpha_0 \), so that \( \alpha_0 \) is a probability measure. From the definition of the Dirichlet process, it follows that

\[
E F(t) = \alpha_0(t). \tag{2.7}
\]

It can then be said that "\( \mathcal{D}_\alpha \) is centered at \( \alpha_0 \)". The parameter \( \alpha(R) \) indicates the concentration of \( \mathcal{D}_\alpha \) around \( \alpha_0 \). For example, it is easy to show that if \( \alpha_0 \) is fixed, then as \( \alpha(R) \to \infty \), \( \mathcal{D}_\alpha \) converges to the point mass at \( \alpha_0 \) in the weak topology.

For a review of the Dirichlet priors see Ferguson (1974). We will state here only the properties of the Dirichlet prior needed in the present paper.

Sethuraman and Tiwari (1982) showed that if \( \alpha_0 \) is fixed and \( \alpha(\omega) \to 0 \), then \( \mathcal{D}_\alpha \) converges weakly to the random distribution which is degenerate at a point whose distribution is \( \alpha_0 \).

Ferguson (1974) showed that if \( F \) is distributed according to \( \mathcal{D}_\alpha \), then \( F \) is neutral to the right, and

\[
-\log E e^{-\lambda Y(t)} = \log \frac{\Gamma(\alpha(\omega)+\lambda)}{\Gamma(\alpha(\omega))} + \log \frac{\Gamma(\alpha(\omega)-\alpha(t))}{\Gamma(\alpha(\omega)-\alpha(t)+\lambda)} \tag{2.8}
\]
for \( \lambda \geq 0 \) and \( t \geq 0 \). In equation (2.6),

\[
m(\cdot) = 0,
\]

(2.9)

and

\[
dN_t(x) = \frac{e^{-(\alpha(\infty) - \alpha(t))x} \cdot (1 - e^{-\alpha(t)x})}{x(1 - e^{-x})} \, dx.
\]

(2.10)

If \( \alpha \) is continuous, then \( F \) is continuous in probability, and the family

\( \{N_t; t \geq 0\} \)

of measures is continuous, i.e., for each Borel set \( A \subset (0, \infty) \),

\( N_t(A) \) is continuous.

Now, let \( \alpha \) be a finite symmetric measure on \( \mathbb{R} \) and let \( \beta \) be the measure

concentrated on \( [0, \infty) \) defined by

\[
\beta(A) = 2\alpha(A \cap (0, \infty)) + \alpha(A \cap \{0\}).
\]

(2.11)

Let \( F \sim D_\alpha \), and let \( D^\alpha \) denote the distribution of the random c.d.f. obtained

from \( F \) through (2.3). (Note: the construction of the symmetrized Dirichlet

prior \( D^\alpha \) given here is not identical with the one given by Dalal (1979a,b).

The difference is, however, nominal.)

**Example 2.2.** Homogeneous neutral to the right processes.

Let \( Y \) be any nonnegative infinitely divisible random variable. Let

\( \psi(\cdot) \) be defined by

\[
\psi(\lambda) = -\log E \, e^{-\lambda Y} \text{ for } \lambda \geq 0.
\]

(2.12)

It is well-known (see Ito and McKean, 1965, pp. 146-149) that

\[
\psi(\lambda) = p\lambda + \int_0^\infty (1 - e^{-\lambda x}) dN(x)
\]

(2.13)
where \( p > 0 \) and \( N \) is a measure on \((0, \infty)\) that satisfies

\[
\int_0^1 x \, N(dx) < \infty \quad \text{and} \quad N(1, \infty) < \infty.
\]  

(2.14)

Assume \( Y \) is "normalized" so that \( \psi(1) = 1 \). Let \( \beta \) be a distribution function on \([0, \infty)\) and let \( \{Y(t); \ t \in [0, \infty)\} \) be the independent increments process with log Laplace transform given by

\[
\log E \, e^{-\lambda Y(t)} = \log(1 - \beta(t)) \, \psi(\lambda) \quad \text{for} \; t \geq 0, \lambda \geq 0.
\]

(2.15)

Then \( E \, F(t) = \beta(t) \) for \( t \geq 0 \). \( F \) is called a homogeneous neutral to the right process.

3. **The Posterior Distribution of \( \theta \).**

Recall that the measure \( \mu \) on \( P^S \) is obtained from a nondecreasing, independent increments process \( \{Y(t); \ t \in [0, \infty)\} \) by (2.4) and (2.3), and that \( P \) is the probability measure defined on \( \mathbb{R}^{n} \times \mathbb{N} \) by equation (2.1). In this section, the posterior distribution of \( \theta \) given a sample \( X_1, \ldots, X_n \) from \( F_\theta \) is given.

For ease of reference, all notation and assumptions used in the rest of the paper are collected here. In what follows, \( X \) denotes the random vector \( (X_1, \ldots, X_n) \) and \( x \) denotes the vector \( (x_1, \ldots, x_n) \).

**Notation:**

N1. \( \psi_t(\lambda) = -\log E \, e^{-\lambda Y(t)} \) for \( t \geq 0 \) and \( \lambda \geq 0 \).

N2. \( m \) is equal to the number of distinct values of the sequence \( \{x_1, \ldots, x_n\} \).

\( x(1) < x(2) < \ldots < x(m) \) denote the ordered values of the sequence \( \{x_i; 1 \leq i \leq n\} \), and \( n_i \) denotes the multiplicity of \( x(i) \).
N3. For each pair \( \{i,j\} \) with \( 1 \leq i, j \leq n \), \( x_i \neq x_j \), \( \#\{i,j\} \) denotes the number of distinct pairs \( \{x_k, x_\ell\} \) with \( 1 \leq k, \ell \leq n \), \( x_k \neq x_\ell \), such that 
\[
\frac{x_k + x_\ell}{2} = \frac{x_i + x_j}{2}.
\] (The pairs \( \{x_k, x_\ell\} \) and \( \{x_s, x_r\} \) are distinct if the sets \( \{x_k, x_\ell\} \) and \( \{x_s, x_r\} \) are distinct.) The number \( \#\{i,j\} \) is called the multiplicity of the point \( \frac{x_i + x_j}{2} \).

N4. \( \#(x) = \max \#\{i,j\} \).

N5. Let \( \theta = \frac{x_k + x_\ell}{2} \) whenever \( x_k \neq x_\ell \). The sequence \( \{|x_i - \theta|; i = 1, \ldots, n\} \) has \( m \) distinct values. Let \( t_1(\theta) < \ldots < t_m(\theta) \) denote its ordered distinct values, and \( \varepsilon_1(\theta), \ldots, \varepsilon_m(\theta) \) denote the multiplicities of \( t_1(\theta), \ldots, t_m(\theta) \), respectively. For notational brevity, \( t_i = t_i(\theta) \), and \( \varepsilon_i = \varepsilon_i(\theta) \), for \( i = 1, \ldots, m \).

N6. Let \( \theta = \frac{x_k + x_\ell}{2} \) for some \( x_k, x_\ell \), with \( x_k \neq x_\ell \). The sequence 
\( \{|x_i - \theta|; i = 1, \ldots, n\} \) has \( m - \#(k, \ell) \) distinct values. Let 
\( v_1(k, \ell) < \ldots < v_{m - \#(k, \ell)}(k, \ell) \) denote its ordered distinct values, and 
\( \delta_1(k, \ell), \ldots, \delta_{m - \#(k, \ell)}(k, \ell) \) their respective multiplicities. The 
\( v_i(k, \ell) \) and \( \delta_i(k, \ell) \) will be abbreviated \( v_i \) and \( \delta_i \) whenever convenient.

N7. \( v_x \) or \( v(d\theta|x) \) denote a regular conditional distribution of \( \theta \) given \( X = x \).

Assumptions:

A1. For \( \lambda \geq 0 \), the function \( \psi(\cdot)(\lambda) \) is continuously differentiable on \([0, \infty)\).

Let \( \dot{\psi}_t(\lambda) \) denote \( \frac{\partial}{\partial t} \psi_t(\lambda) \).

A2. \( v \) is absolutely continuous with respect to Lebesgue measure, with a continuous density \( v' \).
The usual method of computing the posterior distribution, i.e., "the posterior is proportional to the likelihood times the prior", is inapplicable here, since there is no likelihood: there is no $\sigma$-finite measure dominating the family $\{F_\theta; (F, \theta) \in \Pi\}$. Consequently, the posterior distribution of $\theta$ will have to be obtained in a different way.

What is desired is a regular conditional distribution of $\theta$ given $X$. Recall that a regular conditional distribution of $\theta$ given $X$ is a function $\nu(\cdot)(\cdot)$ defined on $\mathbb{R}^n \times \mathcal{B}$ satisfying

(i) For each $x \in \mathbb{R}^n$, $\nu_x$ is a probability measure on $\mathcal{B}$.
(ii) For each $A \in \mathcal{B}$, $\nu_x(\cdot)(A)$ is a measurable function of $x$.
(iii) For each $A \in \mathcal{B}$, $\nu_x(A)$ is a version of $P(\theta \in A | X=x)$; i.e., for each linear Borel set $A$ and $n$-dimensional Borel set $B$,

$$\int_B \nu_x(A) dP = P(\theta \in A; X \in B).$$

One of the properties of regular conditional distributions is that the conditional expectation may be obtained by taking an ordinary expectation relative to the conditional probability distribution (Chow and Teicher, p. 211).

Let $\rho_{A.C.}^x$ be the absolutely continuous measure defined by

$$\frac{d\rho_{A.C.}^x}{dv}(\theta) = \left\{ \begin{array}{ll}
\prod_{i=1}^m \left[ \exp \psi_{t_i} \left( n - \sum_{\ell=1}^i \varepsilon_{\ell} \right) - \psi_{t_i} \left( n - \sum_{\ell=1}^{i-1} \varepsilon_{\ell} \right) \right] \\
\prod_{i=1}^m \sum_{r=0}^{\varepsilon_i} (-1)^{r+1} \binom{\varepsilon_i}{r} \psi_{t_i} \left( n - \sum_{\ell=1}^i \varepsilon_{\ell} + r \right) \end{array} \right\}$$

(3.2)
Consider the points \( \frac{x_k + x_\ell}{2} \) for \( x_k \neq x_\ell \), such that \#(k, \ell) = \#(x). Let \( \rho_x^D \) be the discrete measure concentrated on those points, and defined by

\[
\rho_x^D\left(\frac{x_k + x_\ell}{2}\right) = \left\{\begin{array}{l}
\prod_{i=1}^{m-\#(x)} \exp \left[ \psi_{\nu_i} \left( n - \frac{i}{L} \sum_{s=1}^{L} \delta_s \right) - \psi_{\nu_i} \left( n - \frac{i-1}{L} \sum_{s=1}^{L} \delta_s \right) \right]
\end{array}\right.
\]

\[
\left\{\begin{array}{l}
\prod_{i=1}^{m-\#(x)} \sum_{r=0}^{\delta_i} \left(-1\right)^{r+1} \left(\delta_i\right)_r \psi_{\nu_i} \left( n - \frac{i}{L} \sum_{s=1}^{L} \delta_s \right) \right)^{-\frac{1}{2}} \left(\frac{x_k + x_\ell}{2}\right).
\]

(3.3)

THEOREM. There exists a regular conditional distribution of \( \theta \) given \( X = x \), which will be denoted by \( \nu_x \). Assume A1 and A2. Let \( \rho_x^{A.C.} \) and \( \rho_x^D \) be defined by (3.2) and (3.3), respectively.

If \( \#(x) = 1 \), then \( \nu_x \) is the sum of two measures, an absolutely continuous measure \( \nu_x^{A.C.} \), and a discrete measure \( \nu_x^D \). \( \nu_x^{A.C.} \) is absolutely continuous with respect to \( \nu \), with

\[
\frac{d\nu_x^{A.C.}}{d\nu}(\theta) = c(x) \frac{d\rho_x^{A.C.}}{d\nu}(\theta).
\]

(3.4)

\( \nu_x^D \) is discrete, and

\[
\nu_x^D = \frac{c(x)}{2} \rho_x^D.
\]

(3.5)

c(x) is a normalizing constant.

If \( \#(x) > 1 \), then except in the "pathological cases" discussed in Remark (3.4), \( \nu_x \) is discrete and is equal to \( c(x)\rho_x^D \), where \( c(x) \) is a normalizing constant.
The following corollary gives the posterior distribution of \( \theta \) in the case where \( F \sim \mathcal{D}_{\alpha}^S \). To satisfy A1, \( \alpha_0 \) must be assumed absolutely continuous, with a continuous density \( \alpha_0^* \).

COROLLARY. Suppose that \( F \sim \mathcal{D}_{\alpha}^S \), and assume A2.

If \( \#(X) = 1 \), the posterior distribution of \( \theta \) given \( X \) is the sum of an absolutely continuous measure, denoted \( \nu^{A,C}(d\theta | X) \), and a discrete measure, denoted \( \nu^D(d\theta | X) \). The absolutely continuous measure is given by

\[
\nu^{A,C}(d\theta | X) = c(X) \left( \prod_{i=1}^{n} \alpha_0^*(X_i - \theta) \right) \nu(d\theta), \tag{3.6}
\]

where the * indicates that the product is over distinct \( X_i \)'s only. The discrete measure is concentrated on the points \( \frac{X_k + X_\ell}{2} \), \( X_k \neq X_\ell \), and is given by

\[
\nu^D \left( \frac{X_k + X_\ell}{2} | X \right) = c(X) \frac{w_{k,\ell}}{4 \alpha(\infty)} \left( \prod_{X_i = X_k, X_\ell} \alpha_0^* \left( X_i - \frac{X_k + X_\ell}{2} \right) \right) \alpha_0^* \left( \frac{X_k - X_\ell}{2} \right) \nu^* \left( \frac{X_k + X_\ell}{2} \right) \tag{3.7}
\]

where

\[
w_{k,\ell} = \frac{(n_k + n_\ell - 1)!}{(n_k - 1)! (n_\ell - 1)!}, \tag{3.8}
\]

The product in (3.7) is over distinct \( X_i \)'s such that \( X_i \neq X_k \) and \( X_i \neq X_\ell \).

c(X) is a normalizing constant.

If \( \#(X) > 1 \), and \( \alpha^*(t) \) is positive for all \( t \) (see Remark (3.4)), then \( \nu(d\theta | X) \) is purely atomic, and is concentrated on the midpoints \( \frac{X_k + X_\ell}{2} \) of highest multiplicity. Let \( \frac{X_k + X_\ell}{2} \) be a midpoint of highest multiplicity, i.e., \( \#(k,\ell) = \#(X) \). Let \( (X_{k_1}^1, X_{\ell_1}^1), \ldots, (X_{k_{\#(X)}}^{\#(X)}, X_{\ell_{\#(X)}}^{\#(X)} \) be the \( \#(X) \) distinct
pairs of points having the midpoint \( \frac{X_k + X_{\ell}}{2} \). Then,

\[
\nu\left(\frac{X_k + X_{\ell}}{2} \mid X\right) = c(X)w_{k, \ell} \prod_{r=1}^{\#(X)} \alpha_0^\ast\left(\frac{X_k - X_{\ell}}{2}\right) \alpha_0^\ast\left(\frac{X_k + X_{\ell}}{2}\right),
\]

where

\[
w_{k, \ell} = (\prod(n_i - 1)!) (n_{k_1} + n_{k_1} + \ldots + n_{k_\#(X)} + n_{\ell_\#(X)}) - 1!
\]  

The product in (3.10) is over all \( i, 1 \leq i \leq m \), such that \( i \notin \{k_1, \ell_1, \ldots, k_\#(X), \ell_\#(X)\} \). The first product in (3.9) is over distinct \( X_i \) such that \( X_i \neq X_{k_1}, X_{\ell_1}, \ldots, X_{k_\#(X)}, X_{\ell_\#(X)} \). \( c(X) \) is a normalizing constant.

The behavior of the posterior distribution of \( \Theta \).

We now give some general remarks concerning \( \nu(d\Theta \mid X) \) and then specialize to the case of the priors \( D_\alpha^S \). Throughout the rest of the paper assumptions A1 and A2 are in force. Note that A1 is a smoothness assumption on the expected value of the random c.d.f. \( F(\cdot) \), which for the Dirichlet priors is a smoothness assumption on \( \alpha_0 \). A2 is a smoothness assumption on the prior on \( \Theta \).

**Remark 3.1.** In a Bayesian setup, if there exists a \( \sigma \)-finite measure dominating all the probability measures indexed by the parameter space, then the posterior is absolutely continuous with respect to the prior. In case B, the posterior is clearly not absolutely continuous with respect to the prior. In fact, if \( \#(X) > 1 \), the prior and the posterior are mutually singular.
Remark 3.2. There is an interpretation for the representation of $\nu_x$ as the sum of an absolutely continuous measure and a discrete measure in case B. The appearance of the atoms will be explained heuristically.

Let $F$ be chosen according to $\mu$ and let $Y_1, \ldots, Y_n$ be i.i.d. $\sim F$; let $\theta$, distributed according to $\nu$, be independent of $Y_1, \ldots, Y_n$, and let $X_i = \theta + Y_i$, $i = 1, \ldots, n$. Excepting the case where $\mu$ assigns all its mass to one distribution function, the random distribution function has atoms, and these are symmetrically located about 0. Thus, for $k \neq \ell$, $Y_k = -Y_\ell$ has positive probability, and implies that $\theta = \frac{X_k + X_\ell}{2}$. Therefore, conditional on the event that there exist $k, \ell \in \{1, \ldots, n\}$, $k \neq \ell$, such that $Y_k = -Y_\ell$, the true value of $\theta$ is to be found among the points $\frac{X_k + X_\ell}{2}$, for $k \neq \ell$. Thus, $\nu^{D_x}_x \nu^{D_x}_x(R)$ is the conditional distribution of $\theta$ given $X_1, \ldots, X_n$ and given that there exist $k, \ell \in \{1, \ldots, n\}$, $k \neq \ell$, such that $Y_k = -Y_\ell$.

Furthermore, it is intuitively obvious that if $\#(k, \ell) > \#(i, j)$, then the point $\frac{X_k + X_\ell}{2}$ is "infinitely more likely" to be equal to $\theta$ than is the point $\frac{X_1 + X_j}{2}$. This also explains why $\nu_x$ is purely atomic when there exists a point which is the midpoint of more than one pair of observations.

Remark 3.3. Let $X_1, X_2, \ldots$ be i.i.d. $H_\theta$ where $H_\theta(x) = H(x - \theta)$, $H$ is a fixed symmetric c.d.f. and $\theta$ is fixed. The letter $H$ rather than $F$ is used to emphasize that $H$ is not necessarily to be viewed as being picked from a distribution $P^S$. If $H$ is continuous, it is clear that (with probability one) $\nu(d\theta|X)$ has an absolutely continuous component for all $n$. If $H$ is discrete, it is clear that (with probability one) for large $n$, $\nu(d\theta|X)$ is discrete. (Actually, the existence of at least two atoms to the right of 0 is needed to insure that $\nu(d\theta|X)$ is discrete for large $n$.)
We now turn to the priors $D^s_\alpha$.

Recall that $\alpha^0_0$ is the measure $\alpha$ normalized to be a probability measure. If the measure $\mu$ on $P$ assigns mass 1 to $\alpha_0^0$, then the posterior distribution of $\theta$ is equal to

$$
c(X) \prod_{i=1}^{n} \alpha^0_0(X_i - \theta) \nu(d\theta).
$$

(3.11)

(Now and henceforth, $c(X)$ is a normalizing constant, not necessarily the same in different appearances.)

Let $\alpha^0_0$ be fixed and let $X_1, \ldots, X_n$ be fixed. We have seen that if $\#(X) > 1$, then $\nu(d\theta|X)$ is discrete and does not depend on $\alpha^0(\omega)$. Suppose that $\#(X) = 1$ and that the $X_i$'s are all distinct, as is the case when their common distribution is continuous. Note that the measure $\nu_X^{A.C.}$ normalized to be a probability is precisely (3.11). As $\alpha^0(\omega) \to \infty$ the weight attached to the discrete component vanishes, and $\nu(d\theta|X)$ converges setwise to (3.11).

If $\nu$ is formally replaced by Lebesgue measure, the mean of (3.11) is the Pitman estimate of location. This is the best (under squared error as loss) location invariant estimate of $\theta$ when $X_1, \ldots, X_n$ are i.i.d. $\nu \alpha_0^0(\cdot - \theta)$. Johns (1979) has investigated the robustness of these Pitman estimators. He used positive functions $\alpha_0^0$ which are not necessarily densities, and showed that certain choices of $\alpha_0^0$ yield estimators which have bounded influence curves and have high efficiencies over a wide variety of symmetric densities.

If we now let $\alpha(\omega) \to 0$, we see that the weight attached to the absolutely continuous component vanishes. One way of understanding this is to relate it to the fact that $D^s_\alpha$ converges weakly to the random measure $\frac{\delta_a + \delta_a}{2}$ (where $\alpha$ denotes the point mass at $a$), where $a$ has the distribution $\alpha_0^0$. 
In an unpublished paper, Diaconis and Freedman (1980) investigated some of the frequentist properties of the Bayes estimator \( \hat{\theta}(X) \) of \( \theta \) when the prior on \( F \) is the symmetrized Dirichlet prior \( D^s_{\alpha} \), and the prior on \( \theta \) is an arbitrary smooth prior with all of \( R \) as its support. They were interested in determining conditions on \( \alpha_0 \) such that

(i) \( \hat{\theta}(X) \) is a consistent estimate of \( \theta_0 \) when \( X_1, X_2, \ldots \) are i.i.d. from a continuous distribution \( H \), symmetric about \( \theta_0 \).

(ii) \( \hat{\theta}(X) \) is "robust" in the sense of being resistant to outliers.

With respect to consistency, Diaconis and Freedman obtained the following results. For a given \( \alpha_0 \), if \( -\log \alpha_0^*(\cdot) \) is convex, the estimator \( \hat{\theta}(X) \) is a consistent estimator of \( \theta_0 \) if the data is i.i.d. from a distribution function \( H \) that is continuous, symmetric about \( \theta_0 \) and satisfies

\[
\int |\log \alpha_0^*(x-\theta_0)| dH(x) < \infty. \tag{3.12}
\]

If \( -\log \alpha_0^*(\cdot) \) is not convex, there exists a distribution \( H \) symmetric about \( 0 \), infinitely differentiable, supported by a finite interval and such that if \( X_1, X_2, \ldots \) are i.i.d. \( \sim H \), then the Bayes estimator \( \hat{\theta}(X) \) oscillates indefinitely between two wrong values as \( n \) tends to infinity. The condition that \( -\log \alpha_0^*(\cdot) \) is convex implies roughly that the tails of \( \alpha_0 \) are not larger than the tails of the exponential distribution.

Concerning robustness, Diaconis and Freedman showed roughly that the estimator is robust if and only if the tails of \( \alpha_0 \) are not smaller than the exponential tails. As a particular case, if the Dirichlet prior is centered at a normal distribution, the estimator is consistent but not robust. If it is centered at the exponential distribution, the estimator is both consistent and robust. Here, consistency refers to the class of symmetric continuous distributions.
The following proposition states that the estimator performs very well if \( H \) is picked from a symmetrized Dirichlet prior: it quickly becomes and remains a point mass at the true value.

**Proposition.** Let \( H \) be picked from \( D^S_\alpha \), and let \( X_1, X_2, \ldots \) be i.i.d. \( \sim H_{\theta_0} \), where \( \theta_0 \) is fixed. Assume that \( \nu \) has an everywhere positive density. Then, for \([D^S_\alpha] \) a.e. \( H \), there exists \( a(H) > 0 \), a constant depending on \( H \), such that for all large \( n \),

\[
1 - P\{ \text{for all } k \geq n, \nu(d\theta | X_1, \ldots, X_k) = \delta_{\theta_0} \} \leq e^{-n a(H)}.
\]

**Proof.**

Let \( \beta \) be the measure concentrated on \((0, \infty)\) obtained from \( \alpha \) by

\[
\beta(A) = 2\alpha(A \cap (0, \infty)),
\]

and let \( \beta_0 = \beta/\beta(\infty) \). The Dirichlet process (on \((0, \infty)\)) with parameter \( \beta \) can be represented as

\[
F(t) = \sum_{i=1}^{\infty} P_i \delta_{L_i}(t), \tag{3.13}
\]

where the \( P_i \)'s are random variables satisfying \( P_1 > P_2 > \ldots > 0 \), \( \sum_{i=1}^{\infty} P_i = 1 \), and \( L_1, L_2, \ldots \) are i.i.d. \( \sim \beta_0 \), independent of the \( P_i \)'s; see Ferguson (1973). Thus, \( H(t) = \frac{1}{2} + \frac{1}{2} F(-t) - \frac{1}{2} F(-t^-) \) has the symmetrized Dirichlet distribution \( D^S_\alpha \), and the set of its atoms is \( \Lambda = \{\ldots, -L_2, -L_1, L_1, L_2, \ldots\} \).

Without loss of generality, assume that \( \theta_0 = 0 \). The idea is to show that \( 0 \) is the only point which is the common midpoint of two distinct pairs of points chosen from \( \Lambda \). That is, we want to show that if \( Z_1, Z_2, Z_3, Z_4 \in \Lambda \), then \( P\{Z_1 + Z_2 = Z_3 + Z_4 = 0\} = 0 \). Note that if \( Z_{n_1} + Z_{n_2} \neq 0 \) for \( n_1, n_2 \in \{1, 2, 3, 4\} \), then \( Z_{n_1} \) and \( Z_{n_2} \) are independent. Assume without loss of generality that \( Z_3 < Z_1 < Z_2 < Z_4 \). The event \( \{Z_1 + Z_2 = Z_3 + Z_4 = 0\} \) can
be written as the disjoint union of the three events
\[ \{Z_1 + Z_2 = Z_3 + Z_4 \neq 0; Z_1 = -Z_4; Z_3 = -Z_2\}, \{Z_1 + Z_2 = Z_3 + Z_4 \neq 0; Z_1 \neq Z_4; Z_3 = -Z_2\}, \]
and \[ \{Z_1 + Z_2 = Z_3 + Z_4 \neq 0; Z_1 = -Z_4; Z_3 = -Z_2\}. \] Using the continuity of \( \alpha_0 \), it is easy to see that each of the three events has probability 0.

Now, \( H \) has atoms of size \( \frac{P_i}{2} \) at \( \pm L_i \), for \( i = 1, 2 \). Let \( K \) be the first \( j \) for which all four points, \( \pm L_1, \pm L_2 \) have been observed in the sample \( X_1, X_2, \ldots, X_j \). Clearly, for all \( k \geq K \), \( \nu(\delta \mid X_1, \ldots, X_k) = \delta_0 \). The probability that at time \( n \) one of the points \( \pm L_1, \pm L_2 \) has not been observed is asymptotic to \( 2(1 - P_2/2)^n \). \( \square \)

The proposition is still true if \( H \) is picked from \( \mathcal{D}^S_\gamma \) where \( \gamma \) is any continuous finite measure.

If \( H \) is a distribution symmetric about 0 and with a finite number of atoms, the same result is true. The proof is similar. More precisely, suppose \( H = \sum_{i=1}^{k} b_i (\delta_{a_i} + \delta_{-a_i}) \). Let \( X_1, X_2, \ldots \) be i.i.d. \( \sim H \). As soon as all the values \( \pm a_i, i = 1, \ldots, k \) are observed, 0 is a midpoint of multiplicity \( k \), and all other midpoints have multiplicity less than \( k \). The posterior then becomes and remains a point mass at 0. The probability that one of the values \( \pm a_1, \ldots, \pm a_k \) is not observed by time \( n \) goes down exponentially with \( n \). In fact, an actual (exponential) bound is easy to obtain in terms of the size of the smallest atom.

It would be of interest to determine necessary and sufficient conditions on discrete distributions \( H \) that are symmetric about \( \theta_0 \) for the posterior to be (with probability one) eventually a point mass at \( \theta_0 \).

As indicated by the proposition and the remark following it, the estimator can be highly efficient if the underlying distribution is symmetric
and discrete. However, a very slight deviation from these conditions can cause the estimator to efficiently estimate a wrong value. Consider the following example. Let $\varepsilon > 0$, $a \neq 0$, and $H_c$ be any continuous distribution symmetric about 0. Define

$$H_\varepsilon = (1-\varepsilon)H_c + \varepsilon \left( \sum_{a=2}^{\infty} \left( \frac{\delta_{a-2} + \delta_{a-1} + \delta_{a+1} + \delta_{a+2}}{4 \cdot 2} \right) \right), \tag{3.14}$$

and let $X_1, X_2, \ldots$ be i.i.d. $\sim H_\varepsilon$. For $\varepsilon > 0$ fixed, $\nu(d\theta|X_1, \ldots, X_n)$ becomes and remains a point mass at a exponentially fast.

In addition to this distributional instability, the function $(X_1, \ldots, X_n) \mapsto \nu(d\theta|X_1, \ldots, X_n)$ is highly discontinuous: a few observations can be moved slightly and completely change the posterior distribution of $\theta$.

We close this section with a few technical remarks.

**Remark 3.4.** The quantity $\rho^D_{x}(\frac{X_k + X_{k+1}}{2})$ may turn out to be equal to 0. This can occur, for example, if $\nu'\left(\frac{X_k + X_{k+1}}{2}\right) = 0$, or if the quantity in the second set of braces in (3.3) is equal to 0. For the Dirichlet process, the latter possibility will not occur if $\alpha'(t)$ is always positive. If $\rho^D_{x}(\frac{X_k + X_{k+1}}{2}) = 0$ for all points, then $\nu_x$ need not be purely atomic.

**Remark 3.5.** In section V of his paper, Dalal (1979b) erroneously states that the weight attached to the absolutely continuous component of $\nu(d\theta|X)$ (in the case he is considering) is positive. That is not necessarily correct if $n \geq 4$.

**Remark 3.6.** The theorem is proved under Assumption A1 which, for the Dirichlet process, is equivalent to the condition that $\alpha_0$ has a continuous density. Condition A1 can be weakened to the condition that for all $\lambda \geq 0$, 
\( \psi_t(\lambda) \) is absolutely continuous. For the Dirichlet process, this is the condition that \( \sigma_0 \) is absolutely continuous. However, this slight gain in generality is outweighed by the complications caused in the proof.

4. Proofs of the Theorem and Corollary.

Proof of the theorem.

The general method of proof will first be described. To facilitate the implementation of the method, a lemma is stated and proved. Details and the main part of the proof then follow.

Since the event \( X = x \) does not, in general, have positive probability, it is impossible to define, for \( x \) fixed, \( \nu_x \) on \( \mathcal{B} \) by

\[
\nu_x(A) = \frac{P(\theta \in A; X=x)}{P(X=x)} \quad \text{for } A \in \mathcal{B}. \tag{4.1}
\]

Consider instead, for \( x \) fixed, \( \eta > 0 \), the measure \( \nu_x^\eta \) defined on \( \mathcal{B} \) by

\[
\nu_x^\eta(A) = \frac{P(\theta \in A; X_i \in (x_i - \eta/2, x_i + \eta/2), i = 1, \ldots, n)}{P(X_i \in (x_i - \eta/2, x_i + \eta/2), i = 1, \ldots, n)}. \tag{4.2}
\]

According to a theorem of Pfanzagl (1979), the weak limit as \( \eta \to 0 \) of \( \nu_x^\eta \) exists for [P] a.e. \( x \); furthermore, if \( \nu_x \) denotes the weak limit of \( \nu_x^\eta \), then \( \nu_x \) is a regular conditional distribution of \( \theta \) given \( X \). (This will be explained more precisely later.) Therefore, what is needed is to find the weak limit of \( \nu_x^\eta \) as \( \eta \to 0 \). From (2.1) it follows that for all \( A \in \mathcal{B} \),
\[ \nu_x^n(A) = \frac{\int \int_{-\infty}^{\infty} \prod_{i=1}^{n} \left[ F(x_i - \theta + \eta/2) - F(x_i - \theta - \eta/2) \right] \nu(d\theta) \nu(d\eta)}{\int \int_{-\infty}^{\infty} \prod_{i=1}^{n} \left[ F(x_i - \theta + \eta/2) - F(x_i - \theta - \eta/2) \right] \nu(d\theta) \nu(d\eta)}. \] (4.3)

Defining \( f_x^n(\theta) \) by
\[ f_x^n(\theta) = E \prod_{i=1}^{n} \left[ F(x_i - \theta + \eta/2) - F(x_i - \theta - \eta/2) \right], \] (4.4)

(4.3) may be rewritten as
\[ \nu_x^n(A) = \frac{\int \int f_x^n(\theta) \nu(d\theta)}{\int \int f_x^n(\theta) \nu(d\theta)} \text{ for } A \in \mathcal{B}. \] (4.5)

Thus, it is necessary to find \( f_x^n(\theta) \). Under mild regularity assumptions, the following estimates will be obtained.

There exists a nonnegative \( \nu \)-integrable function \( g_x(\beta) \) defined on \( \mathbb{R} \), and there exist nonnegative numbers \( a_x(k, \ell) \) defined for \( 1 \leq k, \ell \leq n \), with \( x_k \neq x_\ell \), such that
\[
\begin{align*}
\eta \cdot (g_x(\theta) + o(1)) & \text{ as } \eta \to 0 \text{ for } \eta > 0, \text{ and } \theta \text{ such that } \\
\quad & \text{for all } x_k, x_\ell, \text{ with } x_k \neq x_\ell, \\
\quad & \theta \notin \left[ \frac{x_k + x_\ell}{2} - \frac{\eta}{2}, \frac{x_k + x_\ell}{2} + \frac{\eta}{2} \right] \\
\eta^{-\#(k, \ell)} (a_x(k, \ell) + o(1)) & \text{ as } \eta \to 0 \text{ for } \eta > 0 \text{ and } \theta \text{ such that } \\
\quad & \text{that there exist } x_k, x_\ell, \text{ with } x_k \neq x_\ell, \\
\quad & \theta \in \left[ \frac{x_k + x_\ell}{2} - \frac{\eta}{2}, \frac{x_k + x_\ell}{2} + \frac{\eta}{2} \right].
\end{align*}
\] (4.6)

with the \( o \) terms uniform for \( \theta \) bounded.
Let \( V^n_x(k, \ell) = \left( \frac{x_k + x_\ell}{2} - \frac{n}{2}, \frac{x_k + x_\ell}{2} + \frac{n}{2} \right) \). From equations (4.5) and (4.6) we have

\[
v^n_x(A) = \frac{b^n_x(A)}{b^n_x(R)}, \tag{4.7}
\]

where for a Borel set \( B \),

\[
b^n_x(B) = \int_{B \cup \bigcup B} (g_x(\theta) + o(1)) v^*(\theta) \, d\theta + \Sigma^B (a^0_x(k, \ell) + o(1)) n^{-\# \{k, \ell\}} v^n_x(k, \ell). \tag{4.8}
\]

Now and hereafter, the symbols \( B \) and \( \Sigma^B \) signify that the union and sum are taken over all \( k, \ell \) with \( 1 \leq k, \ell \leq n \), \( x_k \neq x_\ell \) and such that \( \frac{x_k + x_\ell}{2} \in B \). We have

\[
\lim_{n \to 0} \frac{1}{n} v^n_x(k, \ell) = v^*\left( \frac{x_k + x_\ell}{2} \right). \tag{4.9}
\]

That the \( o \) terms in (4.8) are uniform only for bounded sets poses a problem that will be dealt with later. Assume for now that the \( o \) terms in (4.8) are uniform.

Suppose that \( \#(x) = 1 \). Then, from (4.7), (4.8) and (4.9), it is apparent that

\[
\lim_{n \to 0} v^n_x(A) = \frac{\int g_x(\theta) v^*(\theta) \, d\theta + \Sigma^A a^0_x(k, \ell) v^*\left( \frac{x_k + x_\ell}{2} \right)}{\int_{-\infty}^{\infty} g_x(\theta) v^*(\theta) \, d\theta + \Sigma^R a^0_x(k, \ell) v^*\left( \frac{x_k + x_\ell}{2} \right)} \tag{4.10}
\]

for \( A \in B \). This yields a representation of \( v_x \) when \( \#(x) = 1 \). In this case,
\( \nu_x \) is the sum of two measures, one absolutely continuous and the other discrete. The absolutely continuous measure has density proportional to \( g_x(\theta) \nu^\ast(\theta) \), and the discrete measure concentrates its mass on the points \( \frac{x_k + x_\ell}{2} \), for \( 1 \leq k, \ell \leq n, x_k \neq x_\ell \).

Suppose that \( \#(x) > 1 \). Then, using similar reasoning, it is seen that

\[
\lim_{n \to 0} \nu^n_x(A) = \frac{\sum_{\{k,\ell\} = \#(x)} a_{x}(k,\ell) \nu^\ast\left(\frac{x_k + x_\ell}{2}\right)}{\sum_{\{k,\ell\} = \#(x)} a_{x}(k,\ell) \nu^\ast\left(\frac{x_k + x_\ell}{2}\right)}.
\]

Thus, if \( \#(x) > 1 \), \( \nu_x \) is a discrete measure concentrated on the midpoints \( \frac{x_k + x_\ell}{2} \) of highest multiplicity. (Of course, we are tacitly assuming that not all the numbers \( a_{x}(k,\ell) \nu^\ast\left(\frac{x_k + x_\ell}{2}\right) \) are 0.)

Let \( \{Y(t); t \in [0,\infty)\} \) be a nondecreasing independent increments process continuous in probability, with \( Y(0) = 0 \), a.s., and \( Y(\infty) = \infty \), a.s.; let \( \psi_t \) be the negative log Laplace transform of \( Y(t) \), and let \( F(t) = 1 - e^{-Y(t)} \).

**Lemma.** Assume A1. Let \( t_1, \ldots, t_m \) be positive numbers such that \( t_{i-1} + n/2 < t_i - n/2 \) for \( i = 1, \ldots, m \). (\( t_0 \) is defined to be 0). Let \( n_1, \ldots, n_m \) be a sequence of positive integers, and let \( n = \sum_{i=1}^{m} n_i \). Then

\[
E \prod_{i=1}^{m} [F(t_i + n/2) - F(t_i - n/2)]^{n_i} = \eta \exp \left[ \psi_{t_1} \left( n - \sum_{\ell=1}^{i} n_\ell \right) - \psi_{t_i} \left( n - \sum_{\ell=1}^{i-1} n_\ell \right) \right]
\]

\[
\left\{ \begin{array}{l}
\prod_{i=1}^{m} \sum_{r=0}^{n_i} (-1)^{r+1} \binom{n_i}{r} \psi_{t_1} \left( n - \sum_{\ell=1}^{i} n_\ell + r \right) + o(n_m),
\end{array} \right.
\]

with the \( o \) uniform for \( t_1', \ldots, t_m' \) ranging over a bounded set.
Proof of Lemma.

Writing

\[
\prod_{i=1}^{m} [F(t_i+n/2) - F(t_i-n/2)]^{n_i} = \\
\prod_{i=1}^{m} \left\{ 1 - \exp[Y(t_i-n/2) - Y(t_i+n/2)] \right\}^{n_i} \\
\exp \left[ \left( n - \sum_{\ell=1}^{i} n_{\ell} \right) [Y(t_i-n/2) - Y(t_i+n/2)] \right] \\
\prod_{i=1}^{m} \exp \left[ \left( n - \sum_{\ell=1}^{i-1} n_{\ell} \right) [Y(t_{i-1}+n/2) - Y(t_i-n/2)] \right]
\]

(4.13)

allows us to use the fact that \( \{Y(t); t \in [0, \omega]\} \) has independent increments in taking expectations.

We have

\[
E \prod_{i=1}^{m} [F(t_i+n/2) - F(t_i-n/2)]^{n_i} = \\
\prod_{i=1}^{m} \left( \frac{n_i}{n} \right)^{n_i} \exp \left[ \psi_{t_i-n/2} - \psi_{t_i+n/2} \left( n - \sum_{\ell=1}^{i} n_{\ell} \right) \right] \\
\prod_{i=1}^{m} \exp \left[ \left( \psi_{t_{i-1}+n/2} - \psi_{t_i-n/2} \right) \left( n - \sum_{\ell=1}^{i-1} n_{\ell} \right) \right].
\]

(4.14)

Consider first the product in the first set of parentheses in the right side of (4.14). The \( i \)th term in the product is equal to

\[
\sum_{r=0}^{n_i} \frac{n_i}{r!} (-1)^{r+1} \psi_{t_i}^{n_i} \left( n - \sum_{\ell=1}^{i} n_{\ell} + r \right) + o(n).
\]

(4.15)

It is easy to see that the \( o(n) \) terms are uniform for \( t_i \) bounded.
The product in the second set of parentheses in the right side of (4.14) is equal to

\[
\prod_{i=1}^{m} \exp \left[ (\psi_{t_{i-1}} - \psi_{t_i}) \left( n - \sum_{\ell=1}^{i-1} n_{\ell} \right) \right] + o(1),
\]

(4.16)

where, again, the \( o \) terms are uniform for \( t_i \) bounded. This completes the proof of the lemma.

Let \( \theta \) satisfy (4.17) and (4.18) below:

\[
\left| \frac{x_i^* + x_\ell^*}{2} - \theta \right| > n/2 \quad \text{for all } x_k \neq x_\ell,
\]

(4.17)

\[
\left| x_i - \theta \right| > n/2 \quad \text{for all } i.
\]

(4.18)

By (4.18), we can write

\[
\mathbb{E}_{x}^{n}(\theta) = \mathbb{E}_{x}^{n} \left[ F\left( \left| x_i - \theta \right| + n/2 \right) - F\left( \left| x_i - \theta \right| - n/2 \right) \right].
\]

(4.19)

By (4.17), the sequence \( \{\left| x_i - \theta \right|; i = 1, \ldots, n\} \) has exactly \( m \) elements. By (4.17) and (4.18), \( t_{i-1} + n/2 < t_i - n/2 \) for \( i = 1, \ldots, n \). Therefore, the lemma is applicable, and we get, for \( \theta \) satisfying (4.17) and (4.18),

\[
\mathbb{E}_{x}^{n}(\theta) = m \prod_{i=1}^{m} \exp \left[ \left( \psi_{t_i} \left( n - \sum_{\ell=1}^{i-1} \epsilon_{\ell} \right) - \psi_{t_i} \left( n - \sum_{\ell=1}^{i-1} \epsilon_{\ell} \right) \right) \right]
\]

\[
\left\{ \prod_{i=1}^{m} \sum_{r=0}^{\epsilon_i} (-1)^{r+1} \left( \frac{\epsilon_i}{r} \right) \psi_{t_i} \left( n - \sum_{\ell=1}^{i} \epsilon_{\ell} + r \right) \right\} + o(n^m)
\]

(4.20)

with the \( o \) term uniform for \( \theta \) bounded.
Now let $\theta$, still satisfying (4.18), be such that 
$$\left| \theta - \frac{x_k + x_\ell}{2} \right| < \eta/2$$
for some $x_k, x_\ell$, with $x_k \neq x_\ell$. Assume temporarily that the vector $x$ satisfies
$$x_i = \frac{x_k + x_\ell}{2} \quad \text{for} \quad 1 \leq i, \quad k, \ell \leq n. \quad (4.21)$$

We consider first the case $\#(k, \ell) = 1$. $\mathbb{F}_x^\eta(\theta)$ is still expressed by (4.19).

The sequence $\{|x_i - \theta|; i = 1, \ldots, n\}$ has exactly $m$ distinct elements (unless $\theta = \frac{x_k + x_\ell}{2}$, which has $\nu$-probability 0). The condition $t_i - \eta/2 < t_i - \eta/2$ for all $i$ is no longer satisfied, and so the lemma cannot be applied directly.

Equation (4.19) will be rewritten so that the lemma can be used. Let
$$\theta = \frac{x_k + x_\ell}{2} + \omega \eta/2, \quad \text{where} \quad \omega \in (-1, 1).$$
Assume without loss of generality that $\omega > 0$. Let

$$T_0 = \prod_{x_i = x_k + x_\ell} \left[ F(|x_i - \theta| + \eta/2) - F(|x_i - \theta| - \eta/2) \right]$$
$$T_1 = F\left( \left| \frac{x_k - x_\ell}{2} \right| + \omega \eta/2 + \eta/2 \right) - F\left( \left| \frac{x_k - x_\ell}{2} \right| - \omega \eta/2 + \eta/2 \right)$$
$$T_2 = F\left( \left| \frac{x_k - x_\ell}{2} \right| - \omega \eta/2 + \eta/2 \right) - F\left( \left| \frac{x_k - x_\ell}{2} \right| + \omega \eta/2 - \eta/2 \right) \quad (4.22)$$
$$T_3 = F\left( \left| \frac{x_k - x_\ell}{2} \right| - \omega \eta/2 + \eta/2 \right) - F\left( \left| \frac{x_k - x_\ell}{2} \right| + \omega \eta/2 - \eta/2 \right)$$
$$T_4 = F\left( \left| \frac{x_k - x_\ell}{2} \right| + \omega \eta/2 - \eta/2 \right) - F\left( \left| \frac{x_k - x_\ell}{2} \right| - \omega \eta/2 - \eta/2 \right).$$

Note that $T_2 = T_3$. We have

$$\mathbb{F}_x^\eta(\theta) = E \ T_0 (T_1 + T_2)^{n_k} (T_3 + T_4)^{n_\ell}. \quad (4.23)$$

In equation (4.23), $n_k$ and $n_\ell$ may need to be interchanged, depending on whether $x_k - x_\ell > 0$ or not. It will soon be clear that this doesn't make
any difference. Expanding \( (T_1 + T_2)^{n_k} \) and \( (T_3 + T_4)^{n_\ell} \) and multiplying, we get

\[
f_x^n(\theta) = E T_0 T_{2}^{n_k + n_\ell} \quad \text{+ other terms.} \tag{4.24}
\]

Now, by the lemma,

\[
E T_0 T_{2}^{n_k + n_\ell} = \eta^{m-1} \frac{1 - |\omega|}{2^n} (a_x(k, \ell) + o(1)) \tag{4.25}
\]

uniformly in \( \omega \), where \( a_x(k, \ell) \) is the expression on the right side of (3.3). The other terms in (4.24) are all \( o(\eta^m) \) uniformly in \( \omega \). Thus, for the case \( \#(k, \ell) = 1 \), \( f_x^n(\theta) \) is equal to the right side of (4.25). The \( \frac{1}{2} \) in (3.5) is caused by the factor \( 1 - |\omega| \) in (4.25).

Suppose now that \( \#(k, \ell) > 1 \). Let \( (x_{k_1, 1}^1, x_{k_1, 1}^2), \ldots, (x_{k_{\#(k, \ell)}, 1}^{1}, x_{k_{\#(k, \ell)}, 1}^{2}) \) be the \( \#(k, \ell) \) distinct pairs of points having the midpoint \( \frac{x_k + x_\ell}{2} \), and let \( M(k, \ell) \) be the set containing the points in these pairs. Let

\[
T_0 = \prod_{x_i \in M(k, \ell)} \left[ F(|x_i - \theta| + n/2) - F(|x_i - \theta| - n/2) \right]
\]

\[
T_{j1} = F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| + \omega \ n/2 + n/2 \right) - F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| - \omega \ n/2 + n/2 \right)
\]

\[
T_{j2} = F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| - \omega \ n/2 + n/2 \right) - F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| + \omega \ n/2 - n/2 \right) \tag{4.26}
\]

\[
T_{j3} = F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| - \omega \ n/2 + n/2 \right) - F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| + \omega \ n/2 - n/2 \right)
\]

\[
T_{j4} = F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| + \omega \ n/2 - n/2 \right) - F\left( \left| \frac{x_{k,j}^1 - x_{k,j}^2}{2} \right| - \omega \ n/2 - n/2 \right)
\]

for \( j = 1, \ldots, \#(k, \ell) \).
Then
\[
\mathcal{E}_x^n(\theta) = E_{T_0} \prod_{j=1}^{\#(k, \ell)} \left[ \frac{n_{k_j}}{(T_{j1} + T_{j2})} \frac{n_{\ell_j}}{(T_{j3} + T_{j4})} \right].
\] (4.27)

As before, in (4.27), the \(n_{k_j}\) and \(n_{\ell_j}\) may need to be interchanged, but this does not affect the final answer. The lemma is applicable, and
\[
\mathcal{E}_x^n(\theta) = E_{T_0} \prod_{j=1}^{\#(k, \ell)} \frac{n_{k_j} + n_{\ell_j}}{T_{j2}} + \text{other terms}. \] (4.28)

\[ E_{T_0} \prod_{j=1}^{\#(k, \ell)} \frac{n_{k_j} + n_{\ell_j}}{T_{j2}} \] is equal to \(\eta^{m - \#(k, \ell)}\) times a constant, plus \(o(\eta^{m - \#(k, \ell)})\) uniformly in \(\omega\). The other terms are all \(o(\eta^{m - \#(k, \ell) + 1})\) uniformly in \(\omega\).

Assumption (4.21) insured that \(v_i(k, \ell) > 0\) for all \(i\). A little thought shows that the assumption can be removed.

Let \(\tau_x\) be the probability measure defined by
\[
\tau_x = \begin{cases} 
(c(x)(\rho_x^{A.C.} + \rho_x^D)/2) & \text{if } \#(x) = 1 \\
(c(x)\rho_x^D) & \text{if } \#(x) > 1 
\end{cases}
\] (4.29)

where \(\rho_x^{A.C.}\) and \(\rho_x^D\) are defined by (3.2) and (3.3), respectively. We will show that the family of measures \(\{\tau_x; x \in \mathbb{R}^n\}\) forms a regular conditional probability distribution of \(\theta\) given \(X\). Note that in the definition of a regular conditional probability distribution (3.1), the measure \(v_x\) needs to be defined only for \([P]\) a.e. \(x\).
As a consequence of the theorem in Pfanzagl (1979), we have

(i) For [P] a.e. \( x \), there exists a measure \( \nu_x \) such that the net of measures \( \{ \nu_x^n; n > 0 \} \) defined by (4.2) converges weakly to \( \nu_x \).

(ii) The family \( \nu_x \) above is a regular conditional probability distribution of \( \theta \) given \( X \).

Let \( N \in \mathcal{B}^n \) be a set of probability 0, as guaranteed by (i), with the property that for all \( x \notin N \), there exists a probability measure \( \nu_x \) such that \( \nu_x^n \) converges weakly to \( \nu_x \). Let \( x \notin N \) be fixed. Supposing first that \( \#(x) = 1 \), let \( g_x(\cdot) \) be the function given by the product of the quantities inside the braces in (4.20), so that (4.20) is rewritten compactly as

\[
f^n_x(\theta) = n^{-m} \frac{1}{2^n} (g_x(\theta) + o(1)).
\]  

An easy argument shows that

\[
\int_{-\infty}^{\infty} g_x(\theta) \nu^-(\theta) d\theta < \infty. \tag{4.31}
\]

Let \( a, b \in \mathbb{R} \) be continuity points of \( \nu_x \), let \( \epsilon > 0 \), and let \( K \) be so large that

\[
\begin{align*}
& \int_a^b g_x(\theta) \nu^-(\theta) d\theta + \Sigma^{(a,b)} a_x(k, \epsilon)/2 \\
& \leq \int_{-K}^K g_x(\theta) \nu^-(\theta) d\theta + \Sigma^R a_x(k, \epsilon)/2 < \tau_x(a, b) + \epsilon. \tag{4.32}
\end{align*}
\]

We then have
\[ v_x((a,b)) = \lim_{\eta \to 0} \nu_x^\eta((a,b)) \]

\[ \leq \lim_{\eta \to 0} \frac{1}{K} \int_{-K}^{K} (g_x(\theta) + o(1)) \nu_x(\theta) \, d\theta + \Sigma(a, b)(a_x(k, \theta)/2 + o(1)) \]

\[ = \lim_{\eta \to 0} \frac{1}{K} \int_{-K}^{K} (g_x(\theta) + o(1)) \nu_x(\theta) \, d\theta + \Sigma^R(a_x(k, \theta)/2 + o(1)) \]  

(4.33)

Since the "little oh" terms are uniform for \( \theta \) bounded, we have by (4.32) and the fact that \( \varepsilon > 0 \) was arbitrary

\[ v_x(a, b) \leq \tau_x(a, b) \]  

(4.34)

for all \( a, b \) continuity points of \( v_x \). An easy argument shows that equality must hold in (4.34), and that is enough to show that \( \tau_x = v_x \). The case \( \#(x) > 1 \) is simpler and is omitted. Thus, \( \tau_x = v_x \) for \( [P] \) a.e. \( x \), and we conclude that \( \tau_x \) is a regular conditional distribution of \( \theta \) given \( X \). \( \square \)

**Proof of Corollary.**

We apply the theorem and (2.8). Consider first the case \( \#(X) = 1 \), and consider the absolutely continuous component of the posterior. We will show that the right side of (3.2) is equal to

\[ (n^2 2a^{-\frac{d}{2}}(X_1 - \theta)) \left( \prod_{i=1}^{m} \frac{(n_i!)}{(n_i - 1)!} \right) \left( \prod_{j=1}^{n} \frac{1}{\alpha(\infty) + j - 1} \right). \]  

(4.35)

Note that the sequences \( \{t_i\}_{i=1}^{m} \) and \( \{e_i\}_{i=1}^{m} \) are permutations of the sequences \( \{X_1 - \theta\}_{i=1}^{m} \) and \( \{n_i \}_{i=1}^{m} \), respectively. Using the formula \( \Gamma(x+1) = x\Gamma(x) \) repeatedly, we see that the quantity in the first set of braces in (3.2) is equal to
\[
\frac{m \sum_{i=1}^{n_i-1} \left( \beta(\infty) - \beta(t_i) + n - \sum_{\ell=1}^{i} \varepsilon_{\ell} + r \right)}{\Pi \Pi \prod_{r=0}^{i} \left( \beta(\infty) + n - \sum_{\ell=1}^{i} \varepsilon_{\ell} + r \right)}.
\]

(4.36)

Let \( i \in \{1, \ldots, m\} \) be fixed, and consider the \( i^{th} \) factor in the product in the second set of braces in (3.2). This factor is equal to

\[
\sum_{r=0}^{\varepsilon_i} (-1)^{r} \binom{\varepsilon_i}{r} \frac{d}{dt} \left[ \log \left( \beta(\infty) - \beta(t_i) \right) 
+ n - \sum_{\ell=1}^{i} \varepsilon_{\ell} + r - 1 \right] + \ldots + \log(\beta(\infty) - \beta(t_i)) \right].
\]

(4.37)

Now since \( \sum_{r=0}^{\varepsilon_i} (-1)^{r} \binom{\varepsilon_i}{r} = 0 \), (4.37) is equal to

\[
\sum_{r=0}^{\varepsilon_i} (-1)^{r} \binom{\varepsilon_i}{r} \frac{d}{dt} \left[ \log \left( \beta(\infty) - \beta(t_i) \right) + n - \sum_{\ell=1}^{i} \varepsilon_{\ell} + r - 1 \right]
+ \ldots + \log(\beta(\infty) - \beta(t_i)) + n - \sum_{\ell=1}^{i} \varepsilon_{\ell} \right].
\]

(4.38)

To simplify the notation, let \( z = \beta(\infty) - \beta(t_i) + n - \sum_{\ell=1}^{i} \varepsilon_{\ell} \). Then, (4.38) is equal to

\[
2 \alpha^{-\varepsilon_i} \frac{\varepsilon_i - r}{t_i} \sum_{r=0}^{\varepsilon_i-1} (-1)^{r} \binom{\varepsilon_i}{r} \frac{d}{dz+r}.
\]

(4.39)
We will have finished showing that the right side of (3.2) is equal to (4.35) as soon as we show that

\[
\frac{\varepsilon_i - r}{\varepsilon_i - 1} \sum_{r=0}^{\varepsilon_i - 1} (-1)^{r} \binom{\varepsilon_i}{r} \frac{(\varepsilon_i - 1)!}{z + r} = \frac{(\varepsilon_i - 1)!}{z(z+1) \ldots (z + \varepsilon_i - 1)}.
\]

(4.40)

To see the last equality, observe that both sides of (4.40) are meromorphic functions, with simple poles at \( z = -r \), for \( r \in \{0, 1, \ldots, \varepsilon_i - 1\} \). That the residues at those poles are equal follows from the formula

\[
\sum_{\ell=0}^{j} (-1)^{\ell} \binom{\varepsilon_i}{\ell} = (-1)^{j} \binom{\varepsilon_i - 1}{j}.
\]

Consider now the discrete part of the posterior. Very similar computations show that the right side of (3.3) is equal to

\[
\left\{ \begin{array}{c} \Pi^* \alpha^\nu \frac{1}{X_i^* - \left( \frac{X_k + X_\ell}{2} \right)} \alpha^\nu \left( \frac{X_k - X_\ell}{2} \right) \\ (\Pi_{\ell=1}^{\Pi} \left( \delta_i(k, \ell) - 1 \right) \prod_{j=1}^{\Pi} \frac{1}{\alpha(\nu) + j - 1}) v^\nu \left( \frac{X_k + X_\ell}{2} \right) \end{array} \right.
\]

(4.41)

This is enough to prove the corollary for the case \( \#(X) = 1 \). The computations for the case \( \#(X) > 1 \) are very similar. \( \Box \)

Acknowledgements

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.
References


