The Annals of Statistics
1985, Vol. 13, No. 4, 1432–1444

BAYESIAN NONPARAMETRIC ESTIMATION OF THE MEDIAN; PART I: COMPUTATION OF THE ESTIMATES[1]

BY HANI DOSS

Florida State University

Let X_i, $i = 1, \dots, n$ be i.i.d. $\sim F_\theta$, where $F_\theta(x) = F(x - \theta)$ for some F that has median equal to 0. F is assumed unknown or only partially known, and the problem is to estimate θ. Priors are put on the pair (F, θ). 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 θ given $X_1, \dots, X_n$ is computed. The mean of the posterior is taken as the estimate of θ.

1. Introduction and summary. Let $X_1, \dots, X_n$ be i.i.d. $\sim F_\theta$ 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 α_0 with a density α_0' symmetric about 0. Suppose that the problem is to estimate θ. Use of the mle $\hat{\theta}^{\alpha_0}$ based on the model $X_1, \dots, X_n$ i.i.d. $\sim \alpha_0'(x - \theta)$ leads to an estimator that is efficient if F is equal to α_0, but that can perform particularly poorly if F differs slightly from α_0 in the heaviness of tails, skewness, etc. Indeed, $\hat{\theta}^{\alpha_0}$ can consistently estimate a wrong value (for example, if α_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 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 α_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).

In a Bayesian approach to estimating the median, one proceeds as follows. Take the parameter space to be $\Pi = \mathscr{P}^* \times \mathscr{R}$, where $\mathscr{P}^*$ is the set of all c.d.f.s with median equal to 0, and put a prior π on the generic point (F, θ) of Π. Compute the marginal posterior distribution of θ given a sample $X_1, \dots, X_n$ from $F(x - \theta)$; with squared error as loss the Bayes estimate is the mean of the

Received July 1983; revised December 1984.

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

AMS 1980 *subject classifications*. Primary 62F15; secondary 62G05.

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

posterior.

In this paper we consider a prior π of the following form. The parameters F and θ are taken to be independent; the distribution of θ is arbitrary and the prior on F is obtained from Doksum's (1974) neutral to the right priors. The simplest of these priors on F are related to the Dirichlet priors, and we now proceed to describe them.

Let $M > 0$, let $\alpha = M\alpha_0$ and let α_- and α_+ denote the restrictions of α to $(-\infty, 0)$ and $(0, \infty)$, respectively. Choose F_- and F_+ independently from $\mathscr{D}_{\alpha-}$ and $\mathscr{D}_{\alpha+}$, respectively ($\mathscr{D}_\beta$ denotes the Dirichlet prior with parameter measure β; see Ferguson, 1973, 1974), and form $F = \frac{1}{2}F_- + \frac{1}{2}F_+$. This F has median equal to 0, but with probability one is not symmetric, although it is symmetric "on the average." For example,

$$EF(t) = \alpha_0(t) \quad \text{for all } t.$$

The distribution of F is denoted $\mathscr{D}_\alpha^*$. It has many of the properties of the ordinary Dirichlet prior: The support of $\mathscr{D}_\alpha^*$ is appropriately large if the support of α is large (see Doss, 1985). Also, the parameter M indicates the degree of concentration of $\mathscr{D}_\alpha^*$ about its "center" α_0. For example, it is easy to see that if $M \to \infty$, then $\mathscr{D}_\alpha^*$ converges to the point mass at α_0 in the weak topology. The measure $\mathscr{D}_\alpha^*$ on $\mathscr{P}^*$ can also be viewed as the conditional distribution of the ordinary Dirichlet prior with parameter α given that the median of F is equal to 0.

The posterior distribution of θ given a sample is obtained. The Bayes estimate of θ turns out to be essentially a convex combination of the mle $\hat{\theta}^{\alpha_0}$ and of the sample median, with the weights depending on the sample.

The formal setup is described in Section 2. The marginal posterior distribution of θ, given a sample, is computed in Section 3. For the case where the prior on F is $\mathscr{D}_\alpha^*$, Section 4 gives a description of the basic features of the posterior distribution of θ. Section 5 extends the results to the problem of estimating quantiles. The consistency properties of the posterior and the Bayes estimator of the median are studied in detail in Doss (1985).

Dalal (1979a, b) and Diaconis and Freedman (1985a, b) considered a "symmetrized Dirichlet distribution" as the prior on F. This prior, denoted $\overline{\mathscr{D}}_\alpha$, is constructed as follows. Let $F_+ \sim \mathscr{D}_{\alpha+}$ as before. Then $F(t) = \frac{1}{2}F_+(t) + \frac{1}{2}(1 - F_+(-t^-))$ has the prior $\overline{\mathscr{D}}_\alpha$. This F is symmetric. Note that if F_1 and F_2 are chosen *independently* from $\mathscr{D}_{\alpha+}$, then $F(t) = \frac{1}{2}F_1(t) + \frac{1}{2}(1 - F_2(-t^-))$ is distributed according to $\mathscr{D}_\alpha^*$. Diaconis and Freedman computed the posterior distribution of θ given a sample $X_1, \ldots, X_n$ when the values $\frac{1}{2}(X_i + X_j)$ are distinct, and obtained the Bayes estimate of θ under squared error as loss. They showed that for certain choices of α_0 the Bayes estimate can be inconsistent. See Diaconis and Freedman (1982) for a discussion of the robustness properties of these Bayes rules. The problem of Bayesian nonparametric estimation of quantiles has also been considered by Ferguson (1973). Doss (1984) gives 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 μ be a probability measure on $\mathscr{P}^*$, where $\mathscr{P}^*$ denotes the set of all c.d.f.s F on $\mathscr{R}$ with median equal to 0 [i.e., $F(0^-) \le \frac{1}{2} \le F(0)$], and let F be distributed according to μ. (Throughout this work, probability measures on $\mathscr{R}$ 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 ν be a probability measure on $\mathscr{R}$, let θ be distributed according to ν, and assume that θ is independent of F. Let the distribution of the random variables $X_1, \dots, X_n$ be as follows: Given (F, θ), $X_1, \dots, X_n$ are i.i.d. $\sim F_\theta$, where F_θ is the distribution function defined by $F_\theta(x) = F(x - \theta)$.

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

$$P\{X_1 \le x_1; \cdots; X_n \le x_n; F \in C; \theta \in A\} = \int_A \int_C \prod_{i=1}^{n} F(x_i - \theta) \mu(dF) \nu(d\theta), \tag{2.1}$$

where $x_1, \dots, x_n \in \mathscr{R}$, $C \in \mathscr{F}^*$, $A \in \mathscr{B}$. Note that (2.1) is sufficient to define P.

It is desired to obtain the marginal posterior distribution of θ given a sample $X_1, \dots, X_n$. From a decision theoretic viewpoint, the Bayes estimate under squared error as loss is $E(\theta | X_1, \dots, X_n)$ (other loss functions can also be used). As the conditional expectation of θ given $X_1, \dots, X_n$ can be envisaged as an ordinary expectation relative to a "regular conditional distribution" of θ given $X_1, \dots, X_n$, it is desired to obtain such a conditional distribution. This is done in Section 3.

If μ is a measure on $\mathscr{P}^*$, then the distribution function F is a random function. It will be very helpful to view μ as a stochastic process $\{F(t); t \in \mathscr{R}\}$. Briefly, any separable stochastic process $\{F(t); t \in \mathscr{R}\}$ that satisfies

$$(2.2)\quad \begin{aligned} &\text{(i) } F \text{ is nondecreasing, a.s.,} \\ &\text{(ii) } \lim_{t \to -\infty} F(t) = 0, \text{ a.s.,} \quad \lim_{t \to \infty} F(t) = 1, \text{ a.s.,} \\ &\text{(iii) } \lim_{t \to s^+} F(t) = F(s) \text{ for each } s \in \mathscr{R}, \text{ a.s.,} \\ &\text{(iv) } F(0^-) \le \tfrac{1}{2} \le F(0), \text{ a.s.,} \end{aligned}$$

induces a measure μ on $(\mathscr{P}^*, \mathscr{F}^*)$. Conversely, any measure μ on $(\mathscr{P}^*, \mathscr{F}^*)$ induces a separable stochastic process $\{F(t); t \in \mathscr{R}\}$ satisfying (i)–(iv) of (2.2). Details are provided in Doksum (1974, pp. 189, 190); actually, Doksum considers the space $\mathscr{P}$ of all c.d.f.s on $\mathscr{R}$, but his results apply to $\mathscr{P}^*$ as well. The measures μ on $\mathscr{P}^*$ to be considered are derived from Doksum's neutral to the right measures. The most important of these are the priors $\mathscr{D}_\alpha^*$, which are treated separately. The general case is discussed in Section 3 after the proof of Theorem 1.

2.2. The prior $\mathscr{D}_\alpha^*$. Let α be a finite nonnull measure on $\mathscr{R}$. For the construction below, α need not be symmetric, but is assumed to have "median equal to 0" in that $\alpha(-\infty, 0) = \alpha(0, \infty)$. Let α_- and α_+ be the restrictions of α to $(-\infty, 0)$ and $(0, \infty)$, respectively, in the following sense:

$$\begin{aligned} \alpha_-\{A\} &= \alpha\{A \cap (-\infty, 0)\} + \tfrac{1}{2}\alpha\{A \cap \{0\}\} \quad \text{and} \\ \alpha_+\{A\} &= \alpha\{A \cap (0, \infty)\} + \tfrac{1}{2}\alpha\{A \cap \{0\}\}. \end{aligned} \tag{2.3}$$

Choose F_- and F_+ independently from $\mathscr{D}_{\alpha-}$ and $\mathscr{D}_{\alpha+}$, respectively, and form

$$F(t) = \tfrac{1}{2}F_-(t) + \tfrac{1}{2}F_+(t). \tag{2.4}$$

This construction defines a measure on $\mathscr{P}^*$, which we denote by $\mathscr{D}_\alpha^*$. Writing $\alpha_0 = \alpha/(\alpha\{\mathscr{R}\})$, it is easy to check that $EF(t) = \alpha_0(t)$ for all t.

Assume that α is continuous at 0. Then, $\mathscr{D}_\alpha^*$ may be viewed as the conditional distribution of $\mathscr{D}_\alpha$ 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 $\mathscr{D}_\alpha$-probability 0. For $\eta > 0$, let $\mathscr{D}_\alpha^\eta$ denote the conditional Dirichlet prior given that $F(0) \in (\frac{1}{2} - \eta, \frac{1}{2} + \eta)$. Then, as $\eta \to 0$, $\mathscr{D}_\alpha^\eta \to \mathscr{D}_\alpha^*$ in the weak topology. A formal proof of this will appear elsewhere. If $0 \in \operatorname{supp}(\alpha)$ the median of F is unique [this is clear; for a rigorous proof, see Proposition 2 of Chapter V of Doss (1983)]. This means that $\mathscr{D}_\alpha^*$ may be viewed as the conditional distribution of $\mathscr{D}_\alpha$ given that the median of F is equal to 0.

3. The posterior distribution of θ. In this section the marginal posterior distribution of θ given a sample $X_1, \dots, X_n$ is computed when the prior on F is $\mathscr{D}_\alpha^*$ (Theorem 1). Theorem 2, stated without proof, gives the posterior distribution of θ given $X_1, \dots, X_n$ when the prior on F is obtained from the neutral to the right priors. In what follows, X denotes the random vector $(X_1, \dots, X_n)$ and x denotes the vector $(x_1, \dots, x_n)$.

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 σ-finite measure dominating the family $\{F_\theta; (F, \theta) \in \Pi\}$. Consequently, the posterior distribution of θ will have to be obtained in a different way.

What is desired is a regular conditional distribution of θ given X. Recall that a regular conditional distribution of θ given X is a function $\nu_{(\cdot)}(\cdot)$ defined on $\mathscr{R}^n \times \mathscr{B}$ satisfying:

(3.1)

(i) For each $x \in \mathscr{R}^n$, ν_x is a probability measure on $\mathscr{B}$;

(ii) For each $A \in \mathscr{B}$, $\nu_{(\cdot)}(A)$ is a measurable function of x;

(iii) For each $A \in \mathscr{B}$, $\nu_x(A)$ 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 (1978), p. 211.]

Before stating Theorem 1, we give some notation.

NOTATION.

N1. m is equal to the number of distinct values of the sequence $\{x_1, \dots, x_n\}$, $x_{(1)} < x_{(2)} < \cdots < x_{(m)}$ denote the ordered values of the sequence, and n_i denotes the multiplicity of $x_{(i)}$.

N2. $\nu_x(\cdot)$ or $\nu(d\theta|x)$ denote a regular conditional distribution of θ given $X = x$.

N3. $\Gamma(\cdot)$ denotes the gamma function.

THEOREM 1. *Let $F \sim \mathscr{D}_\alpha^*$, and assume that α_0 is absolutely continuous, with density α_0', continuous on $\mathscr{R}$. Let $\theta \sim \nu$ be independent of F, and given θ and F let $X_1, \dots, X_n$ be i.i.d. $\sim F(x - \theta)$. Then there exists a regular conditional distribution of θ given X that is absolutely continuous with respect to ν and is given by*

$$\nu(d\theta|X) = c(X)\left(\prod_{i=1}^{n}{}^{*} \alpha_0'(X_i - \theta)\right)M(X,\theta)\nu(d\theta), \tag{3.2}$$

*where $[M(X,\theta)]^{-1} = \Gamma(\frac{1}{2}\alpha(\infty) + nF(\theta))\Gamma(\frac{1}{2}\alpha(\infty) + n(1 - F_n(\theta)))$, with F_n the empirical distribution function of $X_1, \dots, X_n$. 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, ν_x on $\mathscr{B}$ by

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

Consider instead, for x fixed, $\eta > 0$, the measure ν_x^η defined on $\mathscr{B}$ by

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

From (2.1) it follows that for all $A \in \mathscr{B}$,

$$\nu_x^\eta(A) = \frac{\int_A \int_{\mathscr{P}} \prod_{i=1}^n [F(x_i - \theta + \eta/2) - F(x_i - \theta - \eta/2)]\mathscr{D}_\alpha^*(dF)\nu(d\theta)}{\int_{-\infty}^{\infty} \int_{\mathscr{P}} \prod_{i=1}^n [F(x_i - \theta + \eta/2) - F(x_i - \theta - \eta/2)]\mathscr{D}_\alpha^*(dF)\nu(d\theta)}. \tag{3.5}$$

Defining $f_x^\eta(\theta)$ by

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

(3.5) may be rewritten as

$$\nu_x^\eta(A) = \frac{\int_A f_x^\eta(\theta)\nu(d\theta)}{\int_{-\infty}^{\infty} f_x^\eta(\theta)\nu(d\theta)} \quad \text{for } A \in \mathscr{B}. \tag{3.7}$$

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

$$\lim_{\eta \to 0} \eta^{-m} f_x^\eta(\theta) = f_x(\theta). \tag{3.8}$$

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

$$\lim_{\eta \to 0} \nu_x^\eta(A) = \frac{\int_A f_x(\theta)\nu(d\theta)}{\int_{-\infty}^{\infty} f_x(\theta)\nu(d\theta)} \quad \text{for each } A \in \mathscr{B}. \tag{3.9}$$

The task is two-fold. First we need to find a family of functions $\{f_x(\cdot);\, x \in \mathscr{R}^n\}$ satisfying (3.8), and second we need to show that the family of measures given by the right-hand side of (3.9) forms a regular conditional distribution of θ given X.

Consider now $f_x^\eta(\theta)$ defined by (3.6), assume temporarily that $\theta \notin \{x_1, \dots, x_n\}$, and assume that η is sufficiently small so that for all i, $\theta \notin (x_i - \eta/2, x_i + \eta/2)$. By (2.4) and the independence of $F_-(\cdot)$ and $F_+(\cdot)$, we have

$$f_x^\eta(\theta) = \frac{1}{2^n} f_{x,-}^\eta(\theta) f_{x,+}^\eta(\theta), \tag{3.10}$$

where $f_{x,-}^\eta(\theta)$ and $f_{x,+}^\eta(\theta)$ are defined by

$$f_{x,\,\underset{(+)}{-}}^\eta(\theta) = E \prod_{\substack{x_{(i)} - \theta < 0 \\ (>)}} \left[F_{\underset{(+)}{-}}(x_{(i)} - \theta + \eta/2) - F_{\underset{(+)}{-}}(x_{(i)} - \theta - \eta/2) \right]^{n_i}. \tag{3.11}$$

We now use the gamma process representation for the Dirichlet process. Let $\mathscr{G}(u, 1)$ denote the gamma distribution with shape parameter u and scale parameter 1, and let β be a finite nonnull measure on $\mathscr{R}$. If $\{\gamma(t);\, t \in [0, \beta(\infty)]\}$ is a stationary independent increments process with $\gamma(t) \sim \mathscr{G}(t, 1)$, and if $F(t) = \gamma(\beta(t))/\gamma(\beta(\infty))$ for $t \in \mathscr{R}$, then $F \sim \mathscr{D}_\beta$; see Ferguson (1973). It is well known that if $\{\gamma(t);\, t \in [0, \beta(\infty)]\}$ is a gamma process then

(3.12) the process $\left\{ \dfrac{\gamma(t)}{\gamma(\beta(\infty))};\, t \in [0, \beta(\infty)] \right\}$ and the random variable $\gamma(\beta(\infty))$ are independent.

Consider $f_{x,-}^\eta(\theta)$. By (3.12) and (2.3) we have

$$f_{x,-}^\eta(\theta) = \frac{E \prod_{x_{(i)} - \theta < 0} \left[\gamma\big(\alpha(x_{(i)} - \theta + \eta/2)\big) - \gamma\big(\alpha(x_{(i)} - \theta - \eta/2)\big) \right]^{n_i}}{E\left[\gamma\left(\tfrac{1}{2}\alpha(\infty)\right)\right]^{nF_n(\theta)}}. \tag{3.13}$$

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

$$f_{x,-}^{\eta}(\theta) = \frac{\prod_{x_{(i)}-\theta<0}\Gamma(n_i + A_i(\eta))/\Gamma(A_i(\eta))}{\Gamma(\frac{1}{2}\alpha(\infty) + nF_n(\theta))/\Gamma(\frac{1}{2}\alpha(\infty))}, \tag{3.14}$$

where $A_i(\eta)$ is given by

$$A_i(\eta) = \alpha(x_{(i)} - \theta + \eta/2) - \alpha(x_{(i)} - \theta - \eta/2). \tag{3.15}$$

By the continuity of α_0',

$$A_i(\eta) = \eta\big(\alpha'(x_{(i)} - \theta) + o(1)\big) \tag{3.16}$$

uniformly for θ bounded. This, together with the recursion formula $\Gamma(x+1) = x\Gamma(x)$ gives

$$f_{x,-}^{\eta}(\theta) = \frac{\prod_{x_{(i)}-\theta<0}\eta\left[\alpha'(x_{(i)} - \theta)(n_i - 1)! + o(1)\right]}{\Gamma(\frac{1}{2}\alpha(\infty) + nF_n(\theta))/\Gamma(\frac{1}{2}\alpha(\infty))} \tag{3.17}$$

uniformly for θ bounded. Combining this with a similar expression for $f_{x,+}^{\eta}(\theta)$ gives

$$f_x^{\eta}(\theta) = \eta^m 2^{-n}(\Pi^*(n_i - 1)!)\left[\Gamma(\tfrac{1}{2}\alpha(\infty))\right]^2 f_x(\theta) + o(\eta^m) \tag{3.18}$$

uniformly for θ bounded, where

$$f_x(\theta) = M(x, \theta)\Pi^*\alpha'(x_i - \theta). \tag{3.19}$$

Combining this with (3.7) gives

$$\nu_x^{\eta}(A) = \frac{\int_A (f_x(\theta) + o(1))\nu(d\theta)}{\int_{-\infty}^{\infty}(f_x(\theta) + o(1))\nu(d\theta)} \quad \text{for } A \in \mathscr{B}, \tag{3.20}$$

with the "little oh" terms uniform for θ bounded.

Let the measure λ_x be defined by

$$\lambda_x(A) = \frac{\int_A f_x(\theta)\nu(d\theta)}{\int_{-\infty}^{\infty} f_x(\theta)\nu(d\theta)} \quad \text{for } A \in \mathscr{B}. \tag{3.21}$$

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

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

$$\nu^C(A) = \frac{P\{\theta \in A; X \in C\}}{P\{X \in C\}}. \tag{3.22}$$

According to a theorem of Pfanzagl (1979):

(3.23)

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

(ii) The family ν_x above is a regular conditional probability distribution of θ given X.

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

Let $N \in \mathscr{B}^n$ be a set of probability 0, as guaranteed by (i), with the property that for all $x \notin N$, there exists a probability measure ν_x such that the net $\{\nu^C; C \in \mathscr{K}_x\}$ converges weakly to ν_x. Let $x \notin N$ be fixed. By (3.23)(i), we have in particular

$$\nu_x^\eta \to \nu_x \quad \text{weakly.} \tag{3.24}$$

An easy argument shows that this implies that

$$\int_{-\infty}^{\infty} f_x(\theta)\nu(d\theta) < \infty. \tag{3.25}$$

Let $a, b \in \mathscr{R}$ be continuity points of ν_x, let $\varepsilon > 0$, and let K be such that

$$\frac{\int_a^b f_x(\theta)\nu(d\theta)}{\int_{-K}^K f_x(\theta)\nu(d\theta)} \le \lambda_x(a, b) + \varepsilon. \tag{3.26}$$

Then, by (3.24)

$$\begin{aligned} \nu_x(a, b) &= \lim_{\eta \to 0} \frac{\int_a^b (f_x(\theta) + o(1))\nu(d\theta)}{\int_{\infty}^{\infty} (f_x(\theta) + o(1))\nu(d\theta)} \\ &\le \lim_{\eta \to 0} \frac{\int_a^b (f_x(\theta) + o(1))\nu(d\theta)}{\int_{-K}^K (f_x(\theta) + o(1))\nu(d\theta)}. \end{aligned} \tag{3.27}$$

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

$$\nu_x(a, b) \le \lambda_x(a, b) \quad \text{for all } a, b \in \mathscr{R} \tag{3.28}$$

which are continuity points of ν_x. 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 θ was not equal to any of the x_is was made without loss of generality since $P(\theta \in \{X_1, \dots, X_n\}) = 0$, and since the family ν_x needs to be defined only for $[P]$ a.e. x. □

Random c.d.f.s of the neutral to the right type. Let $F_i(\cdot)$, $i = 1, 2$, be two independent neutral to the right random distribution functions on $[0, \infty)$. We can use these to construct a random element $F(\cdot)$ on $\mathscr{P}^*$ as before:

$$F(t) = \tfrac{1}{2}F(t) + \tfrac{1}{2}(1 - F_2(-t^-)) \quad \text{for } -\infty < t < \infty. \tag{3.29}$$

This F is called a random c.d.f. "of the neutral to the right type." The processes $F_i(\cdot)$ can be written $F_i(t) = 1 - e^{-Y_i(t)}$, $t \ge 0$, $i = 1, 2$, where $Y_i(\cdot)$ are nondecreasing independent increments processes. We assume that $Y_i(\cdot)$ are continuous in probability [recall that a stochastic process $\{X(t)\}$ is continuous in probability if $s \to t$ implies $X(s) \to X(t)$ in probability]. This assumption implies the continuity of the c.d.f. defined by $EF(t)$. For simplicity we assume that $Y_1(\cdot)$ and $Y_2(\cdot)$ have the same distribution. Thus, $EF(t)$ is symmetric. Before stating Theorem 2, we give some more notation and state an assumption.

NOTATION.

N4. $\psi_t(\lambda) = -\log Ee^{-\lambda Y_1(t)}$ for $t \geq 0$, $\lambda \geq 0$.

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_is 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 $(\partial/\partial t)\psi_t(\lambda)$.

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 θ given $X = x$, which will be denoted by ν_x. Under A, for $[P]$ a.e. x, ν_x is absolutely continuous with respect to ν, with density equal to*

$$\begin{aligned}\frac{d\nu_x}{d\nu}(\theta) = c(x)&\left\{\prod_{i=1}^{n_1(\theta)}\left(\exp\left[\psi_{\theta - x_{(i)}}\left(\sum_{l=1}^{i-1} n_l\right) - \psi_{\theta - x_{(i)}}\left(\sum_{l=1}^{i} n_l\right)\right]\right)\right.\\ &\quad\left.\cdot\left(\sum_{r=0}^{n_i}(-1)^{r+1}\binom{n_i}{r}\dot{\psi}_{\theta - x_{(i)}}\left(\sum_{l=1}^{i-1} n_l + r\right)\right)\right\}\\ &\cdot\left\{\prod_{i=1}^{n_2(\theta)}\left(\exp\left[\psi_{x_{(m-i+1)}-\theta}\left(\sum_{l=1}^{i-1} n_{m-l+1}\right) - \psi_{x_{(m-i+1)}-\theta}\left(\sum_{l=1}^{i} n_{m-l+1}\right)\right]\right)\right.\\ &\quad\left.\cdot\left(\sum_{r=0}^{n_{m-i+1}}(-1)^{r+1}\binom{n_{m-i+1}}{r}\dot{\psi}_{x_{(m-i+1)}-\theta}\left(\sum_{l=1}^{i-1} n_{m-l+1} + r\right)\right)\right\},\end{aligned} \tag{3.30}$$

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

Theorem 2 is not proved here. The result is extracted easily from the lemma in Doss (1984). A detailed proof appears in Doss (1983).

4. Some remarks about the posterior distribution of θ. Theorem 1 gives the posterior distribution of θ given $X_1, \ldots, X_n$ when the prior on (F, θ) is $\mathscr{D}_\alpha^* \times \nu$. 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 $\alpha_0'(x - \theta)$, and the prior ν is put on θ. This model corresponds to the prior $\delta_{\alpha_0} \times \nu$ on (F, θ). Here, the posterior is proportional to the likelihood times the prior on θ:

$$\nu(d\theta | X_1, \ldots, X_n) = c(X)\prod_{i=1}^n \alpha_0'(X_i - \theta)\nu(d\theta). \tag{4.1}$$

(4.1) and (3.2) differ in two respects, the * in the product and the factor $M(X, \theta)$. The effect of the * in the product is analyzed in Doss (1985). What follows is an analysis of the factor $M(X, \theta)$.

$M(X, \theta)$ is a pseudodensity (it does not integrate to 1) that has a mode at the median of the observations, is constant between observations, and decreases as θ moves away from the median in either direction. Also, $M(X, \theta)$ depends on α only through $\alpha(\infty)$. Thus, $M(X, \theta)$ "shrinks" the posterior towards the sample

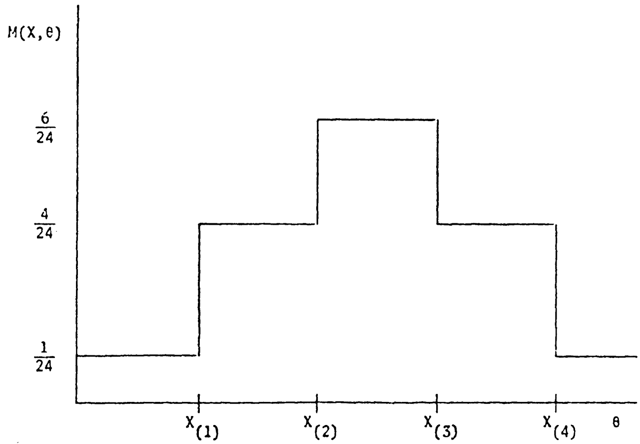

FIG. 1. *Illustration of $M(X,\theta)$ when $n = 4$ and $\alpha(\infty) = 2$.*

median. This effect is more pronounced when $\alpha(\infty)$ is small. Figure 1 gives an illustration of $M(X, \theta)$. In Figure 1, $\alpha(\infty) = 2$, and there are four observations.

Let α_0 be fixed, let $X_1, \dots, X_n$ be fixed, and let $\alpha(\infty) \to \infty$. Then $[\alpha(\infty)]^n M(X, \theta)$ converges to 1 uniformly for $\theta \in \mathscr{R}$. 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 $\mathscr{D}_\alpha^*$ converges to the point mass at α_0 in the weak topology. The posterior $\nu(d\theta | X)$ converges setwise to the probability measure

$$c(X) \prod_{i=1}^{n}{}^{*} \alpha_0'(X_i - \theta) \nu(d\theta). \tag{4.2}$$

The best (under squared error as loss) location equivariant estimator based on n i.i.d. observations $X_1, X_2, \dots, X_n$ with common density $\alpha_0'(x - \theta)$ is the Pitman estimate

$$\hat{\theta}^P(X) = \frac{\int_{-\infty}^{\infty} \theta \prod_{i=1}^n \alpha_0'(X_i - \theta)\, d\theta}{\int_{-\infty}^{\infty} \prod_{i=1}^n \alpha_0'(X_i - \theta)\, d\theta}. \tag{4.3}$$

The formula for the mean of the posterior (3.2) with ν replaced by Lebesgue measure is

$$\frac{\int_{-\infty}^{\infty} \theta \prod{}^* \alpha_0'(X_i - \theta) M(X, \theta)\, d\theta}{\int_{-\infty}^{\infty} \prod{}^* \alpha_0'(X_i - \theta) M(X, \theta)\, d\theta}. \tag{4.4}$$

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 $\prod_{i=1}^n \alpha_0'(X_i - \theta)$ by $\prod_{i=1}^n \alpha_0'((X_i - \theta)/[S(F_n)])$ 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 α_0' that 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 α_0' yield estimators that 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 $\mathscr{D}_\alpha^*$ around the distribution α_0 really involves a specification of the scale parameter. One way around this is to proceed as follows. Let α^σ denote the measure defined by $\alpha^\sigma(A) = \alpha(A/\sigma)$ for linear Borel sets A, and for a given prior λ on σ let $\mathscr{D}_\alpha^* \times \lambda$ be the prior on $\mathscr{P}^*$ defined by

$$(\mathscr{D}_\alpha^* \times \lambda)(E) = \int_0^\infty \mathscr{D}_{\alpha^\sigma}^*(E)\lambda(d\sigma)$$

for measurable sets $E \subset \mathscr{P}^*$. Then, proceed as before to compute the posterior distribution of θ given a sample $X_1, X_2, \dots, X_n$. This posterior turns out to be equal to

$$\nu(d\theta|X) = c(X)\left(\int_0^\infty \prod_{i=1}^{n}{}^{*} \frac{1}{\sigma}\alpha_0'\left(\frac{X_i - \theta}{\sigma}\right)\lambda(d\sigma)\right)M(X,\theta)\nu(d\theta). \tag{4.5}$$

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 ν and λ replaced by the improper priors $\nu(d\theta) = d\theta$ and $\lambda(d\sigma) = d\sigma/\sigma^3$. The mean of the posterior (for continuous data) is then

$$\frac{\int_{-\infty}^\infty \theta \int_0^\infty \frac{1}{\sigma^3} \prod_{i=1}^n \frac{1}{\sigma}\alpha_0'\left(\frac{X_i - \theta}{\sigma}\right) d\sigma\, M(X,\theta)\, d\theta}{\int_{-\infty}^\infty \int_0^\infty \frac{1}{\sigma^3} \prod_{i=1}^n \frac{1}{\sigma}\alpha_0'\left(\frac{X_i - \theta}{\sigma}\right) d\sigma\, M(X,\theta)\, d\theta} \tag{4.6}$$

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.

5. Estimation of quantiles. The theory above can be extended to the problem of quantile estimation. Let α be a finite nonnull measure on $\mathscr{R}$, and assume that α_0 has pth quantile equal to 0. For simplicity, we assume also that α_0 is continuous at 0. Let α_- and α_+ denote the restrictions of α to $(-\infty, 0)$ and $(0, \infty)$, respectively, and choose F_- and F_+ independently from $\mathscr{D}_{\alpha-}$ and $\mathscr{D}_{\alpha+}$, respectively. For $p \in (0,1)$, form

$$F(t) = pF_-(t) + (1-p)F_+(t), \qquad -\infty < t < \infty. \tag{5.1}$$

With probability one, F has pth quantile equal to 0 [in fact, $F(0^-) = p = F(0)$], and $EF(t) = \alpha_0(t)$ for all t. Denote the distribution of F by $\mathscr{D}_{\alpha, p}$. We can use the same setup as before in order to estimate the pth quantile.

THEOREM 3. *Assume the same setup as in Theorem* 1, *except that* $F \sim \mathscr{D}_{\alpha, p}$. *Then there exists a regular conditional distribution of* θ *given* X *that is absolutely continuous with respect to* ν *and is given by*

$$\nu(d\theta|X) = c(X)\left[\prod_{i=1}^{n}{}^{*}\alpha_0'(X_i - \theta)\right]K(X,\theta)\nu(d\theta), \tag{5.2}$$

where

$$K(X,\theta) = \frac{p^{nF_n(\theta)}(1-p)^{n(1-F_n(\theta))}}{\Gamma(p\alpha(\infty) + nF_n(\theta))\Gamma((1-p)\alpha(\infty) + n(1-F_n(\theta)))}.$$

The proof of Theorem 3 consists of minor modifications in that of Theorem 1. Details are omitted.

The factor $K(X,\theta)$ is closely related to binomial probabilities. It has all the properties of the factor $M(X,\theta)$ mentioned in the second and third paragraphs of Section 4, except that its mode is at the pth quantile of the empirical c.d.f. F_n, in a way that the proposition below makes precise.

PROPOSITION. *Let* $j^* = [\![(n-1)p]\!] + 1$, *where* $[\![\;]\!]$ *denotes the integer part. Then* $K(X,\theta)$ *has a mode in the interval* $[X_{(j^*)}, X_{(j^*+1)})$ *and decreases as* θ *moves away from that interval in either direction.*

PROOF. For $j \in \{2,\dots,n\}$, we consider $K(X, X_{(j)})/K(X, X_{(j-1)})$. After simplification, we see that this is greater than or equal to 1 if and only if $j \le (n-1)p + 1$, and the assertion follows. □

In practice, if one has a hypothesized distribution G with pth quantile equal to ξ_p, the distribution α_0 would then be taken to be $\alpha_0(x) = G(x + \xi_p)$.

The posterior distribution of θ can be obtained without the assumption that α_0 have pth quantile equal to 0. However, its form becomes more complicated and more difficult to interpret.

Acknowledgment. I thank Professor Persi Diaconis for his valuable guidance. I am grateful to Fred Huffer and Tom Sellke for some interesting discussions, and to a referee and an Associate Editor for useful suggestions.

REFERENCES

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, New York.

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

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

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

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

DIACONIS, P. and FREEDMAN, D. (1986b). On inconsistent Bayes estimates of location. To appear in *Ann. Statist.* **13**.

DOKSUM, K. A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201.

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

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

DOSS, H. (1985). Bayesian nonparametric estimation of the median; Part II: asymptotic properties of the estimates. *Ann. Statist.* **12** 1445–1464.

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

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

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

JOHNS, M. V. (1979). Robust Pitman-like estimators. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds.) 46–60. Academic, New York.

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

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

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FL 32306