THE UNIVERSITY OF WARWICK

Bayesian Nonparametric Modelling of Spatial Data

M. Kolossiatis & J. Griffin

Introduction

In modelling spatial data there is often the need to use models that are able to capture some form of correlation between some variables. For example, we may want to introduce spatial dependence in the rates of occurrence of an epidemic in some geographical regions. Additionally, we can use the flexible modelling provided by Bayesian nonparametric methods.

Bayesian nonparametric modelling

The term “Bayesian nonparametric models” refers to probability models with infinitely many parameters. One way to construct such models is by using random probability measures (RPM), i.e. probability measures that are themselves random. The mostly used RPM is the Dirichlet process (DP) (Ferguson, 1973). An alternative choice is the normalised inverse-Gaussian process (N-IGP) (Lijoi, Mena and Prünster, 2005).

RPMs are usually used in some middle stage of a hierarchical model, for example:

$$Y_i \sim g(\theta_i),\ i = 1, 2, \ldots, n$$
$$\theta_i \sim F$$
$$F \sim \text{RPM}(\lambda)$$
$$\lambda \sim H$$

For the cases of DP and N-IGP, for example, this structure can help overcome the discreteness of their realisations. Also, a hierarchical structure can lead to mixture models.

Almost all methods for posterior inference proposed in the literature are Markov Chain Monte Carlo (MCMC) methods, and especially Gibbs sampling.

Combining inference

It is the case where we have data that have some form of correlation, as in the example provided in the Introduction. There are many ways to model such data. In the nonparametric context, we might consider using related nonparametric models. An example of such a nonparametric model is proposed in (Müller, Quintana and Rosner, 2004):

The model of Müller, Quintana and Rosner (2004)

Assume that we have J related submodels, each corresponding to the distributions $H_1, H_2, \ldots, H_J$. From each submodel j we have observations $y_{ji},\ i = 1, 2, \ldots, n_j$. Suppose that each distribution H_j can be written as

$H_j = \varepsilon F_0 + (1-\varepsilon) F_j,\ j = 1, 2, \ldots, J,\ 0 \le \varepsilon \le 1$, where $F_0, F_1, \ldots, F_J$ are some nonparametric distributions. We therefore have a common part (F_0) and an idiosyncratic part (F_j), specific for each H_j (and respective submodel j). In this way, we have introduced dependence between the submodels. Additionally, in this model we can directly infer about each model-specific part and about the part that is common in all the submodels. Note that ε is common in all distributions H_j and can be seen as the level of borrowing strength across them.

For example, consider the following hierarchical mixture of Dirichlet processes (MDP) model:

$$y_{ji} \sim N(\mu_{ji}, S),\ j = 1, 2, \ldots, J,\ i = 1, 2, \ldots, n_j$$
$$\mu_{ji} \sim H_j, \text{ where } H_j = \varepsilon F_0 + (1-\varepsilon) F_j \qquad (1)$$
$$F_j \sim \text{DP}(M_j, G_0(m, B)),\ j = 0, 1, 2, \ldots, J$$
$$M_0, M_1, \ldots, M_J \overset{iid}{\sim} \text{Ga}(a_0, b_0),\ \varepsilon \sim p(\varepsilon),$$
$$S \sim IW(q, (qR)^{-1}),\ (m, B) \sim \pi(m, B)$$

where $IW(s, D)$ denotes the inverse Wishart distribution with s degrees of freedom and matrix parameter D,

$G_0 \equiv N(m, B)$, a multivariate normal distribution with parameters m and B, which are given a conjugate hyperprior distribution: $\pi(m, B) = N(m_0, A) \times IW(c, (cC)^{-1})$. The authors suggest a prior for ε that allows for positive probabilities for the two extreme cases $\varepsilon = 0$ and $\varepsilon = 1$:

$$p(\varepsilon) = \pi_0 \delta_0(\varepsilon) + \pi_1 \delta_1(\varepsilon) + (1 - \pi_0 - \pi_1)\text{Be}(\alpha_\varepsilon, \beta_\varepsilon)$$

where $\alpha_\varepsilon, \beta_\varepsilon > 0,\ 0 \le \pi_0, \pi_1 < 1$ and $\pi_0 + \pi_1 < 1$. Note in model (1) that the common and idiosynchratic cases F_0 and F_j have the same base distribution G_0 and what distinguish them are their concentration parameters M_0 and M_j.

Our proposed work

It is known that both the DP and the N-IGP can be constructed by normalising the gamma process and the inverse-Gaussian process, respectively: $\forall\ B \subset \Omega$,

$F(B) = \frac{G(B)}{G(\Omega)}$, where $F \sim$ DP/N-IGP(MH) and

$G \sim$ GammaPr/Inv-GaussianPr(MH).

In fact, normalising a random measure is a general method of constructing random probability measures.

The idea of our project will be to exploit the infinite divisibility of the underlying random measure, in order to construct random probability measures which are identically distributed, but not independent. Those models could be used in modelling spatial data, when it is natural to consider them as identically distributed and dependent.

As a simple example, consider the DP. The gamma process is infinitely divisible, i.e. if

$G^* \sim \text{GammaPr}(MH),\ G_i \sim \text{GammaPr}(M_i H),\ i = 1, 2, \ldots, k$, and $\sum_{i=1}^k M_i = M$, then $\forall\ A \subset \Omega$,

$$G^*(A) \overset{d}{=} \sum_{i=1}^{k} G_i(A).$$

By normalising this expression, we have: $\forall\ A \subset \Omega$,

$$F^*(A) = \frac{G^*(A)}{G^*(\Omega)} = \frac{\sum_{i=1}^k G_i(A)}{\sum_{j=1}^k G_j(\Omega)} = \sum_{i=1}^{k} \frac{G_i(\Omega) G_i(A)}{\sum_{j=i}^k G_j(\Omega) G_i(\Omega)}$$

$\Rightarrow F^*(A) = \sum_{i=1}^k w_i F_i(A)$ where $w_i = \frac{G_i(\Omega)}{\sum_{j=1}^k G_j(\Omega)}$.

Consider now the two-components case $k = 2$:

Let $F_1 \sim \text{DP}(M_1, H)$ and $F_2, F_3 \sim \text{DP}(M_2, H)$.

By normalising F_1 and F_2 we get:

$$F_1^*(A) = wF_1(A) + (1-w)F_2(A)$$

and by normalising F_1 and F_3 we get:

$$F_2^*(A) = wF_1(A) + (1-w)F_3(A)$$

Clearly, the two produced RPMs are identically distributed, but not independent. In fact, it can be shown that $\forall A \in \Omega$,

$$Corr(F_1^*(A), F_2^*(A)) = \frac{M_1}{M_1 + M_2}.$$

Now we can embed the above structure into a hierarchical model that is very similar to model (1) (for $J = 2$) :

$$y_{ji} \sim N(\mu_{ji}, S),\ j = 1, 2\ i = 1, 2, \ldots, n_j$$
$$\mu_{ji} \sim H_j,\ \ H_j = wF_1 + (1-w)F_{j+1}\ , j = 1, 2 \qquad (2)$$
$$F_1 \sim \text{DP}(M_1, G_0(m, B)),\ F_2, F_3 \overset{iid}{\sim} \text{DP}(M_2, G_0(m, B))$$
$$M_1, M_2 \overset{iid}{\sim} \text{Ga}(a_0, b_0),\ w \sim \text{Be}(M_1, M_2),$$
$$S \sim IW(q, (qR)^{-1}), (m, B) \sim N(m_0, A) \times IW(c, (cC)^{-1})$$

We can see that in this model, we have an additional “relationship” between our parameters, $w \sim \text{Be}(M_1, M_2)$. This relationship can cause some complications, both in algebraic calculations and in posterior simulation. However, this is a very special case and the behaviour of such models must be studied in a broader context.

Computational issues

In models (1) and (2), we are mostly interested in the predictive distributions for each dataset, $p(y_{j,n_j+1}|y_{j,1}, \ldots, y_{j,n_j})$ and probably in the posterior distributions of the concentration parameters M_j and of the common weights of the common part, ε or w. Posterior inference for both models is easily implemented using MCMC algorithms:

- Model 1: For conjugate prior distributions of the parameters, this algorithm is a Gibbs sampler. Simulations are also enhanced by using appropriate auxiliary variables.
- Model 2: Due to the prior distribution for ε, the posterior distributions for M_1 and M_2 are not of known form. We therefore use a Metropolis-Hastings step for updating them. For all other parameters we use Gibbs sampling.

Results

We checked the two models using simulated data.

Example: Consider the case

$Y_{1i} \overset{iid}{\sim} 0.2 \cdot \text{N}(-10, 1) + 0.8 \cdot \text{N}(1, 1),\ i = 1, 2, \ldots, 162$

$Y_{2i} \overset{iid}{\sim} 0.5 \cdot \text{N}(8, 1) + 0.5 \cdot \text{N}(1, 1),\ i = 1, 2, \ldots, 162.$

The predictive distributions in both cases were as we would expect them to be and are shown in figure 1:

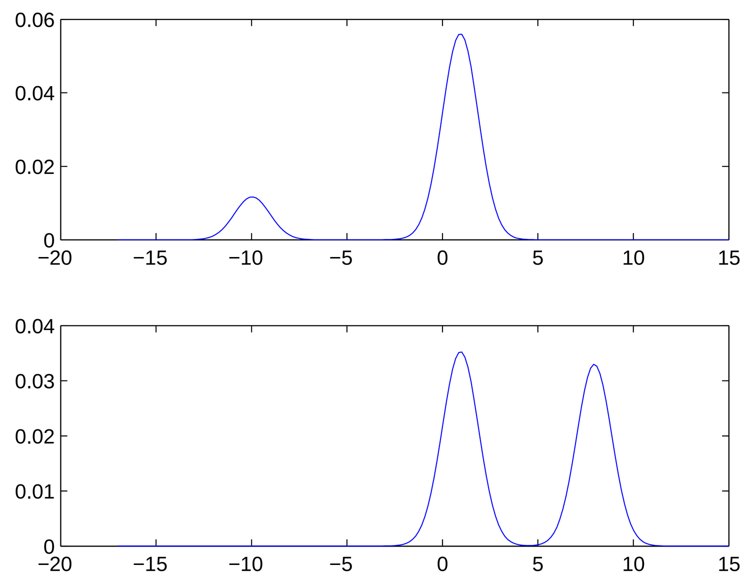

Figure 1: The predictive distribution for $\underline{Y_1}$ and $\underline{Y_2}$

Note also that, in order to have a common weight in the two data sets, this must be between 0 and 0.5, and since the case $\varepsilon = 0.5\ (w = 0.5)$ leads to the most parsimonious allocation of the cases F_j, the Bayesian methodology will tend to favor this one. So,

$Y_{2i} \overset{iid}{\sim} \underbrace{0.2 \cdot \text{N}(-10, 1) + 0.3 \cdot \text{N}(1, 1)}_{F_1} + 0.5 \cdot \underbrace{\text{N}(1, 1)}_{F_0}$. and

$Y_{2i} \overset{iid}{\sim} 0.5 \cdot \underbrace{\text{N}(8, 1)}_{F_2} + 0.5 \cdot \underbrace{\text{N}(1, 1)}_{F_0}$

The predictive distributions for F_0, F_1 and F_2, shown in figure 2 were also as one would expect to be.

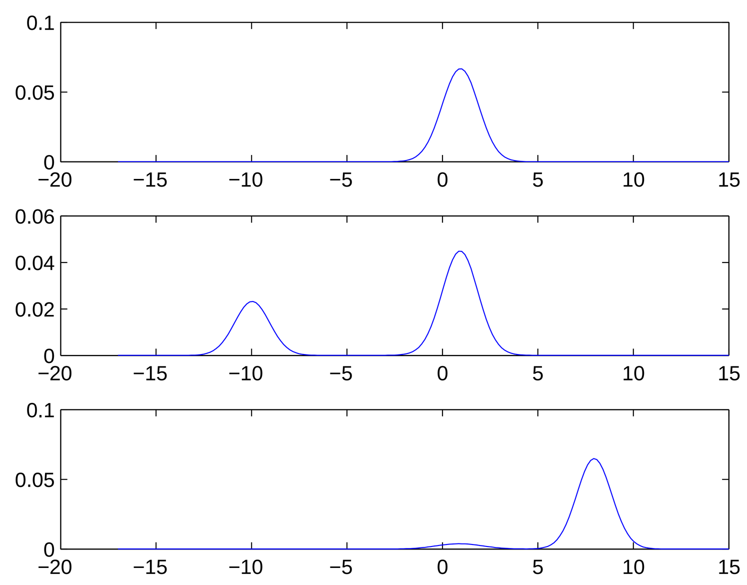

Figure 2: The predictive distribution for F_0, F_1 and F_2

The posterior sample for the weight is centered around the value 0.5, again as expected (figure 3). In this prediction, however, using model (2), we get an additional mode at zero. This drawback is obviously due to the prior of w.

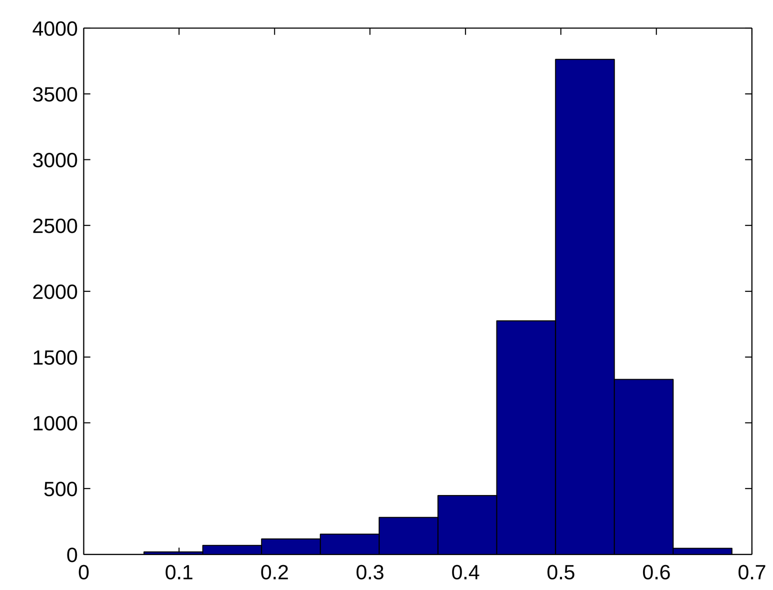

Figure 3: Posterior sample of ε for model 1

Finally, we tried using the first model by fixing the M_j. For small values of them, the results were as before, whereas for larger values (e.g. 25), the results weren't so good.

References

[1] Ferguson, T.S. (1973), A Bayesian Analysis of some nonparametrics problems, *The Annals of Statistics*, **1**, 209-230.

[2] Müller, P., Quintana, F.A. & Rosner, G. (2004), A Method for Combining Inference Across Related Nonparametric Bayesian Models, *Journal of the Royal Statistical Society Series B*, Vol. **66**, No. 3, 735-749.

[3] Lijoi, A., Mena, R.H. & Prünster, I. (2005), Hierarchical Mixture Modelling With Normalized Inverse Gaussian Priors, *Journal of the American Statistical Association*, Vol. **100**, No. 472, 1278-1291.