Mixtures of Gaussian Wells: Theory, Computation, and Application in Immunofluorescence Histology

Ioanna Manolopoulou[a,*], Thomas B. Kepler[b], Daniel M. Merl[c]

[a]Department of Statistical Science, Duke University
[b]Center for Computational Immunology, Duke University
[c]Applied Statistics Group, Lawrence Livermore National Laboratory

Abstract

A primary challenge in unsupervised clustering using mixture models is the selection of a family of basis distributions flexible enough to succinctly represent the distributions of the target subpopulations. In this paper we introduce a new family of Gaussian Well distributions (GWDs) for clustering applications where the target subpopulations are characterized by hollow [hyper-]elliptical structures. We develop the primary theory pertaining to the GWD, including mixtures of GWDs, selection of prior distributions, and computationally efficient inference strategies using Markov chain Monte Carlo. We demonstrate the utility of our approach, as compared to standard Gaussian mixture methods, for the case of immunofluorescence imaging analysis, emphasizing the improved interpretability and parsimony of the GWD-based model.

Keywords: Gaussian mixtures, Poisson point processes, subtractive mixtures, histology

*To whom correspondence should be addressed: Ioanna Manolopoulou, Duke University, BOX 90251, Durham 27708

Email address: im30@stat.duke.edu (Ioanna Manolopoulou)

1. Introduction

Bayesian nonparametric mixture modeling, as exemplified by the Dirichlet process mixture model (DPM) (Müller and Quintana, 2004), has emerged as a standard statistical technique for flexible density estimation and model-based clustering. Much of the appeal of this approach stems from two features of the inference process. First, in contrast to traditional finite mixture models, the DPM specifies a mixture model with infinite component capacity, thus eliminating the requirement that the user specify *a priori* the number of mixture components. Second, the standard data augmentation scheme for performing inference on a mixture model involves the estimation of latent parameters that associate observations with individual components in the mixture model. These indicator variables can be regarded to indicate cluster membership, and as such the DPM has gained popularity as a probabilistic clustering technique.

The success of the DPM as a clustering technique, however, lies in the ability of the mixture model components to efficiently represent the probability distributions of the subpopulations of observations that constitute the individual "clusters". For example, the multivariate Gaussian DPM is often employed for clustering applications involving observations with continuous valued features. In such cases, it is often necessary for the practitioner to perform manual annotation of the inferred mixture components in order to reconcile occurrences of multiple mixture components jointly characterizing individual "clusters."

In this paper, we introduce a new family of Gaussian Well Distributions (GWDs) for use in spatial clustering applications where the aim is to identify hollow regions characterized by density "wells". The Gaussian well components of such a model are themselves *subtractive* mixtures of two multivariate Gaussian distributions. Existing approaches have focused on modelling hollow structures by fitting the data onto the surface of a hyper-ellipse or torus (Mardia and El-Atoum, 1976; Kato, 2010; Di Marzio et al., 2011; Kanatani and Rangarajan, 2011; Jones and Pewsey, 2011; Umbach and Jones, 2003; Watson, 1961; Gander et al., 1994), which can be too restrictive and does not offer an accurate representation of many other types of circular data. In contrast, the Gaussian well distribution provides positive probability density throughout the $\mathbb{R}^p$ space, except for singularities; the mode of this distribution subsequently corresponds to a hyper-ellipse.

Our motivation stems from the analysis of immunofluorescence imaging data. Immunofluorescence imaging provides measurements of the intensities of cellular surface proteins tagged with fluorescent markers. This technique is often applied to tissue samples, in order to characterize the tissue composition in terms of the spatiotemporal distributions of distinct cell types as identified by their unique surface proteins. A fundamental goal of this type of analysis is the unsupervised identification of the emergence of cellular structures called *germinal centers*, which are formed as part of the organism's immune response to some stimulus (Figure 1).

Following Kottas and Sansó (2007), who proposed the use of the Dirichlet process as a random mixing distribution of bivariate Beta distributions to model spatial point processes, Ji et al. (2009) recently demonstrated a novel computational approach for analyzing data of this type by characterizing the incidence of each pixel's fluorescence level for a particular tagged protein as a realization of a spatially inhomogeneous Poisson process. The intensity function of the Poisson process was modelled using a Dirichlet process mixture of bivariate Gaussian distributions, thus enabling the representation of arbitrarily complex spatial arrangements of the marked cells. However, representation of the intensity function as an additive mixture of Gaussian kernels only allowed identification of the germinal centers by manual inspection and annotation due to the inability of the Gaussian components to succinctly represent the density "wells" that characterize the germinal centers.

Here we extend the approach of these authors by modelling the intensity function of the spatially inhomogeneous Poisson process as a mixture of Gaussian well distributions. We demonstrate that in addition to providing greater interpretability, our approach is more parsimonious and computationally efficient, requiring fewer mixture components and therefore fewer model parameters.

The remainder of the paper is organized as follows. Section 2 provides the basic definition and properties of the Gaussian well distribution and mixtures thereof. Section 3 provides details on selection of prior distributions and develops a computationally efficient inference strategy via MCMC. Section 4 discusses the application of a Gaussian well modeling approach to the analysis of immunofluorescence imaging data obtained from a vaccine design experiment. Section 5 concludes with a discussion of alternative applications and future work.

2. The Gaussian Well Distribution

2.1. Basic theory

In this section we develop the basic theory for the Gaussian Well Distribution (GWD). We begin by defining the GWD, in its basic form, as a subtractive, rather than an additive, mixture of two multivariate Gaussian kernels:

Definition 1. *Given non-negative scalars w and v and p-dimensional Gaussian kernels $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$, a Gaussian Well Distribution is a multivariate continuous probability distribution characterized by the density function $f : \mathbb{R}^p \to \mathbb{R}_0^+$ such that*

1. $f(x) = wN(x; \mu_1, \Sigma_1) - vN(x; \mu_2, \Sigma_2)$
2. $\forall x \in \mathbb{R}^p,\ f(x) \geq 0$
3. $\int f(x)\mathrm{d}x = 1$

For appropriately chosen parameters, f resembles the Gaussian distribution $N(\mu_1, \Sigma_1)$ with a region of low density (i.e. a density "well") with shape Σ_2 surrounding μ_2. Figure 2 contains sample density surfaces for several GWDs in one and two dimensions.

The primary theory pertaining to the GWD has to do with defining the necessary constraints on the free parameters w, v, μ_1, μ_2, Σ_1, and Σ_2 so that requirements 2 and 3 above hold. We begin with the following Lemma:

Lemma 1. *The parameters w, v, μ_1, μ_2, Σ_1, and Σ_2 are valid for a Gaussian Well distribution (i.e. the positivity and integrability requirements will be met) if and only if $v = w - 1$, and all eigenvalues of D are non-negative, where*

$$D = \left(\begin{array}{c|c} (-\Sigma_1^{-1} + \Sigma_2^{-1})/2 & 0.5 \times (\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2) \\ \hline 0.5 \times (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1}) & 0.5 \times (\mu_2'\mu_2 - \mu_1'\mu_1) + \log\left(\frac{w}{w-1}\sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\right) \end{array} \right),$$

where μ' denotes the vector transpose.

Proof that these constraints satisfy the positivity and integrability requirements appears in Appendix A.

From Lemma 1, we can immediately derive the following corollary.

Corollary 2. *In the special case* $\mu_1 = \mu_2$, $\Sigma_2 = \tau\Sigma_1$, *the positivity constraint implies* $\frac{w-1}{w} \leq \tau^{p/2} \leq 1$.

Proof of the corollary appears in Appendix B. For this subclass of GWDs, the mode of the distribution is defined by an ellipsoid with equation

$$(x-\mu)'\Sigma^{-1}(x-\mu) \quad = \quad \frac{2\tau \log\left(\frac{w-1}{w}\tau^{-p/2-1}\right)}{1-\tau},$$

revealing the relationship between the shape of the well and the parameters of the distribution. The shape of this distribution is therefore natural for discovering the ring-shaped germinal centers in our immunofluorescence application. In the special case $\frac{w-1}{w} = \tau^{p/2}$, corresponding to a subclass of GWDs which we will refer to as *Ring GWDs*, the equation becomes

$$(x-\mu)'\Sigma^{-1}(x-\mu) = \frac{2\tau}{\tau-1}\log\tau.$$

In the remainder of paper, we focus on the special case where $\mu_1 = \mu_2 := \mu$ and $\Sigma_2 = \tau\Sigma_1 := \tau\Sigma$. This form of the distribution is used below in deriving a representation of mixtures of GWDs.

2.2. Mixtures of Gaussian Wells

We define a mixture of the special class of ring GWDs ($\mu_2 = \mu_1$ and $\Sigma_2 = \tau\Sigma_1$) as a convex combination of Ring GWD components

$$f(x|\boldsymbol{\pi}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) = \sum_{k=1}^{K} \pi_k f_k(x|w_k, \mu_k, \Sigma_k, \tau_k),$$

where

$$f_k(x|w_k, \mu_k, \Sigma_k, \tau_k) = w_k N(x|\mu_k, \Sigma_k) - (w_k - 1)N(x|\mu_k, \tau_k\Sigma_k),$$

and $\{\boldsymbol{\pi}, \boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}\}$ denotes the set of parameters for all components.

The mixture representation can be simplified using ring GWD, which have density exactly 0 at $x = \mu$, occurring when $\tau = \left(\frac{w_k - 1}{w_k}\right)^{2/p}$. This leads to the following lemma:

Lemma 3. *Any ring GWD*

$$f(x|w, \mu, \Sigma) = wN(x|\mu, \Sigma) - (w - 1)N(x|\mu, \tau\Sigma)$$

can be re-written as

$$f(x) = \tilde{\pi}_1 g_1(x) + \tilde{\pi}_2 g_2(x),$$

where $g_1(x)$ is a ring GWD and $g_2(x)$ is a standard [multivariate] Gaussian kernel.

Proof of the Lemma appears in Appendix C.

This result implies that the class of mixture models with ring GWD components is a subset of the class of mixture models with components *either* ring GWDs *or* standard Gaussian densities. Thus, we will be able to perform inference on the ring GWD mixture model by considering a more general mixture model capable of representing *both* ring GWD components *and* standard Gaussian components. Thus, our target mixture model is defined by

$$f(x) = \sum_{k=1}^{K} \pi_k f_k(x),$$

where each component has the following distribution

$$f_k(x) = \begin{cases} N(x|\mu_k, \Sigma_k) & \text{w.p. } (1-\rho), \\ w_k N(x|\mu_k, \Sigma_k) - (w_k - 1) N\left(x \middle| \mu_k, \left(\frac{w_k - 1}{w_k}\right)^{2/p} \Sigma_k\right) & \text{w.p. } \rho, \end{cases}$$

Here ρ is the marginal probability of any individual mixture component being a ring GWD component (as opposed to a standard Gaussian component). Since a standard Gaussian component can be viewed as a Gaussian well component with $a_k = 1$, we can express our mixture model using only Gaussian well components, so that

$$f(x) = \sum_{k=1}^{K} \pi_k w_k N\left(x \middle| \mu_k, \Sigma_k\right) - (w_k - 1) N\left(x \middle| \mu_k, \left(\frac{w_k - 1}{w_k}\right)^{2/p} \Sigma_k\right),$$

such that $w_k = 1$ with probability $1 - \rho$ corresponds to (with a slight abuse of notation) the limiting case of a standard Gaussian kernel.

3. Prior Distributions and Posterior Inference

Having defined the GWD mixture model in the previous section, we now motivate prior distributions for its parameters and describe a Markov chain Monte Carlo algorithm for conducting inference.

3.1. Priors for component mixing weights π

We use a standard truncated Dirichlet process prior for the mixing weights π using a concentration parameter $\alpha \sim Ga(\nu_\alpha, \lambda_\alpha)$. Denoting the latent component configuration indicators as z_i, for each observation x_i we have $z_i|\pi \sim \text{Multinomial}(\pi, 1)$ (see Ishwaran and James, 2002). The mixing weights π therefore follow the standard form for the stick-breaking representation of a Dirichlet process mixture model:

$$\begin{aligned} \pi_k | V_1, \ldots, V_k &= V_k \prod_{j=1}^{k-1} (1 - V_j) \text{ for } k = 1 \ldots K \\ V_j | \alpha &\sim \text{Beta}(1, \alpha) \text{ for } k = 1 \ldots K-1 \\ V_K &= 1. \end{aligned}$$

3.2. Priors for subtractive mixture weights $\{w_k\}$

As mentioned above, the subtractive mixture weight w_k is exactly 1 with probability $1-\rho$, and takes values in $(1, \infty)$ otherwise. The choice of prior on the subtractive mixture weights $w_k \in (1, \infty)$ determines the expected shape of the Gaussian well components. Different values of w_k lead to a variety of differently shaped wells. For example, in one dimension, with $w_k > 1$, the width of the Gaussian well increases with w (Figure 3).

In order to obtain further intuition about the behaviour of w, we observe that any distribution $f(x|\theta) = w f_1(x|\theta) - (w-1) f_2(x|\theta), \quad w > 1$, which satisfies $f(x|\theta) \geq 0 \quad \forall x$, can be realized through the following rejection sampler:.

1. Draw a random variable from distribution $f_1(x)$.
2. Accept that value with probability $1 - \frac{(w-1)f_2(x)}{w f_1(x)}$.
3. The accepted values will have distribution $f(x)$.
4. The rejected values will have distribution $f_2(x)$.

In the case of $f(x) = wN(x|\mu, \Sigma) - (w-1)N(x|\mu, \tau\Sigma)$, given the data X and the parameters μ and Σ, we have exact forms for the envelope distribution $f_1(x) = N(x|\mu, \Sigma)$ and the distribution of the rejected values

$f_2(x) = N(x|\mu, \tau\Sigma)$. This implies that generating values for $Y \sim f_2(x)$ and combining them with the data X, the set $\{X, Y\}$ should be well modelled by $f_1(x)$. The information about w does not come from the shape of the rejected sample, but rather from the *number of rejected observations*, n_2, required to produce $n = |X|$ accepted samples from the target distribution. According to the rejection sampling mechanism, for each accepted observation correspond $\frac{w}{w-1}$ rejected ones. In other words, the number of total observations drawn from $f_1(x)$ to obtain n accepted values is negative binomially distributed with probability $\frac{w-1}{w}$. We therefore use a natural re-parametrization in which we express w through the probability $\tau := \frac{w-1}{w} \in (0, 1)$, resulting in a Beta prior

$$\tau_k = \frac{w_k - 1}{w_k} \sim Be(p_1, p_2)$$

with probability ρ, and $\tau_k = 0$ with probability $1 - \rho$.

3.3. Priors on component-specific parameters $\{\mu_k, \Sigma_k\}$

We denote the means and variances of the Gaussian wells by (μ_k, Σ_k). The conjugate prior for $\{\mu_k, \Sigma_k\}|\{\tau_k = 0\}$ is a Normal-Inverse Wishart distribution such that $\Sigma_k|\{\tau_k = 0\} \sim IW(d, S)$ and $\mu_k|\{\tau_k = 0, \Sigma_k\} \sim N(0, t_0\Sigma_k)$. In the case of the Gaussian wells, we use $\Sigma_k|\{\tau_k > 0\} \sim IW(d^\dagger, S^\dagger)$ and $\mu_k|\{\tau_k > 0, \Sigma_k\} \sim N(0, t_0^\dagger\Sigma_k)$

Standard conjugacy results hold for the standard Gaussian components. In the case of the GWD components, this leads to posteriors of non-standard forms, for which Metropolis steps are required. Details on posterior approximation are provided in the next section.

3.4. Markov chain Monte Carlo sampler

We now describe a Markov chain Monte Carlo algorithm for producing samples from the joint posterior distribution $p(\alpha, \mu, \Sigma, \tau, \pi|X)$. A key novelty of our approach lies in the use of Gibbs and Metropolis-Hastings updates based on tractable approximations for the true posterior distributions of the parameters. This ensures rapid mixing and convergence to the target stationary distribution of the Markov chain. The algorithm is as follows:

1. Update τ_k by proposing $\tau_k = 0$ with probability ρ and $\tau_k \sim U(0,1)$ otherwise, accepting/rejecting according to the corresponding Metropolis-Hastings ratio.
2. The means and covariances μ_k and Σ_k of components with $\tau_k = 0$ can be updated through Gibbs steps.
3. The means of the GWDs can be updates through a variance-discounted version of the standard Gaussian - inverse Wishart form, such that

$$q\left(\mu_k | \left\{X, z, \tau_k > 0, \Sigma_k\right\}\right) = N\left(\frac{\bar{x}_k n_k t_0^\dagger}{1 + n_k t_0^\dagger}, 4\frac{t_0^\dagger \Sigma_k}{1 + n_k t_0^\dagger}\right),$$

   where $\bar{x}_k$ represents the sample mean of component k and n_k the number of observations in component k. The proposed values are then accepted/rejected according to the Metropolis-Hastings ratio.
4. In order to approximate the posterior distribution of Σ_k for GWDs, we use the fact that

$$Var(x_{z_i=k}) \quad = \quad w_k \times \Sigma_k - (w_k - 1) \times \tau_k \Sigma_k$$

   which provides an estimate for the posterior mean of Σ_k. Using the factor $c_k = w_k + (1 - w_k)\tau_k = w_k + (1 - w_k)\left(\frac{w_k - 1}{w_k}\right)^{2/p}$, we use the following distribution in order to approximate the posterior distribution, so that

$$q\left(\Sigma_k \middle| \{X, z, \tau_k > 0\}\right) = IW\left(d^\dagger + n_k, S^\dagger + \sum_{z_i = k} \left(x_i - \bar{x}_k\right)\left(x_i - \bar{x}_k\right)' / c_k\right),$$

where, as before, $\bar{x}_k$ denotes the sample mean of component k and n_k the number of observations in component k. The proposed values are then accepted/rejected according to the corresponding Metropolis-Hastings ratio.

5. Finally, standard Gibbs steps are used for the updates of α, π and z (see Ishwaran and James, 2002).

4. Application in immunofluorescence histology

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. Observed measurements are fluorescent intensities generated from tagged cell-surface proteins that characterize a specific cell type. 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.

Following Ji et al. (2009), our general statistical framework jointly models the fluorescence intensity function of the signal as a spatially inhomogeneous Poisson process and the noisy outcome of the point process; see

Appendix for details. Here we use mixtures of both Gaussian wells and standard Gaussian components, providing a better representation of germinal centre structure using fewer components. The Markov chain Monte Carlo sampler described in Section 3.4 is integrated within the algorithm presented in Ji et al. (2009).

We implement our algorithm on images of size $n_1 \times n_2$. The axes are transformed so that the x and y ranges are both $[-5, 5]$ by dividing them by K_x and K_y respectively, and centering. We assume priors for the image and noise signals:

$$\begin{aligned} (m_y|v, y) &\sim N(m_y|\bar{m}_y, tv_y) \\ v_y^{-1} &\sim Ga(c_v/2, c_v/(2\bar{v}_y)). \end{aligned}$$

Here $m_0 = 3.75,\ m_1 = 4.5,\ t = 0.05,\ \bar{v}_0 = 0.25,\ \bar{v}_1 = 0.5$ and $c_v = 10$.

4.1. Fitting a single Gaussian well structure

We implement our algorithm on a zoomed-in image of simply one germinal centre shown in Figure 4 (a), where a single Gaussian well-like structure is identifiable by eye. We use the following priors for signal and background, and intensity function of the occupied pixels through the Gaussian well mixture. As described in Section 3.1, we use a mixture model where each component is either a GWD or a standard Gaussian distribution, with prior probability of each $\rho = 0.5$. For the prior of the Dirichlet concentration parameter $\alpha \sim Ga(10^{-4}, 10^{-4})$. The prior of the covariance matrices of the outside well is given by $\Sigma_k|\{\tau_k > 0\} \sim IW(10, 100I)$, whereas the covariance matrices of the standard Gaussian components $\Sigma_k|\{\tau_k = 0\} \sim IW(10, 10I)$. The means of the GWDs and standard Gaussians are given by $\mu_k|\{\tau_k >$

$0, \Sigma_k\} \sim N(0, 5\Sigma_k)$ and $\mu_k|\{\tau_k = 0, \Sigma_k\} \sim N(0, 5\Sigma_k)$ respectively. We place a Beta prior on $\tau_k = (w_k - 1)/w_k \sim Be(10, 1)$. We run the sampler for 2000 iterations fixing the maximum number of components at 15, discarding the first 1000 as burn-in.

As is shown in Figure 4 (b) and (c), the data are fitted well by a single Gaussian well component, with some additional structure represented by an extra standard Gaussian component with very low weight; the posterior intensity surface shows an accurate representation of the original image.

4.2. Fitting full image: B220 PA alone day 26

Finally, we fit our algorithms to the whole image shown in Figure 5 (a). In this case there are several regions with Gaussian well shape, most of which represent germinal centre follicles, but also several regions with standard Gaussian shape. Owing to the flexibility of the model, the image is fitted by several standard and well Gaussians. We use a Gamma prior of the Dirichlet concentration parameter $\alpha \sim Ga(10^{-4}, 10^{-4})$. The prior of the covariance matrices of the outside well is given by $\Sigma_k|\{\tau_k > 0\} \sim IW(100, 10I)$, whereas the covariance matrices of the standard Gaussian components $\Sigma_k|\{\tau_k = 0\} \sim IW(5, 0.5I)$. The increased degrees of freedom in the Inverse Wishart prior distribution of the Gaussian Wells allow us to strongly incorporate expert knowledge about the shape and size of germinal centre follicles, as opposed to spurious well-like structures. The means of the GWDs and standard Gaussians are given by $\mu_k|\{\tau_k > 0, \Sigma_k\} \sim N(0, 10\Sigma_k)$ and $\mu_k|\{\tau_k = 0, \Sigma_k\} \sim N(0, 30\Sigma_k)$ respectively, and we place a Beta prior on $\tau_k = (w_k - 1)/w_k \sim Be(10, 1)$. We run the sampler for 2000 iterations fixing the maximum number of components at 15, discarding the first 1000 as burn-in.

Figure 5 (b) and (c) show that our sampler identifies around 10 Gaussian well components, which successfully correspond to germinal centre follicle candidates identifiable by eye. Comparison against the posterior distribution samples obtained using a standard Gaussian mixture model in Figure 6 shows a significant improvement both in the ability to identify these follicles, but also in terms of an accurate representation of the original image.

5. Discussion

We have presented a new class of Gaussian mixture distributions which allow for negative mixing weights, resulting in probability density functions which resemble hollow volumes; in the case of 2 dimensions, these structures look like wells. Our flexible mixture modeling approach can simultaneously model highly unstructured inhomogeneous regions through a very flexible prior and germinal centre follicles with a distinct hollow structure with a more informative prior using experts' knowledge about the shape and structures of the germinal centres. Our approach allows us to identify candidate germinal centres; these can then be used to subsequently compute measures like average and total intensity within each Gaussian well, in order to characterize and compare germinal centres within and across images.

Although we have only implemented our methods on immunofluorescence histology datasets, our models can be applied to a variety of other applied fields, such as medical imaging, astronomy or synthetic biology, where data are often dispersed around, but not restricted to, an ellipse, a sphere or a circle.

Natural extensions to our methods include investigating the properties and devising priors that allow for asymmetric Gaussian well distributions.

Figure 1: A single germinal center of a mouse lymph node stained for B220 cells.

Figure 2: Examples of Gaussian Well distributions in 1 and 2 dimensions. The left-hand column shows "ring wells" in 1 and 2 dimensions, in which the density is exactly 0 at the centre of the well. The rightmost column shows Gaussian wells with equal means; in the top row, the variances are chosen such that the the center of the well has strictly positive density, and the bottom row demonstrates the effect of asymmetric covariance matrices.

While providing a much more general class of distributions, constructing efficient proposals for exploring the sample space is crucial.

Our Markov chain Monte Carlo sampler was implemented in MATLAB; code is available upon request.

6. Acknowledgments

The authors are grateful to Heather E. Lynch, Shelley Stewart, Gregory D. Sempowski, and S. Munir Alam for producing the dataset presented in this manuscript. This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, document release number LLNL-JRNL-504251 (DMM). The work was also partially supported by NIH/NIAID HHSN268200500019C (TBK) and NSF under grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute (IM).

Figure 3: Probability density functions for 1-dimensional ring GWDs using $\mu_1 = \mu_2 = 0$ and $\sigma_1 = 1$ for increasing values of w.

References

Di Marzio, M., Panzera, A., Taylor, C.C., 2011. Kernel density estimation on the torus. Journal of Statistical Planning and Inference 141, 2156 – 2173.

Figure 4: A single germinal centre zoomed in. The leftmost figure shows the original images, representing a zoomed-in version of the lymph node of a mouse stained for B220. The middle figure shows the component structure at a single iteration of the MCMC, and rightmost figure shows the posterior mean intensity function. The Gaussian well successfully shapes around the germinal centre structure using a single Gaussian well component and an additional standard Gaussian component in order to account for the slight asymmetry of the germinal centre.

Figure 5: The full image, on the left, of a B220 stain from a mouse lymph node. The middle figure shows the component structure at a single iteration of the MCMC, and the right one shows the posterior mean intensity function. The Gaussian well structures model the 6/7 germinal centres visibly identifiable in the original image, with more than one component required for some.

Gander, W., Gander, W., G, W., Golub, G.H., Strebel, R., 1994. Fitting of Circles and Ellipses - Least Squares Solution. Technical Report. ETH.

Ishwaran, H., James, L., 2002. Approximate Dirichlet process computing in finite normal mixtures: Smoothing and prior information. Journal of Computational and Graphical Statistics 11, 508–532.

Ji, C., Merl, D., Kepler, T., West, M., 2009. Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology. Bayesian Analysis 4, 297–316.

Jones, M.C., Pewsey, A., 2011. Inverse batschelet distributions for circular data. Biometrics , no–no.

Kanatani, K., Rangarajan, P., 2011. Hyper least squares fitting of circles and ellipses. Computational Statistics and Data Analysis 55, 2197 – 2208.

Kato, S., 2010. A markov process for circular data. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72, 655–672.

Kottas, A., Sansó, B., 2007. Bayesian mixture modeling for spatial poisson

Figure 6: Mixture structure and intensity plot for the same image using a standard Gaussian mixture model.

process intensities, with applications to extreme value analysis. Journal of statistical planning and inference 137, 3151–3163.

Mardia, K.V., El-Atoum, S.A.M., 1976. Bayesian inference for the von mises-fisher distribution. Biometrika 63, 203–206.

Müller, P., Quintana, F., 2004. Nonparametric Bayesian data analysis. Statistical Science 19, 95–110.

Umbach, D., Jones, K.N., 2003. A few methods for fitting circles to data. IEEE Transactions on Instrumentation and Measurement 52.

Watson, G.S., 1961. Goodness-of-fit tests on a circle. Biometrika 48.

Appendix A. Proof of Lemma 1

Lemma. *The parameters w, v, μ_1, μ_2, Σ_1, and Σ_2 are valid for a Gaussian Well distribution (i.e. the positivity and integrability requirements will be met) if and only if $v = w - 1$, and all eigenvalues of D are non-negative, where*

$$D = \left(\begin{array}{c|c} (-\Sigma_1^{-1} + \Sigma_2^{-1})/2 & 0.5 \times (\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2) \\ \hline 0.5 \times (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1}) & 0.5 \times (\mu_2'\mu_2 - \mu_1'\mu_1) + \log\left(\frac{w}{w-1}\sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\right) \end{array} \right),$$

where μ' denotes the vector transpose.

Proof. The function requirement trivially integrates to 1 iff $v = w - 1$. The positivity requirement $f(x) = wf_1(x) - vf_2(x) \geq 0$ is equivalent to $\log(wf_1(x)) - \log(vf_2(x)) \geq 0$. Then we have

$$\begin{aligned} \log(f_1(x)) - \log(f_2(x)) &= \log(w) - 0.5\log(2\pi|\Sigma_1|) - \frac{1}{2}(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1) \\ &\quad - \log(1-w) + 0.5\log(2\pi|\Sigma_2|) + \tfrac{1}{2}(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2) \\ &= \log\left(\tfrac{w}{v}\sqrt{\tfrac{|\Sigma_2|}{|\Sigma_1|}}\right) - \tfrac{1}{2}(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1) + \tfrac{1}{2}(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2) \end{aligned}$$

which can be written as a quadratic form

$$(x_1 \cdots x_p\ 1) D (x_1 \cdots x_p\ 1)'$$

where

$$D = \left(\begin{array}{c|c} (-\Sigma_1^{-1} + \Sigma_2^{-1})/2 & 0.5 \times (\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2) \\ \hline 0.5 \times (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1}) & 0.5\,(\mu_2'\mu_2 - \mu_1'\mu_1) + \log\left(\frac{w}{v}\sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\right) \end{array} \right)$$

In the above symmetric matrix D, the left upper-triangular $p \times p$ matrix is equal to $\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)/2$, the first p entries of the $p+1$ column of the matrix are equal to $\Sigma_1^{-1}\mu_1 - \Sigma_2^{-1}\mu_2$, and the $(p+1, p+1)$ entry of the matrix is equal to $-0.5\mu_1'\mu_1 + 0.5\mu_2'\mu_2 + \log\left(\frac{w}{v}\sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\right)$.

Thus the original inequality reduces to the requirement that D be positive semi-definite, which will be true when all eigenvalues of D are positive. Thus it is necessary that $\Sigma_1^{-1} - \Sigma_2^{-1}$ is positive semi-definite, representing the intuitive condition that $f_2(x)$ must have lighter tails than $f_1(x)$. A further necessary condition is that all diagonal elements are positive, which in this case requires $-0.5\mu_1'\mu_1 + 0.5\mu_2'\mu_2 + \log\left(\frac{w}{v}\sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\right) > 0$. □

Appendix B. Proof of Corollary 2

Corollary. *In the special case $\mu_1 = \mu_2$, $\Sigma_2 = \tau\Sigma_1$, the positivity constraint implies $\frac{w-1}{w} \leq \tau^{p/2} \leq 1$.*

Proof. In the special case $\mu_1 = \mu_2$ and $\Sigma_2 = \tau\Sigma_1$, we have

$$D = \left(\begin{array}{c|c} (1-1/\tau)\Sigma_1^{-1}/2 & 0 \\ \hline 0 & \log\left(\frac{w}{v}\sqrt{\tau^p}\right) \end{array}\right)$$

This implies that, in order for D to be positive semi-definite, we require $\frac{w-1}{w} \leq \tau^{p/2} \leq 1$. □

Appendix C. Proof of Lemma 3

Lemma. *Any ring GWD*

$$f(x|w,\mu,\Sigma) = wN(x|\mu,\Sigma) - (w-1)N(x|\mu,\tau\Sigma)$$

can be re-written as

$$f(x) = \tilde{\pi}_1 g_1(x) + \tilde{\pi}_2 g_2(x),$$

where $g_1(x)$ is a ring GWD and $g_2(x)$ is a standard [multivariate] Gaussian kernel.

Proof.

$$wN(\mu, \Sigma) - (w-1)N(\mu, \tau\Sigma) =$$
$$\tilde{\pi}_1 \left(\frac{1}{1-\tau^{\frac{p}{2}}} N(\mu, \Sigma) - \left(\frac{1}{1-\tau^{\frac{p}{2}}} - 1 \right) N(\mu, \tau\Sigma) \right) + \tilde{\pi}_2 N(\mu, \tau\Sigma),$$

where $\tilde{\pi}_1 = w(1-\tau^{\frac{p}{2}}), \tilde{\pi}_2 = 1 - \tilde{\pi}_1$. □

Appendix D. Indirectly observed point process

Since the cells are much larger than each individual pixels, each pixel is assumed to be either occupied by a cell or is empty, corresponding to signal or noise in the model. Occupied pixels are modeled as an inhomogeneous Poisson process, with Poisson intensity surface of the signal following a random mixture of 2-dimensional Gaussian wells. In other words, the locations $\mathbf{x} = \{x_i\}$ of cells are independent 2-dimensional variates with probability density function of a mixture of Gaussian wells.

Using the de-composition into standard Gaussian and Gaussian Well components (see Section 3), the mixture model can be realized through a set of latent configuration indicators c_i for each observation x_i, allocating each observation to precisely one component of the mixture model (either Gaussian well or standard Gaussian), allowing to improve efficiency of the algorithms Ishwaran and James (2002). Here the number of components is unknown, and we use flexible priors for their location, spread and well parameter $\theta_k = (w_k, \mu_k, \Sigma_k)$.

Following Ji et al. (2009), we define a latent binary indicator $y \in \{0, 1\}$, corresponding to the absence or presence, respectively, of cell-associated immunofluorescence in the pixel region with index point $x_i \in \mathcal{X}$. Then the likelihood function fits the set of observed points $\{x \in \mathcal{X} : y(x) = 1\}$, corresponding to pixels occupied by a part of a cell, such that

$$p(X|\gamma, f) \propto \exp(-\gamma)\gamma^N / a^{2N} \prod_{i=1}^{N} f(x_i).$$

Here γ represents the overall scale parameter $\gamma = \int_{x \in S} \lambda(x)dx$, S being the finite image region, and a^2 is the 2-dimensional pixel region. Given the binary probability indicator for each pixel, $p(z|y = 0, \mu)$ represents background noise in the absence of an object at a specific location, and similarly $p(z|y = 1, \mu)$ represents noise in the presence of a signal. Then if $z = z(y(x))$ represents measured fluorescence at pixel region x, the $p(z|y, \mu)$ is defined by

$$\begin{aligned}
(\log(z)|y = 0) &\sim N(m_0, v_0)I(\log(z) < h_2), \\
(\log(z)|y = 1) &\sim N(m_1, v_1)I(h_1 < \log(z) < h_2),
\end{aligned}$$

where m_0, v_0 relates to the background noise and (m_1, v_1) to signal fluorescence. The upper truncation is inherent to the digital fluorescent image bit depth and the lower truncation h_1 of the image signal is introduced as an identifying criterion which distinguishes between signal and noise.