# 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 11, 2013

# Latent Gaussian Models

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

$\eta_i$: Structured Additive Predictor

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

- Define $\mathbf{x}$ as all $\eta_i, \{f^{(j)}\}, \beta_k$, and $\alpha$ 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

# Possible Applications

These models are *very* flexible

- Regression Models: $\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k z_{ki}$
  e.g. [2]
- Dynamic Models: include temporal dependence by defining $f(\cdot)$ and $\mathbf{u}$ such that $f(u_t) = f_t$
  e.g. [4]
- Spatial Models: include spatial dependence by defining $f(\cdot)$ and $\mathbf{u}$ such that $f(u_s) = f_s$
  e.g. [1]

# An example: London Suicides

Suicide mortality rate in 32 London boroughs (1989-1993)

$$y_i \sim \text{Poisson}(\lambda_i),\ \lambda_i = \rho_i E_i$$

Linear predictor on log scale:

$$\eta_i = log(\rho_i) = \alpha + \mu_i + \nu_i$$

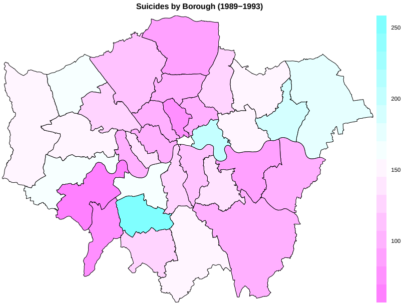
Besag-York-Mollie [1] :

$$\mu_i | \boldsymbol{\mu}_{j \neq i} \sim N(m_i, s_i^2)$$

$$m_i = \frac{\sum_{j \in N(i)} \mu_j}{\#N(i)},\ s_i^2 = \frac{\sigma_\mu^2}{\#N(i)}$$

Unstructured residuals:

$$\nu_i \sim N(0, \sigma_\nu^2)$$



# What we want

- In terms of the original formulation, $\mu_i = f_1(i)$ and $\nu_i = f_2(i)$, are two area specific effects
- We can assume priors to the hyperparameters e.g. $\tau_\mu, \tau_\nu \sim logGamma(1, .005)$

We would like

- Posteriors for the parameters: $\pi(\mathbf{x}|\mathbf{y})$
- Posteriors for the hyperparameters: $\pi(\boldsymbol{\theta}|\mathbf{y})$

# Possible Approaches

- Monte Carlo Markov Chains: construct a markov chain that has the desired distribution (posteriors of the parameters) as the equilibrium distribution
- Variational Bayes [3]: approximate the joint density of $p(\mathbf{x}, \boldsymbol{\theta})$ by minimizing the Kullback-Leibler contrast of $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$ with respect to $p(\mathbf{x}, \boldsymbol{\theta})$
- Expectation Propagation [5]: approximate the joint density of $p(\mathbf{x}, \boldsymbol{\theta})$ by minimizing the Kullback-Leibler contrast of $p(\mathbf{x}, \boldsymbol{\theta})$ with respect to $\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$

# Variational Bayes & Expectation Propagation

- Both often well approximate posterior modes
- VB and EP both require constraints on $p(\mathbf{x}, \boldsymbol{\theta})$
  e.g. $p(\mathbf{x}, \boldsymbol{\theta}) = p_{\mathbf{x}}(\mathbf{x}) p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$
- VB can significantly underestimate posterior variances. This has been seen in latent Gaussian models.
- Similarly, EP can overestimate posterior variances

# Monte Carlo Markov Chains

- Helped to make Bayesian inference tractable
- Asymptotically correct
  $\rightarrow$ MCMC errors can be made arbitrarily small
  $\rightarrow$ characterized by additive $\mathcal{O}_p(N^{-1/2})$ errors
- Often have poor performance (slow) in latent Gaussian models
- Inferential validity assumes convergence of the chain to the equilibrium distribution

# Integrated Nested Laplace Approximations

- Implements numerical integration and analytical approximations to avoid simulation
  $\rightarrow$ dodges convergence issues
- Gaussian approximations are appealing for latent Gaussian models
  $\rightarrow$ often $\pi(\mathbf{x}|\mathbf{y})$ looks 'nearly' Gaussian
- Potentially introduce errors through approximations
  $\rightarrow$ MCMC errors seem preferable
- Number of hyperparameters should be kept small

# 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 $\pi(x_i|\mathbf{y})$

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 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 $\pi(x_i|\mathbf{y})$:

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

# Conclusion

- Latent Gaussian Models are a broad and useful class of models
- Bayesian inference on these models has been difficult, though MCMC can do it
- INLA can provide an accurate and (often faster) solution

# Acknowledgments


- Jon, Patrick, Vladimir
- All of you


Questions?

# 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*