Bayesian analysis of censored data

J.K. Ghosh [a], R.V. Ramamoorthi [b,*] , K.R. Srikanth [b]

[a] *Department of Statistics, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA*

Abstract

We consider consistency of the posterior in the context of right censored data. We establish posterior consistency when the distribution of the lifetime has a Dirichlet distribution and also investigate the case when the prior is generated through a prior for the distribution of the observable (Z, Δ). We also show that a naive extension of these methods to interval censored data leads to peculiar estimates.

Keywords: Lifetime; Posterior consistency; Censored data; Dirichlet prior

1. Introduction

Susarla together with Blum and Van Ryzin initiated the nonparametric Bayesian analysis of right censored data. Starting with Dirichlet process prior they obtained the form of the posterior (Blum and Susarla, 1977) and the Bayes estimate (Susarla and VanRyzin, 1976) and explored the connection with the classical Kaplan–Meier estimate. This note is devoted to a brief exposition and elaboration of their work.

Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be positive i.i.d. random variables with distributions F, G respectively. We view the X's as life times and the Y's as censoring times and also assume that the X's and the Y's are independent. What is observed is the set $(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n)$ where $Z_i = (X_i \wedge Y_i)$ and $\Delta_i = I(X_i \leqslant Y_i)$. The goal is to make inference on F.

Section 1 has a brief discussion of Dirichlet process and Polya trees. In Section 2 we follow Susarla and Blum, and Susarla and VanRyzin and consider a Dirichlet process prior for F. If (F, G) are independent under the prior, the marginal posterior distribution of F given $(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n)$ does not depend on the prior for G and hence the prior for G plays no essential role in the analysis. It is shown that the posterior can be represented as a Polya tree and this representation is used to establish consistency of the posterior (Lavine, 1992; Muliere and Walker, 1997; Srikanth, 1997).

We then generate priors for (F, G) through priors for the distribution of (Z, Δ) and show that the posterior is consistent. This method of constructing priors is closely related to the work of Tsai (1986), who considers Dirichlet process for the distribution of (Z, Δ) to construct self-consistent estimates.

Another popular method is to construct priors for F via the cumulative hazard function. Early examples of such construction can be found in Ferguson and Phadia (1979); Dykstra and Laud (1981). Related processes are the elegant Beta processes (Hjort, 1990) and their generalization to Beta-Stacy processes proposed by Walker and Muliere (1997) and the mixture of Dirichlet models due to Doss (1994). We do not study these processes in this note.

The last section is somewhat tentative. We note that a naive extension of the analysis to interval censored data has serious limitations and recognize that the nice connection with frequentist methods that obtains when $\alpha(\mathbb{R}) \to 0$ in the right censored case, fails here. This is not surprising since Dirichlet process is not a natural conjugate prior in this context.

2. Preliminaries

Let $\mathscr{F}$ denote the space of distribution functions on $\mathbb{R}^+$. If α is a finite measure on $\mathbb{R}^+$ then D_α – the Dirichlet process with base measure α – is a probability measure on $\mathscr{F}$ such that for any $t_1 < t_2 < \cdots < t_k$, $(F(t_1), F(t_2)-F(t_1), \ldots, (1-F(t_k)))$ has a k-dimensional Dirichlet distribution with parameters $(\alpha(-\infty, t_1], \alpha(t_2 - t_1], \ldots, \alpha(t_k, \infty))$. It is well known that if $F \sim D_\alpha$ and if given F, $X_1, X_2, \ldots, X_n$ are i.i.d. with distribution F then the posterior distribution of F given $X_1, X_2, \ldots, X_n$ is $D_{\alpha + \sum \delta_{X_i}}$. It is also well known that the posterior is weakly consistent in the sense that for any F_0 the posterior probability of any weak neighborhood V of F_0 goes to 1 almost surely P_{F_0}. In view of the Glivenko Cantelli theorem it is not surprising that the following stronger result holds. For a proof we refer to Srikanth (1997).

Theorem 1. *Let D_α be the prior on $\mathscr{F}$. If $U = \{F : \sup_{t \geqslant 0} |F(t) - F_0(t)| < \varepsilon\}$, then as $n \to \infty$, the posterior probability of U given $X_1, X_2, \ldots, X_n$ goes to 1 almost surely P_{F_0}.*

For a detailed account of the properties of Dirichlet process we refer to Ferguson (1973,1974), Ferguson et al. (1992) and to Schervish (1995).

Polya tree processes are a generalization of Dirichlet processes. A detailed study of these can be found in Mauldin et al. (1992), Lavine (1992,1994) and Schervish (1995). Here we confine ourselves to just the basic properties.

Let E_j be the set of all sequences of 0's and 1's of length j and let $E^* = \bigcup_j E_j$.

Let $\underline{\underline{T}} = \{\underline{\underline{T}}_n, n \geqslant 1\}$ be a sequence of nested partitions of $\mathbb{R}^+$ into intervals such that $\bigcup_j \underline{\underline{T}}_j$ generates the Borel σ-algebra. Formally, $\underline{\underline{T}}_j = \{B_{\underline{\varepsilon}} : \underline{\varepsilon} \in E_j\}$. At the $(j+1)$th stage each $B_{\underline{\varepsilon}}$ is partitioned into $B_{\underline{\varepsilon}0}$ and $B_{\underline{\varepsilon}1}$. We want each $B_{\underline{\varepsilon}}$ to be an interval and the σ-algebra generated by $\bigcup_{\underline{\varepsilon} \in E^*} B_{\underline{\varepsilon}}$ to be the Borel σ-algebra on $\mathbb{R}^+$.

Definition 1. A prior μ on $\mathscr{F}$ is said to be a Polya tree prior with respect to the partition $\underline{\underline{T}}$ and with parameters $\alpha = \{\alpha_{\underline{\varepsilon}} : \underline{\varepsilon} \in E^*\}$ if under μ

1. $\{P(B_{\underline{\varepsilon}0} | B_{\underline{\varepsilon}}) : \underline{\varepsilon} \in E^*\}$ are a set of independent random variables.
2. For each $\underline{\varepsilon} \in E^*$, $P(B_{\underline{\varepsilon}0} | B_{\underline{\varepsilon}}) \sim \mathrm{Beta}(\alpha_{\underline{\varepsilon}0}, \alpha_{\underline{\varepsilon}1})$.

A Polya tree process ($\mathrm{PT}(\underline{\underline{T}}, \underline{\alpha})$) is determined by both the partition $\underline{\underline{T}}$ and $\underline{\alpha}$. Some of facts about Polya tree processes that we need are:

- Dirichlet process can be characterized as processes that are Polya trees with respect to every sequence of nested partitions.

- If $\mathrm{PT}(\underline{\underline{T}}, \underline{\alpha})$ is the prior on $\mathscr{F}$, and if given $P, X_1, X_2, \ldots, X_n$ are i.i.d. P, then the posterior given $X_1, X_2, \ldots, X_n$ is again a Polya tree process with respect to the partitions $\underline{\underline{T}}$ and with parameters $\alpha_{\underline{\varepsilon}} + \#\{X_i \in B_{\underline{\varepsilon}}\}$.

One feature of interest to us is that Polya tree process permit easy posterior updating even in the presence of partial information. Proof of the following proposition is routine.

Proposition 1. *Let* μ *be a* $\mathrm{PT}(\underline{\underline{T}}, \underline{\alpha})$. *Given* P; $X_1, X_2, \ldots, X_n$ *are i.i.d.* P. *The posterior given* $I_{B_{\underline{\varepsilon}_1}}(X_1)$, $I_{B_{\underline{\varepsilon}_2}}(X_2), \ldots, I_{B_{\underline{\varepsilon}_n}}(X_n)$ *is again a Polya tree with respect to* $\underline{\underline{T}}$ *and with parameters* $\alpha'_{\underline{\varepsilon}} = \alpha_{\underline{\varepsilon}} + \#\{i : B_{\underline{\varepsilon}_i} \subset B_{\underline{\varepsilon}}\}$.

3. Right censored data

The setup that we consider can be described as follows: X and Y are nonnegative random variables corresponding to lifetime and censoring time, with distribution functions F and G, respectively. If F and G are independent under the prior it is enough to specify the prior for F and treat G as fixed and hence we consider priors of the form $D_\alpha \times \delta_{G_0}$.

This model was first investigated by Susarla and Van Ryzin (1976), who obtained the Bayes estimate for F and showed that this Bayes estimate converges to the Kaplan–Meier estimate as $\alpha(\mathbb{R}^+) \to 0$. Blum and Susarla (1977) complemented this result by showing that the posterior distribution is a mixture of Dirichlet processes. Lavine (1992) observed that the posterior can be realized as a Polya tree process. The mixture representation is somewhat cumbersome and we feel that the Polya tree approach is more natural for the censored data framework and makes the computations transparent (Muliere and Walker, 1997; Srikanth, 1997).

Let $\boldsymbol{Z} = (z_1, z_2, \ldots, z_n)$, where $z_1 < \cdots < z_n$. Consider the sequence of nested partitions $\{\pi_m(\boldsymbol{Z})\}_{m \geqslant 1}$ given by

$$\pi_1(\boldsymbol{Z}) : B_0 = (0, z_1], \quad B_1 = (z_1, \infty)$$

$$\pi_2(\boldsymbol{Z}) : B_{00}, B_{01}, B_{10} = (z_1, z_2], \quad B_{11} = (z_2, \infty),$$

and for $l \leqslant (n-1)$, let

$$\pi_{l+1}(\boldsymbol{Z}) : B_{0_l 0}, B_{0_l 1}, \ldots, B_{1_l 0} = (z_l, z_{l+1}], \quad B_{1_l 1} = (z_{l+1}, \infty),$$

where 1_l is a string of 1's of length ℓ, and 0_l is a string of ℓ 0's. The remaining $B_{\boldsymbol{\varepsilon}}$'s are arbitrarily partitioned into two intervals such that $\{\pi_m(\boldsymbol{Z})\}_{m \geqslant 1}$ forms a sequence of nested partitions which generates $\mathscr{B}(\mathbb{R}^+)$.

Let $\alpha_{\varepsilon_1, \ldots, \varepsilon_l} = \alpha(B_{\varepsilon_1, \ldots, \varepsilon_l})$. For any $\{(z_1, \delta_1), \ldots, (z_n, \delta_n)\}$, let $\boldsymbol{Z}^*$ denote the vector of distinct values of the censored observations (those for which the corresponding $\delta = 0$), arranged in an increasing order.

To evaluate the posterior given $(z_1, \delta_1), \ldots, (z_n, \delta_n)$, first look at the posterior given all the uncensored observations among $(z_1, \delta_1), \ldots, (z_n, \delta_n)$. Since the prior on $\boldsymbol{M}(\boldsymbol{X})$ – the space of all distributions for X – is a D_α, the posterior on $\boldsymbol{M}(\boldsymbol{X})$ is Dirichlet with parameter $\alpha + \sum_{(i:\, \delta_i = 1)} \delta_{Z_i}$.

Since a Dirichlet process is a Polya tree with respect to every partition, it is so with respect to $\underline{\underline{T}}^*(\underline{Z}^*)$. Proposition 1 easily leads to the updated parameters $\alpha'_{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k}$. We summarize these observations in the following theorem.

Theorem 2. *Let* $\mu = D_\alpha \times \delta_{G_0}$ *be the prior on* $M(\mathbb{R}^+) \times M(\mathbb{R}^+)$. *Then the posterior distribution* $\mu_1(\cdot | (z_1, \delta_1), \ldots, (z_n, \delta_n))$ *is a Polya tree process with parameters* $\pi_n^{(\boldsymbol{Z}, \boldsymbol{\delta})} = \{\pi_n^*(\boldsymbol{Z}^*)\}_{n \geqslant 1}$ *and* $\alpha_n^{(\boldsymbol{Z}, \boldsymbol{\delta})} = \{\acute{\alpha}_{\varepsilon_1, \ldots, \varepsilon_l}\}$, *where* $\acute{\alpha}_{\varepsilon_1, \ldots, \varepsilon_l} = \alpha_{\varepsilon_1, \ldots, \varepsilon_l} + \sum_{\delta_i = 1} I[Z_i \in B_{\varepsilon_1, \ldots, \varepsilon_l}] + \sum_{\delta_i = 0} I[(z_i, \infty) \subset B_{\varepsilon_1, \ldots, \varepsilon_l}]$.

Remark 1. Note that if $B_{\varepsilon_1, \ldots, \varepsilon_l} = (z_k, \infty)$ then $\alpha'_{\varepsilon_1, \ldots, \varepsilon_l} = \alpha(B_{\varepsilon_1, \ldots, \varepsilon_l})$ + number of individuals surviving at time z_k, and for every other $B_{\varepsilon_1, \ldots, \varepsilon_l}$, $\alpha'_{\varepsilon_1, \ldots, \varepsilon_l} = \alpha(B_{\varepsilon_1, \ldots, \varepsilon_l})$ + number of uncensored observations in $B_{\varepsilon_1, \ldots, \varepsilon_l}$.

We will denote this posterior by $\mathrm{PT}(\pi_n^{(Z,\delta)}, \alpha_n^{(Z,\delta)})$.

For any distribution function F let $\overline{F}(t) := 1 - F(t)$, for $t \in \mathbb{R}$. Since for any z_j with $\delta_j = 0$, the posterior distribution of $\overline{F}(z_j)$ is a product of independent beta random variables, the Bayes estimate of $\hat{\overline{F}}(z_j)$ given the observations $\{(z_1, \delta_1), \ldots, (z_n, \delta_n)\}$ is given by

$$\hat{\overline{F}}_n(z_j) = \prod_{\{i: z^*_{(i)} \leqslant z_j\}} \left[\frac{\alpha(z^*_{(i)}, \infty) + n_i}{\alpha(z^*_{(i-1)}, \infty) + n_i + \lambda_i} \right]$$

where $n_i = \#\{z_k \geqslant z^*_{(i)}\}$, and $\lambda_i = \#\{z_k \in (z^*_{(i-1)}, z^*_{(i)}) : \delta_k = 1\}$.

For $z_{(i-1)} < t < z_{(i)}$, a similar expression can be obtained by viewing the posterior as a Polya tree with respect to a slightly modified sequence of partitions – at the $(i+1)$th stage, instead of the intervals $(z_{(i-1)}, z_{(i)}]$ and $(z_{(i)}, \infty)$, we use the intervals $(z_{(i-1)}, t]$ and (t, ∞) for our partition sequence. Again $\overline{F}(t)$ is a product of independent beta random variables.

Let $\mathscr{F}_0 \subset \mathscr{F}$ be the class of all distribution functions F, such that
1. F and G_0 have no points of discontinuity in common, and
2. $\mathrm{Support}(F) \subset \mathrm{Support}(G_0)$.

Let $F_0 \in \mathscr{F}_0$, and consider the set V_{F_0} of all sequences $\{(z_n, \delta_n)\}_{n \geqslant 1}$ such that
1. For any (z_j, δ_j) with $\delta_j = 0$,
   1.1. $(1/n)\sum_{i=0}^{n} I\{z_i \geqslant z_j\} \to \overline{F}_0(z_j)\overline{G}_0(z_j-)$, and
   1.2. $\hat{\overline{G}}_n(z_j-) \to \overline{G}_0(z_j-)$, where $\hat{G}_n$ is the Kaplan–Meier estimate of G.
2. $\{z_j : \delta_j = 0\}$ is a dense subset in the support of F_0.

It follows from the SLLN that, $(1/n)\sum_{i=0}^{n} I\{Z \geqslant z\} \to \overline{F}_0(z)\overline{G}_0(z-)$, a.s. (F_0, G_0), whenever z is a point of discontinuity of G_0. Also the SLLN for censored data (Stute and Wang, 1993), implies that $\hat{\overline{G}}_n(z-) \to \overline{G}_0(z-)$. Therefore $P^{\infty}_{F_0,G_0}(V_{F_0}) = 1$, where $P^{\infty}_{F_0,G_0}$ is the i.i.d. product measure corresponding to the joint distribution of the entire sequence $(X_1, Y_1)(X_2, Y_2), \ldots$. We next show that the posterior is consistent when the observations are in V_{F_0}.

Theorem 3. *Let α be such that $D_\alpha(\mathscr{F}_0) = 1$. Let $\{(z_n, \delta_n)\}_{n \geqslant 1} \in V_{F_0}$, then the Bayes estimate of F,*

$$\hat{F}_n(\cdot | (z_1, \delta_1), \ldots, (z_n, \delta_n)) = \int F(\cdot) \mathrm{PT}((\pi_n^{(z,\delta)}, \alpha_n^{(z,\delta)})(\mathrm{d}F),$$

converges weakly to $F_0(\cdot)$.

Proof. Consider a fixed sequence $\{(z_j, \delta_j)\}_{n \geqslant 1} \in V_{F_0}$, and let (z_j, δ_j) be a coordinate such that $\delta_j = 0$. By our assumptions on V_{F_0}, it is enough to show that $\overline{F}_n(t) \to \overline{F}_0(t)$, for $t = z_1, z_2, \ldots$. For simplicity of notation we will assume that $\delta_1 = 0$, and show that $\overline{F}_n(z_1) \to \overline{F}_0(z_1)$.

Let $0 = z_{(0)} < z_{(1)} < \cdots < z_{(n(1))} = z_1$, be the z's among $\{(z_1, \delta_1), \ldots, (z_n, \delta_n)\}$ for which the corresponding δ's are 0, and are $\leqslant z_1$. By Theorem 2 (and the remark following it),

$$\hat{\overline{F}}_n(z_1) = \prod_{j=1}^{n(1)} \frac{\alpha(z_{(j)}, \infty) + n_j}{\alpha(z_{(j-1)}, \infty) + n_j + \lambda_j},$$

where $n_j = \sum_k I\{z_k \geqslant z_{(j)}\}$, and $\lambda_j = \sum_{k:\delta_k=1} I\{z_k \in (z_{(j-1)}, z_{(j)})\}$. Rewriting the expression on the right-hand side of the last equation, we get

$$\begin{aligned}\overline{F}_n(z_1) &= \frac{\alpha(z_1,\infty) + \sum I\{z_i \geqslant z_1\}}{\alpha(\mathbb{R}^+) + n} \times \prod_{j=1}^{n(1)-1} \frac{\alpha(z_{(j)},\infty) + n_j}{\alpha(z_{(j)},\infty) + n_{j+1} + \lambda_{j+1}} \\ &= A_n(z_1) \times B_n(z_1),\end{aligned}$$

where

$$A_n(z_1) = \frac{\alpha(z_1,\infty) + \sum I\{z_i \geqslant z_1\}}{\alpha(\mathbb{R}^+) + n} \quad \text{and} \quad B_n(z_1) = \prod_{j=1}^{n(1)-1} \frac{\alpha(z_{(j)},\infty) + n_j}{\alpha(z_{(j)},\infty) + n_{j+1} + \lambda_{j+1}}.$$

Since $\{(z_j,\delta_j)\}_{n\geqslant 1} \in V_{F_0}$ and $\delta_1 = 0$, $A_n(z_1) \to \overline{G}_0(z_1-)\overline{F}_0(z_1)$. Let $d_j = \#\{z_i : \delta_i = 0 \text{ and } z_i = z_{(j)}\}$, then $n_{i+1} + \lambda_{i+1} = n_i - d_i$, and hence

$$B_n(z_1) = \prod_{j=1}^{n(1)-1} \frac{\alpha(z_{(j)},\infty) + n_j}{\alpha(z_{(j)},\infty) + n_j - d_j}.$$

Therefore

$$\prod_{j=1}^{n(1)-1} \frac{n_j - d_j}{n_j} \leqslant (B_n(z_1))^{-1} \leqslant \prod_{j=1}^{n(1)-1} \frac{\alpha(\mathbb{R}^+) + n_j - d_j}{\alpha(\mathbb{R}^+) + n_j}.$$

Since as $n \to \infty$, $\alpha(\mathbb{R}^+)/(\alpha(\mathbb{R}^+) + n_j) \to 0$, and since $\prod_{j=1}^{n(1)-1}(n_j - d_j)/n_j$ is the Kaplan–Meier estimate for $\overline{G}(z_1-)$, $(B_n(z_1))^{-1}$ converges to $\overline{G}_0(z_1-)$. Therefore, $\overline{F}_n(z_1) = A_n(z_1) \times B_n(z_1) \to \overline{F}_0(z_1)$. □

Theorem 4. *Let $\mu = D_\alpha \times \delta_{G_0}$ be the prior on $M(\mathbb{R}^+) \times M(\mathbb{R}^+)$, where $\mathscr{D}_\alpha(\mathscr{F}_0) = 1$. Let $F_0 \in \mathscr{F}_0$, then the marginal posterior on $M(\mathbb{R}^+)$ is weakly consistent at F_0.*

Proof. We will show that the marginal posterior $\mathrm{PT}((\pi_n^{(z,\delta)}, \alpha_n^{(z,\delta)}) \Rightarrow \delta_{F_0}$, for all $\{(z_j,\delta_j)\}_{n\geqslant 1} \in V_{F_0}$. Using an argument along the lines of Theorem 3.1 of Sethuraman and Tiwari (1982) and the last theorem it follows that the posterior sequence $\mathrm{PT}((\pi_n^{(z,\delta)}), \alpha_n^{(z,\delta)})$ is a tight family of probability measures on $M(\mathbb{R}^+)$. To complete the proof, it is enough to show that for any continuous function f on $\mathbb{R}^+$, with a compact support and for any $\delta > 0$, $\mathrm{PT}(\pi_n^{(z,\delta)}, \alpha_n^{(z,\delta)})(U_{F_0}^\delta) \to 1$, where

$$U_{F_0}^\delta = \left\{F : \left|\int f\,\mathrm{d}F - \int f\,\mathrm{d}F_0\right| < \delta\right\}.$$

Let us fix the sequence $\{(z_j,\delta_j)\}_{n\geqslant 1} \in V_{F_0}$. For the remaining portion of the proof, we will write π_n for $\pi_n^{(z,\delta)}$, and α_n for $\alpha_n^{(z,\delta)}$. Also, let $D = \{z_j : \delta_j = 0\}$.

Let f have support $[0,k]$, and let γ be such that $|x - y| < \gamma$ implies $|f(x) - f(y)| < \delta/3$.

Let $0 = a_1 < a_2 < \cdots < a_{l+1} = k$, be such that $|a_{i+1} - a_i| < \gamma/2$, and let $z_{(1)} < z_{(2)} < \cdots < z_{(l)}$, with $z_{(i)} \in D$, and $z_{(i)} \in (a_i, a_{i+1})$.

Let $f_\delta(z) := \sum_1^{l+1} f(z_{(i)}) I\{z \in (z_{(i-1)}, z_{(i)}]\}$, with $z_{(0)} = 0$, and $z_{(l+1)} = k$. Then $\| f - f_\delta \| = \sup_x |f - f_\delta| < \frac{\delta}{3}$.

Further,

$$\left|\int f\,\mathrm{d}F - \int f\,\mathrm{d}F_0\right| \leqslant \frac{2\delta}{3} + \left|\int f_\delta\,\mathrm{d}F - \int f_\delta\,\mathrm{d}F_0\right|.$$

Let $U_\delta^1 = \{F : |\int f_\delta \, dF - \int f_\delta \, dF_0| < \frac{\delta}{3}\}$, then $U_\delta^1 \subset U_{F_0}^\delta$. For any F, $\int f_\delta \, dF = \sum f(z_{(i)})[\overline{F}(z_{(i-1)}) - \overline{F}(z_{(i)})]$, and hence

$$\left| \int f_\delta \, dF - \int f_\delta \, dF_0 \right| \leqslant 2 \parallel f \parallel \sum |\overline{F}_0(z_{(i)}) - \overline{F}(z_{(i)})|.$$

Since that $\mathrm{PT}(\pi_n, \alpha_n)(U_{F_0}^\delta) \geqslant \mathrm{PT}(\pi_n, \alpha_n)(U_\delta^1)$, to complete our proof, it is enough to show that $\mathrm{PT}(\pi_n, \alpha_n)(F : |\overline{F}(z_{(i)} - \overline{F}_0(z_{(i)}| > \eta) \to 0$, for every $\eta > 0$. This would follow by an application of Markov's inequality, if we show that $E(\overline{F}(t)|(z_1, \delta_1), \ldots, (z_n, \delta_n)) \to \overline{F}_0(t)$, and $E((\overline{F}(t))^2|(z_1, \delta_1), \ldots, (z_n, \delta_n)) \to (\overline{F}_0(t))^2$ for all $t \in D$.

We have already seen that $E(\overline{F}(t)|(z_1, \delta_1), \ldots, (z_n, \delta_n)) \to \overline{F}_0(t)$ for all $t \in D$. We will now show that $E((\overline{F}(t))^2|(z_1, \delta_1), \ldots, (z_n, \delta_n)) \to (\overline{F}_0(t))^2$ for all $t \in D$. For simplicity, let us assume that $z_1 \in D$, and we will use the same notation as in the proof of Theorem 3. Using the properties of Polya tree processes and the Polya tree representation for the posterior, we have

$$E((\overline{F}(z_1))^2|(z_1, \delta_1), \ldots, (z_n, \delta_n)) = A_n^*(z_1) \times B_n^*(z_1),$$

where

$$A_n^*(z_1) = \frac{\alpha(z_1, \infty) + \sum I\{z_i \geqslant z_1\}}{\alpha(\mathbb{R}^+) + n} \times \frac{\alpha(z_1, \infty) + \sum I\{z_i \geqslant z_1\} + 1}{\alpha(\mathbb{R}^+) + n + 1}$$

and

$$B_n^*(z_1) = \prod_{j=1}^{n(1)-1} \frac{\alpha(z_{(j)}, \infty) + n_j}{\alpha(z_{(j)}, \infty) + n_{j+1} + \lambda_{j+1}} \times \prod_{j=1}^{n(1)-1} \frac{\alpha(z_{(j)}, \infty) + n_j + 1}{\alpha(z_{(j)}, \infty) + n_{j+1} + \lambda_{j+1} + 1}.$$

Computations similar to the one used in the proof of the last theorem now yields $A_n^*(z_1) \to (\overline{G}_0(z_1-)\overline{F}_0(z_1))^2$ and $B_n^*(z_1) \to (\overline{G}_0(z_1))^{-2}$, and thus $E((\overline{F}(z_1))^2|(z_1, \delta_1), \ldots, (z_n, \delta_n)) \to (\overline{F}_0(z_1))^2$. Similarly $E((\overline{F}(t))^2|(z_1, \delta_1), \ldots, (z_n, \delta_n)) \to (\overline{F}_0(t))^2$ for all $t \in D$. □

It might be argued that in the censoring context, subjective judgements such as exchangeability are to be made on the observables (Z, Δ) and would hence lead to priors for the distribution of (Z, Δ). The model of independent censoring can be used to transfer this prior to the distribution of the lifetime X.

Formally, let $\boldsymbol{M}_0 \subset \boldsymbol{M}(\boldsymbol{X}) \times \boldsymbol{M}(\boldsymbol{Y})$ be the class of all pairs of distribution functions (F, G) such that

1. F and G have no points of discontinuity in common, and
2. for all $t \geqslant 0$, $F(t) < 1$ and $G(t) < 1$.

Denote by T the function $T(x, y) = (x \wedge y, I_{x \leqslant y})$ and by ϕ the function on $\boldsymbol{M}(\boldsymbol{X}) \times \boldsymbol{M}(\boldsymbol{Y})$ defined by $\phi(P, Q) = (P, Q) \circ T^{-1}$, i.e., $\phi(P, Q)$ is the distribution of T under (P, Q). Let $\boldsymbol{M}_0^* = \phi(\boldsymbol{M}_0)$. The following properties of ϕ can be found in Peterson (1977) and Tsai (1986).

1. On $\boldsymbol{M}_0$, ϕ is 1–1.

2. The maps ϕ, ϕ^{-1} (respectively, on $\boldsymbol{M}_0, \boldsymbol{M}_0^*$) are continuous with respect to the supremum distance on distribution functions, i.e.,

$$\sup_t |P_n(t) - P_0(t)| + \sup_t |Q_n(t) - Q_0(t)| \to 0$$

iff

$$\sup_t |\phi(P_n, Q_n)((0,t] \times 0 - \phi(P_0, Q_0((0,t] \times 0)| + \sup_t |\phi(P_n, Q_n)((0,t] \times 1 - \phi(P_0, Q_0((0,t] \times 1)| \to 0.$$

Peterson (1977) and Tsai (1976) provide explicit representation for ϕ, ϕ^{-1}. While we do not need the explicit representation, we note that every prior on $\boldsymbol{M}_0$ gives rise to a prior on $\boldsymbol{M}_0^*$ via ϕ and every prior on $\boldsymbol{M}_0^*$ induces a prior on $\boldsymbol{M}_0$ through ϕ^{-1}.

Theorem 5. *Let μ be a prior on $\boldsymbol{M}_0$ and $\mu^* = \mu \circ \phi^{-1}$ be the induced prior on $\boldsymbol{M}_0^*$. If $\mu^*(.|(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n))$ on $\boldsymbol{M}_0^*$ is weakly consistent at $\phi(P_0, Q_0)$, P_0, Q_0 continuous, then the posterior $\mu(.|(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n))$ on $\boldsymbol{M}_0$ is weakly consistent at (P_0, Q_0).*

Proof. Immediately follows from the continuity of ϕ^{-1} and Theorem 1, by noting that $\mu(.|(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n))$ on $\boldsymbol{M}_0$ is just the distribution of ϕ^{-1} under $\mu^*(.|(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n))$. □

Sethuraman's (1994) construction shows that it is possible to have a Dirichlet process on $\boldsymbol{M}_0^*$ and the last theorem shows that the induced prior on $\boldsymbol{M}_0$ has good consistency properties. A result like this was proved in Ghosh and Ramamoorthi (1995, Theorem 1). However, in that paper $D_{\alpha_1} \times D_{\alpha_2}$ and $D_{\alpha_1 \times \alpha_2}$ were inadvertently mixed up. A careful look at the proof shows that the theorem applies to the set up just considered but it is a little weaker than Theorem 5 since the result of Peterson and Tsai were not fully exploited.

In addition to consistency, if the empirical distribution of (Z, Δ) is a limit of Bayes estimate on $\boldsymbol{M}_0^*$ then so is the Kaplan–Meier estimate. This method of constructing priors on $\boldsymbol{M}_0$ is on the one hand appealing and merits further investigation – for instance the Dirichlet process on $\boldsymbol{M}_0^*$ arises through a Polya urn scheme and it would be of interest to see the corresponding process for the induced prior. On the other hand any prior on $\boldsymbol{M}_0^*$ induces a prior for both F and G. While this prior would be supported by product measures it is not clear that F and G will be independent under this prior.

Going back to the Susarla–Van Ryzin approach since the Dirichlet process picks discrete distributions with probability 1, it is desirable to have a consistency result for priors that would be supported by densities. However, even for Polya tree priors the argument of Theorem 2 does not go through. We next provide an indirect argument which establishes consistency for Polya tree priors with carefully chosen parameters.

Theorem 6. *Let μ be a prior on $\mathscr{L}_0$ – the set of densities of all (P, Q) in $\boldsymbol{M}_0$ which are absolutely continuous with respect to Lebesgue measure. For any f_0, g_0 in $\mathscr{L}_0$ if, $\mu\{(f,g): \int f_0 \log f_0/f + \int g_0 \log g_0/g < \varepsilon\} > 0$ for all $\varepsilon > 0$ then $\mu(|(Z_1, \Delta_1), (Z_2, \Delta_2), \ldots, (Z_n, \Delta_n))$ is weakly consistent at (f_0, g_0).*

Proof. Since under T the Lebesgue measure gets transformed to a non σ-finite measure, we will view the elements of $\mathscr{F}_0$ as densities with respect to an equivalent probability measure λ. If we denote by $\phi(f,g)$ the density of the distribution of T under (f,g) with respect to $\lambda \circ T^{-1}$, then

$$\int f_0 \log f_0/f + \int g_0 \log g_0/g \leqslant \int \phi(f_0, g_0) \log \phi(f_0, g_0)/\phi(f,g).$$

Consequently Schwartz's theorem (Schwartz, 1965; Barron, 1986; Ghosh and Ramamoorthi, 1998) gives the posterior consistency of μ^* at $\phi(f_0, g_0)$, and hence by Theorem 5 consistency of μ at (f_0, g_0). □

Remark 2. It is shown in Lavine (1994) and in Ghosal et al. (1997) that for suitable choice of parameters of the Polya tree the condition of the theorem is satisfied.

4. Interval censored data

Susarla and Van Ryzin showed that the Kaplan–Meier estimate, which is also the *nonparametric MLE*, is the limit of Bayes estimates with a D_α prior for the distribution of X. The observations in this section show that this result does not carry over to other kinds of censored data.

Here our observation consists of n pairs $(L_i, R_i]$; $1 \leqslant i \leqslant n$ where $L_i \leqslant R_i$ and corresponds to the information $X \in (L_i, R_i]$. We assume that $(L_i, R_i]$; $1 \leqslant i \leqslant n$ are independent and also that the underlying censoring mechanism is independent of the lifetime X so that the posterior distribution depends only on $(L_i, R_i]$; $1 \leqslant i \leqslant n$. Let $t_1 < t_2 < \cdots < t_{k+1}$ denote the end points of $(L_i, R_i]$; $1 \leqslant i \leqslant n$ arranged in increasing order and let $I_j = (t_j, t_{j+1}]$. For simplicity we assume that $t_1 = \min_i L_i$ and $t_{k+1} = \max_i R_i$.

Our starting point is a Dirichlet prior $D(c\alpha_1, c\alpha_2, \ldots, c\alpha_k)$ for $(p_1, p_2, \ldots, p_k)$ where $p_j = P\{X \in I_j\}$. Turnbull (1976) suggested the use of the *nonparametric MLE* obtained from the likelihood function

$$\prod_{i=1}^{n} \left(\sum_{I_j \subset (L_i, R_i]} p_j \right).$$

If $(p_1, p_2, \ldots, p_k)$ has a $D(c\alpha_1, c\alpha_2, \ldots, c\alpha_k)$ prior then the posterior distribution of $(p_1, p_2, \ldots, p_k)$ given $(L_i, R_i]; 1 \leqslant i \leqslant n$ is a mixture of Dirichlet distributions.

Call a vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$, where a_i, is an integer an *imputation of* $(L_i, R_i]; 1 \leqslant i \leqslant n$ if $I_{a_i} \subset (L_i, R_i]$. For an imputaion $\boldsymbol{a}$, let $n_j(\boldsymbol{a})$ be the number of observations assigned to the interval I_j. Formally $n_j(\boldsymbol{a}) = \#\{i : a_i = j\}$.

Let the order $O(\boldsymbol{a})$ of an imputation be $\#\{j\colon n_j(\boldsymbol{a}) > 0\}$. Let $\boldsymbol{A}$ be the set of all imputations of $(L_i, R_i]; 1 \leqslant i \leqslant n$ and let $m = \min_{\boldsymbol{a} \in \boldsymbol{A}} O(\boldsymbol{a})$. Call an imputation $\boldsymbol{a}$ minimal if $O(\boldsymbol{a}) = m$.

It is not hard to see that the posterior distribution of $(p_1, p_2, \ldots, p_k)$ given $(L_i, R_i]$; $1 \leqslant i \leqslant n$ is

$$\sum_{\boldsymbol{a} \in \boldsymbol{A}} C_{\boldsymbol{a}} D(c\alpha_1 + n_1(\boldsymbol{a}), c\alpha_2 + n_2(\boldsymbol{a}), \ldots, c\alpha_k + n_k(\boldsymbol{a})),$$

where

$$C_{\boldsymbol{a}} = \frac{\prod_1^k \Gamma(c\alpha_j + n_j(\boldsymbol{a}))}{\sum_{\boldsymbol{a}' \in \boldsymbol{A}} \prod_1^k \Gamma(c\alpha_j + n_j(\boldsymbol{a}'))}.$$

The Bayes estimate of any p_j is

$$\hat{p}_j = \sum_{\boldsymbol{a} \in A} C_{\boldsymbol{a}} \frac{c\alpha_j + n_j(\boldsymbol{a})}{c + n}.$$

As

$$c \downarrow 0, \quad \frac{c\alpha_j + n_j(\boldsymbol{a})}{c + n} \to \frac{n_j(\boldsymbol{a})}{n}.$$

The behavior of $C_{\boldsymbol{a}}$ is given by the next proposition.

Proposition 2. $\lim_{c \to 0} C_{\boldsymbol{a}} > 0$ *iff* $\boldsymbol{a}$ *is a minimal imputation.*

Proof. Suppose $\boldsymbol{a}$ is not minimal. Let $\boldsymbol{a}_0$ be an imputation with $O(\boldsymbol{a}) > O(\boldsymbol{a}_0)$.

$$C_{\boldsymbol{a}} \leqslant \frac{\prod_1^k \Gamma(c\alpha_j + n_j(\boldsymbol{a}))}{\prod_1^k \Gamma(c\alpha_j + n_j(\boldsymbol{a}_0))} = \frac{\prod_{j=1}^k \Gamma(c\alpha_j)}{\prod_{j=1}^k \Gamma(c\alpha_j)} \frac{\prod_{j:n_j(\boldsymbol{a})\neq 0}(\prod_0^{n_j(\boldsymbol{a})}(c\alpha_j + i)}{\prod_{j:n_j(\boldsymbol{a}_0)\neq 0}(\prod_0^{n_j(\boldsymbol{a}_0)}(c\alpha_j + i))}.$$

Since $O(\boldsymbol{a}) > O(\boldsymbol{a}_0)$ the ratio goes to 0. Conversely if $\boldsymbol{a}$ is minimal it is easy to see that

$$\frac{1}{C_{\boldsymbol{a}}} = \sum_{\boldsymbol{a}' \in A} \frac{\prod_1^k \Gamma(c\alpha_j + n_j(\boldsymbol{a}'))}{\prod_1^k \Gamma(c\alpha_j + n_j(\boldsymbol{a}))}$$

converges to a positive limit. □

Thus the limiting behavior is determined by minimal imputations. A few examples clarify these notions.

Example 1. Consider the right censoring case, i.e., for each i either $L_i = R_i$ or $R_i = t_k$. Any minimal imputation is given assigning each censored observation z_j to one of the uncensored $\{z_i\}$ greater than z_j (and to I_k if the last (largest) observation is censored).

Example 2. Consider the case when we have current status or CaseI interval censored data. Here for each i, either $L_i = t_1$ or $R_i = t_{k+1}$ so that all we know is if X_i is to the right of L_i or to the left of R_i.

1. If $\max_i L_i < \min_i R_i$ then the minimal imputation allocates all the observations to the interval $(\max_i L_i, \min_i R_i]$.
2. In general the minimal imputations have order 2. For example, a consistent assignment of the data to $(t_1, \min_i R_i], (\max_i L_i, t_{k+1}]$ would yield a minimal imputation.

A couple of simple numerical examples help clarify the different cases. In the examples below the prior of the distribution is $D_{c\alpha}$ where α is a probability measure. The limit is taken as $c \to 0$. Corresponding to any imputation $\boldsymbol{a}$, we will call the intervals I_j's for which $n_j(\boldsymbol{a}) > 0$, an allocation, and an allocation corresponding to a minimal imputation will be called a minimal allocation.

Example (a). This example illustrates that the limit of Bayes estimates could be supported on a much bigger set than the NPMLE. The observed data consists of the four intervals $(1, \infty)$, $(2, \infty)$, $(0, 3]$, $(4, \infty)$. The limit of Bayes estimates in this case turns out to be,

$$\tilde{F}(0,1] = \tfrac{1}{22}, \qquad \tilde{F}(1,2] = \tfrac{2}{22}, \quad \tilde{F}(2,3] = \tfrac{6}{22}, \quad \tilde{F}(4,\infty] = \tfrac{13}{22},$$

while the NPMLE is given by

$$\hat{F}(2,3] = \tfrac{1}{2}, \qquad \hat{F}(4,\infty] = \tfrac{1}{2}.$$

In the example above, each minimal allocation consists of only 2 subintervals.

1. $(0, 1]$ and $(4, \infty)$, with the corresponding numbers of X_i's in the subintervals being 1 and 3, respectively, represents a minimal allocation.
2. $(2, 3]$ and $(4, \infty)$ with the corresponding numbers of X_i's in the subintervals being 1 and 3, respectively, represents another minimal allocation.
3. $(2, 3]$ and $(4, \infty)$ with the corresponding numbers of X_i's in the subintervals being 2 and 2, respectively, represents yet another minimal allocation.

Example (b). This example shows that the limit of Bayes estimates could be supported on a smaller set than the NPMLE. The observed data consists of the intervals $(0, 1]$, $(2, \infty)$, $(0, 3]$, $(0, 4]$, $(5, \infty)$. The limit of

Bayes estimates in this case turns out to be

$$\tilde{F}(0,1]=\tfrac{3}{5}, \qquad \tilde{F}(5,\infty)=\tfrac{2}{5},$$

while the NPMLE is given by

$$\hat{F}(0,1]=\tfrac{1}{2}, \qquad \hat{F}(2,3]=\tfrac{1}{6}, \qquad \hat{F}(5,\infty)=\tfrac{1}{3}.$$

As $c\to 0$ while Dirichlet priors leads to strange estimates for the current status data the case $c=1$ seems to present no problems. Even when $c\to 0$ we expect that the limiting behavior will be more reasonable when the data are Case 2 interval censored, in the sense described in Groeneboom (1995). In this case the tendency to push the observation to the extremes would be less pronounced.

Acknowledgements

This research is supported in part by the National Institute of General Medical Sciences under Grant 1 R01 GM 49374.

References

Barron, A., 1986. Discussion of Diaconis, P. and Freedman, D., 1986. On the consistency of Bayes estimates (with discussion). Ann. Stat. 14, 1–67.

Blum, J., Susarla, V., 1977. On the posterior distribution of a Dirichlet process given randomly right censored observations. Stochastics Process. Appl. 5, 207–211.

Doss, H., 1994. Bayesian nonparametric estimation for incomplete data via successive substitution sampling. Ann. Stat. 22, 1763–1786.

Dykstra, R.L., Laud, P., 1981. A Bayesian nonparametric approach to reliability. Ann. Stat. 9, 356–367.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Stat. 1, 209.

Ferguson, T.S., 1974. Prior distributions on spaces of probability measures. Ann. Stat. 2, 615–629.

Ferguson, T.S., Phadia, E.G., 1979. Bayesian nonparametric estimation based on censored data. Ann. Stat. 7, 163–186.

Ferguson, T.S., Phadia, E.G., Tiwari, R.C., 1992. Bayesian nonparametric inference. Current Issues in Statistical Inference: Essays in Honor of D. Basu, vol. 17, Lecture Notes Monograph Series, Institute of Mathematical Statistics, pp. 127–150.

Ghosal, S., Ghosh, J.K., Ramamoorthi, R.V., 1997. Consistent semiparametric Bayesian inference about a location parameter. Preprint.

Ghosh, J.K., Ramamoorthi, R.V., 1995. Consistency of Bayesian inference for survival analysis with or without censoring. Analysis of Censored Data, vol. 27, Lecture Notes Monograph Series, Institute of Mathematical Statistics, pp. 95–103.

Ghosh, J.K., Ramamoorthi, R.V., 1998. Lecture notes on Bayesian asymptotics. Under preparation.

Groeneboom, P., 1995. Nonparametric estimators for interval censoring problems. Analysis of censored data, vol. 27, Lecture Notes Monograph Series, Institute of Mathematical Statistics, pp. 105–128.

Hjort, N.L., 1990. Nonparametric estimators based on beta processes in models for life history data. Ann. Stat. 18, 1259–1294.

Lavine, M., 1992. Some aspects of Polya tree distributions for statistical modelling. Ann. Stat. 20, 1222–1235.

Lavine, M., 1994. More aspects of Polya tree distributions for statistical modelling. Ann. Stat. 22, 1161–1176.

Mauldin, R.D., Sudderth, W.D., Williams, S.C., 1992. Polya trees and random distributions. Ann. Stat. 20, 1203–1221.

Muliere, P., Walker, S., 1997. A Bayesian non-parametric approach to survival analysis using Polya trees. Scand. J. Stat. 24, 331–340.

Peterson, A.V., 1977. Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions. J. Am. Stat. Assoc. 72, 854–858.

Srikanth, K.R., 1997. Posterior consistency in some nonparametric problems. Dissertation, Michigan State University.

Schervish, 1995. Theory of Statistics. Springer, New York.

Schwartz, L., 1965. On Bayes' procedures. Z. Wahrsch. Verw. Gebiete 4, 10–26.

Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Stat. Sinica 4, 639–650.

Sethuraman, J., Tiwari, R., 1982. Convergence of Dirichlet measures and the interpretation of their parameter. In: Shanti, S.G., James, O.B. (Eds.), Statistical Decision Theory and Related topics III, vol. 2. Academic Press, New York, pp. 305–315.

Susarla, V., Van Ryzin, J., 1976. Nonparametric Bayesian estimation of survival curves from incomplete observations. J. Am. Stat. Assoc. 71, 897–902.

Stute, W., Wang, J.-L., 1993. The strong law under random censorship. Ann. Stat. 21, 1591–1607.
Tsai, W.Y., 1986. Estimation of survival curves from dependent censorship models via a generalized self-consistency property with nonparametric Bayesian approach. Ann. Stat. 14, 238–249.
Turnbull, B.W., 1976. The empirical distribution function from arbitrarily grouped, censored, and truncated data. J. Roy. Stat. Soc. Ser. B 38, 290–295.
Walker, S., Muliere, P., 1997. Beta-Stacy processes and a generalization of the Plya-urn scheme. Ann. Stat. 25, 1762–1780.