

## Polya Tree Distributions for Statistical Modeling of Censored Data

ANDREW A. NEATH<sup>†</sup>

aneath@siue.edu

*Department of Mathematics and Statistics, Southern Illinois University Edwardsville, IL, USA*

**Abstract.** Polya tree distributions extend the idea of the Dirichlet process as a prior for Bayesian nonparametric problems. Finite dimensional distributions are defined through conditional probabilities in  $P$ . This allows for a specification of prior information which carries greater weight where it is deemed appropriate according to the choice of a partition of the sample space. Muliere and Walker[7] construct a partition so that the posterior from right censored data is also a Polya tree. A point of contention is that the specification of the prior is partially dependent on the data. In general, the posterior from censored data will be a mixture of Polya trees. This paper will present a straightforward method for determining the mixing distribution.

**Keywords:** Bayesian nonparametric problems, finite dimensional distributions, mixture distributions, survival data

### 1. Introduction

In a nonparametric statistical model, the unknown of interest is the probability measure responsible for generating the data at hand. The Bayesian approach to nonparametric problems is to place a prior directly on the class of probability measures. A traditional choice is to use the Dirichlet process.

**Definition.** (Ferguson[3]) Let  $\alpha$  be a non-null finite measure over  $(\mathfrak{R}, \beta)$  where  $\mathfrak{R}$  denotes the real line and  $\beta$  denotes the class of Borel sets. We say that the probability measure  $P$  follows a Dirichlet process with parameter  $\alpha$  if for every measurable partition  $(B_1, \dots, B_k)$  of  $\mathfrak{R}$ ,  $(P(B_1), \dots, P(B_k))$  follows a Dirichlet distribution with parameter vector  $(\alpha(B_1), \dots, \alpha(B_k))$ .

Ferguson has extended the idea of Bayesian estimation for a probability vector, where the conjugate prior would be the Dirichlet distribution, to the case where one wishes to estimate an entire probability measure. By defining the finite dimensional distributions of  $(P(B_1), \dots, P(B_k))$  as Dirichlet with parameters determined through a measure  $\alpha$ , a distribution on the

---

<sup>†</sup> Requests for reprints should be sent to Andrew A. Neath, Department of Mathematics and Statistics, Southern Illinois University Edwardsville, IL, 62026-1653 USA.

probability  $P$  is implied. The weight of the prior information is the constant  $\alpha(\mathfrak{R})$ .

Polya tree distributions ([5], [6]) extend the idea of the Dirichlet process by defining the finite dimensional distributions as conditional probabilities in  $P$ . This allows for a specification of prior information which carries greater weight in subsets of  $\mathfrak{R}$  where it is deemed appropriate according to the choice of a partition of  $\mathfrak{R}$ .

In this paper, we present an analysis based on nonparametric Bayesian methods when the prior is a Polya tree and the available data is subject to censoring. Section 2 contains the needed mathematical background on Polya tree distributions. Muliere and Walker[7] (subsequently referred to as MW) examine Polya tree priors in survival analysis, constructing a partition so the posterior from right censored data will also be from the family of Polya tree distributions. A point of contention is that the specification of the prior is partially dependent on the data. Without this construction, the posterior from censored data will be a mixture of Polya trees. We will present in Section 3 a straightforward method for determining this mixing distribution. Our concluding section contains an example as well as a discussion on how the choice of a partition influences the posterior estimate.

## 2. Polya Tree Distributions

The necessary partitioning of the sample space for defining a Polya tree is done in stages. Actually, we refer to a sequence of ever finer partitions.

Let  $\{B_0, B_1\}$  be a measurable partition of  $\mathfrak{R}$ . Then let  $\{B_{00}, B_{01}\}$  be a measurable partition of  $B_0$  and  $\{B_{10}, B_{11}\}$  be a measurable partition of  $B_1$ . We continue in this manner. Let  $E^m = \{\epsilon = \epsilon_1 \cdots \epsilon_m : \epsilon_k \in \{0, 1\}\}$ ,  $m = 1, 2, \dots$  with  $E^0 = \{\emptyset\}$ . Let  $E^* = \cup_{m=0}^{\infty} E^m$ . At the  $m^{\text{th}}$  stage,  $B_\epsilon$  for  $\epsilon \in E^m$  is partitioned into  $\{B_{\epsilon 0}, B_{\epsilon 1}\}$ . We allow for the possible split of  $B_{\epsilon 0} = \emptyset, B_{\epsilon 1} = B_\epsilon$ . We require that  $\Pi = \{B_0, B_1, B_{00}, B_{01}, \dots\}$  generates the Borel sets.

Now define a sequence of random variables  $Y = \{Y_0, Y_{00}, Y_{10}, \dots\}$ , and a sequence of nonnegative real-valued parameters  $\mathcal{A} = \{\alpha_0, \alpha_1, \alpha_{00}, \alpha_{01}, \dots\}$ , with  $Y_{\epsilon 0} \sim \text{Beta}(\alpha_{\epsilon 0}, \alpha_{\epsilon 1})$  for  $\epsilon \in E^*$ . The subscripts  $\emptyset 0$  and  $\emptyset 1$  are interpreted as 0 and 1, respectively. The random variables in  $Y$  determine conditional probabilities in  $P$  as  $Y_{\epsilon 0} = P(B_{\epsilon 0} | B_\epsilon)$ .

The particular Polya tree distribution is determined by the partitions in  $\Pi$  and the Beta parameters in  $\mathcal{A}$ .

**Definition.** (Lavine[5]) The random probability measure  $P$  on  $\mathfrak{R}$  is said to have a Polya tree distribution with parameter  $(\Pi, \mathcal{A})$  if there exists random variables  $Y = \{Y_0, Y_{00}, Y_{10}, \dots\}$  such that the following hold:

- (i) the random variables in  $Y$  are independent;
- (ii) for every  $\epsilon \in E^*$ ,  $Y_{\epsilon_0} \sim \text{Beta}(\alpha_{\epsilon_0}, \alpha_{\epsilon_1})$ ;
- (iii) for every  $m = 1, 2, \dots$ , and every  $\epsilon \in E^m$ ,

$$P(B_\epsilon) = \left[ \prod_{\substack{j=1 \\ \{\epsilon_j = 0\}}}^m Y_{\epsilon_{1\dots\epsilon_{j-1}0}} \right] \left[ \prod_{\substack{j=1 \\ \{\epsilon_j = 1\}}}^m 1 - Y_{\epsilon_{1\dots\epsilon_{j-1}0}} \right].$$

We will use the notation  $P \sim \text{PT}(\Pi, \mathcal{A})$ .

The weight of the information on  $P(B_{\epsilon_0}|B_\epsilon)$  is given by  $\alpha_{\epsilon_0} + \alpha_{\epsilon_1}$ . Polya trees are generalizations of Dirichlet processes. A Polya tree is a Dirichlet process if for every  $\epsilon \in E^*$ ,  $\alpha_\epsilon = \alpha_{\epsilon_0} + \alpha_{\epsilon_1}$  ([5]). Without this restriction, one has the option of choosing  $\alpha_{\epsilon_0} + \alpha_{\epsilon_1}$  large to reflect greater weight or small to reflect greater uncertainty as to the nature of the true probability  $P$  on the set  $B_\epsilon$ .

The choice of a partition plays no role in defining a Dirichlet process, so long as the above restriction is met. We shall see that the choice of the sets  $B_\epsilon$ ,  $\epsilon \in E^*$ , in the partition can play an important role in specifying a Polya tree prior.

Random variables  $W_1, \dots, W_n$  are said to be a sample of size  $n$  from  $P$  if

$$W_1, \dots, W_n \Big| P \sim \text{iid } P \text{ for } P \sim \text{PT}(\Pi, \mathcal{A}).$$

The unconditional distribution of an observation  $W_i$  is given for  $\epsilon \in E^*$  as

$$\begin{aligned} \mathcal{P}[W_i \in B_\epsilon] &= E \mathcal{P}[W_i \in B_\epsilon | P] \\ &= E[P(B_\epsilon)] \\ &= E[P(B_{\epsilon_1})P(B_{\epsilon_1\epsilon_2}|B_{\epsilon_1}) \dots P(B_\epsilon|B_{\epsilon_1\dots\epsilon_{m-1}})] \\ &= \frac{\alpha_{\epsilon_1}}{\alpha_0 + \alpha_1} \dots \frac{\alpha_{\epsilon_1\dots\epsilon_m}}{\alpha_{\epsilon_1\dots\epsilon_{m-1}0} + \alpha_{\epsilon_1\dots\epsilon_{m-1}1}}. \end{aligned}$$

We can think of this expression as a prior estimate  $\hat{P}_0(B_\epsilon)$  of  $P(B_\epsilon)$ .

The posterior distribution is obtained ([5]) as

$$P \mid \{W_i\} \sim PT(\Pi, \mathcal{A}(W_1, \dots, W_n))$$

where

$$\alpha_\epsilon(W_1, \dots, W_n) = \alpha_\epsilon + n_\epsilon \text{ and } n_\epsilon = \#\{W_i \in B_\epsilon\}. \quad (1)$$

A complete observation  $W$  increases by 1 all parameters corresponding to sets for which  $W \subseteq B_\epsilon$ .

A posterior estimate of the probability measure evaluated at  $B_\epsilon$ ,  $\epsilon \in E^m$ , is

$$\begin{aligned} \hat{P}(B_\epsilon) &= E[P(B_\epsilon) \mid W_1, \dots, W_n] \\ &= \frac{\alpha_{\epsilon_1} + n_{\epsilon_1}}{\alpha_0 + \alpha_1 + n} \dots \frac{\alpha_{\epsilon_1 \dots \epsilon_m} + n_{\epsilon_1 \dots \epsilon_m}}{\alpha_{\epsilon_1 \dots \epsilon_{m-1}0} + \alpha_{\epsilon_1 \dots \epsilon_{m-1}1} + n_{\epsilon_1 \dots \epsilon_{m-1}}}. \end{aligned} \quad (2)$$

Our interest is in using Polya tree distributions for modeling problems with censored data. We look to how the posterior is computed from an observation  $W$  known only to be contained in some set  $A$ . If  $A \in \Pi$ , then the posterior will still be a Polya tree distribution. The updated sequence of parameters is defined through

$$\alpha_\epsilon(A) = \begin{cases} \alpha_\epsilon + 1, & \text{if } A \subseteq B_\epsilon \\ \alpha_\epsilon, & \text{otherwise} \end{cases}. \quad (3)$$

Result (3) will be demonstrated by a small example. Let  $m = 2$  and consider sets  $B_0, B_1$  and  $B_{00}, B_{01}, B_{10}, B_{11}$ . Suppose the Polya tree is parametrized by  $\alpha_0 = 1, \alpha_1 = 3$  and  $\alpha_{00} = 1, \alpha_{01} = 1, \alpha_{10} = 3, \alpha_{11} = 1$ . The prior estimate of  $P(B_0)$  is  $\frac{1}{4}$  and the prior estimates of  $P(B_{00} \mid B_0)$  and  $P(B_{10} \mid B_1)$  are  $\frac{1}{2}$  and  $\frac{3}{4}$ , respectively. Thus, the probability vector  $[P(B_{00}), P(B_{01}), P(B_{10}), P(B_{11})]$  is estimated a priori as

$$\left[ \left(\frac{1}{4}\right)\left(\frac{1}{2}\right), \left(\frac{1}{4}\right)\left(\frac{1}{2}\right), \left(\frac{3}{4}\right)\left(\frac{3}{4}\right), \left(\frac{3}{4}\right)\left(\frac{1}{4}\right) \right] = [.125, .125, .5625, .1875].$$

Suppose we observe two censored observations described as  $[W_1 \in A_1]$ ,  $[W_2 \in A_2]$  with  $A_1 = B_{01}, A_2 = B_{10} \cup B_{11} = B_1$ . Since  $A_1 \subseteq B_{01} \subseteq B_0$ , the updated parameters become  $\alpha_{01} = 2, \alpha_0 = 2$ , and the updated probability vector estimate becomes

$$\left[ \left(\frac{2}{5}\right)\left(\frac{1}{3}\right), \left(\frac{2}{5}\right)\left(\frac{2}{3}\right), \left(\frac{3}{5}\right)\left(\frac{3}{4}\right), \left(\frac{3}{5}\right)\left(\frac{1}{4}\right) \right] = [.133, .267, .45, .15].$$

Note the estimate of  $P(B_{00})$  has increased even though it is known  $W_1 \notin B_{00}$ . This is a consequence of the partitioning, which reflects a belief that an

observation in  $B_{01}$  signals a greater likelihood for event  $B_{00}$  than for events  $B_{10}$  and  $B_{11}$ . For the second observation,  $A_2 \subseteq B_1$ . The only updated parameter is  $\alpha_1 = 4$ . Since no further information is available from this observation, the parameters within the second partition are unchanged. We now have

$$\begin{aligned} & [\hat{P}(B_{00}), \hat{P}(B_{01}), \hat{P}(B_{10}), \hat{P}(B_{11})] \\ &= \left[ \left(\frac{2}{6}\right)\left(\frac{1}{3}\right), \left(\frac{2}{6}\right)\left(\frac{2}{3}\right), \left(\frac{4}{6}\right)\left(\frac{3}{4}\right), \left(\frac{4}{6}\right)\left(\frac{1}{4}\right) \right] = [.111, .222, .5, .167]. \end{aligned}$$

MW[7] show how Polya tree priors update from right censored data provided the partitions are constructed with respect to the censoring times. We take up the problem of determining the posterior in a setting which is slightly broader.

### 3. Mixtures Of Polya Tree Distributions

We begin with the following result for a single censored observation.

**THEOREM 1** *Suppose  $P \sim PT(\Pi, \mathcal{A})$ . Let  $W$  be a sample from  $P$  and suppose that our information consists of  $W \in A$ . Then,*

$$P \Big| \xi, W \in A \sim PT(\Pi, \mathcal{A}(\xi)) \text{ where } \xi = {}^d W \Big| W \in A.$$

We discuss Theorem 1 before presenting a formal proof. Bayesian updating is accomplished after a censored observation by introducing the variable  $\xi$  to represent the unobserved realization  $W$  from  $P$ . To better understand a mixture of Polya trees, we examine our result written as

$$P \Big| \xi \sim PT(\Pi, \mathcal{A}(\xi)), \quad \xi \sim H. \quad (4)$$

Consider the finite-dimensional distributions defined through the random variables in

$$Y^M = \{Y_{\gamma_0} : \gamma \in E_{\star}^M\} \text{ where } E_{\star}^M = \cup_{m=0}^M E^m.$$

Then

$$\mathcal{P} [Y_{\gamma_0} \leq s_{\gamma_0}, \forall \gamma \in E_{\star}^M \mid \xi] = \prod_{\gamma \in E_{\star}^M} \beta(s_{\gamma_0} \mid \alpha_{\gamma_0}(\xi), \alpha_{\gamma_1}(\xi)) \quad (5)$$

where  $\beta(\cdot|a, b)$  denotes the cumulative distribution function for  $\text{Beta}(a, b)$ . So,

$$\begin{aligned} \mathcal{P} [Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_{\star}^M] & \quad (6) \\ &= \int \prod_{\gamma \in E_{\star}^M} \beta(s_{\gamma 0} | \alpha_{\gamma 0}(\xi), \alpha_{\gamma 1}(\xi)) dH(\xi). \end{aligned}$$

Note that  $\beta(s_{\gamma 0} | \alpha_{\gamma 0}(\xi), \alpha_{\gamma 1}(\xi))$  is constant in  $\xi$  over  $B_{\epsilon}$ , for each  $\epsilon \in E^{M+1}$ . Thus,

$$\begin{aligned} \mathcal{P} [Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_{\star}^M] & \quad (7) \\ &= \sum_{\epsilon \in E^{M+1}} \left\{ \mathcal{P}(\xi \in B_{\epsilon}) \prod_{\gamma \in E_{\star}^M} \beta(s_{\gamma 0} | \alpha_{\gamma 0}(B_{\epsilon}), \alpha_{\gamma 1}(B_{\epsilon})) \right\} \end{aligned}$$

where for  $i = 0, 1$

$$\alpha_{\gamma i}(B_{\epsilon}) = \begin{cases} \alpha_{\gamma i} + 1, & \text{if } B_{\epsilon} \subseteq B_{\gamma i} \\ \alpha_{\gamma i}, & \text{otherwise} \end{cases}.$$

The induced posterior on  $Y^M$  is a weighted average of product-Beta distributions, with the weights determined by probabilities on the index variable  $\xi$ .

**Proof:**

We look to determine the induced posterior on  $Y^M$ , with  $M$  arbitrary:

$$\mathcal{P} [Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_{\star}^M | W \in A]$$

For  $\epsilon \in E^{M+1}$ ,

$$\begin{aligned} & \prod_{\gamma \in E_{\star}^M} \beta(s_{\gamma 0} | \alpha_{\gamma 0}(B_{\epsilon}), \alpha_{\gamma 1}(B_{\epsilon})) \\ &= \mathcal{P} [Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_{\star}^M | W \in B_{\epsilon}] \\ &= \mathcal{P} [Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_{\star}^M | W \in B_{\epsilon}, W \in A] \end{aligned}$$

when the event  $[W \in B_{\epsilon}, W \in A]$  is nonempty. Then,

$$\begin{aligned} \mathcal{P} [W \in B_{\epsilon} | W \in A] & \times \prod_{\gamma \in E_{\star}^M} \beta(s_{\gamma 0} | \alpha_{\gamma 0}(B_{\epsilon}), \alpha_{\gamma 1}(B_{\epsilon})) \\ &= \mathcal{P} [Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_{\star}^M, W \in B_{\epsilon} | W \in A]. \end{aligned}$$

Summing over the  $(M + 1)^{\text{th}}$  partition, we obtain

$$\begin{aligned} & \sum_{\epsilon \in E^{M+1}} \left\{ \mathcal{P}[W \in B_\epsilon | W \in A] \times \prod_{\gamma \in E_\star^M} \beta(s_{\gamma 0} | \alpha_{\gamma 0}(B_\epsilon), \alpha_{\gamma 1}(B_\epsilon)) \right\} \\ &= \mathcal{P}[Y_{\gamma 0} \leq s_{\gamma 0}, \forall \gamma \in E_\star^M | W \in A]. \end{aligned} \quad (8)$$

The proof is done since this is the same form as (7) with  $\xi =^d W | W \in A$ .

Now consider a sample of censored observations. An updating scheme will be developed through induction. Let  $\xi_n$  denote the random vector representing the unobserved  $W_1, \dots, W_n$ . Then the distribution on  $P$  prior to the  $(n + 1)^{\text{th}}$  censored observation can be described by the mixture of Polya trees in (4). A direct application of Theorem 1 yields the posterior given  $[W_{n+1} \in A_{n+1}]$  as

$$P | \xi_n, \xi_{n+1} \sim PT(\Pi, \mathcal{A}(\xi_n, \xi_{n+1})) \text{ with } \xi_{n+1} | \xi_n =^d W | W \in A_{n+1}, \xi_n$$

and

$$dH(\xi_n | W_{n+1} \in A_{n+1}) = \frac{\mathcal{P}(W_{n+1} \in A_{n+1} | \xi_n) dH(\xi_n)}{\int \mathcal{P}(W_{n+1} \in A_{n+1} | \xi_n) dH(\xi_n)}. \quad (9)$$

Note the distribution of  $\xi_n$  is changed through the updating procedure. This prevents a stochastic process depiction of the mixing distribution, a technique which has been utilized in similar problems ([1],[2], [8]).

We continue with our example to show the use of result (9). After the first two observations, the Polya tree has parameters  $\alpha_0 = 2$ ,  $\alpha_1 = 4$ ,  $\alpha_{00} = 1$ ,  $\alpha_{01} = 2$ ,  $\alpha_{10} = 3$ ,  $\alpha_{11} = 1$ , and the current estimate of the probability vector is

$$[\hat{P}(B_{00}), \hat{P}(B_{01}), \hat{P}(B_{10}), \hat{P}(B_{11})] = \left[ \frac{1}{9}, \frac{2}{9}, \frac{1}{2}, \frac{1}{6} \right].$$

Recall that this also represents the unconditional distribution for the next observation.

Suppose the next two censored observations are given as  $[W_3 \in A_3]$ ,  $[W_4 \in A_4]$  with  $A_3 = B_{01} \cup B_{10}$ ,  $A_4 = B_{00} \cup B_{01}$ . Since  $A_3$  is not a set in  $\Pi$ , the posterior will become a mixture of Polya trees.

The mixing variable  $\xi_3$  will initially have probabilities

$$\mathcal{P}(\xi_3 \in B_{01}) = \mathcal{P}(W_3 \in B_{01} | W_3 \in A) = \frac{\frac{2}{9}}{\frac{2}{9} + \frac{1}{2}} = \frac{4}{13}$$

and

$$\mathcal{P}(\xi_3 \in B_{10}) = \mathcal{P}(W_3 \in B_{10} | W_3 \in A) = \frac{9}{13}.$$

If  $\xi_3 \in B_{01}$ , then  $\alpha_0$  and  $\alpha_{01}$  increase by 1. Likewise for  $\alpha_1$  and  $\alpha_{10}$  if  $\xi_3 \in B_{10}$ . The newly updated estimate is given as a weighted average:

$$\begin{aligned} & \left(\frac{4}{13}\right)\left[\left(\frac{3}{7}\right)\left(\frac{1}{4}\right), \left(\frac{3}{7}\right)\left(\frac{3}{4}\right), \left(\frac{4}{7}\right)\left(\frac{3}{4}\right), \left(\frac{4}{7}\right)\left(\frac{1}{4}\right)\right] \\ & + \left(\frac{9}{13}\right)\left[\left(\frac{2}{7}\right)\left(\frac{1}{3}\right), \left(\frac{2}{7}\right)\left(\frac{2}{3}\right), \left(\frac{5}{7}\right)\left(\frac{4}{5}\right), \left(\frac{5}{7}\right)\left(\frac{1}{5}\right)\right] \\ & = [.099, .231, .527, .143]. \end{aligned}$$

Now after observing  $[W_4 \in A_4]$ , we must adjust the probabilities on the variable  $\xi_3$  according to (9). So,

$$\begin{aligned} \mathcal{P}(\xi_3 \in B_{01} | W_4 \in A_4) &= \frac{\mathcal{P}(\xi_3 \in B_{01}) \mathcal{P}(W_4 \in A_4 | \xi_3 \in B_{01})}{\mathcal{P}(W_4 \in A_4)} \\ &= \frac{\left(\frac{4}{13}\right)\left(\frac{12}{28}\right)}{\left(\frac{4}{13}\right)\left(\frac{12}{28}\right) + \left(\frac{9}{13}\right)\left(\frac{6}{21}\right)} = \frac{2}{5} \end{aligned}$$

since

$$\begin{aligned} \mathcal{P}(W_4 \in A_4) &= \mathcal{P}(\xi_3 \in B_{01})\mathcal{P}(W_4 \in A_4 | \xi_3 \in B_{01}) \\ &+ \mathcal{P}(\xi_3 \in B_{10})\mathcal{P}(W_4 \in A_4 | \xi_3 \in B_{10}). \end{aligned}$$

The denominator may also be computed from the current estimate of the probability vector.

We see the probability on event  $[W_3 \in B_{01}]$  has increased after observing  $[W_4 \in B_{00} \cup B_{01}]$ . The mathematics follows the logic which says an observation at  $B_{01}$  increases the likelihood of that event.

Conditional on  $\xi_3$ , one can calculate the probabilities on  $\xi_4$  to be

$$\mathcal{P}[\xi_4 \in B_{00} | \xi_3 \in B_{01}] = 1/4 \text{ and } \mathcal{P}[\xi_4 \in B_{00} | \xi_3 \in B_{10}] = 1/3.$$

Then

$$\mathcal{P}[\xi_4 \in B_{01} | \xi_3 \in B_{01}] = 3/4 \text{ and } \mathcal{P}[\xi_4 \in B_{01} | \xi_3 \in B_{10}] = 2/3.$$

The mixing distribution on the variables  $(\xi_3, \xi_4)$  has probabilities

$$\begin{aligned} \mathcal{P}[\xi_3 \in B_{01}, \xi_4 \in B_{00}] &= (2/5)(1/4) = .1, \\ \mathcal{P}[\xi_3 \in B_{01}, \xi_4 \in B_{01}] &= (2/5)(3/4) = .3, \\ \mathcal{P}[\xi_3 \in B_{10}, \xi_4 \in B_{00}] &= (3/5)(1/3) = .2, \\ \mathcal{P}[\xi_3 \in B_{10}, \xi_4 \in B_{01}] &= (3/5)(2/3) = .4. \end{aligned}$$



In this manner, we can determine the mixing distribution for any number of censored observations.

For our example, the observed sets could be written by finite unions of sets in  $\Pi$ . This condition does not impose a significant restriction. As the sample space is partitioned into smaller sets, the observed sets will become closely approximated by sets in  $\Pi$  at some level  $M$ . This is similar motivation to that of sampling an approximate probability measure from a Polya tree distribution by terminating the procedure at a finite level.

#### 4. An Application

We use the data given in Kaplan and Meier[4] and analyzed by MW[7].

deaths occurred at 0.8 3.1 5.4 9.2  
 censoring times at 1.0 2.7 7.0 12.1

The partition defined by MW so that the posterior on the survival distribution will also be a Polya tree is given in Table 1. We consider a partition based upon breaking the interval  $[0, \infty)$  into subintervals of equal length as is seen in Table 2.

Table 1. Muliere, Walker partition

$m = 1$  :  $B_0 = (0, 1.0], B_1 = (1.0, \infty)$   
 $m = 2$  :  $B_{00} = (0, 0.8], B_{01} = (0.8, 1.0], B_{10} = (1.0, 2.7], B_{11} = (2.7, \infty)$   
 $m = 3$  :  $B_{110} = (2.7, 7.0], B_{111} = (7.0, \infty)$   
 $m = 4$  :  $B_{1100} = (2.7, 3.1], B_{1101} = (3.1, 7.0], B_{1110} = (7.0, 12.1],$   
 $B_{1111} = (12.1, \infty)$   
 $m = 5$  :  $B_{11010} = (3.1, 5.4], B_{11011} = (5.4, 7.0], B_{11100} = (7.0, 9.2],$   
 $B_{11101} = (9.2, 12.1]$   
 $m = 6$  :  $B_{110110} = (5.4, 6.0], B_{110111} = (6.0, 7.0]$

Polya tree modeling allows for the prior to be centered on a particular distribution  $P_0$  according to the choice of the parameters in  $\mathcal{A}$ . For  $\epsilon \in E_m$ , let  $\alpha_\epsilon = \gamma_m P_0(B_\epsilon)$  where  $\gamma_m > 0$ . We determine the parameters for our Polya tree prior by letting  $P_0$  correspond to the exponential distribution with failure rate  $\lambda = 0.12$ . The prior estimate of the survival function becomes

$$\hat{S}_0(t) = \exp\{-0.12t\}.$$

Table 2. An alternate partition

$$\begin{aligned}
m = 1 : & \quad B_0 = (0, 8], B_1 = (8, \infty) \\
m = 2 : & \quad B_{00} = (0, 4], B_{01} = (4, 8], B_{10} = (8, 12], B_{11} = (12, \infty) \\
m = 3 : & \quad B_{000} = (0, 2], B_{001} = (2, 4], B_{010} = (4, 6], B_{011} = (6, 8] \\
& \quad B_{100} = (8, 10], B_{101} = (10, 12], B_{110} = (12, 14], B_{111} = (14, \infty)
\end{aligned}$$

The partition follows this pattern as  $m$  gets larger.

Also, we take  $\gamma_m = m^2$ . This choice is sufficient for the Polya tree prior to place probability 1 on distributions which are absolutely continuous.

Aside from differences in the partition  $\Pi$ , our prior is equivalent to the one considered by MW. The posterior estimates of survival can be compared using Tables 3 and 4. Here, we can see the effect of  $\Pi$  as our estimates of survival are larger than those from MW.

We will take the set  $(0, 1.0]$  as a basis of comparison. In the MW partition, this set is  $B_0$ . For the alternate partition, the set  $(0, 1.0]$  is an element of  $\Pi$  for  $m = 4$ , rather than  $m = 1$ .

An observed death represents a complete realization and updates the distribution on  $P$  according to (1). There are four complete observations in our data set. For the MW partition, this results in

$$\alpha_0^{\text{MW}} \rightarrow \alpha_0^{\text{MW}} + 1, \quad \alpha_1^{\text{MW}} \rightarrow \alpha_1^{\text{MW}} + 3.$$

The alternate partition after the four deaths are incorporated gives

$$\begin{aligned}
\alpha_0 & \rightarrow \alpha_0 + 3, \quad \alpha_1 \rightarrow \alpha_1 + 1 \\
\alpha_{00} & \rightarrow \alpha_{00} + 2, \quad \alpha_{01} \rightarrow \alpha_{01} + 1 \\
\alpha_{000} & \rightarrow \alpha_{000} + 1, \quad \alpha_{001} \rightarrow \alpha_{001} + 1 \\
\alpha_{0000} & \rightarrow \alpha_{0000} + 1.
\end{aligned}$$

At this point, the estimates of  $S(1.0)$ , which were equal a priori, are

$$\begin{aligned}
\text{MW: } S(1.0) & = 1 - P(B_0^{\text{MW}}) = .6855 \\
\text{alt: } S(1.0) & = 1 - P(B_{0000}) \\
& = 1 - P(B_0)P(B_{00}|B_0)P(B_{000}|B_{00})P(B_{0000}|B_{000}) \\
& = .867.
\end{aligned}$$

An observation known to be in a set  $B_\epsilon$  for  $\epsilon \in E^m$  will have a greater impact on the probability estimate  $\hat{P}(B_\epsilon)$  when  $m$  is small. Lavine[5] states

Table 3. Survival estimates based upon the MW partition

	$t$	0.8	1.0	2.7	3.1	5.4	
prior:	$\hat{S}_0(t)$	.908	.887	.723	.689	.523	
posterior:	$\hat{S}_{MW}(t)$	.884	.877	.818	.745	.568	
	$t$	6.0	7.0	9.2	12.1		
prior:	$\hat{S}_0(t)$	.487	.432	.332	.234		
posterior:	$\hat{S}_{MW}(t)$	.538	.492	.358	.263		

Table 4. Survival estimates based upon an equal length interval partition

	$t$	1.0	2.0	3.0	4.0	5.0	6.0	7.0
prior:	$\hat{S}_0(t)$	.887	.787	.698	.619	.549	.487	.432
posterior:	$\hat{S}_{MW}(t)$	.902	.843	.788	.695	.647	.561	.530
	$t$	8.0	9.0	10.0	11.0	12.0	13.0	
prior:	$\hat{S}_0(t)$	.383	.340	.301	.267	.237	.210	
posterior:	$\hat{S}_{MW}(t)$	.499	.448	.331	.301	.275	.244	

“choosing  $\alpha_\epsilon$  for small  $m$  typically involves judgments about sets which have large initial probability. For large  $m$ , the consideration is that  $\alpha_\epsilon$  can express a belief about the smoothness of  $P$ .” An indication of this effect is seen here with the observed death at 0.8 having large influence under the MW partition.

The four censored observations are considered next. The MW partition, by design, keeps the posterior within the Polya tree family. The alternate partition gives a mixture of Polya trees for its posterior and thus is not represented as nicely mathematically. It was chosen, however, to reflect true prior uncertainty, not for mathematical tractability.

### Acknowledgments

The author wishes to thank an anonymous referee whose input has led to a much improved version of this manuscript.

## References

1. P. K. Bhattacharya. Posterior distribution of a Dirichlet process from quantal response data. *Annals of Statistics*, 9, 803-811, 1981.
2. J. Blum and V. Susarla. On the posterior distribution of a Dirichlet process given randomly right censored observation. *Stochastic Processes and Their Application*, 5, 207-211, 1977.
3. T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209-230, 1973.
4. E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of American Statistical Association*, 53, 457-481, 1958.
5. M. Lavine. Some aspects of Polya tree distributions for statistical modelling. *Annals of Statistics*, 20, 1222-1235, 1992.
6. R. D. Mauldin, W. D. Sudderth and S. C. Williams. Polya trees and random distributions. *Annals of Statistics*, 20, 1203-1221, 1992.
7. P. Muliere and S. Walker. A Bayesian nonparametric approach to survival analysis using Polya trees. *Scandinavian Journal of Statistics*, 24, 331-340, 1997.
8. A. A. Neath and F. J. Samaniego. On Bayesian estimation of the multiple decrement function in the competing risks problem. *Statistics and Probability Letters*, 31, 75-85, 1996.