

Lecture 7: MCMC and INLA

Gaussian Markov random fields

David Bolin
Chalmers University of Technology
February 16, 2015



Classes of hierarchical GMRF models

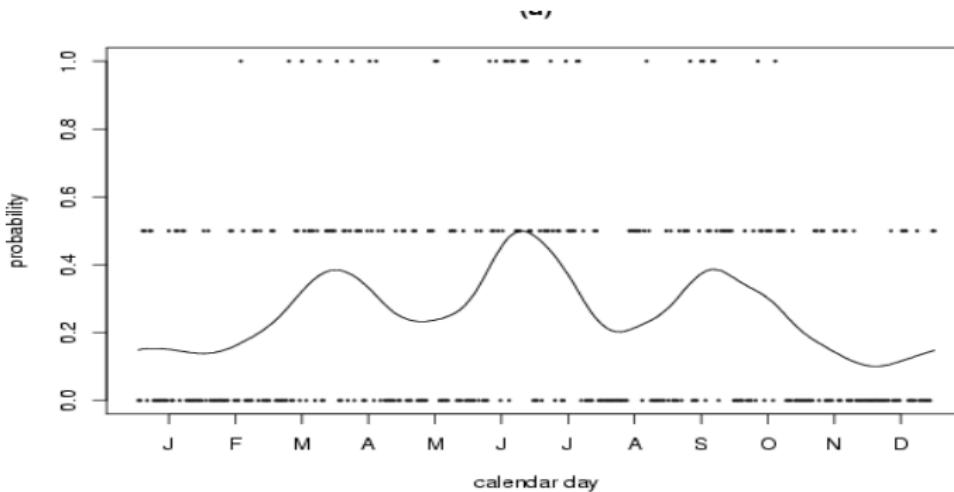
We can divide these models into three classes with increasing difficulty in terms of estimation

- ① Normal data
- ② Non-normal data that allows for a normal-mixture representation
 - Student- t distribution
 - Logistic and Laplace (Binary regression)
- ③ Non-normal data
 - Poisson
 - and others...

Last time, we looked at the case of normal data and normal-mixture representations

Today, we will continue with the normal-mixtures and then cover the non-normal data.

A simple example: Tokyo rainfall



A much analysed binomial time series from Kitawaga (1987).

- Each day during the years 1983 and 1984, it was recorded whether there was more than 1 mm rainfall in Tokyo.
- Of interest is to study the underlying probability p_i of rainfall at calendar day $i = 1, \dots, 366$.

Normal mixtures

The Tokyo model

$$\pi(\mathbf{x} \mid \tau) \pi(\tau) \prod_i \pi(y_i \mid x_i)$$

- $y_i|x_i$ is Binomial with $p_i = \Phi(x_i)$
- $\mathbf{x} \mid \tau$ is Gaussian with dimension 366
- τ is Gamma

Because of the likelihood, the full posterior is non-Gaussian, so we cannot use the methods for normal data directly

However, we can retrieve Gaussian conditionals by introducing auxiliary variables.

Binary regression models

Another important example where Auxiliary variables are useful:

Example: Binary regression

Gaussian \mathbf{x} and Bernoulli data

$$\begin{aligned}y_i &\sim \mathcal{B}(g^{-1}(x_i)) \\ g(p) &= \Phi^{-1}(p) \quad \text{probit link}\end{aligned}$$

Equivalent representation using auxiliary variables \mathbf{w}

$$\begin{aligned}w_i &= x_i + \epsilon_i, \quad \epsilon_i \sim N(0, 1) \\ y_i &= \begin{cases} 1 & \text{if } w_i > 0 \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

Auxiliary variables for the Tokyo data

We have binomial data, where each day

$$y_i \sim \begin{cases} y_{i,1} + y_{i,2}, i \neq 60 \\ y_{i,1}, i = 60 \end{cases}$$

where $y_{i,\bullet} \sim \mathcal{B}(p_i)$.

The data only contain information about y_i , so if $y_i = 1$ we let $y_{i,1} = 1$ and $y_{i,2} = 0$.

Let n_i be the number of observations for day i and introduce one auxiliary variable $w_{i,j}$ for each $y_{i,j}$.

With

$$w_i = \sum_{j=1}^{n_i} w_{i,j}$$

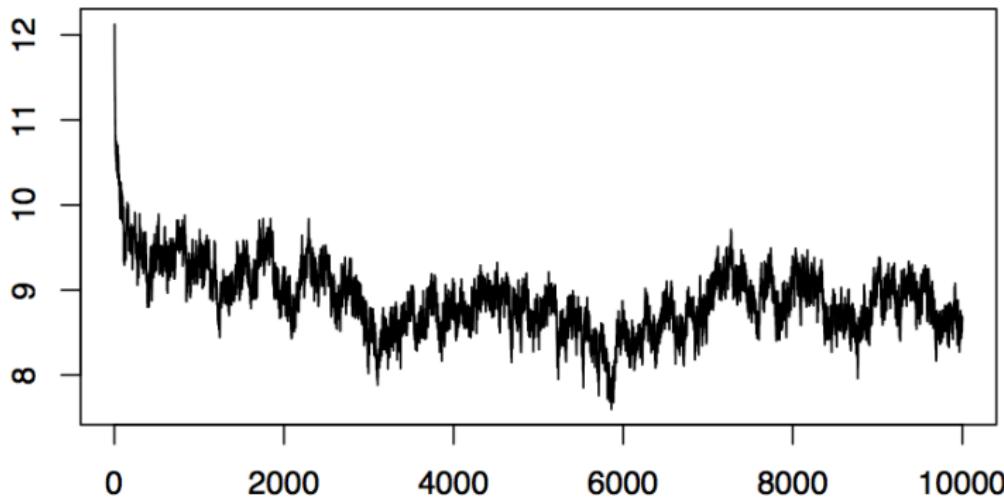
we have that $\mathbf{x} | \text{the rest} \sim N(\boldsymbol{\mu}, \hat{\mathbf{Q}}^{-1})$ where $\hat{\mathbf{Q}} = \tau \mathbf{R} + \text{diag}(\mathbf{n})$ and $\boldsymbol{\mu} = \hat{\mathbf{Q}}^{-1} \mathbf{w}$.

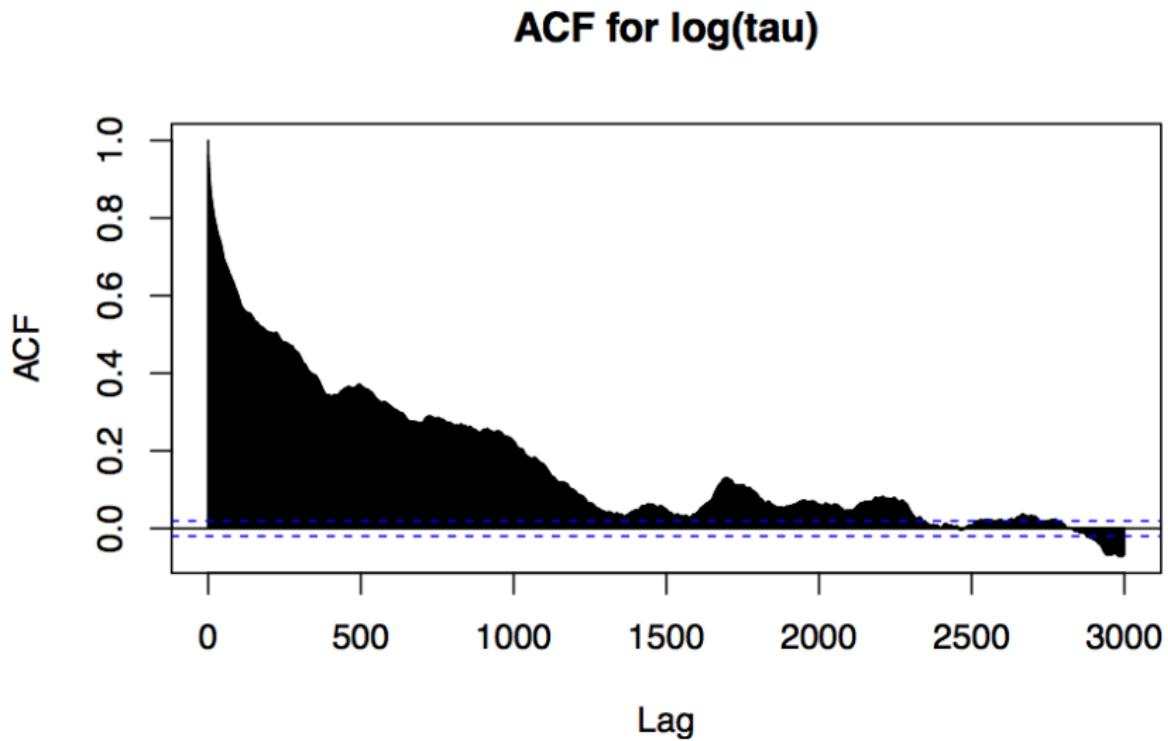
Single-site Gibbs sampling

Using the auxiliary variable formulation, we obtain the following Gibbs sampler for the problem:

- $\tau \sim \Gamma\left(\frac{366-1}{2} + \alpha, \frac{1}{2}\mathbf{x}^\top \mathbf{R}\mathbf{x} + \beta\right)$
- for each i
 - $x_i \sim \mathcal{N}(\mu_i - \hat{Q}_{i,i}^{-1} \sum_{j \neq i} \hat{Q}_{ij}(x_i - \mu_i), \hat{Q}_{i,i}^{-1})$
- for each i
 - $w_i \sim \mathcal{W}(\cdot)$

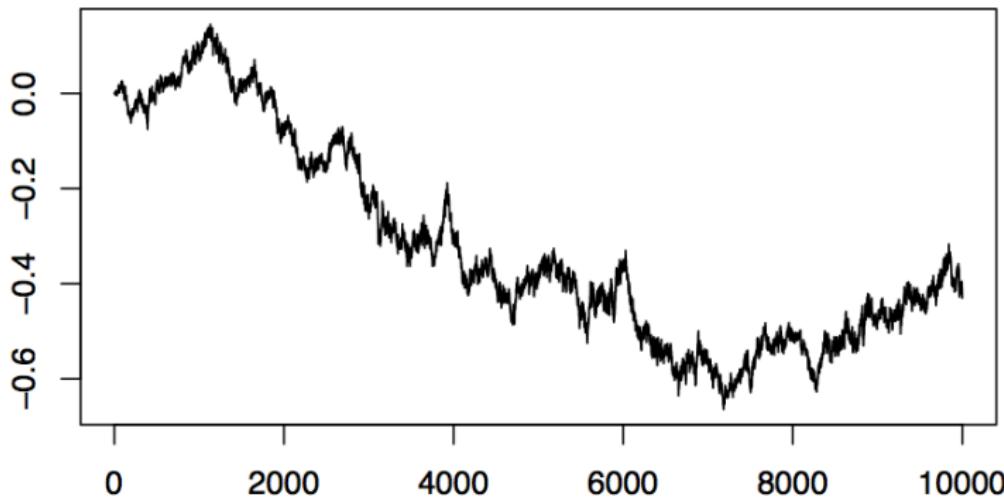
The distribution for w_i is $N(x_i, 1)$ truncated to be positive if $y_i = 1$ and truncated to be negative if $y_i = 0$.

Results: hyper-parameter $\log(\tau)$ **Trace of $\log(\tau)$** 

Results: hyper-parameter $\log(\tau)$ 

Results: x_1

Trace of $x[1]$



Discussion

The single-site method has severe problems also in this case:

- Even *long runs* shows large variation
- “Long” range dependence
- *Very* slowly mixing
- It takes *hours* to obtain even short chains.
- Easy to be “fooled” running shorter chains
- The variability can be underestimated.

Again, we have two issues

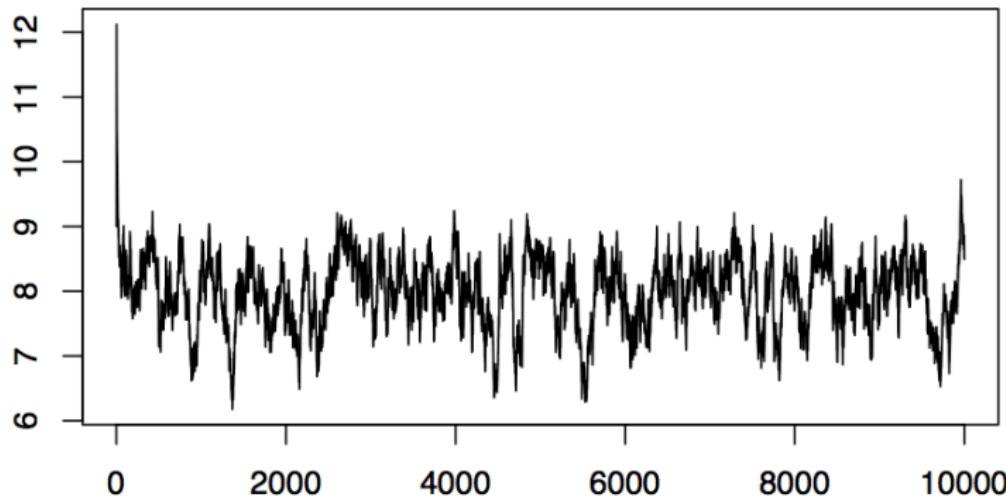
- ① Slow mixing within the latent field \mathbf{x}
- ② Slow mixing between the latent field \mathbf{x} and τ .

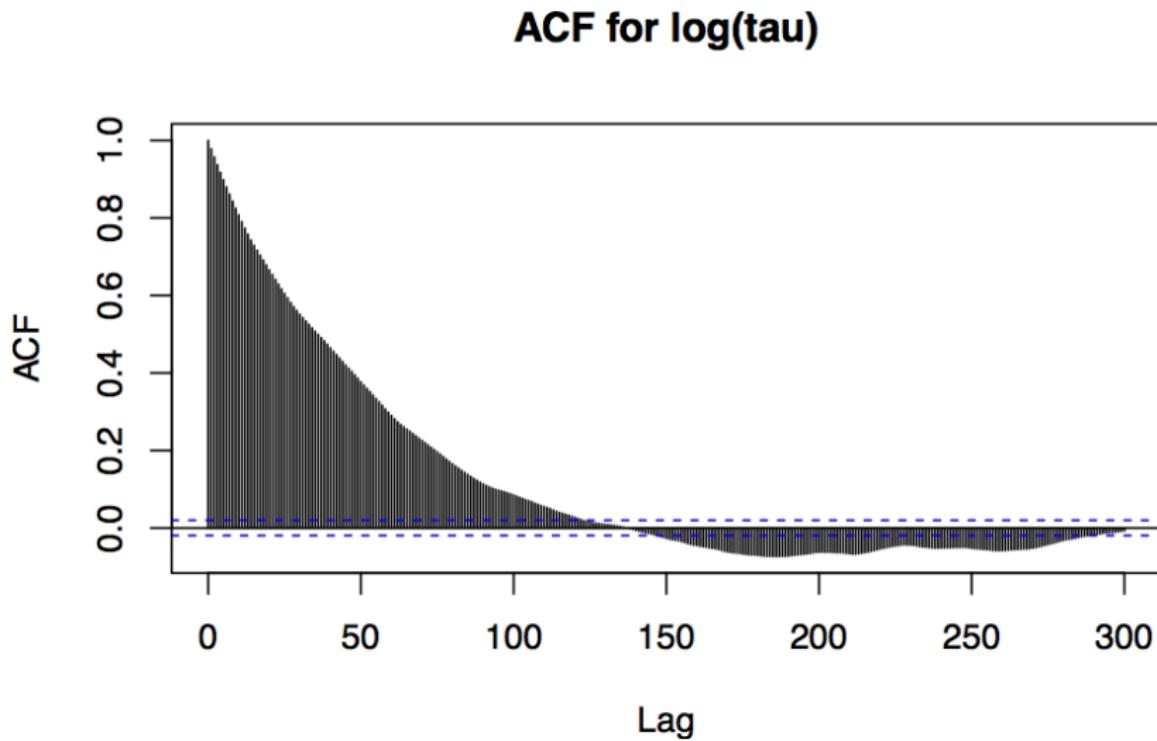
A first strategy for blocking

Since we know the joint distribution for $\mathbf{x} | \text{the rest}$, we can update all components in \mathbf{x} simultaneously.

Thus, we change our sampler to

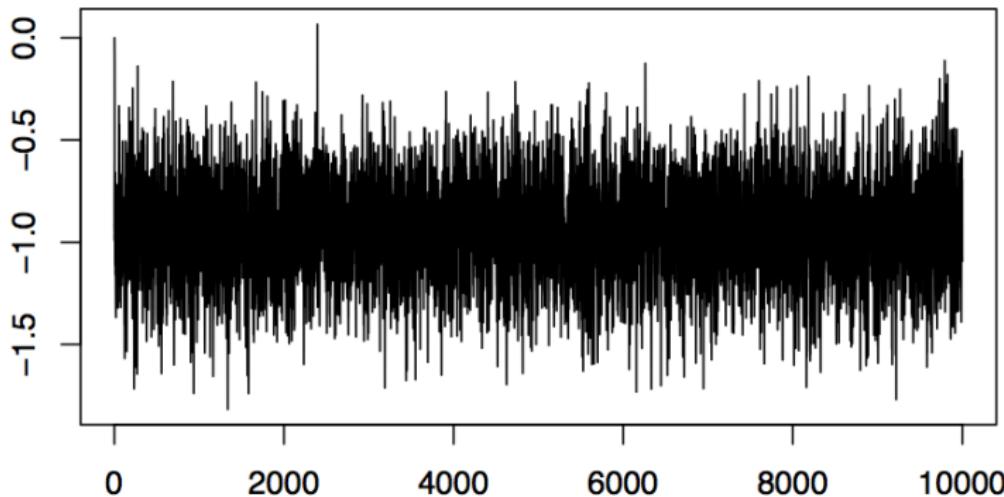
- $\tau \sim \Gamma\left(\frac{366-1}{2} + \alpha, \frac{1}{2}\mathbf{x}^\top \mathbf{R}\mathbf{x} + \beta\right)$
- $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \hat{\mathbf{Q}}^{-1})$
- for each i
 - $w_i \sim \mathcal{W}(\cdot)$

Results: hyper-parameter $\log(\tau)$ **Trace of $\log(\tau)$** 

Results: hyper-parameter $\log(\tau)$ 

Results: x_1

Trace of $x[1]$



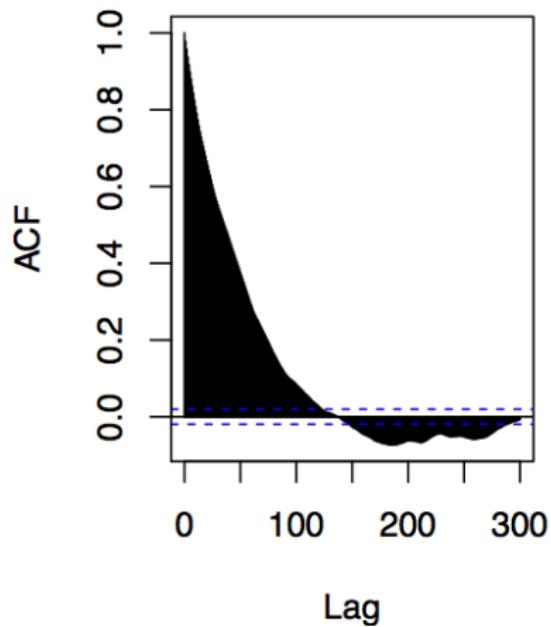
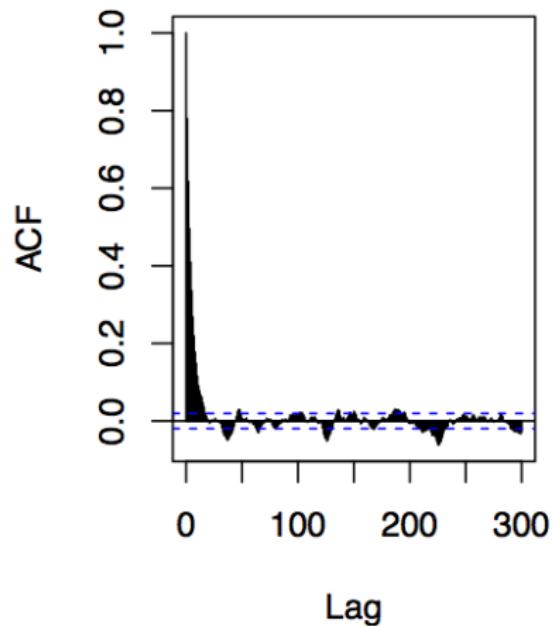
An updated blocking strategy

Also in this case, we have slow mixing due to the interaction between the latent field \mathbf{x} and τ .

Thus, we merge the two sub-blocks and do a joint update

- Sample
 - $\tau' \sim q(\tau'; \tau)$ (e.g. using a random walk proposal)
 - $\mathbf{x}' \sim \mathcal{N}(\boldsymbol{\mu}, \hat{\mathbf{Q}}^{-1})$
- and then accept/reject (\mathbf{x}', τ') jointly
- for each i
 - $w_i \sim \mathcal{W}(\cdot)$

This is easy to do since the normalising constant for $\mathbf{x}|\cdot$ is available.

Results: hyper-parameter $\log(\tau)$ **Scheme 1****Scheme 2**

Non-normal response models

- In the previous case, the auxiliary variables makes the full conditional for \mathbf{x} Gaussian
- Often, we cannot find such variables, the full conditional for \mathbf{x} then looks like

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

- We cannot sample directly from this distribution, but we can construct a Gaussian approximation of the distribution which we use as a proposal distribution in the MH-algorithm.
- This should work fine since the posterior distribution is not too far away from being Gaussian!

The Gaussian approximation (I)

- Let $g_i(x_i) = \log \pi(y_i|x_i)$, and do a Taylor approximation around some value μ_i^0 :

$$g_i(x_i) \approx g_i(\mu_i^0) + g'_i(\mu_i^0)(x_i - \mu_i^0) + \frac{g''_i(\mu_i^0)}{2}(x_i - \mu_i^0)^2$$

- Let $c_i = -g''_i(\mu_i^0)$ and $b_i = g'_i(\mu_i^0) - g''_i(\mu_i^0)\mu_i^0$, we then have that

$$\pi(\mathbf{x} | \dots) \approx \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{Q} + \text{diag}(\mathbf{c})) (\mathbf{x} - \boldsymbol{\mu}) \right)$$

where $\boldsymbol{\mu} = (\mathbf{Q} + \text{diag}(\mathbf{c}))^{-1}\mathbf{b}$.

The Gaussian approximation (II)

- We can repeat the Taylor approximation around μ to improve the approximation.
- Repeat until convergence to obtain a Gaussian approximation, $\pi_G(\mathbf{x} | \dots)$, with precision matrix

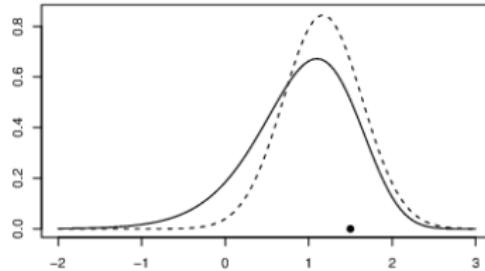
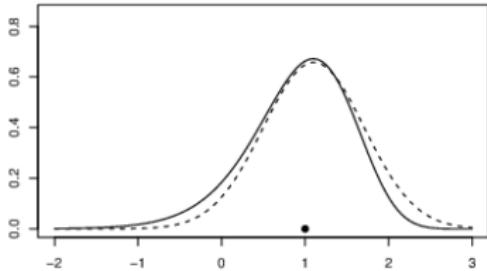
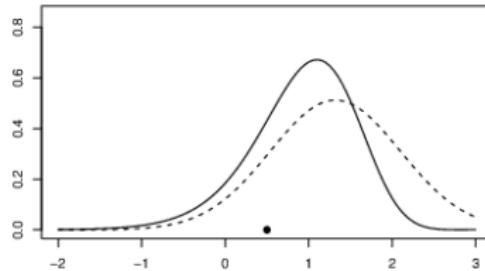
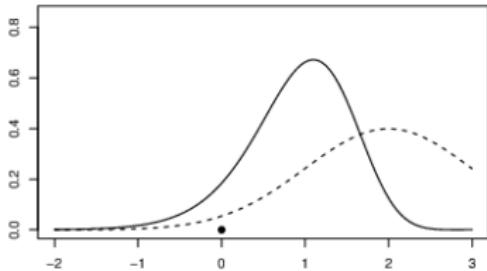
$$\mathbf{Q}^* = \mathbf{Q} + \text{diag}(\mathbf{c}^*)$$

and mean

$$\boldsymbol{\mu}^* = (\mathbf{Q} + \text{diag}(\mathbf{c}^*))^{-1} \mathbf{b}^*$$

- Important: sparsity properties are preserved!
- This is in fact just the Newton-Rapson algorithm. In statistics, variants of it are known as Fisher scoring and iteratively reweighted least squares.
- It is well-known that these quadratic approximations converge (under certain regularity conditions).

Why use the mode?



Example: Tokyo again

- The auxiliary variables makes the full conditional for \mathbf{x} Gaussian
- If we do not use them, the full conditional for \mathbf{x} looks like

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

- Thus, we can construct a GMRF approximation of the distribution which we use as a proposal distribution in the MH-algorithm.

$$\pi(\mathbf{x} \mid \dots) \approx \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T (\mathbf{Q} + \text{diag}(\mathbf{c})) (\mathbf{x} - \boldsymbol{\mu}) \right)$$

Improved one-block scheme

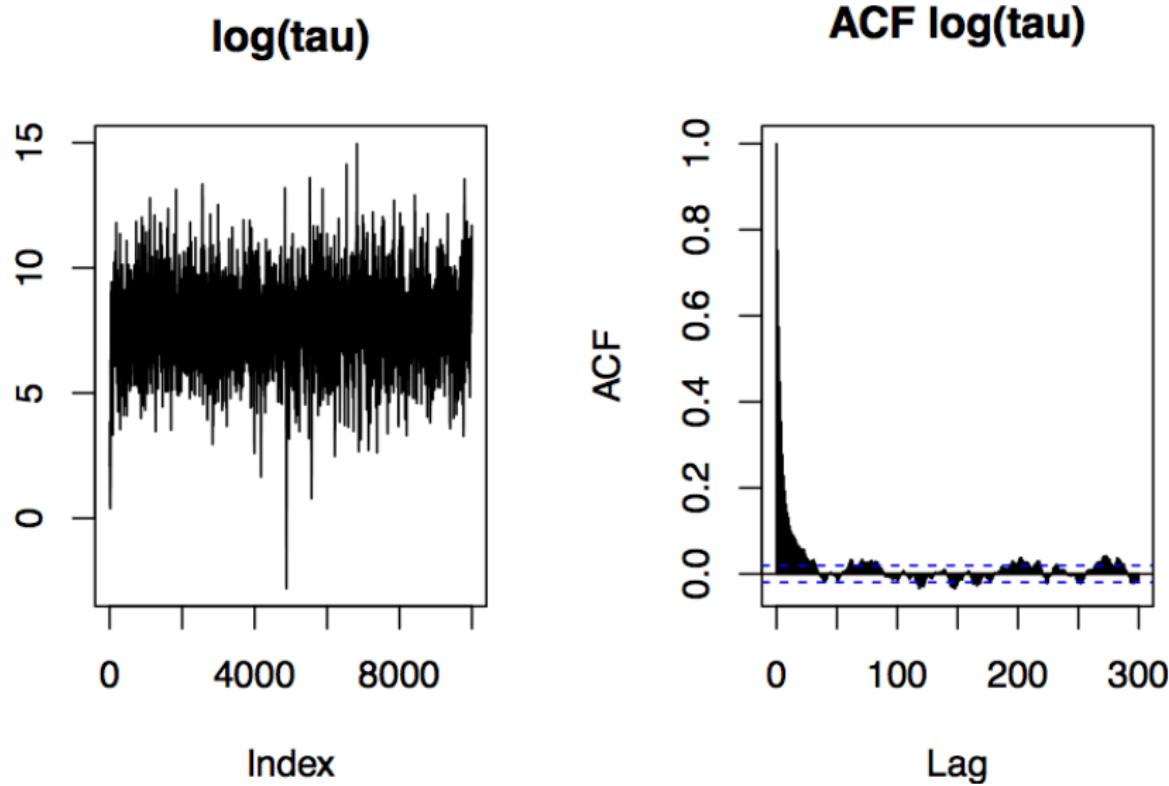
We can now remove the auxiliary variables from our one-block Gibbs sampler:

- $\tau' \sim q(\cdot; \tau)$
- $\mathbf{x}' \sim \pi_G(\mathbf{x} \mid \tau', \mathbf{y})$
- Accept/reject (\mathbf{x}', τ') jointly

We still sample τ using a random walk proposal.

Note: $\pi_G(\cdot)$ is indexed by τ' , hence we need to compute one for each value of τ' .

Results: Independence sampler



Improving the independence sampler (I)

We could improve the sampler by constructing a non-Gaussian approximation of the posterior, which

- we can block-update, and
- know the normalizing constant of the density.

One way of doing this is to use the non-homogeneous AR representation of the GMRF using the Cholesky factor.

Sequentially approximate the conditional densities

$$\pi(x_i | x_{i+1}, \dots, x_n, y_1, \dots, y_n)$$

More accurate non-gaussian approximations can give large improvements compared to the GMRF approximation.

Several methods for doing this is outlined in the course book.

Improving the independence sampler (II)

We can improve the independence sampler by replacing the random walk proposal for τ with an independence proposal.

We want to construct an approximation of $\pi(\tau|y)$, and we do this by starting with the simple identity

$$\begin{aligned}\pi(\tau | y) &= \frac{\pi(x, \tau | y)}{\pi(x | \tau, y)} \\ &\propto \frac{\pi(\tau)\pi(x|\tau)\pi(y|x)}{\pi(x|\tau, y)}\end{aligned}$$

We can evaluate the righthand side for any fixed value of x as the lefthand side does not depend on x .

The only unknown term in the expression is $\pi(x|\tau, y)$, so an approximation is obtained by instead using $\pi_G(x|\tau, y)$.

Laplace approximation

Thus, the approximation of $\pi(\tau|\mathbf{y})$ is

$$\tilde{\pi}(\tau \mid \mathbf{y}) = \frac{\pi(\tau)\pi(\mathbf{x}|\tau)\pi(\mathbf{y}|\mathbf{x})}{\pi_G(\mathbf{x}|\tau, \mathbf{y})}$$

The approximation now depends on \mathbf{x} , and we want to choose \mathbf{x} so that the approximation is as good as possible.

Intuitively, the Gaussian approximation is most accurate in the mode $\mu^*(\tau)$, and we should therefore use $\mathbf{x} = \mu^*(\tau)$ in the approximation.

When $\mathbf{x} = \mu^*(\tau)$ is used, the approximation is in fact a Laplace approximation.

Improved independence sampler

We have

$$\tilde{\pi}(\tau \mid \mathbf{y}) = \frac{\pi(\tau)\pi(\mathbf{x}|\tau)\pi(\mathbf{y}|\mathbf{x})}{\pi_G(\mathbf{x}|\tau, \mathbf{y})} \Big|_{\mathbf{x}=\mu^*(\tau)}$$

To obtain a computationally efficient sampler we can

- Evaluate the Laplace-approximation at some “selected” points
- Build an interpolation log-spline which is used as $\tilde{\pi}(\tau|\mathbf{y})$.

Improved independence sampler

Our final sampler is:

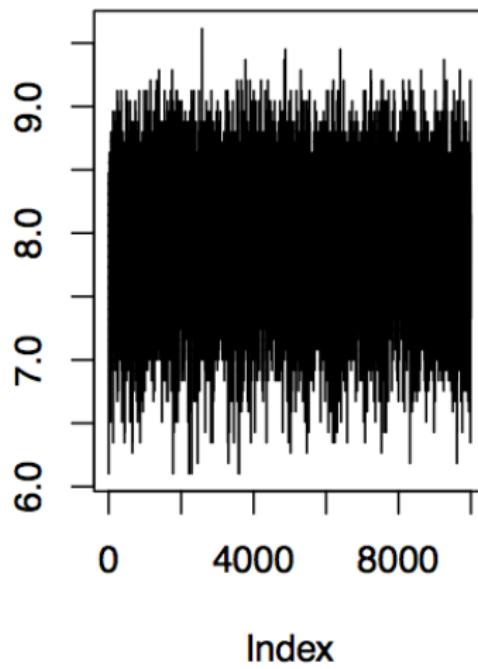
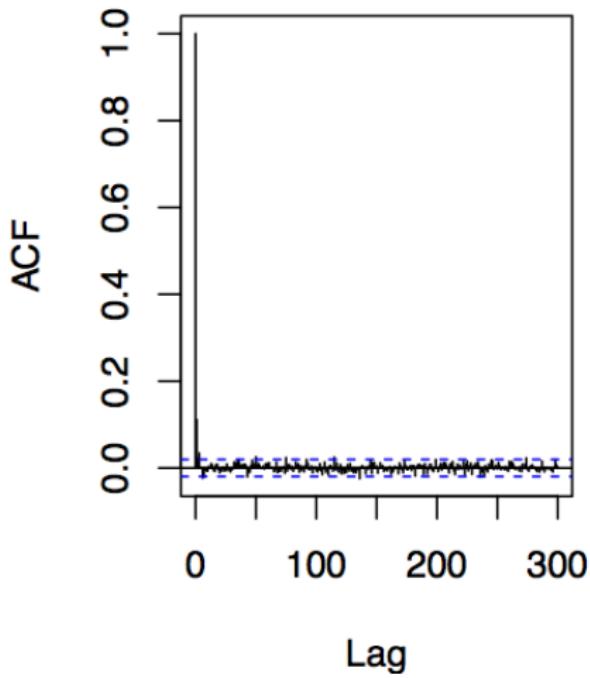
- $\tau' \sim \tilde{\pi}(\tau|\mathbf{y})$
- $\mathbf{x}' \sim \pi_G(\mathbf{x}|\tau', \mathbf{y})$
- Accept/reject (τ', \mathbf{x}') jointly

For an independence sampler with acceptance probability α , the correlation of the chain satisfies

$$\text{Corr}(\tau^i, \tau^{i+k}) \approx (1 - \alpha)^{|k|}$$

In our case we have $\alpha = 0.897!$

Results: Improved independence sampler

log(tau)**ACF log(tau)**

Why does this work?

Everything we need to compute are integrals (or ratios of integrals) of the form

$$\int f(\mathbf{u}) \exp(-ng(\mathbf{u})) d\mathbf{u}$$

and integrals of this form can be well approximated using the Laplace approximation!

Review: The classical Laplace approximation

Compute and approximation to the integral

$$\int \exp(ng(x)) dx$$

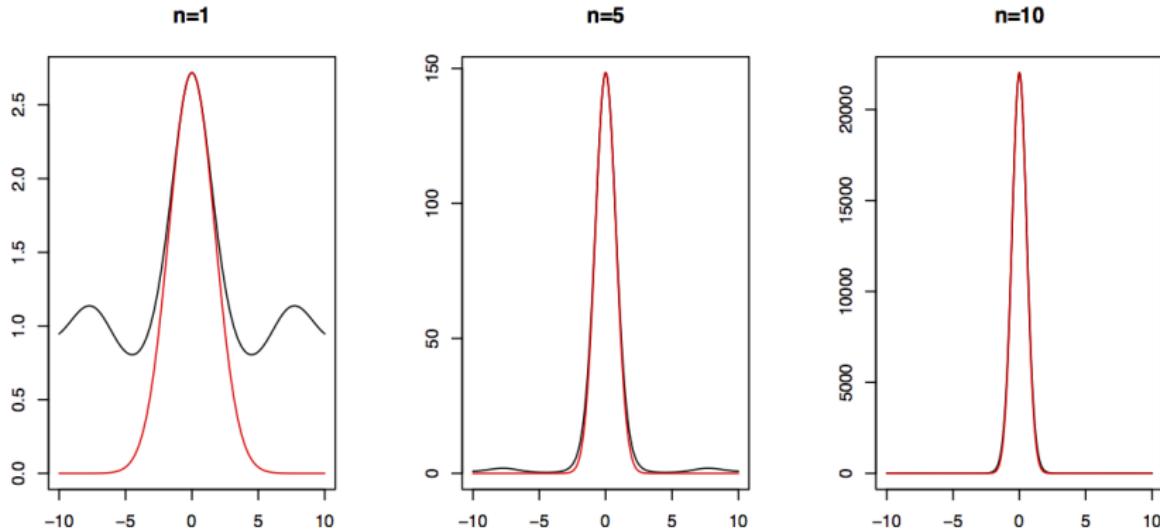
Let x_0 be the mode of $g(x)$, then

$$g(x) = g(x_0) - \frac{1}{2}|g''(x_0)|(x - x_0)^2 + \dots$$

So we have

$$\begin{aligned} \int \exp(ng(x))dx &\approx \exp(ng(x_0)) \int \exp\left(-\frac{1}{2}n|g''(x_0)|(x - x_0)^2\right)dx \\ &= \exp(ng(x_0)) \sqrt{\frac{2\pi}{n|g''(x_0)|}} \end{aligned}$$

Analysis of the error



- As $n \rightarrow \infty$, then the integrand gets more and more peaked.
- Error tends to zero as $n \rightarrow \infty$
- Detailed analysis gives that the relative error is $1 + \mathcal{O}(1/n)$
- The figure above shows an example with $g(x) = \sin(x)/x$

Using it in statistics

In our case, we have

$$g_n(x) = \frac{1}{n} \sum_{i=1}^n g_i(x)$$

then the mode x_0 depends on n as well.

It turns out this isn't a problem and it all still works...

If \mathbf{x} is multivariate,

$$\int \exp(ng(\mathbf{x})) d\mathbf{x} = \sqrt{\frac{(2\pi)^n}{n|-\mathbf{H}|}}$$

where \mathbf{H} is the hessian (matrix) at the mode

$$H_{ij} = \left. \frac{\partial^2}{\partial x_i \partial x_j} g(\mathbf{x}) \right|_{\mathbf{x}=\mathbf{x}_0}$$

Beyond Tokyo

Everything we have done so far generalised beyond the Tokyo data.

So what have we learned so far?

- Hierarchical models are popular but it is tricky to construct good MCMC samplers for them
- Single-site Gibbs samplers are popular
 - and they are very bad if there is strong correlation...
 - If you would use them anyway (and many do), your results will likely be wrong
- For latent GMRF models, we can improve the single-site samplers through
 - Blocking
 - GMRF approximations
 - Laplace approximations

A case for 'approximate' inference...

Even though we can improve the simple MCMC samplers a lot, they still require a lot of computational power.

MCMC is 'exact' in the limit. In practice, however, we can never do exact inference (except for trivial models).

Facetious principle

If we're going to be wrong, we might as well be wrong quickly.

MCMC violates this principle!

Can we fix this somehow?

- In general: No
- For latent GMRF models: Yes!

Deterministic inference

The trick is to remove the sampling part of our algorithm.

Recall that we typically only care about the marginal posteriors $\pi(x_i | \mathbf{y})$ and $\pi(\theta_i | \mathbf{y})$.

For the Tokyo model, having derived the Laplace approximations, we can do numerical integration instead of sampling:

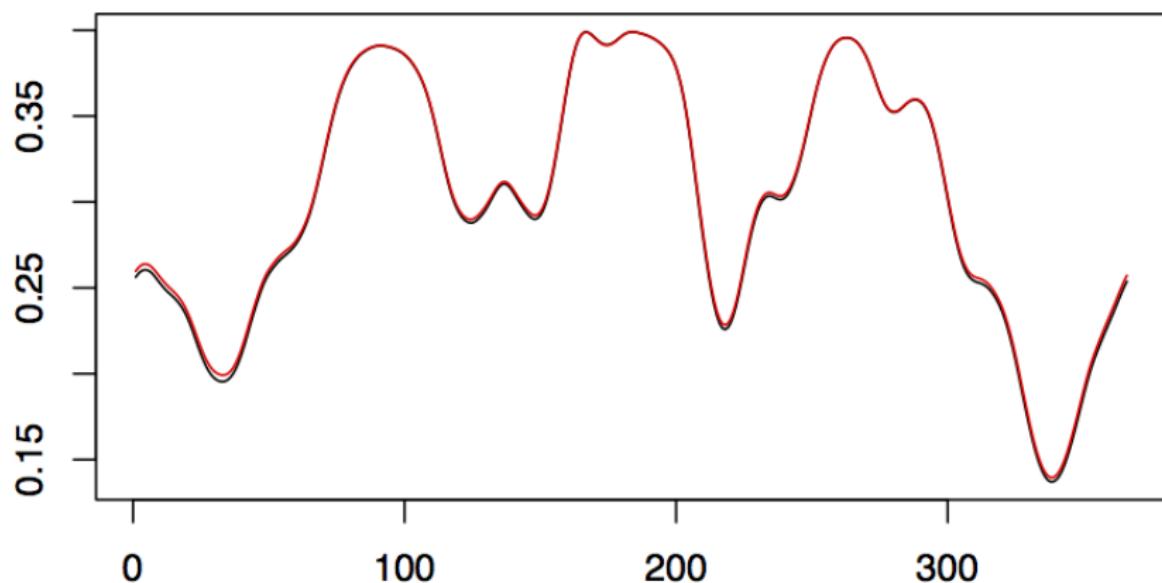
Posterior marginal for τ :

- Compute $\tilde{\pi}(\tau | \mathbf{y})$ (We already have this)

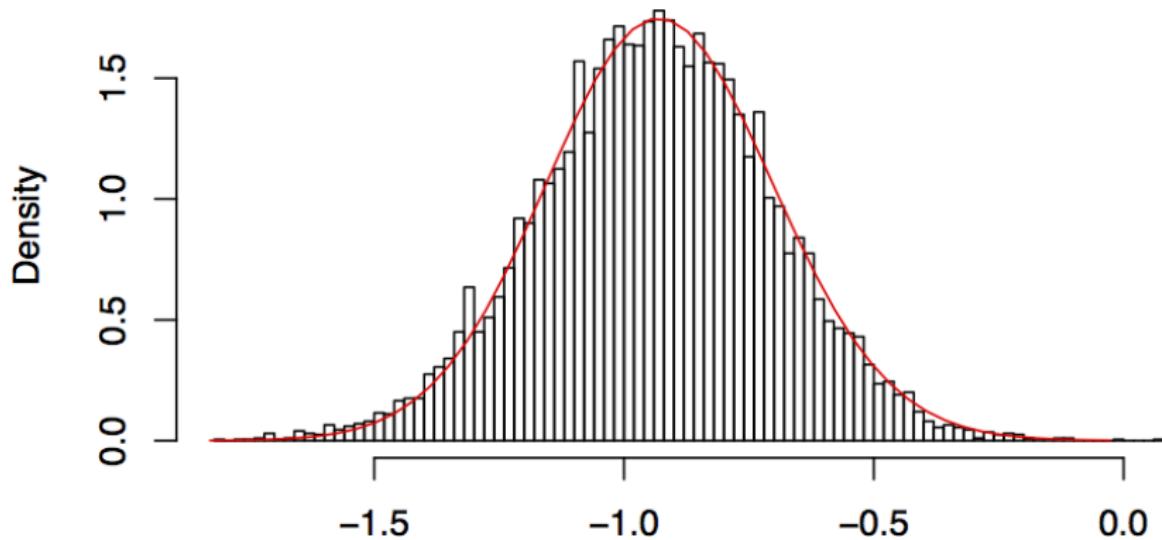
Posterior marginal for x_i :

- Use numerical integration

$$\begin{aligned}\pi(x_i | \mathbf{y}) &= \int \pi(x_i | \mathbf{y}, \tau) \pi(\tau | \mathbf{y}) d\tau \\ &\approx \sum_k \mathcal{N}(x_i; \mu_{\tau_k}, \sigma^2(\tau_k)) \times \tilde{\pi}(\tau_k | \mathbf{y}) \times \Delta_k\end{aligned}$$

Results: Comparison for mean of p 

The red curve shows deterministic estimate and the black curve shows the estimate using the independence sampler.

Results: Comparison for x_1 

Histogram of the samples for x_1 using the improved independence sampler shown together with the estimate of the marginal distribution using numerical integration.

Integrated nested Laplace approximations (INLA)

These simple observations form the basis behind of the Integrated nested Laplace approximation (INLA) method.

Note that we are only interested in posterior marginals.

Steps in the INLA method

- ① Find a **Laplace approximation** to $\pi(\boldsymbol{\theta} | \mathbf{y})$.
- ② Find an approximation to $\pi(x_i | \boldsymbol{\theta}, \mathbf{y})$:
 - Gaussian approximation
 - Laplace approximation
 - Simplified Laplace approximation
- ③ Numerical integration
 - Grid strategy
 - Central composite design (CCD)

The INLA team



The INLA team is Håvard Rue, Sara Martino, Finn Lindgren, Daniel Simpson, Andrea Riebler, Janine Illian, and others (photo 2011)

Step 1: Laplace approximation to $\pi(\boldsymbol{\theta} \mid \mathbf{y})$.

- **Definition of conditional probability**

$$\pi(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})}{\pi(\boldsymbol{\theta} \mid \mathbf{y})} \quad \Leftrightarrow \quad \pi(\boldsymbol{\theta} \mid \mathbf{y}) = \frac{\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})}{\pi(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})}$$

- **Laplace approximation:**

Replace denominator by GMRF approximation

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

and evaluate this at the mode $\mathbf{x}^*(\boldsymbol{\theta})$. Note that

$$\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y}) = \pi(\boldsymbol{\theta}) \pi(\mathbf{x} \mid \boldsymbol{\theta}) \prod_i \pi(y_i \mid x_i, \boldsymbol{\theta})$$

Step 2: Estimate the marginals of the latent field

The marginals for components x_i of the latent field:

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

- Need to find estimates $\tilde{\pi}(x_i | \boldsymbol{\theta}, \mathbf{y})$.
- More challenging!
 - dimension of \mathbf{x} might be very large
- Three alternatives
 - Gaussian approx: fast but inaccurate.
 - Laplace approx: accurate but computationally demanding.
 - Simplified Laplace approx: trade-off between speed and accuracy (default in R-INLA).

Alternative 1: A Gaussian approximation

The obvious, simplest and fastest alternative is to use the GMRF-approximation.

- Remember

$$\tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y}) \propto \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{Q}^*(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))\right\},$$

- Use this to derive Gaussian approximations also to the univariate marginals

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

- Important: Need the efficient method for computing variances.
- However, the Gaussian approximation might be inaccurate in some applications.
 - True density is not necessarily symmetric.
 - Errors in locating the mode and missing skewness.

Alternative 2: The Laplace approximation

Laplace-approximations can be viewed as a correction to Gaussian approximations.

The Laplace approximation is constructed as before:

$$\tilde{\pi}(x_i \mid \boldsymbol{\theta}, \mathbf{y}) \approx \frac{\pi(\mathbf{x}, \boldsymbol{\theta} \mid \mathbf{y})}{\tilde{\pi}_{GG}(\mathbf{x}_{-i} \mid x_i, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}_{-i} = \mathbf{x}_{-i}^*(x_i, \boldsymbol{\theta})}$$

- Very accurate as the exact denominator is almost Gaussian.
- Computationally expensive!
 - Optimization and iteration steps for each x_i , $i = 1, \dots, n$.
 - n factorisations of $(n - 1) \times (n - 1)$ matrices.

Alternative 3: The simplified Laplace Approximation

A series expansion of the Laplace approximation for $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$:

- Calculate

$$\log \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) = -\frac{1}{2}x_i^2 + bx_i + \frac{1}{6}d x_i^3 + \dots$$

- Fit a skew-Normal density.
- This is sufficiently accurate for most application.
- Computationally much faster: $\mathcal{O}(n \log(n))$ for each i .

The skew-normal density

$$\pi_{SN}(z) = \frac{2}{\tau} \phi\left(\frac{z - \psi}{\tau}\right) \Phi\left(a \frac{z - \psi}{\tau}\right),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard Gaussian pdf and cdf.

- The parameters $\psi > 0$, $\tau > 0$ and a represent location, scale and skewness, respectively.

Step 3: Numerical integration

- Step 1 and 2:

$$\begin{aligned}\tilde{\pi}(x_i \mid \mathbf{y}) &= \int \tilde{\pi}(x_i \mid \boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}. \\ \tilde{\pi}(\theta_j \mid \mathbf{y}) &= \int \tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}_{-j}.\end{aligned}$$

- Step 3: Numerical integration

Approximate marginals for the latent field:

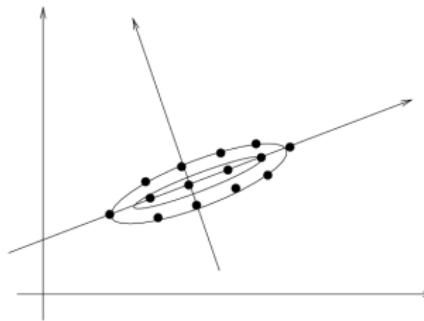
$$\tilde{\pi}(x_i \mid \mathbf{y}) \approx \sum_k \tilde{\pi}(x_i \mid \theta_k, \mathbf{y}) \tilde{\pi}(\theta_k \mid \mathbf{y}) \Delta_k$$

where Δ_k denotes the area weight corresponding to θ_k .

Integration with respect to θ

Need to explore $\tilde{\pi}(\theta|y)$ numerically to find suitable set of integration points $\{\theta_k\}$

- Locate the mode
- Use the Hessian to construct new variables
- Grid-search



Step 3: Strategies, marginals for latent field

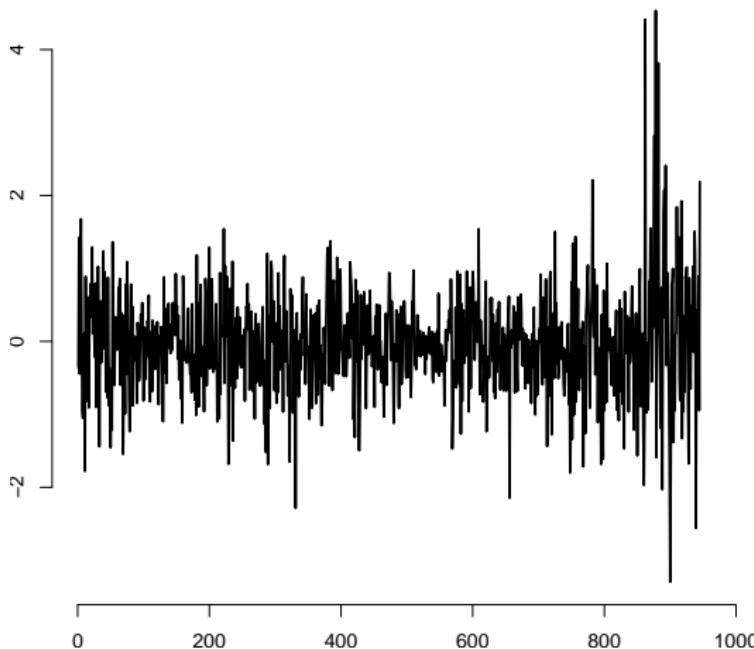
Grid strategy:

- Accurate but might be time-consuming if $m = \dim(\boldsymbol{\theta})$ is large.

Central composite design (CCD):

- Use small amount of support points in the m -dimensional space of $\boldsymbol{\theta}$.
- Augment each center point with a group of points used to estimate the curvature of $\tilde{\pi}(\boldsymbol{\theta} | \mathbf{y})$.
- Much faster than grid-strategy as fewer integration points are needed.
- Default strategy in R-INLA.

Example: Stochastic Volatility



Log of the daily difference of the pound-dollar exchange rate from October 1st, 1981, to June 28th, 1985.

Stochastic Volatility model

The likelihood of the data, conditional on the latent variables is

$$y_t | \eta_t \sim \mathcal{N}(0, \exp(\eta_t))$$

The model for the latent variables is

$$\eta_t = \mu + x_t, t = 0, \dots, T$$

where μ is an unknown mean and x_t is modelled as an AR(1) process:

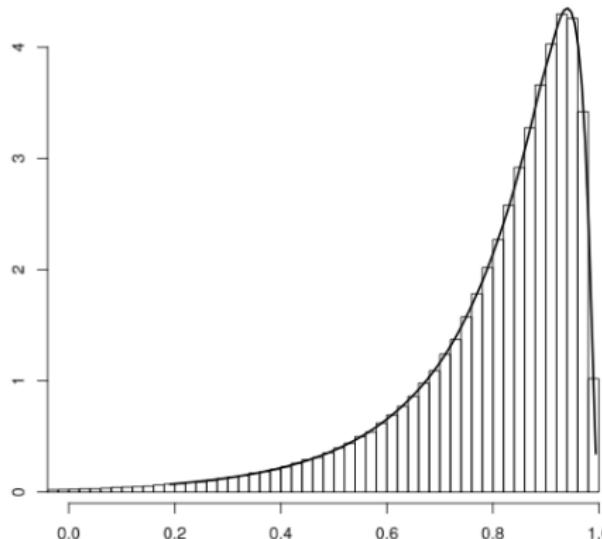
$$x_t = \phi x_{t-1} + \tau^{-1} e_t$$

We assume a Gaussian prior for μ , a Gamma prior for τ , and model ϕ using a logit link (to ensure that $|\phi| < 1$):

$$\kappa = \text{logit}\left(\frac{\phi + 1}{2}\right) \sim \mathcal{N}(0, 10000)$$

Results: $\nu = \text{logit}(2\phi - 1)$

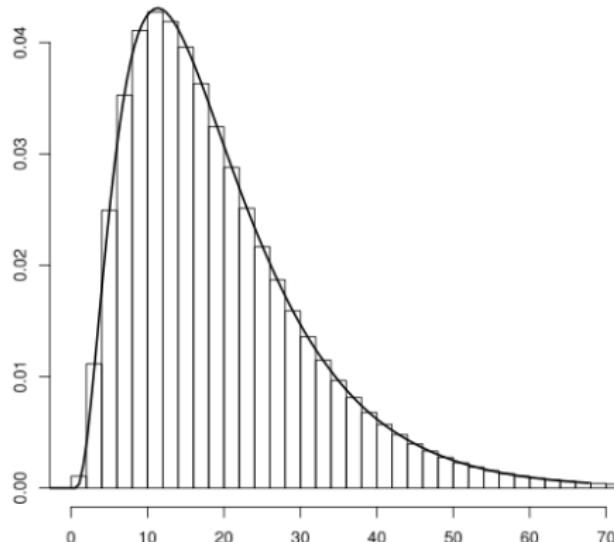
Using just the first 50 data-points only, which makes the problem harder



The histogram is obtained from a long MCMC run.

Results: $\log(\kappa_x)$

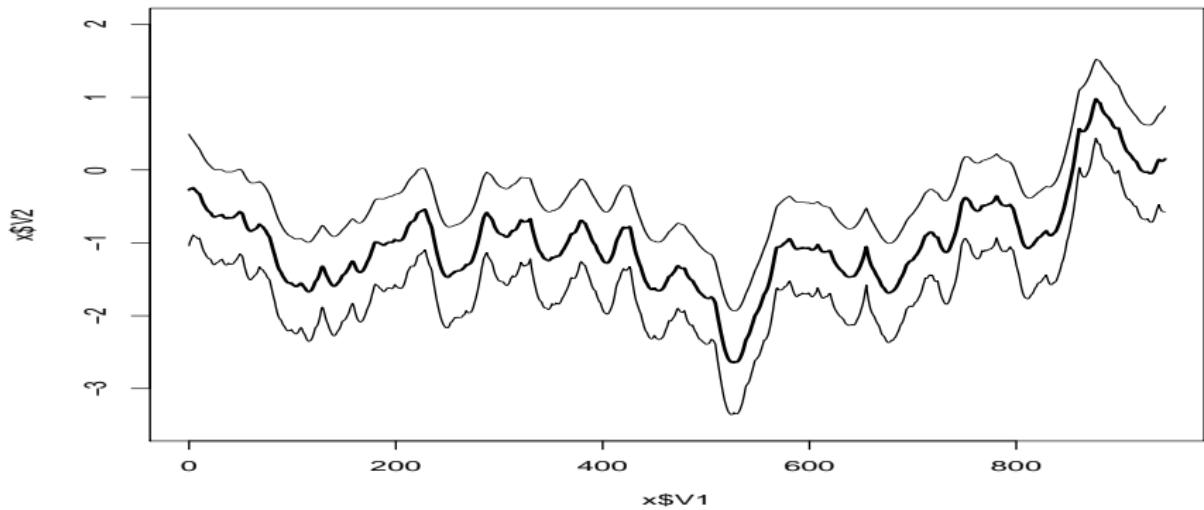
Using just the first 50 data-points only, which makes the problem harder



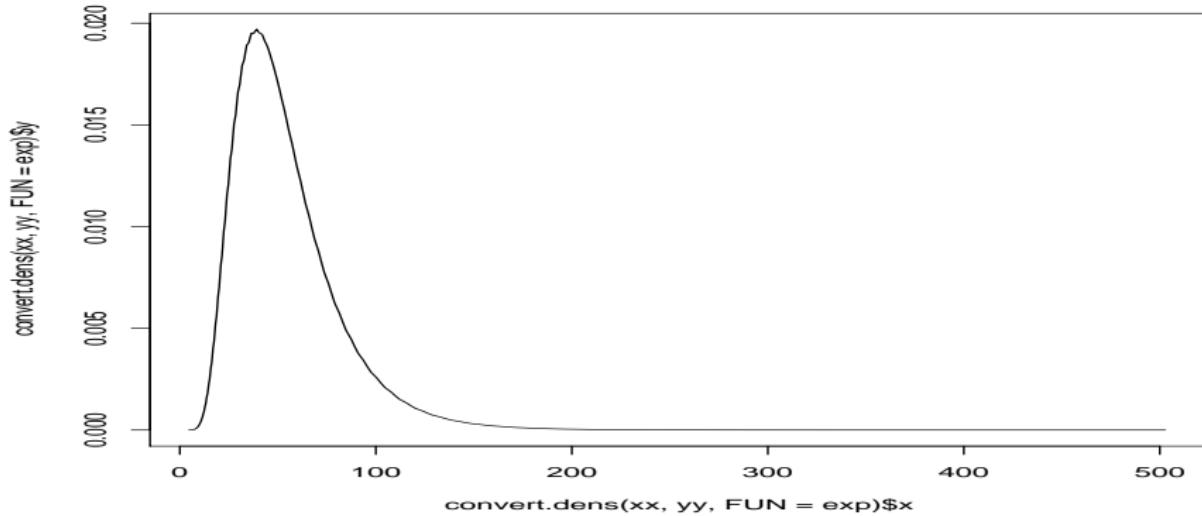
The histogram is obtained from a long MCMC run.

Results: Mean of $x_t + \mu$

Using the full dataset

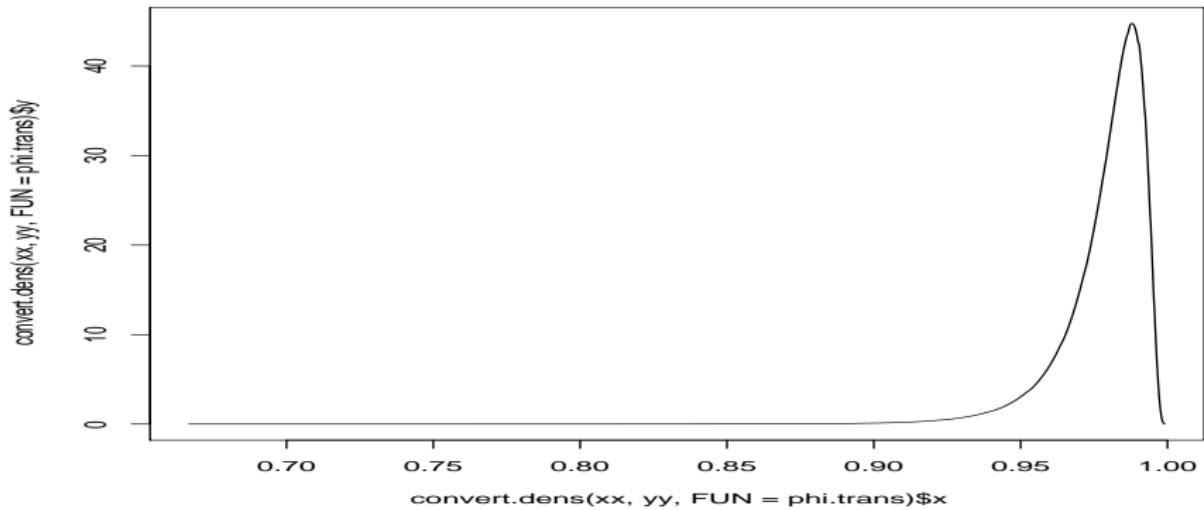


Using the full dataset



The posterior marginal for the precision.

Using the full dataset



The posterior marginal for the lag-1 correlation.