Sparse Log Gaussian Processes via MCMC for Spatial Epidemiology

Aki Vehtari and Jarno Vanhatalo
Helsinki University of Technology
P.O.Box 9203, FI-02015, HUT, Finland

Introduction

In this work a fully independent training conditional (FITC) sparse approximation is used to speed up GP computations in the study of the spatial variations in relative mortality risk in a point referenced health-care data. The sampling of the latent values is sped up with transformations taking into account the approximate posterior precision.

Log Gaussian processes (LGP) are an attractive way to construct intensity surfaces for the purposes of spatial epidemiology. 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 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) sparse approximation is used to speed up GP computations.

To set a golden standard for the uncertainty estimates, we integrate over both the hyperparameters and the latent values using Markov chain Monte Carlo method (MCMC). The sampling the latent values is sped up with transformations taking into account the approximate conditional posterior precision.

Model

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 \( A \) with co-ordinates \((x_i, y_i)\). The mortality in an area \( A_i \) is modeled as a Poisson process with mean \( \lambda_i \), where \( \lambda_i \) is the standardized expected number of deaths in the area \( A_i \). The complete model is

\[
Y \sim \text{Poisson}(\lambda_i) \\
\log(\lambda_i) = \mu + \sigma_u^2 \exp \left( \frac{1}{\tau} \sum_{j \neq i} (x_j - x_i)^2 + (y_j - y_i)^2 \right),
\]

where the relative log rate \( \log(\lambda_i) \) is given a Gaussian process prior with zero mean and a squared exponential covariance function described as

\[
\kappa(x_i, x_j) = \sigma_u^2 \exp \left( -\frac{1}{\tau} \sum_{d \in \{x, y\}} (x_d - x_d)^2 \right).
\]

The covariance function parameters, the characteristic length-scale \( \tau \) and the signal magnitude \( \sigma_u^2 \), are given a half Student's-\( t \) prior.

The Gaussian process prior (Eq. (1)) is characterized by a fully independent training conditional (FITC) sparse approximation in order to speed up the computations. In the FITC a new set of \( n_p \) sparse inducing points \( \{x_1, \ldots, x_n_p\} \) are defined. The inducing variables, are used to determine the inducing conditioning

\[
Q_{\text{FITC}}(\mathbf{Q}) = N \left( K_1 \mathbf{K}_n \mathbf{K}_n \mathbf{q}, \text{diag} \left[ \mathbf{K}_n \mathbf{k}_n - \mathbf{Q}_1 \right] \right),
\]

\[
Q_{\text{FITC}}(\mathbf{f}) = \mathbf{N} \left( \mathbf{K}_n \mathbf{K}_n \mathbf{f}, \mathbf{k}_n - \mathbf{Q}_1 \right),
\]

where \( \mathbf{k}_n \) is the size of \( \mathbf{K}_n \), \( \mathbf{Q}_1 \) and \( \mathbf{f} \) represent the training and the test latent values respectively.

Transformation with approximate posterior variance

Both the hyperparameters and the latent values are sampled with hybrid Monte Carlo (HMC) method separately. The sampling for the latent values \( \mathbf{f} \) is sped up by transformation using a matrix square root of an approximate posterior covariance matrix.

The HMC dynamics is conducted in the space \( \Sigma^{-1/2} \) resulting from the transformation. The approximate posterior precision, \( \Sigma^{-1} = \mathbf{K}^{-1} + \Sigma^{-1}_i \), is obtained as the sum of the precisions of the prior and the likelihood, where the precision of the likelihood is approximated as \( \sum_i = \text{diag}(\mathbf{K}_i) + \Sigma^{-1}_i \).

In the FITC approximation \( Q_{\text{FITC}}(\mathbf{f}) \), where \( \Lambda = \text{diag}(\mathbf{K}_i - \mathbf{Q}_1) \), replaces the prior covariance \( \mathbf{K} \), and the posterior precision transforms to \( \Sigma^{-1} = (\mathbf{K} + \mathbf{Q}_1)^{-1} + \Lambda^{-1} \). To extend the transformation into the FITC approximation a matrix inversion lemma can be used to form a matrix square root of \( \Sigma^{-1} \), and to construct the following transformation equations:

\[
\begin{align*}
\mathbf{u} & = \Lambda^{1/2} \mathbf{K}^{-1/2} \mathbf{k} - \mathbf{k}^{\dagger} \mathbf{K}^{-1} \mathbf{k}, \\
\mathbf{f} & = \Lambda^{1/2} \mathbf{K}_1^{-1/2} \mathbf{K}_n - \mathbf{K}_n \mathbf{K}_1^{-1/2} \mathbf{K}_n, \\
\mathbf{f} & = \Lambda^{1/2} \mathbf{K}_1^{-1/2} \mathbf{K}_n - \mathbf{K}_n \mathbf{K}_1^{-1/2} \mathbf{K}_n.
\end{align*}
\]

\[
\begin{align*}
\mathbf{u} & = \Lambda^{1/2} \mathbf{K}^{-1/2} \mathbf{k} - \mathbf{k}^{\dagger} \mathbf{K}^{-1} \mathbf{k}, \\
\mathbf{f} & = \Lambda^{1/2} \mathbf{K}_1^{-1/2} \mathbf{K}_n - \mathbf{K}_n \mathbf{K}_1^{-1/2} \mathbf{K}_n, \\
\mathbf{f} & = \Lambda^{1/2} \mathbf{K}_1^{-1/2} \mathbf{K}_n - \mathbf{K}_n \mathbf{K}_1^{-1/2} \mathbf{K}_n.
\end{align*}
\]

where \( \mathbf{K} \) and \( \mathbf{k} \) are matrices of eigenvectors and eigenvalues of the right hand side of the Eq. (6) respectively. \( \mathbf{U} = \sqrt{\mathbf{K}^{\dagger}} \mathbf{K}^{-1/2} \mathbf{k} \) and \( \Lambda = \left( \Sigma^{-1} + \Lambda^{-1} \right)^{-1} \).

Figure 1 illustrates the need of scaling in the two problems presented here. The alcohol-related diseases are more rare than the Cerebral vascular diseases. This results in fewer large scale eigenvalues in alcohol-related diseases than in Cerebral vascular diseases. The size differences of eigenvalues are also larger in alcohol-related diseases.

Results

The results of FITC approximation were compared with two sets of health care data and found similar to full model. The sampling time for FITC approximation was about half the time of full model.

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.

Future Development

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 variables.