On the Fully Bayesian Treatment of Latent Gaussian Models using Stochastic Simulations

Maurizio Filippone

School of Computing Science
University of Glasgow
maurizio.filippone@glasgow.ac.uk

May 30th, 2012
Outline of the talk

1. Latent Gaussian Models
2. Inference in Latent Gaussian Models using MCMC
3. An application to neuroimaging data
Latent Gaussian Models - LGMs

- Class of hierarchical models

\[ p(\text{data}|\text{latent}) \quad p(\text{latent}|\text{par}) \quad p(\text{par}) \]

- \( p(\text{latent}|\text{par}) = \mathcal{N}(\text{latent}|\mu(\text{par}), K(\text{par})) \)
LGM - Logistic regression example
LGM - Log-Gaussian Cox example

- $\text{latent}$
- $\text{gain}$
- $\text{length-scale}$
Ideally - predictions for new data\(^1\) \(p(\text{data}_*|\text{data})\):

\[
\int p(\text{data}_*|\text{latent}, \text{par})p(\text{latent}, \text{par}|\text{data}) \, d\text{latent} \, d\text{par}
\]

requires the posterior distribution \(p(\text{latent}, \text{par}|\text{data})\)

\(^1\)here \text{latent} comprises also \text{latent}_*
Inference and predictions

- Posterior:

\[ p(\text{latent, par}|\text{data}) \propto p(\text{data}|\text{latent})p(\text{latent}|\text{par})p(\text{par}) \]

- usually analytically intractable
Stochastic approximations - Monte Carlo integration

- Predictions for new data $p(\text{data}_* \mid \text{data})$ is an expectation

$$\int p(\text{data}_* \mid \text{latent}, \text{par}) p(\text{latent}, \text{par} \mid \text{data}) \, d\text{latent} \, d\text{par}$$

- Monte Carlo estimation:

$$E[f(x)] = \int f(x) p(x) dx \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

with $x_i$ drawn from $p(x)$

- **good** news: asymptotically correct
- **bad** news: the variance of $E[f(x)] \to 0$ in $O(1/N)$
Stochastic approximations - MCMC

- Explore the parameter space according to the density

Often it is not possible to draw samples directly - need to set up a Markov chain
Stochastic approximations - MCMC

- Explore the parameter space according to the density

... sometimes things can go wrong - need for efficient proposal mechanisms
Complexity of LGMs

- updates of $\text{par}$ cost $O(n^3)$ operations

$$\log |K| \cdot K^{-1}$$

unless particular structures are assumed
Approximate inference

- MCMC is usually considered slow/awkward to apply to LGMs
- approximate $^2 \ p(\text{latent}|\text{par, data}) \simeq q(\text{latent}|\text{par, data})$
  - maximizing approximate $\hat{p}(\text{data}|\text{par})$ (using $q$) wrt $\text{par}$
  - numerically integrate out $\text{par}$ by quadrature or MCMC
- usually fast but:
  - still in $O(n^3)$
  - we would like to include the uncertainty on $\text{par}$
  - we might not be happy with the approximation $q$
  - quadrature can’t be employed if $\text{par}$ is large dimensional

---

$^2$Laplace Approximation, Expectation Propagation
Gibbs sampling $p(\text{latent, par} | \text{data})$

- Why is MCMC for LGMs difficult?
- obvious choice (aka Sufficient Augmentation (SA) scheme):
  - $p(\text{latent} | \text{par, data})$ (can be efficiently sampled)
  - $p(\text{par} | \text{latent})$ (bad idea - see figure)
Gibbs sampling $p(\text{latent}, \text{par}|\text{data})$

- Ancillary Augmentation (AA) scheme - reparametrization:
  \[ \text{ancillary} = L^{-1} \text{latent} \quad K = LL^T \]

- $p(\text{latent}|\text{par}, \text{data})$ (can be efficiently sampled)
- $p(\text{par}|\text{ancillary}, \text{data})$ (larger marginal posterior variance)
Other strategies

- Joint sampler by Knorr-Held and Rue (2002): propose $\text{par}'|\text{latent}$ and $\text{latent}'|\text{par}'$, $\text{data}$ and then jointly Accept/Reject
- Interweave AA and SA (ASIS, Yu and Meng 2011)
Some results

- \( \text{par} \) sampled using MH
- \( \text{latent} \) sampled using Elliptical Slice Sampling (Murray et al. 2010)

**Table:** Logistic regression - \( n = 100, \text{par} \in \mathbb{R}^2 \)

<table>
<thead>
<tr>
<th>Sampler</th>
<th>min ESS</th>
<th>( \hat{R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASIS</td>
<td>2.3%</td>
<td>1.01</td>
</tr>
<tr>
<td>SA</td>
<td>0.7%</td>
<td>1.75</td>
</tr>
<tr>
<td>AA</td>
<td>2.0%</td>
<td>1.01</td>
</tr>
<tr>
<td>KHR</td>
<td>0.9%</td>
<td>1.04</td>
</tr>
</tbody>
</table>
Take-home message

- fastest convergence and highest sample independence wrt computations: AA scheme with \texttt{par} sampled using the Metropolis-Hastings algorithm
- clever and complicated proposal are expensive and don’t seem to pay off
Parkinson syndromes data

- 62 subjects
- Early stage prediction of development of
  - Idiopathic Parkinson’s Disease (IPD)
  - Multiple System Atrophy (MSA)
  - Progressive Supranuclear Palsy (PSP)
- Given neuroimages
LGM based multiclass classification with multiple sources

- latent variables $f_c(x)$ with GP prior with covariance

$$\text{cov}(f_c(x_1), f_c(x_2)) = \sum_{s=1}^{q} w_{cs} C_s(x_1, x_2)$$

- Multinomial likelihood

$$p(\text{disease} = c | \text{latent, sources}) = \frac{\exp(f_c(x))}{\sum_{r=1}^{m} \exp(f_r(x))}$$

- this problem is aka Multiple Kernel Learning
Parkinson syndromes data - multi source

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP classifier</td>
<td>0.598</td>
</tr>
<tr>
<td>SimpleMKL</td>
<td>0.418</td>
</tr>
</tbody>
</table>
Analysis of brain regions

- for this analysis we used only the GM data source
- we used an anatomical template as in Shattuck et al. 2008 to parcellate the GM images into:
  - brainstem
  - bilateral cerebellum
  - bilateral caudate
  - bilateral middle occipital gyrus
  - bilateral putamen
  - all other regions
### Parkinson syndromes data - multi region

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP classifier</td>
<td>0.614</td>
</tr>
<tr>
<td>SimpleMKL</td>
<td>0.229</td>
</tr>
</tbody>
</table>
Conclusions and ongoing work

The fully Bayesian treatment of LGMs using MCMC is still an open question but recent advances in the field allow to tackle small to moderately large (up to a thousand) inference problems in reasonable time.

Thank you!

Questions?