Bayesian inference

J. Daunizeau

Brain and Spine Institute, Paris, France
Wellcome Trust Centre for Neuroimaging, London, UK
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 Notes on Bayesian inference
   2.1 Variational methods (ReML, EM, VB)
   2.2 Family inference
   2.3 Group-level model comparison

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 Notes on Bayesian inference
   2.1 Variational methods (ReML, EM, VB)
   2.2 Family inference
   2.3 Group-level model comparison

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 \((D3)\)

\[
\begin{align*}
\text{• normalization:} & \quad \sum_a P(a) = 1 \\
\text{• marginalization:} & \quad P(b) = \sum_a P(a, b) \\
\text{• conditioning :} & \quad P(a, b) = P(a|b) P(b) = P(b|a) P(a)
\end{align*}
\]
Bayesian paradigm

deriving the likelihood function

\[
\theta \rightarrow \text{Distribution of data, given fixed parameters} \rightarrow y
\]
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*

\[ p(\mathcal{Y} | \mathcal{I}, m) \]  
likelihood

\[ p(y | \mathcal{I}, m) \]  
forward problem

\[ p(\mathcal{Y} | y, m) \]  
posterior distribution

inverse problem
Bayesian paradigm

model comparison

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

Model evidence:
\[ p(y|m) = \int p(y|\theta, m)p(\theta|m) \, d\theta \]

"Occam’s razor":

Too simple

Just right

Too complex

space of all data sets
Hierarchical models

principle

\[ p(\theta_2 | \theta_3, m) \]

\[ p(\theta_1 | \theta_2, m) \]

\[ p(y | \theta_1, m) \]

\[ y \]

\[ \theta_1 \]

\[ \theta_2 \]

inference

causality
Hierarchical models

directed acyclic graphs (DAGs)

\[ p(\theta_1|\theta_2, u, m) \]

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

\[ p(\theta|m) = \prod_j p(\theta_j|\text{par}(\theta_j), m) \]
Frequentist versus Bayesian inference

*a (quick) note on hypothesis testing*

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

\[
p(t|H_0)
\]

\[
P(t > t^*|H_0)
\]

- estimate parameters (obtain test stat.)

- apply decision rule, i.e.: 

  \[
  \text{if } P(t > t^*|H_0) \leq \alpha \text{ then reject } H_0
  \]

  classical (null) hypothesis testing
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 Notes on Bayesian inference
   2.1 Variational methods (ReML, EM, VB)
   2.2 Family inference
   2.3 Group-level model comparison

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
Variational methods
VB / EM / ReML

\[
\ln p(y|m) = \left< \ln p(\theta, y|m) \right>_q + S(q) + D_{KL}(q(\theta); p(\theta|y,m))
\]

→ VB : maximize the free energy \( F(q) \) w.r.t. the approximate posterior \( q(\theta) \) under some (e.g., mean field, Laplace) simplifying constraint
Family-level inference

trading inference resolution against statistical power

\[
P(m_1|y) = 0.04
\]

\[
P(m_2|y) = 0.25
\]

\[
P(m_1|y) = 0.01
\]

\[
P(m_2|y) = 0.7
\]

model selection error risk:

\[
P(e = 1|y) = 1 - \max_m P(m|y)
\]

\[
= 0.3
\]
Family-level inference

trading inference resolution against statistical power

\[ P(m_1|y) = 0.04 \]
\[ P(m_2|y) = 0.01 \]
\[ P(f_1|y) = 0.05 \]

\[ P(m_2|y) = 0.25 \]
\[ P(m_2|y) = 0.7 \]
\[ P(f_2|y) = 0.95 \]

model selection error risk:

\[ P(e = 1|y) = 1 - \max_{m} P(m|y) \]
\[ = 0.3 \]

family inference
(pool statistical evidence)

\[ P(f|y) = \sum_{m \in f} P(m|y) \]

\[ P(e = 1|y) = 1 - \max_{f} P(f|y) \]
\[ = 0.05 \]
Group-level model comparison

**preliminary: Poly'a's urn**

\[
\begin{align*}
  m_i &= 1 \quad \rightarrow \text{\(i\)th marble is blue} \\
  m_i &= 0 \quad \rightarrow \text{\(i\)th marble is purple}
\end{align*}
\]

\( r = \text{proportion of blue marbles in the urn} \)

\( \rightarrow \) (binomial) probability of drawing a set of \(n\) marbles:

\[
p(m \mid r) = \prod_{i=1}^{n} r^{m_i} (1 - r)^{1-m_i}
\]

Thus, our belief about the proportion of blue marbles is:

\[
p(r \mid m) \propto p(r) \prod_{i=1}^{n} r^{m_i} (1 - r)^{1-m_i} \quad \Rightarrow \quad E[r \mid m] = \frac{1}{n} \sum_{i=1}^{n} m_i
\]
Group-level model comparison

*what if we are colour blind?*

At least, we can measure how likely is the \( i \)th subject’s data under each model!

\[
p(y_1|m_1) \quad p(y_2|m_2) \quad \cdots \quad p(y_i|m_i) \quad p(y_n|m_n)
\]

$$p(r,m|y) \propto p(r) \prod_{i=1}^{n} p(y_i|m_i) p(m_i|r)$$

Our belief about the proportion of models is:

\[
p(r|y) = \sum_{m} p(r,m|y)
\]

Exceedance probability:

\[
\varphi_k = P(r_k > r_{k \neq k} | y)
\]
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 Notes on Bayesian inference
   2.1 Variational methods (ReML, EM, VB)
   2.2 Family inference
   2.3 Group-level model comparison

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
segmentation and normalisation → realignment → normalisation → smoothing → general linear model → statistical inference → p < 0.05

posterior probability maps (PPMs) → multivariate decoding

dynamic causal modelling

Gaussian field theory
aMIRI segmentation
mixture of Gaussians (MoG) model

class variances

$\sigma_1$, $\sigma_2$, $\ldots$, $\sigma_k$

$i^{th}$ voxel label

$y_i$ $\sigma_1$ $\sigma_2$ $\sigma_3$

$i^{th}$ voxel value

$\mu_1$, $\mu_2$, $\ldots$, $\mu_k$

class frequencies

$\lambda$

class means

grey matter

white matter

CSF
Decoding of brain images
recognizing brain states from fMRI

log-evidence of X-Y sparse mappings:
effect of lateralization

log-evidence of X-Y bilateral mappings:
effect of spatial deployment
fMRI time series analysis
spatial priors and model comparison

PPM: regions best explained by short-term memory model

PPM: regions best explained by long-term memory model

fMRI time series

prior variance of GLM coeff
prior variance of data noise
GLM coeff
AR coeff (correlated noise)
fMRI time series

short-term memory design matrix (X)

long-term memory design matrix (X)
Dynamic Causal Modelling

network structure identification

models marginal likelihood

\[ \ln p(y|m) \]

estimated effective synaptic strengths for best model \( m_4 \)
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

\[
\Lambda = \frac{p(y|H_1)}{p(y|H_0)} \geq u
\]

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

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

**ROC analysis**

- **MVB (Bayes factor)**
  \( u=1.09, \text{ power}=56\% \)

- **CCA (F-statistics)**
  \( F=2.20, \text{ power}=20\% \)