PARTIAL POSTERIOR CONSISTENCY USING DIRICHLET PRIORS IN ESTIMATION OF SURVIVAL FUNCTION

by

N. D. Shyamalkumar
Purdue University

Technical Report #96-21

Department of Statistics
Purdue University
West Lafayette, IN USA

June 1996

PARTIAL POSTERIOR CONSISTENCY USING DIRICHLET PRIORS IN ESTIMATION OF SURVIVAL FUNCTION

N.D. Shyamalkumar

Purdue University *

Abstract

Suppose that one has a random sample from a survival function of the form $(1 - F_0)(1 - G)$ where F_0 is known and G is unknown. We study the problem of Bayesian estimation of G in this paper. It is known that the GMLE of G and the Bayesian estimator, using the Dirichlet prior for G, are inconsistent for estimating G when the distributions are continuous. To derive consistent Bayesian estimators we, under the assumption that G is absolutely continuous unimodal distribution, put a prior on G which concentrates on absolutely continuous unimodal distributions by using the Dirichlet as a prior on the Khinchine measure. It is shown that in this case one can show consistency for a large class of sampling distributions, the class depending on the parameter of the Dirichlet prior. We discuss an importance sampling method for computing the estimate. An example is discussed where this estimator compares favorably with another estimator from the literature.

Key words and phrases: Survival analysis, unimodal distributions, Bayesian nonparametrics, Bayesian consistency, Dirichlet process prior.

*Research supported by Purdue Research Foundation Grant and partially supported by the National Science Foundation, Grants DMS-8923071 and DMS-9303556.

1 Introduction

Let us consider the problem of estimating the distribution function, given a random sample, under the constraint that the true distribution function is in the class of all distributions uniformly stochastically smaller than a known given distribution. For two distribution functions, F and G, on $[0, \infty)$ we shall say that F is uniformly stochastically smaller than G and denote it by $F <_{(+)} G$ if $(1 - G)/(1 - F)$ is non-decreasing on the support of F. This problem is of interest to us as in our problem, we are basically restricting the sampling distribution to not only being uniformly stochastically smaller than F_0 but also that it is of the specific form $\bar{F}_0\bar{G}$, where G is some distribution function and $\bar{G}$ represents its survival function. It is worth noting that, in the definition of uniformly stochastically smaller, if one further imposes the condition that the ratio $\bar{G}/\bar{F}$ is not only nondecreasing but approaches infinity as $\bar{F}$ tends to zero, then we are basically requiring F to have the above form.

There is some literature on estimating the distribution function under a constraint of uniform stochastic ordering. The assumption that F_0 is known is reasonable when the life testing is done in a controlled environment which permits it's estimation with sufficient accuracy. It was shown in Rojo and Samaniego (1990) that the nonparametric maximum likelihood estimator, NPMLE, for H is inconsistent under the constraint $H <_{(+)} F_0$, for F_0 increasing and continuous. There it was shown that the NPMLE for H is of the form $1 - \bar{F}_0 \cdot \bar{H}_n$, where H_n is the empirical distribution function, which converges to the wrong limit $1 - \bar{F}_0\bar{H}$, instead of H.

In Shyamalkumar (1996), it is shown that using the Dirichlet as a prior for G results in an estimator which is inconsistent for estimating G, when the distributions are continuous. The problem was that the Dirichlet puts all its mass on discrete distributions. In the third section we try to find a Bayesian estimator of the survival function when one restricts G to the class of unimodal distributions. We use the method of Lo (1984) to do this, i.e. by making use of the Khintchine theorem. The derivation of the estimator does not pose any problem but its form is pretty unyielding to standard theoretical analysis. Moreover, the consistency of the posterior derived using such a prior is not known, beyond the Doob (1949) result. We give a proof of the consistency for a certain class of distributions. Moreover our condition for consistency is very simple. One way of looking at our results is that, if one knows an upper bound on the tail of the true survival function, then one can use this bound to come up with a parameter for the Dirichlet prior such that posterior consistency holds for all survival functions which satisfy the upper bound. The other way is that, if we use a certain parameter for the Dirichlet prior, then one is guaranteed consistency for all the distributions which satisfy a certain upper bound (a function of the parameter) on their tail. The first way of looking at it is very Bayesian in some sense. If you have prior knowledge, then incorporating it in the prior assures better behavior of the estimator. The proof is long but straightforward though the pre-requisites may be high.

It can be seen that we are in fact working with a prior on densities induced by convolving the distributions from a Dirichlet with a kernel of random width, $K(x,\eta)$ given by

$$K(x,\eta) = \begin{cases} 0 & \eta < x \\ f_0(x) + \frac{\bar{F}_0(x) - x f_0(x)}{\eta} & \eta \geq x \end{cases}$$

where F_0 is a distribution on $(0,\infty)$ satisfying certain conditions, which are satisfied, for example, when F_0 is IFR with continuous density. In this case it is shown that for suitable choices of the parameter of the Dirichlet prior we have strong local matching for a large class of likelihoods. From this we deduce strong consistency for all likelihoods for which the strong local matching is attained with the prior under consideration. This answers, partially, some open questions posed in Barron (1986). We believe that such results could be proved for a more general class of kernels, by following a similar approach, as is adopted in this paper. This generalization would be carried out elsewhere.

In the fourth section, we discuss a computation scheme for the estimator. Though the scheme is base on that mentioned in Escobar (1994), we have to make some important modifications because of the form of our integrand. One such is the choice of the importance sampling density and the other is working after ordering of the data vector, $\underset{\sim}{Z}$. We explain these with an example and compare our estimate with that of Rojo and Samaniego (1993).

2 Pre-requisites in Bayesian Nonparametrics

2.1 Dirichlet Process Prior

The Dirichlet process prior is a prior on the space of probability measures on a measurable space. Let $\mathcal{X}$ be a set and $\mathcal{A}$ be a sigma field of subsets of $\mathcal{X}$. According to the definition in Ferguson (1973), a Dirichlet Process with parameter α, denoted by $\mathcal{D}_\alpha$, whe re α is a finite measure on $(\mathcal{X}, \mathcal{A})$ is a random process, P,

indexed by elements of $\mathcal{A}$ with the property that for every positive integer k, and every measurable partition $\{A_1, \ldots, A_k\}$, the random vector $(P(A_1), \ldots, P(A_k))$ has a k-dimensional Dirichlet distribution with parameter vector $(\alpha(A_1), \ldots, \alpha(A_k))$.

It is worthwhile to know that in the case when $\mathcal{X}$ is the set of natura l numbers and $\mathcal{A}$ is the power set then the Dirichlet process prior is same as *stick breaking* with the proportion of successive breaks being distributed as independent Beta distributions with parameter depending on the parameter of the Dirichlet prior.

One of the most important results on the Dirichlet process prior is the following from Ferguson (1973). This is what makes it mathematically tractable for a Bayesian.

Proposition 2.1 (Ferguson) *If P is a Dirichlet process with parameter α, and if given P, $X_1, \ldots, X_n$ is a random sample from P, then the posterior distribution of P given the random sample is also a Dirichlet process but with parameter $\alpha + \sum \delta_{X_i}$ where $\delta(x)$ represents a point mass distribution at x.*

Now we give an alternative constructive definition of a Dirichlet process prior given in Sethuraman (1994). The following definition is simpler and the existence of the process follows immediately with topological assumptions on the measurable space unlike the Ferguson definition.

Proposition 2.2 (Sethuraman) *Let $Y_1, Y_2, \ldots$ be i.i.d. with a Beta distribution, $\mathcal{B}(M, 1)$, $M > 0$, and $X_1, X_2, \ldots$ be i.i.d. from α_0. Both the Y and Z sequences are independent. Let $P := \sum P_j \delta(X_j)$, where $P_1 := Y_1$ and $P_j := (1 - Y_j) \prod_1^{j-1} Y_k$, $j > 1$. Then P is a Dirichlet process with parameter $M\alpha_0$.*

We end this subsection by giving a fact about Dirichlet process priors which we shall need in the derivation of the estimator we seek.

Fact 2.1 *The joint marginal distribution of X_i's is given by the following.*

$$\begin{aligned} X_1 &\sim \alpha_0 \\ X_i \mid X_1, \ldots, X_{i-1} &\sim \frac{\alpha + \sum_{j=1}^{i-1} \delta_{X_j}}{M + i - 1} \qquad \textit{for } i = 2, \ldots, n. \end{aligned}$$

PROOF. Follows from elementary facts about Dirichlet process priors. □

There are a host of other important and interesting results on the Dirichlet process prior available in the literature. We shall, as and when we need, mention some of them at the appropriate places later in the paper.

2.2 Posterior Consistency

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and $\mathcal{P}$ be the set of all probability measures on $(\mathcal{X}, \mathcal{A})$. Let $\{P_\theta : \theta \in \Theta\}$ be an indexed class of probability measures.

For a probability measure P on $(\mathcal{X}, \mathcal{A})$ we shall denote by P^∞ the countable product of P, a probability measure on $(\mathcal{X}^\infty, \mathcal{A}_\infty)$. Now let π be a probability measure on Θ with some associated sigma-field. Let π_n denote the posterior after observing the first n observations. We would say that (θ, π) is consistent if

$$\pi_n(N_\theta) \to 1 \text{ a.e.}(P_\theta^\infty),$$

for every neighborhood N_θ of θ. When we talk about posterior consistency we must assume some underlying topology on the parameter space. It is natural to endow the parameter space with a topology which is induced by the mapping $\theta \to P_\theta$ where the topology on $\{P_\theta : \theta \in \Theta\}$ is the subset topology from that on $\mathcal{P}$. So it is clear that we should be looking for a reasonable topology on $\mathcal{P}$ so that consistency is meaningful. In the usual case when $(\mathcal{X}, \mathcal{A})$ happens to be a Polish space, i.e. a complete separable metric space with its associated Borel sigma-field, it is natural to endow $\mathcal{P}$ with the weak topology, which also happens to be a complete separable metric topology. If one achieves posterior consistency fo r this weak topology it would be meaningful and moreover it is weaker than all the other natural topologies, such as the total variation, on $\mathcal{P}$. In other words with the weak-topology it is easiest to achieve meaningful consistency. Consistency with weak-topology would imply classical consistency of Bayesian estimators of all functions of θ which would be bounded and continuous in the above described induced topology. In this paper we shall only be working with the consistency based on the weak-topology.

3 Restricted Bayesian Estimator - Continuous Case

3.1 Introduction

In this section we try to construct a Bayesian estimator which has good asymptotic behavior in the sense of consistency. To do this we restrict the support of the prior on G. The restriction we impose is that of unimodality. To put a prior on the space of all unimodal distribution we adopt a scheme of Lo (1984). To describe this scheme we recall a result by Khintchine, see Feller (1965), which states that any unimodal distribution can be represented as a scale mixture of uniform distributions. Note that the above is a type of Choquet representation theorem. For a unimodal distribution we shall call the mixing measure in the above representation as the Khintchine measure associated with the unimodal distribution. Note that an unimodal distribution is absolutely continuous w.r.t. Lebesgue measure on $\mathbb{R}^+$ except possibly for a mass at zero. The scheme of Lo (1984) is to put a Dirichlet prior on the Khintchine measure to arrive at a probability measure on the set of all unimodal distributions. Note that almost surely w.r.t. the Dirichlet prior (with a continuous measure as parameter) the point zero is assigned zero mass and hence almost surely all realizations from our prior are absolutely continuous unimodal probability measures. This prior measure that we assign to G shall be denoted by $\mathcal{D}^u_\alpha$.

Below is the summary of the model we shall be working with in this section.

- $\{Y_i\}_1^n$ a random sample from F_0.
- P has a Dirichlet law with parameter α denoted by $\mathcal{D}_\alpha$. α is assumed to be absolutely continuous w.r.t. Lebesgue measure with density α'.
- Conditioned on P, $\{\eta_i\}_1^n$ is a random sample from P.
- Conditioned on $\{\eta_i\}_1^n$, X_i has a $U_{(0,\eta_i)}$ law for $i = 1, \ldots, n$, $\{X_i\}_1^n$ independent.
- $Z_i := X_i \wedge Y_i$, for $i = 1, \ldots, n$.

The problem is to estimate the law of X denoted by G, which is assumed to be unimodal with no point mass at zero.

Before we go further into the technical details we shall try to motivate why we would expect a more sensible behavior of the estimate of G with this type of restriction and prior. For this we need to recall some facts about the I_0 divergence and the total variation metric, which we give below.

For two probability measures, P and Q, on a measure space, we shall denote by $I_0(P,Q)$, the Kullback-Leibler I_0 divergence between the two measures, defined by,

$$I_0(P,Q) = \int \log\left(\frac{dQ}{dP}\right)\, dQ$$

where the densities are with respect to some dominating measure. It is easy to see that this divergence is invariant w.r.t. to the choice of the dominating measure. Here and in the sequel we understand

$$\log 0 = -\infty, \quad \log\left(\frac{a}{0}\right) = +\infty, \quad 0 \cdot (\pm\infty) = 0,$$

where $a > 0$. $I_0(P,Q)$ is always non-negative and vanishes only when $P = Q$. If $Q \not\ll P$ then $I_0(P,Q) = +\infty$.

Let us recall that the total variation distance between two probability measur es P and Q is given by

$$\|P - Q\| = 2 \cdot \sup_A |P(A) - Q(A)|$$

where the supremum is taken over all sets in the σ-field on which the probability measures are defined. The mysterious factor 2 is included to make it equal to the L_1 distance between the densities of the two measures with respect to any dominating σ-finite measure. Note that the L_1 metri c is invariant to the choice of the dominating measure.

The below given information inequality was independently discovered by many in the late 1960's; among them being Csiszár (1967), Kemperman (1967) and Kullback (1967). We refer to them for its proof which follows essentially from the monotonicity property of the I- divergence.

Lemma 3.1 *Let P and Q be two probability measures. Then with the above notations we have*

$$I_0(P,Q) \geq \frac{(\|P-Q\|)^2}{2}.$$

By symmetry the above is true with I_1 replacing I_0.

PROOF. See Csiszár (1967).

□

As we shall see later one of the factors that guarantees consistency of the posterior is the richness of its support. Now the support of a probability measure is dependent on the topology one prescribes to. The usual convention is to talk of the weak support, i.e. the support w.r.t. weak topology on the space of probability measures. This is necessary and sufficient to guarantee consistency when we are talking about the case of a finite population, see Freedman (1963). But in other cases it can be grossly insufficient. The appropriate *topology* as we shall see later is the one defined by the I_0 divergence. Before we go further let us give two definitions for the support of a probability measure on, say, $(\mathcal{X}, \mathcal{A})$ which we assume to be a Polish space, i.e. $\mathcal{X}$ has a complete seperable metric associated with it and $\mathcal{A}$ is the Borel sigma field w.r.t. this topology.

1. The smallest closed set C which has probability one under P is called the support of P.

2. A point $x \in \mathcal{X}$ is said to belong to the support of P if every open ball around x is assigned a positive probability by P.

The above two definitions are equivalent when the topology is separable and metrizable, see Parthasarathy (1967). But when the space has cardinality of at least the continuum it is easy to see that the topology induced by the total variation metric is not separable. In this case it can happen that under definition 1 the support does not exist. To avoid such problems we shall in what follows use the second definition. We state a lemma below concerning the support of the two priors mentioned so far w.r.t. the total variation metric.

Lemma 3.2 *The total variation support of the $\mathcal{D}_\alpha$ is the empty set and the total variation support of $\mathcal{D}_\alpha^u$, when α has the whole of $\mathbb{R}^+$ as its support, contains the set of all absolutely continuous unimodal distributions on $\mathbb{R}^+$. Hence, the support w.r.t. the I_0 divergence of $\mathcal{D}_\alpha$ is also empty.*

PROOF. Let us prove the statement regarding the support of $\mathcal{D}_\alpha^u$ first. Let U_P be an absolutely continuous unimodal distribution. As the weak support of the Dirichlet process prior is the whole space of probability measures on $\mathbb{R}^+$, and as the mapping $P \rightarrow U_P$ is continuous with the weak topology on the left and the total variation topology on the right (proved using Scheffe's theorem), we have the desired result. Clearly, by the discreteness of the probabilities from $\mathcal{D}_\alpha$, we have to consider only discrete measures as candidates for the support. But, by the representation of the Dirichlet given in Sethuraman (1994), we see that neither does any discrete measure belong to it's support. From the previous fact, Fact 3.1, it is clear that the I_0 divergence shall induce a stronger *topology* than the total variation and hence a smaller support. This ends the proof. □

The above lemma along with the preceding comments explains in some sense the inconsistency observed in Shyamalkumar (1996) and gives us some hope for the the success of the present approach.

We need a few more notations in this section which we list below. Below M is a measure on $\mathbb{R}^+$ and A is a Borel subset of $\mathbb{R}^+$.

- By M_A we shall denote the measure M restricted to the subset A, i.e. $M_A(B) = M(A \bigcap B)$ for any Borel subset B of $\mathbb{R}^+$. By M_{A^c} we shall denote the measure M restricted to A^c.

- By M_a, for $a \in \mathbb{R}^+$, we shall denote the measure M restricted to $[0, a]$. By M_{a^c}, for $a \in \mathbb{R}^+$, we shall denote the measure M restricted to (a, ∞).

- For a finite measure M we shall denote by M^* the normalized measure derived from M.

- For a finite measure M, $M(a) := M([0, a])$ and $\bar{M}(a) := 1 - M(a)$.

- By U_P, for P a probability measure on $\mathbb{R}^+$, we shall denote the unimodal distribution with Khintchine measure P. We shall denote the density of it's absolutely continuous part by u_P.

- h_P is defined by

$$h_P(x) := \bar{F}_0(x) u_P(x) + \bar{U}_P(x) f_0(x), \quad \text{for } x \in \mathbb{R}^+.$$

  When $P = \delta_\eta(\cdot)$, we shall denote h_P simply by h_η.

- For $\{P_n\}$ a sequence of probability measures on some Polish space we shall use the notation

$$P_n \Rightarrow P_\infty$$

  to denote convergence of the sequence in the weak topology to P_∞.

We state below an elementary result on the Dirichlet process which will be of critical use when we prove the consistency of the estimator in the later sub-sections.

Fact 3.1 (Projection lemma) *Let A be a Borel subset of $\mathbb{R}^+$. Let P be a probability from $\mathcal{D}_{\alpha_A}$, Q be a probability from $\mathcal{D}_{\alpha_{A^c}}$ and p be drawn from $Beta(\alpha(A), \alpha(A^c))$. Let P, Q and p be mutually independent. Then the law of $p \cdot P + (1 - p) \cdot Q$ is $\mathcal{D}_\alpha$.*

PROOF. The proof follows easily by working backwards. □

Before we proceed further with the above setup, we shall show below that in a similar setup in the discrete case we do not have tail-free property of the prior, unlike the unrestricted case. This shall also provide justification for the route we take to establish consistency.

We shall, in the remaining part of this subsection, assume that F_0 and G are distributions on $\{1, 2, \ldots\}$ and we shall put a prior on G as a mixture of discrete uniforms, with the mixing distribution having a Dirichlet prior with α, a finite measure on $\{1, 2, \ldots\}$.

First note that $\bar{H} = \bar{F}_0 \bar{U}_Q$, where Q is given a $\mathcal{D}_\alpha$ prior. Now after some simple algebra, one can see that

$$H\{k\} = F_0\{k\} \cdot \bar{U}_Q(k) + U_Q\{k\} \cdot \bar{F}_0^+(k),$$
$$\frac{H\{k\}}{\bar{H}^+(k)} = \frac{U_Q\{k\}}{\bar{U}_Q^+(k)} + \frac{F_0\{k\}}{\bar{F}_0^+(k)} \cdot \frac{\bar{U}_Q\{k\}}{\bar{U}_Q^+(k)}.$$

Now, as in the proof of the tail-free property in the unrestricted case, we fix $\theta_0 := 0$ and θ_i for $i = 1, 2, \ldots$ are independent random variables following $Beta(\alpha\{i\}, \sum_{i+1}^{\infty} \alpha\{j\})$. Then we can write

$$\bar{U}_Q(k) = \prod_0^k (1 - \theta_j) \left[1 - k \cdot \sum_{k+1}^{\infty} \left(\frac{\theta_j (\prod_{m \geq (k+1)}^{j-1} (1 - \theta_m))}{j} \right) \right],$$

$$U_Q\{k\} = \prod_0^{k-1} (1 - \theta_j) \left[\sum_k^{\infty} \left(\frac{\theta_j (\prod_{m \geq k}^{j-1} (1 - \theta_m))}{j} \right) \right].$$

From the above it is clear that the independence that is required for the tail-free property does not hold.

3.2 The Estimator

We shall in this sub-section derive the estimator of G which we shall denote by $\hat{G}_{U,n}$. The loss function has the same form as in the previous section and hence the estimator is the posterior expectation of the corresponding distribution function. Below we shall for an, n-dimensional vector, say $\underset{\sim}{V}$, denote by $\underset{\sim}{V}_{-i}$ the $n-1$-dimensional vector with the i-th coordinate of $\underset{\sim}{V}$ removed. We shall for $u \in \mathbb{R}^+$ and an n-dimensional vector $\underset{\sim}{V}$ denote by $\psi(u; V_i, \underset{\sim}{V}_{-i})$ the expression

$$\frac{\int I_{\{\eta_i \leq u\}}[\prod_{j=1}^{n} h_{\eta_j}(V_j)] \prod_{j=1}^{n} [\alpha(d\eta_j) + \sum_{k=1}^{j-1} \delta_{\eta_k}(d\eta_j)]}{\int [\prod_{j=1}^{n} h_{\eta_j}(V_j)] \prod_{j=1}^{n} [\alpha(d\eta_j) + \sum_{k=1}^{j-1} \delta_{\eta_k}(d\eta_j)]}.$$

Theorem 3.1 *The Bayes estimator for G, denoted by $\hat{G}_{U,n}$, under the forementioned model, is given by*

$$\hat{G}_{U,n}(u) = \frac{\alpha(u)}{M+n} + \frac{\sum_{i=1}^{n} \psi(u; Z_i, \underset{\sim}{Z}_{-i})}{M+n}, \quad \text{for } u \in \mathbb{R}^+.$$

PROOF. Let u be an arbitrary positive real number. Note that

$$\hat{G}_{U,n}(u) = E(P(u)|\underset{\sim}{Z}) = E(E(P(u)|\underset{\sim}{\eta}, \underset{\sim}{Z})|\underset{\sim}{Z}).$$

Observe that, from elementary properties of the Dirichlet prior, we have

$$E(P(u)|\underset{\sim}{\eta}, \underset{\sim}{Z}) = E(P(u)|\underset{\sim}{\eta}) = \frac{\alpha(u)}{M+n} + \frac{\sum_{i=1}^{n} I_{\{\eta_i \leq u\}}}{M+n}.$$

Hence we have

$$\hat{G}_{U,n}(u) = \frac{\alpha(u)}{M+n} + \frac{\sum_{i=1}^{n} Prob(\eta_i \leq u|\underset{\sim}{Z})}{M+n}.$$

We note that $\{(\eta_i, Z_i)\}_1^n$ is an exchangeable sequence. Hence it is sufficient to find $Prob(\eta_1 \leq u|\underset{\sim}{Z})$ to be able to explicitly write down the form of $\hat{G}_{U,n}(u)$. Now by using Fact 2.1 we see that the joint density of $(\underset{\sim}{Z}, \underset{\sim}{\eta})$ is given by

$$C(M,n) \prod_{i=1}^{n} [h_{\eta_i}(dZ_i)[\alpha(d\eta_i) + \sum_{j=1}^{i-1} \delta_{\eta_j}(d\eta_i)]].$$

Hence we get

$$Prob(\eta_1 \leq u|\underset{\sim}{Z}) = \frac{\int I_{\{\eta_1 \leq u\}}[\prod_{i=1}^{n} h_{\eta_i}(Z_i)] \prod_{i=1}^{n} [\alpha(d\eta_i) + \sum_{j=1}^{i-1} \delta_{\eta_j}(d\eta_i)]}{\int [\prod_{i=1}^{n} h_{\eta_i}(Z_i)] \prod_{i=1}^{n} [\alpha(d\eta_i) + \sum_{j=1}^{i-1} \delta_{\eta_j}(d\eta_i)]}.$$

Note that the above is precisely $\psi(u; Z_i, \underset{\sim}{Z}_{-i})$ and the notation is justified because of exchangeability. From this the result of the theorem follows.

Remark 3.1 We note that as we let M approach zero the estimator converges to

$$\frac{\int_0^u \prod_{i=1}^n h_\eta(Z_i)\alpha_0(d\eta)}{\int_0^\infty \prod_{i=1}^n h_\eta(Z_i)\alpha_0(d\eta)}, \quad \text{for } u \in \mathbb{R}^+.$$

The above is precisely the Bayes estimate of $I_{\eta \le u}$ when one is working with the parametric model in which the distribution of Y is an uniform with unknown scale η. The proof of the above is easily accomplished by observing that the required limit is given by

$$\lim_{\alpha(\mathbb{R}^+)\to 0} \left\{ \frac{\int P(u) \cdot \prod_{i=1}^n h_P(Z_i)\mathcal{D}_\alpha(dP)}{\int \prod_{i=1}^n h_P(Z_i)\mathcal{D}_\alpha(dP)} \right\}.$$

As one has weak convergence of the Dirichlet when one lets M approach zero, with the limiting distribution assigning all probability to point masses with the distribution of the point mass being α_0 (see Sethuraman and Tiwari (1982)) and the integrands are bounded continuous functions (actually, the set of continuity points has measure one w.r.t. the limiting distribution) we have the proof. The interesting part is that the above can be consistent only when the true Khintchine measure is degenerate. This is easily seen by the result of Berk (1966) which implies that the above converges to $I_{\eta * \le u}$ where $\eta *$ is the unique value of η for which $I_0(h_\eta, h^0)$ is minimized, where h^0 is the true distribution. Of course one has to have the existence of a unique η^*.

3.3 Consistency of the Estimator

In all that follows whenever we talk about a probability measure which is used as a Khintchine measure then we shall assume that it has no mass at 0. In this sub-section we shall need some assumptions which are given below.

Assumptions

(A1) F_0 has a strictly positive continuous density, denoted by f_0, such that for some $K > 0$ we have

$$x \cdot \bar{F}_0(x) \downarrow, \qquad \text{for } x > K.$$
$$\tfrac{x \cdot f_0(x)}{\bar{F}_0(x)} \uparrow, \qquad \text{for } x > K.$$

(A2) α is such that it's support is the whole of $\mathbb{R}^+$.

Remark 3.2 Note that (A1) holds for all IFR distributions with continuous densities. Note that every distribution with logconcave density is IFR.

Remark 3.3 We shall assume that $\bar{Q}(a) > 0$ for all $a > 0$, as otherwise consistency of the posterior can be achieved easily (because of compactness) under assumption (A4) given later in this sub-section.

Lemma 3.3 *Under (A1) and P any probability measure the following are true.*

$$h_P(x) \leq \bar{P}(x) f_0(x), \quad \forall \quad x > K \tag{1}$$
$$h_P(x) \geq \frac{\bar{F}_0(x) \cdot \bar{P}(x)}{x}, \quad \forall \quad x > K. \tag{2}$$

PROOF. We first note that, for any probability measure P, we have

$$\bar{U}_P(x) = \bar{P}(x) - x \cdot u_P(x).$$

Using the above we have

$$h_P(x) = [\bar{F}_0(x) - x \cdot f_0(x)] u_P(x) + \bar{P}(x) \cdot f_0(x). \tag{3}$$

Now by (A1) we have by calculus that

$$\bar{F}_0(x) - x \cdot f_0(x) \leq 0, \qquad \forall x > K. \tag{4}$$

Now using the above in (3) we have (1). For the lower bound note that

$$u_P(x) = \int_x^\infty \frac{1}{a} dP(a) \leq \frac{\bar{P}(x)}{x}.$$

Using (4) and the above in (3) we have (2). □

Lemma 3.4 *Let P and Q be two probability measures and $\epsilon > 0$. Then there exists a $B(Q,\epsilon) > K$ such that, if for $a \geq B(Q,\epsilon)$ we have*

$$-\log P(a) \leq \epsilon/2 \quad \textit{and} \quad \bar{P}(x) \geq \frac{\bar{Q}(x) \cdot x \cdot f_0(x)}{\bar{F}_0(x)}, \quad \forall x \geq a$$

and

$$h_a(x) := \frac{P(a)}{Q(a)} \cdot h_{Q_a}(x) + h_{P_{a^c}}(x), \quad \textit{for } x \in \mathbb{R}^+,$$

then

$$\int \left[\log\left(\frac{h_Q(x)}{h_a(x)}\right)\right]^+ h_Q(x)\,dx \leq \epsilon.$$

Proof. Choose $B(Q,\epsilon) > K$ such that

$$\left[\log\left(\frac{h_Q(x)}{h_a(x)}\right)\right]^+ \leq \epsilon, \ \forall x \leq K, a \geq B(Q,\epsilon).$$

This is so because

$$\begin{aligned}
\frac{h_Q(x)}{h_a(x)} &\leq \frac{h_{Q_a}(x) + h_{Q_{a^c}}(x)}{\frac{P(a)}{Q(a)} \cdot h_{Q_a}(x)} \\
\inf_{x\in[0,K]} h_{Q_a}(x) &\geq \inf_{x\in[0,K]} f_0(x)\bar{U}_{Q_a}(K) > 0 \uparrow \quad \text{with } a \\
\sup_{x\in[0,K]} h_{Q_a^c}(x) &\leq \frac{\bar{Q}(a)}{a} + \sup_{x\in[0,K]} f_0(x)\bar{Q}(a) \downarrow 0 \quad \text{with } a.
\end{aligned}$$

The dependence of B on only Q and ϵ can be seen from the above. By Lemma 3.3 we have

$$\begin{aligned}
h_Q(x) &\leq \bar{Q}(x) f_0(x), \quad \forall \quad x > K \\
h_P(x) &\geq \frac{\bar{P}(x) \cdot F_0(x)}{x}, \quad \forall \quad x > K.
\end{aligned}$$

Now for $x \in (K,a)$ we have by using the above that

$$\frac{h_Q(x)}{h_a(x)} = \frac{h_{Q_a}(x) + h_{Q_{a^c}}(x)}{\frac{P(a)}{Q(a)} \cdot h_{Q_a}(x) + h_{P_{a^c}}(x)} \leq \frac{h_{Q_a}(x) + \bar{Q}(a) \cdot f_0(x)}{\frac{P(a)}{Q(a)} \cdot h_{Q_a}(x) + \frac{\bar{P}(a)\cdot\bar{F}_0(x)}{x}} \leq \frac{Q(a)}{P(a)}.$$

Similarly for $x \geq a$ we have

$$\frac{h_Q(x)}{h_a(x)} = \frac{h_{Q_{a^c}}(x)}{h_{P_{a^c}}(x)} \leq \frac{\bar{Q}(x) f_0(x)}{\frac{\bar{P}(x)\cdot \bar{F}_0(x)}{x}} \leq 1.$$

Hence we have

$$I_0(h_a, h_Q) \leq \int \left[\log\left(\frac{h_Q(x)}{h_a(x)}\right)\right]^+ h_Q(x)\,dx \leq \left[\epsilon \vee \log\left(\frac{Q(a)}{P(a)}\right)\right] \leq \epsilon.$$

□

Lemma 3.5 *Let*

$$T_Q^a(P) := \int_0^a \log\left(\frac{h_Q(x)}{h_P(x)}\right) h_Q(x)\,dx$$

for two probability measures Q and P and some positive real a. Let Q be such that $\bar{Q}(a) > 0$, $Q\{0\} = 0$ and h_Q is bounded. Let $\{P_n\}$ be a sequence of probability measures such that $\{h_{P_n}\}$ is uniformly bounded away from zero on $[0, a]$ and such that $P_n \Rightarrow P_\infty$, with $P_\infty\{0\} = 0$. Then

$$\lim_{n\to\infty} T_Q^a(P_n) = T_Q^a(P_\infty).$$

Proof. Note that, by definition of weak topology, we have

$$P_n \Rightarrow P_\infty \quad \leadsto \lim_{n\to\infty} h_{P_n}(x) = h_{P_\infty}(x) \quad \text{a.s..}$$

The above convergence, by Scheffe's theorem, also takes place in L_1. Using boundedness of $x\log(x)$ on compacts and writing

$$T_Q^a(P_n) = \int_0^a \left(\frac{h_Q(x)}{h_{P_n}(x)}\right) \log\left(\frac{h_Q(x)}{h_{P_n}(x)}\right) h_{P_n}(x)\,dx,$$

we have the lemma by uniform integrability. □

Lemma 3.6 *Let Q satisfy the conditions of Lemma 3.5 and let $a > B(Q, \epsilon/2)$, $B(Q, \epsilon/2)$ from Lemma 3.4, be such that*

$$-\log P(a) \leq \epsilon/4 \quad \textit{and} \quad \bar{P}(x) \geq \frac{\bar{Q}(x)\cdot x \cdot f_0(x)}{\bar{F}_0(x)}, \quad \forall x \geq a,$$

for some $\epsilon > 0$. Then for R, a probability measure on $\mathbb{R}^+$, if we define

$$h_a^*(R, x) := P(a) h_R(x) + h_{P_{a^c}}(x), \quad \forall x$$

and

$$S_{Q,P}^a(R) := \int_0^a \log\left(\frac{h_Q(x)}{h_a^*(R,x)}\right) h_Q(x)\, dx,$$

we have $\mathcal{D}_{\alpha_a}\{R : S_{Q,P}^a(R) < \epsilon\} > 0$.

PROOF. By Lemma 3.4 we have for $a > B(Q, \epsilon/2)$ that

$$\int \left[\log\left(\frac{h_Q(x)}{h_a(x)}\right)\right]^+ h_Q(x)\, dx \le \frac{\epsilon}{2}.$$

Now note that, using Lemma 3.3, we have

$$h_a^*(R, x) \ge h_{P_{a^c}}(x) \ge \bar{U}_{P_{a^c}}(a) \cdot \inf_{x \in [0,a]} f_0(x) > 0, \quad \forall x \in [0, a].$$

Hence $h_a^*(R, \cdot)$ is uniformly bounded away from zero on $[0, a]$, uniformly in R. Using Lemma 3.5 we see that $S_{Q,P}^a(R)$ is weakly continuous in R at Q_a^*. Hence there exists a weak open ball, say N, around Q_a^*, such that

$$S_{Q,P}^a(R) < \epsilon, \quad \forall R \in N.$$

But since α_a has support as $[0, a]$, we have the whole space of probability measures on $[0, a]$ to be the support of $\mathcal{D}_{\alpha_a}$, see Ferguson (1973). The last fact implies that $\mathcal{D}_{\alpha_a}(N) > 0$. Hence the result. □

We shall need the following assumptions in the following.

Assumptions

(A3) Let Q be a probability measure such that $Q\{0\} = 0$ and

$$\bar{Q}(t) \le \frac{\bar{F}_0(t)}{t \cdot f_0(t)} \cdot \exp\left[\frac{-2 \log|\log \bar{\alpha}_0(t)|}{\bar{\alpha}_0(t)}\right], \quad \forall t > K.$$

(A4) Let Q be a probability measure such that h_Q is bounded.

Proposition 3.1 (Doss & Sellke) *Let Q satisfy (A3). Then for almost every P from $\mathcal{D}_\alpha$, we have*

$$\frac{\bar{Q}(t) \cdot t \cdot f_0(t)}{\bar{F}_0(t)} \leq \bar{P}(t), \quad \textit{for large } t.$$

PROOF. Follows from Doss and Sellke (1982). □

Lemma 3.7 *Define, for an $a \in \mathbb{R}^+$ and probability measure Q, an operator on the space of probability measures on $\mathbb{R}^+$ by,*

$$W_Q^a(P) := \int_a^\infty \left[\log\left(\frac{h_Q(x)}{h_P(x)}\right)\right]^+ h_Q(x)\, dx.$$

Then for a given $\epsilon > 0$ we can choose a $K_1 > 0$ such that

$$\mathcal{D}_\alpha\left\{P : \bar{P}(x) \geq \frac{\bar{Q}(x) \cdot x \cdot f_0(x)}{\bar{F}_0(x)}, \ \ \forall\, x \geq a;\ \bar{P}(a) < \epsilon;\ W_Q^a(P) = 0\right\} > 0, \ \ \forall\, a > K_1.$$

PROOF. Now since the Dirichlet process is a measure on the space of probability measures we have that

$$\lim_{x \to \infty} \mathcal{D}_\alpha\{P : \bar{P}(x) < \epsilon\} \to 1.$$

Hence we can choose a $C_1 > 0$ such that $\mathcal{D}_\alpha\{P : \bar{P}(a) < \epsilon\} > 1 - \epsilon/2$, for $a > C_1$. Now by Lemma 3.3 we have

$$\frac{h_Q(x)}{h_P(x)} \leq \frac{\bar{Q}(x) \cdot x \cdot f_0(x)}{\bar{P}(x) \cdot \bar{F}_0(x)}.$$

By Proposition 3.1 we can choose a $C_2 > K$ such that

$$\mathcal{D}_\alpha\left\{P : \bar{P}(x) \geq \frac{\bar{Q}(x) \cdot x \cdot f_0(x)}{\bar{F}_0(x)} \text{ and } \frac{h_Q(x)}{h_P(x)} \leq 1, \ \ \forall\, x \geq a\right\} > 1 - \epsilon/2, \ \ \forall\, a > C_2.$$

Let $K_1 = max(C_1, C_2)$. Choosing K_1 to be the maximum of C_1 and C_2 we have the result.

□

Theorem 3.2 *Let Q be such that it satisfies (A3). Then, for any chosen $\epsilon > 0$, we have*

$$\mathcal{D}_\alpha\{P : I_0(h_P, h_Q) \leq \epsilon\} > 0.$$

Informally, the above says that h_Q belongs to the I_0 divergence support of the law of h_P when $P \sim \mathcal{D}_\alpha$.

PROOF. Choose and fix an $\epsilon > 0$. Let $\epsilon' > 0$ be such that $-\log(1-\epsilon') \leq \epsilon/4$. Now by Lemma 3.7 we have an $a > B(Q, \epsilon/2)$ such that

$$\mathcal{D}_\alpha \left\{ P : \bar{P}(x) \geq \frac{\bar{Q}(x) \cdot x \cdot f_0(x)}{\bar{F}_0(x)}, \ \forall\, x \geq a;\ \bar{P}(a) < \epsilon';\ W_Q^a(P) = 0 \right\} > 0.$$

Denote the set above by $A_{Q,a,\epsilon'}$. It is clear that it is enough to show that

$$\mathcal{D}_\alpha \{ P : T_Q^a(P) < \epsilon | A_{Q,a,\epsilon'} \} > 0, \tag{5}$$

as

$$0 \leq I_0(h_P, h_Q) \leq T_Q^a(P) + W_Q^a(P).$$

Note that any P can be decomposed into $(P_a^*, P_{a^c}^*, \bar{P}(a))$ for $a \in \mathbb{R}^+$. From the definition of $W_Q^a(P)$ it follows that it depends only on $(P_{a^c}^*, \bar{P}(a))$. Hence it is clear that $A_{Q,a,\epsilon'}$ depends only on $(P_{a^c}^*, \bar{P}(a))$. Now by using the Projection Lemma, i.e. Fact 3.1, and Lemma 3.6 it is clear that (5) holds. Hence the result. □

Below we shall state a fundamental result upon which our proof for posterior consistency depends. The result is implicit in Schwartz (1965). The statement we give along with the proof, can be found in Ghosh & Ramamoorthi (1994).

Proposition 3.2 (Schwartz) *Let U_i's be i.i.d. random variables with common distribution P. Let P belong to $\mathcal{P}$, where $\mathcal{P}$ is a family of probability measures dominated by a σ-finite measure, and let P_0 be the true distribution. Supposing the prior π puts positive mass on every Kullback-Leibler ball B_δ around P_0, namely $\pi(\{P : I_0(P, P_0) < \delta\}) > 0$, then the posterior is consistent at P_0.*

Lemma 3.8 *Let $\mathcal{P}_H$ be the set of probability measures defined by*

$$\{H_P : P \text{ a probability measure}\}.$$

Let $\mathcal{P}_H$ be endowed with the topology induced by the weak topology on the space of probability measures. Let T be the map

$$T:\ \mathcal{P}_H \to \mathcal{P}$$
$$H_P \to P$$

with the space of probability measures being endowed with the weak topology. Then T is a homeomorphism.

PROOF. For all the facts about unimodal distributions used below see Dharmadhikari and Joag-Dev (1988). First we shall show that P is identifiable. So let H_P and H_Q be the same law. By (A1) this implies that U_P and U_Q coincide. Now by the uniqueness of the Khintchine measure the identifiability follows. This guarantees that T is well defined. T is one to one. That T is open follows from an elementary argument and application of Scheffe's theorem. Continuity of T follows from the fact that the set of all unimodal distributions is weakly closed, (A1) and identifiability of the Khintchine measure. This completes the proof. □

Theorem 3.3 *Let all the four assumptions mentioned above hold. Then the posterior for P is consistent when the Z's are sampled from H_Q.*

PROOF. In view of proposition 3.2 it is clear that we only have to show the joint measurability of $h_P(x)$ w.r.t. P and x. Note that for each fixed P the function $h_P(\cdot)$ is right continuous. For each fixed $x > 0$ the function $h_P(x)$ is continuous in P. Hence we have the required measurability. So we have from the above that the posterior asymptotically concentrates on any set of P's such that the corresponding set of H_P's forms a weak ball around H_Q. But from Lemma 3.8 it is clear that this is same for any weak ball around Q. Hence the proof. □

Corollary 3.1 *Let all the four assumptions mentioned above hold. Then the estimator in Theorem 3.1 is consistent for U_Q when the Z's are sampled from H_Q.*

PROOF. Let us define the map S_u from $\mathcal{P}$ to $[0,1]$ by

$$\begin{aligned} S_u :\ & \mathcal{P} \to [0,1] \\ & P \to U_P(u). \end{aligned}$$

It is clear that this map is continuous with $\mathcal{P}$, as usual, endowed with the weak topology for every $u > 0$. That it is bounded, for every $u > 0$, is clear. Since, by Theorem 3.3, the posterior is weakly consistent and Q assigns no mass to 0 we have by the definition of weak topology that the posterior expectation of S_u converges to $U_Q(u)$ for all $u \in \mathbb{R}^+$. Hence the result. □

We note that for a unimodal distribution, U_P, we have

$$\bar{P}(x) = \bar{U}_P(x) + x \cdot u_P(x).$$

Hence if one can get an upper bound on the rate of decay of a set of allowed absolutely continuous unimodal distributions for G, then this could be used to choose a prior parameter α_0 such that consistency is guaranteed for any absolutely continuous unimodal G satisfying this upper bound in its tails. We say *could* because to guarantee the above we need this upper bound to be decaying faster than

$$\frac{\bar{F}_0(x)}{x \cdot f_0(x)}.$$

This factor would usually be polynomial as even for normal, which has a rapidly decaying tail, the factor is of the order x^{-2}. We choose the parameter α_0 such that (A3) is satisfied for all Q we deem possible. Of course we also need that h_Q is bounded, which basically says that u_Q is finite at 0 which is not a very restrictive assumption. An extension of the same kind of argument would help relax this condition. So the results of this section, unlike Doob (1949), guarantee consistency for a certain specific class of parametric values.

An advantage of being able to prove consistency in the above way is that the property of consistency is preserved when we take mixtures of one such prior with any arbitrary priors. This is because the *Kullback-Leibler support* of the resulting prior would be the union of the supports of the constituents of the mixture. This is in contrast to proving consistency via the tail-free property. Diaconis and Freedman (1983) gives many examples pointing out the above.

4 Computation of the Estimator

4.1 The Scheme

In this section we shall discuss the computations of the Bayesian estimator when we are working with a Dirichlet prior on the Khinchine measure of a unimodal distribution. Note that we are talking about evaluations of ratio of integrals of the type

$$\frac{\int I_{\{\eta_i \leq u\}}[\prod_{j=1}^{n} h_{\eta_j}(Z_j)] \prod_{j=1}^{n} [\alpha(d\eta_j) + \sum_{k=1}^{j-1} \delta_{\eta_k}(d\eta_j)]}{\int [\prod_{j=1}^{n} h_{\eta_j}(Z_j)] \prod_{j=1}^{n} [\alpha(d\eta_j) + \sum_{k=1}^{j-1} \delta_{\eta_k}(d\eta_j)]}.$$

Also note that the function h_η is given by

$$h_\eta(y) = \begin{cases} 0 & \eta < y \\ f_0(y) + \frac{\bar{F}_0(y) - y f_0(y)}{\eta} & \eta \geq y. \end{cases}$$

Computation of this type of estimator was discussed in Lo (1984), Kuo (1986) and Escobar (1994). We shall use an importance sampling scheme very similar to that given in Escobar (1994). The difference in the scheme used there and here is that we are forced to use a *different* importance sampling measure because of non-integrability of $h_\eta(y)$ with respect to η (Lebesgue measure).

Our scheme, following Escobar (1994), is as follows: generate

$$\left.\begin{array}{ll} \eta_i & \sim \phi_{z_i} \\ U_i & = 1 \end{array}\right\} \text{ with probability} \frac{\|\alpha\|}{\|\alpha\| + \sum_{1}^{i-1} h_{\eta_j}(z_i)}$$

and, for $j = 1, \ldots, i-1$,

$$\left.\begin{array}{ll} \eta_i & = \eta_j \\ U_i & = 0 \end{array}\right\} \text{ with probability} \frac{h_{\eta_j}(z_i)}{\|\alpha\| + \sum_{1}^{i-1} h_{\eta_l}(z_i)}.$$

Here, the ϕ_{z_i} are densities with respect to the Lebesgue measure of the importance sampling measures. Now to compute the above integral, we shall generate M i.i.d. vectors $(\underset{\sim}{U}, \underset{\sim}{\eta})$ and evaluate the sum

$$\frac{\sum_1^M I_{\eta_i^m \le u} \prod_1^n \{U_i^m \cdot A_i(\eta_i^m) + (1-U_i^m)\}\{\|\alpha\| + \sum_1^{i-1} h_{\eta_j^m}(z_i)\}}{\sum_1^M \prod_1^n \{U_i^m \cdot A_i(\eta_i^m) + (1-U_i^m)\}\{\|\alpha\| + \sum_1^{i-1} h_{\eta_j^m}(z_i)\}},$$

where

$$A_i(\eta) = \frac{h_\eta(Z_i) \cdot \alpha_0(\eta)}{\phi_{Z_i}(\eta)}.$$

Now our goal is to evaluate $\hat{G}_{U,n}(u)$. Below, we give a Monte Carlo estimate of $\hat{G}_{U,n}(u)$,

$$\begin{aligned} \mathrm{MCE}_M(\hat{G}_{U,n}(u)) &= \\ &\frac{\alpha(u)}{M+n} + \frac{n}{M+n} \cdot \left\{ \frac{\sum_1^M \#\{k : \eta_k^m \le u\} \prod_1^n \{U_i^m \cdot A_i(\eta_i^m) + (1-U_i^m)\}\{\|\alpha\| + \sum_1^{i-1} h_{\eta_j^m}(z_i)\}}{\sum_1^M \prod_1^n \{U_i^m \cdot A_i(\eta_i^m) + (1-U_i^m)\}\{\|\alpha\| + \sum_1^{i-1} h_{\eta_j^m}(z_i)\}} \right\}. \end{aligned}$$

Theorem 4.1 *As $M \to \infty$, we have $MCE_M(\hat{G}_{U,n}(u)) \to \hat{G}_{U,n}(u)$, a.s..*

PROOF. Proof is similar to that of Theorem 3 in Escobar (1994). □

The advantage of the above scheme is that it samples $\underset{\sim}{\eta}$ near $\underset{\sim}{Z}$ where the integral is concentrated. Moreover, note that because of the above form of $h_\eta(y)$ it is more efficient to use the scheme after ordering $\underset{\sim}{Z}$ in decreasing order. Of course, the densities, ϕ_{z_i} , $i = 1, 2, \ldots, n.$, must have support on $[z_i, \infty)$. All the usual requirements of an important sampling density apply to the ϕ_{z_i} , $i = 1, 2, \ldots, n$, densities. That is these must be easy to sample from and their tails must be heavier than and, in general a good approximation to the actual sampling density.

The above points will be illustrated more clearly in the example we discuss in the next subsection.

4.2 An Example

We consider an example where F_0 is chosen to be exponential with unit mean. The true G is chosen to be folded normal with scale parameter equal to 4. We chose α_0 to be Gamma with shape parameter equal to 2. This makes the prior mean of G to be exponential with unit mean. The true survival function $\bar{H}$ is shown in the following figure. Also plotted is the prior mean of $\bar{H}$. Note that we have chosen a prior which is markedly different from the true. This was done deliberately. We sampled, using IMSL pseudo-random number generation routines, a random sample of size 100 from the above given H. We then used the estimator derived using the Dirichlet for estimation of $\bar{H}$. To compute the estimate, we used the above importance sampling procedure with $\phi_z(\eta)$ defined as

$$\phi_z(\eta) = \begin{cases} \frac{I_{\{z<\eta\}} \cdot p(\eta)}{\exp(-z)+\exp(-1)} & 0 < z \leq 1 \\ \frac{I_{\{z<\eta\}} \cdot p(\eta)}{\exp(-z)\cdot(1+z)} & 1 < z \end{cases}$$

where, $p(\cdot)$ is given by,

$$p(\eta) = \begin{cases} \exp(-\eta) & 0 \leq \eta \leq 1 \\ \eta \cdot \exp(-\eta) & 1 < \eta. \end{cases}$$

We have chosen the above form for the importance sampling density in order that its tail is similar to that of $h_z(\eta) \cdot \alpha_0(\eta)$, as is its behavior near zero. Moreover, we see that the $A_z(\eta)$ are bounded for our choice of $\phi_z(\eta)$. This was done to satisfy the recommendations of Gweke (1989). Sampling from $\phi_z(\eta)$ is easy as it is either a truncated Exponential or truncated Gamma. In the case of a truncated Exponential, we generated pseudo-random numbers using the inverse distribution function and a uniform random variable generator. In the case of a truncated Gamma, we used acceptance-rejection sampling and a Gamma variable generator from IMSL.

One important step in the computation was the ordering of the data in decreasing order. Since it is a one time procedure, it does not add much to the running time. This increases the efficiency of the computation many times. The reason behind this is that, because of the marginal of $\underset{\sim}{\eta}$ being exchangeable, we can reorder and the reordering helps because of the function $h_\eta(z)$ being zero for $\eta < z$ and monotonically decreasing. This is also permissible because of the fact that the the product,

$$\prod_{j=1}^{n} h_{\eta_j}(Z_j),$$

appears in both the numerator and the denominator. In effect, because of the $\sum_{k=1}^{j-1} \delta_{\eta_k}(d\eta_j)$ term, by ordering we are eliminating sampling $\underset{\sim}{\eta}$ from the region which neither contribute to the numerator nor the denominator.

The program code, to implement the above importance sampling scheme, was written in FORTRAN programming language. The running time was just a couple of minutes for $M = 10000$ and estimates for the distribution function at all points in the range $[0, 10]$ in an interval of period 0.02.

$\|\alpha\|$ was chosen to be 1.25. Infact, we tried varying $\|\alpha\|$ over a wide range. But for exposition purpose we choose the value 1.25. It is clear that as $\|\alpha\|$ is increased the prior concentrates on the prior mean and so the estimate would tend to it. This was also observed in simulation. For the chosen value, see the following figure, the Bayesian estimator for $\bar{H}$ is smoother and very closely matches the true. Also in the following figure is the plot of the estimate, true and prior mean for the Khinchine measure associated with G. It is comforting to see movement of the estimate away from the prior mean towards the true.

For comparison purposes we have chosen the estimator of H given in Rojo and Samaniego (1994). As mentioned before, they consider estimation in the larger class of distributions uniformly stochastically smaller than F_0. But since their estimator is of the form that we consider in our problem, our choice of their estimator for purpose of comparison is reasonable. Their estimate for H, $\hat{H}_{RS}$, is given by

$$\hat{\bar{H}}_{RS}(x) = \bar{F}_0(x) \cdot \inf_{\{y \leq x\}} \left\{ \frac{\bar{H}_n(y)}{\bar{F}_0(y)} \right\}.$$

In the above definition we assumed that F_0 has the whole real line as its support. Else, define the estimate to be zero outside the support of F_0. This estimator for $\bar{H}$ beats the empirical distribution function in the expected Kolmogorov distance. In the third plot we have plotted this estimator along with the true survival function. For the particular choice of $\|\alpha\|$, the Bayesian estimator seems to do better than the Rojo-Samaniego estimator for $\bar{H}$, and is also much smoother.

Finally we could use finite mixtures of the Dirichlet to estimate G with almost no change in the computation procedure. See Escobar (1994) for more details.

5 Concluding Remarks

In this paper we have studied the problem from survival analysis of estimating a survival function in a sub-class of distributions uniformly stochastically smaller than a given distribution. In the case when we define the prior by using the Dirichlet prior for generating the Khinchine measure of a unimodal distribution, we prove that the consistency is attained for a certain class of unimodal distributions which can be defined by a bound on the rate of decay of the tails. In the literature there exists no results on consistency of estimators derived using Dirichlet in such a non-naive fashion. Is this the best one can do? Actually, perhaps no. It should be clear that the condition on the Kullback-Leibler *support* is not necessary. For example, if the Dirichlet is the prior on the sampling distribution function, consistency is achieved for all parametric values, but the Kullback-Leibler support is empty. Moreover, one can prove consistency by trying the following. There are two steps in proving consistency of the posterior. The first is to prove that the integral of the likelihood outside a neighborhood of the true vanishes exponentially fast. Since we are working with weak neighborhoods, that is fine here. The other step is to show that the integral of the likelihood in a neighborhood of the true does not decrease exponentially fast. This is where we used the Kullback-Leibler support condition. We could have tried a different route in this step. Indeed if π is the prior, then we need to find lower bounds on $\pi\{V \cap W_\eta\}$, where W_η is a total variation neighborhood of the true and V is a weak neighborhood of the true. The rate of decay of this probability as η approaches zero could help in proving consistency using Schwartz (1965). Such an approach shall be pursued in the future.

The computation of the estimator was reasonably fast and the code is quite small and simple to write. The estimator compares favorably with the estimator of Rojo-Samaniego. Of course, they do not assume unimodality. The example showed that it performs well even if the prior is wrongly centered. And of course it has the advantage of allowing incorporation of prior beliefs in an easy fashion.

Acknowledgement

I express my sincere thanks to Prof. Jayanta K. Ghosh for the course on *Bayesian Asymptotics and Nonparametrics* of Fall '94 at Purdue which rekindled my interest in the area and improved my understanding of the subject.

References

Berk, R. H. (1966). Limiting behaviour of posterior distributions when the model is incorrect. *Ann. Math. Statist.* **37**, 51-58.

Barron, A. R. (1986). Discussion on "On the consistency of Bayes estimates". *Ann. Statist.* **14**, 26-30.

Csiszár, I. (1967). Information type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica.* **2**, 299-318.

Dharmadhikari, S. and Joag-Dev K. (1988). *Unimodality, convexity and applications.* Academic press, San Diego, California.

Diaconis, P. and Freedman, D. (1983). On Inconsistent Bayes Estimates in the Discrete Case. *Ann. Statist.* **11** 1109-1118.

Doob,J. (1949). Application of the theory of Martingales. *Coll. Int. du CNRS.* Paris,22-28.

Doss, H. and Sellke, T. (1982). The tails of probabilities chosen from a Dirichlet prior. *Ann. Statist.* **10**, 1302-05.

Escobar, M.D. (1994). Estimating Normal Means With a Dirichlet Process Prior. *J. Amer. Statist. Assoc.* **89**, 268-277.

Feller, W. (1965). *An Introduction To Probability Theory And Its Applications Vol. II,* Wiley Eastern Limited, New Delhi.

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

Freedman, D. (1963). On the asymptotic behavior of Bayes estimates in the discrete case. *Ann. Math. Statist.* **34**, 1386-1403.

Ghosh, J. K. and Ramamoorthi, R. V. (1994). Consistency of Bayesian inference for survival analysis with or without censoring. *Preprint.*

Gweke, J. (1989). Acceleration methods for Monte Carlo integration in Bayesian Inference. *Econometrica* **57**, 1317-1339.

Kemperman, J.H. (1967). On the optimum rate of transmitting information. In *Prob. and Inf. Th.*. 126-129. Lecture Notes in Mathematics. Springer-Verlag: Berlin.

Kullback, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory.* **13**, 126-127.

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**, 351-57.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

Rojo, J. and Samaniego, F.J. (1993). On estimating a survival curve subject to uniform stochastic ordering constraint. *J. Amer. Statist. Assoc.* **88**, 566-572.

Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete.* **4**, 10-26.

Shyamalkumar, N. D. (1996). ESTIMATION OF A SURVIVAL FUNCTION IN A SUB-CLASS OF A UNIFORMLY STOCHASTICALLY DOMINATED FAMILY. Technical Report No. 96-20, Purdue University.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors.

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. Gupta and Berger, Academic Press, New York, 305-315.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907

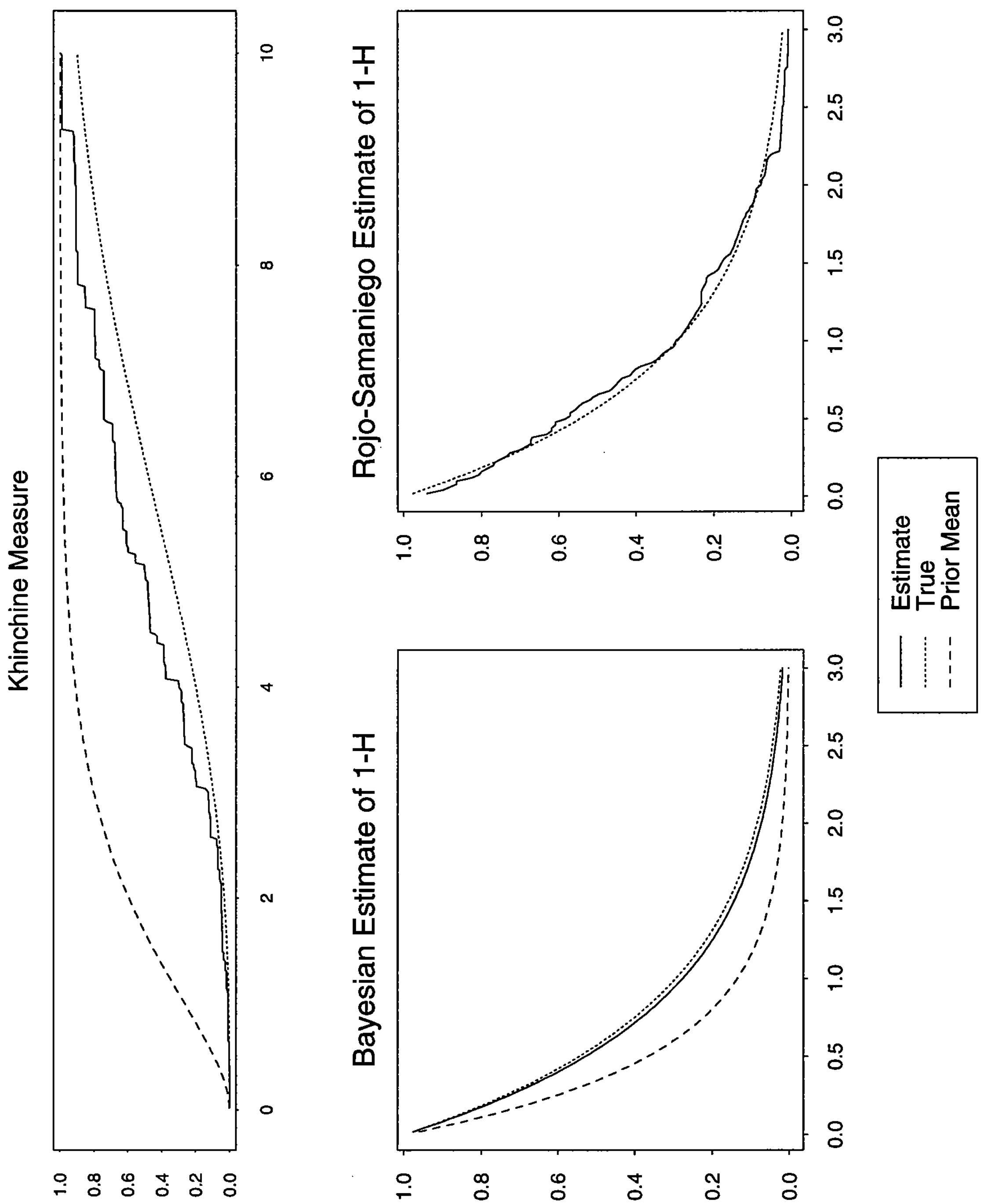

Figure 1: Estimates from the Dirichlet Process Prior