Quantifying Uncertainty

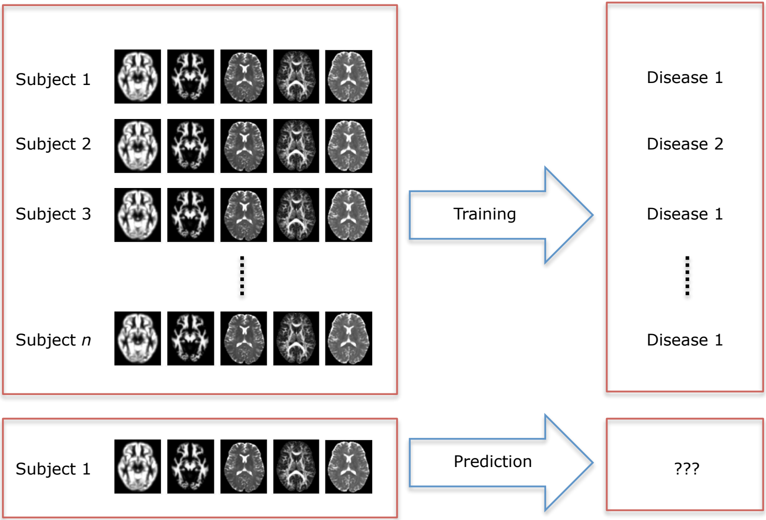
Maurizio Filippone

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

December 5th, 2013

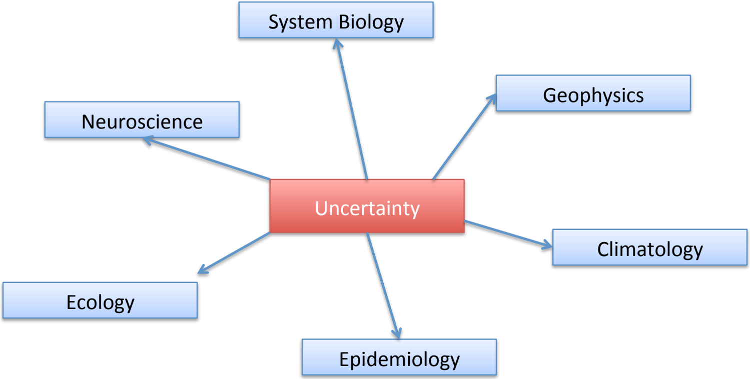
Motivating Application

Neurological disorder from neuroimages



Relevance of the problem

Relevance of the problem



Models

$$\text{label} = \text{function}(\text{input}, \text{par}, \text{noise})$$

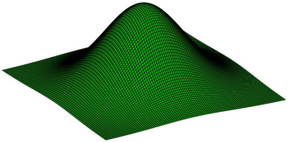
Probabilistic Modeling

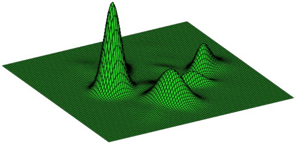
- How do we estimate model parameters?
- How do we assess that a model is preferable over another?
- How do we incorporate knowledge by experts?
- How can we attach confidence intervals to our predictions and parameter estimates?

Probabilistic modeling offers an answer to these questions

- Parameters and data are viewed as random variables

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





- Inference using Bayes theorem:

$$p(\text{par}|\text{data}) = \frac{p(\text{data}|\text{par})p(\text{par})}{\int p(\text{data}|\text{par})p(\text{par})d\text{par}}$$

Inference and predictions

- Predictions for new data

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

- Requires the posterior distribution $p(\text{par}|\text{data})$

Monte Carlo integration

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

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

- Monte Carlo estimation:

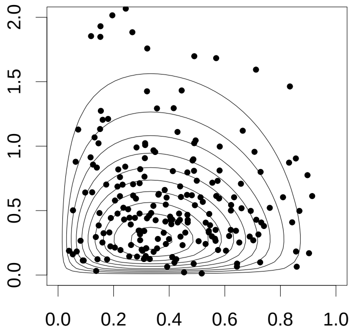
$$\int p(\mathrm{data}_*|\mathrm{par})p(\mathrm{par}|\mathrm{data})\, d\mathrm{par} \simeq \frac{1}{N}\sum_{i=1}^{N} p(\mathrm{data}_*|\mathrm{par}_i)$$

  with par_i drawn from $p(\mathrm{par}|\mathrm{data})$

- Good news: asymptotically correct
- Bad news: the "error" $\rightarrow 0$ in $O(1/\sqrt{N})$

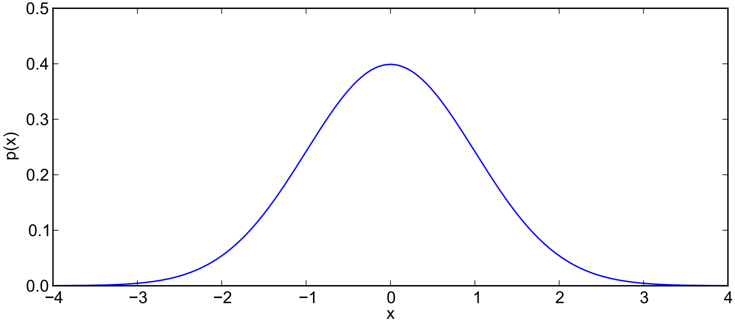
Monte Carlo integration

- Draw samples according to the density



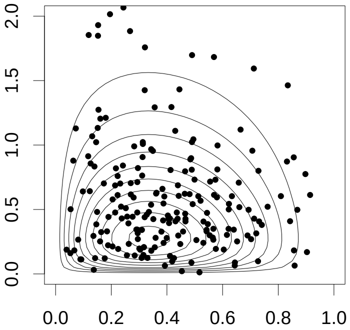
Probability density functions - An important example

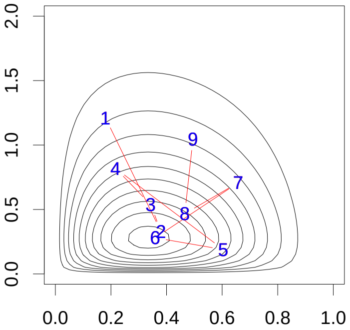
- Gaussian probability density function
- We can draw samples from this directly



Markov chain Monte Carlo

- Explore the parameter space according to the density





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

- MCMC needs the density up to a normalization constant
- Random walk sampler - accept a proposal with probability

$$\min\left(1, \frac{p(\mathrm{par}'|\mathrm{data})}{p(\mathrm{par}|\mathrm{data})}\right)$$

- by Bayes' theorem

$$p(\mathrm{par}|\mathrm{data}) = \frac{p(\mathrm{data}|\mathrm{par})p(\mathrm{par})}{\int p(\mathrm{Data}|\mathrm{Par})p(\mathrm{Par})d\mathrm{Par}}$$

- Therefore:

$$\min\left(1, \frac{p(\mathrm{data}|\mathrm{par}')p(\mathrm{par}')}{p(\mathrm{data}|\mathrm{par})p(\mathrm{par})}\right)$$

A class of hierarchical models

- Models can have more complex structures
- For example:

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

- Note: $p(\text{latent state}|\text{par})$ is actually (for example):

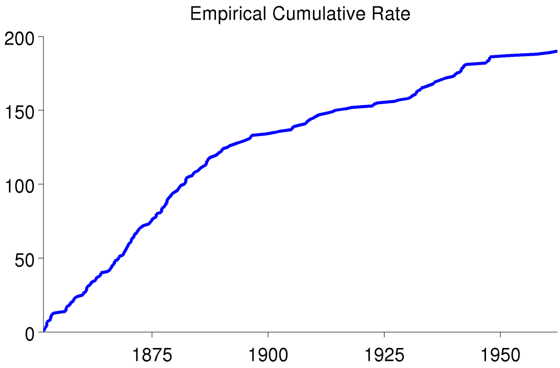
$$p(\text{latent state}|\text{par}, \text{time})$$

$$\text{or}$$

$$p(\text{latent state}|\text{par}, \text{location})$$

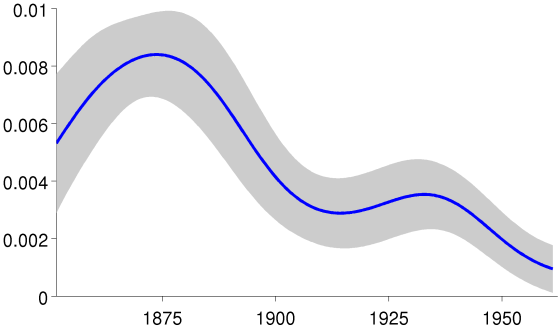
Coal mine disasters data

191 accidents between 1851 and 1962



Coal mine disasters data

191 accidents between 1851 and 1962

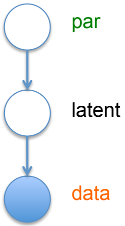


Latent Gaussian Models - LGMs

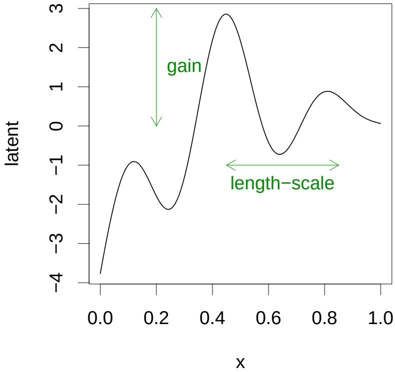
- Class of hierarchical models

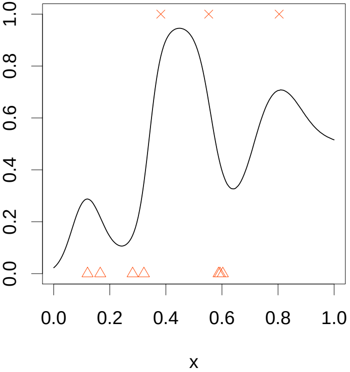
$$p(\text{data}|\text{latent}) \quad p(\text{latent}|\text{par}) \quad p(\text{par})$$

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

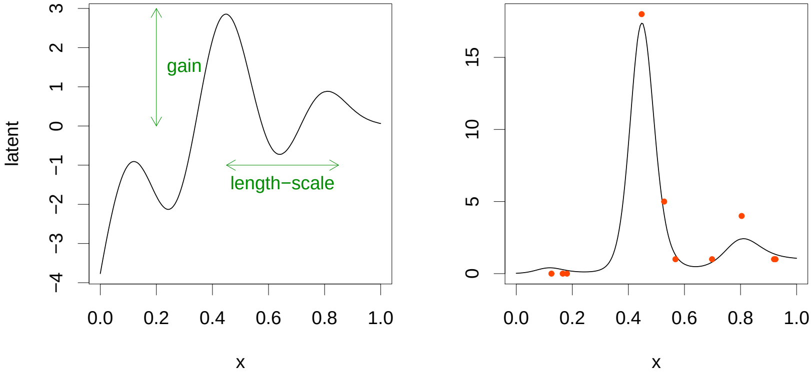


LGMs - Classification example



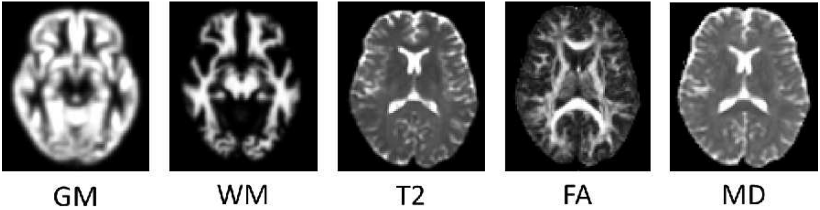


LGMs - Count data example



Parkinson syndromes data

- 62 subjects
- Early stage prediction of development of
  - Parkinson Syndromes
  - Multiple System Atrophy
  - Progressive Supranuclear Palsy
- Given neuroimages



LGM based multiclass classification with multiple sources

- Multiclass classification based on LGMs

$$p(\text{disease} = c|\text{sources}) = \text{unknown function}$$

- For every pair of subjects x_i x_j, we can characterize their similarity (covariance) $C_s(x_i, x_j)$ as given by the imaging modality s
- We do this by looking at the values of the neuroimages

LGM based multiclass classification with multiple sources

- Latent variables $f_c(x)$ with LGM prior with covariance

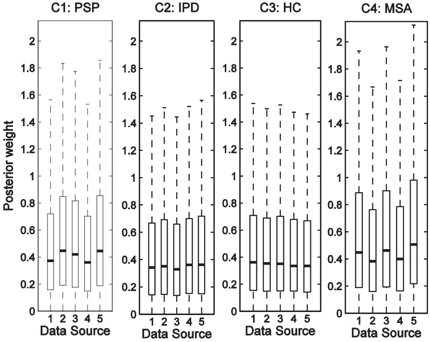
$$\mathrm{cov}(f_c(x_i), f_c(x_j)) = \sum_{s=1}^{q} w_{cs} C_s(x_i, x_j)$$

- Multinomial likelihood - general case of the logistic function

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

Parkinson syndromes data - multi source

Method	Accuracy
GP classifier	0.598
SimpleMKL	0.418



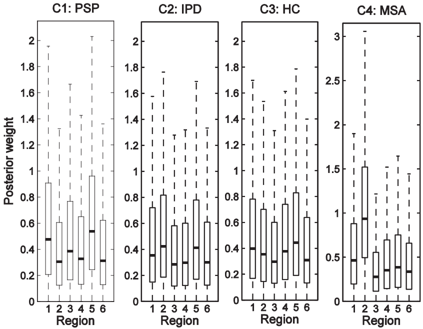
Parkinson syndromes data

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

- Bayesian inference offers the possibility to tackle challenging problems by accurately quantifying uncertainty in predictions and model parameters
- Computing is playing an important part in bridging the gap between theory and application of Bayesian inference

Thank you!

Questions?