Max-and-Smooth: a two-step approach for approximate Bayesian inference in latent Gaussian models

Birgir Hrafnkelsson

University of Iceland

November 18, 2019

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES

Max-and-Smooth: a two-step approach for approximate Bayesian inference in latent Gaussian models

Develop a novel posterior inference scheme for LGMs.
Applications:
(i) linear regression on a lattice
(ii) a spatial LGM for annual maximum flow in rivers

Participants:
Birgir Hrafnkelsson - University of Iceland
Árni Víðir Jóhannesson - University of Iceland
Stefan Siegert - University of Exeter
Haakon Bakka - KAUST
Raphaël Huser - KAUST

Introduction

Data level

$$y_{it} \sim \mathcal{F}(\mu_i, \sigma_i), \quad t = 1, ..., T, \quad i = 1, ..., J$$

Multivariate link function

$$(\psi_i, \tau_i) = h(\mu_i, \sigma_i) = (\log(\mu_i), \log(\sigma_i/\mu_i))$$

Latent level

$$\psi_i = \boldsymbol{x}_{\psi i}\boldsymbol{\beta}_\psi + \boldsymbol{a}_{\psi i}\boldsymbol{u}_\psi + \epsilon_{\psi i}$$

$$\tau_i = \boldsymbol{x}_{\tau i}\boldsymbol{\beta}_\tau + \boldsymbol{a}_{\tau i}\boldsymbol{u}_\tau + \epsilon_{\tau i}$$

Assign Gaussian prior densities to:

$$\boldsymbol{\beta}_\psi, \boldsymbol{u}_\psi, \boldsymbol{\epsilon}_\psi, \boldsymbol{\beta}_\tau, \boldsymbol{u}_\tau, \boldsymbol{\epsilon}_\tau$$

Introduction

Data level: $\boldsymbol{y} \leftarrow$ data $\quad \boldsymbol{\eta} \leftarrow \psi_i, \tau_i$

$$p(\boldsymbol{y}|\boldsymbol{\eta})$$

Latent level: $\boldsymbol{\nu} = (\boldsymbol{\beta}_\psi^\top, \boldsymbol{u}_\psi^\top, \boldsymbol{\beta}_\tau^\top, \boldsymbol{u}_\tau^\top)^\top$

$$p(\boldsymbol{\eta}, \boldsymbol{\nu}|\boldsymbol{\theta}) = p(\boldsymbol{\nu}|\boldsymbol{\theta})p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta})$$

Hyperparameter level: $\boldsymbol{\theta} \leftarrow$ hyperparameters

$$p(\boldsymbol{\theta})$$

Posterior density:

$$p(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{\nu}|\boldsymbol{\theta})p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{\eta})$$

Introduction

Likelihood function:

$$L(\boldsymbol{\eta}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\eta}) \quad \text{(as a fcn of } \boldsymbol{\eta}\text{)}$$

Hessian matrix and mode of $\log(L(\boldsymbol{\eta}|\boldsymbol{y}))$:

$$-Q_{\hat{\eta}y} \quad \& \quad \hat{\boldsymbol{\eta}}$$

Gaussian approximation:

$$L(\boldsymbol{\eta}|\boldsymbol{y}) \approx \tilde{L}(\boldsymbol{\eta}|\boldsymbol{y}) = c\mathcal{N}(\boldsymbol{\eta}|\hat{\boldsymbol{\eta}}, Q_{\hat{\eta}y}^{-1}), \quad c \text{ is a constant indep. of } \boldsymbol{\eta}$$

Approximated posterior density:

$$\begin{aligned} p(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y}) &\approx \tilde{p}(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\boldsymbol{y}) \\ &\propto p(\boldsymbol{\theta})p(\boldsymbol{\nu}|\boldsymbol{\theta})p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta})\tilde{L}(\boldsymbol{\eta}|\boldsymbol{y}) \end{aligned}$$

Introduction

Data level:

$$p(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}) = \mathcal{N}(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}, Q_{\hat{\eta}y}^{-1})$$

Latent level:

$$p(\boldsymbol{\eta}, \boldsymbol{\nu}|\boldsymbol{\theta}) = p(\boldsymbol{\nu}|\boldsymbol{\theta})p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta})$$

Hyperparameter level:

$$p(\boldsymbol{\theta})$$

Posterior density:

$$\begin{aligned} p(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\hat{\boldsymbol{\eta}}) &\propto p(\boldsymbol{\theta})p(\boldsymbol{\nu}|\boldsymbol{\theta})p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta})p(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}) \\ &\propto p(\boldsymbol{\theta})p(\boldsymbol{\nu}|\boldsymbol{\theta})p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta})\tilde{L}(\boldsymbol{\eta}|\boldsymbol{y}) \end{aligned}$$

as

$$p(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}) = \mathcal{N}(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}, Q_{\hat{\eta}y}^{-1}) = \mathcal{N}(\boldsymbol{\eta}|\hat{\boldsymbol{\eta}}, Q_{\hat{\eta}y}^{-1}) = c^{-1}\tilde{L}(\boldsymbol{\eta}|\boldsymbol{y})$$

Introduction

Analyze the latter model as that is a Gaussian-Gaussian model.

Data level:

$$p(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}) = \mathcal{N}(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}, Q_{\hat{\eta}y}^{-1})$$

Latent level:

$$p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\eta}|Z\boldsymbol{\nu}, Q_{\epsilon}^{-1})$$

$$p(\boldsymbol{\nu}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\nu}|\boldsymbol{\mu}_{\nu}, Q_{\nu}^{-1})$$

Hyperparameter level:

$$p(\boldsymbol{\theta})$$

Posterior density:

$$p(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta}|\hat{\boldsymbol{\eta}}) \propto p(\boldsymbol{\theta})\mathcal{N}(\boldsymbol{\nu}|\boldsymbol{\mu}_{\nu}, Q_{\nu}^{-1})\mathcal{N}(\boldsymbol{\eta}|Z\boldsymbol{\nu}, Q_{\epsilon}^{-1})\mathcal{N}(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}, Q_{\hat{\eta}y}^{-1})$$

Introduction

The posterior density of the pseudo model is

$$p(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} | \hat{\boldsymbol{\eta}}) \propto p(\boldsymbol{\theta}) p(\boldsymbol{\nu} | \boldsymbol{\theta}) p(\boldsymbol{\eta} | \boldsymbol{\nu}, \boldsymbol{\theta}) p(\hat{\boldsymbol{\eta}} | \boldsymbol{\eta}).$$

The approximated posterior density of the original model is

$$\tilde{p}(\boldsymbol{\eta}, \boldsymbol{\nu}, \boldsymbol{\theta} | \boldsymbol{y}) \propto p(\boldsymbol{\theta}) p(\boldsymbol{\nu} | \boldsymbol{\theta}) p(\boldsymbol{\eta} | \boldsymbol{\nu}, \boldsymbol{\theta}) \tilde{L}(\boldsymbol{\eta} | \boldsymbol{y}).$$

These two are the same since

$$\tilde{L}(\boldsymbol{\eta} | \boldsymbol{y}) = p(\hat{\boldsymbol{\eta}} | \boldsymbol{\eta}).$$

Hierarchical linear models

- Data level

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_1) = \mathcal{N}(\boldsymbol{y}|Z\boldsymbol{x}, Q_\epsilon^{-1})$$

- Latent level

$$p(\boldsymbol{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\boldsymbol{y}|\boldsymbol{\mu}_x, Q_x^{-1})$$

- Hyperparameter level

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_1)p(\boldsymbol{\theta}_2)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top \; \boldsymbol{\theta}_2^\top)^\top$

Hierarchical linear model

The Gaussian prior densities of $\boldsymbol{y} \mid \boldsymbol{x}$ and $\boldsymbol{x}$ are

$$p\left(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}_1\right) = \mathcal{N}(\boldsymbol{y} \mid Z\boldsymbol{x}, Q_\epsilon^{-1})$$
$$p\left(\boldsymbol{x} \mid \boldsymbol{\theta}_2\right) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_x, Q_x^{-1})$$

The joint prior distribution $(\boldsymbol{y}^\top, \boldsymbol{x}^\top)^\top$ becomes

$$p(\boldsymbol{y}, \boldsymbol{x} \mid \boldsymbol{\theta}) = \mathcal{N}\left(\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} \middle| \begin{pmatrix} Z\boldsymbol{\mu}_x \\ \boldsymbol{\mu}_x \end{pmatrix}, \begin{pmatrix} Q_\epsilon & -Q_\epsilon Z \\ -Z^\top Q_\epsilon & Q_x + Z^\top Q_\epsilon Z \end{pmatrix}^{-1}\right)$$

Hierarchical linear model

The conditional posterior distribution of $\boldsymbol{x}$ (given $\boldsymbol{\theta}$ and $\boldsymbol{y}$) in terms of precision matrices:

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_{x|y}, Q_{x|y}^{-1})$$

where

$$\boldsymbol{\mu}_{x|y} = Q_{x|y}^{-1}(Q_x \boldsymbol{\mu}_x + Z^{\mathsf{T}} Q_\epsilon \boldsymbol{y})$$

$$Q_{x|y} = Q_x + Z^{\mathsf{T}} Q_\epsilon Z$$

Hierarchical linear model

The marginal distribution of $\boldsymbol{y}$ (given $\boldsymbol{\theta}$) in terms of precision matrices:

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{\theta})p(\boldsymbol{y} \mid \boldsymbol{\theta})$$

$$p(\boldsymbol{y} \mid \boldsymbol{\theta}) = \frac{p(\boldsymbol{y}, \boldsymbol{x} \mid \boldsymbol{\theta})}{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} = \frac{p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta})p(\boldsymbol{x} \mid \boldsymbol{\theta})}{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \quad \text{(indep. of } \boldsymbol{x}\text{)}$$

where

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_{x|y}, (Q_x + Z^\mathsf{T} Q_\epsilon Z)^{-1})$$

$$p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y} \mid Z\boldsymbol{x}, Q_\epsilon^{-1})$$

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}_x, Q_x^{-1})$$

Posterior simulation for hierarchical linear models

Simulate samples from $p(\boldsymbol{x}, \boldsymbol{\theta} \mid \boldsymbol{y})$

1. Draw $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta} \mid \boldsymbol{y})$

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto p(\boldsymbol{\theta}) p(\boldsymbol{y} \mid \boldsymbol{\theta}) = p(\boldsymbol{\theta}) \frac{p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) p(\boldsymbol{x} \mid \boldsymbol{\theta})}{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})}$$

2. Draw $\boldsymbol{x}$ from $p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x} | \boldsymbol{\mu}_{x|y}, Q_{x|y}^{-1})$$

where

$$\boldsymbol{\mu}_{x|y} = Q_{x|y}^{-1}(Q_x \boldsymbol{\mu}_x + Z^\mathsf{T} Q_\epsilon \boldsymbol{y})$$

$$Q_{x|y} = Q_x + Z^\mathsf{T} Q_\epsilon Z$$

Hierarchical linear models

- Data level

$$p(\boldsymbol{y}|\boldsymbol{\eta}) = \mathcal{F}(\boldsymbol{y}|\boldsymbol{\eta})$$

- Latent level

$$p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\eta}|Z\boldsymbol{\nu}, Q_{\epsilon}^{-1})$$

$$p(\boldsymbol{\nu}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\nu}|\boldsymbol{\mu}_{\nu}, Q_{\nu}^{-1})$$

- Hyperparameter level

$$p(\boldsymbol{\theta})$$

HIERARCHICAL LINEAR MODELS

- Data level - pseudo model

$$p(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}) = \mathcal{N}(\hat{\boldsymbol{\eta}}|\boldsymbol{\eta}, Q_{\hat{\eta}y}^{-1})$$

- Latent level

$$p(\boldsymbol{\eta}|\boldsymbol{\nu}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\eta}|Z\boldsymbol{\nu}, Q_{\epsilon}^{-1})$$

$$p(\boldsymbol{\nu}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\nu}|\boldsymbol{\mu}_{\nu}, Q_{\nu}^{-1})$$

- Hyperparameter level

$$p(\boldsymbol{\theta})$$