Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations

Håvard Rue, Sara Martino, and Nicolas Chopin

Aaron E. Zimmerman

Department of Statistics
University of Washington - Seattle

April 30, 2013

Latent Gaussian Models

- Assume y_i belongs to exponential family
- Let $E[y_i] = \mu_i$ be linked to a η_i: $g(\mu_i) = \eta_i$

η_i: Structured Additive Predictor

$$\eta_i = \alpha + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \epsilon_i.$$

- ϵ_is are unstructured terms
- $f^{(j)}$s are unknown functions of the u_{ji}s
- β_ks are linear effects of z_{ki}s
- Define $\mathbf{x}$ as all $\eta_i, \{f^{(j)}\}, \beta_k$, and α such that
  $$\pi(\mathbf{x}|\boldsymbol{\theta}) \sim N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta}))$$
- $\boldsymbol{\theta}$ are hyperparameters to which we can assign priors

The Laplace Approximation

- Assume we have a distribution of the following form:

$$p(y) \propto e^{-m*h(y)}b(y).$$

- We'll use the Laplace method to find expectations:

$$E[q(y)] = \frac{\int q(y)b(y)e^{-mh(y)}dy}{\int b(y)e^{-mh(y)}dy}.$$

Let $x = \sqrt{m}(y - \hat{y})$, where $\hat{y}$ is the mode of $q(y)$.

The Laplace Approximation for Expectations

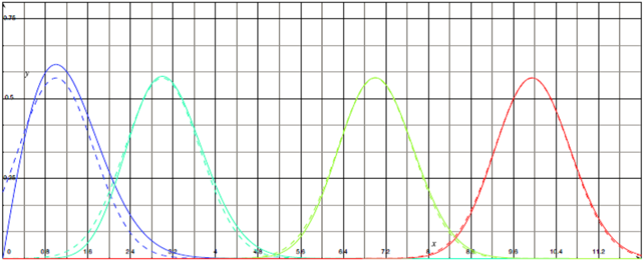
- We expand q, b, and h around $\hat{y}$

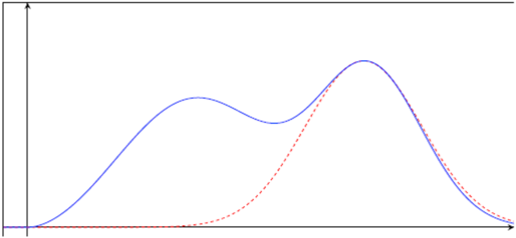
$$E[q(y)] = \frac{\int e^{-\frac{h''(\hat{y})x^2}{2}}[(q(\hat{y}) + \frac{xq'(\hat{y})}{\sqrt{m}} + \ldots)(b(\hat{y}) + \frac{xb'(\hat{y})}{\sqrt{m}} + \ldots)exp\{\frac{-h'''(\hat{y})x^3}{6\sqrt{m}} + \ldots\}]dx}{\int e^{-\frac{h''(\hat{y})x^2}{2}}[(b(\hat{y}) + \frac{xb'(\hat{y})}{\sqrt{m}} + \ldots)exp\{\frac{-h'''(\hat{y})x^3}{6\sqrt{m}} + \ldots\}]dx}$$

- So we can approximate this expectation by using a $\mathcal{N}(0, [h''(\hat{y})]^{-1})$ distribution.
- After some more algebra (and expansions of $\frac{1}{\sqrt{m}}$), you can get the Laplace Approximation:

$$E[q(y)] \approx q(\hat{y}) + \frac{1}{m}\left(\frac{q'(\hat{y})}{h''(\hat{y})}\left[\frac{b'(\hat{y})}{b(\hat{y})} - \frac{h'''(\hat{y})}{2h''(\hat{y})}\right] + \frac{q'(\hat{y})}{h''(\hat{y})}\right).$$

Density Approximations





courtesy of Dr. Cevher Rice University

INLA in 3 steps

Step 1

- Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$ with a Laplace approximation:

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left.\frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}\right|_{\mathbf{x}=\mathbf{x}^*(\theta)}$$

- Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ with a Laplace approximation
- Numerically integrate out $\boldsymbol{\theta}$ from $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ to approximate $\pi(x_i|\mathbf{y})$

Example - Gaussian Approximation

$$\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) = \mathcal{N}(x_i; \mu_i(\boldsymbol{\theta}), \sigma_i^2(\boldsymbol{\theta}))$$

INLA in 3 steps

Step 2

- Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$ with a Laplace approximation
- Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ with a Laplace approximation:

$$\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \left. \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} \right|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(\theta)}$$

- Numerically integrate out $\boldsymbol{\theta}$ from $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ to approximate $\pi(x_i|\mathbf{y})$

INLA in 3 steps

Step 3

- Approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$ with a Laplace approximation
- Approximate $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ with a Laplace approximation
- Numerically integrate out $\boldsymbol{\theta}$ from $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ to approximate $\pi(x_i|\mathbf{y})$:

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k$$

Obtaining $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$

Rearranging $\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y}) = \pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) * \pi(\boldsymbol{\theta}|\mathbf{y}) * \pi(\mathbf{y})$, yields that

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}.$$

Our approximation is then:

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})},$$

where $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional.

This is an application of the Laplace methods for integration.

Obtain $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ for $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$

By assumption, the latent field is a GMRF. As a result:

$$\pi(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \sum \log\pi(y_i|x_i)\right),$$

we match the mode and the curvature at the mode to produce the Gaussian approximation:

$$\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{x}^*)^T(\mathbf{Q}+\mathrm{diag}(c))(\mathbf{x}-\mathbf{x}^*)\right),$$

where $\mathbf{Q}$, c, and $\mathbf{x}^*$ (the mode), are functions of $\boldsymbol{\theta}$.

Exploring $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$

We mainly need $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ to integrate out uncertainty with respect to $\boldsymbol{\theta}$. We use grid exploration, nothing parametric.

- Locate mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ via optimization of $\log\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$. Call it $\boldsymbol{\theta}^*$.
- At $\boldsymbol{\theta}^*$ numerically compute the negative Hessian, $\mathbf{H} > 0$, and define $\boldsymbol{\Sigma} = \mathbf{H}^{-1}$.
- We explore the $\boldsymbol{\theta}-$space on the standardized $\mathbf{z}$-axes.

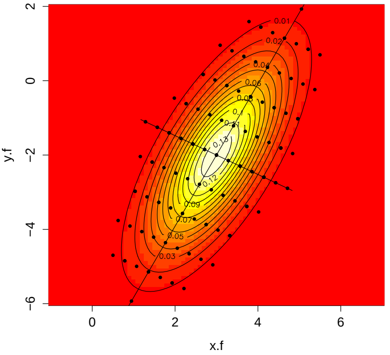
$$\boldsymbol{\theta}(\mathbf{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Delta}^{1/2}\mathbf{z},$$

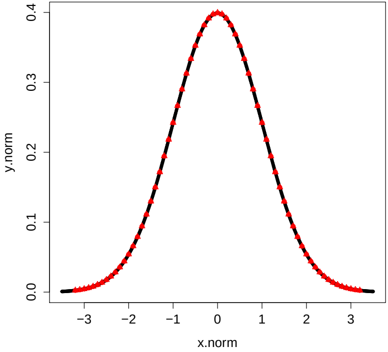
where $\boldsymbol{\Sigma} = \mathbf{V}\boldsymbol{\Delta}\mathbf{V}^T$.

Exploring $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$

Step out along the axes with step size δ_z until the density is too small. We fill in the grid the same way. i.e. until

$$log[\tilde{\pi}(\mathbf{0}|\mathbf{y})] - log[\tilde{\pi}(\boldsymbol{\theta}(\mathbf{z})|\mathbf{y})] < \delta_\pi.$$





Denote the saved grid locations as $\boldsymbol{\theta}_k$.

Obtaining $\tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})$

Unfortunately, the equivalent Laplace approximation

$$\tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y}) \propto \left.\frac{\pi(\mathbf{x}, \boldsymbol{\theta}_k, \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}_k, \mathbf{y})}\right|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta}_k)}$$

is expensive, though it can be done. Instead, two modifications are proposed to speed up computation.

First, we approximate the modal configuration by

$$\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta}) \approx E_{\tilde{\pi}_G}[\mathbf{x}_{-i}|x_i].$$

Obtaining $\tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})$

Second, we assume that only x_j near to x_i have an impact. This eventually leads to a faster Laplace approximation:

$$\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}_k, \mathbf{y}) \propto \mathcal{N}\{x_i, \mu_i(\boldsymbol{\theta}_k), \sigma_i^2(\boldsymbol{\theta}_k)\} * \exp(\text{cubic spline}(x_i)),$$

where the cubic spline is fit to

$$\log\tilde{\pi}_{LA}(x_i|\boldsymbol{\theta}_k, \mathbf{y}) - \log\tilde{\pi}_G(x_i|\boldsymbol{\theta}_k, \mathbf{y}).$$

Putting it all together

Finally we obtain the marginals of interest for the latent field by numerical integration.

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k$$

will do it.

What's Next

- Now that you've suffered through the methodology, next time there will be some examples!
- Comparison of INLA to MCMC - e.g. multimodal situations
- Errors in INLA

Thanks everyone!

References I

Besag et al. (1991) *Bayesian Image Restoration with two applications in spatial statistics*

Dey et al. (2000) *Generalized linear models: a bayesian perspective*

Hinton and van Camp. (1993) *Keeping the neural networks simple by minimizing the description length of the weights*

Kitagawa and Gersch. (1996) *Smoothness priors analysis of time series*

Minka. (2001) *Expectation propagation for approximate bayesian inference*

Rue, Martino, Chopin. (2008) *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*

Tierney and Kadane. (1986) *Accurate approximations for posterior moments and marginal densities*

Wakefield. (2007) *Disease mapping and spatial regression with count data*