

# Spatial Mixture Modelling for Unobserved Point Processes: Examples in Immunofluorescence Histology

Chunlin Ji\*, Daniel Merl†, Thomas B. Kepler‡ and Mike West§

**Abstract.** We discuss Bayesian modelling and computational methods in analysis of indirectly observed spatial point processes. The context involves noisy measurements on an underlying point process that provide indirect and noisy data on locations of point outcomes. We are interested in problems in which the spatial intensity function may be highly heterogenous, and so is modelled via flexible nonparametric Bayesian mixture models. Analysis aims to estimate the underlying intensity function and the abundance of realized but unobserved points. Our motivating applications involve immunological studies of multiple fluorescent intensity images in sections of lymphatic tissue where the point processes represent geographical configurations of cells. We are interested in estimating intensity functions and cell abundance for each of a series of such data sets to facilitate comparisons of outcomes at different times and with respect to differing experimental conditions. The analysis is heavily computational, utilizing recently introduced MCMC approaches for spatial point process mixtures and extending them to the broader new context here of unobserved outcomes. Further, our example applications are problems in which the individual objects of interest are not simply points, but rather small groups of pixels; this implies a need to work at an aggregate pixel region level and we develop the resulting novel methodology for this. Two examples with immunofluorescence histology data demonstrate the models and computational methodology.

**Keywords:** Bayesian computation, blocked Gibbs sampler, Dirichlet process mixture model, inhomogeneous Poisson process, unobserved point process

## 1 Introduction

Parametric and nonparametric approaches to spatial point process modelling have been well-studied in recent years (Diggle 2003; Moller and Waagepetersen 2004), with increased use of mixtures and convolutional methods for modelling heterogeneity in intensity functions (Wolpert and Ickstadt 1998). Recently Kottas and Sanso (2007) proposed the use of the Dirichlet process as a random mixing distribution for mixture-based methods. The full computational machinery of nonparametric Bayesian models has thus been brought to bear on this class of inference problems for point processes. Tradition-

---

\*Department of Statistical Science, Duke University, Durham, NC , <mailto:cj27@stat.duke.edu>

†Department of Statistical Science, Duke University, Durham, NC , <mailto:dan@stat.duke.edu>

‡Department of Immunology, Duke University Medical Center, Durham NC <mailto:kepler@duke.edu>

§Department of Statistical Science, Duke University, Durham, NC , <mailto:mw@stat.duke.edu>

ally, all such methods assume perfect knowledge of the outcome of the point process. However, in situations such as that described below, the outcome of the spatial point process cannot be observed directly, but is measured by some imperfect proxy. In such situations, inference on the underlying intensity function will depend on accounting for the uncertainty surrounding the outcome of the point process, the latter including the number of realized points as well as their spatial locations. Our work here defines models that address these issues and develops computational Bayesian methods for model fitting and analysis.

Our motivating applications are immunological studies of multiple fluorescent intensity images of lymphatic tissue. Observed measurements are fluorescent intensities generated from tagged cell-surface proteins; this generates indirect, noisy observation of cell locations for as many as tens of thousands of cells in the context of background noise. The spatial configurations of cells across the 2 or 3- $d$  tissue region is typically hugely heterogenous, so requiring flexible models for underlying intensities. In any one experiment (of many) a series of images may reflect cellular distributions at different times and/or as a response to different interventions and treatments. For each, we aim to characterize the underlying intensity functions and overall level of abundance of cell types in order to facilitate comparisons across multiple images.

Section 2 describes our nonparametric mixture models of intensity functions, measurement error modelling for a single unobserved point process, and the need to deal with problems in which individual, spatially distributed objects of interest are not simply points, but rather small groups of pixels. Section 3 describes MCMC methods and extensions to our framework of unobserved, pixel region level outcomes. Section 4 develops examples with image data from the immunological context, and we conclude with summary comments in Section 5.

## 2 Latent Spatial Mixture Models

Our general statistical framework jointly models the intensity function of a spatially inhomogeneous Poisson process and the uncertain outcome of the point process. Modelling of the intensity function is similar to that of [Kottas and Sanso \(2007\)](#), but here relying on a Dirichlet process mixture model of multivariate normal densities (rather than beta densities). Incidences of the point process are modelled via a modification of the basic model to represent data on a pixelated grid across image space, and this couples with a generalized linear model for linking noisy measurements (e.g. fluorescence levels, available at the gridded level) to incidences of the point process (e.g. presence of cells).

### 2.1 Basic Spatial Point Process Model

A spatial point process over a finite region  $S \subset \mathbb{R}^d$  (here,  $d = 2$ ) generates realizations  $x_{1:N} = \{x_1, \dots, x_N\}$  of  $N \geq 0$  points  $x_i \in S$ . We regard  $x_{1:N}$  as the outcome of an inhomogeneous Poisson process with intensity function  $\lambda(x) \geq 0$  ( $x \in S$ ), integrable over

$S$ . That is: (a) for any region  $s \subseteq S$ , the number of points  $n(s) = \#\{i = 1 : N \mid x_i \in s\}$  is Poisson with mean  $\Lambda(s) = \int_{x \in s} \lambda(x) dx$ ; and (b) conditional on  $\lambda(\cdot)$ ,  $n(s) \perp\!\!\!\perp n(r)$  for any disjoint subsets  $s, r \subset S$  (Daley and Vere-Jones 2003; Diggle 2003).

Bayesian analysis of *observed* data  $x_{1:N}$  arising from a spatial inhomogeneous Poisson process requires first specifying a prior probability model for the intensity function  $\lambda(\cdot)$ , and then conducting posterior inference on  $\lambda(\cdot)$  in light of the realized outcomes  $x_{1:N}$ . As in Kottas and Sanso (2007), define the overall intensity scale parameter  $\gamma = \int_{x \in S} \lambda(x) dx$  and the probability density (over  $x \in S$ )  $f(x) = \lambda(x)/\gamma$ . Then the likelihood function resulting from observed data  $x_{1:N}$  can be expressed as

$$p(x_{1:N} | \gamma, f) \propto \exp(-\gamma) \gamma^N \prod_{i=1}^N f(x_i) \quad (1)$$

as a function of  $(\gamma, f)$ . The degree to which underlying spatial heterogeneity can be represented in  $\lambda(\cdot)$  is therefore linked to the modelling assumptions surrounding  $f(\cdot)$ .

## 2.2 Dirichlet Process Mixture Models

To provide flexibility in characterizing spatial heterogeneity in the intensity function we employ the Dirichlet process mixture framework in which the normalized intensity function  $f(x)$  is the density of a random mixture of  $d$ -dimensional normal distributions. This follows Kottas and Sanso (2007) who develop models using mixtures of betas rather than normals. Since we are working on problems with very heterogeneous intensity functions in 2 and 3- $d$ , and with sample sizes  $N$  that (though unknown) are large, we very much need the flexibility offered by mixtures of multivariate normals coupled with their relative analytic and computational benefits; we simply truncate and ignore the form of fitted and simulated models outside the finite region  $S$ .

A key observation of Kottas and Sanso (2007) was to note that the likelihood function of equation (1) depends on  $f(\cdot)$  only through the term  $\prod_{i=1}^N f(x_i)$  and is *precisely* the likelihood that would arise from simple random sampling from  $f(\cdot)$  generating data  $x_{1:N}$ . Thus, for computational purposes, we can then use the standard methods of posterior computation based on any assumed model for  $f(\cdot)$ . Use of Dirichlet process mixtures is one example.

In brief,  $f(\cdot)$  is taken as the density of a distribution arising from the following hierarchical model for independent,  $d$ -dimensional variates  $x_i$ , each with its own parameter  $\theta_i = (\mu_i, \Sigma_i)$ , a mean vector and variance matrix, respectively. Then the model for  $f(\cdot)$  is

$$(x_i | \theta_i) \sim N(x_i | \mu_i, \Sigma_i), \quad (\theta_i | G) \sim G, \quad (G | \alpha, G_0) \sim DP(\alpha, G_0) \quad (2)$$

using standard notation. Here  $G(\cdot)$  is an uncertain distribution function,  $G_0(\cdot)$  is the prior mean of  $G(\cdot)$  and  $\alpha > 0$  the total mass, or precision of the DP. For conditional conjugacy, it is convenient and common to take the prior as normal-inverse Wishart. The implied distribution corresponding to the density  $f(x)$  is a discrete mixture of a countably infinite number of normals. The model notation and structural details are

standard and used widely in applied Bayesian inference; key foundational modelling and computational aspects are available in, for example, MacEachern (1994), West et al. (1994), Escobar and West (1995, 1998), MacEachern (1998) and MacEachern and Mueller (1998) and in a broader context in the more recent review paper of Mueller and Quintana (2004). Details that are key to computation are noted below in Section 3.2 and the Appendix.

### 2.3 Discrete Pixel Region Model

In the immunological application as in other studies in spatial modelling, the data arises in terms of images of the region  $S$  within which the individual objects of interest are not simply points, but rather small groups of pixels. This, coupled with the fact that the objects (here, cells) are in any case not directly observed, implies a need to work at an aggregate pixel region level. This can be quite general but, for purposes here, we focus on rectangular pixel regions; in the immunological imaging study, for example, we work at the level of  $3 \times 3$  pixel regions (in  $2-d$ ) and each region is either occupied by a cell, or not.

Generally, in  $d$ -dimensions suppose the overall imaged region  $S = [-s, s]^d$ , for some  $s > 0$ , and that the level of resolution is  $a$  pixels in each dimension.  $S$  is then a rectangular grid of  $a^d$  pixel regions; label these by interior points  $x_i^r$ , ( $i = 1 : a^d$ ), and set  $\mathcal{X} = \{x_i^r : i = 1 : a^d\}$ . Assuming  $a$  to be large and with  $\Delta = (2s/a)^d$ , we have approximate intensity  $\Delta\gamma f(x_i^r)$  for pixel region  $i$  and  $\sum_{x \in \mathcal{X}} \Delta f(x) \approx \int_{x \in S} f(x) dx = 1$ .

Now, for any realization of the point process, each pixel region will be either occupied by an object or not. Define binary variates  $y(x) = 1/0$  to represent presence/absence of an object (e.g., a cell) in the pixel region with index point  $x \in \mathcal{X}$ . Then observing the occurrence of objects at a subset of  $N$  regions is equivalent to observing binary data  $y(x)$  for all  $x \in \mathcal{X}$  with  $y(x) = 1$  at just the  $N$  regions with objects. Suppose the  $N$  regions are indexed by  $x_{1:N} \in \mathcal{X}$ , and write  $Y$  for the full set of  $a^d$  binaries. It then follows that the likelihood of equation (1) is equivalent to

$$p(Y|\gamma, f) \propto \prod_{x \in \mathcal{X}} \{\Delta\gamma f(x)\}^{y(x)} \exp\{-\gamma\Delta f(x)\} \quad (3)$$

and

$$p(Y|\gamma, f) \propto \exp(-\gamma)\Delta^N \gamma^N \prod_{i=1}^N f(x_i). \quad (4)$$

This provides the ability to work at the discretized, pixel region level appropriately, and simply modifies the likelihood with the additional  $\Delta$  term. Most importantly also, this discrete pixel region version also enables easy development for contexts in which the locations  $x_{1:N}$  are not observed directly but are measured with noise, since equation (3) delivers a likelihood function for uncertain locations and number of objects  $x_{1:N}$  along with  $(\gamma, f)$ .

## 2.4 Unobserved Spatial Inhomogeneous Poisson Process

Consider now contexts in which the locations  $x_i$ , and their number  $N$ , are uncertain. The example of fluorescent intensity images of lymphatic tissue is a key motivation and raises broader modelling questions. There, the specific locations of biological cells are not observed, but reflected in terms of fluorescence generated from labelled cell surface proteins. Under the discrete pixel region formulation, we can incorporate uncertainty about  $x_{1:N}$  using equation (3), as follows.

Suppose we have observations  $z(x)$  at each location  $x \in \mathcal{X}$  generated by the measurement process. That is, the measurements represent single pixel region locations with no overlap or interaction. It is practicable to assume that the measurement error distribution depends on  $x$  only through presence or absence of objects, i.e., on the  $y(x)$  binary indicators, and will usually involve uncertain parameters here denoted by  $\delta$ . That is, a measurement error model is defined by two density functions  $p(z|x, \delta) = p(z|y(x), \delta)$  for  $y(x) = 1/0$ , where  $p(z|y = 0, \delta)$  represents background noise in the absence of an object at a specific location, and  $p(z|y = 1, \delta)$  represents noise in the presence of a signal.

We can now combine  $p(Y|\gamma, f)$  of equation (3) as the prior for all  $y(x)$  with the implied likelihood components  $p(z(x)|y(x), \delta)$ , based on recorded data  $Z = \{z(x) : x \in \mathcal{X}\}$ . In terms of posterior odds on  $y(x) = 1$  versus  $y(x) = 0$ , this yields conditionally independent posteriors with

$$\text{Odds}(y(x) = 1|Z, \gamma, f) = r(z(x))\gamma\Delta f(x) \quad (5)$$

where  $r(z(x)) = p(z(x)|y(x) = 1, \delta)/p(z(x)|y(x) = 0, \delta)$  for all  $x \in \mathcal{X}$ .

In the immunological imaging study, appropriate noise models are truncated normals and the posterior odds ratios are trivially evaluated. This is important as we can then embed imputation of  $Y$  in the overall MCMC computations.

## 3 Posterior Inference And Sampling Strategies

The overall posterior inference goals are to explore and summarize aspects of the implied joint posterior for all uncertain quantities based on a complete model specification that now includes independent priors on  $\gamma, \alpha$  and any of the hyperparameters  $\delta$  we may wish to treat as uncertain. In summary, this is the posterior  $p(Y, f, \gamma, \alpha, \delta | Z)$ .

### 3.1 Overall MCMC Framework

The MCMC computational algorithm visits the following components in turn. Each of the imputation steps here draws new variates from the conditional distribution given all other conditioning quantities. In each, only those conditioning quantities that matter are included in the notation.

### Sampling the normalized intensity function $f(x)$ , its parameters and $\alpha$

Each MCMC iterate generates a realized density that is a mixture of a finite number of  $d$ -dimensional normals,  $f(x) \equiv f(x|\Theta) = \sum_{j=1}^k w_j \mathcal{N}(x|\mu_j^*, \Sigma_j^*)$ , with parameters  $\Theta = \{w_{1:k}, \mu_{1:k}^*, \Sigma_{1:k}^*\}$  changing at each MCMC step, being generated via a two-step process discussed in Section 3.2 and the Appendix below. This step also resamples the Dirichlet precision  $\alpha$ . Given  $\Theta$ , the density  $f(\cdot|\Theta)$  can be evaluated at the finite set of points  $x \in \mathcal{X}$  for further use.

At each iterate, the  $\Theta$  parameters are drawn from an implicit conditional  $p(\Theta|N, x_{1:N})$  where the number of imaged objects,  $N$ , and their location indices  $x_{1:N}$ , are set at current values. As the MCMC progresses these values are resampled as the analysis explores the joint posterior that now also includes uncertainty about  $(N, x_{1:N})$ .

### Sampling spatial object location indicators $Y$

Equation (5) leads to resampling of new values of  $Y$  as independent binaries  $y(x)$  at each  $x \in \mathcal{X}$ , based on implied probabilities  $Pr(y(x) = 1|z(x), \gamma, f, \delta)$ . This generates a complete set of binaries from which those values  $y(x) = 1$  identify the new sample size  $N$  and pixel region locations  $x_{1:N}$ . Notice that one by-product is samples from the posterior for  $N$ , i.e., the ability to make inferences about the uncertain number of underlying objects as well as their locations. This step explicitly requires the evaluation  $f(x) \equiv f(x|\Theta)$ , the mixture of normals based on the most recently sampled  $\Theta$ .

### Sampling the overall scale of intensity $\gamma$

The form of equation (4) makes it clear that a gamma prior is conjugate to the conditional likelihood, leading to a gamma distribution  $p(\gamma|N)$ .

### Sampling hyperparameters $\delta$ of the measurement error model

Under a prior  $p(\delta)$ , these parameters may be generated using some form of Gibbs or Metropolis-Hastings component strategy based on the implied conditional

$$p(\delta|Z, Y) \propto p(\delta) \prod_{x \in \mathcal{X}} p(z(x)|y(x), \delta).$$

For example, normal measurement errors would involve normal mean and variance parameters in  $\delta$ , one pair for each of the signal and noise error models; in such a case, conditionally conjugate priors would aid in this computational step. In our immunological studies appropriate error models are truncated normals, which introduces a need for Metropolis-Hastings for sampling  $\delta$  as described in 4.2 and the examples.

### 3.2 Simulation in DP Mixtures

Under the DP model of equation (2),  $G$  is discrete. This results in any realized set of  $N$  parameters  $\theta_{1:N} = (\mu_{1:N}, \Sigma_{1:N})$  being configured into some  $k \leq N$  *distinct* values  $(\mu_{1:k}^*, \Sigma_{1:k}^*)$ . The DP generates configuration indicators  $c_{1:N}$  such that  $c_i = j$  indicates  $(\mu_i, \Sigma_i) = (\mu_j^*, \Sigma_j^*)$ . The original MCMC approaches to generating posterior samples in DP mixtures (MacEachern 1994; West et al. 1994; Escobar and West 1995, 1998; MacEachern 1998; MacEachern and Mueller 1998) utilize this theory to generate samples from the full joint posterior of  $k$ ,  $(\mu_{1:k}^*, \Sigma_{1:k}^*)$  and  $c_{1:N}$ . Most effective among these approaches are the *collapsed* or *configuration* samplers for DP mixture models originating from MacEachern (1994). More recent approaches are based on the innovative strategy using the *blocked Gibbs sampler* (Ishwaran and James 2001) that explicitly includes simulation from approximations to the conditional posteriors for the underlying mixing distribution  $G(\cdot)$  itself.

In many problems with small or moderate sample sizes  $N$ , and when  $f(\cdot)$  is well-behaved to the extent that it may be well-approximated by a small mixture of normals, there is little to choose between the configuration and blocked samplers in terms of either computational or statistical efficiencies. However, as  $N$  increases, and also with densities  $f(\cdot)$  of greater complexity that therefore require larger numbers  $k$  of mixture components for adequate representation, the blocked sampler dominates. Configuration sampling iteratively resamples each configuration indicator conditional on the rest; this one-at-a-time update degrades computational efficiency as  $N$  increase, and difficulties in moving in configuration space induced by the tight conditioning degrade mixing of the MCMC, and hence statistical efficiency. In contrast, the blocked sampling strategy breaks this configuration conditioning at each iterate, resampling the full set of configuration indicators jointly.

In our immunological applications,  $N$  is in the thousands or tens of thousands, and intensity surfaces can be very heterogeneous, so the blocked sampling strategy is really demanded for efficiency reasons. In fact, the approach is almost mandated in the context of measurement error; as we have seen, values of the normalized intensity  $f(x)$  itself are key components of the overall analysis, arising in the conditional posteriors for the latent spatial object location indicators  $Y$  in equation (5). To evaluate values of the density  $f(\cdot)$  requires inference on the underlying mixing distribution  $G(\cdot)$  itself, and this is provided by the blocked sampling strategy. Kottas and Sanso (2007) use this strategy, pointing out that it is needed to generate posterior inferences on aspects of  $f(\cdot)$  in any case; our new framework with latent spatial process outcomes, large  $N$  and heterogeneous intensity patterns very strongly reinforces this choice.

The block sampler involves three linked steps: sampling of the set of configuration indicators  $c_{1:N}$ , sampling of parameters that define an approximation to the mixing distribution  $G(\cdot)$ , and sampling of sets of normal model means and variance matrices. A key element is the truncated approximation to the so-called stick-breaking representation of  $G$  (Sethuraman 1994) that effectively defines a finite mixture model with a specified upper bound  $k$  on the number of components (Ishwaran and James 2001). Importantly, then, this approach actually introduces a theoretical approximation to the

full DP mixture model through this truncation. The practical relevance of the truncation is limited, however, particularly when dealing with problems with large numbers of components. Moreover, the resulting truncated version can in any case be viewed as a directly specified alternative model in its own right, rather than as an approximation to the DP mixture. The MCMC strategy we use follows that of [Kottas and Sanso \(2007\)](#), with some changes in detail related to the resampling steps for parameters, and is briefly outlined in the Appendix here.

## 4 Immunofluorescence Histology Image Analysis

### 4.1 Context

The motivating application for this work arises in immunological studies in mice where multiple images provide data on the spatial configuration of many immune cells in a specific, localized region of lymphatic or spleen tissues. A single experiment views an image as the response to stimulus via injection of a vaccine, the overall context being exploration of responses under candidate vaccine designs. Comparisons involve replicate images from different mice – possibly at different times and under differing treatments – with careful matching and registration of the tissue region across mice. Observed measurements are fluorescent intensities generated from tagged cell-surface proteins that characterize a specific cell type. The pixel region model adopts a very small,  $3 \times 3$  region of pixels as the level of resolution for modelling; this is small enough to be consistent with each region  $x$  being either occupied by a single cell ( $y(x) = 1$ ) or being unoccupied. Interest lies in characterizing the spatial intensity functions underlying observed data in each image, and feeding the statistical summaries and characterizations into visual and numerical comparisons. For the current paper, we simply explore aspects of the analyses of two example images, focussing on statistical aspects.

### 4.2 Measurement Error Models

Based on exploration of past data and experimentation with different measurement error models, a simple truncated lognormal model for the measurement of the fluorescence intensity appears to be adequate. That is, if  $z = z(y(x))$  represents measured fluorescence at pixel region  $x$ , then  $p(z|y, \delta)$  is defined by

$$(\log(z)|y, \delta) \sim N(m_y, v_y)I(\log(z) < h), \quad (y = 0, 1),$$

where  $(m_0, v_0)$  relates to the background noise and  $(m_1, v_1)$  to the distribution of signal fluorescence – i.e., the distribution conditional on the pixel region being occupied by a cell. Then  $\delta = (m_0, m_1, v_0, v_1)$ . Here  $h = \log(255)$ , the truncation being inherent to the digital fluorescent image generation process;  $\log(z)$  is recorded on a scale of  $(0, h)$  with values greater than  $h$  being truncated.

As described in Section 3.1, the overall MCMC can be extended to include a component representing uncertainty about  $\delta$ . Under the truncated normal model for  $(z|\delta)$ , this

can be done with standard priors, although it raises a need for a Metropolis strategy due to the truncation effects. We adopt independent normal-inverse gamma priors; for each of  $y = 0, 1$  independently,  $(m_y|v_y) \sim N(m_y|\bar{m}_y, tv_y)$  and  $v_y^{-1} \sim Ga(c/2, c\bar{v}_y/2)$  based on specified prior estimates  $\bar{m}_y, \bar{v}_y$  of  $m_y, v_y$ , respectively. These values are chosen to reflect known scales of (log) fluorescence and background. Specified values of the hyperparameters  $t, c$  are used to define relatively precise priors while allowing for adaptation to the data in each new image data analysis. In the examples here,  $\bar{m}_0 = 3.5, \bar{m}_1 = 4.5, t = 10, \bar{v}_0 = 0.1, \bar{v}_1 = 0.1$  and  $c = 10$ . We explore aspects of model fit and adequacy in the following discussion.

Conditional on the current  $Y$  and data  $Z$ , the  $(m_0, v_1)$  and  $(m_1, v_1)$  are conditionally independent. Each of the two conditional posteriors is the product of two terms: updated inverse-gamma based on the full set of  $N(\log(z)|m_y, v_y)$  non-truncated terms in the conditional likelihood functions, and the term involving products of cumulative normal distribution function values derived from truncation. The first terms provide suitable Metropolis proposals for the full conditional posteriors, with the contribution to the conditional likelihoods from the truncation normalization providing the terms in the acceptance ratio. Due to the fact that only a relatively small fraction of the data is truncated in these images (almost none, generally, for background levels, and in the region of up to 15% for signal) with much of the data lying well below the limit, these constructed proposal distributions are typically close to the target conditional posterior and the resulting sampler is very effective.

### 4.3 Image Data B220

Figure 1a is the original image of the intensity of emission from the fluorescent dye AF350 conjugated to an antibody that binds to B220, a molecule expressed on the surface of B lymphocytes. Lymph nodes were excised and sectioned 24 hours after subcutaneous alum injection. Figure 1b shows the corresponding heat map of intensity levels, which after logging provide the raw numeric data  $Z$ ; histograms of the fluorescence and logged values  $Z$  are shown in Figure 2. The image indicates typical heterogeneity of spatial distributions of a very large number of cells. On the resolution analyzed, we have  $180 \times 180$  pixel regions, each  $3 \times 3$  pixels representing the locations of individual cells (if present). We take the image region  $S$  as  $[-5, 5]^2$  so  $\Delta = 1/324$ .

We use priors as follows. First  $\alpha \sim Ga(1, 1)$  and  $\gamma \sim Ga(1, 0.001)$  for the two scalar model parameters. The base prior  $G_0(\mu, \Sigma)$  is  $N(\mu|0, t_0\Sigma)IW(\Sigma|s_0, S_0)$  where  $t_0 > 0, s_0 > 0$  is the prior degree-of-freedom and  $E(\Sigma) = S_0/(s_0 - 2)$  when  $s_0 > 2$ . Analysis here adopts  $t_0 = 50, s_0 = 2$  and  $S_0 = 0.4I$ , based on the specified scale of the image region  $S = [-5, 5]^2$  and the expectation of needing a large number of widely dispersed and relatively concentrated normal components to represent a very heterogeneous intensity surface. Further, the truncation of the mixture model uses  $k = 250$  as the upper bound on the number of components.

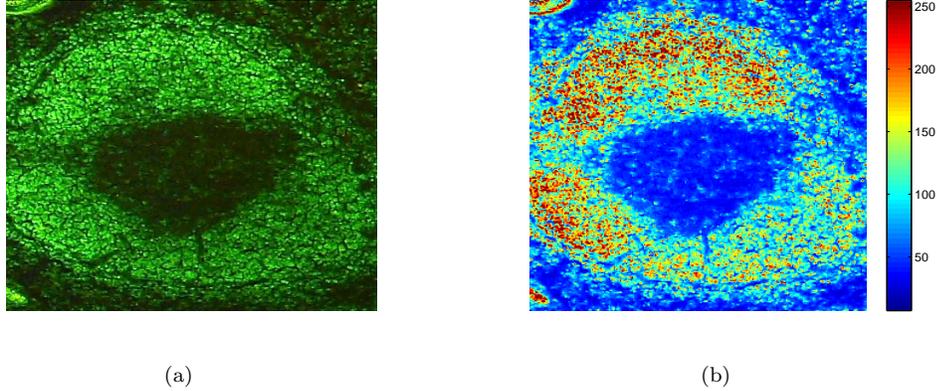


Figure 1: Data image on cell experiment B220 on day 1; (a) shows an image of the original data with fluorescent green tag, and (b) shows the scale of the corresponding intensity data.

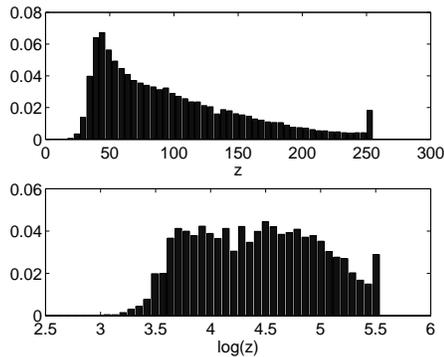


Figure 2: Data from B220, day 1: intensities (upper) and log intensities  $Z$  (lower).

the pixel regions. Evidently, the raw data displayed in the lower frame of Figure 2 shows the signal/noise structure, but without conditioning on signal/noise assignments it is difficult to develop direct graphical or numerical assessments of the normality assumption; repeat exploration across a series of MCMC samples aids measurably in this exploratory model assessment exercise.

Additional snapshots of one of the MCMC iterates are graphed as follows. Figure 4 shows an image of one posterior sample of  $Pr(y(x) = 1|Z)$  over all  $x \in \mathcal{X}$ , and Figure 5 (upper frame) identifies the corresponding current sampled mixture components overlaid on the data.

Some aspects of the analysis with a final 5000 MCMC iterations after burn-in are graphically summarized in the figures. Additional visual confirmation of the relevance of the truncated normal measurement error model is illustrated in Figure 3, showing normal qqplots of the  $Z$  data partitioned into noise and signal samples based on the current indicators  $Y$  at one randomly chosen step on the MCMC analysis. Repeat draws show similar forms. Looking at these graphs for a series of MCMC steps is useful in confirming the stability of the apparent adequacy of the truncated normal model, as reflected in the qqplots across multiple realizations of the signal/noise allocation of

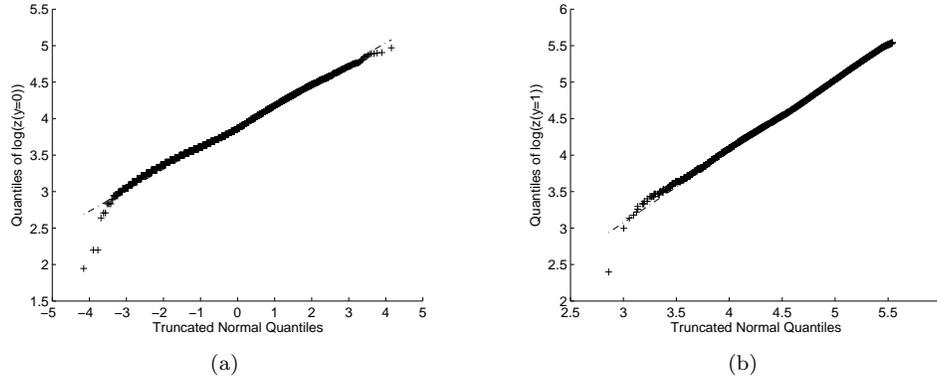


Figure 3: Based on a single, randomly selected draw from the MCMC in analysis of B220, day 1, the data are partitioned into noise ( $y = 0$ ) and signal ( $y = 1$ ) and the two corresponding samples of log intensities  $Z$  are displayed as normal qqplots: (a) noise, and (b) signal. This provides useful, visual insight into the utility and relevance of the truncated noise measurement error model and represents a nice dissection of the full data in the histogram of Figure 2.

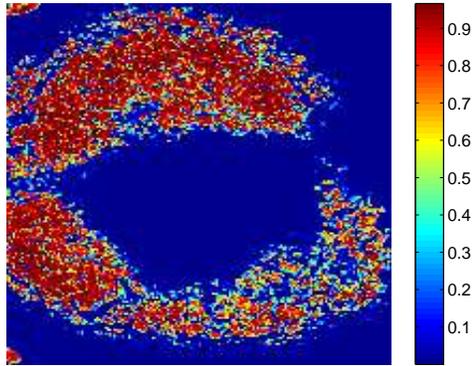


Figure 4: Image plot of  $Pr(y(x) = 1|Z)$  at one randomly chosen step of the MCMC in analysis of B220, day 1.

over the MCMC steps. In essence and up to a constant, this also represents the Monte Carlo posterior mean of the probabilities in Figure 4. Comparisons with Figure 1 begin to indicate the ability of the model to reflect the complexity of the data configuration, and with large numbers of heterogeneous mixture components to adapt to very variable patterns in the underlying spatial intensity.

The supplemental material on the web site contains a series of these “snapshot” figures in a movie, which gives a nice overview of the uncertainty across MCMC steps. Some regions are more stable/less variable than others, and this comes through best by viewing a series of samples through the MCMC. Viewing such figures aids in understanding and illustrating aspects of the model, and comparison of such MCMC snapshots with the real data in Figure 1 is illuminating. In terms of posterior estimates, averaging over MCMC iterates produces relevant summaries. For example, Figure 5 (lower frame) shows an image of the posterior mean intensity estimate based on averaging Monte Carlo samples  $f(x|\Theta)$

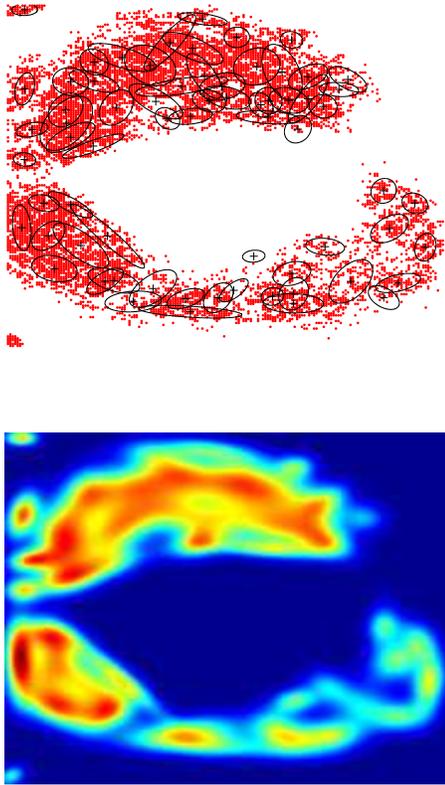


Figure 5: *Upper frame*: Scatter plot of the current sampled locations of cells  $Y$  at one MCMC step in analysis of B220 on day 1, overlaid with contours representing the location, scale and shape of the corresponding posterior sample of the normal mixture components underlying the intensity function. The contours are drawn at one standard deviation from the means in each of the major and minor axes directions. *Lower frame*: Image plot of the posterior estimate of the normalized intensity function  $f(x)$  in analysis of B220 on day 1, based on averages of the sampled surfaces over MCMC steps.

Monitoring of various parameters aids in assessment of convergence, again at the usual informal level. Rapid stabilization of trajectories of key single parameters, including  $\alpha$ ,  $\gamma$ ,  $N$  and  $\delta$  among others, is typically observed, and that is exemplified in Figures 6, 7 and 8. Good mixing is evident in these and other marginal trajectory plots.

From the applied perspective, the samples for  $N$  provide summary approximate posterior inferences on the underlying numbers of occupied pixel regions, i.e. our proxy for the number of cells, as one characteristic of this data set. Figure 8b indicates that  $N$  very likely lies in the range 8,500 – 8,700 for B220 on day 1, corresponding to 26-28% coverage of the image.

For a brief visual comparison, a second image of data illustrates the ability of the model analysis to reflect a diversity of patterns of complexity in image intensities. This comes from data captured from tissue in the same experiment, using the same dye-antibody combination, but now after an additional 10 days. Comparisons between selected posterior summaries between day 1 and day 11 clearly indicate that the distribution of the fluorescently labelled cells has changed significantly between day 1 and day 11. For example, the number of the fluorescent labelled cells has apparently also reduced significantly in the later stages; Figure 9 suggests  $N$  likely lies in the range 8,700 – 9,200 for B220 on day 11, so that the coverage of the overall image region is slightly increased relative to day 1. Further, there are multiple regions with higher intensities on day 1 that dissipate later on, and the overall intensity becomes fragmented; see the images for day 11 in Figure 10 in comparison to those at day 1.

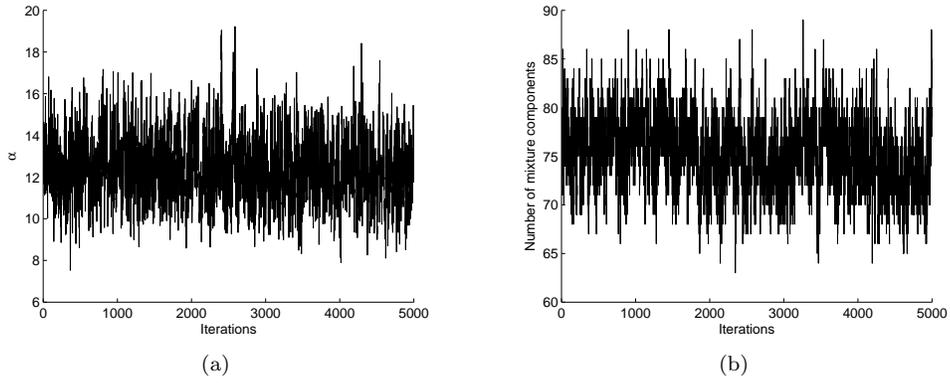


Figure 6: MCMC outputs in analysis of B220, day 1: Trajectories of (a) the Dirichlet process precision parameter  $\alpha$ , and (b) the number of realized, non-empty components in the mixture model.

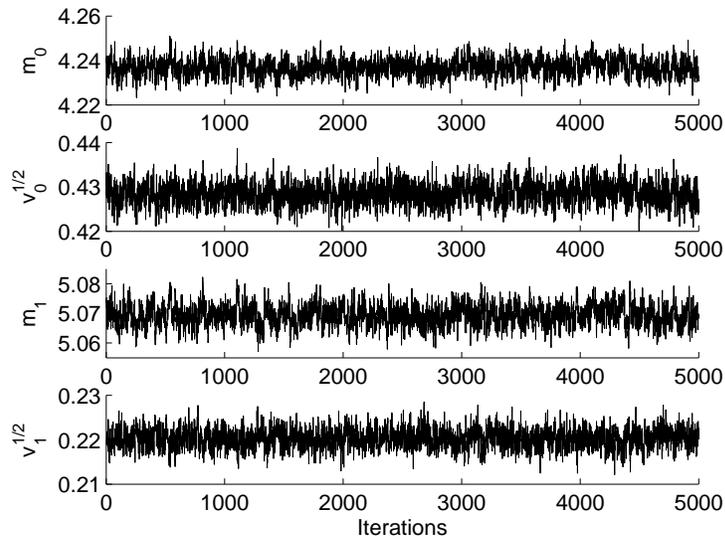


Figure 7: Trajectories of MCMC samples of measurement error model parameters in analysis of day 1 data.

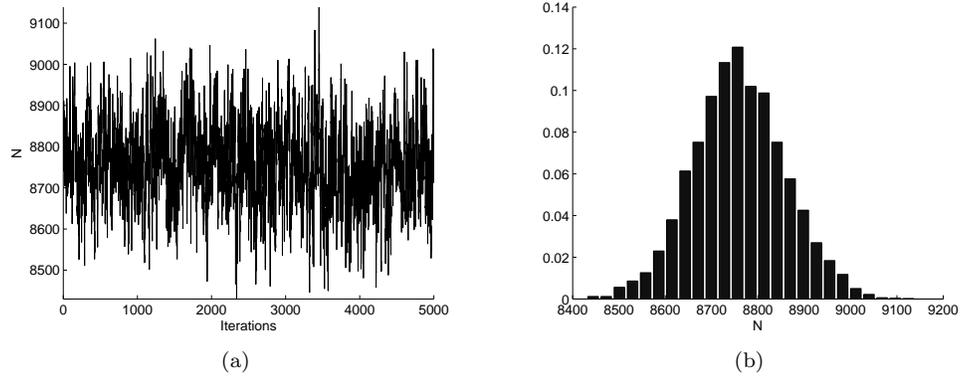


Figure 8: Experiment B220, day 1: Plots to show the number of “cells” in the image. (a) Trajectory of sampled  $N$  values in the MCMC, (b) the resulting histogram.

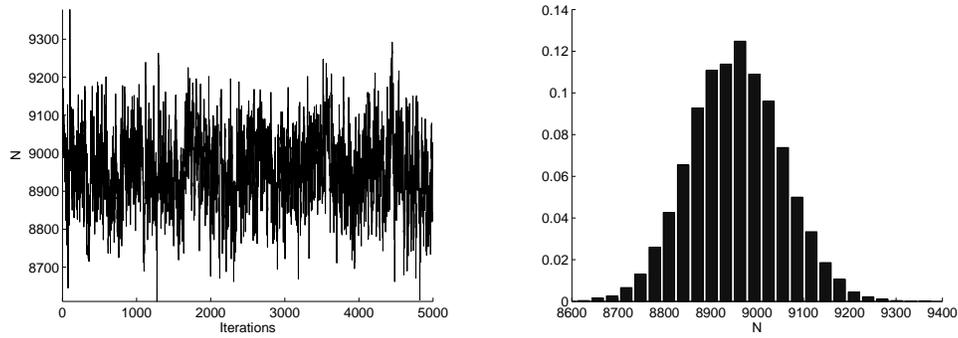


Figure 9: Experiment B220, day 11: Plots to show the number of “cells” in the image. (a) Trajectory of sampled  $N$  values in the MCMC, (b) the resulting histogram.

## 5 Additional Comments

Our applied studies involve large (though unknown) numbers of point occurrences and intensity mixture models with relatively large numbers of mixture model components to represent potentially complex patterns of variation over the spatial region. Coupled with the need for practically relevant measurement error models to link between observed, noisy data and the underlying latent spatial process of biological relevance, this represents a challenging computational as well as modelling problem context. Our examples shown here, and experiences with other data sets, indicate the relevance and utility of the model developed. The use of flexible, nonparametric Bayesian mixture models of intensity functions, pioneered by [Kottas and Sanso \(2007\)](#) and extended here, is central and key in engendering adaptability to wildly heterogeneous intensity patterns coupled with robustness and in-built parsimony. The use of effective MCMC samplers is key, and the blocked sampler for Dirichlet process mixture models is attractive from that viewpoint, but also really necessary as our overlaid measurement error structure demands that we have direct, albeit approximate evaluation of the underlying density-intensity function with the MCMC that generates from conditional posteriors of the underlying latent spatial process. In many spatial point process modelling contexts, lack of complete, direct observation on point outcomes is common, and our new methodology provides examples of how the overall analysis framework can be extended to allow for that. Our immunological context generates data sets for which truncated normal models of fluorescence, under both signal and noise at a point location, are adequate, and our experience suggests that we can robustly include learning on measurement error models within the overall analysis. Other contexts may, of course, require alternative measurement error model choices, but the general strategy will apply.

Our use of mixtures of normals for the spatial intensity builds on the well-known framework of normal mixtures and their ability to represent even very highly irregular surfaces. [Kottas and Sanso \(2007\)](#) used mixtures of bivariate betas. The choice

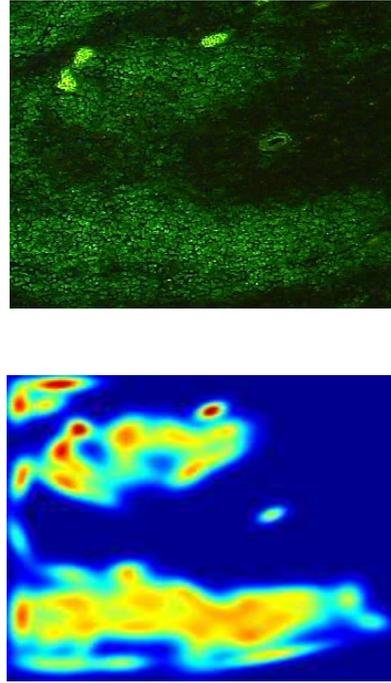


Figure 10: B220 on day 11. *Upper frame*: Image of the original data with fluorescent green tag. *Lower frame*: Image plot of the posterior estimate of the normalized intensity function.

of parametric form of the mixands, or “kernels”, is to some extent arbitrary, and the use of mixtures of betas is mathematically elegant when the spatial region is a specified rectangle. Normal mixtures do offer advantages, however. In modelling terms, the restricted range of correlations and shapes that bivariate betas are able to represent limits their flexibility. To represent an irregular intensity as accurately as a mixture of normals then requires more beta mixture components. We have explored this in studies with simulated data and confirmed a need for 4 or 5 times the number of beta than normal components in some examples. This imposes greater computational burden and decreased flexibility. Looking ahead to 3-dimensional, and possibly higher-dimensional extensions, normal mixtures clearly generalize trivially, in both modelling and computational implementation senses. Further, due to the lack of conditional conjugacy for parameters of the bivariate betas within clusters, the MCMC analysis is complicated. [Kottas and Sanso \(2007\)](#) use a traditional Dirichlet process mixture Gibbs sampler with Metropolis-Hasting (MH) steps for sets of beta parameters. Beyond the difficulties in specifying efficient MH proposals, and subsequent inefficiencies, the standard, Polya urn-based Gibbs sampler for these mixtures is inherently slow mixing and this is increasingly problematic with larger sample sizes and numbers of mixture components, such as in our examples. These issues make the MCMC sampler for beta mixtures very slow compared to approaches based on block sampling and that can analytically integrate over parameters, exploiting conditional conjugacy, as in the normal mixture models. In head-to-head comparisons with the data sets here, we find computations in the normal model to be roughly *20 times faster* per iterate than using the beta model algorithm.

Current and potential future areas for consideration include refined computational strategies to increase computational efficiency and enable at least partial parallel implementation to take advantage of both multi-threading and cluster computation. New statistical directions might include consideration of local spatial dependencies in the 0/1 outcomes process, and also potential dependencies at the observational level due to fluorescence scatter across neighboring pixel regions. Potential refinements of prior specifications over the normal variance matrix parameters may also be of interest; for example, mixtures of priors favoring very different scales of variances in the  $\Sigma_j$  may allow us to more adequately represent very heterogenous images. In the applied context of immunofluorescent images arising in studies of vaccine design, current case studies are focused in part on the context-specific questions of making comparisons between models fitted to two or more images. Further studies are currently exploring extensions of the current approach to deal with problems in which several cells of distinct biological types are marked by fluorescent tags with distinct emission spectra; interest then lies in simultaneously estimating two or more underlying spatial intensity functions for the separate cell types, with a need for dealing with the uncertainties about cell type at any one pixel region location due to frequency interference in recorded intensities.

### Supplementary material and code

The web page <http://ftp.stat.duke.edu/WorkingPapers/08-25.html> provides freely available Matlab code that implements the method described here. This includes support functions and the examples from this paper as templates for other more general models. The site also provides additional information on aspects of posterior uncertainty and predictive fit in the day-1 example. These include contour and image plots of successive samples of the DP mixture-based intensity surface through a series of MCMC iterations, and associated plots of the changes in the implied MCMC-based posterior mean estimate of the intensity function as it is updated through a series of MCMC iterations. Additional supplementary plots show pixel probabilities representing presence/absence of cell-based fluorescence as they vary over a series of MCMC iterations, accompanied by sampled spatial point patterns – i.e., locations of cells – corresponding to the above probabilities as the MCMC progresses.

In terms of computational benchmarks, for each one of the examples presented here each iterate of the MCMC algorithm presented takes roughly 2-3 seconds when running on a 2.80 GHz Intel Pentium 4 laptop with 1024 MB memory. We are investigating multicore and multiscale implementations that will speed up analyses substantially. The current MCMC is of the order of 20 times faster than alternatives using more traditional Polya urn configuration MCMC samplers.

### References

- Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes (2nd edn.)*. New York: Springer Verlag. 299
- Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*. London: Arnold. 297, 299
- Escobar, M. and West, M. (1995). “Bayesian density estimation and inference using mixtures.” *Journal of the American Statistical Association*, 90: 577–588. 300, 303
- (1998). “Computing nonparametric hierarchical models.” In Dey, D., Mueller, P., and Sinha, D. (eds.), *Practical Nonparametric and Semiparametric Bayesian Statistics*, 1–22. New York: Springer Verlag. 300, 303
- Ishwaran, H. and James, L. (2001). “Gibbs Sampling Methods for Stick-Breaking Priors.” *Journal of the American Statistical Association*, 96: 161–173. 303, 315
- Kottas, A. and Sanso, B. (2007). “Bayesian Mixture Modeling for Spatial Poisson Process Intensities, with Applications to Extreme Value Analysis.” *Journal of Statistical Planning and Inference (Special Issue on Bayesian Inference for Stochastic Processes)*, 137: 3151–3163. 297, 298, 299, 303, 304, 311, 312, 314
- MacEachern, S. (1994). “Estimating normal means with a conjugate style Dirichlet process prior.” *Communications in Statistics: Simulation and Computation*, 23: 727–741. 300, 303

- MacEachern, S. and Mueller, P. (1998). “Estimating mixture of Dirichlet process models.” *Journal of Computational and Graphical Statistics*, 7: 223–238. 300, 303
- MacEachern, S. N. (1998). “Computational methods for mixture of Dirichlet process models.” In Dey, D., Mueller, P., and Sinha, D. (eds.), *Practical Nonparametric and Semiparametric Bayesian Statistics*, 23–43. New York: Springer Verlag. 300, 303
- Moller, J. and Waagepetersen, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. London: Chapman and Hall. 297
- Mueller, P. and Quintana, F. (2004). “Nonparametric Bayesian data analysis.” *Statistical Science*, 19: 95–110. 300
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statistica Sinica*, 4: 639–650. 303
- West, M., Mueller, P., and Escobar, M. (1994). “Hierarchical priors and mixture models, with application in regression and density estimation.” In Freeman, P. and Smith, A. (eds.), *Aspects of Uncertainty: A Tribute to D.V. Lindley*, 363–386. London: Wiley. 300, 303
- Wolpert, R. and Ickstadt, K. (1998). “Poisson/Gamma random field models for spatial statistics.” *Biometrika*, 85: 251–267. 297

### Acknowledgments

We are grateful to Shelley Stewart, Heather Lynch, Greg Sempowski and Munir Alam for providing the images for this study, and to the editors and two referees for constructive comments on the original paper. We acknowledge support of the NSF (grant DMS-0342172) and NIH (grant P50-GM081883 and contract HHSN268200500019C). Any opinions, findings and conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of the NSF or NIH.

### Appendix

The block sampler MCMC for the Dirichlet mixture component of the analysis is as follows. This is based on [Kottas and Sanso \(2007\)](#), with some changes in sampling the normal mean and variance matrix parameters under the DP prior. The truncated DP model requires a chosen upper bound  $k$  on the number of mixture components. The sampler then successively resamples values of the parameters  $\theta_{1:k}^* = (\mu_{1:k}^*, \Sigma_{1:k}^*)$  of  $k$  distinct normals, together with the mixture weights  $w_{1:k}$ . Choosing  $k$  large, the posterior over all model parameters puts positive probability on zero values among the  $w_{1:k}$  so allowing for reduction of the number of effective components, and the MCMC explores this along with relevant values of the moments of normal component. Each MCMC iterate involves the following steps:

- Resample configuration indicators  $c_{1:N}$  from  $1 : k$  with probabilities

$$Pr(c_i = j) \propto w_j N(x_i | \mu_j^*, \Sigma_j^*), \quad (j = 1 : k),$$

independently over  $i = 1 : N$ . This reconfigures the  $N$  points independently among the  $k$  components, and delivers counts  $n_j = \#\{c_i = j, i = 1 : N\}$  for  $j = 1 : k$ . Note that some components may be empty, with  $n_j = 0$  for some  $j$ .

- For  $j \in \{1, \dots, k\}$ , independently sample new parameters  $\theta_j^*$  from  $p(\theta_j^* | x_{1:N}, c_{1:N})$ . The model has  $G_0(\mu, \Sigma) = N(\mu | 0, t_0 \Sigma) IW(\Sigma | s_0, S_0)$  where  $t_0 > 0$ ,  $s_0 > 0$  is the prior degree-of-freedom, and  $E(\Sigma) = S_0 / (s_0 - 2)$  when  $s_0 > 2$ . This leads to conditional normal-inverse Wishart distributions for each of the  $k$  parameters. This straightforward step samples a new set of  $k$  parameters, including new draws from  $G_0(\cdot)$  for cases with  $n_j = 0$ .
- For each component  $j = 1 : (k - 1)$ , compute  $\alpha_j = 1 + n_j$  and  $\beta_j = \alpha + \sum_{r=j+1}^k n_r$  and then sample independent beta variates  $v_j \sim Be(\alpha_j, \beta_j)$ ; set  $v_k = 1$ . Compute new values of the component probabilities via  $\pi_1 = v_1$  and  $\pi_j = v_j \prod_{r=1}^{j-1} (1 - v_r)$  for  $j = 2 : k$ .
- Resample the Dirichlet precision  $\alpha$  from its conditional posterior  $p(\alpha | N, \Theta)$  (Ishwaran and James 2001) for which a gamma prior on  $\alpha$  is conjugate.

After each sweep through these steps we have available a sample from the conditional posterior for  $\Theta = (w_{1:k}, \mu_{1:k}^*, \Sigma_{1:k}^*)$  and can therefore evaluate the density function

$$f(x) \equiv f(x | \Theta) = \sum_{j=1}^k w_j N(x | \mu_j^*, \Sigma_j^*)$$

at any chosen set of  $x$  values.

