Sparse Log Gaussian Processes via MCMC for Spatial Epidemiology

Aki Vehtari and Jarno Vanhatalo
Laboratory of Computational Engineering,
Helsinki University of Technology,
P.O.Box 9203, FIN-02015 TKK, Espoo, Finland
{Aki.Vehtari,Jarno.Vanhatalo}@tkk.fi

Abstract

Log Gaussian processes (LGP) are an attractive manner to construct intensity surfaces for the purposes of spatial epidemiology. The intensity surfaces are naturally smoothed by placing a GP prior over the relative log Poisson rate. In this work a fully independent training conditional (FITC) sparse approximation is used to speed up GP computations. The sampling of the latent values is sped up with transformations taking into account the approximate conditional posterior precision.

1 Introduction

Log Gaussian processes (LGP) are an attractive way to construct intensity surfaces for the purposes of spatial epidemiology (see, e.g., Richardson (2003) for review of spatial models for point referenced data). The spatial correlations between areas are included in an explicit and a natural way into the model via a covariance function. The drawback of GP is the computational burden of the required covariance matrix inversion. The computation time becomes prohibitive as the data amount increases up to around a few thousand of cases, limiting the study either to very small areas or a sparsely populated grid. To overcome the computational limitations a number of sparse approximations for GP have been suggested in the literature. Here a fully independent training conditional (FITC) (Snelson and Ghahramani, 2006; Quiñonero-Candela and Rasmussen, 2005) sparse approximation is used to speed up GP computations.

In spatial epidemiology it is very important to have good estimates whether the spatial variation is significant. To set a golden standard for the uncertainty estimates, instead of using faster variational type approximations for the latent values and point estimates for the hyperparameters, we integrate over both the hyperparameters and the latent values using Markov chain Monte Carlo methods (MCMC). The sampling of the latent values is sped up with transformations taking into account the approximate conditional posterior precision. The sparse method is also compared to a full GP approach with two moderate size datasets.

Some results depicting the spatial variations in the relative mortality risk in Finland are illustrated.

2 Models and Methods

The spatial variations in relative mortality risk in a point referenced health-care data are studied with a log Gaussian process with Poisson likelihood. The data is aggregated into areas A_i with co-ordinates $(x_{i,1}, x_{i,2})$. The mortality in an area A_i is modeled as a Poisson process with mean $E_i \mu_i$, where E_i is the standardised expected number of deaths in the area A_i. The complete model is

$$\mathbf{Y} \sim \text{Poisson}(\mathbf{E}\mu) \tag{1}$$

$$\log(\mu) = \mathbf{f}(\mathbf{x}_i, \mathbf{x}_j) \sim \mathcal{GP}(\mathbf{0}, \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j)), \tag{2}$$

where the relative log rate $\log(\mu)$ is given a Gaussian process prior with zero mean and a squared exponential covariance function described as (Rasmussen and Williams, 2006)

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma^2_{\text{exp}} \exp\left(-\frac{1}{l^2} \sum_{p=1}^{P} (x_{i,p} - x_{j,p})^2 \right). \tag{3}$$

The covariance function parameters, the characteristic length-scale l and the signal magnitude σ^2_{exp}, are given a half Student's-t prior.

The Gaussian process prior presented in equation (2) is approximated by a fully independent training conditional (FITC) sparse approximation in order to speed up the computations. In the FITC a new set of latent variables $\mathbf{u} = [u_1, ..., u_m]^{\text{T}}$, called the inducing variables, are used to determine the inducing conditionals

$$q_{\text{FITC}}(\mathbf{f}|\mathbf{u}) = N\left(\mathbf{K}_{\text{f,u}}\mathbf{K}^{-1}_{\text{u,u}})\mathbf{u}, \text{diag}\left[\mathbf{K}_{\text{f,f}} - \mathbf{Q}_{\text{f,f}}\right]\right) \tag{4}$$

$$q_{\text{FITC}}(\mathbf{f}_*|\mathbf{u}) = N(\mathbf{K}_{*,\text{u}}\mathbf{K}^{-1}_{\text{u,u}})\mathbf{u}, \mathbf{K}_{*,*} - \mathbf{Q}_{*,*}), \tag{5}$$

where $\mathbf{Q}_{\text{a,b}} = \mathbf{K}_{\text{a,u}}\mathbf{K}^{-1}_{\text{u,u}}\mathbf{K}_{\text{u,b}}$, and $\mathbf{f}$ and $\mathbf{f}_*$ represent the training and the test latent values respectively.

Both the hyperparameters and the latent values are sampled with hybrid Monte Carlo (HMC) method separately. In FITC approximation the gradient evaluations, required in HMC, are constructed in such a manner that the explicit evaluation of the full covariance matrix is avoided.

The sampling for the latent values f is sped up by transformation using a matrix square root of an approximate posterior covariance matrix similar to Christensen et al. (2006) and conducting HMC dynamics in the resulting $\tilde{\mathbf{f}} = \Sigma^{-1/2}\mathbf{f}$ space. The approximate posterior precision, $\Sigma^{-1} = \mathbf{K}^{-1} + \Sigma_{\text{l}}^{-1}$, is obtained as the sum of the precisions of the prior and the likelihood, where the covariance of the likelihood is approximated as $\Sigma_{\text{l}} \approx -\frac{\partial^2 \log(\text{Poisson}(E\lambda))}{\partial f^2} = E\mu$, and μ is approximated with its prior mean 1.

In the FITC approximation $\mathbf{Q}_{\text{f,f}} + \Lambda$, where $\Lambda = \text{diag}\left[\mathbf{K}_{\text{f,f}} - \mathbf{Q}_{\text{f,f}}\right]$, replaces the prior covariance $\mathbf{K}$, and the posterior precision transforms to $\Sigma^{-1}_{\text{FITC}} = \left(\mathbf{Q}_{\text{f,f}} + \Lambda\right)^{-1} + \Sigma_{\text{l}}^{-1}$. To extend the transformation into the FITC approximation a matrix inversion lemma can be used to form a matrix square root of $\Sigma^{-1}_{\text{FITC}}$ and to construct the following transformation equations:

$$\mathbf{USU}^{\mathrm{T}} = \hat{\Lambda}^{1/2}\Lambda^{-1}\mathbf{K}_{\mathrm{f,u}}\left(\mathbf{K}_{\mathrm{u,u}} + \mathbf{K}_{\mathrm{u,f}}\Lambda^{-1}\mathbf{K}_{\mathrm{f,u}}\right)^{-1}\mathbf{K}_{\mathrm{u,f}}\Lambda^{-1}\hat{\Lambda}^{1/2} \quad (6)$$

$$\mathbf{f} = \hat{\Lambda}^{1/2}(\tilde{\mathbf{f}} + \mathbf{UD}^{-1}\mathbf{U}^{\mathrm{T}}\tilde{\mathbf{f}} - \mathbf{UU}^{\mathrm{T}}\tilde{\mathbf{f}}) \quad (7)$$

$$\tilde{\mathbf{f}} = \hat{\Lambda}^{-1/2}\mathbf{f} + \mathbf{UDU}^{\mathrm{T}}\hat{\Lambda}^{-1/2}\mathbf{f} - \mathbf{UU}^{\mathrm{T}}\hat{\Lambda}^{-1/2}\mathbf{f}, \quad (8)$$

where $\mathbf{U}$ and $\mathbf{S}$ are matrices of eigenvectors and eigenvalues of the right hand side of the Eq. (6) respectively. $\mathbf{D}_{ii} = \sqrt{1 - \mathbf{S}_{ii}}$ and $\hat{\Lambda} = \left(\Sigma_l^{-1} + \Lambda^{-1}\right)^{-1}$.

3 Results

The spatial population and mortality data used in the study were obtained from Statistics Finland. The reference population representing the population at risk was formed of the 1995 population. This included approximately 4.9 million people in the age range of 0–110 years after insufficient records were removed. The mortality data had been collected during 1995–2000. The standardised expected number of deaths E_i was computed using the reference population with covariates age, sex, and education in each area A_i. Two types of mortality were studied: the mortality due to cerebral vascular diseases with roughly 18 000 deaths and the mortality due to alcohol-related diseases bringing forth around 5200 deaths. The data was aggregated into a grid cell size of 20 km x 20 km.

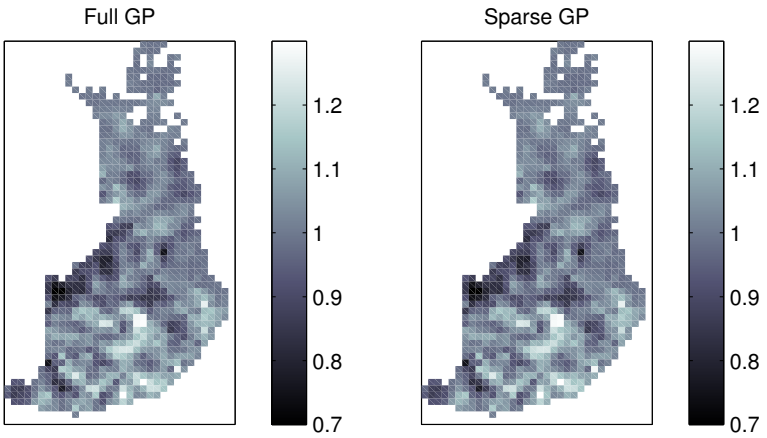
Both datasets consisted of 915 data points. The inducing variables for the sparse model were set uniformly over the grid giving total of 221 inducing points. The sampling took a couple of hours and the full model was approximately two times slower than the sparse model. The posterior mean of the relative risk μ for two datasets with full and sparse GP are show in Fig.1. The posterior probability that the risk differs from the national average can also be evaluated from the model, but is not shown due to space restrictions. The results are at the moment preliminary. The code is written in Matlab and it is not yet fully optimized.

4 Conclusion

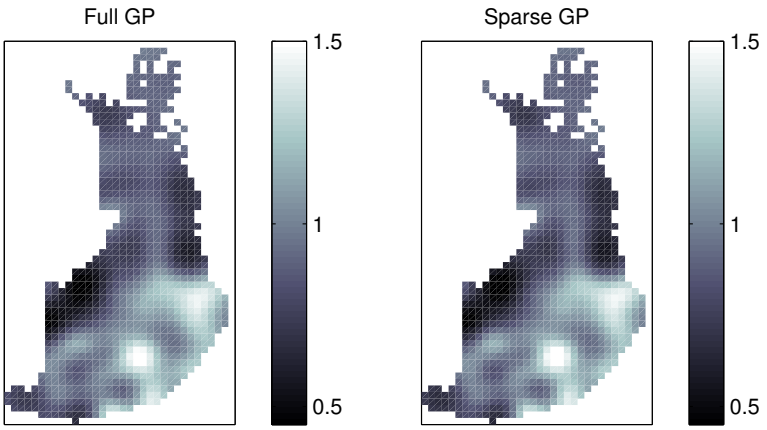
We constructed a log Gaussian process with a FITC sparse approximation with uniformly distributed inducing points and a full model. The results were compared with two sets of health care data and found similar. The simulation time for FITC approximation was about half the time of full model. As a future development, we will study practical limit of the number of regions which can be handled, sampling of the locations of inducing points, various covariance functions, and accuracy of variational type approximations for marginalizing over latent values.

Acknowledgments

Authors would like to thank Harri Valpola for helpfull comments on latent variable transformation and Markus Siivola for data manipulation.



(a) Cerebral vascular diseases. The length-scale is approximately 30km.



(b) Alcohol-related diseases. The length-scale is approximately 74km.

Figure 1: The posterior mean of the relative risk μ estimated with full GP (left) and sparse GP (right).

References

Christensen, O. F., Roberts, G. O., and Sköld, M. (2006). Robust Markov chain Monte Carlo methods for spatial generalised linear mixed models. *Journal of Computational and Graphical Statistics*, 15:1–17.

Quiñonero-Candela, J. and Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *Journal Of Machine Learning Research*, 6(3):1939–1959.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

Richardson, S. (2003). Spatial models in epidemiological applications. In Green, P. J., Hjort, N. L., and Richardson, S., editors, *Highly Structured Stochastic Systems*, pp. 237–259. Oxford University Press.

Snelson, E. and Ghahramani, Z. (2006). Sparse Gaussian Process Using Pseudo-inputs. *Advances in Neural Information Processing Systems*, 18.