ESTIMATION OF A SURVIVAL FUNCTION IN A SUB-CLASS OF A UNIFORMLY STOCHASTICALLY DOMINATED FAMILY

by

N. D. Shyamalkumar
Purdue University

Technical Report #96-20

Department of Statistics
Purdue University
West Lafayette, IN USA

June 1996

ESTIMATION OF A SURVIVAL FUNCTION IN A SUB-CLASS OF A UNIFORMLY STOCHASTICALLY DOMINATED FAMILY

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 estimation of G in this paper. The GMLE for G is known to be inconsistent when the sampling distribution is continuous. The same is shown to hold true for the Bayes estimator of G when one assigns it a Dirichlet prior. Even though there are a lot of examples where using the Dirichlet prior results in an inconsistent estimator, the heuristic in this case is lucid and the inconsistency stark. The latter inconsistency arises because of the discreteness of the evaluations from a Dirichlet prior. In the discrete situation consistency is proved for both the GMLE and the Bayesian estimator using the Dirichlet prior.

Key words and phrases: Survival analysis, 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.

We mainly consider two different approaches to our problem, namely, the Maximum Likelihood and the Bayesian approaches to inference. Both these approaches have been used to tackle a wide variety of statistical inference problems in the literature. They usually give rise to estimators which a satisfy a variety of different *reasonability* criteria, one of the least stringent being consistency. Below we shall be dealing with two different kinds of consistency, namely, Classical and Bayesian consistency. In this paper we shall study the above *infinite-dimensional* estimation problem and investigate the consistency property of the estimators derived. We shall see, as has been often observed, that the *infinite-dimensional* setting is a very different scenario than to the *finite-dimensional* one.

One type of estimation procedure we do not talk about in this paper is the method of minimum distance estimation.What one does here is to find an appropriate distance on the space of probability measures and find the distribution in the class under consideration which is closest to the empirical distribution function. So for example one could use the Kolmogorov distance to start with and find the estimator. Even though the estimator cannot be written down explicitly, very fast algorithms do exist to find approximate minimum distance estimators. It is easy to see that the distance of the estimator from the truth is bounded above by twice the distance of the empirical distribution function from the truth. But in Rojo and Samaniego (1993), an estimator for H is found such that the distance (Kolmogorov metric) from the truth is less than the distance of the empirical from the truth. Their estimator can be written in a simple *closed* form. The interesting part is that, even though they consider H to be in the larger class of distributions which are uniformly stochastically smaller than F_0, their estimator is of the form we are considering in this paper. Hence the minimum distance estimator for H using the Kolmogorov metric is of no interest. Other distances could be tried, but we shall not consider this in this paper.

The maximum likelihood approach to our estimation problem is dealt with in the second section. In this section we discuss the consistency of the estimator derived in two cases, namely, the unrestricted case under both the continuous and discrete assumptions.

In the third section we give some alternate definitions of the Dirichlet process prior and some basic facts about it. We also define Bayesian consistency, or rather, posterior consistency.

In the next two sections we derive Bayesian estimators which, even though they can be blamed for being ad hoc, due to the arbitrariness in the choice of the prior, usually prove to be attractive estimators. Since the problem is intrinsically non-parametric, by which we mean there is no *natural* finite dimensional parametrization, it would be very non-robust to restrict oneself to a finite-dimensional parametric class for the sake of inference. The Dirichlet being a prior with many attractive properties, was chosen as a prior on the unknown distribution G. In the fourth section we use such a prior and discuss the discrete situation. In the fifth section we derive the estimator for G that results from using such a prior and prove that it is inconsistent for continuous distributions. It so happens that the heuristic for this inconsistency are lucid and the proof is elementary, in contrast to other inconsistencies observed using the Dirichlet process.

2 Maximum Likelihood Estimator

Let us recall that we are trying to find an estimate for the parameter G which can possibly be any distribution on the positive half of the real line. This, in particular, implies that the class of distribution functions H which arise by G assuming all possible distributions would be an undominated class. By an undominated class of distributions we mean that there does not exist a sigma-finite measure which dominates every member of the class. But to apply the usual principle of maximum likelihood estimation one requires a likelihood, which can only exist when the class of laws of the data considered is dominated. Hence we cannot apply the usual maximum likelihood principle in our case. The definition of MLE that applies to our case, and in almost total generality, is the one given in Kiefer and Wolfowitz (1956). Their definition, as one expects, agrees with the usual one when the class is dominated. We shall call the estimator as derived by their definition as the GMLE, generalized maximum likelihood estimator. To define a GMLE let us assume that we observe a vector $\underset{\sim}{X}$ which has an unknown law P which belongs to a class of probability measures $\mathcal{P}$. Let the observed data vector be $\underset{\sim}{x}$. Define, for two probability measures P_1, P_2 in $\mathcal{P}$, $f(\underset{\sim}{x}, P_1, P_2)$, by

$$f(\underset{\sim}{x}, P_1, P_2) := \frac{dP_1}{d(P_1 + P_2)}(\underset{\sim}{x}),$$

the Radon-Nikodym derivative of P_1 with respect to $P_1 + P_2$. Then $\hat{P}$ is said to be a GMLE if for all P in $\mathcal{P}$

$$f(\underset{\sim}{x}, \hat{P}, P) \geq f(\underset{\sim}{x}, P, \hat{P}).$$

The Kaplan-Meier PL estimator is a GMLE for the relevant problem, see Kaplan and Meier (1958) and Johansen (1978). Also, it is easy to check that, when there are no restrictions on the distribution function in the i.i.d. sampling case, the empirical distribution function is the GMLE for the unknown distribution function.

Now we shall precisely state what we shall mean by classical consistency of an estimator for a distribution function. Let $X_i,\ i = 1, 2, \ldots$ be i.i.d. from a distribution function F. Let $\hat{F}_n$ be the estimator of F based on the first n observations. Let F_0 be the true law of X. Then the sequence of estimators $F_n,\ n = 1, 2, \ldots$ is said to be consistent for F at F_0 if

$$\hat{F}_n(t) \to F_0(t) \text{ a.e.}(F_0),$$

at all continuity points t of F_0. We shall simply say that a sequence of estimators is consistent for F if it is consistent at all F_0.

In the following for a distribution function G, right continuous, and real x we shall denote by $G\{x\}$ the probability of the point x, by $\bar{G}(x)$ we shall mean $1 - G(x)$ and by $\bar{G}^+(x)$ we shall mean $\bar{G}(x) + G\{x\}$.

2.1 Unrestricted - Continuous Case

Lemma 2.1 *Assume that F_0 is continuous and that all the data points $Z_i,\ i = 1, 2, \ldots, n$ are distinct. Then the GMLE of $\hat{H}_{GMLE}$ is given by*

$$\hat{H}_{GMLE} = 1 - \bar{F}_0 \bar{H}_n,$$

where H_n denotes the empirical distribution based on $Z_i,\ i = 1, 2, \ldots, n$.

PROOF. The proof of the fact that we can restrict ourselves to G's which give mass only to points of the data is similar to that given in the derivation of the Kaplan-Meier estimator, see Johansen (1978) and Kaplan and Meier (1958). Now it is easy to see that, for any such G, the likelihood would be proportional to

$$L = \prod_{i=1}^{n} [P_G\{Z_i\} \cdot \bar{F}_0(Z_i)], \tag{1}$$

where by $P_G\{\cdot\}$ we mean the probability under G. Note that in the above we use the fact that F_0 is continuous in a crucial way. Now since we are maximizing L w.r.t. G, it is clear from (1) that F_0 plays no role and that the maximum is attained when G is the empirical distribution based on Z_i, $i = 1, 2, \ldots, n$, H_n, which is the GMLE for H with no restrictions. Hence $\hat{H}_{GMLE} = 1 - \bar{F}_0 \bar{H}_n$. $\square$

Corollary 2.1 *The above defined estimator $\hat{H}_{GMLE}$ is an inconsistent estimator of H.*

PROOF. Follows trivially from the Glivenko-Cantelli theorem or the SLLN. $\square$

Remark 2.1 As mentioned in the introduction, $\hat{H}_{GMLE}$ is also the GMLE of H when one considers the larger class of distributions which are uniformly stochastically smaller than F_0, see Rojo and Samaniego (1990). Of course, their result does imply the above lemma but the reason we have given the proof is because it is much simpler than it is in their case.

2.2 Unrestricted - Discrete Case

The following theorem completes the picture by discussing the discrete case and showing that consistency is obtained in this case unlike the former. In the next section we shall show that Bayes estimator with respect to any Dirichlet prior also behaves similarly.

Theorem 2.1 (Discrete case) *Let F_0 and G be distributions on $\{1, 2, \ldots\}$. F_0 is assumed to be known and G unknown. Let Z_i $i = 1, \ldots, n$ be i.i.d. from $H = 1 - \bar{F}_0 \bar{G}$. Then the M.L.E. of G is consistent.*

PROOF. Let $\{d_1, d_2, \ldots, d_k\}$ be the distinct observed values in ascending order and let $n_i := \#\{j : z_j = d_i\}, i = 1, 2, \ldots, k$. The log-likelihood $L(G)$ can be seen to be

$$L(G) = \sum_1^n \{n_i \cdot \log(\bar{G}^+(d_i) \cdot F_0\{d_i\} + \bar{F}_0(d_i) \cdot G\{d_i\})\}.$$

From the above, it is easy to see that the MLE of G, say $\hat{G}$, would be supported on $\{d_1, d_2, \ldots, d_k\}$. Let $\{\theta_0, \theta_1, \ldots, \theta_k\}$ be defined such that $\theta_0 := 0$ and

$$G\{d_i\} = \begin{cases} \theta_1 & i = 1 \\ \theta_i \cdot \prod_1^{i-1} (1 - \theta_j) & i = 2, 3, \ldots, k. \end{cases}$$

Similarily we shall define $\{\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_k\}$. Then the log-likelihood, $L(G)$, can be written as

$$L(G) = \sum_1^k \{ (\sum_{j \geq (i+1)} n_j) \cdot \log(1 - \theta_i) + n_i \cdot \log(F_0\{d_i\} + \theta_i F_0\{d_i\}) \}.$$

It is clear from the above that $\hat{\theta}_k = 1$ and the sum is maximized by maximizing each term separately. To this end note that

$$\frac{\partial}{\partial \theta_i}(L(G)) = \frac{n_i \cdot \bar{F}_0(d_i)}{F_0\{d_i\} + \theta_i \cdot \bar{F}_0(d_i)} - \frac{\sum_{j \geq (i+1)} n_j}{(1 - \theta_i)}, \; 0 < i < k$$

It is clear that the maximizing value $\hat{\theta}_i$ would be

$$\hat{\theta}_i = \left\{ \frac{n_i \cdot \bar{F}_0(d_i) - F_0\{d_i\} \cdot \sum_{i+1}^k n_j}{\bar{F}_0(d_i) \cdot \sum_i^k n_j} \right\}^+ , i = 1, 2, \ldots, k - 1.$$

Consistency of the above estimator follows from the SLLN. □

3 Pre-requisites in Bayesian Nonparametrics

3.1 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 $\mathcal{D}_\alpha$, where 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 natural 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 3.1 (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 3.2 (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 $M\alpha_0$.*

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.

3.2 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 $\theta \rightarrow P_\theta$ 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 for 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.

4 Bayesian Estimator : Unrestricted - Discrete Case

4.1 Introduction

Let us suppose that conditioned on G, $X_i, i = 1, 2, \ldots, n$ are i.i.d. with common distribution G and let G be assigned a Dirichlet process prior with parameter α which we shall denote by $\mathcal{D}_\alpha$. Note that α is a positive finite measure and we shall sometimes write it as $M\alpha_0$ where M is a positive non-zero constant and α_0 is a probability measure. Let $Y_i, i = 1, 2, \ldots, n$ be i.i.d. with common distribution F_0 and let $\underset{\sim}{Y}$ be independent of $(\underset{\sim}{X}, G)$. Let $Z_i := X_i \wedge Y_i$ for $i = 1, 2, \ldots, n$. Let $d_i := I_{\{X_i \leq Y_i\}}$ for $i = 1, 2, \ldots, n$. In this section we assume that both F_0 and G_0 are probabilities on $\{1, 2, ..\}$ with whole support.

Also, as usual, empty sums will be taken as 0, empty products as 1 and 0^0 as 1. We shall also make use of the following notations.

- $N^+(u) = \sum\limits_{i=1}^{n} I_{\{Z_i > u\}}$.

- $C(M,n) = 1/\prod_{i=0}^{n-1}(M+i)$.

Whenever we say Bayes estimator we shall mean with respect to a squared error type of loss. In this case the Bayes estimator of a probability measure would be the barycenter of the posterior. More specifically, the Bayes estimator of the survival function evaluated at a point shall be its posterior expectation. 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 4.1 *The joint marginal distribution of $X_i^{\prime}s$ 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. □

Remark 4.1 Note that the distribution of $\underset{\sim}{X}$ is exchangeable which follows from the fact that it is a mixture of i.i.d. probability measures.

Remark 4.2 Blackwell and MacQueen called the above a generalized Polya sequence with parameter α. The Dirichlet process prior has an interpretation through these (generalized) Polya sequences, see Blackwell and MacQueen (1973).

4.2 Consistency of the Estimator

Following Freedman (1963), we shall say that a probability μ on the space of probability measures on $\{1, 2, ..\}$ is tail free if

1. $\mu\{F : F(k) < 1\} = 1$ for all k.

2. There is a natural number N such that the random vector $\{F\{k\} : k \leq N\}$ and the random variables $[1 - F(N+k-1)]^{-1} F\{N+k\}, 1 \leq k < \infty$ are mutually independent.

Below we state a result regarding tail free priors from Freedman (1963).

Proposition 4.1 (Freedman) *Let μ be a tail free probability on the space of all probability measures on $\{1, 2, ..\}$. Then (μ, F) is consistent if and only if F is in the topological carrier of μ.*

Now we state and prove the main result of this section on the consistency of the Bayesian estimator in the discrete setting.

Theorem 4.1 *Let F_0 and G_0 be distribution functions on $\{1, 2, \ldots\}$. F_0 is assumed to be known and G_0 unknown. Let Z_i, $i = 1, \ldots, n$, be i.i.d. from $H_0 = 1 - \bar{F}_0\bar{G}_0$. We put a prior on G as $\mathcal{D}_\alpha$, such that $\alpha\{i\} > 0$, $\forall i \geq 1$ and $\|\alpha\| = \sum \alpha\{i\} < \infty$. Then*

1. *The prior on the unknown distribution of the Z's, $\bar{H} = \bar{F}_0\bar{G}$, is tail-free.*

2. *The Bayesian estimator for G is consistent.*

PROOF. Note that

$$G\{i\} = \theta_i \prod_0^{i-1}(1 - \theta_j), i = 1, 2, \ldots$$

and

$$\bar{G}^+(i) = \prod_0^{i-1}(1 - \theta_j), i = 1, 2, \ldots$$

where $\theta_0 := 0$ and θ_i, for $i = 1, 2, \ldots$, are independent random variables following $Beta(\alpha\{i\}, \sum_{i+1}^{\infty} \alpha\{j\})$. Also note that

$$H\{i\} = \bar{F}_0(i) \cdot G\{i\} + F_0\{i\} \cdot \bar{G}^+(i)$$

and

$$\bar{H}^+(i) = \bar{F}_0^+(i) \cdot \bar{G}^+(i).$$

Using the above representation we get

$$\frac{H\{i\}}{\bar{H}(i-1)} = \frac{F_0\{i\}}{\bar{F}_0(i-1)} + \theta_i \cdot \frac{\bar{F}_0(i)}{\bar{F}_0(i-1)}$$

and

$$H\{i\} = \prod_{0}^{i-1}(1-\theta_j)\cdot(\bar{F}_0(i)\cdot\theta_i + F_0\{i\}).$$

From the definition of tail-free priors and the above expression we see that the prior on H is tail-free. Hence, by proposition 4.1, the Bayes estimator for H is consistent. As H is linearly related to G and F_0 has whole support, it follows that the Bayes estimator for G is also consistent. □

5 Bayesian Estimator : Unrestricted - Continuous Case

5.1 The Estimator

The set up is as in the previous section except that, in the rest of the paper, we shall work with the following two assumptions unless specified otherwise.

- F_0 and α_0 are continuous probabilities with positive densities f_0 and α_0', respectively, w.r.t. Lebesgue measure on $\mathbb{R}^+$.
- $Z_i, i = 1, 2, \ldots, n$, is a sample from a survival function $\bar{F}_0\bar{G}_0$ where G_0 is continuous which, in particular, implies the distinctness of the data points almost surely.

Now let us consider the problem where apart from the vector $\underset{\sim}{Z}$ we also observe the vector $\underset{\sim}{d}$. This is a regular censoring problem. Susarla and Van Ryzin (1976) found the Bayes estimator using the Dirichlet process prior in the fashion mentioned above. Their estimator is given below.

Proposition 5.1 (Susarla & Van Ryzin) *The Bayes estimator of $\bar{G}$ when one has observed $(\underset{\sim}{Z}, \underset{\sim}{d})$, denoted by $\hat{\bar{G}}_{d,n}$, is given by*

$$\hat{\bar{G}}_{d,n}(u) = \frac{\bar{\alpha}(u)+N^+(u)}{\alpha(\mathbb{R}^+)+n}\prod_{i=1}^{n}\left[\frac{\bar{\alpha}(Z_i)+N^+(Z_i)+1}{\bar{\alpha}(Z_i)+N^+(Z_i)}\right]^{\{(1-d_i)I_{\{Z_i\le u\}}\}} \tag{2}$$

Now it is clear that to find the estimate for $\bar{G}$ on the basis of just having observed $\underset{\sim}{Z}$ we have to find the conditional expectation of the above estimator w.r.t. $\underset{\sim}{d}$ conditioned on $\underset{\sim}{Z}$. So for this purpose we find below the joint distribution of $\underset{\sim}{d}$ and $\underset{\sim}{Z}$.

Lemma 5.1 *The joint density of $\underset{\sim}{d}$ and $\underset{\sim}{Z}$ w.r.t. the product of the counting measure on $\{0,1\}^n$ and Lesbegue measure on $\mathbb{R}^n$ is given by,*

$$C(M,n)\prod_{i=1}^{n}[\bar{F}_0(Z_i)\alpha'(Z_i)]^{d_i}[(\bar{\alpha}(Z_i)+N^+(Z_i))f_0(Z_i)]^{1-d_i} \tag{3}$$

where α' is the density of α w.r.t. to Lebesgue measure.

PROOF. Since $\{(Z_i,d_i)\}_1^n$ is a finite exchangeable sequence let us assume without loss of generality that $(d_1,\ldots,d_k)=(1,\ldots,1)$ and $(d_{k+1},\ldots,d_n)=(0,\ldots,0)$ with obvious modification for extreme values of k. Now since $d_i=1 \rightsquigarrow Z_i=X_i$, we have the joint marginal density with respect to the natural dominating measure of $\{(Z_i,d_i)\}_1^k$ would be

$$\frac{\prod_{i=1}^{k}[\bar{F}_0(Z_i)\alpha'(Z_i)]}{\prod_{i=0}^{k-1}[M+i]}.$$

Now all that remains is to find the joint conditional density of $\{(Z_i,d_i)\}_{k+1}^n$ given $\{(Z_i,d_i)\}_1^k$. Since $(d_{k+1},\ldots,d_n)=(0,\ldots,0)$ we have the required conditional density to be

$$Prob(X_i\in(Z_i,\infty),\ i=k+1,\ldots,n \mid X_i=Z_i,\ i=1,\ldots,k)\prod_{i=k+1}^{n}[f_0(Z_i)]. \tag{4}$$

Now note that the first term is precisely the *joint survival function* of a generalized Polya sequence with parameter $\alpha+\sum_{i=1}^{k}\delta_{Z_i}$. Again by use of exchangeability we shall assume without loss of generality that $Z_{k+1}>\ldots>Z_n$. We shall denote the survival function of a generalized Polya sequence with parameter μ by $\bar{\mathcal{P}}_\mu$. With this notation the first term in (4) is

$$\bar{\mathcal{P}}_{\alpha+\sum_{i=1}^{k}\delta_{Z_i}}(Z_{k+1},\ldots,Z_n). \tag{5}$$

By distinctness of the $Z_i's$ we have that (5) is the same as

$$\frac{1}{M+k}\int_{Z_{k+1}}^{\infty}\bar{\mathcal{P}}_{\alpha+\delta_Z+\sum_{i=1}^{k}\delta_{Z_i}}(Z_{k+2},\ldots,Z_n)\,d\alpha(z)$$

$$+\frac{1}{M+k}\sum_{i=1}^{k}I_{Z_i>Z_{k+1}}\bar{\mathcal{P}}_{\alpha+\delta_{Z_i}+\sum_{j=1}^{k}\delta_{Z_j}}(Z_{k+2},\ldots,Z_n)$$

and by assumption that $Z_{k+1}>\ldots>Z_n$ we can write the above as

$$\left[\frac{\bar{\alpha}(Z_{k+1})+N^+(Z_{k+1})}{M+k}\right]\bar{\mathcal{P}}_{\alpha+\sum_{i=1}^{k+1}\delta_{Z_i}}(Z_{k+2},\ldots,Z_n).$$

Now proceeding similarly it is clear that the joint density we seek can be written as

$$C(M,n)\prod_{i=1}^{k}[\bar{F}_0(Z_i)\alpha'(Z_i)]\prod_{i=k+1}^{n}[(\bar{\alpha}(Z_i)+N^+(Z_i))f_0(Z_i)].$$

Note that the expression in (3) is same as above under the assumption of ordering of $\underset{\sim}{d}$ and $\underset{\sim}{Z}$ and does not depend on the ordering. Hence by exchangeability the proof. □

Remark 5.1 The posterior for G is a mixture of Dirichlet. This can be easily seen since the conditional law of G given $(\underset{\sim}{\delta},\underset{\sim}{Z})$ is a mixture of Dirichlet, see Blum and Susarla (1977) for a precise representation. For the representation here one has to just use their representation and the above lemma.

Theorem 5.1 *The Bayes estimator for $\bar{G}$ when one has observed only $\underset{\sim}{Z}$, denoted by $\hat{\bar{G}}_n$, is given by*

$$\hat{\bar{G}}_n(u)=\frac{\bar{\alpha}(u)+N^+(u)}{\alpha(\mathbf{R}^+)+n}\prod_{i=1}^{n}\left[1+\frac{f_0(Z_i)I_{\{Z_i\le u\}}}{\bar{F}_0(Z_i)\alpha'(Z_i)+(\bar{\alpha}(Z_i)+N^+(Z_i))f_0(Z_i)}\right]. \tag{6}$$

PROOF. Note first that (2) can be written as

$$\hat{\bar{G}}_{d,n}(u)=\frac{\bar{\alpha}(u)+N^+(u)}{\alpha(\mathbf{R}^+)+n}\prod_{i=1}^{n}\left[\frac{\bar{\alpha}(Z_i)+N^+(Z_i)+I_{\{Z_i\le u\}}}{\bar{\alpha}(Z_i)+N^+(Z_i)}\right]^{(1-d_i)}. \tag{7}$$

From (3) we get that the marginal density of $\underset{\sim}{Z}$ when the coordinates are distinct is given by

$$C(M,n)\prod_{i=1}^{n}[\bar{F}_0(Z_i)\alpha'(Z_i)+(\bar{\alpha}(Z_i)+N^+(Z_i))f_0(Z_i)].$$

By doing similar algebra as needed to derive the above we get the numerator for the required conditional expectation to be

$$C(M,n)\frac{\bar{\alpha}(u)+N^{+}(u)}{\alpha(\mathbf{R}^{+})+n}\prod_{i=1}^{n}[\bar{F}_0(Z_i)\alpha'(Z_i)+(\bar{\alpha}(Z_i)+N^{+}(Z_i)+I_{\{Z_i\le u\}})f_0(Z_i)].$$

The ratio of the last two expressions is precisely (6). Hence the result. □

Remark 5.2 Note that from Sethuraman and Tiwari (1982) we know that the limiting prior does not intuitively correspond to an *ignorance prior* but the posterior sometimes becomes free of the parameter α_0 and hence in some sense corresponds to a noninformative posterior. So it is interesting to find the limit of the above Bayesian estimator as we let M approach zero. It does happen many times that the Bayesian estimator using the Dirichlet process prior tends to the GMLE of the problem as we let M approach zero, see Pruitt (1990a). In Pruitt (1990b, 1992) an example of inconsistency is given in the situation of estimating a bivariate survival function where the *limiting* estimator derived in this fashion using a Dirichlet process depends on the Dirichlet parameter. In this situation the GMLE is not uniquely defined and consistency of the GMLE depends on the choice. This makes one wonder if there is a connection between non-uniqueness of the GMLE and dependence of the limiting estimator on the parameter. In this context it is worth noting that the situation of truncated data, see Gasparini (1995), is an example where the limiting estimator does depend on the parameter, but asymptotically has no effect and, in fact, coincides with the GMLE, and the GMLE is uniquely defined. In the above, consistency is achieved. In our problem we have similar behavior except for the fact that the estimator is very different from the GMLE and also is grossly inconsistent. To this end note that this limit is given by the survival function given below.

$$\lim_{\alpha(\mathbf{R}^{+})\to 0}\hat{\bar{G}}_n(u)=\begin{cases}1 & u<Z_{(n)}\\ \frac{\bar{\alpha}_0(u)\cdot f_0(Z_{(n)})}{\bar{F}_0(Z_{(n)})\cdot\alpha_0'(Z_{(n)})+\bar{\alpha}_0(Z_{(n)})\cdot f_0(Z_{(n)})} & u\ge Z_{(n)}\end{cases}$$

where $Z_{(n)}$ is the sample maximum. It is, of course, clear that with increasing sample size the above would tend to the point mass at infinity when the true G_0 has full support. Actually, as we shall see in the next sub-section, this is what happens even with the estimator derived with any amount of prior information.

5.2 Inconsistency

Below we prove that the estimator derived above is inconsistent.

Theorem 5.2 *Under the standing assumptions the Bayes estimator of $\bar{G}$, $\hat{\bar{G}}$, is inconsistent. More specifically, we have*

$$\hat{\bar{G}}_n(u) \to 1 \ a.s.(H_0).$$

PROOF. First note from (6) that the first term goes to $\bar{H}_0(u)$ by the SLLN. Choose and fix a $u \in \mathbb{R}^+$. Let the logarithm of the second term be denoted by $W_n(u)$. Using the fact that $\mid \log(1+x) - x \mid \leq x^2 \ \forall \ x > 0$ we have

$$W_n(u) = \sum_{i=1}^{n} \frac{f_0(Z_i) I_{\{Z_i \leq u\}}}{\bar{F}_0(Z_i)\alpha'(Z_i) + (\bar{\alpha}(Z_i) + N^+(Z_i)) f_0(Z_i)} + R_n(u), \tag{9}$$

where

$$\mid R_n(u) \mid \leq \frac{n}{[N^+(u)]^2} \to 0 \text{ a.s.}(H_0),$$

by the SLLN. Now the first term in (9) can be written as

$$\sum_{i=1}^{n} \frac{I_{\{Z_i \leq u\}}}{N^+(Z_i)} - \sum_{i=1}^{n} \frac{[\bar{F}_0(Z_i)\alpha'(Z_i) + \bar{\alpha}(Z_i) f_0(Z_i)] I_{\{Z_i \leq u\}}}{N^+(Z_i)[\bar{F}_0(Z_i)\alpha'(Z_i) + (\bar{\alpha}(Z_i) + N^+(Z_i)) f_0(Z_i)]}. \tag{10}$$

Now note that

$$\sum_{i=1}^{n} \frac{I_{\{Z_i \leq u\}}}{N^+(Z_i)} = \int_0^u \frac{dH_n(s)}{1 - H_n(s)}.$$

By assumption of continuity and full support of H_0 we have by the Glivenko-Cantelli theorem and Theorem 5.5 (due to H. Rubin) of Billingsley (1968) that the above converges to $-\log \bar{H}_0(u)$ a.s.(H_0). Now all that remains to be shown is that the second term in (10) converges to zero almost surely. To this end note that this term is less than

$$\sum_{i=1}^{n} \frac{I_{\{Z_i \leq u\}}}{N^+(u)} \left[\frac{2}{N^+(u)\left(\frac{f_0(Z_i)}{\alpha'(Z_i)} \wedge 1\right)} \wedge 1 \right].$$

The above can be seen to go to zero a.s.(H_0) by using the assumption that F_0 has as support the whole of $\mathbb{R}^+$, and an elementary argument. $\square$

5.3 Discussion and heuristics

In this subsection we try to explain what we think is a very lucid heuristic for the inconsistency observed in the previous subsection. Before doing so we shall review some of the literature which study the consistency properties of Bayes Estimators and explain why we claim the above to be interesting. See Diaconis and Freedman(1986a) for reasons as to why even a Bayesian should be interested in consistency of his estimators.

We shall start with the old and well known result of Doob (1949) which states under mild conditions (like existence of consistent estimators) that for almost every value of the parameter with respect to the prior one has consistency of the Bayes estimators. The crucial thing to observe is that consistency need not hold for *all* values of the parameter. Moreover, since sets with probability one vary with prior and that such sets can be topologically very bad (nowhere dense) it is desirable to check whether consistency holds for every value of the parameter. So in the following whenever we say that the Bayes estimator is consistent we shall mean it is consistent for all values of the parameter. Even in the case when sampling from a countably infinite population with unknown distribution there does exist priors assigning positive probability to every open set such that the posterior is consistent for only a set of parameters of the first category, see Freedman (1963). Moreover in Freedman (1965) it is shown that when sampling from countably infinite population with unknown distribution, for all but a set of priors of the first category, Bayes estimates are consistent only at a set of distributions of the first category. Even so it is *fair* to say that Bayes estimators are *usually consistent* in finite dimensional problems, see Berliner and MacEachern (1993) for an exception. In infinite dimensional examples it has been known for sometime now that Bayes estimators can be inconsistent and sometimes also behave in a rather erratic fashion see for eg. Diaconis and Freedman (1986a) where it is shown that the Bayes estimate for a location parameter oscillates between two wrong values. We should mention here that we are loosely classifying the dimensionality of the problems, see Hartigan (1986).

The Dirichlet process is a probability measure on the space of all probability measures such that its weak support is the whole space of probability measures when the parameter has support as the whole space, see Ferguson (1973). Even though the former is true, the Dirichlet process assigns probability one to the set of all discrete measures, which of course is dense in any seperable metric space. There are different proofs of the discreteness of the Dirichlet, for two of them see Blackwell and MacQueen (1973) and Sethuraman and Tiwari (1982). So in view of the above and our assumption that the underlying G_0 is continuous, the result just observed is precisely a form of the phenomenon which is not guarded against by the Doob (1949).

Now we mention what we believe is the reason behind the inconsistency just observed; we have not been able to carry out a proof using this heuristic as it seems to be very messy. Let us consider the situation where one is sampling the $Z's$ from a survival function $\bar{F}_0 \cdot \bar{G}$ where F_0 is as above and G is discrete. Now let us suppose that we have $\bar{G}(u) < 1$. Then, if we observe a large number of observations we are bound to have quite a few observations which are not distinct. Reversing this argument, if we have a large number of observations and all of them are distinct then we would think that G assigns almost all its mass far away. Of course, we agree that there is quite a bit of hand-waving here, but after all this is only an heuristic and the phenomenon that it addresses has been proven in the last section.

6 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. It has been shown that the behavior of the GMLE and the Bayesian estimator derived using the Dirichlet prior in a naive fashion, in the continuous case, is unreasonable, in the sense that they converge to wrong values. We observe that the proof of inconsistency is simple and there is a nice heuristic for the phenomenon. In the discrete case, both the

MLE and Bayesian estimator, using the Dirichlet naively, are consistent.

In Shyamalkumar (1996), addressing the same problem, using a prior which puts all its mass on the set of absolutely continuous unimodal distributions, we construct Bayesian estimators for which consistent is exhibited for a large set of G. The prior is based on the Dirichlet prior; we use the distribution sampled from a Dirichlet as the Khinchine measure of the unimodal distribution, following Lo (1984). Hence the caveat is that it is the Kullback-Leibler *support* which is pertinent to the consistency of Bayesian estimators and not the weak support.

Acknowledgement

Prof. Dongchu Sun pointed out the inconsistency of the GMLE which led to this paper. I express my sincere thanks to him.

References

Blackwell, D. and MacQueen, J.B. (1973). Ferguson distributions via Polya urn schemes. *Ann. Statist.* **1**, 356-358.

Berliner, L.M. and MacEachern, S.N. (1993). Examples of inconsistent Bayes procedures based on observations on dynamical systems. *Statist. Probab. Lett.* **17**, 355-360.

Blum J. and Susarla V. (1977). On the posterior distribution of a Dirichlet process given randomly right censored observations. *Stoch. Proc. Appl.* **5**, 207-11.

Billingsley, P. (1968). *Convergence of Probability Measure.* John Wiley & Sons, Inc., New York.

Diaconis, P. and Freedman, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1-26.

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

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.

Freedman, D. (1965). On the asymptotic behavior of Bayes estimates in the discrete case II. *Ann. Math. Statist.* **36**, 454-56.

Gasparini, M. (1995). Nonparametric Bayes estimation of a distribution function with truncated data. To appear in *J. Statist. Plann. Inference* .

Hartigan, J.A. (1986). In discussion of "On the consistency of Bayes estimates". *Ann. Statist.* **14**, 48.

Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* **5**, 195-99.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations.*J. Amer. Statist. Assoc.* **53**, 457-481.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27**, 887-906.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density Estimates. *Ann. Statist.* **12**, 351-57.

Pruitt, R. C. (1990a). Some issues in Bayesian nonparametrics. *Proc. on the interface of Computer Science and Statistics.* **22**, 243-248.

Pruitt, R. C. (1990b). Bayesian survival curve estimation for incompletely observed multivariate data. *Preprint.* University of Minnesota.

Pruitt, R. C. (1992). An inconsistent Bayes estimator in bivariate survival curve analysis. *Statist. Probab. Lett.* **15**, 177-180.

Rojo, J. and Samaniego, F.J. (1990). On nonparametric maximum likelihood estimation of distribution uniformly stochastically smaller than a standard. *Statist. Probab. Lett.* **11**, 267-271.

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.

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.

Shyamalkumar, N. D. (1996). PARTIAL POSTERIOR CONSISTENCY USING DIRICHLET PRIORS IN ESTIMATION OF SURVIVAL FUNCTION. Technical Report No. 96-21, Purdue University.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc.* **71**, 897-902.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907