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

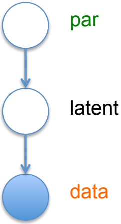
Outline of the talk

Latent Gaussian Models - LGMs

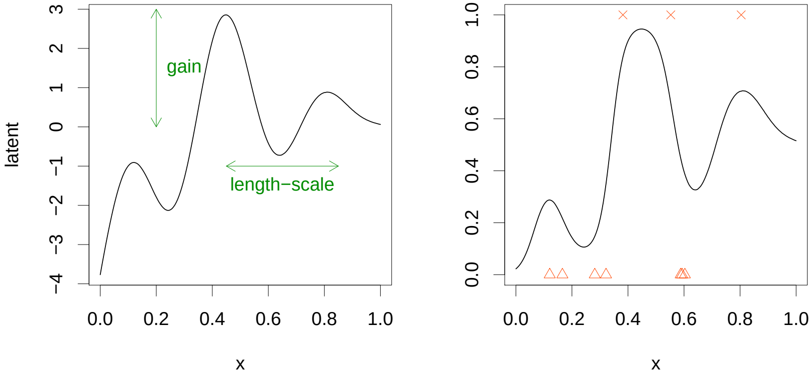
- Class of hierarchical models

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

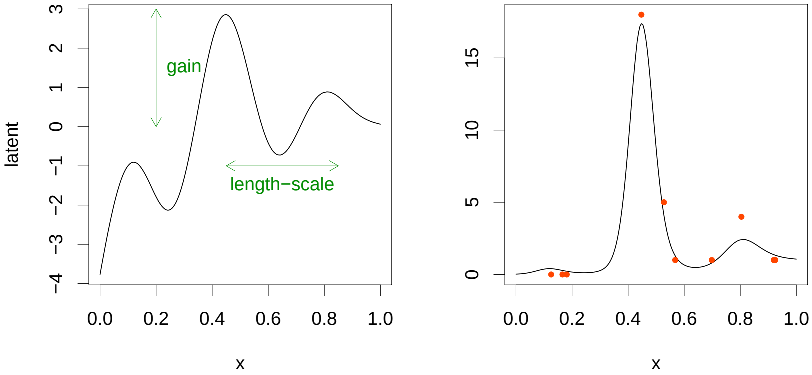
- $p(\text{latent}|\text{par}) = \mathcal{N}(\text{latent}|\boldsymbol{\mu}(\text{par}), K(\text{par}))$



LGM - Logistic regression example



LGM - Log-Gaussian Cox example

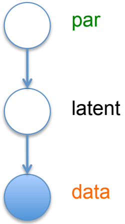


Inference and predictions

- **Ideally** - predictions for new data[1] $p(\mathrm{data}_*|\mathrm{data})$:

$$\int p(\mathrm{data}_*|\mathrm{latent}, \mathrm{par})p(\mathrm{latent}, \mathrm{par}|\mathrm{data})\, d\mathrm{latent}\, d\mathrm{par}$$

- requires the posterior distribution $p(\mathrm{latent}, \mathrm{par}|\mathrm{data})$



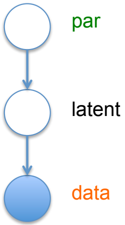
[1] here latent comprises also latent_*

Inference and predictions

- Posterior:

$$p(\mathrm{latent}, \mathrm{par} | \mathrm{data}) \propto p(\mathrm{data} | \mathrm{latent}) p(\mathrm{latent} | \mathrm{par}) p(\mathrm{par})$$

- usually analytically intractable



Stochastic approximations - Monte Carlo integration

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

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

- Monte Carlo estimation:

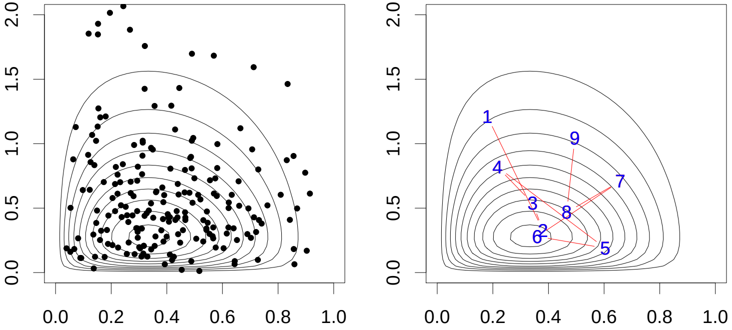
$$\mathrm{E}[f(x)] = \int f(x)p(x)dx \simeq \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

with x_i drawn from $p(x)$

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

Stochastic approximations - MCMC

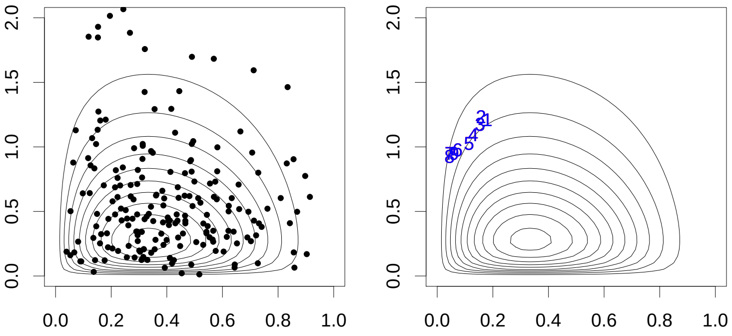
- Explore the parameter space according to the density



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 par cost $O(n^3)$ operations

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

unless particular structures are assumed

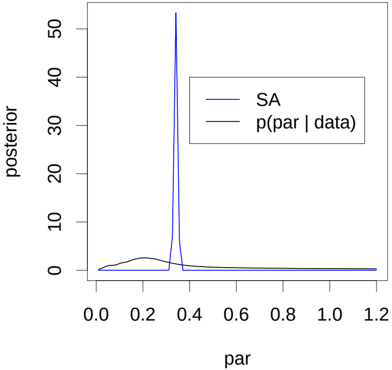
Approximate inference

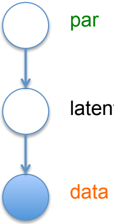
- MCMC is usually considered slow/awkward to apply to LGMs
- approximate[2] $p(\mathrm{latent}|\mathrm{par}, \mathrm{data}) \simeq q(\mathrm{latent}|\mathrm{par}, \mathrm{data})$
  - maximizing approximate $\hat{p}(\mathrm{data}|\mathrm{par})$ (using q) wrt par
  - numerically integrate out par by quadrature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

[2] Laplace Approximation, Expectation Propagation

Gibbs sampling $p(\text{latent}, \text{par}|\text{data})$

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



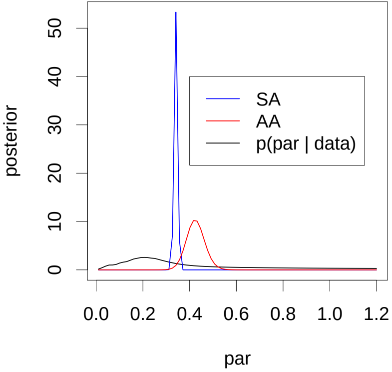


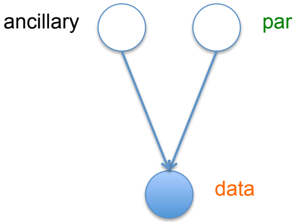
Gibbs sampling $p(\text{latent}, \text{par} | \text{data})$

- Ancillary Augmentation (AA) scheme - reparametrization:

$$\text{ancillary} = L^{-1}\,\text{latent} \qquad K = LL^{\mathrm{T}}$$

  - $p(\text{latent}|\text{par}, \text{data})$ (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

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

Table: Logistic regression - $n = 100$, par $\in \mathbb{R}^2$

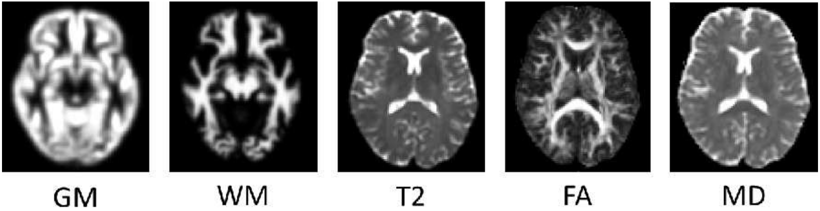
Sampler	min ESS	$\hat{R}$
ASIS	2.3%	1.01
SA	0.7%	1.75
AA	2.0%	1.01
KHR	0.9%	1.04

Take-home message

- fastest convergence and highest sample independence wrt computations: AA scheme with 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

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

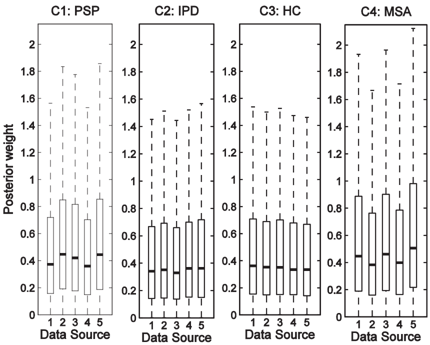
- Multinomial likelihood

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

- this problem is aka Multiple Kernel Learning

Parkinson syndromes data - multi source

Method	Accuracy
GP classifier	0.598
SimpleMKL	0.418



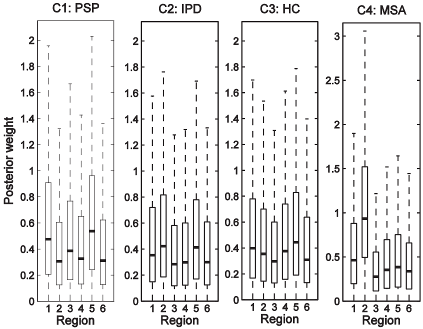
Parkinson syndromes data

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

Method	Accuracy
GP classifier	0.614
SimpleMKL	0.229



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

References

[1] M. Filippone, A.F. Marquand, C.R.V. Blain, S.C.R. Williams, J. Mourão-Miranda, and M. Girolami. **Probabilistic prediction of neurological disorders with a statistical assessment of neuroimaging data modalities**. *Annals of Applied Statistics. To appear.*

[2] M. Filippone, M. Zhong, and M. Girolami. **On the fully Bayesian treatment of latent Gaussian models using stochastic simulations**. Technical Report TR-2012-329, School of Computing Science, University of Glasgow, February 2012.

Thank you!

Questions?