Published in final edited form as:

Comput Stat Data Anal. 2012 December 1; 56(12): 3809–3820. doi:10.1016/j.csda.2012.03.027.

Mixtures of Gaussian Wells: Theory, Computation, and Application

Ioanna Manolopoulou^{a,*}, Thomas B. Kepler^b, and Daniel M. Merl^c

^aDepartment of Statistical Science, Duke University

^bDepartment of Microbiology, Boston University

^cApplied 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 on a synthetic dataset, and exemplify its applicability on an example from immunofluorescence imaging, emphasizing the improved interpretability and parsimony of the GWD-based model.

Keywords

Gaussian mixtures; Poisson point processes; subtractive mixtures; histology

Non-Homogeneous Poisson Processes (NHPPs) are a natural and powerful tool in inference for data consisting of point patterns because they provide interpretable representations while allowing for tractable computation. Its combination with 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 in point processes. Finally, the extension of those methods to pixelized imaging datasets has allowed for flexible statistical analysis of a wide variety of images (see Ji et al., 2009).

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’. Here 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

© 2012 Elsevier B.V. All rights reserved.

*To whom correspondence should be addressed: Ioanna Manolopoulou, Duke University, BOX 90251, Durham 27708
im30@stat.duke.edu (Ioanna Manolopoulou).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

components of such a model are themselves *subtractive* mixtures of two multivariate Gaussian distributions.

Our motivation stems from the analysis of immunofluorescence imaging data, although similar structures are applicable beyond imaging in areas such as astronomy, systems biology and demographics. Immunofluorescence imaging provides measurements of the intensities of cellular surface proteins tagged with fluorescent markers. A fundamental goal of this type of analysis is the unsupervised identification of the emergence of ‘hollow’ cellular structures called *germinal centers*, which are formed as part of the organism’s immune response to stimulus (Figure 1).

The remainder of the paper is organized as follows. Section 1 provides an introduction to the NHPP, its combination with DPM models, and their use in indirectly observed imaging. 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. In Section 4 we compare our methods with standard Gaussian DPM models on a synthetic dataset, and finally apply them on an immunofluorescence imaging dataset obtained from a vaccine design experiment. Section 5 concludes with a discussion of alternative applications and future work.

1. Indirectly observed Non-Homogeneous Poisson process

A NHPP, denoted by $PoP(S, \lambda)$, is defined in an observation region S with intensity $\lambda(x)$ for $x \in S$, which is a non-negative and locally integrable function for all bounded $s \subseteq S$. The Poisson intensity function $\lambda(x) \geq 0$ ($x \in S$) is defined such that in any region $s \subseteq S$, the number of event-points $n(s)$ is Poisson distributed with mean $\gamma(s) = \int_{x \in s} \lambda(x) dx$, implying that disjoint regions are independent given $\lambda(x)$. Then, conditional on $n(s)$, point locations within s are distributed independently with density $\lambda(x)/\gamma(s)$.

Modeling a non-homogeneous Poisson process then amounts to employing flexible representations of the intensity function such as nonparametric mixtures. 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’.

Spatial point process modeling can be extended to imaging datasets representing pixelized observations, in which point-events represent pixels occupied by cells or parts of cells. Ji et al. (2009) employed a flexible Bayesian Gaussian mixture model approach for the analysis

of indirectly observed point processes. Their method relies upon using a pixel-discretized Gaussian mixture to model the Poisson intensity surface of a partially observed signal.

The 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 Ji et al., 2009). We assume that each pixel of size $a \times a$ represents either part of a cell or background. The distribution of cell-occupied pixels is modeled as an inhomogeneous Poisson process, the Poisson intensity surface of which gives the probability of a pixel's being associated with a cell. The intensity surface of the signal is modeled as a random mixture of 2-dimensional distributions using Dirichlet process mixing weights Ishwaran and James (2002). The mixture model allows the intensity function to be represented as an inhomogeneous smooth surface, representing regions of high and low probability respectively. In other words, the locations $X = \{x_i\}$ of cells are independent 2-dimensional variates with probability density function

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

where $f_k(x_i)$ denotes the probability density function of a base distribution evaluated at x_i , and the mixing weights π_k are such that $0 \leq \pi_k \leq 1$, $\sum \pi_k = 1$. 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, 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 parameters.

A latent binary indicator $y \in \{0, 1\}$ is defined (see Ji et al., 2009), 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 area. 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. The distribution $\log(z)|y=1$ can be viewed as a mark distribution on the Poisson observations such that it's independent of the position of the point process. A

natural extension of the proposed model would include a joint distribution of the location of cells and fluorescence of the image (see Taddy and Kottas, 2010).

2. The Gaussian Well Distribution

We define a new class of distributions aimed at modelling relatively hollow structures. The Gaussian well components of such a model are themselves *subtractive* mixtures of two multivariate Gaussian distributions. Existing approaches have focused on 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 when data cannot be accurately represented by a lower-dimensional manifold. 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.

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 $f: \mathbb{R}^p \rightarrow \mathbb{R}_0^+$ 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. $\int f(x)dx = 1$

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, with

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

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

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

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

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

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

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

where

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

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

The mixture representation can be simplified using a combination of standard Gaussian and

GWDs, 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 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 GWD components is a subset of the class of mixture models with components *either* ring GWDs *or* standard Gaussian densities. Our target mixture model is then defined by

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

where each component has the following distribution

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

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 $w_k = 1$, we can express our mixture model using only Gaussian well components, so that

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

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 $\alpha \sim \text{Ga}(\nu_\alpha, \lambda_\alpha)$. Denoting the latent component configuration indicators as z_i for each observation x_i we have $z_i^\top \pi \sim \text{Multinomial}(\pi, 1)$ (see Ishwaran and

James, 2002). Using similar notation as Ji et al. (2009) for convenience, the mixing weights π therefore follow the standard form for the stick-breaking representation of a Dirichlet process mixture model:

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

The Dirichlet Process can also be viewed through the corresponding Polya Urn representation, and can be extended to a two-dimensional Poisson-Dirichlet prior distribution (see Teh, 2006).

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, and 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 $f(x|\theta) = wf_1(x|\theta) - (w-1)f_2(x|\theta)$, $w > 1$, which satisfies $f(x|\theta) \geq 0 \forall x$, can be realized through the following rejection sampler.

1. Draw a random variable from distribution $f_1(x)$.
2. Accept that value with probability $1 - \frac{(w-1)f_2(x)}{wf_1(x)}$.
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, to 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

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

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 efficient mixing and convergence to the target stationary distribution of the Markov chain. The algorithm is as follows:

1. Update $z_0, z_1, m_0, m_1, v_0, v_1, \gamma$ according to Ji et al. (2009).
2. 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.
3. The means and covariances μ_k and Σ_k of components with $\tau_k = 0$ can be updated through Gibbs steps.
4. The means of the GWDs can be updated through a variance-discounted version of the standard Gaussian - inverse Wishart form, such that

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

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.

5. In order to approximate the posterior distribution of Σ_k for GWDs, we use the fact that

$$\text{Var}(x_{c_i=k}) = w_k \times \sum_k - (w_k - 1) \times \tau_k \sum_k$$

which provides an estimate for the posterior mean of Σ_k . Using the factor

$v_k = w_k + (1 - w_k) \tau_k = w_k + (1 - w_k) \left(\frac{w_k - 1}{w_k}\right)^{2/p}$, we use the following distribution in order to approximate the posterior distribution, so that

$$q\left(\sum_k | \{X, c, \tau_k > 0\}\right) = IW\left(d^\dagger + n_k, S^\dagger + \sum_{c_i=k} (x_i - \bar{x}_k)(x_i - \bar{x}_k)' / v_k\right),$$

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.

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

4. Implementation

We first implement our methods on a synthetic datasets in order to demonstrate the superiority of the Gaussian well distribution at modeling well-like structures in Section 4.1 in a simplified example of vanilla mixture modeling in a standard setting (not as a point process or pixelized image). We then continue in Section 4.2 to implement the indirectly observed imaging point process combined with our Gaussian well mixtures on an example from immunofluorescence histology, and again compare the performance with the use of a standard Gaussian mixture model (instead of the GWD).

Both of the implementations assume symmetric Gaussian well distributions, such that both inner and outer wells have the same mean and same covariance shape (up to a constant of proportionality). This is a reasonable assumption in the case of the immunofluorescence histology, since germinal centres have a strong symmetric structure.

4.1. Synthetic example

We generate data from a symmetric Gaussian well distribution, obtaining a pattern shown in the scatter plot of panel (a) and (d) of Figure 4. We first use a standard Gaussian mixture model in order to fit the data. For the prior of the Dirichlet concentration parameter we assume $\alpha \sim \text{Ga}(10^{-4}, 10^{-4})$. The prior of the covariance matrices of the standard Gaussian components is given by $\Sigma_k | \{\tau_k = 0\} \sim IW(10, 10I)$, and the prior for the means by $\mu_k | \{\tau_k = 0, \Sigma_k\} \sim N(0, 5\Sigma_k)$. We run the sampler for 4000 iterations fixing the maximum number of components at 15, discarding the first 2000 as burn-in.

The results of the analysis are shown on the top three panels of Figure 4. Although the posterior mean intensity plot shows a reasonable representation of the underlying structure, there is no direct way of ‘extracting’ that structure. In addition, the model requires a large number of components (see Figure 5) in order to accurately model the data

We then use our GWD mixture model to fit the data. 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 $\rho = 0.5$. The prior of the covariance matrices of the outside well is given by $\Sigma_k | \{\tau_k > 0\} \sim IW(10, 100I)$, and the prior for the means of the GWDs is given by $\mu_k | \{\tau_k > 0, \Sigma_k\} \sim N(0, 5\Sigma_k)$. We place a Beta prior on $\tau_k = (w_k - 1)/w_k \sim \text{Be}(10, 1)$. We run the sampler for 4000 iterations fixing the maximum number of components at 15, discarding the first 2000 as burn-in.

The model successfully identifies the single well-like component, consistently using only a single GWD component in order to accurately represent the data as shown in Figure 4. This confirms that, in cases when hollow structures are present in the data, Gaussian well distributions provide a flexible and powerful tool for parsimoniously fitting the data and automatically extracting the hollow features.

4.2. 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.

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:

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

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

We fit our algorithms to the image shown in Figure 6(a). A few regions with potential Gaussian well shape are identifiable by eye, 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 for 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 200000 iterations fixing the maximum number of components at 50, discarding the first 100000 as burn-in.

Figure 6(b) and (c) show that our sampler identifies around 10 Gaussian well components, which successfully correspond to germinal centre follicle candidates identifiable by eye. The lower light-hand panel of Figure 6 shows the posterior distribution of (i) the total number of components (both Gaussian and GWDs), and (ii) the number of Gaussian wells.

In order to assess the convergence of the MCMC sampler, Figure 7 shows aspects of the trajectories. The top panels of the Figure show trace plots of the total number of components and number of GWD components respectively. The lower panels show trace plots of the Dirichlet concentration parameter and marginal log-likelihood, respectively. Finally, ordering the components by decreasing weights at each iteration and calculating corresponding acceptance rates for w_k , Σ_k and μ_k for each components yields a minimum of 0.39, 0.37 and 0.27 post burn-in, respectively. The sampler was implemented in MATLAB and complete runtime on a desktop computer was about 20 hours.

We then repeat the analysis, this time forcing only standard Gaussian components. Comparison against the posterior distribution samples obtained using a standard Gaussian mixture model in Figure 8 shows a significant improvement both in the ability to identify these follicles, but also in terms of an accurate representation of the original image. The lower right-hand panel of Figure 6, Boxplot (iii), shows the posterior distribution of the number of standard Gaussian components, which significantly exceed the total number of components of the GWD mixture shown in Boxplot (i) of the same Figure.

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 as described in Section 2. While providing a much more general class of distribution as, constructing efficient proposals for exploring the sample space is crucial.

Finally, introducing a mark distribution with spatial structure (see Taddy and Kottas, 2010) will allow for spatial correlations of the fluorescence intensity to provide borrowing of information across neighboring pixels.

Our Markov chain Monte Carlo sampler was implemented in MATLAB; code is available at ftp.stat.duke.edu/WorkingPapers/11-22.html.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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).

References

- Di Marzio M, Panzera A, Taylor CC. Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*. 2011; 141:2156–2173.
- Gander, W.; Gander, WGW.; Golub, GH.; Strebel, R. Technical Report. ETH; 1994. Fitting of Circles and Ellipses - Least Squares Solution.
- Ishwaran H, James L. Approximate Dirichlet process computing in finite normal mixtures: Smoothing and prior information. *Journal of Computational and Graphical Statistics*. 2002; 11:508–532.
- Ji C, Merl D, Kepler T, West M. Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology. *Bayesian Analysis*. 2009; 4:297–316. [PubMed: 21037943]
- Jones, MC.; Pewsey, A. *Biometrics*. 2011. Inverse Batschelet distributions for circular data.
- Kanatani K, Rangarajan P. Hyper least squares fitting of circles and ellipses. *Computational Statistics and Data Analysis*. 2011; 55:2197–2208.
- Kato S. A Markov process for circular data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2010; 72:655–672.
- Mardia KV, El-Atoum SAM. Bayesian inference for the von Mises-Fisher distribution. *Biometrika*. 1976; 63:203–206.
- Müller P, Quintana F. Nonparametric Bayesian data analysis. *Statistical Science*. 2004; 19:95–110.

- Taddy, M.; Kottas, A. Mixture modeling for marked Poisson processes. 2010. Arxiv preprint arXiv: 1012.2105
- Teh, Y. A hierarchical bayesian language model based on pitmanyor processes. Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, Association for Computational Linguistics; 2006. p. 985-992.
- Umbach, D.; Jones, KN. IEEE Transactions on Instrumentation and Measurement. 2003. A few methods for fitting circles to data; p. 52
- Watson GS. Goodness-of-fit tests on a circle. Biometrika. 1961:48.

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

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

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

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

which can be written as a quadratic form

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

where

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

In the above symmetric matrix D , the left upper-triangular $p \times p$ matrix is equal to $(\Sigma_1^{-1} - \Sigma_2^{-1})/2$, the first p entries of the $p+1$ column of the matrix are equal to

$\sum_1^{-1} \mu_1 - \sum_2^{-1} \mu_2$, 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

$\sum_1^{-1} - \sum_2^{-1}$ 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

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

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

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.

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 .

Figure 4a

Figure 4b

Figure 4c

Figure 4d

Figure 4e

Figure 4f

Figure 4.

Top row indicates results using a standard Gaussian mixture model, whereas lower row shows results using mixture of Gaussian wells. Left hand panels: a single sample of the MCMC, with black ellipses centred about the mean of a component at 1 standard deviation in each direction, whereas red ellipses (where present) indicate 1 standard deviation of inner Gaussian well. Middle panels: posterior mean of the NHPP intensity function. Right-hand panels: trace plots of marginal log-likelihood.

Figure 5. Boxplot representing the posterior distribution for the total number of components of the standard Gaussian mixture model.

Figure 6a

Figure 6b

Figure 6c

Figure 6d

Figure 6.

On the top left, a B220 stain from a mouse lymph node. The top right figure shows the component structure at a single iteration of the MCMC, with black ellipses centred about the mean indicated by '+' of a component at 1 standard deviation in each direction, whereas red ellipses (where present) indicate 1 standard deviations of inner Gaussian well. The lower left-hand panel shows the posterior mean intensity function. Finally, the lower right-hand figure shows the posterior distribution of the total number of components. Boxplot (i) shows the total number of components, (ii) shows the number of Gaussian wells, and (iii) shows the posterior distribution of the total number of components using a standard Gaussian mixture model only.

Figure 7a

Figure 7b

Figure 7c

Figure 7d

Figure 7.

Trace plots of thinned Markov chain Monte Carlo samples for (a) the total number of components, (b) number of Gaussian well components, (c) Dirichlet concentration parameter and (d) marginal log-likelihood of the data.

Figure 8a

Figure 8b

Figure 8. Single MCMC sample of the mixture structure, with '+'s indicating component means and ellipses at one standard deviation, and intensity plot for the same image using a standard Gaussian mixture model.