

More Nonparametric Bayesian Inference in Applications

Michele Guindani · Wesley O. Johnson

the date of receipt and acceptance should be inserted later

1 Introduction

We are delighted to have the opportunity to discuss this paper. We are long time believers in the BNP approach to modeling statistical data. The authors have an extensive and distinguished record of accomplishments in this area and it is fitting that they would feature some of their work that displays the utility and outright advantage of the BNP method in a variety of complex clinically relevant settings, as they have done masterfully in the work displayed in this article.

We take this opportunity to augment the discussion of the authors' work by mentioning some of our own; part of which also includes the authors of this article. The authors discussed novel applications to survival analysis. We mention the work of De Iorio et al (2009), that uses a DP mixture of log normal distributions in order to provide a semi-parametric survival model that allows survival curves to cross, thus avoiding the assumption of proportional hazards. The papers Hanson and Johnson (2002, 2004), Hanson et al (2009), Hanson et al (2011) constitute a body of work that embeds parametric survival families of distributions into broader non-parametric families using Mixtures of Finite Polya Trees (MFPT). The models discussed in these papers allow for considerable flexibility compared with their parametric counterparts. An additional theme involves consideration of several alternative semi-parametric families, for example, they model baseline survival distributions using MFPTs for proportional hazards, accelerated failure time, proportional odds and Cox and Oakes models. Some of the work focusses on fixed time dependent covariates, and other work develops joint models for survival and longitudinal processes that are related

M. Guindani
Department of Statistics
University of California, Irvine
E-mail: mguindan@uci.edu

Wesley O. Johnson
Department of Statistics
University of California, Irvine
E-mail: wjohnson@ics.uci.edu

to survival. Competing models are compared using the LPML statistic (Geisser and Eddy, 1979) in order to select the one with the greatest predictive ability.

Another related theme that may be of interest involves the development of BNP methodology for receiver operating characteristic (ROC) curve estimation. Branscum et al (2008, 2015) used MFPTs to model biomarker distributions for individuals known to have a specified condition/disease, and for individuals known to not have the condition. They also developed identifiable semi-parametric regression models that are similar to the survival models discussed above with the purpose of generalizing parametric methods to semi-parametric methods for assessing the quality of biomarkers. We discuss another biomarker assessment problem in more detail below. We also note that the work mentioned above and more is discussed in the survey article by Johnson and de Carvalho (2015).

Bayesian Nonparametric methods have been also employed in large-scale multiple hypotheses testing, and for selecting relevant predictors in a regression. We review some recent contributions, namely spike-and-slab DP processes for variable selection, mixtures of DP processes for large-scale screening of differential genes, and discovery test statistics which approximate optimal decision rules.

The authors provide an extensive illustration of the Bayesian Nonparametric literature in the analysis of spatial data. Spatio-temporal data arising from brain imaging studies have received increased interest recently. These data are particularly challenging, since they are high-dimensional, highly noisy and heterogeneous across subjects. We discuss some application of Bayesian Nonparametric methods to this type of data.

Finally, the authors point out that in many Bayesian Nonparametric models the main target of inference is a partition of the n samples into more homogeneous subsets. Typically, such random partitions are exchangeable. However, natural dependencies in the data may go against the exchangeability assumption. We review a recent class of models that defines non-exchangeable partitions, and its application to the analysis of array comparative genomic hybridization (CGH) data and the detection of copy number aberrations.

2 Factors Affecting and Clustering of Hormone Curves for Women in Menopause

Quintana et al (2016) developed a novel statistical model that generalizes standard mixed models for longitudinal data and which allows for flexible mean functions in addition to combined compound symmetry (CS) and autoregressive (AR) covariance structures. AR structure was specified using a Gaussian process (GP) with an exponential covariance function. This structure was extended to a Dirichlet Process Mixture (DPM) over the covariance parameters of the GP, which allows the possibility to estimate a variety of covariance structures. They illustrated that models that fail to incorporate CS or AR structure can result in very poor estimation of a covariance or correlation matrix.

Quintana et al (2016) analyzed a subset of patients from the Study of Women's Health Across the Nation (SWAN) with 9 yearly responses during the menopausal transition on 162 women. They focused on the hormone follicle stimulating hormone

(FSH) serum concentration profiles with the goal of assessing the effect of Age at entry (≤ 46 , > 46) and Ethnicity (African American, Caucasian, Chinese, Japanese) on profile shape. Time 0 corresponds to the final menstrual period.

Sample profiles are shown in Quintana et al (2016) and they display considerable variability with no regular profile shape by individual. However, empirical data and biology suggest that, on average, these profiles start out relatively flat, then increase to a new level, and then flatten.

The Quintana et al (2016) model generalizes the Zeger and Diggle (1994) model, which is itself a generalization of the Laird and Ware (1982) linear mixed model (LMM). The Zeger and Diggle (1994) model is:

$$y_i(t) = \mu(t) + f_i(t) + x_i(t)\beta + z_i(t)v_i + w_i(t) + \varepsilon_i(t), \quad (1)$$

with $w_i(t)$ representing an Ornstein-Uhlenbeck (Gaussian) process (OUP). There are many variations of Model (1) e.g. using various basis functions for the overall mean function $\mu(t)$ and the random deviations from it $f_i(t)$, modeling the distribution of the random effects vector v_i using a DP or DPM (see Li et al, 2010) and/or using a variety of covariance structures for the OUP. Quintana et al (2016) extended this model by taking a DPM of OUPs in order to generalize the correlation structure from AR to Toeplitz. The primary goal of such models is to account for heterogeneity across individuals, account for longitudinal correlation structure and extend mixed models to allow for a more flexible correlation structure.

Since it was believed that sigmoid structure for the means was appropriate, the authors considered a 5 parameter generalization of sigmoid functions. Figure 1 shows predictive curves for the SWAN data using 4 models, with the solid curves corresponding to the Quintana et al (2016) mixture of OUP models. The LPML statistic was used to select among 6 candidate models, which included a parametric version with simple random effects and no OUP, a DDP mixture on the v s with no OUP, and other variants of Model (1). Results from the Quintana et al (2016) analysis are displayed in Figure 1, where it can be seen that the basic shapes of the curves are sigmoidal and increasing until the end of the time window when they decline. Some of the curves are noticeably different from one another; in fact, statistically different. For example, the posterior probability that the maximum curve value achieved for younger Japanese women is greater than the maximum for Chinese women in the same age category is 0.9994. In addition, posterior probabilities that the timing of the maximum for younger Chinese women would be greater than that for African Americans, Caucasians and Japanese are 0.987, 0.9998 and 0.999, respectively. The estimate for Chinese women is on the order of 2 years greater. Moreover, the estimated correlations between responses for women that are 1-8 years apart were: (0.43, 0.27, 0.21, 0.17, 0.15, 0.14, 0.14, 0.13), indicating a clear departure from AR structure.

A different subset of SWAN hormone data was performed by He (2014). He modeled log Estradiol (E2) profiles for 11 years of data on 928 women. Using a DPM of orthogonal (Legendre) polynomials (levels ≤ 4). The purpose of this analysis was to find clusters of women who had different shaped profiles. Figure 2 shows three clusters with distinct shapes.

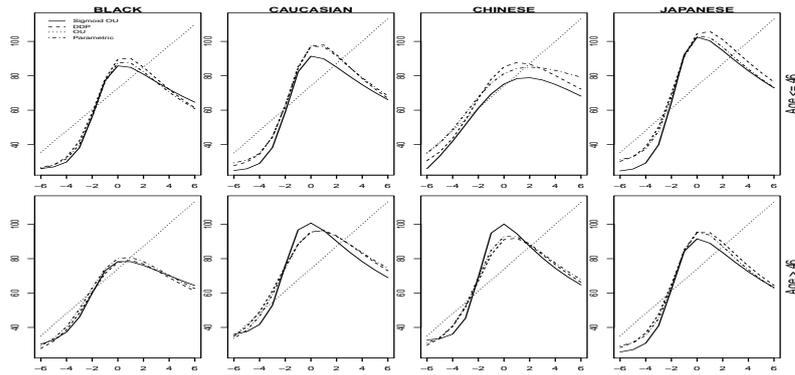


Fig. 1 Predictive FSH profiles using DPM of OUP model, and other models

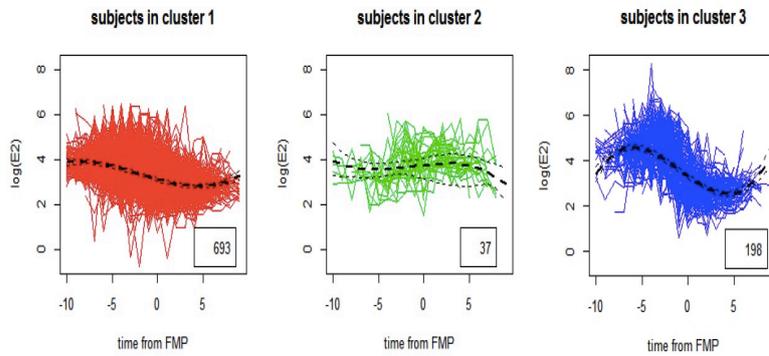


Fig. 2 Clustering of $\log(E_2)$ curves using DPM of orthogonal polynomials

3 Estimating the Quality of a Biomarker for Johnes Disease in Cattle

Diagnostic testing involves an assessment of whether or not a particular condition is present. A typical goal is to assess the quality of one or more biomarkers for the condition. With a single continuous biomarker, a cutoff is set so that outcomes larger than the cutoff are classified as having the condition, and values below are classified as free of it. The cutoff is selected to strike a balance between the false positive and false negative rates.

Let $D+$ denote that the condition of interest is present and let $T+$ denote that the outcome of a diagnostic test is positive in the sense that a continuous biomarker exceeded a selected cutoff, or a categorical outcome indicated that the condition was present. Similarly define $D-$ and $T-$. Denote the sensitivity of the test to be $Se = Pr(T+ | D+)$, which is one minus the false positive rate or the true positive rate, and the specificity of the test to be $Sp = Pr(T- | D-)$, which is the true negative rate. Acceptable diagnostic tests have $Se + Sp > 1$. HIV tests for example are highly

accurate with Se and Sp greater than 0.99. In animal testing, it is often the case that the Se is somewhat low, while the Sp is quite high, near one, thus leading to many false negatives but few false positives.

The sensitivity of a diagnostic test generally depends on how long the individual being tested has had the condition. For example, it is impossible to detect HIV immediately after the infection has occurred; testing is not performed until there has been sufficient time for a detectable antibody response. Since most statistical assessments of the test accuracy are performed based on cross-sectional data, the estimated Se and Sp are necessarily dependent on the distribution of times of acquisition of the condition in the population sampled.

This brings us to the current study involving Johnes Disease (JD) in cattle. JD is caused by infection with bacterium *Mycobacterium avium* subspp. *paratuberculosis* (Map), the agent of association with biomarkers designed to react to its presence. Norris et al (2014) analyzed a longitudinal data set consisting of two diagnostic outcomes on 365 cows. Cows were tested on average every 6 months over several years for the presence of MAP using a continuous serologic (antibody detection) outcome, and a dichotomous (organism detection) outcome. The two biomarkers are serology (S) and Fecal Culture (FC).

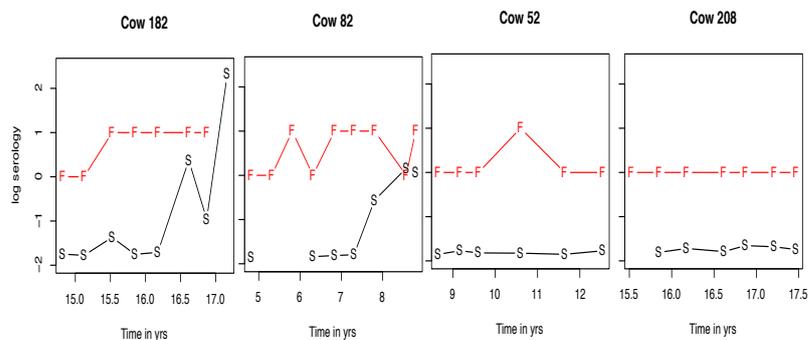


Fig. 3 Fecal culture and Serology profiles for four cows

Data on several cows are depicted in Figure 3. The FC test appears to detect the organism in cow 182 around age 15.5 years, while the serologic response to the infection is delayed for about a year. The FC test appears to detect the organism in cow 82 around age 6, but that test is followed by a possible false negative and then another positive. The serologic response appears after a delay of more than one year from the initial FC+ outcome. The third and fourth plots indicate animals that are not infected over the time frame considered, but with one probably false FC+ outcome.

The statistical model for the data involves conditionally independent Bernoulli($\theta(t)$) outcomes for FC where for all t less than the time of infection, $\theta(t) = 1 - Sp$, and after infection, $\theta(t) = Se$. Serology is modeled in three parts involving times: (i) before infection, (ii) after infection if infection occurred within the lag time just before the last observation on that cow (in which case there is no time for a serologic response),

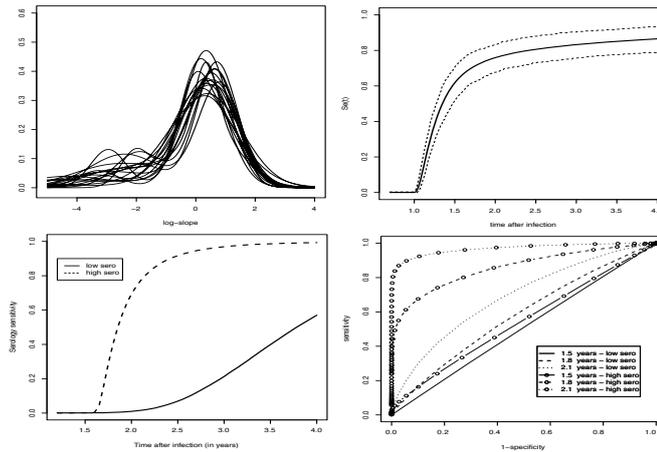


Fig. 4 Upper left: Iterates of slope distribution. Upper right: Estimated $Se(t)$ with probability intervals. Lower left: Cluster based estimates of $Se(t)$. Lower right: Cluster by time since infection estimates of ROC curves

and (iii) after infection if infection occurred before the last time of observation minus an unknown lag time (in which case there is time for there to be a serologic response). The model for S before infection is a simple mixed effects model that allows for correlation between repeated observations on the same cow. The model for S in the second situation is the same as the first, and the model for S in the third situation involves modeling an unknown change point when the cow became infected, and adding a positive random slope in time for each cow, after the infection plus lag time. The Norris et al (2014) analysis implemented reversible jump methodology due to the differing model dimensions of these cases.

The parametric version of the model anticipates that cows will have differing slopes. However, biology suggests that there may be two or more groups of cows, each with similar rates of serologic response. Consequently, Norris et al (2014) modeled the random slopes with a DPM of log-normal distributions. Figure 4 (Upper left) shows a plot of a number of iterates of the slope distribution from the Norris et al (2014) analysis, where we see two different types of slope iterate: one that is bimodal with a steeper slope mode and a more gradual slope mode, and the other that is unimodal. The posterior probability of 1 mode for the slope distribution was 0.62, and for 2 modes was 0.30, indicating a moderately strong case for the possibility of two or more groups of cows that we might care to distinguish.

Figure 4 (Upper right) shows a plot of the primary inference of interest, the posterior estimates and 95% pointwise probability intervals for $Se(t)$, the sensitivity of a test as a function of time based on S using a cutoff of -1.29 (data are on log scale). Figure 4 (Lower left) shows estimated $Se(t)$ for two clusters identified with rapid and slower serologic responses. Figure 4 (Lower right) shows estimated ROC curves for the two clusters categorized by times 1.5, 1.8 and 2.1 years after the lag. Obviously it

is much easier to detect MAP for the group that has the more rapid serologic response and after longer times since infection plus lag.

4 Multiple Hypothesis Testing and Variable Selection.

Kim et al (2009) have proposed a Bayesian method for multiple hypothesis testing based on the use of spiked distributions for Bayesian variable selection. We exemplify their proposal with reference to a single population, although their framework applies more generally to a collection of populations. Let us consider the linear model $Y = X\beta$, with β a $p \times 1$ parameter vector. In variable selection, we consider a sequence of hypotheses $H_{0i} : \beta_i = 0, i = 1, \dots, p$. Kim et al (2009) propose to model the regression coefficients as:

$$\beta_1, \dots, \beta_p | G \sim G, \quad G \sim DP(\alpha_\beta, G_\beta^*)$$

with

$$G_\beta^*(\cdot) = \pi \delta_0(\cdot) + (1 - \pi) G_0(\cdot),$$

where π is a mixing weight with prior $\pi \sim p(\pi)$. The mixture $G_\beta^*(\cdot)$ is a ‘‘spiked’’ mixture of a point mass at 0 (the ‘‘spike’’) and a continuous distribution with large support, $G_0(\cdot)$. These spiked centering priors accommodate sharp null hypotheses and allow for the estimation of the posterior probabilities of each hypothesis. Increased power is obtained by borrowing information across hypotheses through the use of Dirichlet process mixture models.

Do et al (2005) have discussed a nonparametric Bayesian model for multiple hypotheses testing and applied it to the screening of differential genes. Here, the reference framework is the two groups model developed by Efron (2004). For simplicity, we assume that test statistics z_i are used to assess if gene i is differentially expressed or not, $i = 1, \dots, n$. More precisely, the z_i ’s are assumed as independent samples from a mixture of two distributions

$$z_i \sim \pi f_0 + (1 - \pi) f_1,$$

where f_0 is the unknown distribution for the non-differentially expressed genes and f_1 is the unknown distribution of the differentially expressed genes. The unknown distributions $f_j, j = 0, 1$ are then characterized as DPM models. Guindani et al (2014) extend this framework to compare DPM models from samples collected across different conditions, in the analysis of T-cell sequence abundances with a Poisson likelihood.

Shahbaba and Johnson (2013) similarly propose a latent random partition model based on Dirichlet process mixtures (DPM) as an exploratory tool for data analysis in large scale inference problems. Variables of interest (say, genes) are ranked according to the magnitude of posterior cluster variances, with a threshold to divide genes into relevant and not relevant groups. The method can be viewed in the context of variable selection where a very large number of covariates could be potentially included in the model, but where there is a belief in sparsity, which translates to parsimony.

Assuming a Bayesian decision theoretic framework, the multiple comparison problem can also be characterized by a set of actions (decisions) and a loss function for all possible outcomes of an experiment. Let $d_i \in \{0, 1\}$ denote the decision for the i -th hypothesis, with $d_i = 1$ indicating a decision against H_{0i} , and let $d = (d_1, \dots, d_n)$. The optimal rule $d_i^*(z)$ is defined by minimizing the loss function $L(d, \theta)$ with respect to the posterior $p(\theta | z)$. Müller et al (2004) and Müller et al (2007) discuss the optimal decision rule corresponding to loss functions defined as linear combinations of the false negatives and false positive counts, say $L = FN + \lambda FP$, for some constant $\lambda > 0$. The optimal rule is a threshold on the marginal posterior probability of the alternative hypothesis, $v_i = P(H_{1i}|z)$, i.e. $d_i^* = I(v_i > t)$.

Guindani et al (2009) consider a Dirichlet Process Mixture of normals model and describe a Bayesian discovery procedure for large scale multiple testing of hypotheses on the means μ_i 's, $H_{0i} : \mu_i \in A$ vs $H_{1i} : \mu_i \in A^c$. The Bayesian testing procedure is obtained by approximating the marginal posterior probabilities, v_i , using the properties of the conditional posterior distribution $p(G | z)$. More specifically, for large n , the posterior $p(G | \mu, z)$ can be approximated by a degenerate distribution at $F_n = \frac{1}{n} \sum \delta_{\hat{\mu}_i}$, where the $\hat{\mu}_i$'s are centroids of clusters estimated when fitting the Bayesian nonparametric model. Hence, v_i can be approximated by

$$v_i \approx \frac{\int_{A^c} f(z; \mu) dF_m(\mu)}{\int f(z; \mu) dF_m(\mu)} = \frac{\sum_{\hat{\mu}_j \in A^c} f(z_i; \hat{\mu}_j)}{\sum_{j=1}^m f(z_i; \hat{\mu}_j)}. \quad (2)$$

The Bayesian Nonparametric model borrows strength across comparisons by means of the multiple shrinkage induced by the DP clustering, thus improving the power of the testing procedure.

Multiple testing issues arise also in the context of spatial data. For example, in geostatistical applications, one may be interested in isolating regions where the process has values above a given threshold. Guindani et al (2009) describe how the spatial DP model of Gelfand et al (2005) could be used together with a loss function that penalizes isolated discoveries. However, properties of Bayesian nonparametric models in spatial testing have not been thoroughly explored, especially for clusterwise inference, and in a compound decision theoretic framework to control the proportion of false discoveries. See Sun et al (2015) for a discussion of the latter set up.

5 Applications to Brain Imaging Data

Bayesian nonparametric techniques have been widely employed to capture heterogeneity in brain structures as well as brain functions. Jbabdi et al (2009) use a hierarchical mixture of DPs to segment brain regions based on tractography data in multiple-subjects. More recently, Durante et al (2017) proposed a Bayesian nonparametric approach for the estimation of the distribution of brain connectivity structures from white matter tractography data in a population of subjects.

Functional magnetic resonance imaging (fMRI) is a noninvasive neuroimaging method that provides an indirect measure of neuronal activity by detecting blood flow changes over the course of an experiment. fMRI data provide an accurate spatial mapping of brain responses. Furthermore, the sequence of whole-brain scans, which

has been acquired over the duration of the experiment, enables to explore the temporal dynamics of brain functioning. In an fMRI experiment, it is often of interest to study the patterns of activation in response to a stimulus and the interactions between brain regions, both within a single subject and across groups of subjects (say, healthy controls and cases). Zhang et al (2014) describe an analytical framework that allows detection of regions of the brain in response to a stimulus by using variable selection spike-and-slab mixture priors and a Markov random field (MRF) prior to account for the complex spatial correlation structure of the brain. In order to infer association of the voxel time courses, they assume temporally-correlated long memory errors and achieve clustering of the voxels by imposing a DP prior on the parameters of the long memory process. The clustering of fMRI time series captures the so-called functional connectivity among the brain regions (Friston, 2011).

In a multi-subject approach, Zhang et al (2016) employ a hierarchical DP prior to induce clustering among voxels within and across subjects in the analysis of fMRI time series. The hierarchical DP captures spatial correlation among potential activations of distant voxels, within a subject, while simultaneously borrowing strength in the estimation of the parameters from subjects with similar activation patterns. Since a single fMRI experiment can yield hundreds of thousands of high frequency time series for each subject, there is a need to devise efficient computational algorithms for posterior inference. Zhang et al (2016) show that a variational Bayes implementation of the BNP model achieves robust estimation results at reduced computational cost.

As a further example of the potential role of BNP in the analysis of imaging data, Li et al (2015) discuss a scalar-on-image regression to identify imaging biomarkers for predicting individual biological or behavioral traits. More specifically, they propose a joint Ising and DP prior for selecting brain voxels. The Ising component incorporates existing structural spatial information of brain region contiguities, whereas the DP component clusters the regression coefficients to reduce the computational burden of posterior sampling.

In their review, Müller, Quintana and Page have provided a comprehensive discussion of flexible BNP models for the analysis of spatial data. The application of those methods to the analysis of brain imaging data can open new avenues of applied research in the field. For example, product partition models dependent on covariates could be used to determine patterns of the brain activations varying across group of individuals. However, the main challenge will be to develop fast computational algorithms in order to ensure the scalability of the methods to the dimensions typical of voxel-based brain data.

6 Non-Exchangeable Partitions.

Airoldi et al (2014) consider array CGH data, which involve copy number gains or losses over several genomic regions. Genomic abnormalities are more likely to occur and persist over neighboring regions. Thus, for detecting regions of the DNA with copy number amplifications and deletions, one needs to take into account the spatial dependency of genomic aberrations. Du et al (2010) have proposed a sticky Hierarchical DP-HMM (Fox et al, 2011; Teh et al, 2006) to infer the number of states in

an HMM, while also imposing state persistence to capture the persistence of aberrations. Airolidi et al (2014) follow a different approach, by explicitly considering non-exchangeable random partition models. The starting point is the representation of the Pólya urn prior (eq (9) in the paper), $p_{DP}(s|\boldsymbol{\alpha})$, as a species sampling prior. In this representation, the DPM model is characterized as:

$$y_i|\mu_i \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, n,$$

with

$$\mu_{i+1}|\mu_1, \mu_2, \dots, \mu_i \sim \sum_{j=1}^i q_{i,j} \delta_{\mu_j}(\cdot) + q_{i,i+1} G^*, \quad (3)$$

where $\delta_x(\cdot)$ denotes a point mass at x , and $q_{n,i} = \frac{1}{\alpha+i}$ $q_{i,i+1} = \frac{\alpha}{\alpha+i}$. The sequence (3) implicitly defines the (exchangeable) random partition $\{s_1, \dots, s_n\}$ associated to the DP prior, with $s_i = k$ if and only if $\mu_i = \mu_k^*$. The predictive rule (3) can be generalized to take into account more complex types of dependence in the data, by defining the weights in terms of a sequence of independent (not necessarily identically distributed) latent random variables. In particular, Basseti et al (2010) have introduced a class of generalized species sampling sequences, which are not exchangeable, but only *conditionally identically distributed* (CID, Berti et al, 2004). That is, $\mu_{i+1}, \mu_{i+2}, \dots$ are identically distributed conditionally on the values of the process before observation i (i.e., given $\mu_i, \mu_{i-1}, \dots, \mu_1$, for all $i = 1, \dots, n$). The μ_i 's are marginally identically distributed, $\mu_i \sim G^*$, similarly as for the exchangeable DPM model. For the analysis of array CGH data, Airolidi et al (2014) consider a CID process where the weights of the species sampling sequence (3) are obtained as the product of independent latent variables, $W_j \sim \text{Beta}(\alpha_j, \beta_j)$,

$$q_{i,j} = (1 - W_j) \prod_{r=1}^j W_r, \quad \text{and} \quad q_{i,i+1} = \prod_{r=1}^i W_r.$$

The choice of Beta latent variables allows for a flexible specification of the species sampling weights, while still retaining simplicity and interpretability of the sequence allocation scheme. Figure 5 shows the fit to array CGH data for a single chromosome for two samples of breast tumors (Chin et al, 2006). Note how contiguous clones tend to be clustered together, in a pattern typical of these chromosomal aberrations.

Non-exchangeable partitions provide a flexible way to take into account complex dependencies in the data. Fortini et al (2016) have recently introduced a notion of *partially conditionally identically distributed* sequences. Partial CID sequences generalize the notion of partial exchangeability, which characterize the use of hierarchical models in Bayesian statistics to borrow information across related experiments. Dependent random partitions could then be defined through interacting reinforced-urn processes. Muliere et al (2006), Hu and Zhang (2004) are among those who proposed early applications of this type of dependent random partition schemes to the design of clinical trials. For a recent overview, see Flournoy et al (2012).

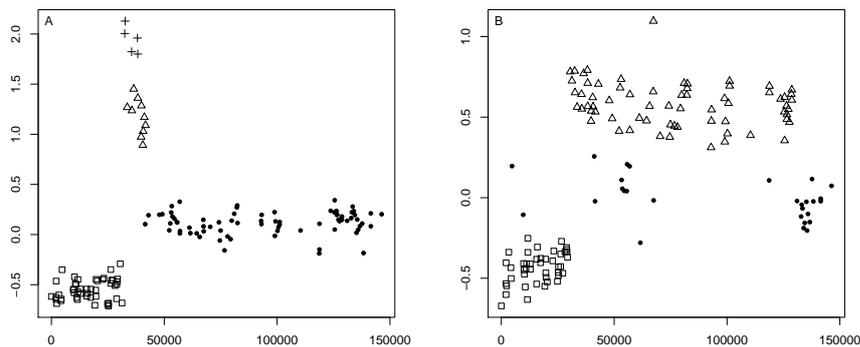


Fig. 5 Model fit overview: Array CGH gains and losses on chromosome 8 for two samples of breast tumors in the dataset in Chin et al (2006). Points with different shapes denote different clusters.

References

1. Airolidi E, Costa T, Bassetti F, Guindani M, Leisen F (2014) Generalized species sampling priors with latent beta reinforcements. *Journal of the American Statistical Association* 109(508):1466–1480
2. Bassetti F, Crimaldi I, Leisen F (2010) Conditionally identically distributed species sampling sequences. *Advances in Applied Probability* 42(2):433–459
3. Berti P, Pratelli L, Rigo P (2004) Limit theorems for a class of identically distributed random variables. *The Annals of Probability* 32(3):2029–2052
4. Branscum AJ, Johnson WO, Hanson TE, Gardner IA (2008) Bayesian semi-parametric ROC curve estimation and disease diagnosis. *Statistics in Medicine* 27(13):2474–2496
5. Branscum AJ, Johnson WO, Hanson TE, Baron AT (2015) Flexible regression models for ROC and risk analysis, with or without a gold standard. *Statistics in Medicine* 34(30):3997–4015
6. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10(6):529 – 541
7. De Iorio M, Johnson WO, Müller P, Rosner GL (2009) Bayesian nonparametric non-proportional hazards survival modelling. *Biometrics* 65(3):762–771
8. Do K, Müller P, Tang F (2005) A Bayesian mixture model for differential gene expression. *Journal of the Royal Statistical Society, Series C* 54(3):627–644
9. Du L, Chen M, Lucas J, Carlin L (2010) Sticky hidden Markov modelling of comparative genomic hybridization. *IEEE Transactions on signal processing* 58(10):5353–5368

10. Durante D, Dunson DB, Vogelstein JT (2017) Nonparametric Bayes modeling of populations of networks. *Journal of the American Statistical Association* In press
11. Efron B (2004) Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association* 99(465):96–104
12. Flournoy N, May C, Secchi P (2012) Asymptotically optimal response-adaptive designs for allocating the best treatment: An overview. *International Statistical Review* 80(2):293–305
13. Fortini S, Petrone S, Sporysheva P (2016) On a notion of partially conditionally identically distributed sequences. Tech. rep., Arxiv.org
14. Fox E, Sudderth E, Jordan M, Willsky A (2011) A sticky HDP-HMM with application to speaker diarization. *Ann Appl Stat* 5(2A):1020–1056
15. Friston KJ (2011) Functional and effective connectivity: a review. *Brain Connectivity* 1(1):13–36
16. Geisser S, Eddy WF (1979) A predictive approach to model selection. *Journal of the American Statistical Association* 74(365):153–160
17. Gelfand A, Kottas A, MacEachern S (2005) Bayesian nonparametric spatial modeling with Dirichlet processes mixing. *J Am Statist Ass* 100:1021–1035
18. Guindani M, Müller P, Zhang S (2009) A Bayesian discovery procedure. *Journal of the Royal Statistical Society, Series B* 71(5):905–925
19. Guindani M, Sepúlveda N, Paulino CD, Müller P (2014) A Bayesian semiparametric approach for the differential analysis of sequence counts data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 63(3):385–404
20. Hanson T, Johnson WO (2002) Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* 97(460):1020–1033
21. Hanson T, Johnson WO (2004) A bayesian semiparametric AFT model for interval-censored data. *Journal of Computational and Graphical Statistics* 13(2):341–361
22. Hanson T, Johnson W, Laud P (2009) Semiparametric inference for survival models with step process covariates. *Canadian Journal of Statistics* 37(1):60–79
23. Hanson TE, Branscum AJ, Johnson WO (2011) Bayesian semi-parametric methods for joint modeling of event time and longitudinal data, with discussion. *Lifetime Data Analysis* 17(3-18)
24. He Y (2014) Bayesian cluster analysis with longitudinal data. PhD thesis, Department of Statistics. University of California - Irvine
25. Hu F, Zhang LX (2004) Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann Statist* 32(1):268–301
26. Jbabdi S, Woolrich M, Behrens T (2009) Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. *NeuroImage* 44(2):373 – 384
27. Johnson W, de Carvalho M (2015) Bayesian nonparametric biostatistics. In: Mitra R, Müller P (eds) *Nonparametric Bayesian Methods in Biostatistics and Bioinformatics.*, Springer, New York, pp 15–53
28. Kim S, Dahl DB, Vannucci M (2009) Spiked Dirichlet process prior for Bayesian multiple hypothesis testing in random effects models. *Bayesian Anal* 4(4):707–732
29. Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38(4):963–974

30. Li F, Zhang T, Wang Q, Gonzalez M, Maresh E, Coan J (2015) Spatial Bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *The Annals of Applied Statistics* 9(12):687–713
31. Li Y, Lin X, Müller P (2010) Bayesian inference in semiparametric mixed models for longitudinal data. *Biometrics* 66(1):70–78
32. Muliere P, Paganoni AM, Secchi P (2006) A randomly reinforced urn. *Journal of Statistical Planning and Inference* 136(6):1853 – 1874
33. Müller P, Parmigiani G, Robert CP, Rousseau J (2004) Optimal sample size for multiple testing: the case of gene expression microarrays. *J Am Statist Assoc* 99:990–1001
34. Müller P, Parmigiani G, Rice K (2007) FDR and Bayesian multiple comparisons rules. In: Bernardo J, Bayarri M, Berger J, Dawid A, Heckerman, D A Smith, West M (eds) *Bayesian Statistics 8*, Oxford, UK: Oxford University Press.
35. Norris M, Johnson W, Gardner I (2014) A semiparametric model for bivariate longitudinal diagnostic outcome data in the absence of a gold standard. *Statistics and its interface* 7:417–438
36. Quintana FA, Johnson WO, Waetjen LE, Gold EB (2016) Bayesian nonparametric longitudinal data analysis. *Journal of the American Statistical Association* 111(515):1168–1181
37. Shahbaba B, Johnson WO (2013) Bayesian nonparametric variable selection as an exploratory tool for discovering differentially expressed genes. *Statistics in Medicine* 32(12):2114–2126
38. Sun W, Reich BJ, Tony Cai T, Guindani M, Schwartzman A (2015) False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77(1):59–83
39. Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical Dirichlet processes. *Journal of the American Statistical Association* 101(476):1566–1581
40. Zeger SL, Diggle PJ (1994) Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* 50(3):689–699
41. Zhang L, Guindani M, Versace F, Vannucci M (2014) A spatio-temporal nonparametric Bayesian variable selection model of fMRI data for clustering correlated time courses. *NeuroImage* 95:162–175
42. Zhang L, Guindani M, Versace F, Engelmann JM, Vannucci M (2016) A spatiotemporal nonparametric bayesian model of multi-subject fMRI data. *Ann Appl Stat* 10(2):638–666