# Extending INLA to a class of near-Gaussian latent models

Thiago G. Martins

**Joint work with Håvard Rue**
Norwegian University of Science and Technology
Trondheim, Norway

**LGM2012**

May, 2012

# Latent Gaussian Models (LGMs)

- Hierarchical representation

**Stage 1.** $\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1 \sim \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}_1)$.

**Stage 2.** $\mathbf{x}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2))$,

**Stage 3.** $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

- Our aim

**Stage 2$^{\text{new}}$.** $\underbrace{(\mathbf{x}_G, \mathbf{x}_{NG})}_{\mathbf{x}}|\theta_2 \sim \pi(\mathbf{x}|\theta_2)$

$$\pi(\mathbf{x}|\theta_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\theta_2)) \times \prod_i \pi(\mathbf{x}_{NGi}|\theta_2)$$

# Latent Gaussian Models (LGMs)

- Hierarchical representation

**Stage 1.** $\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1 \sim \pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1) = \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}_1)$.

**Stage 2.** $\mathbf{x}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2))$,

**Stage 3.** $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

- Our aim

**Stage 2$^{\text{new}}$.** $\underbrace{(\mathbf{x}_G, \mathbf{x}_{NG})}_{\mathbf{x}} |\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2)$

$$\pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

# near-Gaussian latent models

$$\textbf{Stage 2}^{\textbf{new}}\textbf{.}\ \underbrace{(\mathbf{x}_G, \mathbf{x}_{NG})}_{\mathbf{x}}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2)$$

$$\pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

Some assumptions:

- $\mathbf{x}_G$ independent of $\mathbf{x}_{NG}$
- $\mathbf{x}_{NG}$ with independent components

$$\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2) = \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

- $\pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$ not very far from Gaussian.

# near-Gaussian latent models

$$\textbf{Stage 2}^{\textbf{new}}\textbf{.}\ \underbrace{(\mathbf{x}_G, \mathbf{x}_{NG})}_{\mathbf{x}} | \boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2)$$

$$\pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

Some assumptions:

- $\mathbf{x}_G$ independent of $\mathbf{x}_{NG}$
- $\mathbf{x}_{NG}$ with independent components

$$\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2) = \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

- $\pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$ not very far from Gaussian.

# near-Gaussian latent models

$$\textbf{Stage 2}^{\textbf{new}}\textbf{.}\ \underbrace{(\mathbf{x}_G, \mathbf{x}_{NG})}_{\mathbf{x}}|\boldsymbol{\theta}_2 \sim \pi(\mathbf{x}|\boldsymbol{\theta}_2)$$

$$\pi(\mathbf{x}|\boldsymbol{\theta}_2) = \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

Some assumptions:

- $\mathbf{x}_G$ independent of $\mathbf{x}_{NG}$
- $\mathbf{x}_{NG}$ with independent components

$$\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2) = \prod_i \pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$$

- $\pi(\mathbf{x}_{NGi}|\boldsymbol{\theta}_2)$ not very far from Gaussian.

# Motivation

1. For some specific reason, people need such models

   - E.g. gamma frailty models in survival analysis literature - log-gamma random effects.

2. More flexible modeling

   - Challenge the Gaussian distribution of certain components in your model.

   - E.g. robust mixed effect models, with Student's t rather than Gaussian random effects. [Pinheiro et al., 2001]

# Motivation

1. For some specific reason, people need such models

   - E.g. gamma frailty models in survival analysis literature - log-gamma random effects.

2. More flexible modeling

   - Challenge the Gaussian distribution of certain components in your model.

   - E.g. robust mixed effect models, with Student's t rather than Gaussian random effects. [Pinheiro et al., 2001]

# Motivation

1. For some specific reason, people need such models

   - E.g. gamma frailty models in survival analysis literature - log-gamma random effects.

2. More flexible modeling

   - Challenge the Gaussian distribution of certain components in your model.

   - E.g. robust mixed effect models, with Student's t rather than Gaussian random effects. [Pinheiro et al., 2001]

# INLA - main objective

Posterior

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\, \pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta})$$

Compute the posterior marginals:

$$\pi(x_i|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\, \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\, d\boldsymbol{\theta}$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}_{-j}$$

# INLA - main objective

Posterior

$$\pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\, \pi(\mathbf{x}|\boldsymbol{\theta}) \prod_{i \in \mathcal{I}} \pi(y_i|x_i, \boldsymbol{\theta})$$

Compute the posterior marginals:

$$\pi(x_i|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\, \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\, d\boldsymbol{\theta}$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}_{-j}$$

# INLA - tasks

1. Build an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$: $\widetilde{\pi}(\boldsymbol{\theta}\ |\mathbf{y})$
2. Build an approximation to $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$: $\widetilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$

$$\pi(x_i|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\ \pi(x_i|\boldsymbol{\theta},\mathbf{y})\ d\boldsymbol{\theta}$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta}_{-j}$$

3. Do the integration wrt $\boldsymbol{\theta}$ numerically.

# INLA - tasks

1. Build an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$: $\widetilde{\pi}(\boldsymbol{\theta}\ |\mathbf{y})$
2. Build an approximation to $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$: $\widetilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$

$$\pi(x_i|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\ \pi(x_i|\boldsymbol{\theta},\mathbf{y})\ d\boldsymbol{\theta}$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta}_{-j}$$

3. Do the integration wrt $\boldsymbol{\theta}$ numerically.

# INLA - tasks

1. Build an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$: $\widetilde{\pi}(\boldsymbol{\theta}\ |\mathbf{y})$
2. Build an approximation to $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$: $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$

$$\pi(x_i|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\ \pi(x_i|\boldsymbol{\theta}, \mathbf{y})\ d\boldsymbol{\theta}$$

$$\pi(\theta_j|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})\ d\boldsymbol{\theta}_{-j}$$

3. Do the integration wrt $\theta$ numerically.

# INLA - tasks

1. Build an approximation to $\pi(\boldsymbol{\theta}|\mathbf{y})$: $\widetilde{\pi}(\boldsymbol{\theta}\ |\mathbf{y})$
2. Build an approximation to $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$: $\widetilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$

$$\widetilde{\pi}(x_i|\mathbf{y}) = \int \widetilde{\pi}(\boldsymbol{\theta}|\mathbf{y})\, \widetilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})\, d\boldsymbol{\theta}$$
$$\widetilde{\pi}(\theta_j|\mathbf{y}) = \int \widetilde{\pi}(\boldsymbol{\theta}|\mathbf{y})\, d\boldsymbol{\theta}_{-j}$$

3. Do the integration wrt $\boldsymbol{\theta}$ numerically.

# The GMRF-approximation

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

Constructed as follows:

- Locate the mode $\mathbf{x}^*$
- Expand to second order

Markov and computational properties are preserved

# The GMRF-approximation

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

Constructed as follows:

- Locate the mode $\mathbf{x}^*$
- Expand to second order

Markov and computational properties are preserved

# The GMRF-approximation

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \sum_i \log\pi(y_i|x_i)\right)$$
$$\approx \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T(\mathbf{Q}+\text{diag}(c_i))(\mathbf{x}-\boldsymbol{\mu})\right) = \widetilde{\pi}(\mathbf{x}|\mathbf{y})$$

Constructed as follows:

- Locate the mode $\mathbf{x}^*$
- Expand to second order

Markov and computational properties are preserved

# INLA and latent Gaussian field

**The challenge:**

- The success of INLA depends heavily on the Gaussian assumption of the latent field

- Accuracy of approximations
- Computational point of view

# INLA and latent Gaussian field

**The challenge:**

- The success of INLA depends heavily on the Gaussian assumption of the latent field

- Accuracy of approximations

  - Gaussian prior on $\mathbf{x}$ contribute to $\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}$ to be well behaved and close to a Gaussian.

- Computational point of view

# INLA and latent Gaussian field

**The challenge:**

- The success of INLA depends heavily on the Gaussian assumption of the latent field

- Accuracy of approximations

  - Gaussian prior on $\mathbf{x}$ contribute to $\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}$ to be well behaved and close to a Gaussian.

- Computational point of view

  - Conditional independence properties in the Gaussian latent field translate into a sparse precision matrix which speed up the GMRF approximation considerably.

# The extension

1. Approximate $\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$ by a Gaussian $\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$
2. Correct for this approximation in the likelihood function with a correction term

$$CT_i = \pi(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2), \;\; i = 1, ..., k$$

- **But why?**

**Stage 1.** $\mathbf{y}|\mathbf{x}, \theta_1 \sim \prod_{i=1}^{n_d} \pi(y_i|x_i, \theta_1) \prod_{i=1}^{k} CT_i(\mathbf{x}_{NG,i}, \theta_2)$.

**Stage 2.** $\mathbf{x}|\theta_2 \sim \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\theta_2)) \times \pi_G(\mathbf{x}_{NG}|\theta_2)$,

**Stage 3.** $\theta \sim \pi(\theta)$.

- Independent of the chosen $\pi_G(\mathbf{x}_{NG}|\theta_2)$, the model represented above is equivalent to our model of interest.

# The extension

1. Approximate $\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$ by a Gaussian $\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$
2. Correct for this approximation in the likelihood function with a correction term

$$CT_i = \pi(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2), \; i = 1, ..., k$$

- **But why?**

**Stage 1.** $\mathbf{y}|\mathbf{x}, \theta_1 \sim \prod_{i=1}^{n_d} \pi(y_i|x_i, \theta_1) \prod_{i=1}^{k} CT_i(\mathbf{x}_{NG,i}, \boldsymbol{\theta}_2)$.

**Stage 2.** $\mathbf{x}|\theta_2 \sim \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\theta_2)) \times \pi_G(\mathbf{x}_{NG}|\theta_2)$,

**Stage 3.** $\theta \sim \pi(\theta)$.

- Independent of the chosen $\pi_G(\mathbf{x}_{NG}|\theta_2)$, the model represented above is equivalent to our model of interest.

# The extension

1. Approximate $\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$ by a Gaussian $\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$
2. Correct for this approximation in the likelihood function with a correction term

$$CT_i = \pi(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2),\ \ i = 1, ..., k$$

- **But why?**

**Stage 1.** $\mathbf{y}|\mathbf{x}, \theta_1 \sim \prod_{i=1}^{n_d} \pi(y_i|x_i, \theta_1) \prod_{i=1}^{k} CT_i(\mathbf{x}_{NG,i}, \boldsymbol{\theta_2})$.

**Stage 2.** $\mathbf{x}|\theta_2 \sim \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\theta_2)) \times \pi_G(\mathbf{x}_{NG}|\theta_2)$,

**Stage 3.** $\theta \sim \pi(\theta)$.

- Independent of the chosen $\pi_G(\mathbf{x}_{NG}|\theta_2)$, the model represented above is equivalent to our model of interest.

# The extension

1. Approximate $\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$ by a Gaussian $\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$
2. Correct for this approximation in the likelihood function with a correction term

$$CT_i = \pi(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2),\ \ i = 1, ..., k$$

- **But why?**

**Stage 1.** $\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1 \sim \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}_1) \prod_{i=1}^{k} CT_i(\mathbf{x}_{NG,i}, \boldsymbol{\theta}_\mathbf{2})$.

**Stage 2.** $\mathbf{x}|\boldsymbol{\theta}_2 \sim \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$,

**Stage 3.** $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

- Independent of the chosen $\pi_G(\mathbf{x}_{NG}|\theta_2)$, the model represented above is equivalent to our model of interest.

# The extension

1. Approximate $\pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$ by a Gaussian $\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$
2. Correct for this approximation in the likelihood function with a correction term

$$CT_i = \pi(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG,i}|\boldsymbol{\theta}_2),\ i = 1, ..., k$$

- **But why?**

**Stage 1.** $\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}_1 \sim \prod_{i=1}^{n_d} \pi(y_i|x_i, \boldsymbol{\theta}_1) \prod_{i=1}^{k} CT_i(\mathbf{x}_{NG,i}, \boldsymbol{\theta}_\mathbf{2})$.

**Stage 2.** $\mathbf{x}|\boldsymbol{\theta}_2 \sim \mathcal{N}(\mathbf{x}_G; \mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)) \times \pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$,

**Stage 3.** $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$.

- Independent of the chosen $\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)$, the model represented above is equivalent to our model of interest.

# The extension (cont.)

- Affect mainly the Gaussian approximation to $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$

**Before:**

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

**Now:**

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \sum_{i=1}^{n_d} \log \pi(y_i|x_i) + \sum_{j=1}^{k} h_j(x_{NG,j})\right)$$

where $h_j(x_{NG,j}) = \log[\pi(\mathbf{x}_{NG,j})/\pi_G(x_{NG,j})]$.

- A good Gaussian approximation will depend heavily on the shape of $h_j(x_{NG,j})$.

# The extension (cont.)

- Affect mainly the Gaussian approximation to $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$

**Before:**

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

**Now:**

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \sum_{i=1}^{n_d} \log \pi(y_i|x_i) + \sum_{j=1}^{k} h_j(x_{NG,j})\right)$$

where $h_j(x_{NG,j}) = \log[\pi(\mathbf{x}_{NG,j})/\pi_G(x_{NG,j})]$.

- A good Gaussian approximation will depend heavily on the shape of $h_j(x_{NG,j})$.

# The extension (cont.)

- Affect mainly the Gaussian approximation to $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$

**Before:**

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

**Now:**

$$\pi(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \sum_{i=1}^{n_d} \log \pi(y_i|x_i) + \sum_{j=1}^{k} h_j(x_{NG,j})\right)$$

where $h_j(x_{NG,j}) = \log[\pi(\mathbf{x}_{NG,j})/\pi_G(x_{NG,j})]$.

- A good Gaussian approximation will depend heavily on the shape of $h_j(x_{NG,j})$.

# Correction term

- It is desirable to have a bounded correction term

$$CT = \pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2) < \infty$$

  at least on the region that concentrates the bulk of probability mass.
- Although not necessary, it would be nice to have a log-concave correction term, at least around the mode of $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$.
  - More stability to the optimization step.
- In our examples, we have chosen $\pi_G(x_{NG,j}|\theta)$ to be a Gaussian with zero mean and low precision $\tau \to 0$.

# Correction term

- It is desirable to have a bounded correction term

$$CT = \pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2) < \infty$$

  at least on the region that concentrates the bulk of probability mass.
- Although not necessary, it would be nice to have a log-concave correction term, at least around the mode of $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$.
  - More stability to the optimization step.
- In our examples, we have chosen $\pi_G(x_{NG,j}|\theta)$ to be a Gaussian with zero mean and low precision $\tau \to 0$.

# Correction term

- It is desirable to have a bounded correction term

$$CT = \pi(\mathbf{x}_{NG}|\boldsymbol{\theta}_2)/\pi_G(\mathbf{x}_{NG}|\boldsymbol{\theta}_2) < \infty$$

  at least on the region that concentrates the bulk of probability mass.
- Although not necessary, it would be nice to have a log-concave correction term, at least around the mode of $\pi(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$.
  - More stability to the optimization step.
- In our examples, we have chosen $\pi_G(x_{NG,j}|\boldsymbol{\theta})$ to be a Gaussian with zero mean and low precision $\tau \to 0$.

# Survival analysis with gamma frailty

1. $y_{i,j} \sim \exp(\lambda_{i,j}), \quad i = 1, ..., I, \ j = 1, ..., J$

2. $\eta_{i,j} = \log(\lambda_{i,j}) = \beta_0 + \beta_1 x_{i,j} + \log w_i$
   $w_i \sim gamma(\kappa, \kappa), \ E(w_i) = 1$
   $\beta_i \sim N(0, 10^2)$

3. $\kappa \sim gamma(1, 0.1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$, with $\mathbf{b} = \log(\mathbf{w})$.
- Cannot be applied directly with INLA.

# Survival analysis with gamma frailty

1. $y_{i,j} \sim \exp(\lambda_{i,j}), \quad i = 1, ..., I,\ j = 1, ..., J$

2. $\eta_{i,j} = \log(\lambda_{i,j}) = \beta_0 + \beta_1 x_{i,j} + \log w_i$
   $w_i \sim gamma(\kappa, \kappa),\ E(w_i) = 1$
   $\beta_i \sim N(0, 10^2)$

3. $\kappa \sim gamma(1, 0.1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$, with $\mathbf{b} = \log(\mathbf{w})$.
- Cannot be applied directly with INLA.

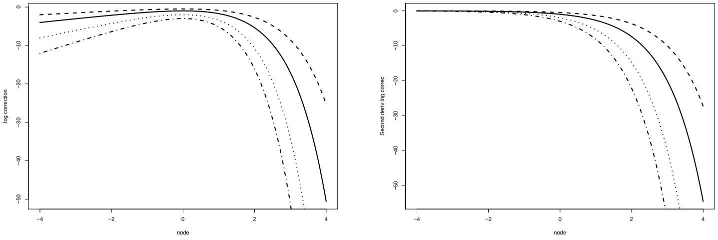
# Survival analysis with gamma frailty

1. $y_{i,j} \sim \exp(\lambda_{i,j}), \quad i = 1, ..., I, \; j = 1, ..., J$
2. $\eta_{i,j} = \log(\lambda_{i,j}) = \beta_0 + \beta_1 x_{i,j} + \log w_i$
   $w_i \sim gamma(\kappa, \kappa), \; E(w_i) = 1$
   $\beta_i \sim N(0, 10^2)$
3. $\kappa \sim gamma(1, 0.1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$, with $\mathbf{b} = \log(\mathbf{w})$.
- Cannot be applied directly with INLA.

# Survival analysis with gamma frailty (cont.)

- Apply our approach with a low precision Gaussian.
- In this case we get a log-concave correction term.




(a) (b)

Figura: Plot of the log correction term (a) and of the second derivative of the log correction term (b) against $b_i$ for different values of $\kappa$

# Survival analysis with gamma frailty (cont.)

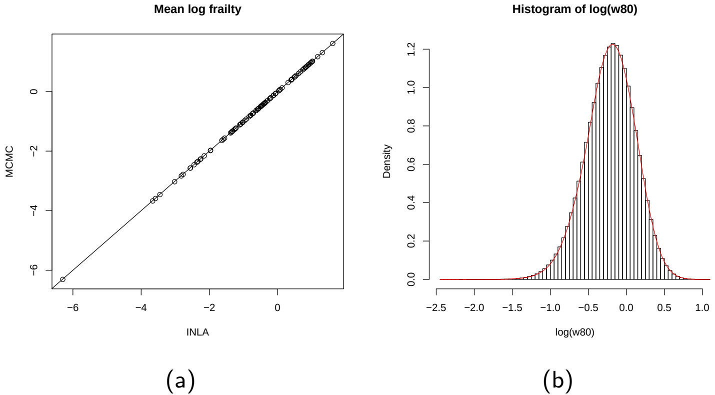



Figura: (a) Plot of the posterior mean of the log frailties returned by INLA (x-axis) vs. MCMC (y-axis). (b) Approximate posterior density for $\log w_{80}$ obtained by INLA (solid line) and by MCMC (histogram).

# Survival analysis with gamma frailty (cont.)

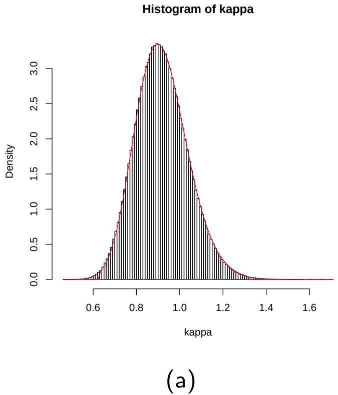



Figura: (a) Approximate posterior density for $\kappa$ obtained by INLA (solid line) and by MCMC (histogram) (b) Approximate posterior density for $\beta_1$ obtained by INLA (solid line) and by MCMC (histogram)

# Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim N(0, \psi_b^2), \quad \mathbf{e}_i \sim N(0, \psi_e^2\mathbf{I}),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

1. $y_{ij}|\mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i\beta + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\eta, \beta, \mathbf{b})$.

## Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim t(0, \psi_b^2, \nu), \quad \mathbf{e}_i \sim t(0, \psi_e^2 \mathbf{I}, \nu),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

1. $y_{ij}|\mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i\beta + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$.

# Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim t(0, \psi_b^2, \nu), \quad \mathbf{e}_i \sim t(0, \psi_e^2 \mathbf{I}, \nu),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

1. $y_{ij} | \mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i \boldsymbol{\beta} + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$.

# Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim t(0, \psi_b^2, \nu), \quad \mathbf{e}_i \sim t(0, \psi_e^2\mathbf{I}, \nu),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

1. $y_{ij}|\mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i\beta + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$.

# Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim t(0, \psi_b^2, \nu), \quad \mathbf{e}_i \sim t(0, \psi_e^2\mathbf{I}, \nu),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

1. $y_{ij}|\mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i\boldsymbol{\beta} + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$.

# Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim t(0, \psi_b^2, \nu), \quad \mathbf{e}_i \sim t(0, \psi_e^2\mathbf{I}, \nu),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

1. $y_{ij}|\mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i\boldsymbol{\beta} + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$.

# Robust mixed-effects model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + b_i + \mathbf{e}_i, \quad \dim(\mathbf{y}_i) = k_i$$

$$b_i \sim t(0, \psi_b^2, \nu), \quad \mathbf{e}_i \sim t(0, \psi_e^2\mathbf{I}, \nu),$$

- Gaussian distributions vulnerable to outliers
- Provides outlier accommodation together with outlier identification.

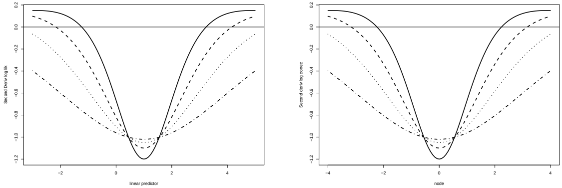
1. $y_{ij}|\mathbf{x}, \boldsymbol{\theta} \sim t(\eta_{ij}, \psi_e^2, \nu), \quad i = 1, ..., n$ and $j = 1, ..., k_i$,
2. $\eta_{ij} = \mathbf{X}_i\boldsymbol{\beta} + b_i$
3. $\psi_b^2, \psi_e^2 \sim gamma(1, 0.1)$, $\nu^* = \log(\nu - 5) \sim N(3, 1)$

- For this model $\mathbf{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \mathbf{b})$.

# Robust mixed-effects model (cont.)

- Again, if we use a zero mean and low precision Gaussian ($\mu = 0$, $\tau \to 0$)

$$\log CT_i = -\frac{(\nu + 1)}{2} \log \left\{ 1 + \frac{b_i^2}{\psi_b^2 \nu} \right\} + \text{const.}$$

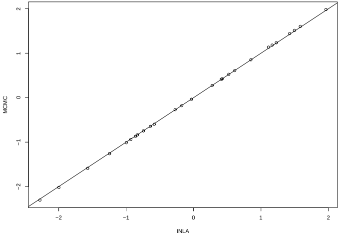



(a) (b)

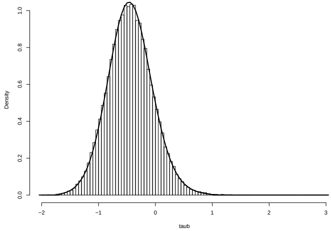
Figura: (a) Plot of the second derivative of the log likelihood against linear predictor and (b) plot of the second derivative of the log correction term for different values of $\nu$.

# Robust mixed-effects model (cont.)

- Again, similar accuracy with only a fraction of computational time.

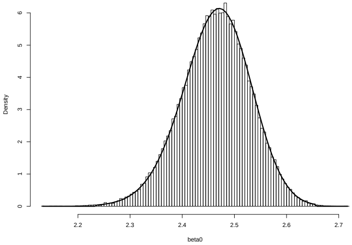



(a) log(**b**) returned by INLA (x-axis) vs. MCMC (y-axis).

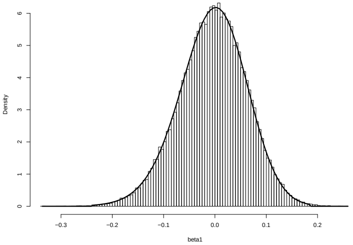



(b) $\log \tau_b = \log 1/\psi_b^2$.

# Robust mixed-effects model (cont.)




(c) $\log \beta_0$.




(d) $\log \beta_1$.

# Conclusion

- Extends the INLA method to latent models, where independent components of the latent field can have a *near-Gaussian* distribution.
- For the previous examples, when compared with MCMC, our method shows similar accuracy with only a small fraction of computation time.
- No precise definition of *near-Gaussian* distribution is given, the main interest is on how our approximation works in practice.

# Future

- Which class of distributions can be used in this framework.
- Other useful classes of models deserve further investigation. E.g. dynamic models with non-Gaussian error term for the observation and/or system equation, as in [Kitagawa, 1987].
- Develop a systematic approach to model selection using this framework to test standard assumptions of a given model, e.g. Gaussian random effects.

THANK YOU

Kitagawa, G. (1987).
Non-Gaussian state-space modeling of nonstationary time series.
*Journal of the American statistical association*, 82(400):1032–1041.

Pinheiro, J., Liu, C., and Wu, Y. (2001).
Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution.
*Journal of Computational and Graphical Statistics*, 10(2):249–276.