Latent Gaussian models

Sara Martino
Department of Mathematical Sciences
NTNU, Norway

November 2008
Outline I

Latent Gaussian models - Definition
- Three examples in details
  - Stochastic volatility model
  - Longitudinal mixed effect model
  - Disease Mapping
- Merging GMRF using conditioning
- Generalized additive (mixed) models

Examples of latent Gaussian Models
- Examples: 1D
- Examples 2D
- Examples 2D+

Latent Gaussian models: Tasks
A simple time series case: The Tokyo rainfall data

We want to model the probability of rain for every day of the year having observed if it rained or not for 2 years.
Latent Gaussian models

Characterised through several stages of observables and parameters.

A typical scenario is as follows.

**Stage 1** Formulate a distributional assumption for the observables, dependent on latent parameters.

- Time series of binary observations $y$, we may assume

  $$y_i, \quad i = 1, \ldots, n : y_i \sim \mathcal{B}(p_i)$$

- We assume the observations to be *conditionally independent*
A simple time series case: The Tokyo rainfall data

Stage 1  Binomial data

\[ y_i \sim \begin{cases} 
\text{Binomial}(2, p(x_i)) \\
\text{Binomial}(1, p(x_i)) 
\end{cases} \]
Latent Gaussian models

Characterised through several *stages* of observables and parameters.

A typical scenario is as follows.

**Stage 2** Assign a prior model, i.e. a Gaussian model, for the unknown parameters, here $p_i$.

- Chose an autoregressive model for the logit-transformed probabilities $x_i = \text{logit}(p_i)$. 
Tokyo rainfall data

Stage 2 Assume a smooth latent $x$,

$$x \sim RW2(\kappa), \quad \text{logit}(p_i) = x_i$$
Latent Gaussian models

Characterised through several *stages* of observables and parameters.

A typical scenario is as follows.

**Stage 3** Assign to unknown parameters (or hyperparameters) of the GMRF

- precision parameter $\kappa$
- “strength” of dependency.

Further stages if needed.
Tokyo rainfall data

Stage 3  Gamma($\alpha, \beta$)-prior on $\kappa$
Latent Gaussian Models - A general set-up

Stage 1  Hyperparameters $\theta$ (low dimension)

Stage 2  Gaussian model $x|\theta$ of size $n$ (large)

Stage 3  Observe some of the $x$ through data $y$

We are interested in the posterior distribution of $x, \theta$ given the observed data $y$

$$\pi(x, \theta|y) \propto \pi(\theta)\pi(x|\theta)\pi(y|x, \theta)$$
Latent Gaussian Markov Models

- We focus on Latent Gaussian Markov Models i.e. models where the latent Gaussian field is endowed with Markov properties.
- We specify the precision matrix $Q(\theta) = \Sigma^{-1}(\theta)$. This matrix is sparse.
- Approximate inference is possible also for non-Markov random fields using different computational tools (will not consider this.)
Latent Gaussian Models

We focus on Latent Gaussian Markov Models i.e. models where the latent Gaussian field is endowed with Markov properties.

We specify the precision matrix $Q(\theta) = \Sigma^{-1}(\theta)$. This matrix is sparse.

All computations take advantage of the sparseness of $Q$.

Approximate inference is possible also for non-Markov random fields using different computational tools (will not consider this.)
Latent Gaussian Markov Models

- We focus on Latent Gaussian Markov Models i.e. models where the latent Gaussian field is endowed with Markov properties.
- We specify the precision matrix $Q(\theta) = \Sigma^{-1}(\theta)$. This matrix is sparse.
- All computations take advantage of the sparseness of $Q$.
- Approximate inference is possible also for non-Markov random fields using different computational tools (will not consider this.)
Latent Gaussian Models

Latent Gaussian Markov Models

- We focus on Latent Gaussian Markov Models i.e. models where the latent Gaussian field is endowed with Markov properties.
- We specify the precision matrix $Q(\theta) = \Sigma^{-1}(\theta)$. This matrix is sparse.
- All computations take advantage of the sparseness of $Q$
- Approximate inference is possible also for non-Markov random fields using different computational tools (will not consider this.)
Latent Gaussian Models

This apparently simple structure includes many often used statistical models. We will see three examples in details:

- Stochastic volatility model
- Generalized mixed model for longitudinal data
- Model for disease mapping

...but many more models can be seen as part of latent GMRF.
Stochastic volatility model

Observed daily difference of the pound-dollar exchange rate from October 1st 1981 to June 28th 1985
Stochastic volatility model

- Conditional independent data:
  \[ y_t \mid h_t \sim \mathcal{N}(0, \exp(h_t)); \quad t = 1, \ldots, T \]

- Model the log-variance using a AR(1) model:
  \[ h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t \]

- Assign prior to all element of the model:
  \[
  \begin{align*}
  \mu & \sim \mathcal{N}(0, \tau_{\mu}) \\
  \phi & \sim \text{Unif}(0, 1) \\
  \tau & \sim \text{Gamma}(a, b)
  \end{align*}
  \]
Stochastic volatility model

- Conditional independent data:
  \[ y_t | h_t \sim \mathcal{N}(0, \exp(h_t)); \ t = 1, \ldots, T \]

- Model the log-variance using a AR(1) model:
  \[ h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t \]

- Assign prior to all element of the model:
  \[
  \begin{align*}
    \mu & \sim \mathcal{N}(0, \tau_\mu) \\
    \phi & \sim \text{Unif}(0, 1) \\
    \tau & \sim \text{Gamma}(a, b)
  \end{align*}
  \]
Latent Gaussian Models

Three examples in details

Stochastic volatility model

- Conditional independent data:
  \( y_t | h_t \sim \mathcal{N}(0, \exp(h_t)); \ t = 1, \ldots, T \)

- Model the log-variance using a AR(1) model:
  \[ h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t \]

- Assign prior to all element of the model:
  \[ \mu \sim \mathcal{N}(0, \tau_\mu) \]
  \[ \phi \sim \text{Unif}(0, 1) \]
  \[ \tau \sim \text{Gamma}(a, b) \]
Stochastic volatility model

- Conditional independent data:
  \[ y_t | h_t \sim N(0, \exp(h_t)); \quad t = 1, \ldots, T \]

- Model the log-variance using a AR(1) model:
  \[ h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t \]

- Assign prior to all element of the model:
  \[ \mu \sim N(0, \tau_\mu) \]
  \[ \phi \sim \text{Unif}(0, 1) \]
  \[ \tau \sim \text{Gamma}(a, b) \]

What is our latent field? Which are the hyperparameters?
Stochastic volatility model

Stage 1 \( \pi(y_t|x_t) = \mathcal{N}(0, \exp(x_t)); \ t = 1, \ldots, T \)

Stage 2 \( x|\theta \sim \mathcal{N}(0, Q(\theta)), \ |x| = T + 1 \)

Stage 3 \( \pi(\theta) \)
Longitudinal mixed effects model: Epil-example from BUGS

<table>
<thead>
<tr>
<th>Patient</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
<th>Trt</th>
<th>Base</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>11</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>20</td>
<td>21</td>
<td>12</td>
<td>0</td>
<td>52</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>12</td>
<td>37</td>
</tr>
<tr>
<td>....</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>37</td>
</tr>
</tbody>
</table>
Longitudinal mixed effects model: Epil-example from BUGS

\[ y_{ij} \sim \text{Poisson}(\exp(\eta_{ij})); \quad i = 1, \ldots, 59; \quad j = 1, \ldots, 4 \]

\[ \eta_{ij} = \beta_0 + \beta_1 \log(\text{Base}_i/4) + \beta_2 \text{Trt}_j + \beta_3 \text{Trt}_j \log(\text{Base}_y/4) + \beta_4 \text{Age}_j + \beta_5 V_4 + \epsilon_j + \nu_{ij} \]

\[ \beta_k \sim \mathcal{N}(0, \tau_k^{-1}); \quad k = 0, \ldots, 5 \]

\[ \epsilon_i \sim \mathcal{N}(0, \tau_\epsilon^{-1}); \quad i = 1 \ldots, 59 \]

\[ \nu_{ij} \sim \mathcal{N}(0, \tau_\nu^{-1}); \quad i = 1 \ldots, 59; \quad j = 1, \ldots, 4 \]

\[ \tau_k \quad \text{Known} \]

\[ \tau_\epsilon, \tau_\nu \quad \text{Unknown} \]
Longitudinal mixed effects model: Epil-example from BUGS
Latent Gaussian Models

Latent Gaussian models - Definition

Three examples in details

Longitudinal mixed effects model: Epil-example from BUGS

- What is the latent GMRF?
- Which are the hyperparameters?
- How does the $Q$ matrix look like?
Longitudinal mixed effects model: Epil-example from BUGS

Non-zero structure of the Q matrix
Longitudinal mixed effects model: Epil-example from BUGS

Stage 1  Data: $\pi(y_{ij}|\eta_{ij}) \sim \text{Poisson} (\exp(\eta_{ij}))$, $i = 1, \ldots, 59; j = 1, \ldots, 4$

Stage 2  Latent GMRF: $\mathbf{x} = \{\eta, \epsilon, \beta_0, \ldots, \beta_5\}$ with $|\mathbf{x}| = 59 \times 4 + 59 + 5 = 300$

$\pi(\mathbf{x}|\theta) \sim \mathcal{N}(0, \mathbf{Q}(\theta)^{-1})$

NB: Latent GMRF parametrised using $\eta$ instead of $\nu$.

Stage 3  Hyperparameters $\theta = \{\tau_\eta, \tau_\epsilon\}$
In each of the 544 region of Germany the number of oral cancer cases is registered.
Disease Mapping - Germany cancer data

\[ y_i \sim \text{Poisson}(E_i \exp(\eta_i)), \quad i = 1, \ldots, 544 \]

\[ \eta_i = s_i + u + i \]

where

- \( s = \{s_1, \ldots, s_{544}\} \) is a Spatially Structured term (CAR model)
- \( u = \{u_1, \ldots, u_{544}\} \) is a Spatially Unstructured term (IID model)
Disease Mapping - Germany cancer data

- Spatially structured term:

\[ \pi(s|\tau_s) \propto \exp\left\{ -\frac{\tau_s}{2} s^T Q_s s \right\} \]

with

\[ Q_s(i, j) = \begin{cases} 
\text{numb. neig. of } i & \text{if } i = j \\
-1 & \text{if } i \sim j \\
0 & \text{otherwise}
\end{cases} \]

- Spatially unstructured term:

\[ \pi(u|\tau_\eta) \propto \exp\left\{ -\frac{\tau_\eta}{2} u^T I u \right\} \]
Disease Mapping - Germany cancer data

Non-zero structure of the $Q$ matrix
Disease Mapping - Germany cancer data

Stage 1 Data: \( \pi(y_{ij}|\eta_{ij}) \sim \text{Poisson}(\exp(\eta_{ij})), \ i = 1, \ldots, 544 \)

Stage 2 Latent GMRF: \( x = \{\eta, s\} \) with \( |x| = 2 \times 544 \)

\[ \pi(x|\theta) \sim \mathcal{N}(0, Q(\theta)^{-1}) \]

**NB:** Latent GMRF parametrised using \( \eta \) instead of \( u \).

Stage 3 Hyperparameters \( \theta = \{\tau_\eta, \tau_s\} \)
Merging GMRFs using conditioning (I)

In general it is very easy to create more and more complex GMRF using conditioning....
Latent Gaussian Models

- Latent Gaussian models - Definition
- Merging GMRF using conditioning

Merging GMRFs using conditioning (II)

\[
\mu \sim \mathcal{N}(0, 1)
\]

- \( x^* = (\mu, x, z, y) \) is a GMRF
- \( x^* \mid y \) is a GMRF
Latent Gaussian Models

- Latent Gaussian models - Definition
- Merging GMRF using conditioning

Merging GMRFs using conditioning (II)

\[ \mu \sim \mathcal{N}(0, 1) \]
\[ \mathbf{x} - \mu | \mu \sim \text{AR}(1) \]

- \( \mathbf{x}^* = (\mu, \mathbf{x}, \mathbf{z}, \mathbf{y}) \) is a GMRF
- \( \mathbf{x}^* | \mathbf{y} \) is a GMRF
Merging GMRFs using conditioning (II)

\[ \mu \sim \mathcal{N}(0,1) \]
\[ \mathbf{x} - \mu \mid \mu \sim \text{AR}(1) \]
\[ \mathbf{z} \mid \mathbf{x} \sim \mathcal{N}(\mathbf{x}, \mathbf{I}) \]

- \( \mathbf{x}^* = (\mu, \mathbf{x}, \mathbf{z}, \mathbf{y}) \) is a GMRF
- \( \mathbf{x}^* \mid \mathbf{y} \) is a GMRF
Merging GMRFs using conditioning (II)

\[ \mu \sim \mathcal{N}(0, 1) \]
\[ x - \mu \mid \mu \sim \text{AR}(1) \]
\[ z \mid x \sim \mathcal{N}(x, I) \]
\[ y \mid z \sim \mathcal{N}(z, I) \]

\[ x^* = (\mu, x, z, y) \] is a GMRF

\[ x^* \mid y \] is a GMRF
Latent Gaussian Models

Latent Gaussian models - Definition
Merging GMRF using conditioning

Merging GMRFs using conditioning (II)

\[
\begin{align*}
\mu & \sim \mathcal{N}(0, 1) \\
x - \mu | \mu & \sim \text{AR}(1) \\
z | x & \sim \mathcal{N}(x, I) \\
y | z & \sim \mathcal{N}(z, I)
\end{align*}
\]

- \( x^* = (\mu, x, z, y) \) is a GMRF
- \( x^* | y \) is a GMRF
Latent Gaussian Models

- Latent Gaussian models - Definition
- Merging GMRF using conditioning

Merging GMRFs using conditioning (II)

\[
\begin{align*}
\mu & \sim \mathcal{N}(0, 1) \\
x - \mu | \mu & \sim \text{AR}(1) \\
z | x & \sim \mathcal{N}(x, I) \\
y | z & \sim \mathcal{N}(z, I)
\end{align*}
\]

- \( x^* = (\mu, x, z, y) \) is a GMRF
- \( x^* | y \) is a GMRF
Latent Gaussian Models

- Latent Gaussian models - Definition
- Merging GMRF using conditioning

Merging GMRFs using conditioning (II)

\[
\begin{align*}
\mu & \sim \mathcal{N}(0, 1) \\
\mathbf{x} - \mu | \mu & \sim \text{AR}(1) \\
\mathbf{z} | \mathbf{x} & \sim \mathcal{N}(\mathbf{x}, \mathbf{I}) \\
\mathbf{y} | \mathbf{z} & \sim \mathcal{N}(\mathbf{z}, \mathbf{I})
\end{align*}
\]

- \( \mathbf{x}^* = (\mu, \mathbf{x}, \mathbf{z}, \mathbf{y}) \) is a GMRF
- \( \mathbf{x}^* | \mathbf{y} \) is a GMRF
- Additional hyperparameters \( \theta \)
Latent Gaussian Models

Merging GMRF using conditioning

If

\[ x \sim \mathcal{N}(0, Q^{-1}) \]
\[ y \mid x \sim \mathcal{N}(x, K^{-1}) \]
\[ z \mid x, y \sim \mathcal{N}(y, H^{-1}) \]

then

\[
\text{Prec}(x, y, z) = \begin{bmatrix}
Q + K & -K & 0 \\
-K & K + H & -H \\
0 & -H & H
\end{bmatrix}
\]

which is sparse if \( Q, K \) and \( H \) are.
Generalised additive (mixed) models

All models seen untill now can be written as:

\[
g(\mu_i) = \sum_j f_j(z_{ji}) + \sum_k \beta_j \tilde{z}_{ji} + \epsilon_i
\]

where

- each \( f_j(\cdot) \), is a smooth (random) function
- \( \beta_j \) is the linear effect of \( z_j \)

Observations \( \{y_i\} \) from an exponential family with mean \( \{\mu_i\} \)
(NB: in the stoch. vol. example we model the variance instead of the mean...)
Examples of latent Gaussian Models

1D Smoothing count data, general spline smoothing, semi-parametric regression, GLM(M), GAM(M), etc

2D Disease mapping, log-Gaussian Cox-processes, model-based geostatistics, 1D-models with spatial effect(s)

3D Time-series of images, spatio-temporal models.

Features

- Dimension of the latent Gaussian field, \( n \), is large, \( 10^2 - 10^5 \), but often Markov.
- Dimension of the hyperparameters \( \text{dim}(\theta) \) is small, \( 1 - 5 \), say.
- Dimension of the data \( \text{dim}(y) \) might vary, but is often non-Gaussian.
Examples of latent Gaussian Models

1D  Smoothing count data, general spline smoothing, semi-parametric regression, GLM(M), GAM(M), etc

2D  Disease mapping, log-Gaussian Cox-processes, model-based geostatistics, 1D-models with spatial effect(s)

3D  Time-series of images, spatio-temporal models.

Features

▶ Dimension of the latent Gaussian field, $n$, is large, $10^2 - 10^5$, but often Markov.
▶ Dimension of the hyperparameters $\dim(\theta)$ is small, $1 - 5$, say.
▶ Dimension of the data $\dim(y)$ might vary, but is often non-Gaussian.
Examples of latent Gaussian Models

1D Smoothing count data, general spline smoothing, semi-parametric regression, GLM(M), GAM(M), etc

2D Disease mapping, log-Gaussian Cox-processes, model-based geostatistics, 1D-models with spatial effect(s)

3D Time-series of images, spatio-temporal models.

Features

- Dimension of the latent Gaussian field, $n$, is large, $10^2 - 10^5$, but often Markov.
- Dimension of the hyperparameters $\dim(\theta)$ is small, 1 – 5, say.
- Dimension of the data $\dim(y)$ might vary, but is often non-Gaussian.
Examples of latent Gaussian Models

1D Smoothing count data, general spline smoothing, semi-parametric regression, GLM(M), GAM(M), etc

2D Disease mapping, log-Gaussian Cox-processes, model-based geostatistics, 1D-models with spatial effect(s)

3D Time-series of images, spatio-temporal models.

Features

▶ Dimension of the latent Gaussian field, \( n \), is large, \( 10^2 \) \( \cdots \) \( 10^5 \), but often Markov.

▶ Dimension of the hyperparameters \( \text{dim}(\theta) \) is small, \( 1 \) \( \cdots \) \( 5 \), say.

▶ Dimension of the data \( \text{dim}(y) \) might vary, but is often non-Gaussian.
Examples of latent Gaussian Models

1D  Smoothing count data, general spline smoothing, semi-parametric regression, GLM(M), GAM(M), etc
2D  Disease mapping, log-Gaussian Cox-processes, model-based geostatistics, 1D-models with spatial effect(s)
3D  Time-series of images, spatio-temporal models.

Features

▶ Dimension of the latent Gaussian field, \( n \), is large, \( 10^2 - 10^5 \), but often Markov.
▶ Dimension of the hyperparameters \( \dim(\theta) \) is small, \( 1 - 5 \), say.
▶ Dimension of the data \( \dim(y) \) might vary, but is often non-Gaussian.
Examples of latent Gaussian Models

1D  Smoothing count data, general spline smoothing, semi-parametric regression, GLM(M), GAM(M), etc

2D  Disease mapping, log-Gaussian Cox-processes, model-based geostatistics, 1D-models with spatial effect(s)

3D  Time-series of images, spatio-temporal models.

Features

- Dimension of the latent Gaussian field, $n$, is large, $10^2 - 10^5$, but often Markov.
- Dimension of the hyperparameters $\dim(\theta)$ is small, $1 - 5$, say.
- Dimension of the data $\dim(y)$ might vary, but is often non-Gaussian.
Examples of latent Gaussian models: 1D
Examples of latent Gaussian models: 2D

Disease mapping: Poisson data
Examples of latent Gaussian models: 2D

Joint disease mapping: Poisson data
Examples of latent Gaussian models: 2D

Spatial GLM with Binomial data
Examples of latent Gaussian models: 2D

Log-Gaussian Cox-process; Oaks-data
Spatial logit-model with semiparametric covariates
Latent Gaussian Models

Latent Gaussian models: Tasks

**Tasks**

Compute from

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

the posterior marginals:

$$\pi(x_i \mid y), \quad \text{for some or all } i$$

and/or

$$\pi(\theta_i \mid y), \quad \text{for some or all } i$$
Latent Gaussian models are not necessarily Markov. GMRF have nice computing properties which make them easier to handle. Approximate inference is possible also with non-Markov models using different computing tools.