Bayesian inference

J. Daunizeau

Institute of Empirical Research in Economics, Zurich, Switzerland
Brain and Spine Institute, Paris, France

Overview of the talk

1 Probabilistic modelling and representation of uncertainty
- *1.1 Bayesian paradigm*
- *1.2 Hierarchical models*
- *1.3 Frequentist versus Bayesian inference*

2 Numerical Bayesian inference methods
- *2.1 Sampling methods*
- *2.2 Variational methods (ReML, EM, VB)*

3 SPM applications
- *3.1 aMRI segmentation*
- *3.2 Decoding of brain images*
- *3.3 Model-based fMRI analysis (with spatial priors)*
- *3.4 Dynamic causal modelling*

Overview of the talk

1 Probabilistic modelling and representation of uncertainty
- *1.1 Bayesian paradigm*
- *1.2 Hierarchical models*
- *1.3 Frequentist versus Bayesian inference*

2 Numerical Bayesian inference methods
- *2.1 Sampling methods*
- *2.2 Variational methods (ReML, EM, VB)*

3 SPM applications
- *3.1 aMRI segmentation*
- *3.2 Decoding of brain images*
- *3.3 Model-based fMRI analysis (with spatial priors)*
- *3.4 Dynamic causal modelling*

Bayesian paradigm

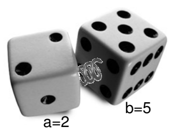
probability theory: basics

Degree of plausibility *desiderata*:

- should be represented using real numbers (D1)
- should conform with intuition (D2)
- should be consistent (D3)

a=2

• normalization: $\sum_a P(a) = 1$



a=2 b=5

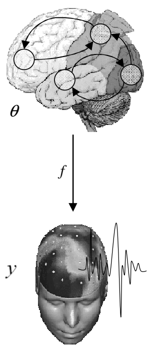
• marginalization: $P(b) = \sum_a P(a,b)$

• conditioning : (*Bayes rule*)

$$P(a,b) = P(a|b)\,P(b) = P(b|a)\,P(a)$$

Bayesian paradigm

deriving the likelihood function



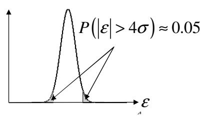
- Model of data with unknown parameters:

$y = f(\theta)$ e.g., GLM: $f(\theta) = X\theta$

- But data is noisy: $y = f(\theta) + \varepsilon$

- Assume noise/residuals is ‘small’:

$$p(\varepsilon) \propto \exp\left(-\frac{1}{2\sigma^2}\varepsilon^2\right)$$

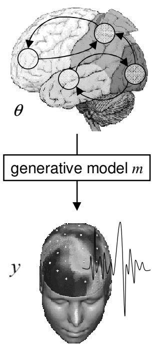


→ Distribution of data, *given fixed parameters*:

$$p(y|\theta) \propto \exp\left(-\frac{1}{2\sigma^2}\left(y - f(\theta)\right)^2\right)$$

Bayesian paradigm

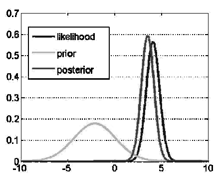
likelihood, priors and the model evidence



Likelihood: $p(y|\theta,m)$

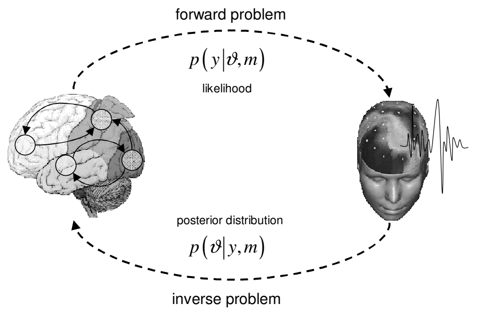
Prior: $p(\theta|m)$

Bayes rule: $$p(\theta|y,m) = \frac{p(y|\theta,m)\,p(\theta|m)}{p(y|m)}$$



Bayesian paradigm

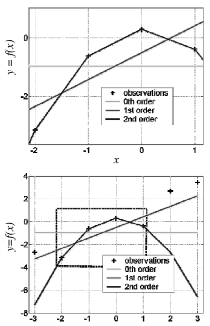
forward and inverse problems



Bayesian paradigm

model comparison

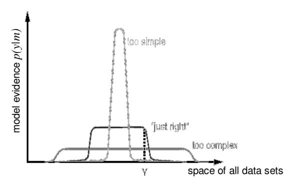
Principle of parsimony :
« plurality should not be assumed without necessity »



Model evidence:

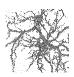
$$p\left(y|m\right)=\int p\left(y|\theta,m\right)p\left(\theta|m\right)d\theta$$

"Occam's razor" :



Hierarchical models

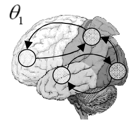
principle

θ_2 

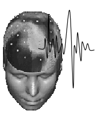
$\vdots$

$p\left(\theta_2|\theta_3,m\right)$

hierarchy

θ_1 

$p\left(\theta_1|\theta_2,m\right)$

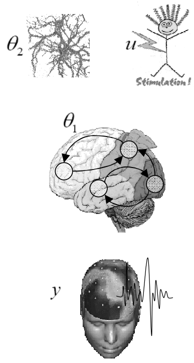
y 

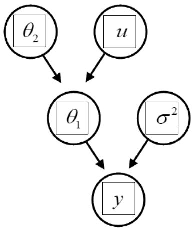
$p\left(y|\theta_1,m\right)$

causality

Hierarchical models

directed acyclic graphs (DAGs)





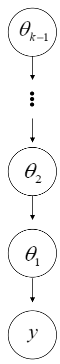
$p(\theta_1|\theta_2, u, m)$

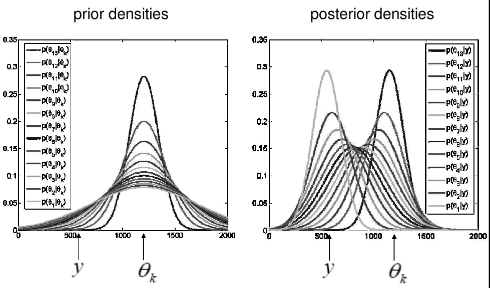
$p(y|\theta_1, \sigma^2, m)$

$$p(\theta|m) = \prod_j p(\theta_j | par(\theta_j), m)$$

Hierarchical models

univariate linear hierarchical model

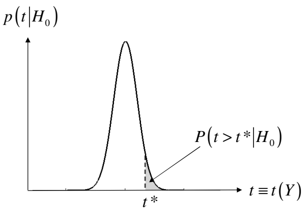




Frequentist versus Bayesian inference

a (quick) note on hypothesis testing

• define the null, e.g.: $H_0 : \theta = 0$



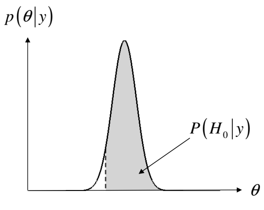
• estimate parameters (obtain test stat.)

• apply decision rule, i.e.:

if $P(t > t^* | H_0) \le \alpha$ then reject H0

classical SPM

• invert model (obtain posterior pdf)



• define the null, e.g.: $H_0 : \theta > 0$

• apply decision rule, i.e.:

if $P(H_0|y) \ge \alpha$ then accept H0

Bayesian PPM

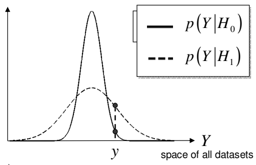
Frequentist versus Bayesian inference

what about bilateral tests?

• define the null and the alternative hypothesis *in terms of priors*, e.g.:

$$H_0 : p(\theta|H_0) = \begin{cases} 1 & \text{if } \theta = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$H_1 : p(\theta|H_1) = N(0,\Sigma)$$



• apply decision rule, i.e.: if $\dfrac{P(H_0|y)}{P(H_1|y)} \leq 1$ then reject H0

• Savage-Dickey ratios (nested models, i.i.d. priors):

$$p(y|H_0) = p(y|H_1)\boxed{\frac{p(\theta=0|y,H_1)}{p(\theta=0|H_1)}}$$

Overview of the talk

1 Probabilistic modelling and representation of uncertainty
- *1.1 Bayesian paradigm*
- *1.2 Hierarchical models*
- *1.3 Frequentist versus Bayesian inference*

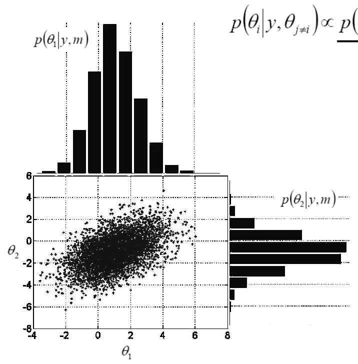
2 Numerical Bayesian inference methods
- *2.1 Sampling methods*
- *2.2 Variational methods (ReML, EM, VB)*

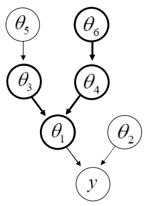
3 SPM applications
- *3.1 aMRI segmentation*
- *3.2 Decoding of brain images*
- *3.3 Model-based fMRI analysis (with spatial priors)*
- *3.4 Dynamic causal modelling*

Sampling methods

MCMC example: Gibbs sampling

$$p(\theta_i|y,\theta_{j\neq i}) \propto \underline{p(\theta_i|par(\theta_i))}\ \underline{\prod_{j=ch(i)} p(\theta_j|par(\theta_j))}$$





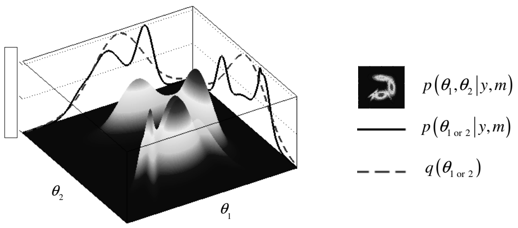
$$\frac{1}{N}\sum_{n=1}^{N} p(y|\theta^{(n)},m) \approx p(y|m)$$

Variational methods

VB / EM / ReML

$$\ln p(y|m) = \underbrace{\left\langle \ln p(\theta, y|m) \right\rangle_q + S(q)}_{\text{free energy } F(q)} + D_{KL}\left(q(\theta); p(\theta|y,m)\right)$$

→ **VB** : maximize the free energy $F(q)$ w.r.t. the "variational" posterior $q(\theta)$ under some (e.g., *mean field*, *Laplace*) approximation



Overview of the talk

1 Probabilistic modelling and representation of uncertainty

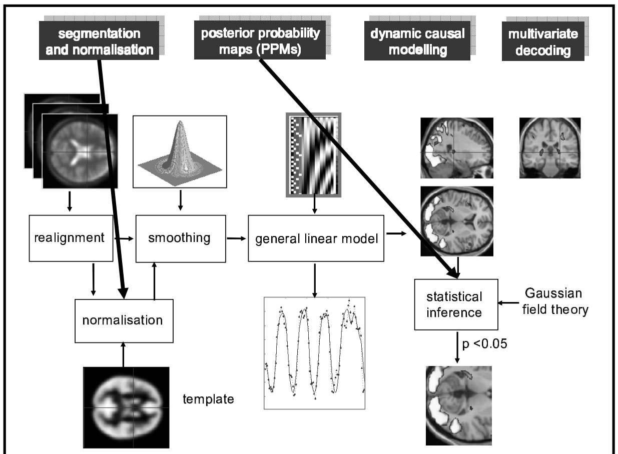
- *1.1 Bayesian paradigm*
- *1.2 Hierarchical models*
- *1.3 Frequentist versus Bayesian inference*

2 Numerical Bayesian inference methods

- *2.1 Sampling methods*
- *2.2 Variational methods (ReML, EM, VB)*

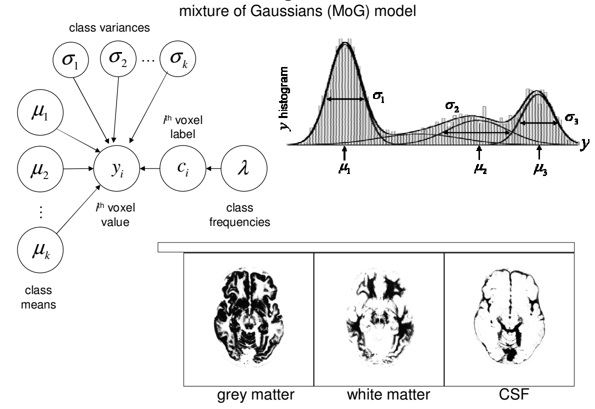
3 SPM applications

- *3.1 aMRI segmentation*
- *3.2 Decoding of brain images*
- *3.3 Model-based fMRI analysis (with spatial priors)*
- *3.4 Dynamic causal modelling*



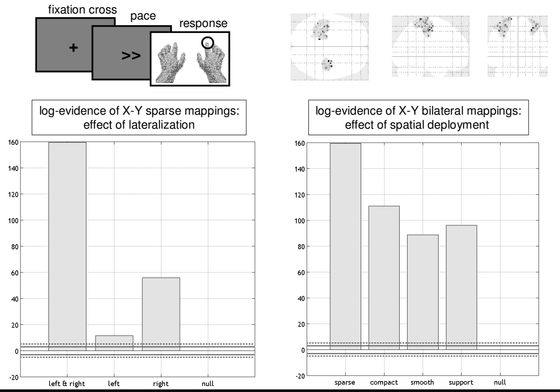
aMRI segmentation

mixture of Gaussians (MoG) model



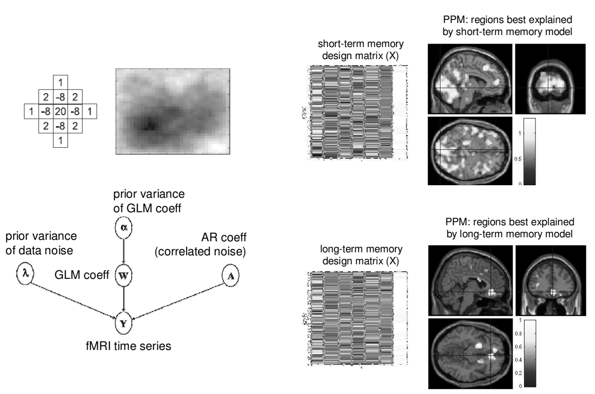
Decoding of brain images

recognizing brain states from fMRI

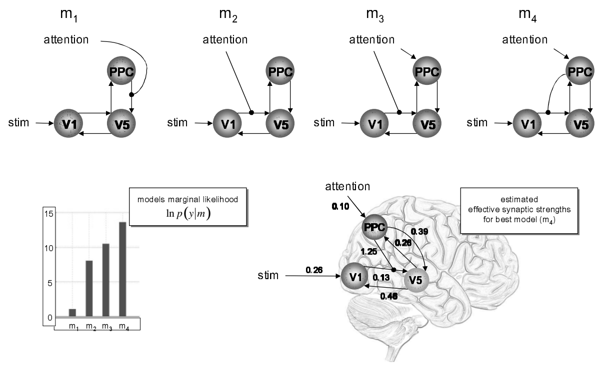


fMRI time series analysis

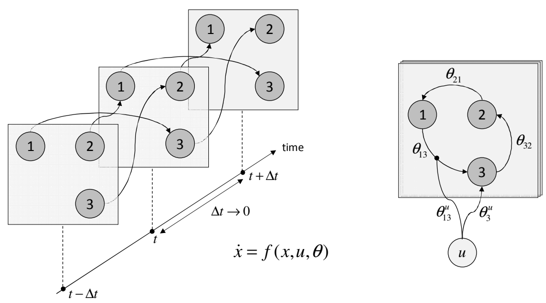
spatial priors and model comparison



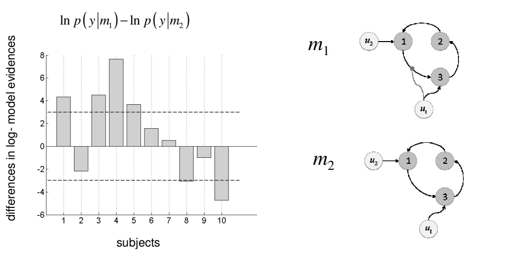
Dynamic Causal Modelling
network structure identification



DCMs and DAGs
a note on causality



Dynamic Causal Modelling
model comparison for group studies



fixed effect assume all subjects correspond to the same model

random effect assume different subjects might correspond to different models

I thank you for your attention.

A note on statistical significance

lessons from the Neyman-Pearson lemma

• Neyman-Pearson lemma: the likelihood ratio (or Bayes factor) test

$$\Lambda = \frac{p\left(y \middle| H_1\right)}{p\left(y \middle| H_0\right)} \geq u$$

is the most powerful test of size $\alpha = p\left(\Lambda \geq u \middle| H_0\right)$ to test the null.

• what is the threshold u, above which the Bayes factor test yields a error I rate of 5%?