

SPECIAL ISSUE “IN MEMORY OF PILAR LORETO IGLESIAS ZUAZOLA”  
RESEARCH PAPER

## Semiparametric Bayesian inference for repeated fractional measurement data

Ying Yang<sup>1</sup>, Peter Müller<sup>2,\*</sup>, and Gary L. Rosner<sup>3</sup>

<sup>1</sup>Bristol-Myers Squibb Company, Plainsboro, NJ 08536, U.S.A,

<sup>2</sup>Department of Biostatistics & Applied Mathematics, The University of Texas, M. D. Anderson Cancer Center, Houston, TX 77030, U.S.A,

<sup>3</sup>Division of Oncology Biostatistics, Johns Hopkins University, Baltimore, MD 21205, U.S.A

(Received: 30 March 2010 · Accepted in final form: 06 April 2010)

### Abstract

We discuss inference for repeated fractional data, with outcomes between 0 to 1, including positive probability masses on 0 and 1. The point masses at the boundaries prevent the routine use of logit and other commonly used transformations of (0, 1) data. We introduce a model augmentation with latent variables that allow for the desired positive probability at 0 and 1 in the model. A linear mixed effect model is imposed on the latent variables. We propose a Bayesian semiparametric model for the random effects distribution. Specifically, we use a Polya tree prior for the unknown random effects distribution. The proposed model can capture possible multimodality and skewness of random effect distribution. We discuss implementation of posterior inference by Markov chain Monte Carlo simulation. The proposed model is illustrated by a simulation study and a cancer study in dogs.

**Keywords:** Fractional data · Linear mixed model · MCMC algorithm · Polya tree · Repeated measurement data · Semiparametric Bayesian inference.

**Mathematics Subject Classification:** Primary 62G99 · Secondary 62P10.

### 1. INTRODUCTION

Random effects are used to model dependence of repeated measurements on patients or other experimental units. Random effects can be thought of as unmeasured covariates whose values can be considered randomly distributed amongst study individuals. For continuous outcomes with normal errors, Laird and Ware (1982) proposed a normal linear random effects model. In this model, random effects are assumed to be centered around the mean regression coefficients for the populations, also known as the fixed effects. Conditional on random effects, repeated observations on a subject are considered independent. Goldstein (1986) and Longford (1987) developed a model that incorporates nested random effects, representing nested group-specific as well as individual-specific sources of heterogeneity (uncontrolled variation) to be modeled. Gilks et al. (1993) presented a linear

---

\*Corresponding author. Email: pmueller@mdanderson.org

multiple-random-effects model that simultaneously accommodates group-specific sources of heterogeneity for several groupings of individuals. They use Gibbs sampling to implement posterior inference. Kleinman and Ibrahim (1998a) described a semiparametric Bayesian generalization of the normal linear random effect model, where a nonparametric prior distribution is specified for the random effects.

In this article, we consider the practically important case when the outcome variable is fraction that is continuous between 0 and 1 plus positive point masses at 0 and 1. We call such data “fractional data”. Statistical modeling on fractional data of this type has not been investigated so far for longitudinal or repeated measurement data. Zero-inflated distributions have been used to model count data (Ridout et al., 1998) but would not be appropriate here. Fractional data can be seen in the cancer studies. For example, investigators may study the proportion of hypoxic cells in the cancer microenvironment study or the proportion of stained cells in immunochemistry (ICH) studies. Such data may take the values between 0 and 1.

A standard approach for constrained data is the use of transformations to remove the constraint, such as a logit or probit transformation. Albert and Chib (1993) propose a probit regression model for binary and polychotomous outcomes. They impose a normal regression structure on latent continuous data. Values of the latent data are simulated from suitable truncated normal distributions. After the latent data have been generated, the posterior distributions of the parameters are computed using standard results from normal linear models. Draws from these posteriors are used to sample new latent data. The process is iterated leading to a Gibbs sampling scheme. For fractional data, however, we are considering, a complication arises from the fact that 0 and 1 are included in the range of possible values, with positive probabilities. This complicates the use of conventional logit or probit transformations.

In this paper, we propose a simulation-based approach to implement posterior inference for the parameters of interest in a model for fractional data. The key idea is to introduce additional latent variables to represent the awkward point masses at 0 and 1. A mixed-effects model is imposed on these latent variables. The model accommodates individual-specific sources of heterogeneity. We start with standard mixed normal linear model assumptions, as usual for continuous data, and then extend to a nonparametric Bayesian model.

The paper is organized as follows. In Section 2 we describe the proposed model for fractional data. In Section 3 we present the normal prior for the random effects in the mixed-effect model. In Section 4 we discuss the Polya tree prior for random effects distribution. We present a simulation study in Section 5 and a cancer study in Section 6. We conclude with a summary discussion in Section 7.

## 2. MODEL FORMULATION ON FRACTIONAL DATA

Suppose that a fractional outcome vector,  $y_i = (y_{i1}, \dots, y_{in_i}), y_{ij} \in [0, 1]$  with  $n_i$  repeated measurements is observed in individual  $i$ . Responses  $y_{ij}$  can also be 0 or 1 with positive probability. Latent variables  $z_{ij}$  are introduced to address this data structure by including point masses at 0 and 1 in the model.

$$y_{ij} = \begin{cases} 0, & \text{if } z_{ij} \leq 0; \\ z_{ij}, & \text{if } 0 < z_{ij} < 1, \quad i = 1, \dots, n, j = 1, \dots, n_i; \\ 1, & \text{if } z_{ij} \geq 1. \end{cases} \quad (1)$$

The  $z_{ij}$  are unknown. The distribution of  $z_{ij}$  is unconstrained and continuous at 0 and 1. We can therefore proceed with standard linear mixed model assumptions as is usual for

continuous data, including normal distribution assumptions. If the application demands, one can use other functional forms for the latent continuous variable. Let  $N_p(t, S)$  denote a  $p$ -dimensional normal probability density function with moments  $(t, S)$ . We construct the following model for  $i$ th individual:

$$z_i = X_i\beta + U_i\theta_i + e_i, \quad (2)$$

where  $\beta$  is a  $p \times 1$  vector of regression coefficients, commonly called fixed effects. The matrix  $X_i$  is an  $n_i \times p$  design matrix of known covariates for  $i$ th individual.  $U_i$  is an  $n_i \times q$  matrix of covariates for the  $q \times 1$  random effect vector  $\theta_i$ , and  $e_i$  is an  $n_i \times 1$  vector of residuals. We assume that  $e_i$  and  $\theta_i$  are independent and  $e_i \sim N_{n_i}(0, \sigma^2 I_{n_i})$ .

For the distribution of random effects, we initially assume a normal distribution. Later, we will introduce alternative and generalized models, as and if indicated by model diagnostics and criticism. We will assign nonparametric priors to the distribution of random effects.

The model includes a monotonicity assumption. We assume that  $P(y_{ij} = 1)$  increases as the mean of subdensity  $p(y_{ij}|0 < y_{ij} < 1)$  increases. We make an analogous assumption for  $P(y_{ij} = 0)$ . We feel this is reasonable in most applications. For example, in an application with  $y_{ij}$  being the fraction of stained cells in immunohistochemistry data, it is reasonable to assume that the probability of all cells being stained (i.e.,  $y_{ij} = 1$ ) increases as the average fraction of stained cells rises.

### 3. NORMAL LINEAR RANDOM EFFECT MODEL

The linear model given in Equation (2) defines the top level sampling model. Without loss of generality we assume conditional independence within experimental units (e.g., dogs), that is, a diagonal variance-covariance matrix. Little would change in the following discussion if we were to assume a non-diagonal variance-covariance matrix.

#### 3.1 PRIOR SPECIFICATION

We complete the model with conjugate priors. For the fixed effects, we assume a conjugate multivariate normal prior

$$\beta \sim N_p(\mu_0, \Sigma_0). \quad (3)$$

Random effects are assumed to arise from the normal random effects model given by

$$\theta_i \stackrel{\text{iid}}{\sim} N_q(0, \Sigma_\theta).$$

The prior on the residual variance is specified as

$$\tau = (\sigma^2)^{-1} \sim \text{Ga}\left(\frac{\gamma_0}{2}, \frac{\lambda_0}{2}\right), \quad (4)$$

where  $\text{Ga}(a, b)$  denotes a gamma distribution with mean  $a/b$  and variance  $a/b^2$ . The conjugate priors of Equation (3) through Equation (4) are chosen for technical convenience. Substantial prior information might require different prior distributions. Finally,  $\mu_0, \Sigma_0, \Sigma_\theta, \gamma_0$ , and  $\lambda_0$  are fixed hyperparameters.

### 3.2 POSTERIOR INFERENCE

We implement posterior simulation by Gibbs sampling, resampling each of the indicated parameters conditional on the currently imputed values of all other parameters and the data. We did not use analytic forms because:

- (i) All priors are only conditionally conjugate. They are not conjugate for the joint posterior distribution;
- (ii) Sampling one parameter at a time avoids manipulating an excessively large design matrix;
- (iii) Additionally, data are truncated by 0 and 1, which breaks joint conjugacy. Moreover, the normal linear mixed model is assumed on  $z$ , not on  $y$ .

**RESAMPLING  $z$**  Conditional on other parameters, resampling the latent variable  $z$  requires truncated normal sampling. From Equations (1) and (2), we find

$$z_{ij} \begin{cases} \sim N(X_{ij}\beta + U_{ij}\theta_i, \sigma^2)I(z_{ij} \leq 0), & \text{if } y_{ij} = 0; \\ = y_{ij}, & \text{if } 0 < y_{ij} < 1; \\ \sim N(X_{ij}\beta + U_{ij}\theta_i, \sigma^2)I(z_{ij} \geq 1), & \text{if } y_{ij} = 1; \end{cases} \quad (5)$$

where  $X_{ij}$  and  $U_{ij}$  are the  $j$ th row of matrices  $X_i$  and  $U_i$ , respectively. Generating random samples from a truncated normal is straightforward. Gelfand et al. (1990) provided an algorithm for sampling from truncated normal distribution.

**RESAMPLING  $\beta$**  We proceed as in a standard normal linear regression. Equation (2) can be re-written as

$$\underbrace{z_i - U_i\theta_i}_{z_i^*} = X_i\beta + e_i.$$

Thus conditional on other parameters,  $z_i^* \sim N(X_i\beta, \tau^{-1})$ . Conditioning on  $z_i$  and  $\theta_i$ ,  $i = 1, \dots, n$ , and combining with the conjugate prior in Equation (3), we find  $p(\beta|\dots) = N(\mu_\beta, v_\beta)$  with moments  $v_\beta = (\Sigma_0^{-1} + \tau \sum_i X_i'X_i)^{-1}$  and  $\mu_\beta = v_\beta(\Sigma_0^{-1}\mu_0 + \tau \sum_i X_i'z_i^*)$ .

We implement Gibbs sampling posterior simulation by iterating over the complete conditional posterior distributions given in the above expressions, starting with initial values for  $\beta, \theta_i, i = 1, \dots, n$ , and  $\tau$ . For later reference, we summarize the Gibbs sampling algorithm:

- (i) Generate  $z$  using Equation (5).
- (ii) Draw  $\beta \sim p(\beta|\dots)$  from the multivariate normal distribution  $N(\mu_\beta, v_\beta)$ .
- (iii) Draw  $\tau \sim p(\tau|\dots)$  from the inverse sigma gamma conditional posterior distribution.
- (iv) For  $i = 1, \dots, n$ , under model (1) through Equation (2),  $\theta_i \sim p(\theta_i|\dots)$  can be easily generated from a normal distribution.

Ergodic averages over the simulated parameter values approximate posterior integrals, including posterior means, posterior predictive distributions, etc. Geweke's test (Geweke, 1992) could be used to verify the convergence of the simulations.

## 4. POLYA TREE RANDOM PRIOR FOR THE RANDOM EFFECT DISTRIBUTION

The normal assumption on the random effects distribution can be restrictive. It is possible for the distribution of random effects to be multimodal and/or with unpredictable types of skewness. In biomedical data, multimodality frequently arises from patient heterogeneity, presence of outliers, exclusion of unknown but important covariates, alternative biologic mechanisms, etc. Examples of heterogenous populations include adult vs. pediatric populations, invasive vs. non-invasive tumors, strokes arising from bursting blood vessels vs. blocked vessels, and so forth. For accurate accounting of uncertainties and for improved prediction, it is clearly critical to account for such heterogeneity. Also, the nature of the heterogeneity might be of interest in itself, as, for example, in the discovery of new (sub-) types of cancer. We achieve the desired generalization with a nonparametric Bayesian model.

A commonly used technical definition of nonparametric Bayesian models is probability models with infinitely many parameters (Bernardo and Smith, 1994). In other words, a nonparametric Bayesian model is a probability model on a function space. Nonparametric Bayesian models are used to avoid critical dependence on parametric assumptions, to make parametric models more robust, and to define model diagnostics and sensitivity analysis for parametric models by embedding them in a larger encompassing nonparametric model (Müller and Quintana, 2004). Bayesian nonparametric and semiparametric approaches include mixture models (West, 1992), Dirichlet process and Dirichlet process mixture models (Ferguson, 1973; Antoniak, 1974; Escobar and West, 1995; MacEachern and Müller, 1998), and Polya tree priors (Lavine, 1992, 1994). For a recent review of nonparametric Bayesian models, see Walker et al. (1999).

In this section, we will explore the Polya tree model as a prior for the random effects distribution in models (1) and (2). Polya trees were proposed as a generalization of Dirichlet process in Bayesian data analysis by Lavine (1992, 1994). The definition, properties, and construction of Polya trees can be found in Appendix.

A random probability measure  $G$ , which is said to have a Polya tree distribution, or a Polya tree prior, with parameter  $(\Pi, \mathcal{A})$ , is written as  $G \sim \text{PT}(\Pi, \mathcal{A})$ . The set  $\Pi$  determines the partition structure of the Polya tree. The parameters  $\alpha_\varepsilon$  in  $\mathcal{A}$  determine the smoothness of a realization of  $G$  and control how quickly the posterior predictive distribution moves from its prior mean to the empirical distribution.

## 4.1 POSTERIOR PREDICTIVE SIMULATION

The joint marginal distribution of a sample  $(X_1, \dots, X_n)$  generated from a random distribution with a Polya tree prior has a closed form. The random probability measure  $G$  can analytically be integrated out. Suppose  $x_i \stackrel{\text{iid}}{\sim} G, i = 1, \dots, n$  and  $G \sim \text{PT}(\Pi, \mathcal{A})$ . Let  $g$  be the density function of  $G$  and  $g_0$  be the density function of the centering distribution  $G_0$  such that  $E(G) = G_0$  (See the appendix for the choice of  $g_0$ ). The marginal joint density of  $(X_1, \dots, X_n)$  is given by

$$\begin{aligned} p(x_1, \dots, x_n) &= \int p(x_1, \dots, x_n | G) dp(G) \\ &= \int \prod_{i=1}^n g(x_i) dp(G) \\ &= f(x_1) \prod_{i=2}^n f(x_i | x_1, \dots, x_{i-1}), \end{aligned}$$

where  $f(x_1) = g_0(x_1)$  and

$$f(x_i|x_1, \dots, x_{i-1}) = g_0(x_i) \lim_{M \rightarrow \infty} \prod_{m=2}^M \frac{\alpha'_{\underline{\varepsilon}_m} (\alpha_{\underline{\varepsilon}_{m-1}0} + \alpha_{\underline{\varepsilon}_{m-1}1})}{\alpha_{\underline{\varepsilon}_m} (\alpha'_{\underline{\varepsilon}_{m-1}0} + \alpha'_{\underline{\varepsilon}_{m-1}1})}, \quad (6)$$

where  $\underline{\varepsilon}_m = \varepsilon_1, \dots, \varepsilon_m$  identifies the level  $m$  subset containing  $x_i$ , i.e.,  $x_i \in B_{\varepsilon_1, \dots, \varepsilon_m}$ . Here  $\alpha'_{\underline{\varepsilon}_m}$  is equal to  $\alpha_{\underline{\varepsilon}_m}$  plus the number of observations among  $x_1, \dots, x_{i-1}$  that belong to  $B_{\varepsilon_1, \dots, \varepsilon_m}$ .

In practice, we can always reduce the right side of Equation (6) to a finite product. Starting at a sufficiently large level  $M_{x_i}$ , the sets  $B_{\varepsilon_1, \dots, \varepsilon_m}$ ,  $m \geq M_{x_i}$ , contain no data points  $x_1, \dots, x_{i-1}$  and thus  $\alpha'_{\underline{\varepsilon}_m} = \alpha_{\underline{\varepsilon}_m}$ , reducing the product to  $M = 2, \dots, M_{x_i}$ . Also for any fixed  $M$ , the right side of Equation (6) can be approximated by a finite product when  $M$  is large.

#### 4.2 A SEMIPARAMETRIC MODEL FOR REPEATED FRACTIONAL DATA

In this section, we improve on certain aspects of the normal linear mixed effect model introduced earlier. In particular, we discuss posterior inference for a model that assumes a Polya tree prior for an unknown random effect distribution. The Polya tree prior is centered around a parametric probability distribution. With this approach, random effects can be directly sampled and inference will be based on the predictive density.

As before, let the observed outcome be  $y_{ij}$  and assume that the distribution of the latent variable  $z_{ij}$  for the  $j$ th measurement from  $i$ th individual follows models (1) and (2). Given  $\beta$  and  $\theta_i$ ,  $z_i$  is normally distributed. The prior specification in a conjugate model for  $\beta$  and  $\tau$  are the same as before, namely, Equations (3) and (4), respectively.

The random effect model is generalized by assuming a nonparametric random effect prior. That is,

$$\theta_i \stackrel{\text{iid}}{\sim} G, \quad G \sim \text{PT}(\Pi, \mathcal{A}).$$

We center the Polya tree distribution around a normal distribution  $G_0$  with median (mean) 0 and a fixed large variance  $\sigma_\theta^2$  (i.e.  $G_0 = \text{N}(0, \sigma_\theta^2)$ ). A fixed partition at level  $m$  is generated by taking

$$B_{\varepsilon_1, \dots, \varepsilon_m} = \left( G_0^{-1} \left( \frac{j}{2^m} \right), G_0^{-1} \left( \frac{j+1}{2^m} \right) \right],$$

where  $j = 0, 1, \dots, 2^m - 1$ .  $\mathcal{A}$  is assumed as

$$\mathcal{A} = \{ \alpha_{\varepsilon_1, \dots, \varepsilon_m} = \text{cm}^2 : \varepsilon_1, \dots, \varepsilon_m \in \{0, 1\}^m \},$$

with a pre-specified value of  $c$ .

To implement Gibbs sampling posterior simulation, we need to generate from the conditional posterior distribution as in Section 3.2. The relevant full conditional posterior distributions include  $p(\beta|z, \theta, \sigma^2)$ ,  $p(\sigma^2|z, \beta, \theta)$ , and  $p(\theta_i|z, \theta_{-i}, \beta, \sigma^2)$ , where  $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$ ,  $i = 1, \dots, n$ . Samples can be obtained using an MCMC algorithm and in particular a Metropolis-Hastings within Gibbs methods.

Conditional on  $\beta$  and the  $\theta_i$ 's,  $z_{ij}$  can be generated by Equation (5) as before. Similarly,  $\beta$  can be sampled from  $\text{N}(\mu_\beta, v_\beta)$  given  $z, \theta = (\theta_1, \dots, \theta_n)$ , and  $\sigma^2$ . The full conditional posterior distribution for  $\tau$  is easily found to be an inverse gamma distribution, allowing random variable generation.

Now we describe how to update  $\theta$ . Let  $z_i = (z_{ij}, j = 1, \dots, n_i)$ . The likelihood for  $\theta_i$ , given  $z, \sigma^2, \beta$ , and  $\theta_{-i}$ , is given by

$$l(z_i|\beta, \theta_i, \theta_{-i}, \sigma^2) = \prod_{j=1}^{n_i} N(z_{ij}|\beta, \theta_i, \sigma^2).$$

The conditional posterior distribution for  $\theta_i$  can be written, up to a constant of proportionality, as

$$p(\theta_i|z, \theta_{-i}, \beta, \sigma^2) \propto l(z_i | \beta, \theta_i, \sigma^2)p(\theta_i|\theta_{-i}), \tag{7}$$

for each  $i = 1, \dots, n$ . This motivates the following Metropolis-Hastings transition probability.

A proposed value  $\theta_i^{(p)}$  is generated from the prior predictive distribution  $p(\theta_i|\theta_{-i})$ . The prior predictive can be written as

$$\theta_i^{(p)} \sim p(\theta_i|\theta_{-i}) = \int G(\theta^{(p)})dp(G | \theta_{-i}) \quad \text{with} \quad p(G | \theta_{-i}) = \text{PT}(\mathcal{A}^*, \Pi),$$

using the posterior PT, conditional on  $\theta_{-i}$ , with  $\mathcal{A}^* = \{\alpha_\varepsilon^* = \alpha_\varepsilon + n_\varepsilon\}$ . Here  $n_\varepsilon = \sum_{h \neq i} I(\theta_h \in B_\varepsilon)$  counts the number of  $\theta_h, h \neq i$ , that fall into the partitioning subset  $B_\varepsilon$ .

We proceed as follows to generate  $\theta_i^{(p)}$ . Recall the definition of  $Y_\varepsilon$  as the random splitting probabilities in the constructive definition of the PT. Starting with  $Y_0$ , we generate the sequence of random probabilities for  $G \sim \text{PT}(\mathcal{A}^*, \Pi)$ , and set  $\varepsilon_0 = 0$  with probability  $Y_0$ , etc. We continue generating random probabilities  $Y_{\varepsilon_0, \dots, \varepsilon_m}$  and  $\varepsilon_m$  until we reach a level  $m$  with  $\alpha_{\varepsilon_0, \dots, \varepsilon_m}^* = \alpha_{\varepsilon_0, \dots, \varepsilon_m}$ , i.e., until we reach a partitioning subset  $B_{\varepsilon_0, \dots, \varepsilon_m}$  that does not contain any of the currently imputed random effects  $\theta_h, h \neq i$ . Finally, we generate  $\theta_i^{(p)}$  from the base measure  $G_0$  restricted to this set  $B_{\varepsilon_0, \dots, \varepsilon_m}$ . This generates  $\theta_i^{(p)} \sim p(\theta_i|\theta_{-i})$ .

The Markov chain moves to the candidate point  $\theta_i^{(p)}$  with probability

$$\min \left\{ 1, \frac{l(z_i|\beta, \theta_i^{(p)}, \sigma^2)}{l(z_i | \beta, \theta_i^{(c)}, \sigma^2)} \right\},$$

where  $\theta_i^{(c)}$  is the current sample.

## 5. A SIMULATION STUDY

In this section, we fit a mixed fractional data model to simulated data. Normal and Polya tree priors are considered as priors of the distributions of random effects, and we compare the resulting inference. We simulated  $n = 100$  random effects from

$$\theta_i \stackrel{\text{iid}}{\sim} 0.5 N(-0.2, 0.15^2) + 0.5 N(0.2, 0.15^2), i = 1, \dots, n.$$

We introduce one covariate,  $X_{ij} \stackrel{\text{iid}}{\sim} N(0.5, 0.3^2), i = 1, \dots, n, j = 1, 2$ . The simulation truth for the regression coefficient was set at  $\beta = (\beta_0, \beta_1) = (0.8, -0.6)$ , and the random residual  $e$  was generated from  $N(0, 0.1^2)$ . Let  $z_{ij} = \beta_0 + X_{ij}\beta_1 + \theta_i + e_{ij}, i = 1, \dots, n, j = 1, 2$ . The observed fractional data are recorded as  $y_{ij} = 1$  if  $z_{ij} \geq 1$ ;  $y_{ij} = z_{ij}$  if  $0 < z_{ij} < 1$ ; and  $y_{ij} = 0$  if  $z_{ij} \leq 0$ .

The histogram of the observed data  $y$  is displayed in Figure 1. From this figure, note the point masses at the boundaries. We analyze the data using model (1) through Equations (3) and (4), with  $z_{ij} = \beta_0 + X_{ij}\beta_1 + \theta_i + e_{ij}$ , and random effects  $\theta_i \stackrel{\text{iid}}{\sim} G$ , and residuals  $e_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ ,  $i = 1, \dots, n$ ,  $j = 1, 2$ .

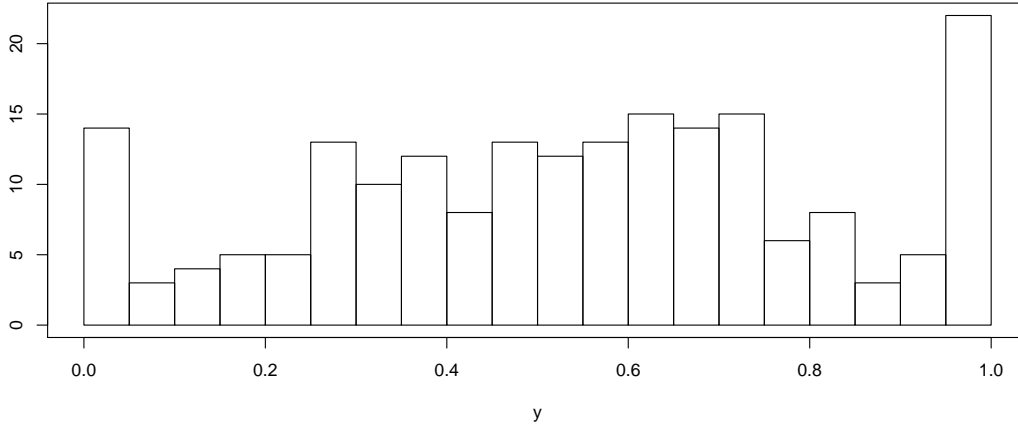


Figure 1. Histogram of  $y$ .

Assuming a flat normal prior on  $\beta = (\beta_0, \beta_1)$ , and  $\beta \sim N_2((0, 0), 100I_2)$ , the random effects distribution is modeled by the following two different approaches.

The distribution of the random effect  $\theta$  is first modeled with normal distribution, that is,  $G(\theta) = N(0, \sigma_\theta^2)$  with a hyper-prior  $1/\sigma_\theta^2 \sim \text{Ga}(0.01, 0.01)$ . We carry out the posterior simulation as described in Section 3.2. The MCMC simulation was repeated 20,000 times. The first 5,000 iterations were discarded as burn-in. The point estimates and 95% posterior intervals of  $\beta$  and selected random effects are presented in Table 1. Figure 2(a) shows the posterior predictive density for  $y_{\text{new},1}$  under normal random effect prior when  $X_{\text{new},1}$  takes the average of all observed  $X$ 's. It shows normal appearance but does not capture the bimodal feature in the data, which suggests that a normal random effects prior is not appropriate in this example.

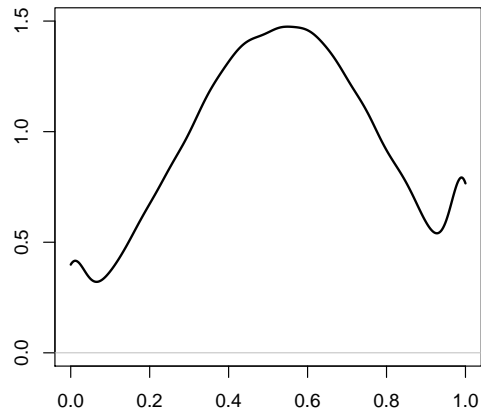
We specify a Polya tree prior for the random effect distribution to allow for multimodality. Priors for other parameters remain unchanged.

We construct the Polya tree prior with a Polya tree centered around  $N(0, 5^2)$ . At the  $m$ th partition level,  $\alpha_{\varepsilon_1, \dots, \varepsilon_m}$  is taken to be  $c m^2$ , with  $c$  fixed at 0.1. Posterior simulation was run for 25,000 iterations, with the first 5,000 being discarded as burn-in period. Every 20th iteration was saved. Posterior summary statistics for the estimated parameters are presented in Table 1. The point estimates for  $\beta_0$  and  $\beta_1$  are fairly close to the simulation truth. Figure 2(b) shows the posterior predictive density for  $y_{\text{new},1}$  under the Polya tree prior for the random effects distribution. It demonstrates the bimodal feature in the data. The posterior predictive density of  $\theta_{n+1}$  is plotted in Figure 2(c). It shows that inference on  $G$  captures the bimodal nature of random effect distribution very well.

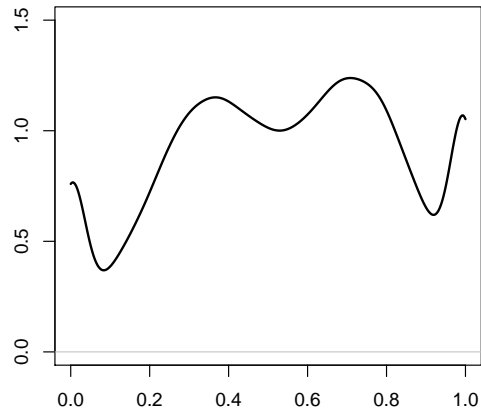
As expected, point estimates of the fixed effects are similar for both models. Posterior intervals for  $\beta_1$  are wider under the Polya tree priors. The increased posterior uncertainty under the Polya tree random effects model is due to the additional randomness in the later model.

The posterior medians for two  $\theta_i$ , shown in Table 1, are slightly different under the two models. This is to be expected. The semiparametric model allows multiple shrinkage. Under the normal model, all estimates for  $\theta_j$  are shrunk towards the common mean,  $E(\theta_j) = 0$ .

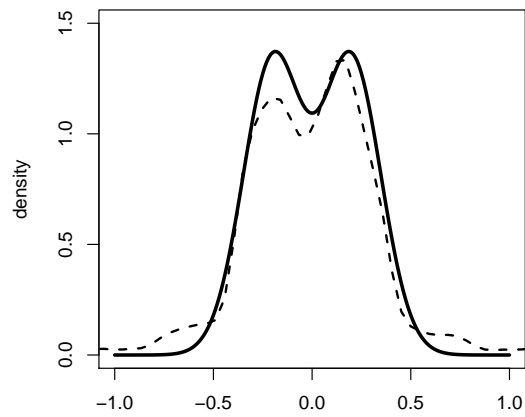




(a)



(b)



(c)

Figure 2. (a) Predictive density of  $y_{\text{new},1}$  under normal random effect distribution, (b) Predictive density of  $y_{\text{new},1}$  under PT prior for random effects distribution (note the bimodal feature) and (c) Estimated random effect distribution  $\bar{G} = E(G|z)$  (dashed line) vs. the simulation truth (solid line)

Table 1. Posterior 2.5%th, 50%th, and 97.5%th percentiles for selected parameters under semiparametric and the parametric normal random effects model.  $\beta_0$  is the intercept,  $\beta_1$  is the fixed effect, and  $\sigma$  is the residual standard deviation.  $\theta_i$  is random effect for  $i$ th subject.

Parameter	Normal prior	Polya tree prior ( $c = 0.1$ )
$\beta_0$	0.830(0.768, 0.890)	0.847(0.723, 0.943)
$\beta_1$	-0.596(-0.659, -0.534)	-0.599(-0.661, -0.533)
$\sigma$	0.093(0.080, 0.108)	0.093(0.081, 0.109)
$\theta_{69}$	0.272(0.133, 0.401)	0.259(0.085, 0.396)
$\theta_{99}$	-0.200(-0.338, -0.051)	-0.219(-0.378, -0.039)
$\sigma_\theta$	0.247(0.194, 0.350)	

## 6. APPLICATION

In cancer studies, some characteristics of the tumor microenvironment are often assessed by histologic evaluation of tumor biopsies. These include oxygenation and proliferation, two important features of the tumor environment that may influence the response to treatment. One approach to measure tumor oxygenation is to measure bound nitroimidazoles based on information from tumor biopsies. The accuracy of biopsy-based methods, however, is related to how precisely the information derived from the biopsies represents the overall tumor microenvironment. Thrall et al. (1997) studied binding of CCI-103F and pimonidazole, both 2-nitroimidazole compounds, in canine solid tumors to assess pretreatment oxygenation and changes in oxygenation during irradiation.

The study reports data for  $n = 9$  dogs, each with a primary solid tumor. Twenty-four hours before the first radiation treatment, CCI-103F was administered intravenously. Immediately prior to the first radiation treatment, up to eight biopsies were obtained from different geographic regions of the tumor. One to four sections from each biopsy sample were placed on glass slides. Slides from four out of eight biopsies were measured 20 minutes after injecting the dye. Slides from the rest of biopsies in the same tumor were measured 24 hours after injecting the dye. The volume fraction of hypoxic tumor tissue was reported by measuring the CCI-103F labelled area in each slide. Responses are recorded as

$$y = \frac{(\text{CCI} - 103\text{F labelled counts})}{(\text{CCI} - 103\text{F labelled counts}) + (\text{unlabelled counts})}.$$

An important feature of the data is that the volume fraction of hypoxic tumor tissue can be 0 or 1 in some slides. Raw cell counts are not reported, only the fraction  $y$ , resulting in the fractional data format discussed earlier. The questions of interest include: “What is the average fraction of hypoxic cells in a tumor?”; “How variable is the fraction within a tumor and between dogs?”; and “How is the measurement of hypoxic cells in the same tumor affected by different measuring times?”.

Figure 3 presents the average fraction of hypoxic cells in each biopsy for each dog. The  $x$ -axis denotes the biopsy label. Biopsies labelled as 1, 2, 3, and 4 were measured 20 minutes after injecting the dye. The fraction of hypoxic cells from biopsies labelled as 5, 6, 7, and 8 were measured 24 hours after injecting the dye.

We use conventional Bayesian and Bayesian semiparametric random effect models to fit the above fractional data. Let  $y_{ijk}$  denote the fraction of hypoxic cells in  $k$ th slide from  $j$ th biopsy of  $i$ th dog. Time  $t_{ij}$  is the measurement time for the  $j$ th biopsy from the  $i$ th dog,  $v_i$  is the volume of tumor from the  $i$ th dog,  $d_i$  denotes a random dog effect and  $b_{j(i)}$  a random biopsy effect nested within the  $i$ th dog. To include point mass probabilities for  $y_{ijk} = 0$

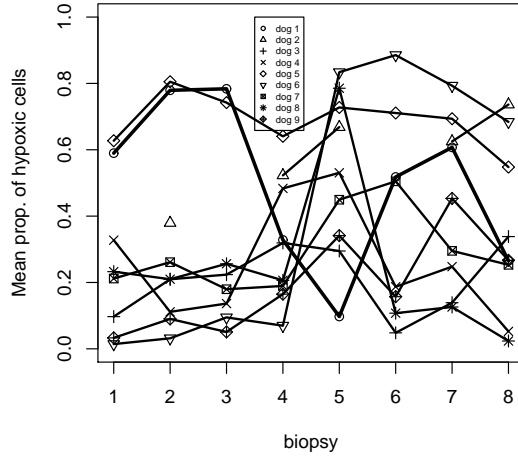


Figure 3. Average fraction of hypoxic cells in each biopsy for each dog.

and  $y_{ijk} = 1$ , we introduce a latent variable  $z_{ijk}$  into the model. As in Equation (1):

$$y_{ijk} = \begin{cases} 0, & \text{if } z_{ijk} \leq 0; \\ z_{ijk}, & \text{if } 0 < z_{ijk} < 1; \\ 1, & \text{if } z_{ijk} \geq 1. \end{cases}$$

The mixed-effects model (2) becomes  $z_{ijk} = \beta_0 + \beta_1 t_{ij} + \beta_2 v_i + d_i + b_{j(i)} + \varepsilon_{ijk}$ , with residuals  $\varepsilon_{ijk} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ ,  $i = 1, \dots, 9, j = 1, \dots, 8, k = 1, \dots, 4$ .

The model is completed with conjugate priors:  $\beta = (\beta_0, \beta_1, \beta_2) \sim N(0, 4I_3)$ ,  $d_i \stackrel{\text{iid}}{\sim} N(0, \sigma_d^2)$ , and a random effect distribution for the biopsy effects:  $b_{j(i)} \stackrel{\text{iid}}{\sim} F$ . As hyper priors, we assume  $1/\sigma_d^2 \sim \text{Ga}(0.01, 0.01)$  and  $1/\sigma^2 \sim \text{Ga}(0.01, 0.01)$ .

Because of the small number of dogs, the distribution of random dog effect is assumed to be normal. Two types of random effects distributions are considered for the distribution of biopsy effect. First, a normal prior  $F = N(0, \sigma_b^2)$  is specified. From other studies, we expect more variability within tumors than between subject-specific averages. Then the parametric prior assumption is relaxed by specifying a nonparametric Polya tree prior, i.e.,  $F \sim \text{PT}(\Pi, \mathcal{A})$ .

### 6.1 NORMAL PRIOR FOR THE BIOPSY EFFECT

Assume  $b_{j(i)} \stackrel{\text{iid}}{\sim} N(0, \sigma_b^2)$  and  $1/\sigma_b^2 \sim \text{Ga}(0.01, 0.01)$ . We ran the Gibbs sampler over 35,000 iterations, with the first 5,000 being discarded as a burn-in period. In addition, because of high autocorrelation, only every 12th iteration was saved and the rest discarded, leading to a total Monte Carlo sample size of 2,500.

The histogram of posterior mean estimated biopsy effects is presented in Figure 4. The distribution of biopsy effects does not show normal appearance. The non-normality may be due to heterogeneity across dogs, or biopsies, or other covariates that are not recorded.

Posterior estimates of model parameters are summarized in Table 2. The results indicate that the proportion of hypoxic cells is strongly related to time but not significantly associated with tumor volume.

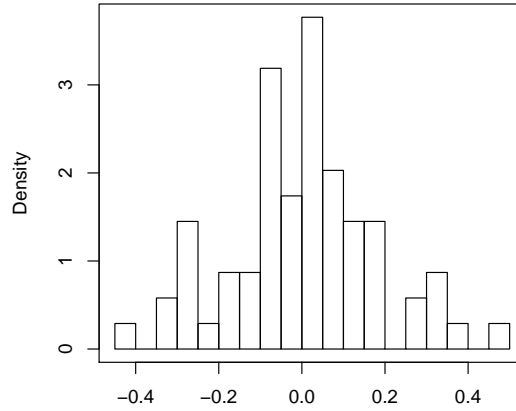


Figure 4. Histogram of mean biopsy effects under the parametric model.

## 6.2 POLYA TREE PRIOR FOR THE BIOPSY EFFECT

Suppose the biopsy effect arises from a random probability measure  $F$ , and we assume a Polya tree prior for  $F$ . We use a Polya tree with centering distribution  $F_0 = N(0, 3^2)$ . We place the same priors on fixed effects and random dog effects as before. The parameters of the Polya tree prior are fixed as follows. The partitioning points to define  $\Pi$  are chosen to be the percentiles of  $F_0$ , and  $c$  is set to 0.1.

Posterior simulation was run for 38,000 iterations, with the first 4,000 being discarded as burn-in period. Every 20th iteration was saved. Convergence of the posterior distributions of parameters was assessed using Geweke's method. The parameter estimates are presented in Table 2. Note that the time effect is no longer significant under the Polya tree prior.

The posterior mean of  $F$  is shown in Figure 5, which indicates high between-biopsy variability. The posterior mean is centered around 0, as we would expect.

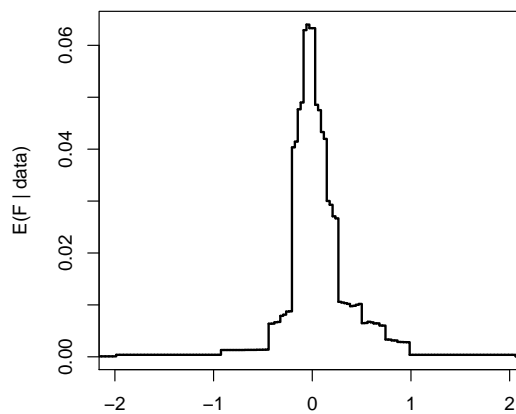
Figure 5. Estimated random biopsy effect distribution  $\bar{F} = E(F|z)$ .

Table 2. Posterior median and central 95% posterior intervals for various parameters from the parametric normal model and Polya tree model. Here,  $\beta_0$  is the intercept,  $\beta_1$  is the slope over time,  $\beta_2$  is the volume of tumor effect,  $\sigma$  is the error standard deviation,  $\sigma_d$  is the standard deviation of dog effect, and  $\sigma_b$  is the standard deviation of biopsy.

Parameter	Normal model	Polya tree model
$\beta_0$	0.303(0.127, 0.507)	0.347(0.173, 0.553)
$\beta_1$	0.113(0.019, 0.212)	0.054(-0.051, 0.169)
$\beta_2$	-0.001(-0.003, 0.003)	-0.001(-0.003, 0.002)
$\sigma$	0.093(0.085, 0.103)	0.093(0.085, 0.102)
$\sigma_d$	0.252(0.154, 0.527)	0.263(0.158, 0.510)
$\sigma_b$	0.197(0.163, 0.241)	

## 7. DISCUSSION

We have proposed Bayesian nonparametric modeling for a class of important inference problems arising in biomedical data analysis. We developed a model and corresponding inference for repeated fractional data model. Proposed techniques included specifying a nonparametric Polya tree prior for the random effects distribution. We provided appropriate posterior simulation schemes.

In the fractional data model, Polya tree priors avoid assuming a specific parametric distribution for random effects. This allows us to estimate the distribution of random effects, which provides insight into population heterogeneity and honestly accounts for related uncertainties.

Polya trees have some practical limitations, however. First, the resulting random probability measure depends on the specific partition sequence adopted. Second, using a fixed partitioning sequence  $\Pi$  results in discontinuities in the predictive distributions. Third, implementations for higher dimensional distributions require extensive housekeeping and are impractical.

Paddock et al. (2003) and Hanson and Johnson (2002) introduced randomized Polya trees to mitigate problems related to the discontinuities. The idea is based on dyadic rational partitions, but instead of taking the nominal half-point, Paddock et al. (2003) randomly chose a cutoff centered around the cutoff point. This construction is shown to mitigate the first two limitations noted above. Hanson and Johnson (2002) consider a mixture with respect to a hyperparameter that defines the partitioning tree. A prior was placed on the spread of the prior distribution of  $G$ , which can be learned from the data. Doing so eliminates the need to choose a centering distribution with adequate spread in ad hoc way.

In the fractional data discussed in the paper, the fractional response outcome was measured at two time points. In other applications the fractional measurements might be obtained from the same biopsy or sample at more than two time points, or the measurement might be made at several time points per cycle for multiple cycles for each subject. The proposed model can be generalized to incorporate the additional random effects, such as nested cycle effect, which is a topic of future research.

## ACKNOWLEDGEMENTS

This work was supported by grant CA075981 from the National Cancer Institute.

## APPENDIX

## A.1 DEFINITION AND BASIC PROPERTIES OF POLYA TREES

Let  $E = [0, 1]$ ,  $E^0 = \emptyset$ ,  $E^m$  be the  $m$ -fold product  $E \times E \times \cdots \times E$ ,  $E^* = \bigcup_{m=0}^{\infty} E^m$  and  $E^N$  be the set of infinite sequences of elements of  $E$ . Let  $\Omega$  be a separable measurable space,  $\pi_0 = \Omega$  and  $\Pi = \{\pi_m; m = 0, 1, \dots\}$  be a separating binary tree of partitions of  $\Omega$ ; that is, let  $\pi_0, \pi_1, \dots$  be a sequence of partitions such that  $\bigcup_{m=0}^{\infty} \pi_m$  generates the measurable sets and such that every  $B \in \pi_{m+1}$  is obtained by splitting some  $B' \in \pi_m$  into two subsets. Let  $B_0 = \Omega$  and, for all  $\varepsilon = \varepsilon_1, \dots, \varepsilon_m \in E^*$ , let  $B_{\varepsilon_0}$  and  $B_{\varepsilon_1}$  be the two subsets into which  $B_{\varepsilon}$  is split.

DEFINITION 7.1 [Lavine (1992)] A random probability measure  $G$  is said to have a Polya tree distribution, or a Polya tree prior, with parameter  $(\Pi, \mathcal{A})$ , written as  $G \sim \text{PT}(\Pi, \mathcal{A})$ , if there exist non-negative numbers  $\mathcal{A} = (\alpha_0, \alpha_1, \alpha_{00}, \dots)$  and random variables  $\mathcal{Y} = (Y_0, Y_1, Y_{00}, \dots)$  such that all random variables in  $\mathcal{Y}$  are independent; for every  $\varepsilon$ ,  $(Y_{\varepsilon_0}, Y_{\varepsilon_1}) \sim \text{Beta}(\alpha_{\varepsilon_0}, \alpha_{\varepsilon_1})$ ; and for every  $m = 1, 2, \dots$ , and every  $\varepsilon = \varepsilon_1, \dots, \varepsilon_m$ ,

$$G(B_{\varepsilon_1, \dots, \varepsilon_m}) = \left( \prod_{j=1, \varepsilon_j=0}^m Y_{\varepsilon_1, \dots, \varepsilon_{j-1}0} \right) \left( \prod_{j=1, \varepsilon_j=1}^m (1 - Y_{\varepsilon_1, \dots, \varepsilon_{j-1}0}) \right)$$

where the first term, i.e. for  $j = 1$ , is interpreted as  $Y_0$  or  $1 - Y_0$ .

The random variable  $Y_{\varepsilon_0}$  is the conditional probability of partition subset  $B_{\varepsilon_0}$  given  $B_{\varepsilon}$ . For instance, for  $m = 2$ ,  $G(B_{00}) = Y_0 Y_{00}$ ,  $G(B_{01}) = Y_0(1 - Y_{00})$ ,  $G(B_{10}) = (1 - Y_0)Y_{10}$ , and  $G(B_{11}) = (1 - Y_0)(1 - Y_{10})$ . The set  $\Pi$  determines the partition structure of the Polya tree. The parameters  $\alpha_{\varepsilon}$  in  $\mathcal{A}$  determine the smoothness of a realization of  $G$  and control how quickly the posterior predictive distribution moves from its prior mean to the empirical distribution.

Several properties facilitate the use of the Polya tree for nonparametric Bayesian inference. Polya trees are conjugate under i.i.d sampling. Assume  $p(G|\Pi, \mathcal{A}) = \text{PT}(\Pi, \mathcal{A})$  and  $x_i \stackrel{\text{iid}}{\sim} G, i = 1, \dots, n$ . Then  $p(G|x, \Pi, \mathcal{A}) = \text{PT}(\Pi, \mathcal{A}^*)$ . The parameters  $\alpha_{\varepsilon}$  are updated by adding the count of points in  $B_{\varepsilon}$ , i.e.,  $\mathcal{A}^* = \{\alpha_{\varepsilon}^* = \alpha_{\varepsilon} + n_{\varepsilon} : \varepsilon \in E^*\}$  and  $n_{\varepsilon}$  is the number of  $x_i$ 's in  $B_{\varepsilon}$ . In words, the posterior distribution of  $G$  under i.i.d sampling is also Polya tree with the same fixed partition sequence. The partitioning probability ( $Y_{\varepsilon}$ ) is generated from beta distribution with updated parameters  $\alpha_{\varepsilon}^*$ , where  $\alpha_{\varepsilon}^*$  is equal to  $\alpha_{\varepsilon}$  plus the number of  $x_1, \dots, x_n$  in subset  $B_{\varepsilon}$ .

The Polya tree includes the Dirichlet process as a special case. A Polya tree is a Dirichlet process if  $\alpha_{\varepsilon} = \alpha_{\varepsilon_0} + \alpha_{\varepsilon_1}$  for every  $\varepsilon \in E^*$  (Ferguson, 1974).

The parameters of a Polya tree can be chosen such that  $G$  is absolutely continuous with probability 1. In particular, any  $\alpha_{\varepsilon_1, \dots, \varepsilon_m} = \rho(m)$  such that  $\sum_{m=1}^{\infty} \rho(m)^{-1} < \infty$  guarantees  $G$  to be absolutely continuous. For example, Walker and Mallick (1999) and Paddock et al. (2003) consider  $\alpha_{\varepsilon_1, \dots, \varepsilon_m} = cm^2$ , where  $c > 0$ .

## A.2. CONSTRUCTION OF A POLYA TREE

Two parameters specify a Polya tree prior, the partition  $\Pi$  and the set  $\mathcal{A}$ . Through choosing  $\Pi$ , we may center the Polya tree prior around a particular continuous distribution  $G_0$ . To do so, we take the partition points to align with percentiles of  $G_0$ . For instance, if  $B_0 = (-\infty, G_0^{-1}(\frac{1}{2})]$  (hence  $B_1 = (G_0^{-1}(\frac{1}{2}), \infty)$ ),  $B_{00} = (-\infty, G_0^{-1}(\frac{1}{4})]$ ,  $\dots$  and  $\alpha_0 = \alpha_1$ ,  $\alpha_{00} = \alpha_{01}, \dots$ , then since  $G(B_0) = Y_0 \sim \text{Be}(\alpha_0, \alpha_1)$ ,  $E(G(B_0)) = \frac{1}{2} = G_0(B_0)$ . Also, e.g.,  $G(B_{00}) = Y_0 Y_{00}$  implies  $E(G(B_{00})) = \frac{1}{4} = G_0(B_{00})$  and for any  $B \in \Pi$ ,  $E(G(B)) = G_0(B)$ .

We need not confine ourselves to quartiles of the form  $G_0^{-1}\left(\frac{j}{2^r}\right)$ .

Figure 6 shows an example of the construction of a Polya tree prior on  $(0, 1] = \Omega$  (Ferguson, 1974). At the top level of the tree,  $\Omega$  is split in half at the dyadic rational, 0.5. Thus  $B_0 = (0, 0.5]$ ,  $B_1 = (0.5, 1]$  and  $\Omega = B_0 \cup B_1$ . At the second level,  $B_0$  and  $B_1$  are again split at 0.25 and 0.75, respectively, which result in subsets  $B_{00} = (0, 0.25]$ ,  $B_{01} = (0.25, 0.5]$ ,  $B_{10} = (0.5, 0.75]$ , and  $B_{11} = (0.75, 1]$ , so on.

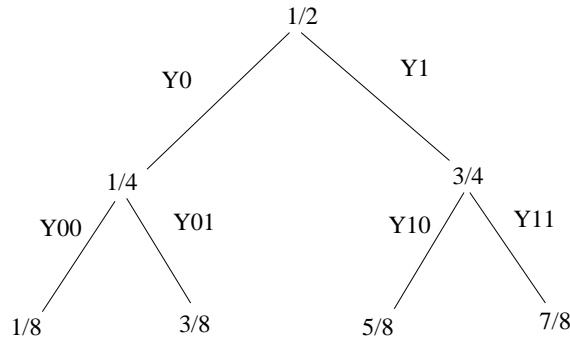


Figure 6. Construction of a Polya tree prior on  $(0, 1]$  (Ferguson, 1974)

One also has to choose the parameters in  $\mathcal{A}$ . The parameters  $\alpha_\varepsilon$  in  $\mathcal{A}$  control how quickly the updated predictive distribution moves from the centering distribution  $G_0$  to the empirical distribution. If the  $\alpha_\varepsilon$ 's are large, then the distribution of  $x_{n+1}|x_1, \dots, x_n$  is close to  $G_0$ . If the  $\alpha_\varepsilon$ 's are small, then the distribution of  $x_{n+1}|x_1, \dots, x_n$  is close to the empirical distribution function. The parameters  $\alpha_\varepsilon$  also express the belief about the smoothness of  $G$ . Ferguson (1974) provides conditions on  $\mathcal{A}$  which yield discrete, continuous singular, and absolutely continuous distributions with probability one. For instance, for level  $m = 1, 2, \dots$ ,  $\alpha_{\varepsilon_1, \dots, \varepsilon_m} = 2^{-m}$  implies a Dirichlet process,  $\alpha_\varepsilon = 1$  yields a random probability  $G$  of a type considered by Dubins and Freedman (1966) and shown to be continuous singular with probability one, and  $\alpha_{\varepsilon_1, \dots, \varepsilon_m} = m^2$  implies an absolutely continuous distribution with probability 1. Walker and Mallick (1999) and Paddock et al. (2003) considered  $\alpha_{\varepsilon_1, \dots, \varepsilon_m} = cm^2$ , where  $c > 0$ .

Therefore, through selection of  $\mathcal{A}$  and  $G_0$ , one can center the Polya tree prior around  $G_0$  arbitrarily close, as determined by  $\mathcal{A}$ , in a manner analogous to the specification of baseline measure and precision parameter in the Dirichlet process.  $\mathcal{A}$  can be thought of as a precision parameter and  $G_0$  as a base measure.

## REFERENCES

- Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669-680.
- Antoniak, C.E., 1974. Mixtures of Dirichlet process with applications to nonparametric problems. *Annals of Statistics*, 2, 1152-1174.
- Bernardo, J.M., Smith, A.F.M., 1994. *Bayesian theory*. Wiley, New York.
- Dubins, L. E., Freedman, D.A., 1966. Random distribution functions. *Proceeding of Fifth Berkeley Symposium Mathematical Statistics and Probability*, 3, 183-214.
- Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577-580.
- Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1, 209-230.

- Ferguson T.S., 1974. Prior distributions on spaces of probability measures. *Annals of Statistics*, 2, 615-629.
- Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). *Bayesian Statistics*, 4, 169-193.
- Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M., 1990. Illustration of Bayesian inference in normal data model using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972-985.
- Gilks, W.R., Wang, C.C., Yvonnet, B., Coursaget, P., 1993. Random-effects models for longitudinal data using Gibbs sampling. *Biometrics*, 49, 441-453.
- Goldstein, H., 1986. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.
- Hanson, T., Johnson, W.O., 2002. Modelling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, 97, 1020-1033.
- Kleinman, K.P., Ibrahim, J.G., 1998a. A semiparametric Bayesian approach to the random effects model. *Biometrics*, 54, 921-938.
- Laird, N.M., Ware, J.M., 1982. Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.
- Lavine, M., 1992. Some aspects of Polya tree distributions for statistical modelling. *Annals of Statistics*, 20, 1222-1235.
- Lavine, M., 1994. More aspects of Polya tree distributions for statistical modelling. *Annals of Statistics*, 22, 1161-1176.
- Longford, N.T., 1987. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.
- MacEachern, S.N., Müller, P., 1998. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 7, 223-338.
- Müller, P., Quintana, F., 2004. Nonparametric Bayesian data analysis. *Statistical Science*, 19, 95-110.
- Paddock, S., Ruggeri, F., Lavine, M., West, M., 2003. Randomised Polya tree models for nonparametric bayesian Inference. *Statistica Sinica*, 13, 443-460.
- Ridout, M.S., Demetrio, C.G.B., Hinde, J.P., 1998. Models for counts data with many zeros. *International Biometric Conference*, 179-192.
- Thrall, D.E., Rosner, G.L., Azuma, C., McEntee, M.C., Raleigh, J.A., 1997. Hypoxia marker labelling in tumor biopsies: quantification of labeling variation and criteria for biopsy sectioning. *Radiotherapy and Oncology*, 44,171-176.
- Walker, S.G., Mallick, B.K., 1999. Semiparametric accelerated life time model. *Biometrics*, 55, 477-483.
- Walker, S.G., Damien, P., Laud, P.W., Smith, A. F. M., 1999. Bayesian nonparametric inference for random distributions and related functions. *Journal of The Royal Statistical Society Series B – Statistical Methodology*, 61, 485-509.
- West, M., 1992. Hyperparameter estimation in Dirichlet process mixture. ISDS discussion paper 92-A03, Duke University, USA.