RT-MAE 9714

BAYES ESTIMATION OF DIRICHLET PROCESS PARAMETER

by

José Galvão Leite, Victor Hugo Salinas Torres, Ram C. Tiwari and Jyoti N. Zalkikar

Palavras-Chave: Jensen inequality, Prior sample size; Maximum likelihood estimator.

Classificação AMS: 62F15, 62G05.
(AMS Classification)

- Agosto de 1997 -

Bayes Estimation of Dirichlet Process Parameter

José Galvão Leite
Departamento de Estatística do
Instituto de Matemática e Estatística
Universidade de São Paulo

Victor Hugo Salinas Torres
Departamento de Probabilidade e Estatística da
Universidade de Chile

Ram C. Tiwari
Department of Mathematics
University of North Carolina, Charlotte-USA

Jyoti N. Zalkikar
Department of Statistics - Florida International
University, Miami-USA

Bayes Estimation of Dirichlet Process Parameter

Abstract

Dirichlet process serves as a prior for the unknown distribution function in nonparametric Bayes estimation of various parameters. The present article addresses the problem of finding Bayes estimator of Dirichlet process parameter itself, with respect to the discrete prior distribution concentrating on finite number of fixed points. This problem is related to the estimation of the parameter of independent and non-identically distributed Benoulli model. We show that in the limiting case as the sample size increases to infinity, the Bayes estimator converges to the prior supported point "closest" to the "true" parameter value, which may not be a prior-supported point. This result examplifies a typical Bayesian phenomenon that the posterior distribution is dominated by the prior distribution.
KEY WORDS: Jensen inequality; Prior sample size; Maximum likelihood estimator.

1. INTRODUCTION

Ferguson (1973) introduced Dirichlet process as a prior distribution on a set of probability measures in Bayesian nonparametric framework. Since then many authors have considered applications of Dirichlet process and studied its properties. An excellent source of references is a recent survey article by Ferguson, Phadia and Tiwari (1992).

Let P be a Dirichlet process on $(R, \mathcal{B}(R))$ with parameter $\theta\,\overline{\alpha}(\cdot)$, where R is the real line and $\mathcal{B}(R)$ is its Borel σ-field, $\overline{\alpha}$ is the prior guess at P, and $\theta > 0$ is a measure of confidence in the prior when viewed as "prior sample size" (see, for example, Ferguson 1973; and Korwar and Hollander, 1976). This interpretation is, however, not always the case as shown by Sethuraman and Tiwari (1982) in a probabilistic setup. They have established that the smaller values of θ correspond to more definitive information in the process. Concisely, these authors have proved that as θ tends to zero, the Dirichlet process converges weakly to a degenerate random probability measure and the point of degeneracy has distribution $\overline{\alpha}$.

In non Bayesian setup, estimation of θ has been discussed in the literature. Korwar and Hollander (1973) derived maximum likelihood estimator, whereas in an empirical Bayes setup, Korwar and Hollander (1976) and Zehnwrith (1981) proposed a consistent estimator of θ in terms of F-statistic. In this paper we consider Bayes estimation of θ. In Section 2, under the assumption that the prior distribution concentrates on $k+1$ fixed points $\theta_0 < \dots < \theta_k$, we show that the parameter space can be broken up into $k+1$ intervals $(0,\xi_0),(\xi_0,\xi_1),\dots(\xi_{k-2},\xi_{k-1}),(\xi_{k-1},\infty)$, where $0<\theta_0<\xi_0<\theta_1<\xi_1<\dots<\xi_{k-1}<\theta_k<\infty$, and that the posterior mean of θ (that is, the Bayes estimator under squared error loss) converges almost surely to θ_j as long as the "true" parameter value is within the ξ-interval containing θ_j. One should not misconstrue this result as inconsistency of Bayes estimator as discussed in Diaconis and Freedman(1986 a, b). Some discussion of this point is contained in Section 3.

2. BAYES ESTIMATOR OF θ AND ITS ALMOST SURE CONVERGENCE

Let $X_1,\dots,X_n$ be a sample of size n from P, where P is a Dirichlet process on $(R,\mathcal{B}(R))$ with parameter $\theta\overline{\alpha}$. One can view the observation $X_1,\dots,X_n$ as being obtained sequentially as follows (Korwar and Hollander, 1973): Let X_1 be a sample of size 1 from P; having obtained X_1, let X_2 be a sample of size 1 from the conditional distribution of P given X_1; and so on until $X_1,\dots,X_n$ are obtained. Set $D_1=1$, and for $i=2,\dots,n$ set $D_i=0$ if $X_i=X_j$ for some $j=1,\dots,i-1$ and 1 otherwise and let $D=\sum_{i=1}^{n} D_i$ denote the number of distinct observations among $X_1,\dots,X_n$. Here the dependence of D on n is ignored for notational ease. The induced distribution Q of $(X_1,\dots,X_n)$ on the sample space $(R^n,\mathcal{B}(R^n))$ is given by (Blackwell, 1973) the marginal of X_1:

$$Q(X_1\le x_1)=\overline{\alpha}((-\infty,x_1]) \tag{2.1}$$

and the predictive (i.e., the conditional) distribution of X_i given $X_1=x_1,\dots,X_{i-1}=x_{i-1}$:

$$Q(X_i\le x_i|X_1=x_1,\dots,X_{i-1}=x_{i-1})$$

$$= \frac{\theta}{\theta+i-1}\overline{\alpha}((-\infty, x_i]) + \frac{i-1}{\theta+i-1}F_{i-1}(x_i)\ ,\ i = 2, ..., n, \tag{2.2}$$

where F_{i-1} is the empirical distribution of $(X_1, ..., X_{i-1})$. Notice from (2.2) that if θ is close to zero, then the predictive distribution of X_i given $X_1, ..., X_{i-1}$ is close to the empirical distribution F_{i-1} and, on the other hand, if θ is large the predictive distribution of X_i given $X_1, ..., X_{i-1}$ is close to the prior guess $\overline{\alpha}$. Thus a large value of θ corresponds to the higher degree of belief in the process leading to the interpretation of θ as the "prior sample size". We shall assume throughout that the prior guess $\overline{\alpha}$ at P is absolutely continuous with respect to Lebesgue measure μ on R with density $d\overline{\alpha}/d\mu = f_0(x)$. Proceeding as in Korwar and Hollander (1973) and Antoniak (1974), the joint distribution Q of $(X_1, ..., X_n)$ in (2.1) and (2.2) can be expressed as the sum of measures which are mutually singular:

$$Q(x_1, ..., x_n) = \sum_{D} \sum_{(n_1,...n_D)} Q_{(n_1,...,n_D)} \quad . \tag{2.3}$$

where the first summation is over the possible number of distinct values D among the sample observations $(X_1, ..., X_n)$ and the second summation is over $(n_1, ..., n_D)$, n_i being the number of replicates of ith distinct observation. Here each $n_i > 0$, $\sum_{i=1}^{D} n_i = n$ and $Q_{(n_1,...,n_D)}$ is a probability measure that is absolutely continuous with respect to Lebesgue measure $\mu_{(n_1,...,n_D)}$ on the set $C(n_1, ..., n_D) = \{(x_1, ..., x_n)$: there are D distinct values with replications $(n_1, ..., n_D)$, $n_i > 0$, $\sum_{i=1}^{D} n_i = n\}$.

Then, Q is absolutely continuous with respect to

$$\lambda = \sum_{D} \sum_{(n_1,...,n_D)} \mu_{(n_1,...,n_D)} \tag{2.4}$$

and for $(x_1, ..., x_n) \in C(n_1, ..., n_D)$,

$$\begin{aligned} L(\theta | X_1 &= x_1, ..., X_n = x_n) = \frac{dQ\ (x_1, ..., x_n)}{d\lambda} \\ &= \frac{\theta^D}{\theta_{(n)}} \prod_{i=1}^{D} (n_i - 1)! \prod_{i=1}^{D} f_0(x_i^*) \\ &\propto \frac{\theta^D}{\theta_{(n)}}, \end{aligned} \tag{2.5}$$

where $\theta_{(n)} = \theta(\theta+1)...(\theta+n-1)$ and $x_1^*, ..., x_D^*$ denote the D distinct values with multiplicities $n_1, ..., n_D$ respectively. From (2.5) it follows that the likelihood function of θ given the sample $(X_1, ..., X_n)$ depends on θ only through D, the number of distinct observations in the sample. Also the conditional distribution of the sample $(X_1, ..., X_n)$ given D is proportional to $\prod_{i=1}^{D} f_0(x_i^*)$ and does not involve the parameter θ. Thus D is a sufficient statistic for θ and has all the relevant information about this parameter contained in the sample. Differentiating the logarithm of the likelihood function in (2.5) with respect to θ and setting this derivative equal to zero yields the maximum likelihood estimator(mle) $\widehat{\theta}_M$ defined by the equation

$$D = \sum_{i=1}^{n} \widehat{\theta}_M / (\widehat{\theta}_M + i - 1) \tag{2.6}$$

Define, $f(\theta) = \left(\sum_{i=1}^{n} \theta/(\theta + i - 1) \right) - D$. Then for $\theta \in [0, \infty)$, $f'(\theta) > 0$ and $f''(\theta) < 0$, and $f(\theta) \to -D$ as $\theta \to 0$ and $f(\theta) \to n - D$ as $\theta \to \infty$. For $\theta \in [0, \infty)$, the function is monotonically increasing. If $n - D > 0$ (i.e. $n - D \geq 1$), then the function $f(\theta)$ crosses the $\theta = 0$ line uniquely from below, and the mle of θ exists and is the unique solution of $f(\theta) = 0$. If $n - D = 0$ i.e. when all the observations in the sample are distinct, then the line $\theta = 0$ is an asymptote of the function, and the mle of θ does not exist. The graphs of the function $f(\theta)$ for $n = 15$ and $D = 3, 9, 15$ are plotted in Figure 1, and the graphs reveal the above features of the function $f(\theta)$. Note that $\widehat{\theta}_M$ is a function of sufficient statistic D. Korwar and Hollander (1973) obtained $\widehat{\theta}_M$ in (2.6) by showing that under the assumption that $\bar{\alpha}$ is non-atomic, $D_1, D_2, ..., D_n$ are independent Bernoulli random variable with $Q(D_i = 1) = \theta/\theta + i - 1$, $i = 1, ..., n$. One may, therefore, regard the problem of estimation of the Dirichlet process parameter θ to be identical to the estimation of θ arising in the context of independent but not identically distributed Bernoulli random variables D_i.

A result of Korwar and Hollander (1973) about the sequence D_i's that will be of use in this paper is as follows. They showed that although the probability of getting new observation decreases with n, nonetheless we are assured of an infinite number of distinct observations and proved that $D/\log n$ converges to θ almost surely as n increases to infinity.

Let $\pi(\theta)$ be the prior density (discrete or continuous) of θ. Then from Bayes' theorem the posterior density of θ is given by

$$\pi(\theta|D; n) \;\; = \;\; \left[\theta^D/\theta_{(n)}\right] \pi(\theta) \; / \; \int \left[\theta^D/\theta_{(n)}\right] \pi(\theta) d\theta$$

$$= \quad C(D;n)(\theta^D/\theta_{(n)})\pi(\theta), \tag{2.7}$$

where $[C(D;n)]^{-1} = \int(\theta^D/\theta_{(n)})\pi(\theta)d\theta$.

The Bayes estimator of θ with respect to squared error loss is given by

$$\begin{aligned}\widehat{\theta}(D;n) &= E_\pi(\theta|D;n)\\ &= C(D;n)/C(D+1;n)\end{aligned} \tag{2.8}$$

The effect of D and n on the Bayes estimator is given in the following result.

Theorem 2.1. For $n \geq 1$, $\widehat{\theta}(D+1;n) \geq \widehat{\theta}(D;n) \geq \widehat{\theta}(D;n+1)$ a.s.

Proof The first inequality follows by an application of Jensen's inequality (cf. Chung, 1974, p.47):

$$\begin{aligned}\widehat{\theta}(D+1;n) &= E_\pi(\theta^2|D;n)/E_\pi(\theta|D;n)\\ &\geq \{E_\pi(\theta|D;n)\}^2/E_\pi(\theta|D;n) \text{ a.s.}\\ &= \widehat{\theta}(D;n) \text{ a.s.}\end{aligned}$$

Define $h(\theta) = (\theta+n)^{-1}$. Then, since $h(\theta)$ is a decreasing function, it follows from Lehmann (1966) that

$$E_\pi(h(\theta) \mid D;n)\; E_\pi(\theta \mid D;n) \geq E_\pi(\theta h(\theta) \mid D;n) \text{ a.s.} \tag{2.9}$$

Clearly,

$$E_\pi(h(\theta) \mid D;n) \;= C(D;n) \;/\; C(D;n+1) \tag{2.10}$$

and

$$E_\pi(\theta h(\theta) \mid D;n) \;= C(D;n) \;/\; C(D+1;n+1), \tag{2.11}$$

where in (2.10) we have interpreted $C(D;n+1) = C(D,\ D_{n+1} = 0; n+1)$ and in (2.11) we have interpreted $C(D+1;n+1) = C(D+1, D_{n+1} = 0; n+1)$. Now the second inequality follows from substituting (2.8), (2.10), (2.11) in (2.9). □

If $\pi(\theta)$ is discrete with $\pi(\theta = \theta_j) = \pi_j, j = 0, ..., k$ and $\sum_{j=0}^{k} \pi_j = 1$, then

$$\widehat{\theta}(D;n) = \frac{\sum_{j=0}^{k} \left(\theta_j^{D+1}/(\theta_j)_{(n)}\right)\pi_j}{\sum_{j=0}^{k} \left(\theta_j^{D}/(\theta_j)_{(n)}\right)\pi_j}, \tag{2.12}$$

where $(\theta_j)_{(n)} = \theta_j(\theta_j+1)...(\theta_j+n-1)$, $j = 0, ..., k$. Also the posterior variance of θ is given by

$$Var_\pi(\theta \mid D; n) = \frac{\sum_{j=0}^{k} \left(\theta_j^{D+2}/(\theta_j)_{(n)}\right)\pi_j}{\sum_{j=0}^{k} \left(\theta_j^{D}/(\theta_j)_{(n)}\right)\pi_j} - \left[\widehat{\theta}(D;n)\right]^2. \tag{2.13}$$

For binomial prior $\pi(\theta = j+1) = \pi_j = \binom{k}{j} p^j(1-p)^{k-j}$, $j = 0, 1, ..., k$,and $p = 0.1, 0.5, 0.9$, the prior and posterior distribution functions of θ are plotted in Figures 2, 3, and 4 for $k = 10$, n =15 and D = 3, 9, 15. For these values of k, n, D and for discrete uniform prior $\pi(\theta = j+1) = \pi_j = 1/(k+1)$, $j = 0, 1, ..., k$, the prior and posterior distribution functions of θ are plotted in Figure 5. The mle of θ, and the prior and posterior means and variances of θ are given in Tables 1, 2, and 3 for each choice of (k, n, D), where $k = 10, 20$, $n = 15, 25$, and $D = 3, 9, 15$. The computations in Table 3 show that the posterior mean of θ demonstrates the effects of D and n as shown in Theorem 2.1; however, this result does not hold for the posterior variance.

The following result is useful in proving our main result given in Theorem 2.3.

Lemma 2.2. If $\theta_0 < ... < \theta_k (k \geq 2)$, then

$$\theta_j < \frac{\theta_{j+1} - \theta_j}{\log(\theta_{j+1}/\theta_j)} < \theta_{j+1}\ , j = 0, ..., k-1$$

Proof. For every $x \neq 1$. we have

$$x - 1 > \log x$$

and hence

$$\frac{\theta_{j+1}}{\theta_j} - 1 > \log\left(\frac{\theta_{j+1}}{\theta_j}\right)$$

and

$$\frac{\theta_j}{\theta_{j+1}} - 1 > \log\frac{\theta_j}{\theta_{j+1}}$$

giving

$$\theta_j < \frac{\theta_{j+1} - \theta_j}{\log(\theta_{j+1}/\theta_j)} < \theta_{j+1}. \square$$

Set $\xi_j = \frac{\theta_{j+1}-\theta_j}{\log(\theta_{j+1}/\theta_j)}$, $j = 0, ..., k-1$. Then Lemma 2.2 implies that the parameter space $(0, \infty)$ can be broken up into $k+1$ ξ-intervals $(0, \xi_0), (\xi_0, \xi_1), ..., (\xi_{k-2}, \xi_{k-1})$, (ξ_{k-1}, ∞) such that $0 < \theta_0 < \xi_0 < \theta_1 < ... < \theta_{j-1} < \xi_{j-1} < \theta_j < ... < \theta_{k-1} < \xi_{k-1} < \theta_k < \infty$. Imposing the restriction on the prior distribution $\pi(\theta)$ that it concentrates on $k+1$ fixed points $\theta_0 < ... < \theta_k$, where θ_i's are contained in $\xi-$ intervals defined above, we have the following main result of the paper.

Theorem 2.3. Under the assumptions that $\pi(\cdot)$ is discrete with support $0 < \theta_0 < ... < \theta_k (k \geq 2)$,

(*i*) if $0 < \theta < \xi_0$, then $\widehat{\theta}(D; n) \to \theta_0$ a.s. as $n \to \infty$

(*ii*) if for $j (j = 1, ..., k-1)$, $\xi_{j-1} < \theta < \xi_j$, then $\widehat{\theta}(D; n) \to \theta_j$ a.s. as $n \to \infty$,

(*iii*) if $\theta > \xi_{k-1}$, then $\widehat{\theta}(D; n) \to \theta_k$ a.s. as $n \to \infty$,

The proof of this theorem is given in the Appendix.

3. DISCUSSION

Under a finitely supported assumption on the prior, strong law of large numbers for Bayesian statistics together with $D/\log n$ converges a.s. to the "true" parameter (Korwar and Hollander, 1973) imply that the posterior mean converges a.s. to θ_j as long as the "true" parameter is θ_j, for each $j = 0, ..., k$. The main contribution of this paper (Theorem 2.3) is that the posterior mean still converges a.s. to θ_j as long as the "true" parameter is within the ξ-interval containing θ_j. This result should not be construed as an example of Bayesian inconsistency of the type discussed by Diaconis and Freedman (1986 a, b). Rather, it rediscovers a typical phenomenon that the posterior distribution is dominated by the prior distribution, so that if the prior distribution has a finite support, the posterior distribution is also finitely supported and the posterior mean tends to the prior supported point "closest" to the "true" parameter (which may not be a prior supported point).

The ξ-intervals define "closeness" property for the Bernoulli model naturally arising in this paper: θ_j is close to the "true" parameter if and only if they are within the same ξ-interval. It remains to be seen what happens when the true parameter falls on the boundary of the ξ-interval. Also further investigation calls for relaxing the restriction that the prior distribution is finitely supported.

4. APPENDIX: Proof of Theorem 2.3

i) From (2.12),

$$\hat{\theta}(D;n) = \theta_0 \left[\frac{\pi_0 + \sum_{j=1}^{k} \left(\frac{\theta_j}{\theta_0}\right)\left(\frac{\theta_j}{\theta_0}\right)^D \frac{(\theta_0)_{(n)}}{(\theta_j)_{(n)}} \pi_j}{\pi_0 + \sum_{j=1}^{k} \left(\frac{\theta_j}{\theta_0}\right)^D \frac{(\theta_0)_{(n)}}{(\theta_j)_{(n)}} \pi_j} \right],$$

hence it suffices to show that, for $j = 1, ..., k$,

$$\left(\frac{\theta_j}{\theta_0}\right)^D \frac{(\theta_0)_{(n)}}{(\theta_j)_{(n)}} \to 0 \text{ a.s. as } n \to \infty. \tag{A.1}$$

But, for $j = 1, ..., k$,

$$\begin{aligned} \left(\frac{\theta_j}{\theta_0}\right)^D \frac{(\theta_0)_{(n)}}{(\theta_j)_{(n)}} &= \prod_{i=1}^{j} \left(\frac{\theta_i}{\theta_{i-1}}\right)^D \frac{(\theta_{i-1})_{(n)}}{(\theta_i)_{(n)}} \\ &= \prod_{i=1}^{j} \left(\frac{\theta_i}{\theta_{i-1}}\right)^D n^{\theta_{i-1}-\theta_i} \frac{(\theta_{i-1})_{(n)}}{(\theta_i)_{(n)}} n^{\theta_i - \theta_{i-1}}, \end{aligned}$$

and since for any two positive numbers a and b (cf. Feller 1968, Vol. I, p. 66, eq. (12.22)),

$$\frac{a_{(n)}}{b_{(n)}} n^{b-a} \to \frac{\Gamma(b)}{\Gamma(a)} \text{ as } n \to \infty, \tag{A.2}$$

we have that, for $i = 1, ..., j (j = 1, ..., k)$,

$$\frac{(\theta_{i-1})_{(n)}}{(\theta_i)_{(n)}} n^{\theta_i - \theta_{i-1}} \to \frac{\Gamma(\theta_i)}{\Gamma(\theta_{i-1})} \text{ as } n \to \infty.$$

Hence to show (A.1) it suffices to show that, for $i = 1, ..., j$,

$$\left(\frac{\theta_i}{\theta_{i-1}}\right)^D n^{\theta_{i-1}-\theta_i} \to 0 \text{ a.s. as } n \to \infty$$

or equivalently

$$\left[\frac{D}{\log n} \log\left(\frac{\theta_i}{\theta_{i-1}}\right) + (\theta_{i-1} - \theta_i)\right] \log n \to -\infty \text{ a.s. as } n \to \infty \tag{A.3}$$

Now, from Lemma 2.2 and the hypothesis that $\theta < \xi_0$, it follows that

$$\theta \log(\theta_i/\theta_{i-1}) + (\theta_{i-1} - \theta_i) < 0, i = 1, ..., k,$$

and hence from the result that $D/\log n \to \theta > 0$ a.s. as $n \to \infty$, we have

$$\frac{D}{\log n} \log\left(\frac{\theta_i}{\theta_{i-1}}\right) + (\theta_{i-1} - \theta_i) \to \theta \log\left(\frac{\theta_i}{\theta_{i-1}}\right) + (\theta_{i-1} - \theta_i) < 0, \text{ a.s. as } n \to \infty.$$

This establishes (A.3).

ii) After some algebra, we can write

$$\widehat{\theta}(D;n) = \theta_j \left[\frac{\sum_{i=0}^{j-1} \left(\frac{\theta_i}{\theta_j}\right)\left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_i)_{(n)}} \pi_i + \pi_j + \sum_{i=j+1}^{k} \left(\frac{\theta_i}{\theta_j}\right)\left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_i)_{(n)}} \pi_i}{\sum_{i=0}^{j-1} \left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_i)_{(n)}} \pi_i + \pi_j + \sum_{i=j+1}^{k} \left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_j)_{(n)}} \pi_j} \right]$$

Thus, it suffices to show that, for $i = 0, ..., k, i \neq j$, $\left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_i)_{(n)}} \to 0$ a.s. as $n \to \infty$.

But, for $i = 0, ..., j-1$,

$$\begin{aligned} \left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_i)_{(n)}} &= \prod_{s=i+1}^{j} \left(\frac{\theta_{s-1}}{\theta_s}\right)^D \frac{(\theta_s)_{(n)}}{(\theta_{s-1})_{(n)}} \\ &= \prod_{s=i+1}^{j} \left(\frac{\theta_{s-1}}{\theta_s}\right)^D n^{\theta_s - \theta_{s-1}} \frac{(\theta_s)_{(n)}}{(\theta_{s-1})_{(n)}} n^{\theta_{s-1} - \theta_s} \end{aligned}$$

and, for $i = j+1, ..., k$,

$$\begin{aligned} \left(\frac{\theta_i}{\theta_j}\right)^D \frac{(\theta_j)_{(n)}}{(\theta_i)_{(n)}} &= \prod_{w=j+1}^{i} \left(\frac{\theta_w}{\theta_{w-1}}\right)^D \frac{(\theta_{w-1})_{(n)}}{(\theta_w)_{(n)}} \\ &= \prod_{w=j+1}^{i} \left(\frac{\theta_w}{\theta_{w-1}}\right)^D n^{\theta_{w-1} - \theta_w} \frac{(\theta_{w-1})_{(n)}}{(\theta_w)_{(n)}} n^{\theta_w - \theta_{w-1}}. \end{aligned}$$

In view of (A.2), we have, for $1 \leq i+1 \leq s \leq j$,

$$\frac{(\theta_s)_{(n)}}{(\theta_{s-1})_{(n)}} n^{\theta_{s-1} - \theta_s} \to \frac{\Gamma(\theta_{s-1})}{\Gamma(\theta_s)} \text{ as } n \to \infty,$$

and, for $j+1 \le w \le i \le k$,

$$\frac{(\theta_{w-1})_{(n)}}{(\theta_w)_{(n)}} n^{\theta_w-\theta_{w-1}} \to \frac{\Gamma(\theta_w)}{\Gamma(\theta_{w-1})} \text{ as } n \to \infty.$$

Thus, it suffices to show that, for $1 \le (i+1 \le) s \le j$,

$$\left(\frac{\theta_{s-1}}{\theta_s}\right)^D n^{\theta_s-\theta_{s-1}} \to 0 \text{ a.s. as } n \to \infty.$$

or equivalently,

$$-\left[\frac{D}{\log n}\log\left(\frac{\theta_s}{\theta_{s-1}}\right) + (\theta_{s-1}-\theta_s)\right]\log n \to -\infty \text{ a.s. as } n \to \infty \tag{A.4}$$

and, for $j+1 \le w (\le i) \le k$,

$$\left(\frac{\theta_w}{\theta_{w-1}}\right)^D n^{\theta_{w-1}-\theta_w} \to 0 \text{ a.s. as } n \to \infty$$

or equivalently,

$$\left[\frac{D}{\log n}\log\left(\frac{\theta_w}{\theta_{w-1}}\right) + (\theta_{w-1}-\theta_w)\right]\log n \to -\infty \text{ a.s. as } n \to \infty. \tag{A.5}$$

The hypothesis that, for $j (j = 1, ..., k-1)$,

$$\xi_{j-1} < \theta < \xi_j,$$

and Lemma 2.2,

$$\theta < \frac{\theta_i - \theta_{i-1}}{\log(\theta_i/\theta_{i-1})}, \; i = j+1, ..., k$$

and

$$\theta > \frac{\theta_i - \theta_{i-1}}{\log(\theta_i/\theta_{i-1})}, \; i = 1, ..., j,$$

imply that

$$\theta \log\left(\frac{\theta_i}{\theta_{i-1}}\right) + (\theta_{i-1}-\theta_i)\begin{cases} > 0, \; i = 1, ..., j \\ < 0, \; i = j+1, ..., k \end{cases} \tag{A.6}$$

Now, (A.6) and the result that $D/\log n \to \theta > 0$ a.s. as $n \to \infty$ imply (A.4) and (A.5).

iii) Writing

$$\widehat{\theta}(D;n) = \theta_k \left[\frac{\sum_{j=0}^{k-1} \left(\frac{\theta_j}{\theta_k}\right)\left(\frac{\theta_j}{\theta_k}\right)^D \frac{(\theta_k)_{(n)}}{(\theta_j)_{(n)}}\pi_j + \pi_k}{\sum_{j=0}^{k-1} \left(\frac{\theta_j}{\theta_k}\right)^D \frac{(\theta_k)_{(n)}}{(\theta_j)_{(n)}}\pi_j + \pi_k} \right],$$

it suffices to show that, for $j = 0, ..., k-1$,

$$\left(\frac{\theta_j}{\theta_k}\right)^D \frac{(\theta_k)_{(n)}}{(\theta_j)_{(n)}} \to 0 \text{ a.s. as } n \to \infty.$$

But, for $j = 0, ..., k-1$,

$$\left(\frac{\theta_j}{\theta_k}\right)^D \frac{(\theta_k)_{(n)}}{(\theta_j)_{(n)}} = \prod_{w=j+1}^{k} \left(\frac{\theta_{w-1}}{\theta_w}\right)^D n^{\theta_w - \theta_{w-1}} \frac{(\theta_w)_{(n)}}{(\theta_{w-1})_{(n)}} n^{\theta_{w-1}-\theta_w}$$

and (cf. eq (A.2)), for $1 \le j+1 \le w \le k$,

$$\frac{(\theta_w)_{(n)}}{(\theta_{w-1})_{(n)}} n^{\theta_{w-1}-\theta_w} \to \frac{\Gamma(\theta_{w-1})}{\Gamma(\theta_w)} \text{ as } n \to \infty.$$

Hence, it suffices to show that, for $1 \le j+1 \le w \le k$,

$$\left(\frac{\theta_{w-1}}{\theta_w}\right)^D n^{\theta_w - \theta_{w-1}} \to 0 \text{ a.s. as } n \to \infty$$

or equivalently,

$$-\left[\frac{D}{\log n} \log\left(\frac{\theta_w}{\theta_{w-1}}\right) + (\theta_{w-1} - \theta_w)\right] \log n \to -\infty \text{ a.s. as } n \to \infty. \tag{A.7}$$

The hypothesis,

$$\theta > \xi_{k-1},$$

and Lemma 2.2, for j = 1, ..., k,

$$\theta > \frac{\theta_j - \theta_{j-1}}{\log(\theta_j/\theta_{j-1})}, \text{for } j = 1, ..., k,$$

imply that, for j = 1, ..., k,

$$\theta \log\left(\frac{\theta_j}{\theta_{j-1}}\right) + (\theta_{j-1} - \theta_j) > 0.$$

Hence, (A.7) and the result that $D/\log n \to \theta > 0$ a.s. as $n \to \infty$ imply (A.6).□

REFERENCES

Antoniak, C.E. (1974), "Mixtures of Dirichlet Processes With Application to Bayesian Nonparametric Problems, *The Annals of Statistics,* 2, 1152-1174.

Blackwell, D. (1973), "Discreteness of Ferguson Selections," *The Annals of Statistics,* 1, 356-358.

Chung, K. (1974), *A Course in Probability Theory,* New York: Academic Press.

Diaconis, P., and Freedman, D. (1986a), "On the Consistency of Bayes Estimates," *The Annals of Statistics,* 14, 1-26.

Diaconis, P., and Freedman, D. (1986b), "On Inconsistency of Bayes Estimates of Location," *The Annals of Statistics,* 14, 68-87.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications,* Vol. I, New York, John Wiley.

Ferguson, T.S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics,* 1, 209-230.

Ferguson, T.S., Phadia, E.G., and Tiwari, R.C. (1992), "Bayesian Nonparametric Inference," in *Current Issues in Statistical Inference: Essays in Honor of D. Basu,* eds. M. Ghosh and P.K. Pathak, IMS Lecture Notes & Monograph Series, Vol. 17, pp. 127-150.

Korwar, R.M., and Hollander, M. (1973), "Contributions to the Theory of Dirichlet Processes," *The Annals of Probability,* 1, 705-711.

Korwar, R.M., and Hollander, M. (1976), " Empirical Bayes Estimation of a Distribution Function," *The Annals of Statistics,* 4, 581-588.

Lehmann, E.L. (1966), " Some Concepts of Dependence," *The Annals of Mathematical Statistics,* 37, 1137-1153.

Sethuraman, J., and Tiwari, R.C. (1982), "Convergence of Dirichlet Measures and the Interpretation of Their Parameter," in *Statistical Decision Theory and Related Topics III 2,* eds. S. Gupta and J.O. Berger, New York: Academic Press, pp. 305-315.

Zehnwirth, B. (1981), " A Note on the Asymptotic Optimality of the Empirical Bayes Distribution Function," *The Annals of Statistics,* 9, 221-224.

Table 1. Maximum Likelihood Estimator of θ

$D \setminus n$	15	25
3	.8241	.6551
9	8.5784	4.5962
15	does not exist	14.9148
25	not applicable	does not exist

Table 2. Prior Means and Variances

Prior	Mean	Variance
$U\{1, ..., 11\}$	6	10
$U\{1, ..., 21\}$	11	20.95
Bin(10,.1)	2	.90
Bin(10,.5)	6	2.50
Bin(10,.9)	10	.90
Bin(20,.1)	3	1.80
Bin(20,.5)	11	5
Bin(20,.9)	19	1.80

Note: Under a Bin(k,p) prior $E(\theta) = kp + 1, Var(\theta) = kp(1-p)$, and under a discrete uniform prior, $U\{1, ..., k+1\}$, $E(\theta) = (k+2)/2$, and $Var(\theta) = k(k+2)/12$.

Table 3. Posterior Means and Variances. The entries for Variances are given in parentheses below the Means.

Prior \ (n, D)	(15,3)	(15,9)	(15,15)	(25,3)	(25,9)	(25,15)
$U\{1,...,11\}$	1.7699	7.7657	10.0049	1.3920	5.7081	9.3095
	(1.2801)	(5.1064)	(1.4467)	(0.5145)	(4.9525)	(2.5322)
$U\{1,...22\}$	1.7733	11.2368	18.0590	1.3920	6.1315	14.7109
	(1.3201)	(21.3792)	(7.6080)	(0.5150)	(8.1887)	(15.2410)
$Bin(10,.1)$	1.4877	3.3025	4.8215	1.3153	2.9366	4.3861
	(.4354)	(.8984)	(1.1022)	(.2929)	(.7467)	(.9726)
$Bin(10,.5)$	3.9359	6.4629	7.9845	3.0977	5.7059	7.3790
	(1.1510)	(1.9482)	(1.5849)	(1.6788)	(1.8606)	(1.6479)
$Bin(10,.9)$	9.1839	9.9722	10.4036	8.5323	9.6145	10.1921
	(1.5704)	(.8958)	(.5354)	(2.0489)	(1.1594)	(0.7017)
$Bin(20,.1)$	1.9266	4.2989	6.2843	1.6077	3.7168	5.5853
	(.8169)	(1.5169)	(1.9087)	(.5543)	(1.2151)	(1.6277)
$Bin(20,.5)$	7.4382	10.7604	12.9298	5.5683	9.2555	11.6804
	(5.2356)	(4.4614)	(3.8940)	(4.3204)	(4.2829)	(3.9413)
$Bin(20,.9)$	18.0058	18.7760	19.3307	17.1821	18.1863	18.8934
	(2.6645)	(1.9707)	(1.4893)	(3.3611)	(2.4477)	(1.8425)

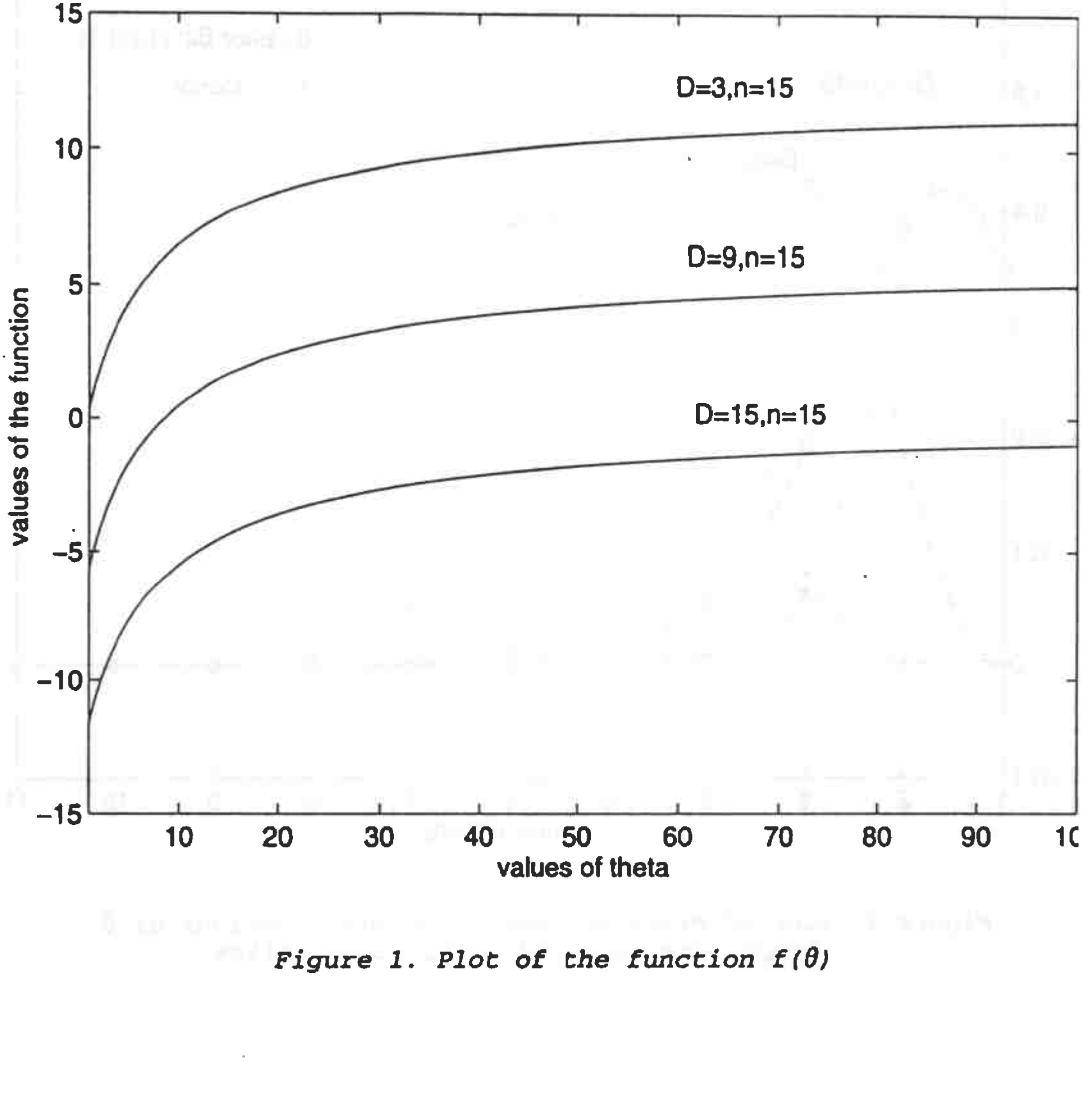

Figure 1. Plot of the function $f(\theta)$

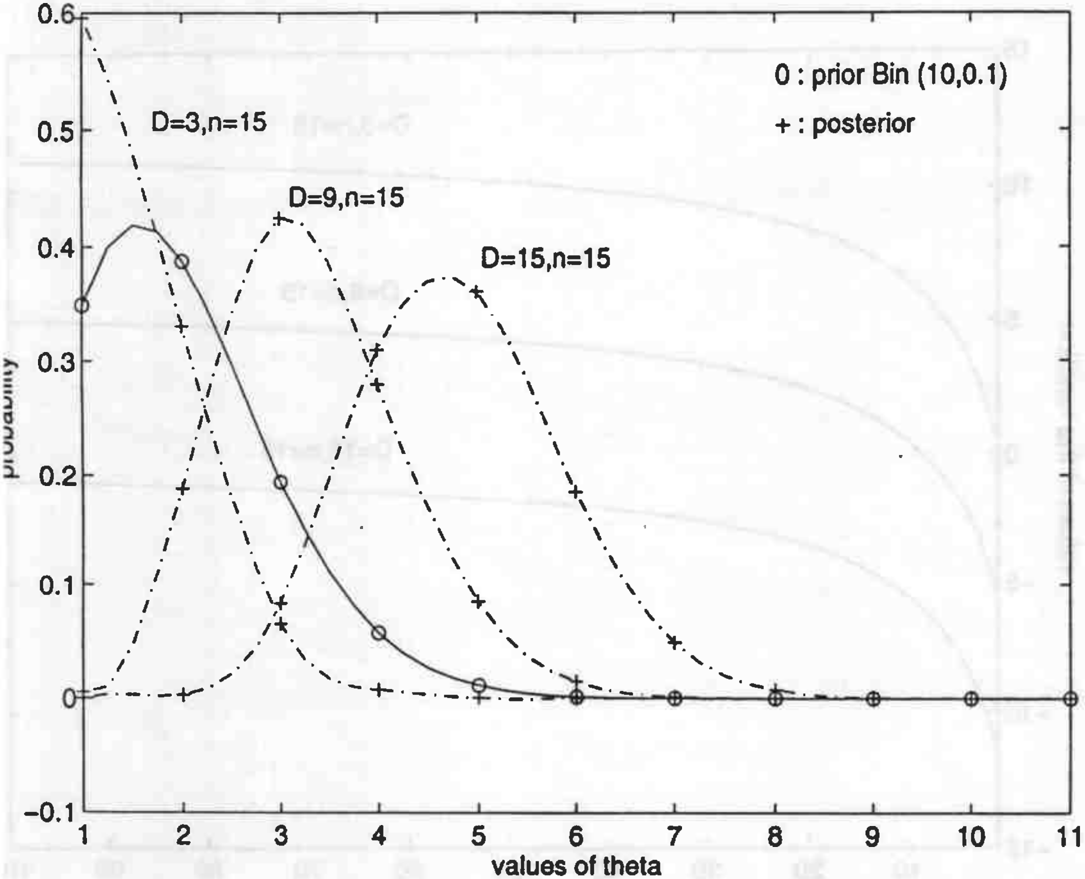

Figure 2. Plot of Prior and Posterior Distributions of θ.
Graphs are smoothed using cubic spline.

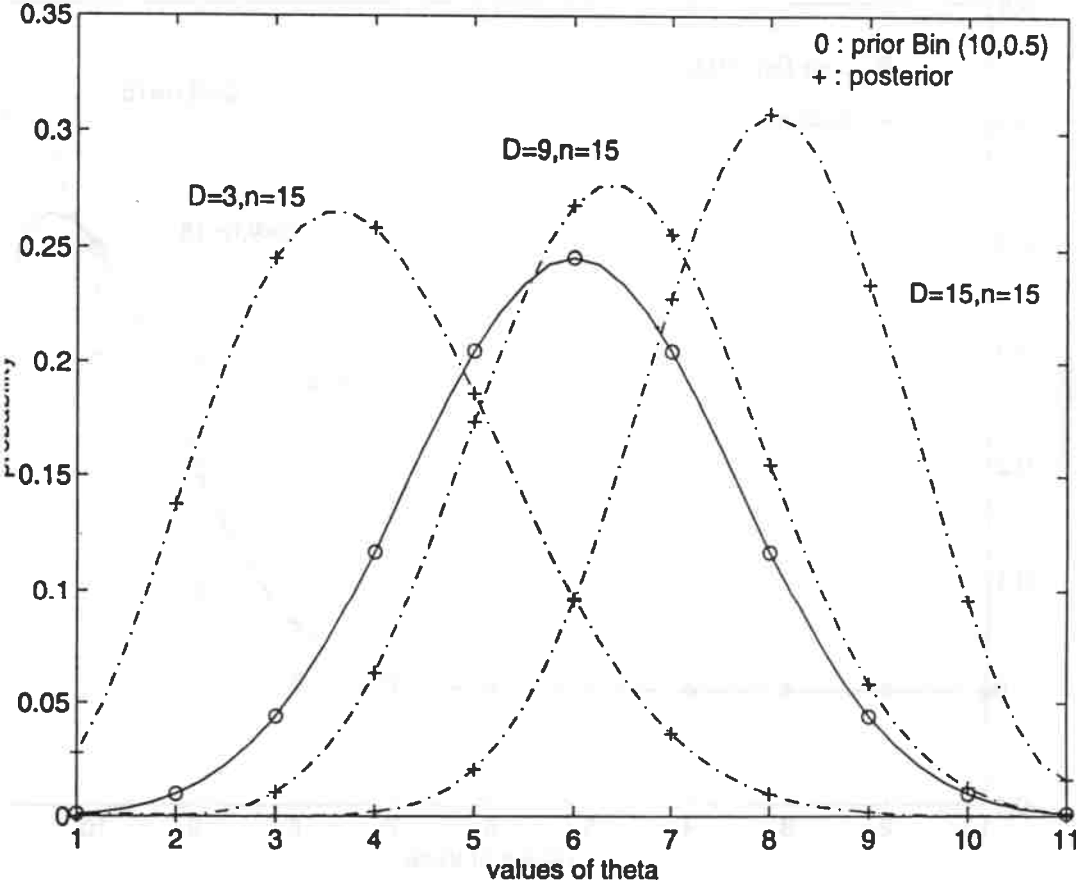

Figure 3. Plot of Prior and Posterior Distributions of θ.
Graphs are smoothed using cubic spline.

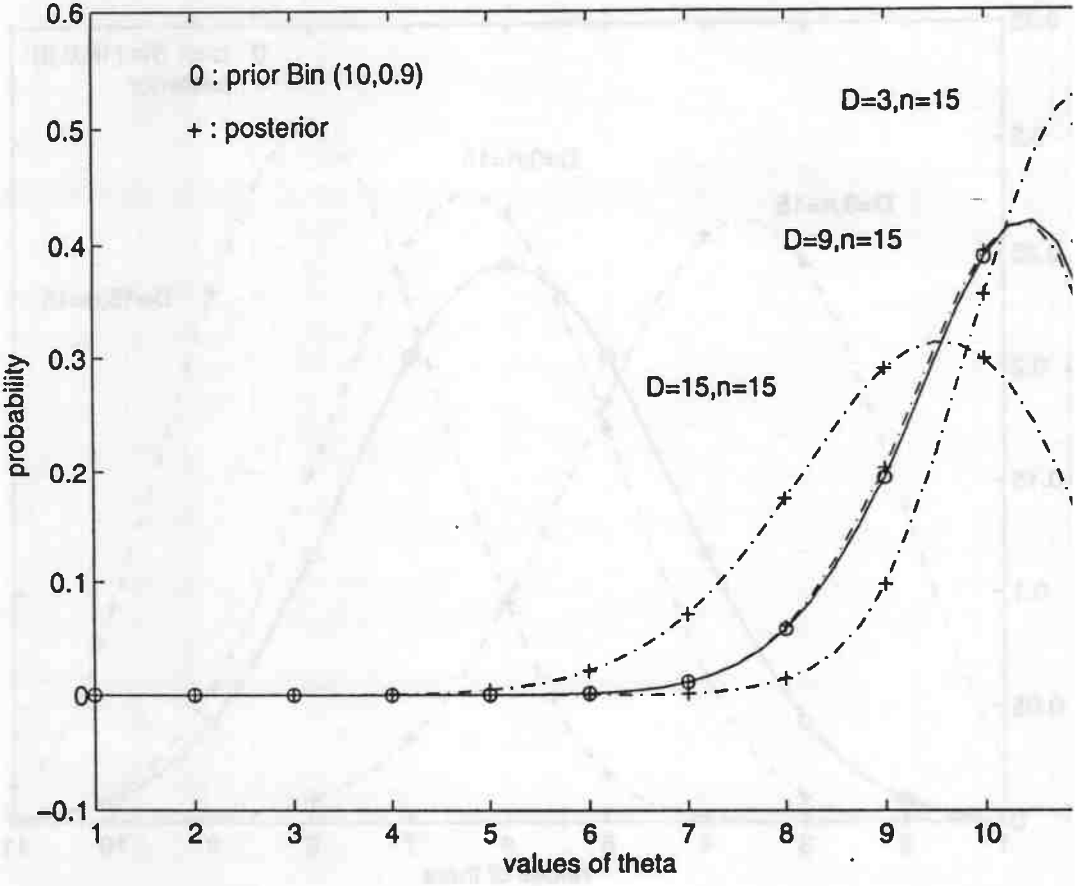

Figure 4. Plot of Prior and Posterior Distributions of θ. Graphs are smoothed using cubic spline.

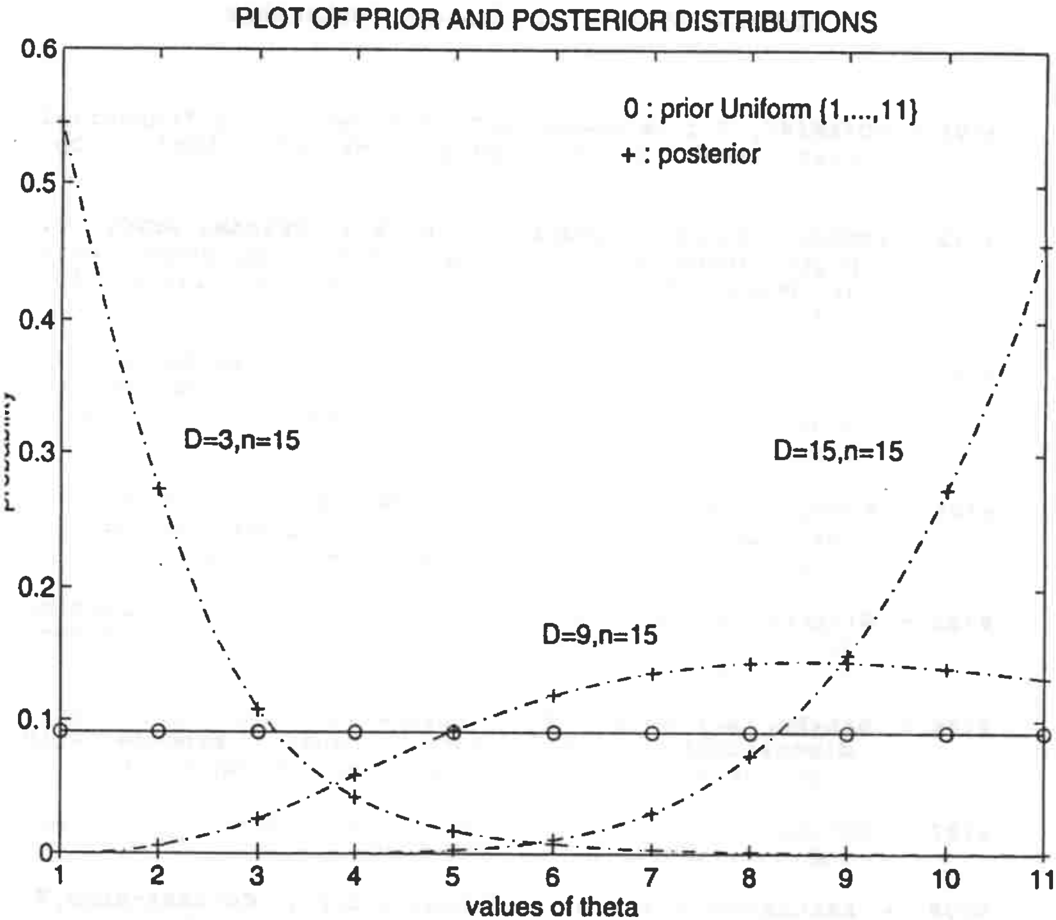

Figure 5. Plot of Prior and Posterior Distributions of θ.
Graphs are smoothed using cubic spline.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

9701 - **BOFARINE, H.; ARELLANO-VALE, R.B.** Weak nondifferential measurement error models. IME-USP, 1997. 12p. (RT-MAE-9701)

9702 - **FERRARI, S.L.P.; CORDEIRO, G. M.; CRIBARI-NETO, F.** Higher Order Asymptotic Refinements for Score Tests in Proper Dispersion Models. IME-USP, 1997. 14p. (RT-MAE-9702)

9703 - **DOREA, C.; GALVES, A.; KIRA, E.; ALENCAR, A. P.** Markovian modeling of the stress contours of Brazilian and European Portuguese. IME-USP, 1997. 11p. (RT-MAE-9703)

9704 - **FONTES, L.R.G.; ISOPI, M.; KOHAYAKAWA, Y.; PICCO, P.** The Spectral Gap of the REM under Metropolis Dynamics. IME-USP, 1997. 24p. (RT-MAE-9704)

9705 - **GIMENEZ, P.; BOLFARINE, H.; COLOSIMO, E.A.** Estimation in weibull regression model with measurement error.IME-USP, 1997. 17p. (RT-MAE-9705)

9706 - **GALVES, A.; GUIOL, H.** Relaxation time of the one-dimensional symmetric zero range process with constant rate. IME-USP, 1997. 10p. (RT-MAE-9706)

9707 - **GUIOL, H.** A note about a burton keane's theorem. IME-USP, 1997. 7p. (RT-MAE-9707)

9708 - **ARELLANO-VALLE,R.B.; FERRARI,S.L.P.; CRIBARI-NETO,F.** Bartlett and barblett-type corrections for testing linear restrictions. IME-USP, 1997. 5p. (RT-MAE-9708)

9709 - **PAULA, G.A.** One-Sided Tests in Dose-Response Models. IME-USP, 1997. 29p. (RT-MAE-9709)

9710 - **BRESSAUD, X.** Subshift on an infinite alphabet. IME-USP, 1997. 34p. (RT-MAE-9710)

9711 - **YOSHIDA, O.S.; LEITE, J.G.; BOLFARINE, H.** Stochastic Monotonicity properties of bayes estimation of the population size for capture-recapture data. IME-USP. 1997. 12p. (RT-MAE-9711)

9712 - **GUIMENEZ,P.;COLOSIMO,E.A.; BOLFARINE,H.** Asymptotic relative efficiency of wald tests in measurement error models. IME-USP. 1997. 14p. (RT-MAE-9712)

9713 - **SCHONMANN, R.H.; TANAKA, N.I.** Lack of monotonicity in ferromagnetic ising model phase diagrams. IME-UPS. 1997. 12p. (RT-MAE-9713)

The complete list of ''Relatórios do Departamento de Estatística'', IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05315-970 - São Paulo, Brasil