Analysis of Censored Data
IMS Lecture Notes – Monograph Series (1995) Volume 27

Consistency of Bayesian inference for survival analysis with or without censoring

Jayanta K. Ghosh*
Indian Statistical Institute and Purdue University
and
R.V. Ramamoorthi†
Michigan State University

Abstract

We study the convergence of the posterior distribution to the true distribution in the context of survival analysis data. In the presence of censoring, when the prior is a Dirichlet process, we establish the consistency when the true distribution satisfies a bounded support assumption. We provide a sufficient condition for consistency for general priors. In the uncensored case we prove a similar result when the prior for the survival distribution arises through a Dirichlet Process prior for the hazard rate.

1 Introduction

Dirichlet process priors D_α for non-parametric Bayesian inference about an unknown distribution function F were introduced by Ferguson(1973,1974). For a recent review see Ferguson, Phadia and Tiwari(1993). Ferguson (1973) proved the consistency of the Bayes estimate for F. The approach of Sethuraman and Tiwari (1982) can be used to prove the consistency of the posterior, in the sense of Freedman, i.e., the posterior converges to δ_{F_0} a.s F_0, where F_0 is the true distribution and δ_{F_0} is the probability measure putting all its mass at F_0. A careful and detailed presentation of this notion of consistency, which can be traced back to the work of Laplace, is available in Freedman(1963) and in Diaconis and Freedman (1986).

*Research supported by NSF grant No.9307727
†Research supported by NIH grant no.1 RO1 GM49374

Dirichlet process priors were introduced in survival analysis by Susarla and Van Ryzin (1976). They found an explicit form for the Bayes estimate of F in the presence of random,right censoring, and showed that in their theory the Kaplan-Meier estimate of F appears as a non-informative Bayes estimate. They showed that the Kaplan-Meier estimate can be obtained by holding the data fixed and taking the limit of a sequence of Bayes estimates corresponding to D_{α_i}, where for all i, $\frac{\alpha_i(A)}{\alpha_i(R^+)} = \frac{\alpha(A)}{\alpha(R^+)}$ is a fixed non-zero probability measure and $\alpha_i(R^+)$ goes to 0 as $i \to \infty$. Susarla and Van Ryzin (1978) prove the consistency and other asymptotic properties of the Bayes estimate for a fixed prior D_α.

It is interesting that the posterior in Susarla and Van Ryzin's work has no simple form though a representation is implicit in their work and given explicitly in unpublished lecture notes of Sethuraman (1986). An interesting question is whether, in this set up, the posterior is consistent in the sense of Freedman. In section 2 we answer this question in the affirmative under an assumption of compact support for both F and the censoring distribution G.

The Bayes estimate of Susarla and Van Ryzin has discontinuities and so it cannot be used for estimating a hazard rate. To avoid this problem Dykstra and Laud (1981) have introduced a prior for the hazard rate itself, and given explicit representation for the posterior distribution of F, with or without censoring. However these expressions are so unwieldy that no consistency properties are known for their posterior or Bayes estimate. We prove a posterior consistency theorem in this setting when a multiple of the hazard rate has a Dirichlet process prior. In the last section we provide a sufficient condition for consistency.

2 Dirichlet Process prior for P_F, P_G

In the following (X_i, Y_i) 's are i.i.d. non negative random variables and X_i and Y_i are independent with distributions P, Q. We view P as the survival distribution and Q as the censoring distribution . For $1 \le i \le n$, let

$$\begin{aligned} Z_i &= min(X_i, Y_i) \quad \text{and} \\ \Delta_i &= 1 \quad \text{if} X_i < Y_i, \\ \Delta_i &= 0 \quad \text{if} X_i > Y_i. \end{aligned}$$

The assumptions made later ensure that $X_i = Y_i$ has zero probability.

In this section and in section 4 , we investigate consistency in the presence of censoring and thus only (Z_i, Δ_i) are observed and we consider the posterior of (P,Q) given $(Z_i,\Delta_i), I = 1,2,\cdots,n$. Given a prior for (P,Q) we say that the posterior for P is consistent at the true (P_0,Q_0) if the marginal posterior distribution of P given the data at stage n , converges weakly, as $n \to \infty$ to the degenerate measure δ_{P_0} almost surely (P_0,Q_0).

We need conditions on the distribution P of X and Q of Y . We formulate these as

C 1 *P and Q do not have any common point of discontinuity, in the sense that the associated distribution functions F_P, F_Q, have no points of discontinuity in common.*

C 2 *For all a , $P(X > a) > 0$ implies $Q(Y > a) > 0$.*

Fix a_1, a_2 such that $a_1 \le a_2$. Let $\mathcal{P}$ be all (P,Q) satisfying c.1 and c.2 and such that $P[0,a_1] = Q[0,a_2] = 1$.

Let T be the map $T(x,y) = (Z,\Delta)$ and let T_1 be the function defined on $\mathcal{P}$ which takes (P,Q) to the distribution of T , i.e. $T_1(P,Q) = (P \times Q)T^{-1}$. The range $T_1(\mathcal{P})$ is thus a family of distributions of (Z,Δ). We set $\tilde{\mathcal{P}} = T_1(\mathcal{P})$ and will denote the elements of $\tilde{\mathcal{P}}$ by P_H. We equip $\mathcal{P}, \tilde{\mathcal{P}}$ with the weak topology. Under assumption c.1 , T is continuous with probability 1 under each (P,Q) and hence T_1 is continuous on $\mathcal{P}$. Further c.2 ensures that T identifies P in the sense that $P_1 \ne P_2$ implies $T_1(P_1,Q_1) \ne T_1(P_2,Q_2)$. This enables us to define " inverse " of T_1 - we define T_2 on $\tilde{\mathcal{P}}$ by $T_2(T_1(P,Q)) = P$.. An explicit representation of T_2 is given in Tsai(1986)

In this and the next section we use a Dirichlet process prior. The basic references are Ferguson (1973,1974) who introduced and studied its basic properties. An excellent treatment with many new results is the unpublished notes of Sethuraman's lectures on nonparametric priors. Other very useful recent reviews are Ferguson, Phadia and Tiwari (1993) and Sethuraman(1994).

Let α_i be non zero finite measures on $[0,a_i]$ Let P,Q be independent with Dirichlet process prior distribution $D_{\alpha_1}, D_{\alpha_2}$.

Theorem 1 *Suppose that the support of α_i is $[0,a_i], i = 1,2$ and that the true (P_0,Q_0) is in $\mathcal{P}$. If P_0 and Q_0 are both continuous then the posterior distribution of P given $(Z_i,\Delta_i), i = 1,2,\cdots,n$ is consistent at (P_0,Q_0).*

Proof. It can be seen via Sethuraman's construction [Sethuraman ,1994] of the Dirichlet process that $D_{\alpha_1\times\alpha_2}(\mathcal{P}) = 1$. Note first that P_H is distributed as $D_{\alpha T_1^{-1}}$ where $\alpha = \alpha_1 \times \alpha_2$ Thus we are in the case where

$\tilde{Z}_i = (Z_i, \Delta_i)$ are R^2 valued random variables i.i.d. as P_H, with P_H itself having a $D_{\alpha T_1^{-1}}$ prior. The posterior of P_H given $\tilde{Z}_1, \tilde{Z}_2, \cdots, \tilde{Z}_n$ is $D_{\alpha+\sum_1^n \delta_{\tilde{Z}_i}}$ Letting $\alpha_n^* = \alpha + \sum_1^n \delta_{\tilde{Z}_i}$, we note that for each borel set B, $\bar{\alpha_n^*}(B)$ converges to $P_{H_0}(B)$ almost surely P_{H_0} and $\alpha_n^*(R^2) \to \infty$ as n goes to infinity. It follows from Theorem 4 in Dalal and Hall (1980) which is a variation of Theorem 3.3 in Sethuraman and Tiwari(1982) that $D_{\alpha_n^*}$ converges almost surely to $\delta_{P_{H_0}}$. Thus the posterior distribution of P_H given $(Z_i, \Delta_i), i = 1, 2, \cdots, n$ converges weakly to $\delta_{P_{H_0}}$ a.s P_{F_0,G_0}.

We now argue that T_2 is continuous at (P_0, Q_0) which means that T_2 is continuous a.s $\delta_{P_{H_0}}$ and the conclusion of the theorem would follow from the observations made in the last paragraph.

To establish continuity at P_{H_0} suppose that $T_1(P_n, Q_n)$ converges to $T_1(P_0, Q_0)$. Since $\{P_n\}, \{Q_n\}$ are tight, we may suppose by going through a subsequence that, (P_n, Q_n) converges to (P^*, Q^*). Since $Z = min(X.Y)$ is continuous, the distribution of Z under (P_n, Q_n) converges to the distribution of Z under (P^*, Q^*). Hence the distribution of Z under (P^*, Q^*) and (P_0, Q_0) are same. By assumption the latter is continuous and it follows easily that (P^*, Q^*) satisfies c.1. This in turn implies that the distribuiton of (Z, Δ) is same under (P^*, Q^*) and $(P_{0,}, Q_0)$; from which it can be easily concluded that (P^*, Q^*) satisfies c.2. Thus (P^*, Q^*) is in $\mathcal{P}$ and $T_1(P^*, Q^*) = T_1(P_0, Q_0)$ and from the identifiability of P by T, we have $P_0 = P^*$

Remark 1 *We believe that consistency would obtain even when* (F_0, G_0) *are not constrained to have compact support as in theorem 1.*

3 Dirichlet Process prior for hazard rate

A basic theorem used in both this and the next sections is the following result which is implicit in Schwartz (1965) and explicitly attributed to her by Barron (1986). A streamlined proof appears in unpublished lecture notes of Ghosh and Ramamoorthi(1994).

Theorem 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 $\sigma-$ finite measure μ and let P_0 be the true distribution. Suppose the prior π puts positive mass on every Kullback-Leibler ball B_δ around P_0 namely $\pi(\{P : I(P_0 \,||\, P) < \delta)\}) > 0$ where $I(P_0 \,||\, P) = \int p_0 \log \frac{p_0}{p} d\mu$ where p_0, p are the densities of P_0, P with respect to μ, then the posterior is consistent at P_0*

Let F have density f. Following Dykstra and Laud (1981) we assume the hazard rate $q(x) = f(x)/\bar{F}(x)$ is monotone. To fix ideas suppose it is non decreasing in x. To put a prior on the space of q's , let A(x) be a random distribution function such that the probability P_A is distributed as D_α, where α is a finite non zero measure on $[0,\infty]$ and set $q(x) = WA(x)$ where W is a positive random variable which is independent of P_A. If W has a Gamma distribution our prior would reduce to a special case of the Gamma Process prior introduced by Dykstra and Laud (1981). We do not know if our method can be adapted to cover their general case.

Assumption A.

Assumption 1 *P_{F_0} has support in the interval $[0,a_1]$ and has density f_0 with respect to the Lebesgue measure and further the density f_0 is strictly positive on $[0,a_1]$ and bounded above on $[0,a_1]$.*

Assumption 2 *The hazard rate $q_0 = f_0/\bar{F}_0$ is non decreasing and*

Assumption 3 $\int_0^{a_1} \mid log f_0 \mid dF_0 < \infty$

Theorem 3 *Suppose the true distribution P_{F_0} satisfies Assumption A, the support of α is $[0,\infty)$ and $\alpha\{0\} > 0$ and the support of the distribution of W is $[0,\infty)$. Then the posterior is consistent at F_0*

Proof: In view of Schwartz's theorem , it is enough to verify that, for all δ,

$$D_\alpha(B_\delta) > 0, \tag{1}$$

where

$$B_\delta = \{P : I(P_{F_0} \mid\mid P) < \delta\}. \tag{2}$$

First note

$$-\log \bar{F}(x) = \int_0^x q(t)dt. \tag{3}$$

Choose $a_0 < a_1$ such that

$$\int_{a_0}^{a_1} \mid \log f_0 \mid dF_0 < \frac{\delta}{4}, \tag{4}$$

$$q_0(a_0) > 1, \tag{5}$$

$$(-\log \bar{F}_0)(\bar{F}_0(a_0)) < \frac{\delta}{4}, \tag{6}$$

and
$K(a_1 - a_0) < \frac{\delta}{4}$, where

$$K \geq 2 \sup_{0\leq x\leq a_1} f_0(x).$$

The second condition above can be ensured since q_0 must be unbounded as $x \to a_1$, otherwise $-\log \bar{F}_0(x)$ will not tend to infinity as $x \to a_1$.

Since the support of α is $[0, \infty)$ and $\alpha\{0\} > 0)$, given any $0 = t_0 < t_1 < t_2 < \cdots < t_n = a_0$, $[A(t_0), A(t_1 - A(t_0), \cdots, A(t_k) - A(t_{k-1})]$ has a non degenerate Dirichlet distribution . Thus given $\hat{\epsilon}$ and A_0 the set $\{A :\mid A(t_i) - A_0(t_i) < \hat{\epsilon}, i = 1, 2, \cdots, n\}$ has positive D_α measure. Since A_0 is a non decreasing continuous function given $\epsilon > 0$, by choosing $\hat{\epsilon}$ and $0 = t_0 < t_1 < t_2 < \cdots < t_n = a_0$ it is easily seen that

$$D_\alpha\{A : \sup_{0 \le x \le a_0} \mid A(x) - A_0(x) \mid < \epsilon\} > 0. \tag{7}$$

Define

$$A_0(x) = \frac{q_0(x)}{W}, \qquad 0 \le x \le a_0. \tag{8}$$

Note that on the set $q_0(a_0) < W < 2q_0(a_0)$, $A_0(a_0) < 1$.
Hence the prior probability of

$$A_\epsilon = \{(W, A) : q_0(a_0) < W < 2q_0(a_0), \sup_{0 \le x \le a_0} \mid A(x) - A_0(x) \mid < \frac{\epsilon}{W}\} \tag{9}$$

is positive for all $\epsilon > 0$.

To complete the proof of (1), we will now show that given $\delta > 0$, for sufficiently small $\epsilon > 0$,

$$(W, A) \in A_\epsilon \text{ implies } P_F \in B_\delta$$

where F satisfies (3) with $q(x) = WA(x)$.

If (W, A) are in A_ϵ, then

$$\sup_{0 \le x \le a_0} \mid q(x) - q_0(x) \mid < \epsilon \tag{10}$$

which implies (via (3))that for a suitable η

$$\sup_{0 \le x \le a_0} \mid \log \bar{F}(x) - \log \bar{F}_0(x) \mid < \eta. \tag{11}$$

and hence using $f = q\bar{F}$. and using the fact that q_0 is bounded away from zero in $[0, a_0]$ by assumption A , we can choose η and hence ϵ, such that

$$\sup_{0 \le x \le a_0} \mid \, log f(x) - \, log f_0(x) \mid < \frac{\delta}{4} \tag{12}$$

Now,

$$0 \le I(P_{F_0} \mid\mid P_F) = \int_0^{a_0} f_0 \log \frac{f_0}{f} + \int_{a_0}^{a_1} f_0 \log f_0 - \int_{a_0}^{a_1} f_0 \log q - \int_{a_0}^{a_1} f_0 \log \bar{F}$$

The sum of the first two terms is bounded by $\frac{\delta}{2}$, by (12) and (4). The third term can be made less than zero by (5 , 10) and the fact that q is non-decreasing .

Finally, by (3,6,11) the last term

$$\begin{aligned}\int_{a_0}^{a_1} f_0 \log \bar{F} &= \int_{a_0}^{a_1}[f_0(\log \bar{F} - \log \bar{F}(a_0)] - log\bar{F}(a_0)(\bar{F_0}(a_0 - \bar{F_0}(a_1)) \\ &\leq W \int_{a_0}^{a_1} f_0(x)(x - a_0) + \frac{\delta}{4} \\ &\leq 2\frac{f_0(a_0)}{\bar{F_0}(a_0)}.(a_1 - a_0).\bar{F_0}(a_0) + \frac{\delta}{4} \\ &\leq \frac{\delta}{2}.\end{aligned}$$

This completes the proof of the fact that for sufficiently small $\epsilon > 0, (W, A) \in A_\epsilon$ implies $I(P_0 \mid\mid P_F) < \delta$. An application of Schwartz's theorem now completes the proof.

4 A sufficient condition for consistency

Let $\mathcal{P}_1$ be the set of all probabilities P_F , absolutely continuous with respect to Lebesgue measure such that the support of P_F is contained in $[0, a_1]$ and let $\mathcal{P}_2$ be the set of all probabilities P_G, absolutely continuous with respect to Lebesgue measure such that the support of P_G is contained in $[0, a_2]$ and also such that a_2 belongs to the support of P_G. Let π be a prior under which P_F and P_G are independent with marginals π_1 and π_2.

Theorem 4 *Let F_0, G_0 be the true distributions in $\mathcal{P}_1 \times \mathcal{P}_2$. Let*

$$B_\delta = \{P_F \in \mathcal{P}_1 : \int_0^{a_1} f_0 \log \frac{f_0}{f} < \delta\}$$

. *Suppose $\pi_1(B_\delta) > 0$ for all $\delta > 0$. Then the posterior for F given $(Z_i, \Delta_i), i = 1, 2, \cdots, n$, is consistent.*

Proof. Suppose first that the following additional assumption holds.

$$\pi_2\{P_G \in \mathcal{P}_2 : \int_0^{a_1} g_0 \log \frac{g_0}{g} < \delta\} > 0, \qquad \forall\delta.$$

As before denoting the elements of $T_1(\mathcal{P}_1 \times \mathcal{P}_2)$ by P_H the above assumptions imply

$$\pi_1 \times \pi_2\{P_H : \int h_0 \log \frac{h_0}{h} < \delta\} > 0, \qquad \forall\delta.$$

It follows that the posterior for P_H is consistent. Further the "inverse" map T_2 exists and is continuous . This follows from the argument in the proof of Theorem 1 since $\mathcal{P}_1 \times \mathcal{P}_2 \subset \mathcal{P}$. It follows that the marginal of the posterior for P_F is consistent. This proves the theorem under the stronger assumption made above. To relax this, note that the posterior for P_F does not depend on the prior π_2 for G. Hence, the behaviour of the posterior for P_F does not depend on the additional assumption made above on π_2. This completes the proof.

Remark 2 *Ghosh and Ramamoorthi discuss in [8] how one can choose π satisfying the above condition for a rich collection of true P_{F_0,G_0}'s.*

Acknowledgement: We are indebted to Prof. Dipak Dey, who let us have copies of Sethuraman's notes and to Prof. Sethuraman for reference [12] and other information. We thank Mr. Shyamal Kumar for comments which helped us eliminate some inaccuracies .

References

[1] BARRON A.R.(1986) Disussion: On the consistency of Bayes estimates *Ann. Statist.* **14** 1-26

[2] DALAL S.R and HALL G.J.(1980) On approximating parametric Bayes models by nonparametric Bayes models. it Ann. Statist . **8** 664-672.

[3] DIACONIS P and FREEDMAN D.A. (1986) On the consistency of Bayes estimates *Ann. Statist* **14** 1-26

[4] DYKSTRA R.L. and LAUD P.(1981) A Bayesian nonparametric approach to reliability *Ann. Statist* **9** 356-367

[5] FREEDMAN D.A.(1963) On the asymptotic behaviour of Bayes estimates in the discrete case I *Ann. Math.Stat* **34** 1386-1403

[6] FERGUSON T.(1973) A Bayesian analysis of some nonparametric problems *Ann. Statist* **1** 209-230

[7] FERGUSON T.(1974) Prior distributions on the space of probability measures *Ann.Statist* **2** 615-629

[8] FERGUSON T., PHADIA E.G. and TIWARI R.C. (1993) Bayesian nonparametric inference *Current Issues in Statistical Inference: Essays in honor of D.Basu (edited by M.Ghosh and P.K.Pathak).* IMS Lecture Notes -Monograph series **34** 127-150

[9] GHOSH J.K. and RAMAMOORTHI R.V.(1994) *Unpublished lecture notes on Bayesian Asymptotics*

[10] SCHWARTZ L.(1965) On Bayes procedures *Z. Wahrscheinlichkeitstheorie verw.Geb* **4** 10-26

[11] SETHURAMAN J.(1986) *Unpublished notes of CBMS-NSF lectures on Nonparametric priors: notes taken by B.G. Lindsay*

[12] SETHURAMAN J. and TIWARI R.C.(1982) Convergence of Dirichlet measures and the interpretation of their parameter *Statistical Decision Theory and Related Topics III ; edited by S.S. Gupta and J.Berger* **2** 305-315

[13] SETHURAMAN J.(1994) A constructive definition of Dirichlet priors *Statistica Sinica* **4** 639-650

[14] SUSARLA V. and VAN RYZIN J.(1976) Nonparametric Bayesian estimation of survival curves from incomplete observations *J. Amer.Statist.Assoc.* **71** 897-902

[15] SUSARLA V. and VAN RYZIN J (1978) Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples *Ann. Statist* **6** 755-768

[16] TSAI W.Y. (1986) Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian application *Ann.Statist* **14** 238-249