The Annals of Statistics
1989, Vol. 17, No. 4, 1550–1566

BAYES METHODS FOR A SYMMETRIC UNIMODAL DENSITY AND ITS MODE[1]

BY LAWRENCE J. BRUNNER AND ALBERT Y. LO

State University of New York at Buffalo

Bayes solutions for i.i.d. sampling from a decreasing density on the half line and for i.i.d. sampling from a symmetric unimodal density with an unknown mode are given. Posterior quantities are obtained as finite sums; Monte Carlo methods based on sampling finite Markov chains are developed for their evaluation.

1. Introduction. The problem of estimating an unknown symmetric unimodal density and its mode can be described as follows: We are given observations

$$X_i = \theta + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1.1}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent random variables having an unknown symmetric unimodal density f with mode at the origin; θ is an unknown location parameter. The object is to estimate θ and f based on $X_1, \ldots, X_n$.

This problem and the related one of estimating a decreasing density on the half line have been studied and discussed extensively from a frequentist viewpoint [Grenander (1956), Robertson (1967), Prakasa Rao (1969) and Wegman (1970)]; the results have been conveniently summarized in a text [Barlow, Bartholomew, Bremner and Brunk (1972)]. A Bayes solution to this problem is also desirable on at least three counts. First of all, Bayes methods allow the incorporation of prior information in statistical inference. Second, under natural identifiability and measurability conditions, Bayes estimates are consistent for almost all parameter values [Doob (1949)]. Third, Bayes procedures very often enjoy small-sample optimalities. Despite these advantages, Bayes solutions to the problem of estimating a symmetric unimodal density and its mode are unavailable. This paper is an attempt to provide such solutions.

A Bayesian approach to our problem consists of assuming a joint prior distribution on the pair (θ, f) and computing the posterior quantities of θ and f given $X_1, \ldots, X_n$. Our technique is based on the fact that any symmetric unimodal density is a scale mixture of symmetric uniform densities [Feller (1971), page 158]. A Dirichlet process prior [Ferguson (1973)] is assumed for the mixing distribution, a prior distribution is assumed for the mode and the posterior distribution of the pair (θ, f) is derived in closed form. Technically, this prior is concentrated on a space of symmetric unimodal densities (and a location parame-

Received July 1987; revised November 1988.
[1]This research was supported in part by NSF Grant MCS-81-02523-01.
AMS 1980 *subject classifications*. Primary 62A15, 62G05.
Key words and phrases. Bayes methods, symmetric unimodal densities, mode, simulation of a Markov chain, mixture models, Dirichlet process.

ter) and avoids the potholes encountered by workers using a raw Dirichlet process prior [for example, see Diaconis and Freedman (1986)].

A Bayes treatment of mixture models in general has been provided by one of the authors [Lo (1984, 1986)], and posterior means were found to be sums over all partitions of the first n (the sample size) integers. The difficulty of evaluating such sums over partitions is well known, and Kuo (1986) has suggested a Monte Carlo evaluation based on the Chinese restaurant process [Aldous (1985), page 92]. The present paper exploits the fine stucture of symmetric unimodal densities and obtains a simpler solution for this problem. The posterior mean is found to be a sum over all vectors $\mathbf{m} = (m_1, \ldots, m_n)$ where the m's are nonnegative integers, $\Sigma_{1 \le i \le j} m_i \le j$ for $j = 1, \ldots, n-1$ and $\Sigma_{1 \le i \le n} m_i = n$. The structure of this sum resembles its relative, the posterior mean of a monotone rate function in a reliability model reported by Dykstra and Laud (1981) and in Aalen's (1978) martingale-based counting process model discussed by Lo and Weng (1989).

Even though summing over all possible $\mathbf{m}$-vectors is simpler than summing over all partitions of the set $\{1, \ldots, n\}$, the evaluation of the former sum for sample sizes larger than 13 is still a formidable problem [the same comment obviously applies to the evaluation of the posterior means of monotone rate functions studied by Dykstra and Laud (1981) and Lo and Weng (1989)]. Hence, the development of numerical methods to evaluate sums over $\mathbf{m}$-vectors for sample sizes larger than 13 is important both for the model considered here and in the reliability or point process model; this development is also of practical interest. Section 5 develops Monte Carlo approximations for sums over all m-vectors based on sampling $(n-1)$-term Markov chains; a uniform distribution on the space of all $\mathbf{m}$-vectors is introduced to facilitate the approximations.

Section 2 is the backbone of our method. In this section, i.i.d. sampling from an unknown decreasing density on the half line (or equivalently, from a symmetric unimodal density with a known mode) is considered. The posterior distribution is found to be a mixture of Dirichlet processes, and the posterior mean of the decreasing density is given in a closed form as a sum over the $\mathbf{m}$-vectors. This sum can then be separated into two parts. The first part is a decreasing step function which jumps at the data sites, a property also shared by the maximum likelihood estimate reported by Grenander (1956) and Wegman (1970). The second part is a mixture of inverse moment functions which contributes to the smooth property of the posterior mean and which may not have a frequentist derivation.

Section 3 discusses the choice of the shape probability of the Dirichlet process based on the idea of conjugate priors [DeGroot (1970)]. A straightforward application of this concept suggests the choice of a Pareto distribution to be the shape probability. However, a moderate or large location index for the Pareto distribution results in a loss of observed data, while a small index gives rise to a bad spike for the posterior mean of the density at abscissas close to the origin; we call this the *Pareto syndrome* [the maximum likelihood estimate is also known to have a peak close to the origin; see Wegman (1970)]. We find that a posterior mean based on an inverted gamma shape probability is spike-free and

is more palatable. The section is concluded with plots of the Monte Carlo evaluations of the posterior means of decreasing densities corresponding to Pareto and inverted gamma shape probabilities.

Bayes procedures for an unknown symmetric unimodal density and its mode based on i.i.d. sampling are studied in Section 4. Section 4.1 discusses the posterior density of the mode, which is a sum over all **m**-vectors (the symmetric unimodal density is treated as a nuisance parameter); we present Monte Carlo evaluations of several posterior densities of the mode based on nested samples. Section 4.2 discusses the posterior distribution of the symmetric unimodal density (here the mode is a nuisance parameter); we display Monte Carlo evaluations of several posterior means of the density based on nested samples.

2. Bayes methods for a decreasing density on the half line. This section is concerned with the Bayes estimation of a decreasing density on the half line. It is well known [Feller (1971), page 158] that any such density f corresponds to a distribution G on $[0, \infty)$ such that

$$f(x|G) = \int v^{-1} I_{\{0 \le x < v\}} G(dv). \tag{2.1}$$

The collection of G's is denoted by Θ, which serves as the parameter space for this model. Note that the map $G \to f$ is one to one, and so the model $\{f(\cdot|G): G \in \Theta\}$ is identifiable. A result of Doob (1949) states that the Bayes procedures for the model (2.1) are consistent for almost all parameter values. Assume a Dirichlet process prior with shape measure α on G, denoted by $P(dG|\alpha)$, and given G, let $X_1, \ldots, X_n$ be i.i.d. random variables from the decreasing density $f(x|G)$. A result of Lo (1984), Theorem 1, specializes and states that the posterior distribution of G is a mixture of Dirichlet processes defined by, for all nonnegative (and measurable) functions h,

$$\int_\Theta h(G) P(dG|X) = \frac{\int_{R^n} \int_\Theta h(G) P(dG|\alpha + \Sigma_i \delta_{v_i}) \mu_n(d\mathbf{v}|\mathbf{X})}{\mu_n(R^n|X)}, \tag{2.2}$$

where $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{v} = (v_1, \ldots, v_n)$, Σ_i denotes $\Sigma_{1 \le i \le n}$ and the measure μ_n on $\mathscr{B}^n$ is defined by

$$\mu_n(C|\mathbf{X}) = \int_C \left[\prod_i v_i^{-1} I_{\{0 \le X_i < v_i\}} \right] \prod_i \left(\alpha + \sum_{1 \le j \le i-1} \delta_{v_j} \right)(dv_i) \tag{2.3}$$

for any $C \in \mathscr{B}^n$, where Π_i denotes $\Pi_{1 \le i \le n}$.

The above posterior distribution summarizes a statistician's posterior opinion of the parameter G and, from a Bayes viewpoint, is the estimate of the unknown parameter. However, due to the complexity of the posterior distribution, a discussion of some posterior quantities (especially the posterior mean) could be revealing and in fact is necessary. In the rest of this section, we derive the posterior mean of the mixing distribution $G(u)$ and the decreasing density $f(t|G)$, which could be used to estimate the mixing distribution and the decreas-

ing density, respectively. The derivations of the corresponding posterior variances and higher moments proceed similarly and will not be repeated. The posterior means turn out to be mixtures of "posterior" quantities when sampling from the family of uniform densities $\{v^{-1}I_{[0 \le x < v\}}: 0 < v < \infty\}$, where v serves as a parameter which has the "prior" distribution $\alpha(dv)/\alpha(R)$. The following combinatoric result forms the basis of our derivation. First we need some notation.

Let $\mathbf{p}$ be a partition of $\{1, \ldots, n\}$, $\{C(i): i = 1, \ldots, n(\mathbf{p})\}$ be the cells of the partition $\mathbf{p}$, e_i be the number of elements in $C(i)$ and $k(\mathbf{p}) = \prod_{1 \le i \le n(p)}(e_i - 1)!$. Let $0 < x(1) < x(2) < \cdots < x(n) < \infty$ be the ordered values of n distinct real numbers $x_1, \ldots, x_n$ and $\max(i) = \max\{x_k: k \in C(i)\}$. Let m_i be nonnegative integers such that $\sum_{1 \le i \le j} m_i \le j$ for $1 \le j \le n - 1$ and $\sum_{1 \le i \le n} m_i = n$. Let $\mathbf{m} = (m_1, \ldots, m_n)$, $s_j = \sum_{1 \le i \le j} m_i$ and $k^*(\mathbf{m}) = \prod_{i^*}(i - 1 - s_{i-1})!/(i - s_i)!$, where the product $\prod_{i^*}$ is over the set of i such that $m_i \ge 1$.

LEMMA 2.1. *For any nonnegative function h,*

$$\int_{R^n}\left[\prod_i h(v_i) I_{\{0 \le x_i < v_i\}}\right]\prod_i\left(\alpha + \sum_{1 \le j \le i-1}\delta_{v_j}\right)(dv_i)$$

$$= \sum_{\mathbf{p}} k(\mathbf{p}) \prod_{1 \le i \le n(\mathbf{p})} \int [h(v)]^{e_i} I_{\{0 \le \max(i) < v\}} \alpha(dv) \tag{2.4}$$

$$= \sum_{\mathbf{m}} k^*(\mathbf{m}) \prod_{i^*} \int [h(v)]^{m_i} I_{\{0 \le x(i) < v\}} \alpha(dv);$$

in particular,

$$\Gamma(\alpha(R) + n)/\Gamma(\alpha(R)) = \sum_{\mathbf{p}} \alpha(R)^{n(\mathbf{p})} k(\mathbf{p}) = \sum_{\mathbf{m}} \alpha(R)^{\#(\mathbf{m})} k^*(\mathbf{m}), \tag{2.5}$$

where $\#(\mathbf{m})$ is the number of $m_i \ge 1$ and $\Gamma(\cdot)$ is the gamma function.

PROOF. This lemma follows from combinatorial results in Dykstra and Laud (1981) and Lo (1984). For clarity, the following direct proof is inserted. The first expression reduces to the second one by Lemma 2 in Lo (1984); see also Ghorai and Rubin (1982). That the second expression reduces to the third one can be seen as follows. Relabel the cells $\{C(i)\}$ by letting C_i be the cell whose maximum element is the integer i; the number of elements in C_i is m_i. Note that each of the summands in $\Sigma_{\mathbf{p}}$ depends on the partition $\mathbf{p}$ only through $\mathbf{m}$ and, for a given $\mathbf{m}$, it reduces to

$$\prod_{i^*}(m_i - 1)! \int [h(v)]^{m_i} I_{\{0 \le x(i) < v\}} \alpha(dv). \tag{2.6}$$

It remains to show that the number of partitions corresponding to each $\mathbf{m}$ is $k^*(\mathbf{m})/\prod_{i^*}(m_i - 1)!$. If $m_1 \ge 1$, the possible number of elements in C_1 is 1. Assume that $m_1, \ldots, m_{i-1}$ are determined for $i \ge 2$. If $m_i \ge 1$, the possible number of elements in C_i is equal to the binomial

coefficient $C(i-1-s_{i-1}, m_i-1)$. The number of partitions corresponding to $\mathbf{m}$ is then

$$\prod_{i^*} C(i-1-s_{i-1}, m_i-1) = k^*(\mathbf{m}) \Big/ \prod_{i^*} (m_i-1)!. \tag{2.7}$$

The equalities in (2.5) follow from (2.4) by equating the integrands to 1. □

The next theorem gives the posterior means of G and f as sums over the $\mathbf{m}$-vectors.

THEOREM 2.1. *For each* $x > 0$,

$$\begin{aligned}\int_\Theta f(x|G)P(dG|\mathbf{X}) &= [\alpha(R)/(\alpha(R)+n)]\,d_0(x) \\ &\quad + [n/(\alpha(R)+n)] \sum_{\mathbf{m}} W(\mathbf{m}) \sum_i (m_i/n) d_i(x|\mathbf{m}),\end{aligned} \tag{2.8}$$

where $W(\mathbf{m}) = \varphi(\mathbf{m})/\Sigma_{\mathbf{m}}\varphi(\mathbf{m})$, $\varphi(\mathbf{m}) = k^*(\mathbf{m})\prod_{i^*}\int v^{-m_i} I_{\{X(i)<v\}}\alpha(d\mathbf{v})$,

$$d_0(x) = \int v^{-1} I_{\{0 \le x < v\}} \alpha(dv)/\alpha(R), \tag{2.9}$$

$$d_i(x|\mathbf{m}) = \frac{\int v^{-(m_i+1)} I_{\{\max[x,\, X(i)]<v\}}\alpha(dv)}{\int v^{-m_i} I_{\{X(i)<v\}}\alpha(dv)}. \tag{2.10}$$

The posterior mean of $G(u)$ *is obtained by replacing* $d_0(x)$ *by* $G_0(u) = \alpha(u)/\alpha(R)$, *and* $d_i(x|\mathbf{m})$ *by*

$$G_i(u|\mathbf{m}) = \frac{\int v^{-m_i} I_{\{X(i)<v\le u\}}\alpha(dv)}{\int v^{-m} i I_{\{X(i)<v\}}\alpha(dv)} I_{\{X(i)\le u\}}. \tag{2.11}$$

PROOF. We compute the posterior mean of $f(x|G)$ only; the posterior mean of $G(u)$ can be derived similarly. Put $h(G) = f(x|G) = \int v^{-1} I_{\{0\le x<v\}} G(dv)$ in (2.2). We obtain

$$\begin{aligned} E[f(x|G)|\mathbf{X}] &= \frac{\int_{R^n}\int_\Theta \int v^{-1} I_{\{0\le x<v\}} G(dv) P\big(dG|\alpha + \Sigma_i \delta_{v_i}\big)\mu_n(d\mathbf{v}|\mathbf{X})}{\mu_n(R^n|\mathbf{X})} \\ &= \frac{\mu_{n+1}(R^{n+1}|\mathbf{X})}{[\alpha(R)+n]\mu_n(R^n|\mathbf{X})} \\ &= \frac{\Sigma_{\mathbf{m}}\varphi(\mathbf{m})\int v^{-1} I_{\{0\le x<v\}}\alpha(dv) + \Sigma_{\mathbf{m}}\varphi(\mathbf{m})\Sigma_i m_i d_i(x|\mathbf{m})}{[\alpha(R)+n]\Sigma_{\mathbf{m}}\varphi(\mathbf{m})}, \end{aligned} \tag{2.12}$$

where the second equality is obtained by regarding x as X_{n+1}, the last equality by applying Lemma 2.1 to the numerator and the denominator separately, and then simplifying. □

The following alternative representation of d_i is more interesting:

$$d_i(x|\mathbf{m}) = \frac{\int v^{-(m_i+1)} I_{\{X(i)<v\}} \alpha(dv)}{\int v^{-m_i} I_{\{X(i)<v\}} \alpha(dv)} I_{\{x<X(i)\}} + \frac{\int v^{-(m_i+1)} I_{\{x<v\}} \alpha(dv)}{\int v^{-m_i} I_{\{X(i)<v\}} \alpha(dv)} I_{\{X(i)\le x\}}. \tag{2.13}$$

Each d_i is a constant function on $[0, X(i)]$; beyond this interval, it decreases to zero as an inverse moment function. Substituting (2.13) into the expression (2.8) results in a posterior mean of the decreasing density as a sum of two parts. The first part is the result of a mixture of the first term in (2.13) and is a decreasing step function which decreases only at the data sites. It is quite similar to the step-wise maximum likelihood estimate of a decreasing density reported by Grenander (1956) [see also Wegman (1970)]. The second part is a mixture of inverse moment functions and serves as a smoothing device.

Let us consider the problem of choosing $\alpha(R)$ and the shape probability $\alpha/\alpha(R)$. The size of $\alpha(R)$ reflects the strength of our prior belief in the shape probability and can be regarded as the prior sample size.

The determination of the shape probability might proceed as follows. Suppose the prior data result in a decreasing step density

$$\pi_0(x) = \sum_{1\le k<\infty} w_k I_{\{b_{k-1}\le x<b_k\}}, \tag{2.14}$$

where $0 < w_{j+1} < w_j$ for all j and $b_0 = 0$. The previously mentioned result in Feller (1971) implies that $\pi_0(x)$ is a scale mixture of uniform densities with a discrete mixing distribution $\mathbf{P} = \sum_{1\le k<\infty} p_k \delta_{b_k}$, given by

$$\int v^{-1} I_{\{0<x<v\}} \mathbf{P}(dv) = \sum_{1\le k<\infty} b_k^{-1} I_{\{0<x<b_k\}} p_k. \tag{2.15}$$

Hence, $p_k = b_k(w_k - w_{k+1})$ for $k = 1, 2, \ldots$. Clearly, $\mathbf{P}$ is the shape probability. Similar considerations may be used to determine the shape probability in Section 4.

REMARK 2.1. The problem of estimating a symmetric unimodal density with mode at the origin can be reduced to that of estimating a decreasing density on the half line by noting that Y has such a symmetric density, $X = |Y|$ is a sufficient statistic and has a decreasing density on the half line. The preceding results apply.

3. The Pareto syndrome and a gamma alternative. The idea of conjugate priors [DeGroot (1970), page 172] suggests that letting the shape probability $\alpha/\alpha(R)$ be a Pareto (α_0, r_0) distribution in Theorem 2.1 would lead to an integral-free (2.8). This choice of α does simplify the expressions somewhat and the resulting posterior mean is computationally efficient. Indeed, the usual prior-to-posterior updating step in sampling a uniform density prevails here. For

example, $\alpha/\alpha(R)$ is Pareto (α_0, r_0) implies that G_0 is Pareto (α_0, r_0) and d_0 is a mixture of a uniform $(0, r_0)$ density and a Pareto (α_0, r_0) density with mixing weights $\alpha_0/(\alpha_0 + 1)$ and $1/(\alpha_0 + 1)$, respectively. The corresponding $G_i(\cdot|\mathbf{m})$'s and $d_i(\cdot|\mathbf{m})$'s can be obtained by updating α_0 to $\alpha_i = \alpha_0 + m_i$ and r_0 to $r_i = \max\{r_0, X(i)\}$. The form of the d's is worth noting:

$$
\begin{aligned}
d_0(x|\mathbf{m}) &= \frac{\alpha_0(r_0)^{\alpha_0}}{(\alpha_0 + 1)[\max(x, r_0)]^{\alpha_0+1}} \\
&= \frac{\alpha_0}{(\alpha_0 + 1)} \times \frac{1}{r_0} I_{\{0 \le x < r_0\}} + \frac{1}{(\alpha_0 + 1)} \times \frac{\alpha_0(r_0)^{\alpha_0}}{x^{\alpha_0+1}} I_{\{r_0 \le x\}}.
\end{aligned} \tag{3.1}
$$

The role of the location index r_0 is critical. First of all, it washes away the effect of any observed $X(i)$ which is less than r_0. Hence a moderate-to-large r_0 is not desirable [this "Pareto syndrome" is very well known in the case of sampling a parametric uniform distribution; see the above reference to DeGroot (1970)]. One way out is to choose a small r_0, yet a small r_0 invites a large spike on $[0, r_0]$ and is not palatable. Figure 1 illustrates the association between r_0 and the size of the spike when $\alpha_0 = 1$. The computation employs the Monte Carlo method developed in Section 5.

The foregoing discussion indicates that the choice of a Pareto (α_0, r_0) shape probability results in a trade-off between a loss of data (large r_0) and a bad spike close to the origin (small r_0); a different shape probability $\alpha/\alpha(R)$ might perform better. An inspection of (2.8)–(2.11) suggests that the next candidate is an inverted gamma (α_0, β_0) distribution and, in this regard, the following reparametrization of the submodel $\{v^{-1}I_{\{0 \le t < v\}}\colon\ 0 < v < \infty\}$ is convenient. Upon reparametrization, the d's can be rewritten as

$$
\begin{aligned}
d_0(x|\mathbf{m}) &= \int y I_{\{0 < y < 1/x\}} \alpha^*(dy), \\
d_i(x|\mathbf{m}) &= \frac{\int y^{(m_i+1)} I_{\{0 < y < \min[1/x,\, 1/X(i)]\}} \alpha^*(dy)}{\int y^{m_i} I_{\{0 < y < 1/X(i)\}} \alpha^*(dy)} \\
&= \frac{\int y^{(m_i+1)} I_{\{0 < y < 1/X(i)\}} \alpha^*(dy)}{\int y^{m_i} I_{\{0 < y < 1/X(i)\}} \alpha^*(dy)} I_{\{0 \le x < X(i)\}} \\
&\quad + \frac{\int y^{(m_i+1)} I_{\{0 < y < 1/x\}} \alpha^*(dy)}{\int y^{m_i} I_{\{0 < y < 1/X(i)\}} \alpha^*(dy)} I_{\{X(i) \le x\}},
\end{aligned} \tag{3.2}
$$

where α^* is a probability on the half line and $\alpha/\alpha(R)$ is the "inverted" distribution corresponding to α^* in the sense that $v \sim \alpha/\alpha(R)$ is equivalent to $y = v^{-1} \sim \alpha^*$.

Let α^* be a gamma (α_0, β_0) distribution. $\alpha/\alpha(R)$ is then called an inverted gamma (α_0, β_0) shape probability. Denote the cumulative distribution function

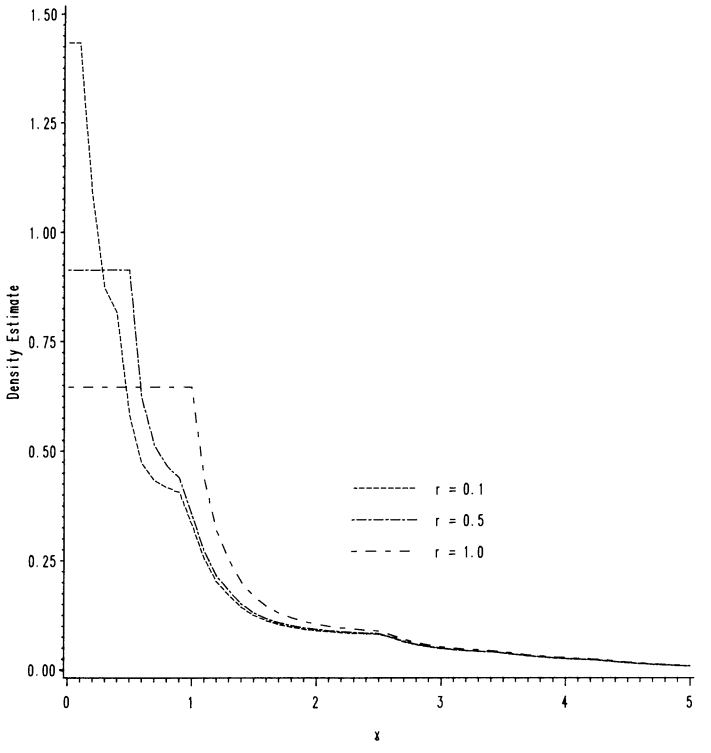

FIG. 1. *The Pareto syndrome. Monte Carlo simulation of* (2.12) *based on* 200 *uniform* $\mathbf{m}$ *vectors and* $n = 50$ *observations from a standard exponential density.* $\alpha/\alpha(R)$ *is Pareto* (α_0, r_0) *with* $\alpha_0 = 1.0$ *and* $r_0 = 0.1, 0.5$ *and* 1.0; $\alpha(R) = 1.0$.

of a gamma (a, b) random variable evaluated at t by $\gamma(t|a, b)$. Then

$$
\begin{aligned}
d_0(x|\mathbf{m}) &= (\alpha_0/\beta_0)\gamma\big(1/x|\alpha_0, \beta_0\big), \\
d_i(x|\mathbf{m}) &= \frac{(\alpha_i/\beta_0)\gamma\big(1/X(i)|\alpha_0 + m_i + 1, \beta_0\big)}{\gamma\big(1/X(i)|\alpha_0 + m_i, \beta_0\big)} I_{\{x < X(i)\}} \\
&\quad + \frac{(\alpha_i/\beta_0)\gamma\big(1/x|\alpha_0 + m_i + 1, \beta_0\big)}{\gamma\big(1/X(i)|\alpha_0 + m_i, \beta_0\big)} I_{\{X(i) \le x\}}.
\end{aligned} \tag{3.3}
$$

The d_i is constant on the interval $[0, X(i)]$ and then tapers off very smoothly to zero beyond this interval. Note the spike-free nature of the second term.

Figure 2 uses the Monte Carlo method developed in Section 5 to plot posterior means of a decreasing density with respect to Pareto and inverted gamma shape probabilities (i.e., α^* is a gamma distribution). The sample consists of 50 observations from a standard exponential density. Notice the lack of spike when an inverted gamma shape probability is used.

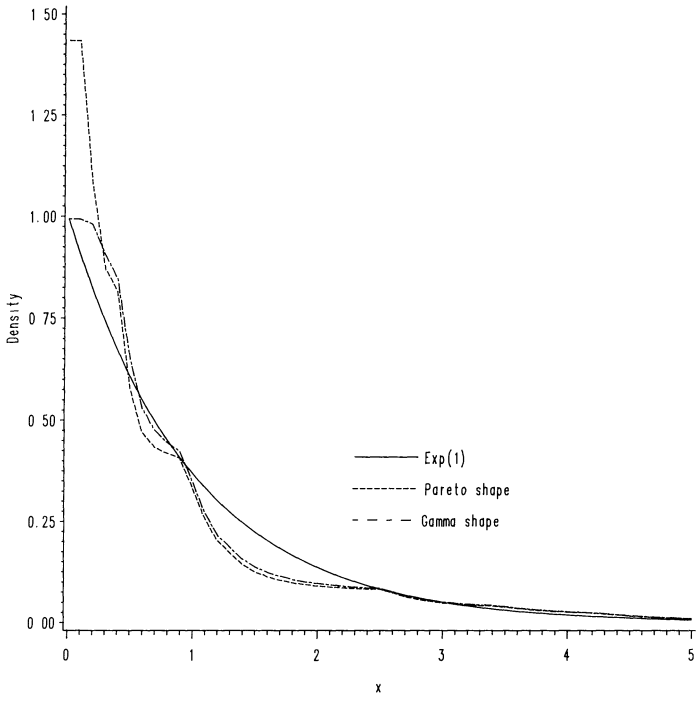

FIG. 2. *Estimates of decreasing density: Comparison of Pareto and inverted gamma shape probabilities. Monte Carlo simulation of* (2.12) *based on* 200 *uniform* $\mathbf{m}$ *vectors and* $n = 50$ *observations from a standard exponential density. Pareto* (α_0, r_0) *shape probability has* $\alpha_0 = 1.0$ *and* $r_0 = 0.1$. *Inverted gamma* (α_0, β_0) *shape probability has* $\alpha_0 = 1.0$ *and* $\beta_0 = 1.0$. $\alpha(R) = 1.0$ *in both cases.*

Next, we turn to the question of deflating the prior. Ferguson's posterior mean of a distribution function [Ferguson (1973)] has the interesting property that letting $\alpha(R) \to 0$ reduces it to the maximum likelihood estimate of the distribution function (i.e., the empirical distribution function). Lo (1986) has shown that letting $\alpha(R) \to 0$ or $\alpha(R) \to \infty$ does not yield useful estimates in mixture models. In particular, if $\alpha(R) \to 0$, the dominating term in (2.8), i.e., $\Sigma_{\mathbf{m}} W(\mathbf{m}) \Sigma_i (m_i/n) d_i(x|\mathbf{m})$, tends to the posterior mean of a $U(0, \theta)$ density based on a sample $\{X_1, \ldots, X_n\}$, obtained by i.i.d. sampling from the parametric model $\{U(0, \theta)\colon \theta > 0\}$ with respect to the prior $\alpha(d\theta)/\alpha(R)$. On the other hand, if $\alpha(R) \to \infty$, this quantity tends to the average of the same posterior quantities based on samples of size 1; neither limiting estimate is useful. In contrast, deflating the shape probability $\alpha/\alpha(R)$ while keeping $\alpha(R)$ fixed yields nontrivial estimates. These limiting estimates are not, in general, the maximum likelihood estimate. For example, in the case of Pareto shape probability, the limiting estimate is not constant between adjacent order statistics.

4. Bayes methods for a symmetric unimodal density and its mode. This section provides Bayes solutions for a symmetric unimodal density with an

unknown mode. The random variable X has a symmetric unimodal density with a mode at θ if and only if $|X - \theta|$ has a decreasing density on the positive line. Therefore, according to (2.1), the model can be defined as follows:

(4.1) Given the pair (θ, G), $X_1, \ldots, X_n$ are i.i.d. with density given by $f(x|\theta, G) = 2^{-1}\int v^{-1} I_{\{|x-\theta<v\}} G(dv)$.

where θ is the unknown mode and G is an unknown probability on the half line. The parameter space Θ is the collection of (θ, G) pairs. Note that since the map $(\theta, G) \to f$ is one to one, the model $\{f(\cdot|\theta, G)\colon (\theta, G) \in \Theta\}$ is identifiable. Again, Doob's (1949) result guarantees consistent Bayes procedures for almost all (θ, G).

Assume any prior on the pair (θ, G). The usual double expectation formula

$$(4.2)\quad E[h(\theta, G)|\mathbf{X}] = \iint h(\theta, G) P(dG|\theta, \mathbf{X}) P(d\theta|\mathbf{X}) \quad \text{for any positive } h,$$

determines the posterior distribution of the pair (θ, G). Suppose

(4.3) the prior distribution of θ is $\pi(d\theta)$, and given θ, the prior distribution of G is a Dirichlet process $P(dG|\alpha_\theta)$

where α_θ is the shape measure. Then $P(dG|\theta, \mathbf{X})$ is defined by (2.2), with α and X_i replaced by α_θ and $|X_i - \theta|$, respectively. It remains to identify the posterior distribution of the mode $P(d\theta|\mathbf{X})$ in (4.2); as a consequence of (4.3), this distribution will be denoted by $P_\pi(d\theta|\mathbf{X})$.

4.1. *Posterior distribution of the mode.* The posterior distribution of the mode θ has an explicit form, which can be derived using the technique developed in Section 2.

THEOREM 4.1. *Let the prior and the model be defined by* (4.3) *and* (4.1). *The posterior distribution of* θ, *given* $\mathbf{X} = (X_1, \ldots, X_n)$, *is defined by*

$$(4.4)\qquad P_\pi(A|\mathbf{X}) = \frac{\sum_{\mathbf{m}} k^*(\mathbf{m}) \int_A \prod_{i^*} \int v^{-m_i} I_{\{X(i,\theta)<v\}} \alpha_\theta(dv) \pi(d\theta)}{\sum_{\mathbf{m}} k^*(\mathbf{m}) \int \prod_{i^*} \int v^{-m_i} I_{\{X(i,t)<v\}} \alpha_t(dv) \pi(dt)},$$

where $0 < X(1, \theta) < \cdots < X(n, \theta) < \infty$ *are the ordered values of* $|X_i - \theta|$, $i = 1, \ldots, n$, *and* A *is any Borel set on the line.*

PROOF. Note that G is a nuisance parameter to be averaged out by the conditional distribution of G given θ, i.e., $P(dG|\alpha_\theta)$. This integration can be carried out by repeatedly applying Lemma 1 in Lo (1984) and then Lemma 2.1, resulting in the following conditional density of $\mathbf{X}$ given θ:

$$(4.5)\quad \begin{aligned} &[2^n \Gamma(\alpha_\theta(R) + n)]^{-1} \int_{R^n} \left[\prod_i v_i^{-1} I_{\{|X_i-\theta|<v_i\}}\right] \prod_i \left(\alpha_\theta + \sum_{1 \le j \le i-1} \delta_{v_j}\right)(dv_i) \\ &\quad = [2^n \Gamma(\alpha_\theta(R) + n)]^{-1} \sum_{\mathbf{m}} k^*(\mathbf{m}) \prod_{i^*} \int v^{-m_i} I_{\{X(i,\theta)<v\}} \alpha_\theta(dv). \end{aligned}$$

Theorem 4.1 follows from the last equality and the assumption that the prior distribution of θ is $\pi(d\theta)$. □

Expression (4.4) can be simplified slightly by assuming that the index measure α_θ is independent of θ and that θ has a prior density $\pi'(\theta)$ with respect to Lebesgue measure. Then the posterior density of θ, $p_\pi(\theta|\mathbf{X})$, is given by

$$p_\pi(\theta|\mathbf{X}) = \frac{\Sigma_{\mathbf{m}} k^*(\mathbf{m})\pi'(\theta)\Pi_{i^*}\int v^{-m_i} I_{\{X(i,\theta)<v\}}\alpha(dv)}{\Sigma_{\mathbf{m}} k^*(\mathbf{m})\int \pi'(t)\left\{\Pi_{i^*}\int v^{-m_i} I_{\{X(i,t)<v\}}\alpha(dv)\right\} dt}. \tag{4.6}$$

For a flat prior on θ, i.e., $\pi'(\theta) = 1$, the posterior density reduces to

$$p(\theta|\mathbf{X}) = \frac{\Sigma_{\mathbf{m}} k^*(\mathbf{m})\Pi_{i^*}\int v^{-m_i} I_{\{X(i,\theta)<v\}}\alpha(dv)}{\Sigma_{\mathbf{m}} k^*(\mathbf{m})\int\left\{\Pi_{i^*}\int v^{-m_i} I_{\{X(i,t)<v\}}\alpha(dv)\right\} dt}. \tag{4.7}$$

The idea of a conjugate prior suggests a Pareto (α_0, r_0) distribution for the shape probability $\alpha/\alpha(R)$. Expression (4.7) becomes

$$p(\theta|\mathbf{X}) = \frac{\Sigma_{\mathbf{m}}\alpha(R)^{\#(\mathbf{m})}k^*(\mathbf{m})\Pi_{i^*}r_0^{\alpha_0}\max\{r_0, X(i,\theta)\}^{-(\alpha_0+m_i)}\alpha_0/(\alpha_0+m_i)}{\Sigma_{\mathbf{m}}\alpha(R)^{\#(\mathbf{m})}k^*(\mathbf{m})\int\Pi_{i^*}r_0^{\alpha_0}\max\{r_0, X(i,t)\}^{-(\alpha_0+m_i)}\alpha_0/(\alpha_0+m_i)\,dt}. \tag{4.8}$$

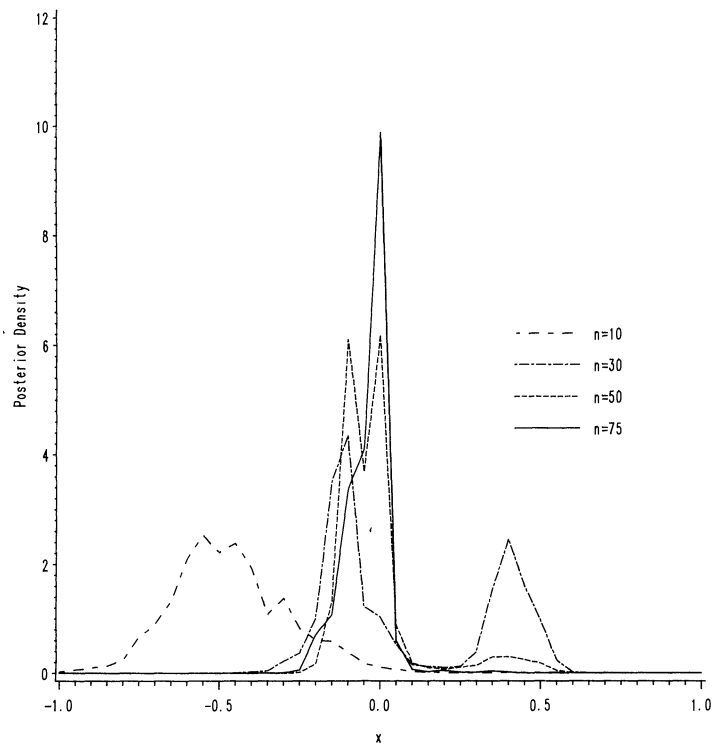

FIG. 3. *Posterior density of the mode: Pareto shape probability. Monte Carlo simulation of* (4.8) *based on* 200 *uniform* $\mathbf{m}$ *vectors and* $n = 10$, 30, 50 *and* 75 *nested observations from a standard Cauchy density.* $\alpha/\alpha(R)$ *is Pareto* (α_0, r_0) *with* $\alpha_0 = 1.0$ *and* $r_0 = 0.1$; $\alpha(R) = 1.0$.

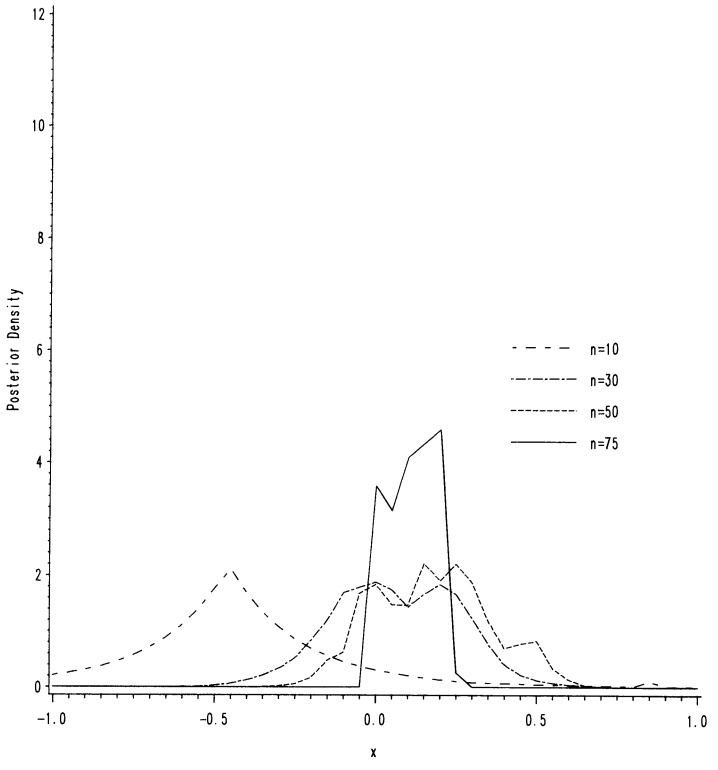

FIG. 4. *Posterior density of the mode: Inverted gamma shape probability. Monte Carlo simulation of* (4.9) *based on* 200 *uniform* $\mathbf{m}$ *vectors and* $n = 10, 30, 50$ *and* 75 *nested observations from a standard Cauchy density.* $\alpha/\alpha(R)$ *is inverted gamma* (α_0, β_0) *with* $\alpha_0 = 1.0$ *and* $\beta_0 = 1.0$; $\alpha(R) = 1.0$.

In the reparametrized form with an inverted gamma (α_0, β_0) shape probability, the posterior density of the mode is given by

$$p(\theta|\mathbf{X}) = \left\{\Sigma_{\mathbf{m}}\alpha(R)^{\#(\mathbf{m})}k^*(\mathbf{m})\Pi_{i^*}\gamma(1/X(i,\theta)|\alpha_0 + m_i, \beta_0)\frac{\Gamma(\alpha_0 + m_i)}{\Gamma(\alpha_0)\beta_0^{m_i}}\right\} \tag{4.9}$$
$$\div \left\{\Sigma_{\mathbf{m}}\alpha(R)^{\#(\mathbf{m})}k^*(\mathbf{m})\int\Pi_{i^*}\gamma(1/X(i,t)|\alpha_0 + m_i, \beta_0)\frac{\Gamma(\alpha_0 + m_i)}{\Gamma(\alpha_0)\beta_0^{m_i}}\,dt\right\}.$$

It is tempting to "deflate" the prior, i.e., taking limits for α_0, r_0 and β_0. Yet this limiting procedure does not lead to simplified and reasonable expressions; for example, if $r_0 \to 0$ in (4.8), the denominator blows up to infinity.

Figure 3 plots the Monte Carlo approximation (developed in Section 5) of (4.8), which is the posterior density based on a Pareto shape probability. For comparison, Figure 4 plots the approximation of (4.9). Nested samples of sizes 10, 30, 50 and 75 from a standard Cauchy density are used to show a peaking posterior density.

REMARK 4.1. The fact that the choice of a Pareto shape probability invariably results in a posterior density with a more profound (and desirable) peak at the true mode is worth noting. This phenomenon does not contradict the results in the last section in which the defect of a Pareto shape probability is discussed. Instead, it suggests that estimating a symmetric density and estimating its mode are "independent" affairs and what is a disadvantage for one could very well be an advantage for the other.

4.2. *Posterior mean of the symmetric unimodal density.* Posterior quantities of the model (4.1) with respect to the prior (4.3) can be obtained by appropriate choices of $h(\theta, G)$ in (4.2). In particular, one can obtain the posterior mean of the symmetric unimodal density evaluated at x by letting $h(\theta, G) = f(x|\theta, G)$. In this case, the inner integral $\int f(x|\theta, G)P(dG|\theta, \mathbf{X})$ in (4.2) is given by (2.8) multiplied by 0.5, with $X(i)$ replaced by $X(i, \theta)$, x replaced by $|x - \theta|$ and α

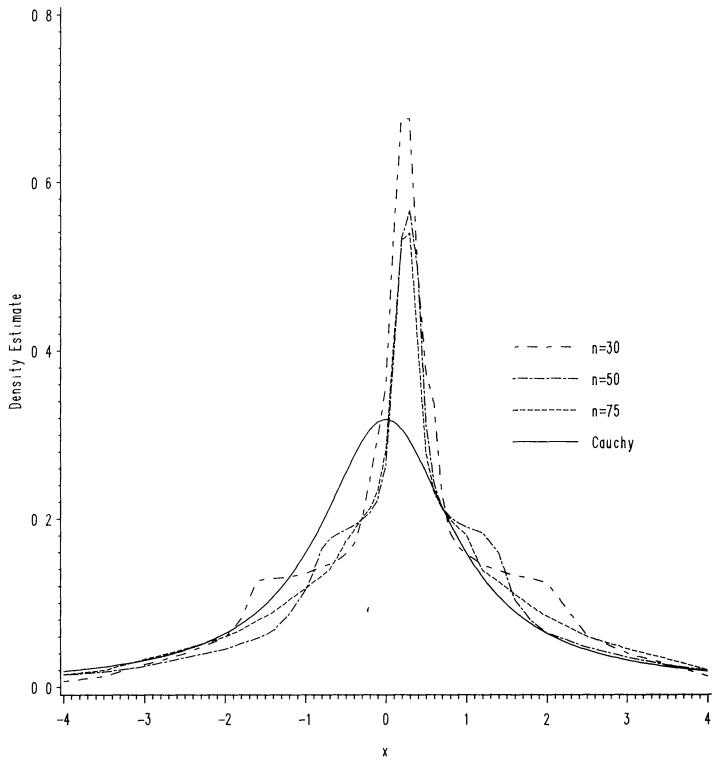

FIG. 5. *Estimate of symmetric unimodal density: Pareto shape probability. Monte Carlo simulation of* (4.2) *with* $h(\theta, G) = f(x|\theta, G)$, *based on* 200 *uniform* $\mathbf{m}$ *vectors and* $n = 30$, 50 *and* 75 *nested observations from a standard Cauchy density.* $\alpha/\alpha(R)$ *is Pareto* (α_0, r_0) *with* $\alpha_0 = 1.0$ *and* $r_0 = 0.1$; $\alpha(R) = 1.0$.

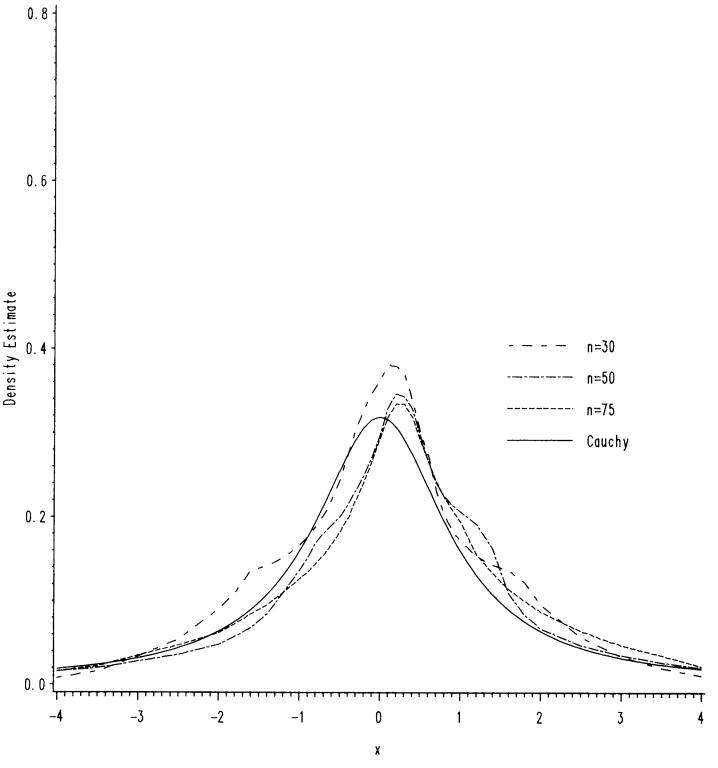

FIG. 6. *Estimate of symmetric unimodal density: Inverted Gamma shape probability. Monte Carlo simulation of* (4.2) *with* $h(\theta, G) = f(x|\theta, G)$, *based on* 200 *uniform* $\mathbf{m}$ *vectors and* $n = 30, 50$ *and* 75 *nested observations from a standard Cauchy density.* $\alpha/\alpha(R)$ *is inverted gamma* (α_0, β_0) *with* $\alpha_0 = 1.0$ *and* $\beta_0 = 1.0$; $\alpha(R) = 1.0$.

replaced by α_θ. $P_\pi(d\theta|\mathbf{X})$ is given by Theorem 4.1. Therefore, $E[\, f(x|\theta, G)|\mathbf{X}]$ is a two-stage sum over $\mathbf{m}$-vectors and can be evaluated by the Monte Carlo method developed in the next section.

We conclude this section by a Monte Carlo evaluation of a posterior mean of the symmetric unimodal density based on a uniform distribution on the $\mathbf{m}$'s developed in Section 5. For simplicity, α_θ is assumed to be independent of θ. A Pareto shape probability for the prior results in Figure 5, where $P_\pi(d\theta|\mathbf{X})$ is given by (4.8). Since $\pi'(\theta) = 1$ for all θ, the prior mean of the random density is the constant 1. Note the presence of the Pareto syndrome in Figure 5. For comparison, an inverted gamma shape probability results in Figure 6; $P_\pi(d\theta|\mathbf{X})$ is given by (4.9). Nested samples of sizes 30, 50 and 75 from a standard Cauchy density are employed to indicate the convergence of posterior quantities. The presence of outliers in a Cauchy sample for a moderate-to-large sample size ($n = 75$) retards the convergence, resulting in a plot shifted to the right. This phenomenon does not prevail when sampling from light-tailed densities.

5. Numerical evaluation of the posterior quantities. The evaluation of the posterior quantities obtained in the previous sections is based on computing sums of the form $\Sigma_{\mathbf{m}} g(\mathbf{m})$ where the g's are known functions of $\mathbf{m}$. A sum of the form $\Sigma_{\mathbf{m}} g(\mathbf{m})$ was computed exactly by Dykstra and Laud (1981) for a sample size (n) of 10. However, we must point out that as far as exact computation is concerned, the case of $n = 12$ or 13 is the limit. The recursive formula (5.3), which follows, can be used to generate the numbers of $\mathbf{m}$-vectors corresponding to different sample sizes. For example:

n =	8	10	12	14	20	50	100
# of $\mathbf{m}$'s =	1430	16796	208012	2674440	6.564×10^9	1.98×10^{27}	8.97×10^{56}

In view of the preceding list, the evaluation of $\Sigma_{\mathbf{m}} g(\mathbf{m})$ is impractical even for modest sample sizes, say $n \geq 14$. This section develops Monte Carlo methods to approximate $\Sigma_{\mathbf{m}} g(\mathbf{m})$ based upon a probability distribution $p(\mathbf{m})$ on the $\mathbf{m}$'s. A random sample of size M, say $\mathbf{m}_1, \ldots, \mathbf{m}_M$, is selected according to this distribution. By the law of large numbers, $M^{-1}\Sigma_{1 \leq j \leq M} g(\mathbf{m}_j)/p(\mathbf{m}_j)$ is a Monte Carlo approximation of $\Sigma_{\mathbf{m}} g(\mathbf{m})$.

Note, however, that each posterior quantity can be written as a ratio $\Sigma_{\mathbf{m}} g(\mathbf{m})/\Sigma_{\mathbf{m}} h(\mathbf{m})$ for given functions g and h [this is obvious for (4.4) and its consequences; it is also easy to show that (2.8) can be so written] and it is more efficient to approximate this ratio directly. A Monte Carlo approximation of $\Sigma_{\mathbf{m}} g(\mathbf{m})/\Sigma_{\mathbf{m}} h(\mathbf{m})$ is given by

$$\left[\sum_{1 \leq j \leq M} g(\mathbf{m}_j)/p(\mathbf{m}_j)\right] \Bigg/ \left[\sum_{1 \leq j \leq M} h(\mathbf{m}_j)/p(\mathbf{m}_j)\right]. \tag{5.1}$$

It is well know that the standard error of this approximation is $O(M^{-1/2})$ [Rubenstein (1981)].

A distribution on the $\mathbf{m}$'s can be conveniently defined using the Markov chain method. Each $\mathbf{m}$ uniquely defines a path of partial sums $\mathbf{s} = (s_0, \ldots, s_n)$ of length $n + 1$ which increases from $s_0 = 0$ to $s_n = n$ subject to $s_j = \Sigma_{1 \leq i \leq j} m_i \leq j$ for $j = 1, \ldots, n - 1$. Since the correspondence between $\mathbf{m}$ and $\mathbf{s}$ is one to one, it suffices to define a distribution $q(\mathbf{s})$ on the $\mathbf{s}$-paths. This can be easily done as follows: Let $\mathbf{S} = (S_0, S_1, \ldots, S_{n-1}, S_n)$ be a Markov chain such that $S_j | S_{j-1} = s_{j-1}$ has a uniform distribution on $\{s_{j-1}, s_{j-1} + 1, \ldots, j\}$ for $j = 1, \ldots, n - 1$; let $S_0 = s_0 = 0$ and $S_n = s_n = n$. Then

$$\begin{aligned} q(\mathbf{s}) &= \prod_{1 \leq i \leq n-1} (i + 1 - s_{i-1})^{-1}, \\ p(\mathbf{m}) &= \prod_{1 \leq i \leq n-1} \left(i + 1 - \sum_{1 \leq j \leq i-1} m_j\right)^{-1}. \end{aligned} \tag{5.2}$$

Let $\mathbf{s}_1, \ldots, \mathbf{s}_M$ be M independent realizations of the Markov chain $\mathbf{S} \sim q(\mathbf{s})$. The corresponding $\mathbf{m}_1, \ldots, \mathbf{m}_M$ is then a sample from $p(\mathbf{m})$.

When the sample size n is large, the $p(\mathbf{m})$ in (5.1) is likely to be very small, creating the usual computational difficulty of evaluating a ratio with a very small denominator. Therefore, it is desirable to select an $\mathbf{m}$-vector according to a

uniform distribution on the **m**'s; in this case $p(\mathbf{m})$ is a constant which cancels in (5.1), and the problem of division by a number close to zero does not arise. To describe this uniform distribution on the **m**'s, we let $t(i, j)$ be the number of paths of length $i + 2$ such that $S_i = j$ (by definition, $S_{i+1} = i + 1$ and $t(n, n)$ is the number of **m**-vectors corresponding to a sample of size n). A moment of reflection, taking into account the nondecreasing property of the **s**-paths, shows that the t's satisfy the recursive relation

$$t(i, j) = 0 \quad \text{for } i < j, \qquad t(i, 0) = 1 \quad \text{for } i = 0, 1, \ldots,$$

$$t(i, j) = t(i - 1, j) + t(i, j - 1) = \sum_{0 \le k \le j} t(i - 1, k) \tag{5.3}$$

$$\text{for } 1 \le j \le i = 1, 2, \ldots .$$

Define a Markov chain $\mathbf{S} = (S_0, S_1, \ldots, S_n)$ as follows: $S_n = n$, $S_{n-1} | S_n = n$ equals x with probability $t(n - 1, x)/t(n, n)$ for $x = 0, 1, \ldots, S_n$. For $j = 2, \ldots, n - 1$, $S_{n-j} | S_{n-j+1}$ equals x with probability $t(n - j, x)/t(n - j + S_{n-j+1})$ for $x = 0, 1, \ldots, S_{n-j+1}$.

PROPOSITION 5.1. *For the preceding Markov chain* $\mathbf{S} = (S_0, S_1, \ldots, S_n)$, *all paths are equally likely. Equivalently, all corresponding* **m**-*vectors are equally likely.*

PROOF. It suffices to note that

$$\begin{aligned} q(\mathbf{s}) &= \left[t(n - 1, s_{n-1})/t(n, n)\right] \\ &\quad \times \prod_{2 \le j \le n} t(n - j, s_{n-j})/t(n - j + 1, s_{n-j+1}) \\ &= 1/t(n, n), \end{aligned} \tag{5.4}$$

proving that **S** has a uniform distribution. □

Acknowledgment. The authors are grateful to Dr. Peisen Zhang for a helpful conversation.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ALDOUS, D. J. (1985). Exchangeability and related topics. *Ecole d'Eté de Probabilités de Saint-Flour XIII—1983. Lecture Notes in Math.* **1117** 1–198. Springer, New York.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.

DEGROOT, M. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

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

DOOB, J. (1949). Application of the theory of martingales. *Colloq. Internat. CNRS* No. 13, 23–27.

DYKSTRA, R. L. and LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

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

GHORAI, J. K. and RUBIN, H. (1982). Bayes risk consistency of nonparametric Bayes density estimates. *Austral. J. Statist.* **24** 51–66.

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153.

KUO, L. (1986). Computations of mixtures of Dirichlet processes. *SIAM J. Sci. Statist. Comput.* **7** 60–71.

LO, A. Y. (1984). On a class of Bayesian nonparametric estimates. I. Density estimates. *Ann. Statist.* **12** 351–357.

LO, A. Y. (1986). Bayes method for mixture models. Research Report 86-3, Dept. Statistics, SUNY Buffalo.

LO, A. Y. and WENG, C. S. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math. Ser. A.* **41** (2).

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

ROBERTSON, T. (1967). On estimating a density which is measurable with respect to a σ-lattice. *Ann. Math. Statist.* **38** 482–493.

RUBENSTEIN, R. Y. (1981). *Simulation and the Monte Carlo Method.* Wiley, New York.

WEGMAN, E. J. (1970). Maximum likelihood estimation of a unimodal density function. I. *Ann. Math. Statist.* **41** 457–471.

DEPARTMENT OF STATISTICS
STATE UNIVERSITY OF NEW YORK
AMHERST, NEW YORK 14261