A Note on Dirichlet Process based Semiparametric Bayesian Models

Arpita Chatterjee

Department of Mathematical Sciences Georgia Southern University, Statesboro, GA, USA
Email: `achatterjee@georgiasouthern.edu`

Abstract Parametric models have been the dominant paradigm for Bayesian inferential works. This is mainly due to its simplicity and straightforward computations. However, given recent computational advances, Semiparametric Bayesian models have become increasingly popular to fit models under flexible distributional assumption. Dirichlet process mixture models form a particular class of Bayesian semiparametric models by assuming a random mixing distribution, taken to be a realization from a Dirichlet process. In this research, we show that even though hierarchical DP models provide flexibility in model fit, they may not perform uniformly better in other aspects as compared to the parametric models. If the DP model gives a better fit, then it should be used regardless of any effect it might have on the power. However, if it results in a reduction in power, then that is just the price of doing a good statistical analysis.

Keywords: Semiparametric Bayesian, Dirichlet Process, Meta Analysis, Binary Data.

1 Introduction

Parametric models have been the dominant paradigm for Bayesian inferential works. This is mainly due to its simplicity and straightforward computations. However, due to recent computational advances, semiparametric Bayesian models are also becoming increasingly popular. These models allow for a greater flexibility in capturing the parameter uncertainty and provide robust alternatives to the conventional parametric models. The Dirichlet process (DP) was first introduced by Ferguson [5] and is arguably the most widely used prior in Bayesian nonparametrics. It is a distribution over probability distributions. A constructive definition of DP is given in Section 2, which leads to the characterization of the DP as a stick-breaking process. For more details, one may refer to Sethuraman and Tiwari [12] and Sethuraman [11].

Dirichlet process mixture (DPM) has been extensively used in Bayesian nonparametrics. Linero and Daniels [10] implemented DP to the MAR model. Ghosh et al. [6] proposed a semiparametric extension of three arm non-inferiority model based on DPM prior. Zhao et al. [13] adopted DPM model to survival data. Burr and Doss [2] introduced a DP based Bayesian semiparametric model for random effects meta-analysis.

In this article, we compare the posterior variance of the mean of the random effects under parametric and DP based semiparametric models. The posterior variance plays a crucial role in Bayesian inference. Large posterior variance results in a wider credible interval which results in a less precise estimate. Also, the posterior variance has huge impact on the power of a statistical test procedure and in Bayesian sample size estimation. It can be shown that under squared error loss, the Bayes risk turns out to be the posterior variance. We proved that, under the setup of linear models with random effects, the semiparametric Bayesian models based on the DP prior result in a larger posterior variance as compared to their parametric counterparts. We establish this result following the notion of Schur Concave function and the theory of majorization. However, for hierarchical Bayesian models under non-conjugate priors, a similar ordering in posterior variance is validated through an extensive simulation study.

The remainder of this paper is designed as follows. In Section 2, we present a brief description about Dirichlet Process (DP) prior. In Section 3 we consider different parametric and semiparametric Bayesian models based on DP prior and establish an ordering between the resulting posterior variances. Section 4 presents a simulation study to examine the performances of the selected models in estimating the mean of the random effects. Section 5 illustrates one application of the meta analysis model using a real data set involving adolescents treated with antidepressants. Some discussions are given in Section 6.

2   Dirichlet Process Prior

In the Bayesian framework the parameters are assumed to be random quantities generated from a probability model, known as prior distribution. Let's consider a random effects model, where the random effects $\theta_1, \theta_2, \ldots, \theta_n \overset{iid}{\sim} G_0$. The parametric distribution G_0 may depend on additional parameters. In practice, G_0 is assumed to be a Normal distribution. On the other hand, under nonparametric Bayesian models, the random effects are assumed to follow an unknown probability distribution, characterized by G, which can be modeled through a DP. For a DP prior we write,

$$\theta_1, \theta_2, \ldots, \theta_n \overset{iid}{\sim} G,$$
$$\text{where} \quad G \sim DP(\kappa, \ G_0(\cdot|\xi)). \tag{1}$$

The parameters G_0 and κ play crucial roles in the definition of DP. $G_0(\cdot|\xi)$ denotes the baseline probability measure, and κ behaves as a precision parameter. The baseline distribution $G_0(\cdot|\xi)$ approximates the true non-parametric shape of G, this can also be viewed as mean parameter of the DP. On the other hand, the positive scalar κ reflects our prior belief about how similar the non-parametric distribution G is to the baseline measure $G_0(\cdot|\xi)$. Moreover, as $\kappa \to \infty$, G looks more like $G_0(\cdot|\xi)$.

Alternatively according to Sethuraman [11] the DP can also be defined as follows:

$$G(\cdot) = \sum_{r=1}^{\infty} p_r \delta_{m_r}(\cdot),$$

where

$$p_1 = \beta_1, \ p_r = \beta_r \prod_{j=1}^{r-1} (1 - \beta_j), r = 1, 2, \cdots,$$

$$\beta_r \overset{iid}{\sim} Beta(1, \alpha), \ r \geq 1, \tag{2}$$

and independent of (β_r), $m_r \overset{iid}{\sim} G_0(\cdot|\xi)$.

This is known as the stick-breaking representation of DP. The term stick-breaking is used because this construction of the weights can be visualized through sequential breaks of a stick of length one. We can obtain a reasonable approximation to G by truncating the infinite sum at a large integer R. Ishwaran and Zarepour [8] suggested that, we can take R to be n, when the number of random effects n is small. Otherwise, $R = \sqrt{n}$. This finite dimensional form of Dirichlet Process, can be easily implemented to OpenBUGS following the blocked Gibbs sampler approach.

Blackwell and MacQueen [1] proposed another interesting construction of the DP known as the Pólya urn representation, which was later used by Escobar [3], and Escobar and West [4] to develop Markov chain samplers. Following the Pólya urn representation, the full conditional of θ_{i+1} is given by

$$\theta_1 \sim G_0,$$
$$\theta_{i+1}|(\theta_1, \ldots, \theta_i) \sim \frac{\kappa}{\kappa + i} G_0(\cdot|\xi) + \sum_{j=1}^{m_i} \frac{n_{j,i}^*}{\kappa + i} \delta_{\theta_j^*} \quad \text{if } i \geq 1$$

where $\{n_{j,i}^*\}$ is the number of θ_is that are equal to the jth unique value θ_j^*. Also, $\delta_{\theta_j^*}$ denotes the degenerate measure at θ_j^*. The Pólya urn representation reveals the inherent clustering property of DP. In particular, for small values of κ, samples from DP, are likely to be composed of a small number of unique θ values, which results in less clusters with relatively large sizes. On the other hand, for large κ values, the DP results in mostly distinct θ values, which corresponds to a large number of comparatively smaller clusters.

3   Random Effects Models

In this section we consider both parametric and semiparametric Bayesian random effects model. The semiparametric model is constructed by assigning a DP prior to the random effects parameter. Suppose $y_1, y_2, \cdots, y_n$ are independent realizations from Normal distributions with y_i associated to mean μ_i. From now on we specify Normal distributions in terms of its mean and precision parameter, unless otherwise stated.

3.1 Normal Location Model

Let's consider the case of unequal precision, that is, y_i follows a Normal distribution with mean μ_i and precision τ_i. Under the parametric set up, the random effects parameter μ_i is assumed to follow $Normal(\mu_0, \tau_\mu)$. On the other hand, to construct the semiparametric Bayesian model based on DP prior we follow the steps as described in Equation (1).

$$\mu_i | G_0, \kappa \sim DP(G_0, \kappa),$$
$$G_0 = \text{Normal } (\mu_0, \tau_\mu). \tag{3}$$

Finally, for both models, we assign a Normal prior on μ_0. In particular, $\mu_0 \sim \text{Normal } (\mu_{00}, \tau_{00})$. Also, for simplicity, the precision τ_i is assumed to be fixed. The next theorem establishes the ordering between the posterior variances of μ_0 under both models.

Theorem 3.1. *For independent normal random variates, the DP model results in larger posterior variance of μ_0 as compared to that of the parametric model, that is, $Var_{DP}(\mu_0|y) \geq Var_N(\mu_0|y)$, where,* $y_i|\mu_i, \tau_i \overset{iid}{\sim} Normal (\mu_i, \; precision = \tau_i), \; i = 1, 2, \ldots, n.$

Proof. We start with obtaining the posterior precision of μ_0. Under the case of unequal precisions, the posterior precision of μ_0 based on the parametric model can be obtained as

$$prec_N(\mu_0|y) = \tau_{00} + \tau_\mu \sum_{i=1}^{n} \frac{\tau_i}{\tau_\mu + \tau_i}. \tag{4}$$

To derive the posterior precision under the DP model we first consider a clustering C_k consists of k clusters, where the kth cluster has size n_k. It is quite evident that within a specific cluster the observations share the same parameter value. However, these values are different across clusters. Given a clustering C_k, the DP mixture model with baseline distribution $Normal(\mu_0, \tau_\mu)$, can be viewed as a parametric model with k groups. Suppose, y_{ij} denotes the jth observation in the ith group, then we can write

$$prec_{DP}(\mu_0|y, C_k) = \tau_{00} + \tau_\mu \sum_{j=1}^{k} \frac{\sum_{l=1}^{n_j} \tau_{jl}}{\tau_\mu + \sum_{l=1}^{n_j} \tau_{jl}}. \tag{5}$$

We first adopt the notion of Schur Concave Function to verify $Var_N(\mu_0|y) \leq Var_{DP}(\mu_0|y, C_k)$, conditional on a clustering C_k. Let's define,

$$g(x) = \frac{x}{\tau_\mu + x}; \quad \text{g is concave for } x > 0, \text{and}$$

$$z = (\sum_{l=1}^{n_1} \tau_l, \sum_{l=1}^{n_2} \tau_l, ..., \sum_{l=1}^{n_k} \tau_l).$$

where,

$$x = (\tau_1, \tau_2, ..., \tau_n).$$

Then by definition $x \prec z$, or we can say z majorizes x, which implies, $\sum g(x_i) \geq \sum g(z_i)$. For more details, see Hardy et al. [7]. Now,

$$\sum g(\tau_i) = \sum_{i=1}^{n} \frac{\tau_i}{\tau_\mu + \tau_i} \quad ,$$

and

$$\sum g(\tau_i^{\star}) = \sum_{j=1}^{k} \frac{\sum_{l=1}^{n_j} \tau_{jl}}{\tau_\mu + \sum_{l=1}^{n_j} \tau_{jl}},$$

where $\tau_i^{\star} = \sum_{l=1}^{n_i} \tau_l$. Then, from Equations (4) and (5), conditional on the clustering C_k, we get $Var_N(\mu_0|y) \leq Var_{DP}(\mu_0|y, C_k)$. Now,

$$Var_{DP}(\mu_0|y) = E[Var_{DP}(\mu_0|y, C_k)] + Var[E_{DP}(\mu_0|y, C_k)],$$
$$\geq E[Var_N(\mu_0|y)] + Var[E_{DP}(\mu_0|y, C_k)],$$
$$\geq Var_N(\mu_0|y).$$

□

We conclude this section with the following Corollary, which is a special case of Theorem 3.1 assuming equal precision for y_i.

Corollary 3.1.1. *Under the case of equal precision, the posterior variance of μ_0 for the Dirichlet process (DP) model is always larger than that of the parametric model, that is, $Var_{DP}(\mu_0|y) \geq Var_N(\mu_0|y)$.*

3.2   Longitudinal Model

Let us consider $\boldsymbol{y}_i$ be a vector of n observations related to the ith subject, and $E[\boldsymbol{Y}_i|\boldsymbol{b}_i] = X\boldsymbol{b}_i$, where X is an $n \times p$ matrix of covariates. We assume the rows of X to be same for each subject. For example, we can consider a clinical trial in which the observations on each subject are recorded at the baseline, 4th week, 8th week and 12th week. Hence in this case, each row of X is a vector of length four. This model can be written as

$$\boldsymbol{y}_i|\boldsymbol{b}_i \sim \text{ Normal } (X\boldsymbol{b}_i, T),$$

where T is the precision matrix. Next, we can assign a normal prior on the coefficient vector $\boldsymbol{b}_i$ and call it as the parametric model. That is, $\boldsymbol{b}_i|\boldsymbol{\beta_0} \sim \text{ Normal } (\boldsymbol{\beta_0}, T_b)$. On the contrary, the semiparametric model based on DP prior can be constructed as follows,

$$\boldsymbol{b}_i|G_0, \kappa \sim DP(G_0, \kappa),$$
$$G_0 = \text{ Normal } (\boldsymbol{\beta_0}, T_b),$$

where under both models, the hyper parameter $\boldsymbol{\beta_0}$ is distributed according to $Normal(\boldsymbol{\beta_{00}}, T_{00})$.

Theorem 3.2. *The posterior variance of $\boldsymbol{\beta_0}$ under the Dirichlet process model is always larger than that of the parametric model, that is, $Var_{DP}(\boldsymbol{\beta_0}|y) \geq Var_N(\boldsymbol{\beta_0}|y)$, where, $\boldsymbol{y}_i|\boldsymbol{b}_i \sim \text{ Normal } (X\boldsymbol{b}_i, T)$.*

Proof. The posterior precisions of $\boldsymbol{\beta_0}$ under the parametric model with normal priors can be derived as:

$$prec_N(\boldsymbol{\beta_0}|\boldsymbol{y}) = T_{00} + nX'(T^{-1} + XT_b^{-1}X')^{-1}X,$$

Similarly, for a given clustering C_k, the semiparametric model based on DP prior results in the following posterior precision

$$prec_{DP}(\boldsymbol{\beta_0}|\boldsymbol{y}, C_k) = T_{00} + \sum_{j=1}^{k} X'(\frac{1}{n_j}T^{-1} + XT_b^{-1}X')^{-1}X.$$

Applying results from Matrix Algebra and after some easy but tedious calculations, we can show that $prec_N(\boldsymbol{\beta_0}|\boldsymbol{y}) \leq prec_{DP}(\boldsymbol{\beta_0}|\boldsymbol{y}, C_k)$ conditional on a clustering C_k. The rest of the proof follows on similar lines to that of Theorem 3.1.   □

3.3   Meta Analysis Model for Binary Response Data

Consider K randomized trials each with two arms. Let y_{ij} denote the number of binary events and n_{ij} represent the number of patients enrolled in the jth arm of the ith trial, where $j \in \{0, 1\}$ represents the control and treatment arms respectively. Then event counts y_{ij} assumes

$$y_{ij} \sim \text{Binomial}(n_{ij}, p_{ij}); \ j = 0, 1 \text{ and } i = 1, 2, \cdots, K.$$

We can define a pair of study specific random effects as, $\mu_i = logit(\pi_{i1}) - logit(\pi_{i0})$ and $\nu_i = 0.5 * (logit(\pi_{i1}) + logit(\pi_{i0}))$. Under the parametric model the random effect $\mu_i \sim Normal(\mu_0, \tau_\mu)$. Whereas, in case of semiparametric model μ_i is assumed to follow an unknown distribution G, which we later model through DP prior. The hierarchical representation of the semiparametric model can be obtained following Equation (3). Finally for both models $\nu_i \sim Normal(\nu_0, \tau_\nu)$. We complete the model specification by assigning appropriate prior distributions. We like to observe whether a similar ordering between posterior
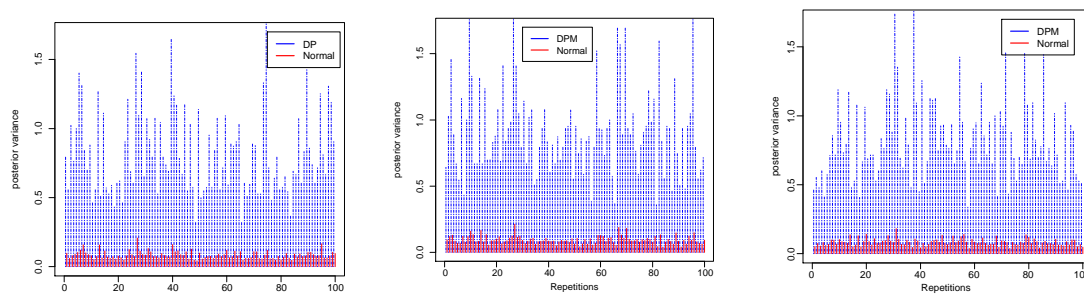
Figure 1. Posterior variance of μ_0, obtained from both models, for each of the 100 simulated datasets, generated under $\mu_0 = 0$ (extreme left), $\mu_0 = 2.5$ (middle), and $\mu_0 = -1.2$ (extreme right).

variances still exists under this setup. However, we are unable to find a closed form expression for the posterior variance of μ_0. Hence, we try to illustrate this through simulations. For the illustration purpose, we consider $\mu_0 = -1.2, 0$, and 2.5. Also, the study sizes n_{ij} are chosen to be 100 for both arms. For a given μ_0, we simulate 100 data sets, each consisting of 20 studies with binary count on both arms. Here, μ_i and μ_0 represent the study specific log-odds ratio and overall log-odds ratio.

Figure 1 gives the estimated posterior variance of μ_0 under both parametric and DP based semiparametric models. For every μ_0 values specified in the data generation, the DP model results in much larger posterior variance as compared to that of the parametric model. Such ordering in posterior variances of μ_0 is reflected in every simulated data.

4   Simulation

In this section we describe a simulation study designed to compare the performances of parametric and DP based semiparametric Bayesian models. The data were generated according to the normal location model defined in Section 3.1. We consider μ_0 to be the parameter of interest, which can also be considered as the overall mean parameter. These models are then compared with respect to the posterior standard deviation of μ_0, average length of it's credible intervals and the resulting power associated to the alternative $H_1 : \mu_0 > 0$.

We now describe the configuration of the simulation study in greater detail. The samples were generated based on a normal location model with mean μ_0 and standard deviation 2. The sample sizes were chosen to be $n = 20, 40, 60$ and 80. For the power comparison, we set $\mu_0 = \xi$, with ξ ranging from 0.2 to 0.8. For every ξ and sample size combination, we generated 1000 simulated samples. The Bayesian models were constructed by assigning a flat normal prior on μ_0 and flat gamma priors on both the precision parameters, τ and τ_μ. These models are then implemented through OpenBUGS. The power under each model can be calculated as follows,

1. Calculate the posterior probability

$$P(H_1|y) = \frac{1}{M} \sum_{m=1}^{M} [I(\mu_0 > 0)],$$

where M is the number of MCMC iteration after burn-in.
2. If $P(H_1|y)$ greater than a pre-defined threshold p^*, increase the COUNTS by 1; otherwise 0.
3. Finally the power can be obtained through the proportion COUNTS/B, where B denotes the number of simulated data.

The threshold p^* was considered to be 0.95. The results are depicted in Table 1. Both models performed an equally good job in estimating the overall mean μ_0. They both result in very little bias. However, in case of DP model, the estimate is less precise, as it presents a comparatively larger posterior variance than that of the parametric model. Such ordering between posterior variances holds for every ξ and sample

Table 1. Posterior summaries of simulated data based on parametric and DP based semiparametric Bayesian models.

ξ	n	Parametric Model			Semi parametric Model		
		Bias (SD)	ALCI	Power	Bias (SD)	ALCI	Power
0.2	20	-0.234 (0.47)	1.85	0.09	-0.233 (1.12)	4.64	0.003
	40	-0.180 (0.32)	1.26	0.18	-0.184 (0.93)	3.85	0.01
	60	-0.138 (0.27)	1.06	0.25	-0.147 (0.84)	3.43	0.05
	80	-0.177 (0.22)	0.88	0.27	-0.190 (0.80)	3.30	0.09
0.4	20	-0.042 (0.47)	1.85	0.18	-0.063 (0.97)	3.79	0.01
	40	0.014 (0.32)	1.27	0.36	-0.002 (0.62)	2.61	0.03
	60	0.013 (0.26)	1.02	0.43	-0.009 (0.49)	1.99	0.09
	80	0.023 (0.22)	0.88	0.56	0.019 (0.43)	1.74	0.12
0.6	20	0.027 (0.46)	1.84	0.41	0.021 (0.93)	3.81	0.02
	40	-0.030 (0.32)	1.27	0.54	-0.051 (0.64)	2.60	0.09
	60	0.009 (0.26)	1.02	0.75	-0.003 (0.49)	1.99	0.22
	80	0.006 (0.23)	0.89	0.84	-0.016 (0.44)	1.76	0.30
0.8	20	0.0006 (0.48)	1.88	0.43	-0.069 (0.96)	3.90	0.02
	40	-0.013 (0.32)	1.25	0.79	-0.042 (0.64)	2.59	0.23
	60	0.038 (0.26)	1.02	0.96	0.022 (0.49)	1.99	0.58
	80	0.002 (0.23)	0.89	0.96	-0.009 (0.43)	1.75	0.68

size combination. In particular, the posterior variances, reduce with increasing sample sizes while keeping the ordering intact. We also observe that, the average length of 95% credible interval (ALCI) based on DP model is much higher than the one obtained for parametric model. A wide confidence interval indicates that we are less sure about the estimation of the unknown parameter and perhaps a larger sample is needed to increase our confidence. For the power comparison we let ξ vary from 0.2 to 0.8 by the increments of 0.2 with varying sample sizes. For $\xi = 0.2$ both models result in smaller even for the case of $n = 80$. To detect small effect sizes with large power, we need to increase the sample size. In particular under the parametric model, when $\xi = 0.4$ the maximum power one can achieve is 0.56 with $n = 80$. However, it increased with both ξ and n and for $\xi = 0.8$, even with a sample of size $n = 40$, an 80% power can be achieved. But in case of the semiparametric model, the powers are consistently low with a maximum power of 0.68 with $\xi = 0.8$ and $n = 80$.

5   Illustrative Example

In this section, we use a data set consisting of 24 randomized controlled trials involving 4487 children and adolescents treated with antidepressants. This example is taken from Kaizar et al. [9]. We analyze this data using both parametric and semi-parametric models, as defined in Section 3.3. These models are implemented through OpenBUGS and R2OpenBUGS package of R. The results listed in Tables 2 and 3 are obtained based on 20,000 Markov chain iterations after a burn-in of 10,000.

Table 2. The posterior summaries for the overall log-odds ratio μ_0 obtained under both parametric and semi-parametric models

Model	Posterior mean	Posterior variance	95% Credible interval
Parametric	0.8184	0.0874	(0.2409 , 1.389)
Semi Parametric	0.8159	0.2787	(-0.1915 , 1.892)

Table 3. Posterior summaries of μ_0 and the estimated number of distinct cluster obtained from the DP model

value of α	avg. # of distinct clusters	Posterior mean	Posterior variance	95% Credible Interval
9	11	0.8276	0.1260	(0.1106 , 1.528)
4	8	0.8052	0.1575	(0.03279 , 1.569)
2	5	0.816	0.5856	(-0.7368 , 2.202)
1	3	0.7195	0.9850	(-1.5700 , 2.574)

Table 2, lists the posterior mean, variance and 95% credible interval of μ_0 under both parametric and semiparametric models. The resulting posterior means are quite similar under both models. However, DP based model results in larger posterior variance as well as wider credible interval as compared to the other. Moreover, according to parametric model, we may conclude that the suicidal ideation is more likely to occur in the treatment group as the credible interval falls on the right side of origin. But, we hesitate to make such comments for the second model, as the resulting 95% credible interval includes the origin.

Also, we fix α at 1, 2, 4, and 9 and fit the semiparametric model for each α. α, denotes the hyperparameter of Beta distribution, in Equation (2), which indirectly determines the number of distinct clusters generated through a DP. According to Table 3, as α decreases the DP model results in less number of distinct clusters, and as a result the 95% credible intervals become wider. However, the posterior estimates of μ_0 do not change that much. There is a logical explanation for this behavior. The parametric model with normally distributed random effects can be viewed as clustering where each study represents a unique cluster. Now, as the number of estimated clusters increase, we can roughly say that the DP model gradually approaches to a full parametric model, which explains the smaller posterior variance and shorter credible interval for large α.

6   Concluding Remarks

Dirichlet process mixture based semiparametric models have been widely cited as a robust and flexible alternative in Bayesian applications. However, it is a well known fact, that such models account for larger uncertainty. We have studied the posterior variance of mean of the random effects distribution assuming different models and proved that DP based semiparametric model always accounts for a higher posterior variability. However, the simulation study suggests that both models bring about very little bias for every choice of ξ and n. But, due to the larger posterior variability, the DP model generates a wider credible interval as compared to its parametric counterparts. That is, we get less precise estimates. This can have serious consequence in Bayesian inference. For example, as we have seen in Table 2, in case of DP model, the 95% credible interval includes the origin, which leads to an inconclusive result. Also, the DP model results in less power. Finally, this behavior of the DP-based semiparametric model can be described as an honest behavior. If the DP model is a better description of reality, then it should be used regardless of any effect it might have on the power. However, if it results in a reduction in power, then that is just the price of doing a good statistical analysis.

References

1. Blackwell, D., and MacQueen, J. B. (1973). Ferguson distribution Via Pólya urn scheme. *The Annals of Statistics*, **1**, 363–365.
2. Burr, D., and Doss, H. (2005). A Bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association*, **100**, 242–251.
3. Escobar, M. D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, **89**, 268–277.
4. Escobar, M. D., and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577–588.
5. Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.

6. Ghosh, P., Nathoo, F., Gönen, M., and Tiwari, R.C. (2011). Assessing noninferiority in a three-arm trial using the Bayesian approach. *Statistics in Medicine* **30**, 1795–1808.
7. Hardy, G. H., Littlewood J. E., and Pólya G. Inequalities. *Cambridge Mathematical Library.*
8. Ishwaran, H., and Zarepour, M. (2002). Dirichlet prior sieves in finite normal mixtures. *Statistica Sinica*, **12**, 941–963.
9. Kaizar, E. E., Greenhouse, J. B., Seltman, H., and Kelleher, K. (2006). Do antidepressants cause suicidality in children? A Bayesian meta-analysis. *Clinical Trials*, **3**, 73–98.
10. Linero, A., and Daniels, M. J. (2015) A flexible Bayesian approach to monotone missing data in Longitudinal studies with nonignorable missingness With Application to an acute schizophrenia clinical trial. *Journal of the American Statistical Association*, **110**.
11. Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **4**, 639–650.
12. Sethuraman, J., and Tiwari, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. *Statistical Decision Theory and Related Topics III,* eds. S. Gupta and J. O. Berger, New York: Springer-Verlag, pp. 305–315.
13. Zhao, L., Shi, J., Shearon, T.H., and Li, Y. (2016). A Dirichlet process mixture model for survival outcome data: assessing nationwide kidney transplant centers. *Statistics in Medicine*, **34**, 1404–1416.