BAYESIAN NONPARAMETRIC ESTIMATION
OF THE MEDIAN; PART I: COMPUTATION
OF THE ESTIMATES

by

Hani Doss

FSU Statistics Report M654

May, 1983

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Key words: Bayes estimator, Dirichlet process priors, estimation of the median, regular conditional distribution.

AMS (1980) Subject Classification: Primary 62F15; Secondary 62G05

Abbreviated title: Bayesian estimation of the median.
BAYESIAN NONPARAMETRIC ESTIMATION OF THE MEDIAN
PART I: COMPUTATION OF THE ESTIMATES

by

Hani Doss
Florida State University

ABSTRACT

Let $X_i$, $i = 1, \ldots, n$ be i.i.d. $\sim F_\theta$, where $F_\theta(x) = F(x-\theta)$ for some $F$ which has median equal to 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 $F$ concentrate all their mass on c.d.f.'s with median equal to 0. These priors include "Dirichlet-type" priors. The marginal posterior distribution of $\theta$ given $X_1, \ldots, X_n$ is computed. The mean of the posterior is taken as the estimate of $\theta$. 

1. Introduction.

Let \( X_1, \ldots, X_n \) be i.i.d. \(~ F_\theta \) where \( F_\theta(x) = F(x - \theta) \), the median of \( F \) is 0, and \( F \) is suspected to be approximately equal to a known distribution \( \alpha_0 \) with a density \( \alpha_0^\prime \) symmetric about 0. Suppose that the problem is to estimate \( \theta \). Use of the m.l.e. \( \hat{\theta}^{\alpha_0} \) based on the model \( X_1, \ldots, X_n \) i.i.d. \(~ \alpha_0(x - \theta) \) leads to an estimator that is efficient if \( F \) is equal to \( \alpha_0 \), but that can perform particularly poorly if \( F \) differs slightly from \( \alpha_0 \) in the heaviness of tails, skewness, etc. Indeed, \( \hat{\theta}^{\alpha_0} \) can consistently estimate a wrong value (for example, if \( \alpha_0 \) is the normal distribution and the mean of \( F \) is not 0).

On the other hand, the nonparametric estimate of the median, i.e. the sample median makes no use at all of any information that is available concerning the shape of \( F \). In fact, there is no known estimate of the median which takes any such information into account.

In the problem of robust estimation of a location parameter, one approach that has been used by many authors is to take a specified neighborhood of \( \alpha_0 \) and find an estimator having certain optimality properties for that neighborhood, e.g. minimax asymptotic mean squared error. An important question is whether or not to let the neighborhoods contain only symmetric distributions. If the neighborhoods contain only symmetric distributions, then the location is a well-defined parameter. The assumption of symmetry may not be realistic, but leads to mathematical convenience and to positive results; see for example Stone (1975) and the references cited in Huber (1981). On the other hand, Pfanzagl (1974) showed that without the assumption of symmetry, one cannot expect asymptotically to improve upon the sample median in estimating the true median of a distribution of unknown shape.
The problem of robust estimation of location has been approached from a Bayesian point of view by Dalal (1979b) and later independently by Diaconis and Freedman in an as yet unpublished paper (but see Diaconis and Freedman, 1982). Dalal and Diaconis and Freedman took the parameter space to be \( \Pi = P^S \times R \) where \( P^S \) denotes the set of all symmetric c.d.f.'s on \( R \). They put a prior \( \mu \) on \( P^S \) (\( \mu \) being a "symmetrized Dirichlet prior"; see Dalal 1979a, 1979b, or Diaconis and Freedman, 1982). They took a prior \( \nu(d\theta) \) on \( R \) and put the product measure \( \mu \times \nu \) on \( P^S \times R \). They computed the posterior distribution of \( \theta \) and \( F \) given a sample \( X_1, \ldots, X_n \) when the values \( \frac{X_i + X_j}{2} \) are distinct, and computed the Bayes estimate of \( \theta \) under squared error as loss. Diaconis and Freedman obtained some interesting results concerning the robustness and consistency properties of the Bayes estimators. These are summarized in Doss (1983c).

In this paper, the parameter space is taken to be \( \Pi = P^* \times R \) where \( P^* \) is the set of distributions \( F \) with median equal to 0. Thus, the median is the parameter to be estimated.

When a prior is put on \( P^* \), then for each \( t \in R \), the quantity \( F(t) \) is a random variable, and the function \( G \) defined by \( G(t) = EF(t) \) is a distribution function on \( R \). \( G \) 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 \( P^* \), it is also important that a prior on \( P^* \) have large support with respect to this topology, so that, hopefully, 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 \( R \). Let \( \alpha \) be a finite non-null measure on \( R \), let \( \Delta_{\alpha} \) 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 \( \Delta_{\alpha} \), then \( EF(t) = \alpha_0([-\infty,t]) \), while the quantity \( \alpha(R) \) expresses the degree of concentration of \( \Delta_{\alpha} \) around its "center" \( \alpha_0 \). The Dirichlet process priors have the attractive feature that if the support of \( \alpha \) is \( R \) then \( \Delta_{\alpha} \) is supported by all of \( P \). Another property is that they give mass 1 to the set of discrete distributions.

The Dirichlet priors on \( P \) can be used to construct a mechanism for selecting a random c.d.f. \( F \) from \( P^* \), as follows. Let \( \alpha \) be a finite symmetric measure on \( R \), and let \( \alpha_0 = \frac{\alpha}{\alpha(R)} \). Let \( \bar{\alpha} \) be the measure concentrated on \( [0, \infty) \) defined by \( \bar{\alpha}(A) = 2 \alpha(A \cap (0, \infty)) + \alpha(A \cap \{0\}) \), and let \( F_1 \) and \( F_2 \) be chosen independently from \( \Delta_{\bar{\alpha}} \). Then, \( F(t) = \frac{1}{2} + \frac{1}{2} F_1(t) - \frac{1}{2} F_2(-t) \) is a random c.d.f. with median equal to 0 a.s., and \( EF(t) = \alpha_0([-\infty,t]) \). The distribution of \( F \) is denoted by \( \Delta_{\bar{\alpha}}^* \). The measure \( \Delta_{\bar{\alpha}}^* \) on \( P^* \) can also be viewed as the conditional distribution of \( \Delta_{2\alpha} \) given that the median of \( F \) is equal to 0; this is explained in Section 2. If the support of the measure \( \alpha \) is all of \( R \), then the support of \( \Delta_{\bar{\alpha}}^* \) is \( P^* \), although \( \Delta_{\bar{\alpha}}^* \) is "concentrated" around the symmetric c.d.f. \( \alpha_0 \).

In this paper a class of priors arising from Doksum's (1974) neutral to the right priors is considered. These priors include the priors \( \Delta_{\bar{\alpha}}^* \). 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 joint posterior distribution of \( \theta \) and \( F \) is not computed, as \( F \) is viewed as a nuisance parameter. In Doss (1983b) the asymptotic behavior of the estimates is analyzed when the prior on \( F \) is \( \Delta_{\bar{\alpha}}^* \).
Suppose now that \( \alpha_0 \) is absolutely continuous with a continuous density \( \alpha_0' \), and that \( \nu \) is an arbitrary prior on \( \theta \). If the prior \( D_a^* \times \nu \) is put on \((F, \theta)\), the marginal posterior distribution of \( \theta \) given \( X = (X_1, \ldots, X_n) \), is shown to be equal to

\[
\nu(d\theta | X) = \frac{\left\{ \prod_{i=1}^{n} \alpha_0'(X_i - \theta) \right\} M(X, \theta) \nu(d\theta)}{\int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{n} \alpha_0'(X_i - \theta) \right\} M(X, \theta) \nu(d\theta)} \tag{1.1}
\]

where \( [M(X, \theta)]^{-1} = \Gamma(\alpha^{(n)} + n F_n(\theta)) \Gamma(\alpha^{(n)} + n(1 - F_n(\theta))) \). \( F_n \) is the empirical distribution function of \( X_1, \ldots, X_n \), and \( \Gamma \) denotes the gamma function. The * in the product indicates that the product is to be taken over distinct values of \( X_i \) only.

The Bayes rule under squared error loss is

\[
\hat{\theta}(X) = \int_{-\infty}^{\infty} \theta \nu(d\theta | X). \tag{1.2}
\]

This should be compared with the simple model in which the true underlying error distribution is known to have the density \( \alpha_0' \) and the prior \( \nu \) is put on \( \theta \). For this model the posterior distribution of \( \theta \) is

\[
\nu(d\theta | X) = \frac{\left\{ \prod_{i=1}^{n} \alpha_0'(X_i - \theta) \right\} \nu(d\theta)}{\int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{n} \alpha_0'(X_i - \theta) \right\} \nu(d\theta)} \tag{1.3}
\]
Equation (1.1) differs from (1.3) in two respects: the factor $M(X, \theta)$ and the * in the product. The factor $M(X, \theta)$, which does not depend on $\alpha_0$, is largest when $\theta$ is equal to the median of the observations and decreases as $\theta$ moves away from the sample median. Thus, roughly speaking (when the $X_i$'s are all distinct), $M(X, \theta)$ has the effect of "shrinking" the posterior (1.3) towards the sample median.

The basic problem of consistency is analyzed in detail in Doss (1983b). To summarize the results, it is necessary to first give a clear notion of what is meant by consistency in the Bayesian context.

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

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

For a loss function $L$, the corresponding Bayes rule $\hat{\psi}$ is consistent at $\psi_0$ if $X_1, X_2, \ldots$ are i.i.d. $\sim P_{\psi_0}$ implies that $\hat{\psi}(X_1, \ldots, X_n)$ converges to $\psi_0$ a.s. $[P_{\psi_0}^\infty]$.

$\hat{\psi}$ is consistent for the family $F$ if $\hat{\psi}$ is consistent at $\psi_0$ for every $\psi_0$ such that $P_{\psi_0} \in F$. The last two notions of consistency can of course be applied to any estimator $\tilde{\psi}$ of $\psi$.

A desirable property for an estimator $\hat{\psi}$ is obviously consistency for the entire parametric family $\{P_\psi \; ; \psi \in \Pi\}$. 
The main results obtained in Doss (1983b) can be summarized as follows. Let $X_1, X_2, \ldots$ be i.i.d. $\sim H_{\theta_0}$ for some $H \in \mathcal{P}^*$, $\theta_0 \in \mathbb{R}$. The posterior (1.1) and the estimator (1.2) have the peculiar property that asymptotically, their behavior depends heavily on whether $H$ is discrete, continuous, or a combination. It is shown that if the underlying error distribution $H$ is discrete, the factor $\Pi^* \alpha_0^*(X_i - \theta)$ is asymptotically negligible relative to $M(X, \theta)$. It should be noted that the sample median is the nonparametric estimate of the population median, and is a consistent estimate of the population median when the latter is unique. It is shown that asymptotically, the prior guess of $\alpha_0$ has no influence on the behavior of the posterior, and both the posterior and the estimator are consistent at $H_{\theta_0}$.

The situation is entirely different if $H$ is continuous. In this case, $\prod_{i=1}^{n} \alpha_0^*(X_i - \theta)$ is equal to $\prod_{i=1}^{n} \alpha_0^*(X_i - \theta)$ with probability one. Let $\hat{\alpha}_0$ denote the maximum likelihood estimate of $\theta$ when the underlying error distribution is assumed to have the density $\alpha_0^*$. Roughly speaking, under sufficient regularity, the factor $\prod_{i=1}^{n} \alpha_0^*(X_i - \theta)$ is largest when $\theta$ is close to the m.l.e. $\hat{\theta}_0^*$ and decreases as $\theta$ moves away from $\hat{\theta}_0^*$. Furthermore, $\prod_{i=1}^{n} \alpha_0^*(X_i - \theta)$ and $M(X, \theta)$ are of the same order of magnitude. If the m.l.e. $\hat{\theta}_0^*$ is a consistent estimate of the population median, then both the posterior and the Bayes estimate $\hat{\theta}$ are consistent at $H_{\theta_0}$. However, it is shown that if the m.l.e. $\hat{\theta}_0^*$ is a consistent estimate of a quantity that is not the population median, then both the posterior and the Bayes estimate are inconsistent at $H_{\theta_0}$.

Doob (1949) has proved that under very general conditions, the posterior $\pi(\theta \mid X)$ is consistent at $\psi_0$ for $[\pi]$ a.e. $\psi_0$. Doob's result raises the question of what is the $\pi$-null set, and more importantly, when is it empty.
LeCam (1953, 1958), Freedman (1963) and Schwartz (1965) have shown that under strong regularity on \( \Pi \) and \( \pi \), the answer is that
\[
\pi(d\psi|X) \text{ is consistent at } \psi_0
\]
if and only if
\[
\psi_0 \in \text{supp}(\pi).
\] (1.4)

The assumptions required for the validity of (1.4) sometimes are severe enough to essentially restrict the result to finite dimensional \( \Pi \). Indeed, Freedman (1963) presented a counterexample involving priors on the set of distributions on the natural numbers.

Outside the Bayesian framework some positive results are obtained. First of all, the estimator based on the prior \( D_\alpha^* \) is consistent if the data is i.i.d. according to the prior guess \( \alpha_0 \). If \( \alpha_0 \) is the double exponential distribution, the m.l.c. \( \hat{\theta}^{\alpha_0} \) is the sample median, and the resulting estimator is consistent for all continuous distributions with a unique median. If \( \alpha_0 \) has the property that \(-\log \alpha^*(\cdot)\) is convex, then roughly speaking, the Bayes estimator is consistent for any underlying symmetric continuous distribution.

The formal setup is described in Section 2. The marginal posterior distribution of \( \theta \) given a sample is computed in Section 3. For the case where the prior on \( F \) is \( D_\alpha^* \) Section 4 gives a description of the basic features of the posterior distribution of \( \theta \). The consistency properties of the posterior and the Bayes estimator are studied in detail in Doss (1983b). In Doss (1983c) are results concerning a class of priors that give probability one to the symmetric c.d.f.'s. This class contains the "symmetrized Dirichlet" priors used by Dalal and Diaconis and Freedman.
2. Preliminaries

2.1. The basic setup.

Let \( \mu \) be a probability measure on \( P^* \), where \( P^* \) denotes the set of all c.d.f.'s \( F \) on \( \mathbb{R} \) with median equal to 0 (i.e., \( F(0^-) = 1 - F(0) \)), and let \( F \) be distributed according to \( \mu \). (Throughout this work, probability measures on \( \mathbb{R} \) are identified with their cumulative distribution functions, and the same symbol is used to denote both the measure and its distribution function whenever convenient.) Let \( \nu \) be a probability measure on \( \mathbb{R} \), let \( \theta \) be distributed according to \( \nu \), and assume that \( \theta \) is independent of \( F \). Let the distribution of the random variables \( X_1, \ldots, X_n \) be as follows: given \( (F, \theta) \), \( X_1, \ldots, X_n \) are i.i.d. \( \sim F_\theta \), where \( F_\theta \) is the distribution function defined by \( F_\theta(x) = F(x-\theta) \).

Formally, the setup is as follows. Let \( F^* \) denote the \( \sigma \)-field on \( P^* \) generated by the topology of weak convergence, and let \( \mu \) be a probability measure on \((P^*, F^*)\). Let \( \nu \) be a probability measure on \((\mathbb{R}, \mathcal{B})\), where \( \mathcal{B} \) denotes the Borel sets of \( \mathbb{R} \). Consider the product space \( \Pi = P^* \times \mathbb{R} \) with the product measure on the product \( \sigma \)-field. This space induces random variables \( X_1, \ldots, X_n \) and a probability measure \( P \) on \( \mathbb{R}^n \times \Pi \) with the product \( \sigma \)-field as follows:

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

\[= \int_A \int_C \Pi \prod_{i=1}^n F(x_i-\theta) \mu(dF) \nu(d\theta)
\]

where \( x_1, \ldots, x_n \in \mathbb{R}, C \in F^*, A \in \mathcal{B} \). Note that (2.1.1) is sufficient to define \( P \).
It is desired to obtain the marginal posterior distribution of $\theta$ given a sample $X_1, \ldots, X_n$. From a decision theoretic viewpoint, the Bayes estimate under squared error as loss is $E(\theta|X_1, \ldots, X_n)$ (other loss functions can also be used). As the conditional expectation of $\theta$ given $X_1, \ldots, X_n$ can be envisaged as an ordinary expectation relative to a "regular conditional distribution" of $\theta$ given $X_1, \ldots, X_n$, it is desired to obtain such a conditional distribution. This is done in Section 3.

If $\mu$ is a measure on $P^*$, 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 separable stochastic process $\{F(t); t \in \mathbb{R}\}$ which satisfies

\begin{align}
(i) \quad & F \text{ is nondecreasing, a.s.} \\
(ii) \quad & \lim_{t \to -\infty} F(t) = 0, \text{ a.s., } \lim_{t \to \infty} F(t) = 1, \text{ a.s.} \\
(iii) \quad & \lim_{t \to s^+} F(t) = F(s) \text{ for each } s \in \mathbb{R}, \text{ a.s.} \\
(iv) \quad & F(0^-) = 1 - F(0), \text{ a.s.}
\end{align}

induces a measure $\mu$ on $(P^*, F^*)$. Conversely, any measure $\mu$ on $(P^*, F^*)$ induces a separable stochastic process $\{F(t); t \in \mathbb{R}\}$ satisfying (i) - (iv) of (2.1.2). Details are provided in Doksum (1974, pp. 189, 190); actually, Doksum considers the space $P$ of all c.d.f.'s on $\mathbb{R}$, but his results apply to $P^*$ as well. The measures $\mu$ on $P^*$ to be considered are described below.

2.2 The Dirichlet priors.

Let $\alpha$ be a finite nonnull measure on $\mathbb{R}$. 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 $\mathbb{R}$ 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

\[
E\alpha(t) = \alpha_0(t).
\]  

(2.2.1)

It can then be said that "\( \mathcal{D}_\alpha \) is centered at \( \alpha_0 \)" (although this is somewhat misleading; the random distribution \( F \) differs a.s. from \( \alpha_0 \) in certain ways, for example, in the tail behavior. See Doss and Sellke (1982)). 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.

In his 1973 paper, Ferguson demonstrated the existence of Dirichlet process priors and showed that they could be used to solve a variety of nonparametric problems. Ferguson (1973, 1974) obtained the following results:

(1) The conditional distribution of \( F \) given a sample \( X_1, \ldots, X_n \) has the Dirichlet distribution with parameter \( \alpha + \sum_{i=1}^{n} \delta_{X_i} \).

(2) \( \mathcal{D}_\alpha \) has a large class of probabilities as its support in the topology of weak convergence. In particular, if the support of \( \alpha \) is \( R \), then \( \mathcal{D}_\alpha \) is supported by all of \( P \).

(3) Let \( G \) denote the gamma distribution with shape parameter \( \alpha \) and scale parameter \( 1 \). If \( \{\gamma(t); t \in [0, \infty)\} \) is a stationary independent increments process with \( \gamma(t) \sim G(t, 1) \), and if \( F(t) = \frac{\gamma(\alpha(t))}{\gamma(\alpha(R))} \) for \( t \in R \), then \( F \) has the Dirichlet distribution with parameter \( \alpha \).

Sethuraman (see Tiwari 1980) showed that

(4) \( F \) may be represented as

\[
F = \sum_{j=1}^{\infty} p_j \delta_{\gamma_j}.
\]
where $V_1, V_2, \ldots$ are i.i.d. $\sim \alpha_0$ and $P_j = B_j \prod_{i=1}^{j-1} (1-B_i)$, with the $B_j$'s i.i.d. $\sim \text{Be}(1, \alpha(\cdot))$ and independent of the $V_j$'s.

As is evident from (4), for small $\alpha(\cdot)$ the largest jump dominates. In fact as $\alpha(\cdot) \to 0$, $D_\alpha$ converges in the weak topology to the random c.d.f. that is degenerate at a point whose distribution is $\alpha_0$. Also as a consequence of (4),

$$D_\alpha \{ F \text{ is discrete} \} = 1.$$  (5)

Let $\alpha$ be a finite symmetric measure on $\mathbb{R}$, and let $\tilde{\alpha}$ be the measure concentrated on $[0, \infty)$ defined by

$$\tilde{\alpha}\{A\} = 2\alpha\{A \cap (0, \infty)\} + \alpha\{A \cap \{0\}\}. \quad (2.2.2)$$

Let $F_1$ and $F_2$ be independently chosen from $D_{\tilde{\alpha}}$, and let $F(\cdot)$ be defined by

$$F(t) = \frac{1}{2} + \frac{1}{2} F_1(t) - \frac{1}{2} F_2(-t^+).$$

This construction defines a measure on $P^*$, which will be denoted $D^{\ast}_{\alpha}$. If $F \sim D^{\ast}_{\alpha}$, then with probability one $F$ is not symmetric, although $EF(t) = \alpha_0(t)$, which is symmetric. The assumption that $\alpha_0$ be symmetric is not necessary for the construction of $D^{\ast}_{\alpha}$.

Assume that $\alpha$ is continuous at 0. Then, $D^{\ast}_{\alpha}$ may be viewed as the conditional distribution of $D_{2\alpha}$ given that $F(0) = \frac{1}{2}$. It is necessary to give this statement more meaning, since the event $\{F(0) = \frac{1}{2}\}$ has $D_{2\alpha}$-probability 0. For $\eta > 0$, let $D_{2\alpha}^\eta$ denote the conditional Dirichletlet prior given that $F(0) \in (\frac{1}{2} - \eta, \frac{1}{2} + \eta)$. Then, as $\eta \to 0$, $D_{2\alpha}^\eta \to D^{\ast}_{\alpha}$ in distribution. A formal proof of this will appear elsewhere. If $0 \in \text{supp}(\alpha)$ the median of $F$ is
unique (this is clear; for a rigorous proof, see Proposition 2 of Chapter V of Doss (1983a)). This means that $D^*_a$ may be viewed as the conditional distribution of $D_{2a}$ given that the median of $F$ is equal to 0.

2.3 Neutral to the right priors.

Let $P^+$ denote the set of all c.d.f.'s concentrated on $[0,\infty)$, that is, the set of all c.d.f.'s $F$ on $R$ with $F(0^-) = 0$, and let $F^+$ be the $\sigma$-field generated by the topology of weak convergence. A measure $\mu^+$ on $(P^+, F^+)$ induces a measure $\mu$ on $(P^*, F^*)$ as follows: let $F_1(\cdot)$ and $F_2(\cdot)$ be chosen independently from $\mu^+$; if $F(\cdot)$ is defined by (2.2.3), then $F(\cdot)$ is a random element of $P^*$.

The measures $\mu^+$ on $P^+$ to be considered are the neutral to the right measures introduced by Doksum. Only random distribution functions continuous in probability will be considered. (Recall that a stochastic process \{X(t)\} is continuous in probability 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 (2.3.1)

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 (2.3.1) 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 function \( F \) defined by (2.3.1) 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 \( 0 \leq t_1 < t_2 \),

\[
\frac{1-F(t_2)}{1-F(t_1)} \quad \text{and} \quad \{F(t); \ t \leq t_1\}
\]

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 is restricted to 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

\[
-\log \mathbb{E} e^{-\lambda Y(t)} = m(t)\lambda + \int_0^\infty (1.e^{-\lambda x}) \, dN_t(x) \quad \text{for} \quad \lambda \geq 0,
\]

(2.3.3)

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 \( t \geq 0 \),

\[
\int_0^1 x \, dN_t(x) < \infty \quad \text{and} \quad N_t[1,\infty) < \infty.
\]
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 (2.3.1), 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 2.3.1.** The Dirichlet process.

Let $\alpha$ be a finite nonnull measure on $[0, \infty)$. Ferguson (1974) showed that if $F$ is distributed according to $D_\alpha$, then $F$ is neutral to the right, and

$$-\log E e^{-Y(t)} = \log \frac{\Gamma(\gamma(\alpha)+\lambda)}{\Gamma(\alpha)} + \log \frac{\Gamma(\alpha(\omega)-\alpha(t))}{\Gamma(\alpha(\omega)-\alpha(\omega)+\lambda)}$$

(2.3.5)

for $\lambda \geq 0$ and $t \geq 0$. In equation (2.3.3),

$$m(\cdot) = 0,$$

(2.3.6)

and

$$dN_t(x) = \frac{e^{-(\alpha(\omega)-\alpha(t))x}(1-e^{-\alpha(t)x})}{x(1-e^{-x})} \, dx.$$  

(2.3.7)
If $\alpha$ is continuous, then $F$ is continuous in probability, and the family $
abla_{x_0}$ of measures is continuous, i.e., for each Borel set $A \subset (0, \infty)$, $N_t(A)$ is continuous.

Example 2.3.2. Homogeneous neutral to the right processes.

Let $Y$ be any nonnegative infinitely divisible random variable. Let $\psi(\cdot)$ be defined by

$$\psi(\lambda) = -\log \mathbb{E} e^{-\lambda Y} \quad \text{for} \quad \lambda \geq 0. \quad (2.3.8)$$

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

$$\psi(\lambda) = p\lambda + \int_0^\infty (1-e^{-\lambda x}) \, dN(x) \quad (2.3.9)$$

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

$$\int_0^1 x \, N(dx) < \infty \quad \text{and} \quad N(1, \infty) < \infty. \quad (2.3.10)$$

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

$$\log \mathbb{E} e^{-\lambda Y(t)} = \log(1-\beta(t)) \psi(\lambda) \quad \text{for} \quad t \geq 0, \lambda \geq 0. \quad (2.3.11)$$

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

3. The Posterior Distribution of $\theta$.

In this section the marginal posterior distribution of $\theta$ given a sample $X_1, \ldots, X_n$ is computed when the prior on $F$ is $D_\alpha$ (Theorem 1). Theorem 2, stated without proof, gives the posterior distribution of $\theta$ given $X_1, \ldots, X_n$ when the prior on $F$ is obtained from the neutral to the right priors as described in section 2.3. In what follows, $X$ denotes the random vector $(X_1, \ldots, X_n)$ and $x$ denotes the vector $(x_1, \ldots, x_n)$. 
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 \( \nu_{X}(\cdot) \) 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 \( A \in \mathcal{B} \), \( \nu_{X}(A) \) 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.)

For ease of reference, all notation and the assumption used in this section are collected here.

Notation.

N1. \( m \) 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, and \( n_i \) denotes the multiplicity of \( x_{(i)} \).

N2. \( \nu_{X}(\cdot) \) or \( \nu(\theta | X) \) denote a regular conditional distribution of \( \theta \) given \( X = x \).
N3. $\Gamma(\cdot)$ denotes the gamma function.

The rest of the notation and the assumption pertain to the neutral to the right priors. Let \( \{Y(t); t \in [0, \infty]\} \) be an independent increments process continuous in probability.

N4. $\psi_t(\lambda) = -\log E e^{-\lambda Y(t)}$ for $t \geq 0$, $\lambda > 0$.

N5. $n_1(\theta) = \sum_{i=1}^{m} I\{x_{(i)} \leq \theta\}$ and $n_2(\theta) = m - n_1(\theta)$. ($I$ is the indicator function.) $n_1(\theta)$ and $n_2(\theta)$ should not be confused with the $n_i$'s defined in N1.

**Assumption.**

A. For each $\lambda \geq 0$, the function $\psi(\cdot)(\lambda)$ is continuously differentiable on $[0, \infty)$. Let $\dot{\psi}_t(\lambda)$ denote $\frac{d}{dt} \psi_t(\lambda)$.

**Theorem 1.** Let $F \sim D^*$, and assume that $\alpha_0$ is absolutely continuous, with density $\alpha_0^*$, continuous on $R$. Then there exists a regular conditional distribution of $\theta$ given $X$ which is absolutely continuous with respect to $\nu$ and is given by

$$\nu(d\theta|X) = c(X) \left( \prod_{i=1}^{n} \alpha_0^*(x_{i} - \theta) \right) M(X, \theta) \nu(d\theta)$$

where $[M(X, \theta)]^{-1} = \Gamma(\alpha(\infty) + nF_n(\theta)) \Gamma(\alpha(\infty) + n(1-F_n(\theta)))$,

with $F_n$ the empirical distribution function of $X_1, \ldots, X_n$. The * in the product indicates that the product is to be taken over distinct values only, and $c(X)$ is a normalizing constant.
Proof. 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

\[
\nu_x(A) = \frac{P(\theta \in A; X=x)}{P(X=x)} \quad \text{for } A \in \mathcal{B}. \tag{3.3}
\]

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

\[
\nu_x^n(A) = \frac{P(\theta \in A; X_i \in (x_i-n/2, x_i+n/2), i=1, \ldots, n)}{P(X_i \in (x_i-n/2, x_i+n/2), i=1, \ldots, n)}. \tag{3.4}
\]

From (2.1.1) it follows that for all \( A \in \mathcal{B} \),

\[
\nu_x^n(A) = \frac{\int \int_{A} \prod_{i=1}^{n} \left[ F(x_i-\theta+n/2) - F(x_i-\theta-n/2) \right] d\alpha (dF) \nu(d\theta)}{\int_{-\infty}^{\infty} \int_{A} \prod_{i=1}^{n} \left[ F(x_i-\theta+n/2) - F(x_i-\theta-n/2) \right] d\alpha (dF) \nu(d\theta)} \tag{3.5}
\]

Defining \( f_x^n(\theta) \) by

\[
f_x^n(\theta) = \prod_{i=1}^{n} \left[ F(x_i-\theta+n/2) - F(x_i-\theta-n/2) \right], \tag{3.6}
\]

(3.5) may be rewritten as

\[
\nu_x^n(A) = \frac{\int_{A} f_x^n(\theta) \nu(d\theta)}{\int_{-\infty}^{\infty} f_x^n(\theta) \nu(d\theta)} \quad \text{for } A \in \mathcal{B}. \tag{3.7}
\]

Suppose that we can find a function \( f_x(\cdot) \) defined on \( \mathbb{R} \) such that for each \( \theta \),

\[
\lim_{\eta \to 0} \eta^{-m} f_x^n(\theta) = f_x(\theta) \tag{3.8}
\]

Assuming that questions involving the uniformity of the convergence have been settled, this gives

\[
\lim_{\eta \to 0} \nu_x^n(A) = \frac{\int_{A} f_x(\theta) \nu(d\theta)}{\int_{-\infty}^{\infty} f_x(\theta) \nu(d\theta)} \quad \text{for each } A \in \mathcal{B}. \tag{3.9}
\]
The task is two-fold. First we need to find a family of functions 
\( \{ f_x(\cdot); x \in \mathbb{R}^n \} \) satisfying (3.8), and second we need to show that the family of measures given by the right side of (3.9) form a regular conditional distribution of \( \theta \) given \( X \).

Consider now \( f^n_x(\theta) \) defined by (3.6), and assume temporarily that \( \theta \notin \{ x_1, \ldots, x_n \} \). By (2.2.5) and the independence of \( F_1(\cdot) \) and \( F_2(\cdot) \), we have

\[
 f^n_x(\theta) = \frac{1}{2^n} f^n_{x,-}(\theta) f^n_{x,+}(\theta)
\]

(3.10)

where

\[
 f^n_{x,-}(\theta) \quad \text{and} \quad f^n_{x,+}(\theta)
\]

are defined by

\[
 f^n_{x,-}(\theta) = \mathbb{E} \prod_{x(i) > \theta} \left[ F_2(x(i)^{-}\theta + \eta/2) - F_2(x(i)^{-}\theta - \eta/2) \right]^{n_i}.
\]

(3.11)

We now use the gamma process representation for the Dirichlet process (see (4) of Section 2). It is well-known that if \( \{ \gamma(t); t \in [0, \alpha(\infty)) \} \) is a gamma process then

\[
 \text{The process } \{ \frac{\gamma(t)}{\gamma(\alpha(\infty))}; t \in [0, \alpha(\infty)) \} \text{ and the random variable } \gamma(\alpha(\infty)) \text{ are independent.}
\]

(3.12)

Consider \( f^n_{x,-}(\theta) \). By (3.12) we have

\[
 f^n_{x,-}(\theta) = \frac{\mathbb{E} \prod_{x(i) < \theta} \left[ \gamma(\tilde{\alpha}(\lvert x(i)^{-}\theta \rvert + \eta/2)) - \gamma(\tilde{\alpha}(\lvert x(i)^{-}\theta \rvert - \eta/2)) \right]^{n_i} \mathbb{E} [\gamma(\tilde{\alpha}(\infty))]^{n} F^n_\theta(\theta)}{\mathbb{E} [\gamma(\tilde{\alpha}(\infty))]^{n} F^n_\theta(\theta)}.
\]

(3.13)
By the independent increments property of the gamma process, for sufficiently small \( n \), (3.13) may be rewritten as

\[
\mathbf{f}^n_{x,\theta}(\theta) = \frac{\prod_{\theta < 0} \frac{\Gamma(n_i + A_i(n))}{\Gamma(A_i(n))}}{\Gamma(\alpha(\infty) + n F_n(\theta))}, \tag{3.14}
\]

where \( A_i(n) \) is given by

\[
A_i(n) = \bar{a}(|x(i) - \theta| + n/2) - \bar{a}(|x(i) - \theta| - n/2). \tag{3.15}
\]

By the continuity of \( a_0^\circ \),

\[
A_i(n) = 2n (a_0^\circ (x(i) - \theta) + o(1)) \tag{3.16}
\]

uniformly for \( \theta \) bounded. This, together with the recursion formula

\[
\Gamma(x+1) = x\Gamma(x)
\]

gives

\[
\mathbf{f}^n_{x,\theta}(\theta) = \frac{\prod_{\theta < 0} \left[2n a_0^\circ(x(i) - \theta)(n_i - 1)! + o(1)\right]}{\Gamma(\alpha(\infty) + n F_n(\theta))} \tag{3.17}
\]

uniformly for \( \theta \) bounded. Combining this with a similar expression for \( \mathbf{f}^n_{x,\theta}(\theta) \) gives

\[
\mathbf{f}^n_{x,\theta}(\theta) = n^m (2^{m-n} \Pi^* (n_i - 1)!) f_x(\theta) + o(n^m) \tag{3.18}
\]

uniformly for \( \theta \) bounded, where

\[
f_x(\theta) = M(x, \theta) \Pi^* a_0^\circ(x_i - \theta). \tag{3.19}
\]

Combining this with (3.7) gives

\[
\nu^n_x(A) = \frac{\int_A (f_x(\theta) + o(1)) \nu(d\theta)}{\int_{-\infty}^{\infty} (f_x(\theta) + o(1)) \nu(d\theta)} \text{ for } A \in \mathcal{B}, \tag{3.20}
\]
with the "little oh" terms uniform for $\theta$ bounded.

Let the measure $\lambda_x$ be defined by

$$
\lambda_x(A) = \frac{\int_A f_x(\theta) \, \nu(d\theta)}{\int_{-\infty}^{\infty} f_x(\theta) \, \nu(d\theta)} \quad \text{for} \quad A \in \mathcal{B}.
$$

(3.21)

We will show that the family $\{\lambda_x; x \in \mathbb{R}^n\}$ is a regular conditional distribution of $\theta$ given $X$.

For $x \in \mathbb{R}^n$, let $H_x$ denote the set of all open cubes of $\mathbb{R}^n$ containing $x$. For any $C \in H_x$ such that $P\{X \in C\} > 0$, let $\nu^C$ denote the probability measure on $(\mathbb{R}, \mathcal{B})$ defined by

$$
\nu^C(A) = \frac{P\{\theta \in A; X \in C\}}{P\{X \in C\}}.
$$

(3.22)

According to a theorem of Pfanzagl (1979),

(i) For $[P]$ a.e. $x$, there exists a probability measure $\nu_x$ such that the net of measures $\{
u^C; C \in H_x\}$ converges weakly to $\nu_x$.

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

(3.23)

(Note that in the definition of a regular conditional probability distribution (3.1), the measure $\nu_x$ needs to be defined only for $[P]$ a.e. $x$.)

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 the net $\{
u^C; C \in H_x\}$ converges weakly to $\nu_x$. Let $x \notin N$ be fixed. By (3.23) (i) we have in particular

$$
\nu^n_x \rightarrow \nu_x \quad \text{weakly}.
$$

(3.24)

An easy argument shows that this implies that

$$
\int_{-\infty}^{\infty} f_x(\theta) \, \nu(d\theta) < \infty.
$$

(3.25)
Let \( a, b \in \mathbb{R} \) be continuity points of \( \nu_x \), let \( \epsilon > 0 \), and let \( K \) be such that

\[
\int_a^b f_x(\theta) \nu(d\theta) \leq \lambda_x(a,b) + \epsilon. \tag{3.26}
\]

Then, by (3.24)

\[
\nu_x(a,b) = \lim_{n \to 0} \frac{\int_a^b (f_x(\theta) + o(1)) \nu(d\theta)}{\int_{-K}^K (f_x(\theta) + o(1)) \nu(d\theta)} \leq \lim_{n \to 0} \frac{\int_a^b (f_x(\theta) + o(1)) \nu(d\theta)}{\int_{-K}^K (f_x(\theta) + o(1)) \nu(d\theta)} \tag{3.27}
\]

Combining (3.27), the fact that the "little oh" terms are uniform for \( \theta \) bounded, (3.26) and the fact that \( \epsilon \) was arbitrary gives

\[
\nu_x(a,b) \leq \lambda_x(a,b) \text{ for all } a, b \in \mathbb{R} \text{ which are continuity points of } \nu_x. \tag{3.28}
\]

An easy argument now shows that equality actually holds in (3.28), and this is enough to show that \( \nu_x = \lambda_x \). The assumption that \( \theta \) was not equal to any of the \( x_i \)'s was made without loss of generality since \( P(\theta \in \{X_1, \ldots, X_n\}) = 0 \), and since the family \( \nu_x \) needs to be defined only for \( [P] \) a.e. \( x \).

Let \( \{Y(t); t \in [0,\infty]\} \) be an independent increments process continuous in probability; as was explained in Section 2.3, this process gives rise to a random distribution function \( F \in \mathcal{P}^* \). \( F \) will be called a random c.d.f. "of the neutral to the right type".
Theorem 2. When $F(\cdot)$ is a random c.d.f. of the neutral to the right type, there exists a regular conditional distribution of $\theta$ given $X = x$, which will be denoted by $\nu_x$. Under $A_1$, for $[P]$ a.e. $x$, $\nu_x$ is absolutely continuous with respect to $\nu$, with density equal to

$$\frac{d\nu_x}{d\nu}(\theta) = c(x) \left\{ \prod_{i=1}^{n_1(\theta)} \exp \left[ \theta - x(i) \left( \sum_{k=1}^{i-1} n_k + \sum_{k=1}^{i} n_k \right) \right] \right\}

\left\{ \sum_{r=0}^{r+1} \left( \begin{array}{c} n_i \\ r \end{array} \right) \psi_{\theta - x(i)} \left( \sum_{k=1}^{i-1} n_k + r \right) \right\}$$

$$\left\{ \prod_{i=1}^{n_2(\theta)} \exp \left[ \psi_{x(m-i+1)-\theta} \left( \sum_{k=1}^{i-1} n_{m-k+1} \right) \right] \right\}

\left\{ \sum_{r=0}^{r+1} \left( \begin{array}{c} n_{m-i+1} \\ r \end{array} \right) \psi_{x(m-i+1)-\theta} \left( \sum_{k=1}^{i-1} n_{m-k+1} + r \right) \right\},$$

(3.29)

where $c(x)$ is a normalizing constant.

Theorem 2 is not proved here. The result is extracted easily from the lemma in Doss (1983c). A detailed proof appears in Doss (1983a).

4. Some remarks about the posterior distribution of $\theta$.

Theorem 1 gives the posterior distribution of $\theta$ given $X_1, \ldots, X_n$ when the prior on $(F,\theta)$ is $D^*_\alpha \times \nu$. It is useful to make a comparison with the "parametric" model where it is assumed that $X_1, \ldots, X_n$ are i.i.d. with density $\alpha^*(x-\theta)$, and the prior $\nu$ is put on $\theta$. This model corresponds to the prior $\delta^*_{\alpha_0} \times \nu$ on $(F,\theta)$. Here, the posterior is proportional to the likelihood times the prior on $\theta$: 
\[ \nu(d\theta | x_1, \ldots, x_n) = c(X) \prod_{i=1}^{n} \alpha_0(x_i - \theta) \nu(d\theta). \]  

(4.1) and (3.2) differ in two respects, the $\ast$ in the product and the factor $M(X, \theta)$. The effect of the $\ast$ in the product is analyzed in Doss (1983b).

In what follows is an analysis of the factor $M(X, \theta)$.

$M(X, \theta)$ is a pseudo-density (it does not integrate to 1) that has a mode at the median of the observations, is constant between observations, and decreases as $\theta$ moves away from the median in either direction. Also, $M(X, \theta)$ depends on $\alpha$ only through $\alpha(\infty)$. An illustration of $M(X, \theta)$ appears below. In the illustration, $\alpha(\infty) = 1$, and there are four observations.

![Illustration of $M(X, \theta)$](image)

**Figure 1.** Illustration of $M(X, \theta)$ when $n = 4$ and $\alpha(\infty) = 1$. 
Let \( \alpha_0 \) be fixed, let \( X_1, \ldots, X_n \) be fixed, and let \( \alpha(\infty) \to \infty \). Then 
\[ [\alpha(\infty)]^{nM(X, \theta)} \] converges to 1 uniformly for \( \theta \in \mathbb{R} \). Thus as \( \alpha(\infty) \to \infty \), the factor \( M(X, \theta) \) disappears, as one would expect intuitively. The * in the product, however, does not disappear, even though \( \mathcal{D}_*^{\alpha} \) converges setwise to the probability measure
\[
c(X) \prod_{i=1}^{n} \alpha_0^*(X_i - \theta) \nu(d\theta).
\]
(4.2)

The best (under squared error as loss) location equivariant estimator

based on \( n \) i.i.d. observations \( X_1, X_2, \ldots, X_n \) with common density \( \alpha_0^*(x - \theta) \) is the Pitman estimate
\[
\theta^p(X) = \frac{\int_{-\infty}^{\infty} \prod_{i=1}^{n} \alpha_0^*(X_i - \theta) d\theta}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} \alpha_0^*(X_i - \theta) d\theta}.
\]
(4.3)

The formula for the mean of the posterior (3.2) with \( \nu \) replaced by Lebesgue measure is
\[
\int_{-\infty}^{\infty} \frac{\theta \prod_{i=1}^{n} \alpha_0^*(X_i - \theta) M(X, \theta) d\theta}{\prod_{i=1}^{n} \alpha_0^*(X_i - \theta) M(X, \theta) d\theta}.
\]
(4.4)

This estimator is location equivariant. Both (4.3) and (4.4) are not scale equivariant.

A simple way to obtain a scale invariant estimate is to replace
\[
\prod_{i=1}^{n} \frac{\alpha_0^*(X_i - \theta)}{S(F_n)}
\]
by
\[
\prod_{i=1}^{n} \frac{\alpha_0^*(X_i - \theta)}{S(F_n)}
\]
where \( S(F) \) is a suitable scale functional, for example, \( S(F) \) is equal to a constant times the MAD (MAD is the median absolute deviation from the median.). This is the way \( M \)-estimates are made scale equivariant.
Johns (1979) has investigated the robustness of Pitman estimates made scale equivariant in this way. He used functions \( \alpha_0^\ast \) which are not necessarily densities (they need not be integrable) and called the corresponding estimates P-estimates. He performed simulations to show that various choices of \( \alpha_0^\ast \) yield estimators which have high efficiencies over a wide variety of symmetric densities.

The method described above for making the estimators scale equivariant is ad hoc. Notice that centering the prior \( D_\alpha^\ast \) around the distribution \( \alpha_0 \) really involves a specification of the scale parameter. One way around this is to proceed as follows. Let \( \alpha^\sigma \) denote the measure defined by \( \alpha^\sigma(A) = \alpha(A/\sigma) \) for linear Borel sets \( A \), and for a given prior \( \lambda \) on \( \sigma \) let \( D_\alpha^\ast \times \lambda \) be the prior on \( P^\ast \) defined by

\[
(D_\alpha^\ast \times \lambda)(E) = \int_0^\infty D_\alpha^\ast(E) \lambda(d\sigma)
\]

for measurable sets \( E \subset P^\ast \). Then, proceed as before to compute the posterior distribution of \( \theta \) given a sample \( X_1, X_2, \ldots, X_n \).

This posterior turns out to be equal to

\[
\nu(d\theta | X) = c(X) \left( \prod_{i=1}^n \frac{1}{\sigma} \alpha_0^\ast \left( \frac{X_i - \theta}{\sigma} \right) \right) \lambda(d\sigma) M(X, \theta) \nu(d\theta).
\]

(4.5)

The calculations and formal justification necessary to obtain (4.5) are very similar to those used in the proof of Theorem 1. Consider (4.5) with \( \nu \) and \( \lambda \) replaced by the improper priors \( \nu(d\theta) = d\theta \) and \( \lambda(d\sigma) = \frac{d\sigma}{\sigma^3} \). The mean of the posterior (for continuous data) is then

\[
\int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{1}{\sigma^3} \prod_{i=1}^n \frac{1}{\sigma} \alpha_0^\ast \left( \frac{X_i - \theta}{\sigma} \right) d\sigma \ M(X, \theta) d\theta
\]

\[
\int_{-\infty}^{\infty} \int_{0}^{\infty} \frac{1}{\sigma^3} \prod_{i=1}^n \frac{1}{\sigma} \alpha_0^\ast \left( \frac{X_i - \theta}{\sigma} \right) d\sigma \ M(X, \theta) d\theta
\]

(4.6)
If the factor $M(X, \theta)$ is removed from (4.6) the result is the so-called location and scale equivariant Pitman estimator of location.

Johns (1979) found that the location and scale equivariant P-estimates performed significantly better than the location equivariant P-estimates made scale equivariant by the *ad hoc* method of division by a scale factor.

Preliminary calculations have shown that both the location equivariant and the location and scale equivariant estimators lie between the corresponding Pitman estimators and the sample median. For large $\alpha(\infty)$, the estimators lie closer to the Pitman estimators. It would be interesting to see if the estimators obtained here retain such efficiencies, and especially, to see how much the factor $M(X, \theta)$ protects against asymmetric contamination.

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


