Bayesian inference

Justin Chumbley
ETH and UZH

(Thanks to Jean Denizeau for slides)

Overview of the talk

- ✓ Introduction: Bayesian inference
- ✓ Bayesian model comparison
- ✓ Group-level Bayesian model selection

Overview of the talk

- ✓ Introduction: Bayesian inference
- ✓ Bayesian model comparison
- ✓ Group-level Bayesian model selection

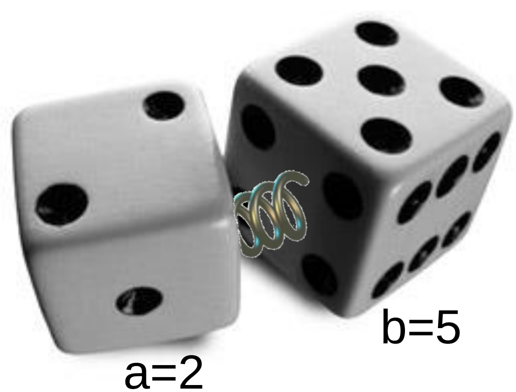
Introduction: Bayesian inference

probability theory: basics

***Degree of plausibility** desiderata*:

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

→ normalization: $$\sum_a P(a) = 1$$



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

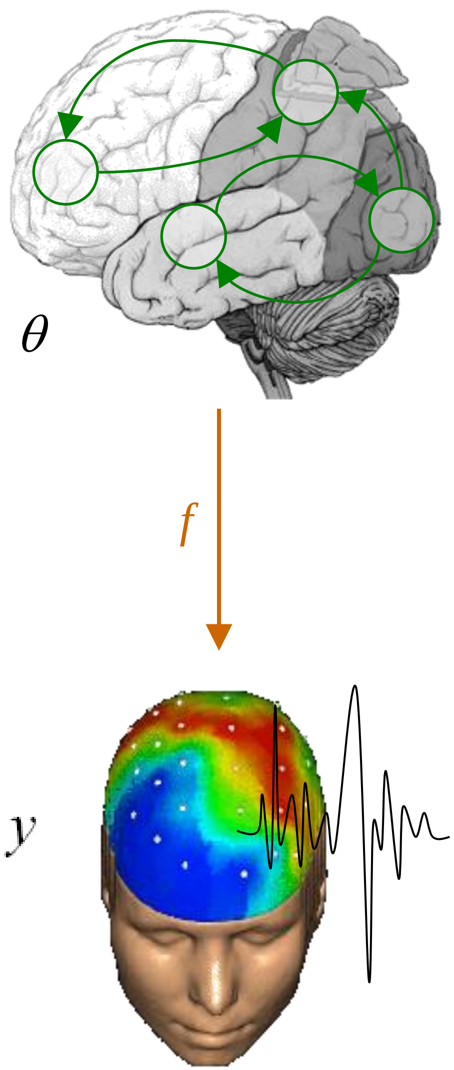
→ conditioning :

(*Bayes rule*)

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

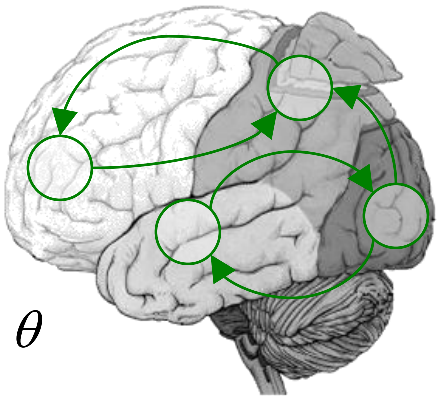
Introduction: Bayesian inference

deriving the likelihood function



Introduction: Bayesian inference

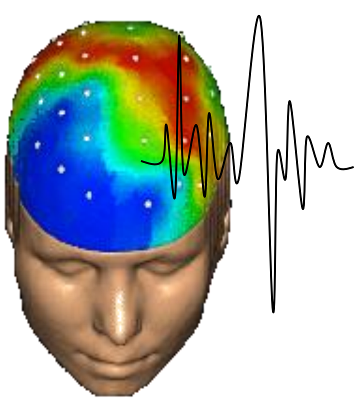
likelihood, priors and Bayes' rule



θ

generative model m

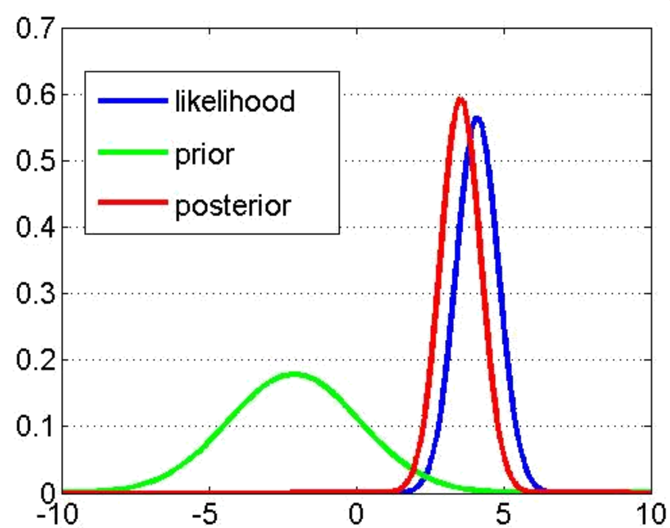
y



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

Prior: $p(\theta|m)$

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



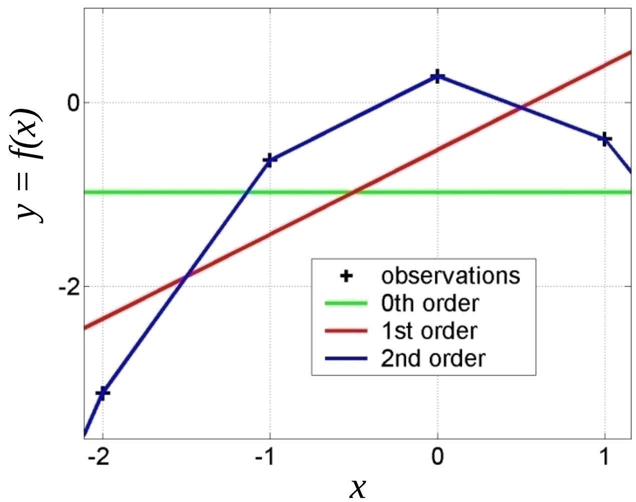
Overview of the talk

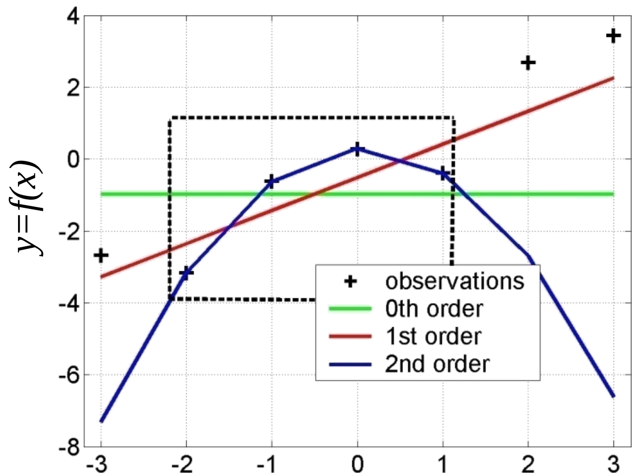
- ✓ Introduction: Bayesian inference
- ✓ Bayesian model comparison
- ✓ Group-level Bayesian model selection

Bayesian model comparison

model evidence

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

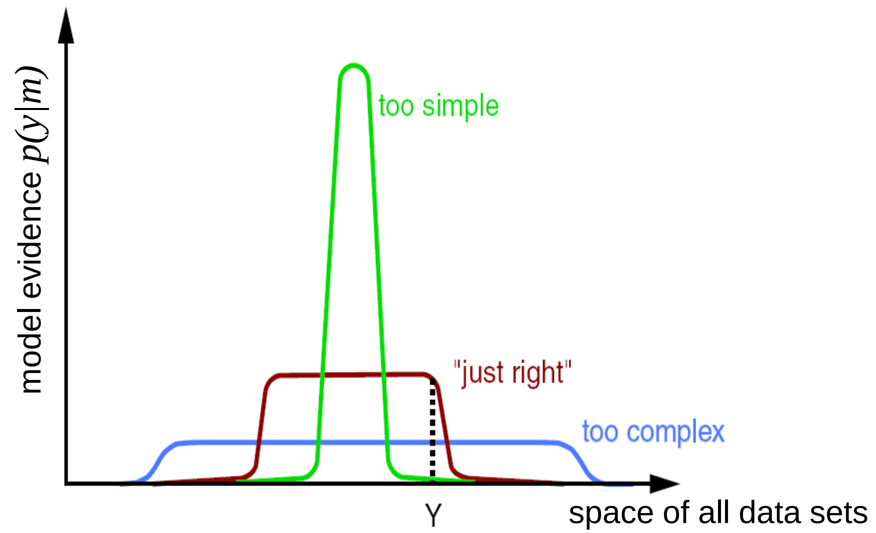




Model evidence:

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

"Occam's razor" :

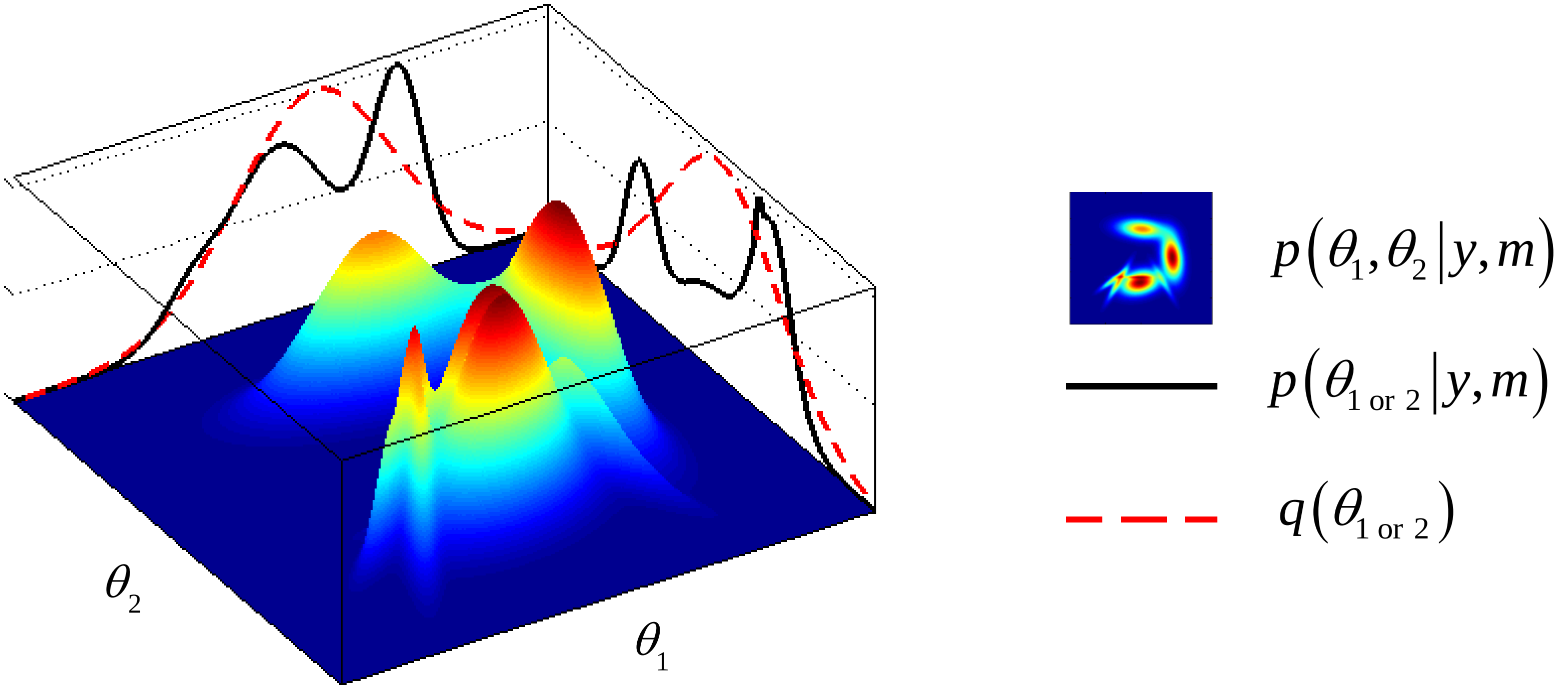


Bayesian model selection

VB and the Free Energy

$$\ln p(y|m) = \underbrace{\langle \ln p(y,\theta|m) \rangle_q + S(q)}_{\text{free energy } F(q)} + \underbrace{D_{KL}\left(p(\theta|y,m);q(\theta)\right)}_{\geq 0}$$

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



Bayesian model selection

Laplace approximation and BIC

→ Laplace approximation

$$q(\theta) \approx N(\mu, \Sigma)$$

$$F \approx \underbrace{\ln p(y|\mu,m) + \ln p(\mu|m) + \frac{p}{2}\ln 2\pi + \frac{1}{2}\ln|\Sigma|}_{F_{\text{Laplace}}}$$

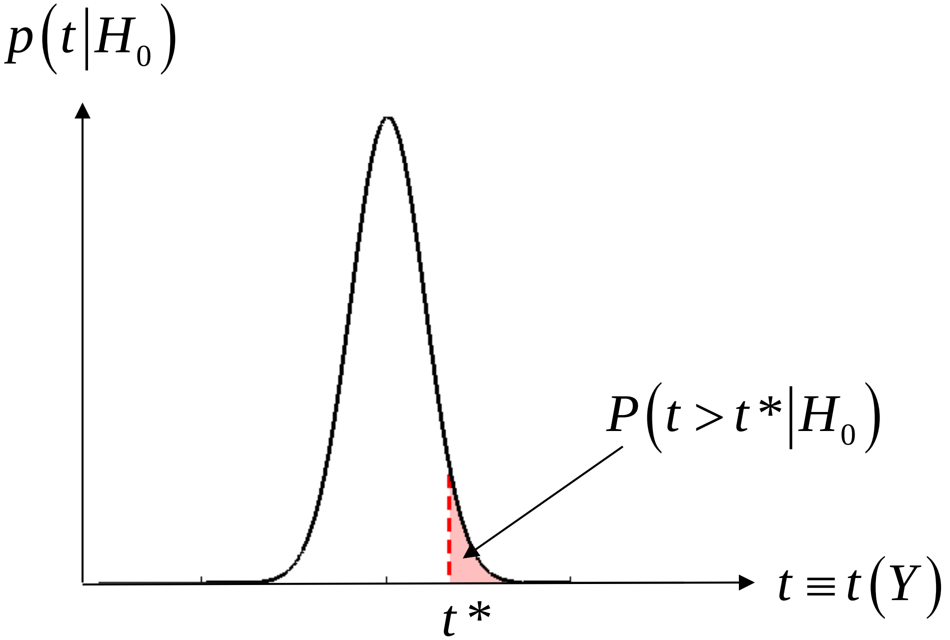
→ BIC: Laplace approximation at the asymptotic limit

$$\Sigma \xrightarrow{n\to\infty} \frac{1}{n} I_p \quad \Rightarrow \quad F_{\text{Laplace}} \xrightarrow{n\to\infty} \underbrace{\ln p(y|\mu,m) - \frac{p}{2}\ln n}_{\text{BIC}}$$

Bayesian model comparison

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.:

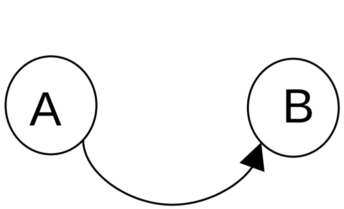
if $P\left(t > t^* \middle| H_0\right) \leq \alpha$ then reject H0

classical (null) hypothesis testing

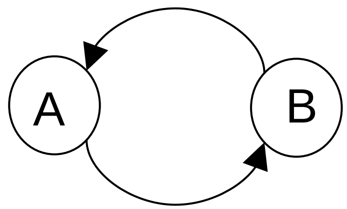
Bayesian model comparison

Family-level inference

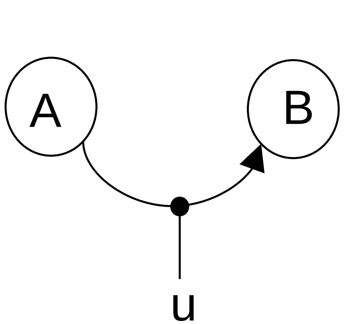
$P(m_1|y) = 0.04$



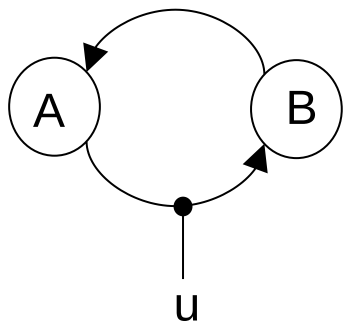
$P(m_2|y) = 0.25$



$P(m_2|y) = 0.01$



$P(m_2|y) = 0.7$



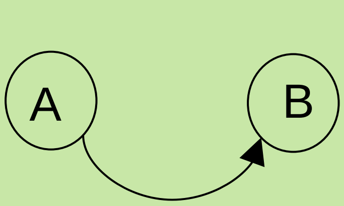
model selection error risk:

$$
\begin{aligned}
P\left(e=1 \middle| y\right) &= 1 - \max_m P\left(m \middle| y\right) \\
&= 0.3
\end{aligned}
$$

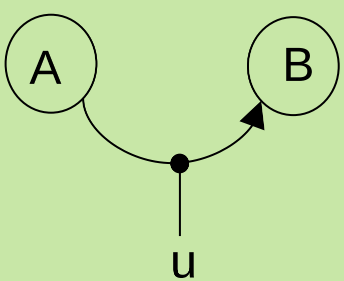
Bayesian model comparison

Family-level inference

$P(m_1|y) = 0.04$

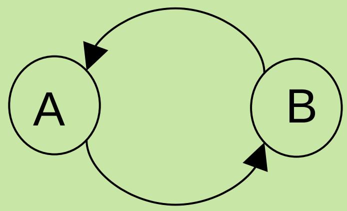


$P(m_2|y) = 0.01$

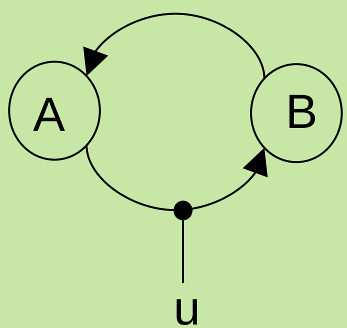


$P(f_1|y) = 0.05$

$P(m_2|y) = 0.25$



$P(m_2|y) = 0.7$



$P(f_2|y) = 0.95$

model selection error risk:

$$P(e=1|y) = 1 - \max_m P(m|y)$$
$$= 0.3$$

family inference
(pool statistical evidence)

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

$$P(e=1|y) = 1 - \max_f P(f|y)$$
$$= 0.05$$

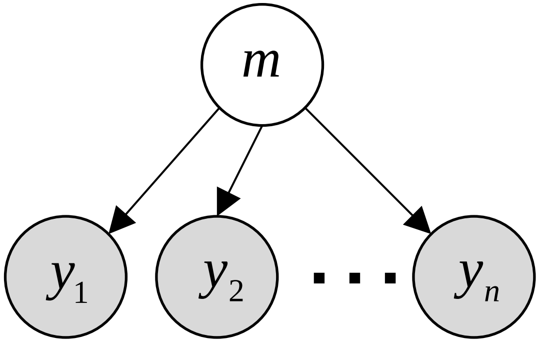
Overview of the talk

- ✓ Introduction: Bayesian inference
- ✓ Bayesian model comparison
- ✓ Group-level Bayesian model selection

Group-level model selection

FFX-BMS analysis

→ *FFX-BMS*: all subjects are best described by a unique (unknown) model



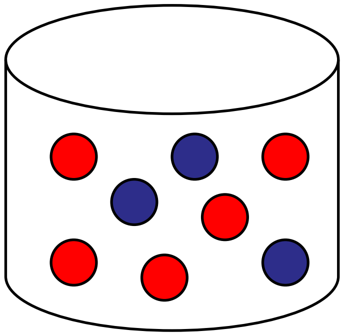
$$p\left(y \mid m_j\right) = \prod_{i=1}^{n} p\left(y_i \mid m_j\right)$$

$$\Rightarrow \ln p\left(m_j \mid y\right) \approx -\ln K + \sum_{i=1}^{n} F_{ij}$$

- ❑ *FFX-BMS* still assumes that model parameters are different across subjects!
- ❑ *FFX-BMS* is not invalid, but main assumption has to be justifiable.
- ❑ What if different subjects are best described by different models? → *RFX-BMS*

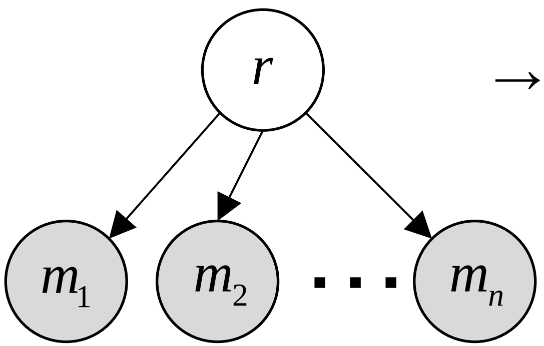
Group-level model selection

RFX-BMS: preliminary (Polya's urn)



$$\begin{cases} m_i = 1 & \rightarrow i^{\text{th}} \text{ marble is blue} \\ m_i = 0 & \rightarrow i^{\text{th}} \text{ marble is purple} \end{cases}$$

r = proportion of blue marbles in the urn



→ (binomial) probability of drawing a set of n marbles:

$$p(m|r) = \prod_{i=1}^{n} r^{m_i} (1-r)^{1-m_i}$$

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

$$p(r|m) \propto p(r) \prod_{i=1}^{n} r^{m_i} (1-r)^{1-m_i} \quad \overset{p(r)\propto 1}{\Rightarrow} \quad E[r|m] = \frac{1}{n} \sum_{i=1}^{n} m_i$$

Group-level model selection

RFX-BMS: the group null

- H1: “reasonable” prior assumption = [the urn is unbiased]

$$E\left[r_k \middle| H_1\right] = 1/K$$

$\Rightarrow$ Exceedance probability: $\varphi_k = P\left(r_k > r_{k'\neq k} \middle| m, H_1\right)$

- H0: “null” prior assumption = [all frequencies are equal]

$$H_0 : r_k = 1/K$$

- Bayesian “omnibus risk”: $P_o = p\left(H_0 \middle| m\right) = \dfrac{p\left(m \middle| H_0\right)}{p\left(m \middle| H_0\right) + p\left(m \middle| H_1\right)}$

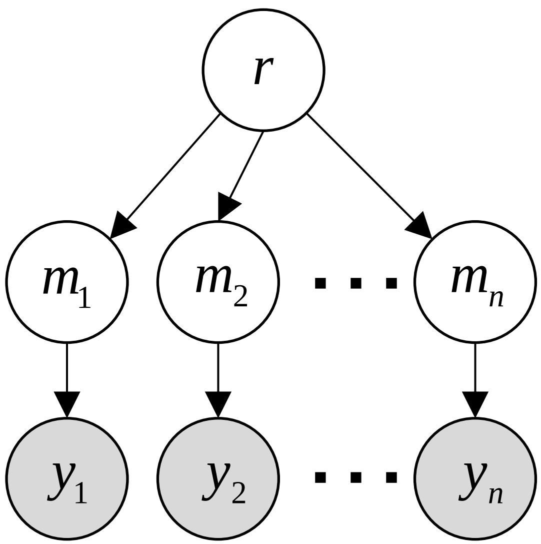
$\Rightarrow$ *Protected* exceedance probability: $\tilde{\varphi}_k = \left(1 - P_0\right)\varphi_k + P_0/K$

Group-level model selection

RFX-BMS: what if we are colour blind?

At least, we can measure how likely is the i^{th} subject's data under each model!

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



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

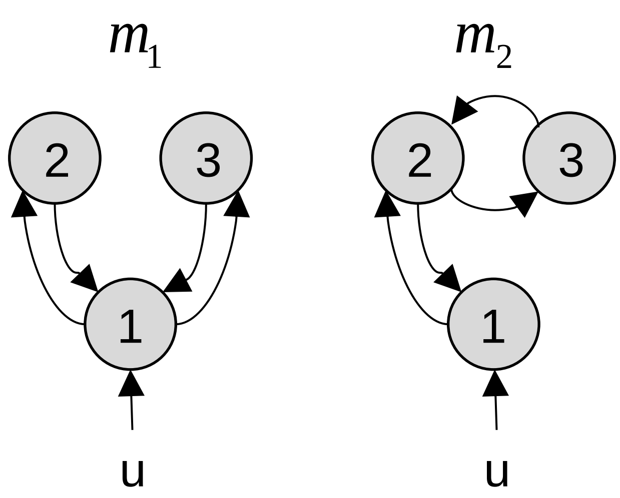
Our belief about the proportion of models is:

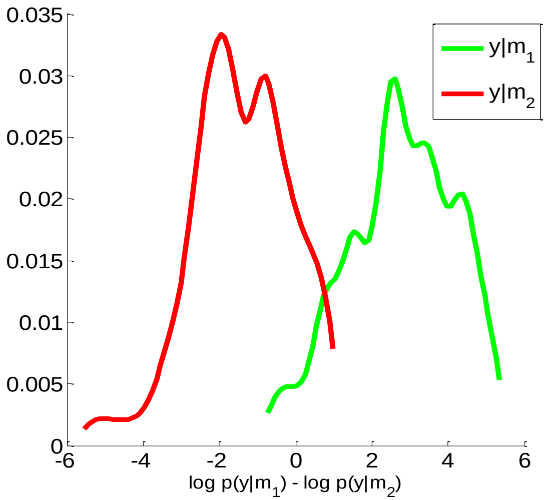
$$p(r|y) = \sum_{m} p(r,m|y)$$

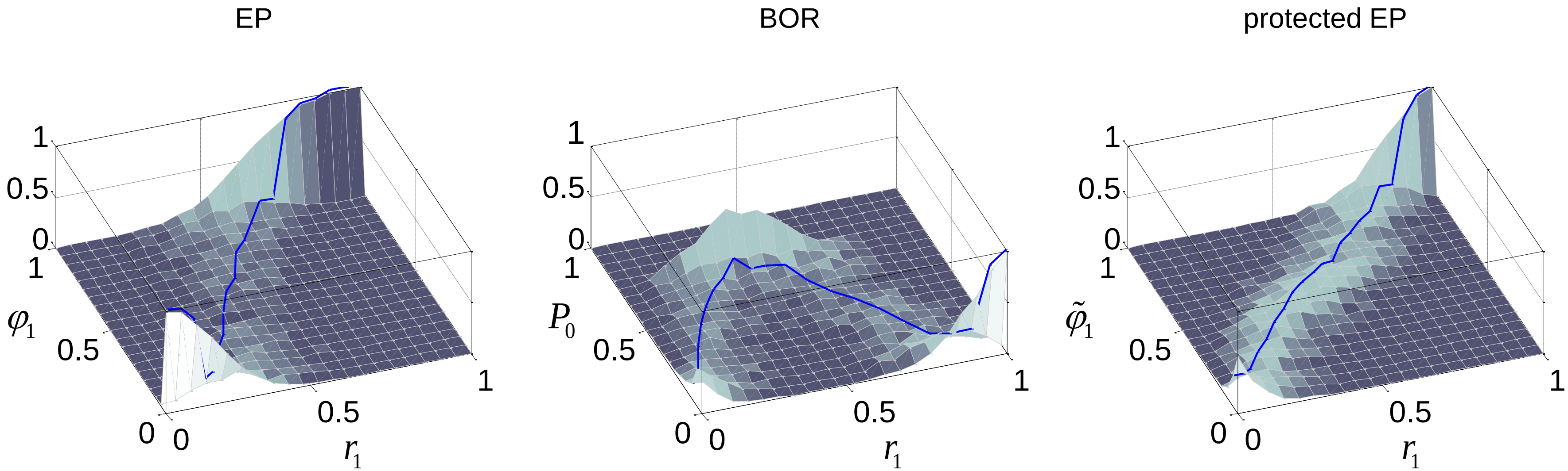
Exceedance probability: $\varphi_k = P\left(r_k > r_{k'\neq k}\,|\,y\right)$

Group-level model selection

RFX-BMS: protecting from DCM overconfidence

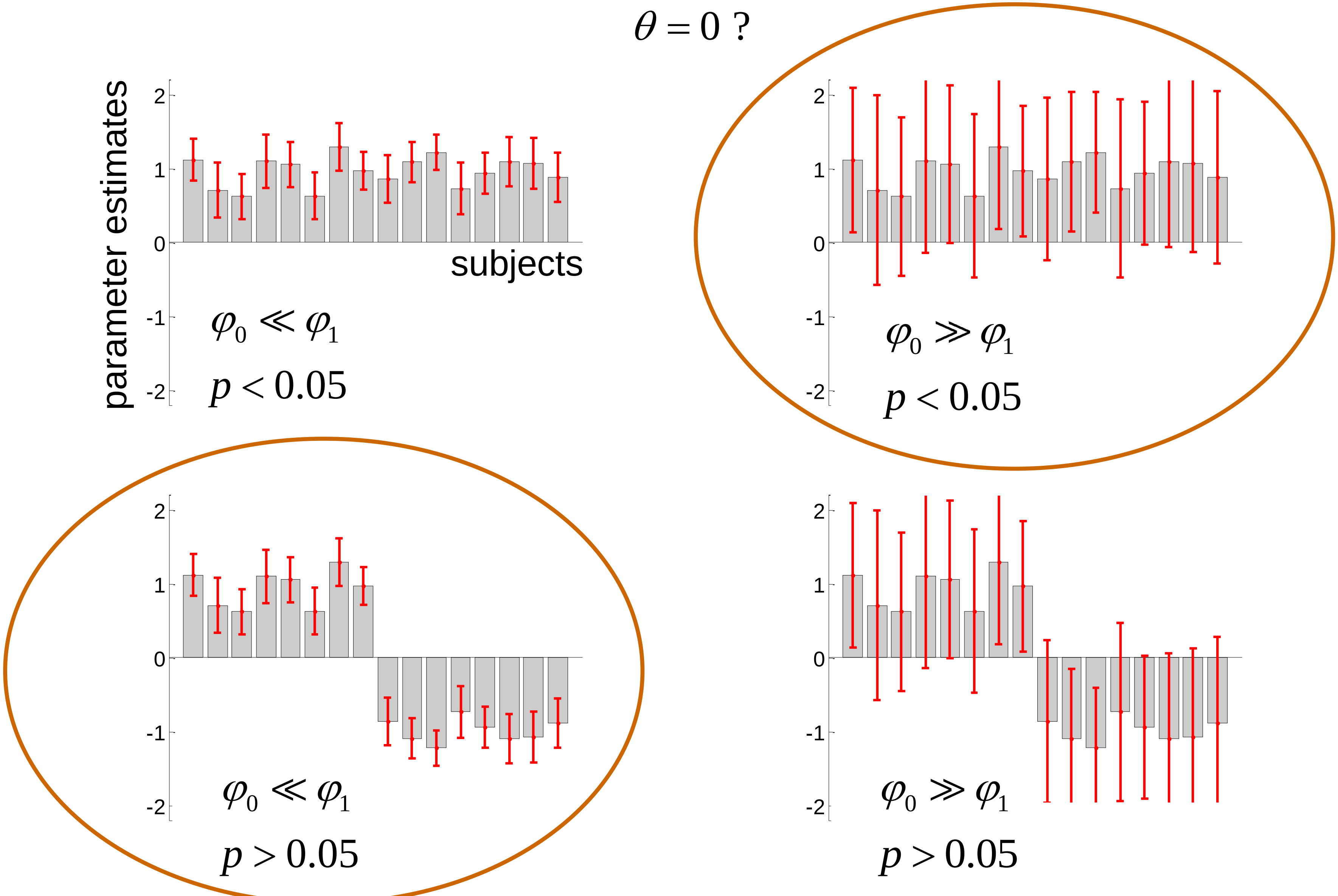






Group-level model selection

frequentist versus Bayesian RFX analyses



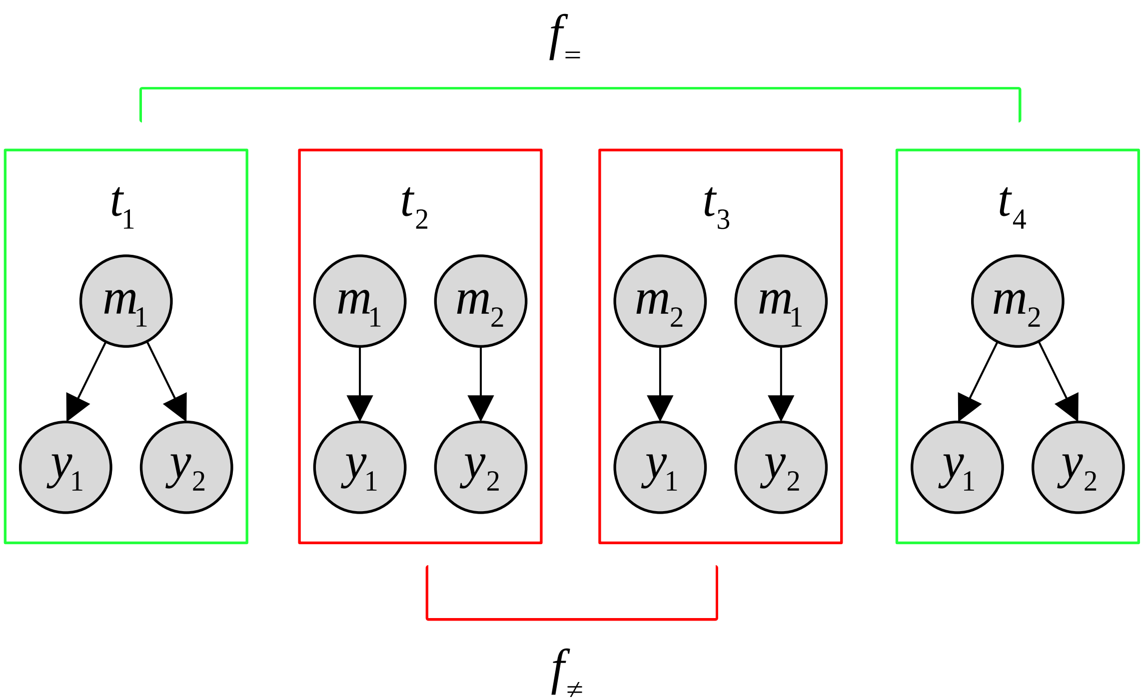
Group-level model selection

RFX-BMS: between-condition comparison

- within-subject design: *n* subjects in 2 conditions

  → statistical evidence for a difference between conditions?

- compare 2 different hypotheses (at the group level):
  - ✓ $f_{=}$: same model across conditions
  - ✓ $f_{\neq}$: different models across conditions



	$y_1 \mid m_1$	$y_1 \mid m_2$
$y_2 \mid m_1$	$y \mid t_1$	$y \mid t_3$
$y_2 \mid m_2$	$y \mid t_2$	$y \mid t_4$

Group-level model selection

RFX-BMS: between-group comparison

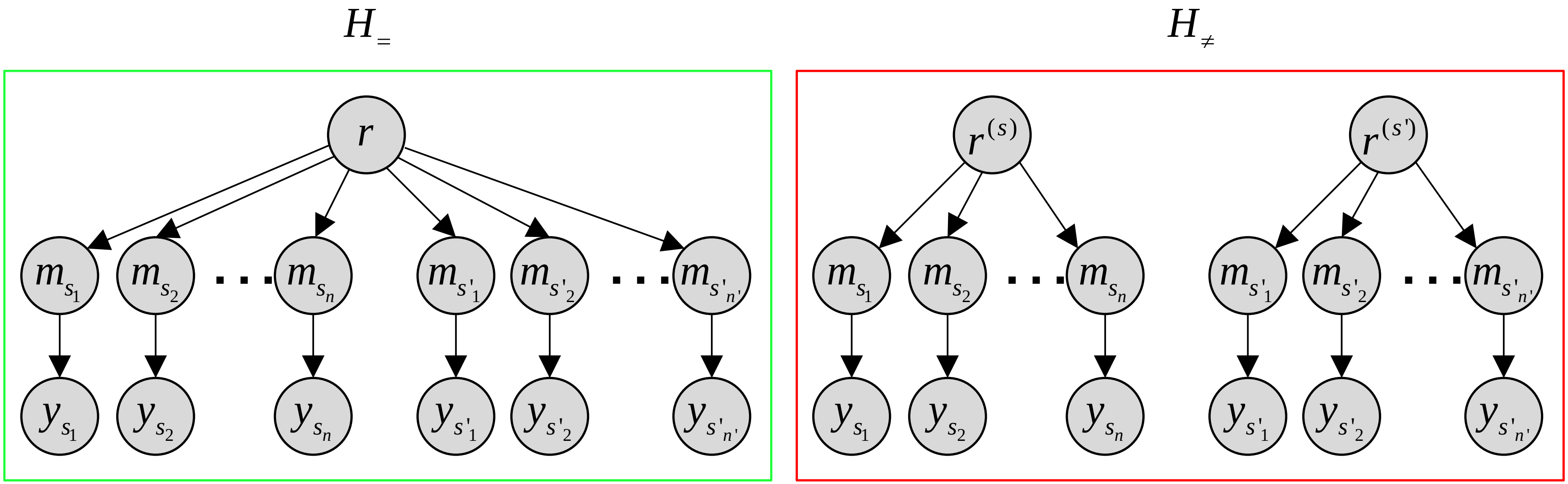
- between-subject design: 2 groups of *n* subjects each

  → statistical evidence for a difference between groups?

- compare 2 different hypotheses (at the group level):

  ✓ $H_=$: different groups come from the same population

  ✓ $H_{\neq}$: different groups come from different populations



Overview of the talk

- ✓ Introduction: Bayesian inference
- ✓ Bayesian model comparison
- ✓ Group-level Bayesian model selection

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%?