Bayesian Analysis

DuBois Bowman, Ph.D.
Gordana Derado, M. S. Shuo Chen, M. S.

Outline

Bayesian Analysis:

I. Introduction to Bayesian Methods

Perspectives

- Parameter of interest: θ
  - The population difference in BOLD activity between two tasks.
- Collect data from a sample
  - Regarded as realizations of *random* variables.

Bayesian
- Determine the distribution of the unknown *random* parameter θ, given the data.

Classical (Frequentist)
- Use the data to estimate the unknown *fixed* parameter θ.

Notation

- Let θ represent the parameter of interest
  - Group (or population) mean BOLD response for a given experimental task or contrast
  - Assume Bayesian interpretation of θ here, unless otherwise noted
- Let ***Y*** represent the data
  - Second level modeling: $\mathbf{Y} = [Y_1, \ldots, Y_n]'$.
    - E.g. contrasts of interest

  Also applicable to first level modeling: $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{iT}]'$.

Bayesian Inference

- Bayesian methods concern one's belief about θ.

[Current Belief (**Posterior**)] $\propto$

(**Prior** Belief) x (**Data**)

Bayes Theorem

$$\Pr(\theta \mid y) = \frac{\Pr(\theta)\Pr(y \mid \theta)}{\Pr(y)}$$

Posterior : $\Pr(\theta \mid y)$ **Prior :** $\Pr(\theta)$

Marginal : $\Pr(y)$ **Data :** $\Pr(y \mid \theta)$

Bayes Theorem

Case I: Assume θ has m possible states:

$$\Pr(y) = \sum_{i=1}^{m} \Pr(\theta_i)\Pr(y \mid \theta_i)$$

Case II: Assume θ is continuous:

$$\Pr(y) = \int_{\theta} \Pr(\theta)\Pr(y \mid \theta)d\theta$$

Bayes Theorem

In either case,

$$\Pr(\theta \mid y) = \frac{\Pr(\theta)\Pr(y \mid \theta)}{\cancel{\Pr(y)}}$$

Denominator is constant with respect to θ.

Bayes Theorem

$$\Pr(\theta \mid y) \propto \Pr(\theta)\Pr(y \mid \theta)$$

Posterior : $\Pr(\theta \mid y) \rightarrow \pi(\theta \mid y)$

Prior : $\Pr(\theta) \rightarrow \pi(\theta)$

Data : $\Pr(y \mid \theta) \rightarrow p(y \mid \theta)$

Bayesian Inference

Posterior $\propto$ **Prior** x **Data**

Example: Normal-Normal Model

- Let: $\mathbf{Y} = [Y_1, \ldots, Y_n]'$
- Data: $Y_i \mid \theta \sim Normal(\theta, \sigma^2), \quad i = 1, \ldots, n$
- Prior: $\theta \sim Normal(\theta_0, \phi_0)$
- Posterior: $\theta \mid \mathbf{Y} \sim Normal(\theta_1, \phi_1)$

$$\phi_1^{-1} = \phi_0^{-1} + (\sigma^2/n)^{-1} \quad \text{and} \quad \theta_1 = \theta_0\left(\frac{\phi_1}{\phi_0}\right) + \bar{Y}\left(\frac{\phi_1}{\sigma^2/n}\right)$$

Example: Normal-Normal Model

- **Posterior**: $\theta \mid \mathbf{Y} \sim Normal(\theta_1, \phi_1)$

$$\phi_1 = \left[\phi_0^{-1} + \left(\sigma^2/n\right)^{-1}\right]^{-1} \quad \text{and} \quad \theta_1 = \theta_0\left(\frac{\phi_1}{\phi_0}\right) + \bar{Y}\left(\frac{\phi_1}{\sigma^2/n}\right)$$

- Consider as a function of **prior uncertainty**:

$$\theta \mid \mathbf{Y} \to Normal\left(\bar{Y}, \frac{\sigma^2}{n}\right) \quad \text{as} \quad \phi_0 \to \infty$$

  – Data will dominate when ϕ_0 is large compared to σ^2/n

Prior Probabilities

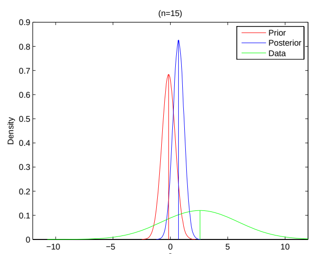
- **Noninformative Prior:**
  – Conveys little knowledge about parameters
  – Also called diffuse or vague prior
- **Informative Prior :**
  – Precise knowledge about parameters
- **Conjugate Prior:**
  – Posterior distribution has the same parametric form as the prior
  – Convenient mathematical form and often justifiable scientifically

Influence of Prior Beliefs

- Normal-Normal Model: $\theta \mid \mathbf{Y} \sim Normal(\theta_1, \phi_1)$

$\theta \sim Normal(\theta_0, \phi_0)$,
$\phi_0 = V(Y_v)$

- Empirically-based prior

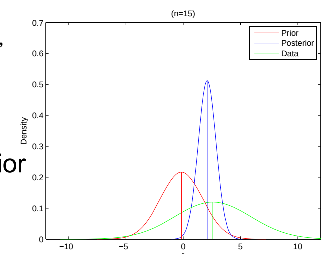


Influence of Prior Beliefs

- Normal-Normal Model: $\theta \mid \mathbf{Y} \sim Normal(\theta_1, \phi_1)$

$\theta \sim Normal(\theta_0, \phi_0)$,
$\phi_0 = 10 \times V(Y_v)$

- More weakly informative prior

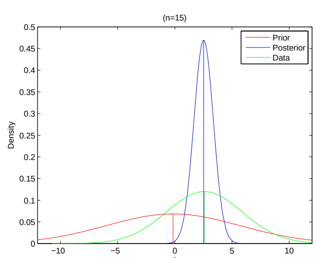


Influence of Prior Beliefs

- Normal-Normal Model: $\theta \mid \mathbf{Y} \sim Normal(\theta_1, \phi_1)$

$\theta \sim Normal(\theta_0, \phi_0)$,
$\phi_0 = 100 \times V(Y_v)$

- Diffuse or Vague prior



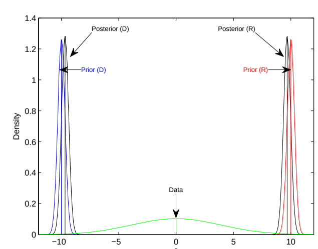
Caution about Prior Beliefs!!

- Normal-Normal Model: $\theta \mid \mathbf{Y} \sim Normal(\theta_1, \phi_1)$

$\theta \sim Normal(\theta_0, \phi_0)$

– U.S. Political Debate

– Biased Jurors

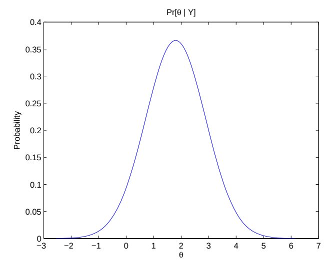


Bayesian Inference

- After analysis, all information about a parameter θ is contained in the **posterior**
- Useful summaries
  - Posterior mean/median/mode
  - Posterior exceedance probabilities
  - Highest (posterior) density regions (HDR)
    - Credible intervals
    - Bayesian confidence intervals
  - Bayes factors (*not discussed here)

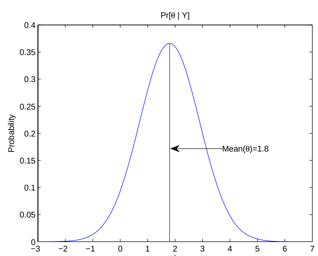
Bayesian Inference

Posterior distribution



Bayesian Inference

Posterior mean

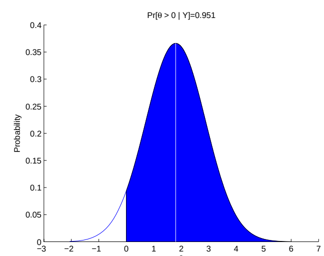


Bayesian Inference

Posterior Probability Maps (PPM)

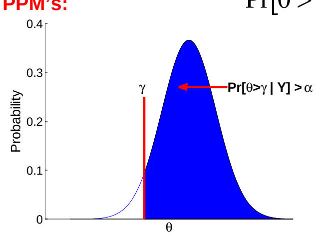
- Posterior (exceedance) probability:

$\Pr[\theta > \gamma | Y]$,
E.g. $\gamma = 0$



Posterior Probability Maps

Threshold PPM's: $\Pr[\theta > \gamma | \mathbf{Y}] > \alpha$



Two thresholds:
- probability α: (e.g. 90%)
- activation threshold γ: percentage of global mean signal (physiologically relevant size of effect)

Posterior Sampling

- Joint posterior follows a known distribution, leading to direct sampling
- Markov Chain Monte Carlo (MCMC)
  - Gibbs sampler (known parametric forms for conditional posterior distributions, which collectively yield samples from the joint posterior distribution)
  - Metropolis-Hastings algorithm
- Variational Bayes Approximation
- Laplace Approximations

Bayesian Learning

- Natural framework for *learning* and *updating*
- Study 1: $\pi_1(\theta \mid y_1) \propto \pi(\theta) \times p(y_1 \mid \theta)$
- Study 2: $\pi_2(\theta \mid y_2) \propto \pi_1(\theta \mid y_1) \times p(y_2 \mid \theta)$

Bayesian Analysis:

II. Implementation in fMRI Studies

Two-stage modeling

- Use conventional two-stage modeling, emulating a random effects analysis
  - Stage 1: Subject-specific analyses producing estimates of task-related activity (or contrasts)
  - Stage 2: Group-level estimates of task-related activity (or contrasts), using stage 1 estimates (contrasts) of regression coefficients as data
- Choice of Bayesian versus classical (frequentist) modeling at each of the two stages

Estimation in SPM

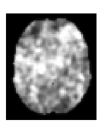
		Stage 2: Classical	Stage 2: Bayesian
Stage 1	Classical	1	2, 4
	Bayesian		3

4. Classical/Spatial Bayesian hierarchical model (BHM) [*Not in SPM]

Posterior Probability Maps

For fMRI applications,

- It is useful to compute PPM's
  - Each voxel contains an exceedance probability, e.g. that the difference between two tasks is greater than zero.



  - Can display thresholded PPM's, e.g. with p>0.80.

Spatial Bayesian Hierarchical Model

- Use anatomical (or functional) parcellations (e.g. Brodmann)
  - # regions<# subjects per group
- Models spatial correlations in activity within and between regions
- Inferences:
  - Voxel-level
  - Regional
  - Task-related functional connectivity

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH — CBIS Center for Biomedical Imaging Statistics

Bayesian Analysis:

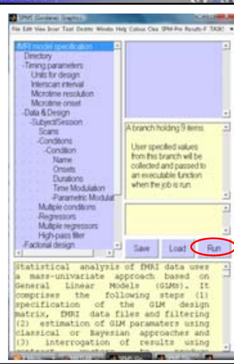
III. Data Example and Analyses

Data Example

Inhibitory Control in Cocaine Addicts

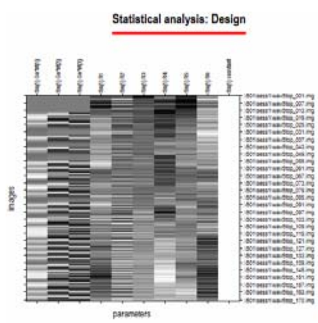
- N=27 subjects: 12 patients (cocaine addicts) and 15 healthy controls (matched on several variables)
- Two sessions: (170 scans per subject in each session)
  - Cocaine addicts: Pre and post-treatment (behavioral therapy)
  - Controls: Baseline and follow-up
- **fMRI Tasks: Stop signal task (Inhibitory control)**
  - Subjects were presented with visual cues and directed to respond to a "GO stimulus" (an uppercase alphabetical letter) by pressing a button as quickly as possible
  - A "STOP signal" – a brief auditory tone lasting 0.5 seconds – was presented randomly in 16% of the trials, and subjects were instructed to refrain from pushing the button if the GO stimulus was followed by a STOP signal
  - A successful performance required prepotent behaviors to be inhibited.
- **Objective:** (Post-treatment – Pre-treatment activity in addicts) > (Follow-up – baseline activity in controls)

SPM: Stage 1 Model Specification



Same for classical and Bayesian

SPM: Stage 1 Design Matrix

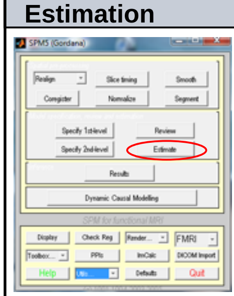


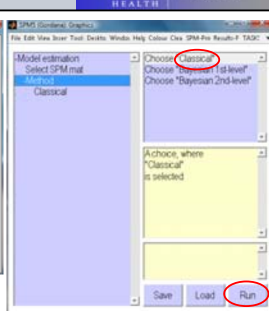
Design description:

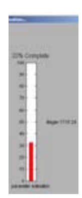
- Basis function: HRF (with time and dispersion derivatives)
- Inter-scan Interval: 2.02
- Grand mean scaling: session specific
- Global normalization: scaling
- Note: Full design matrix included three additional conditions

Same for classical and Bayesian

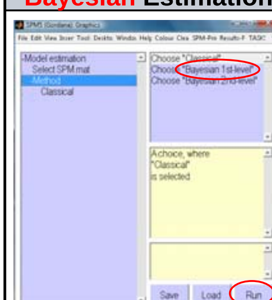
SPM: Stage 1 Classical Estimation

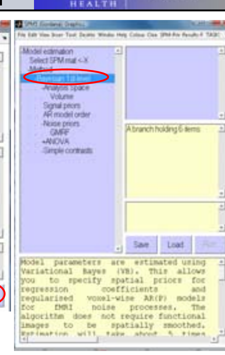


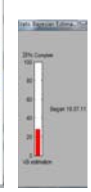




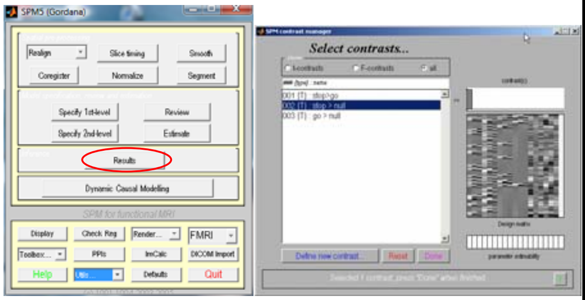
SPM: Stage 1 Bayesian Estimation





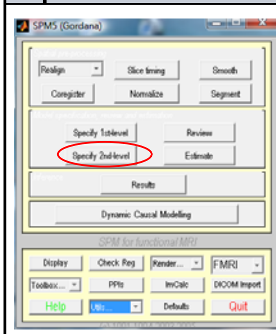


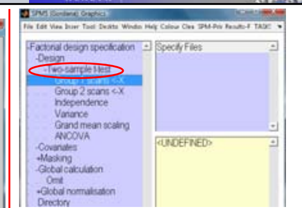
SPM: Stage 1 Results

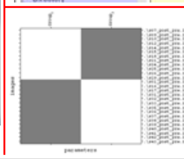


Similar for **classical** and **Bayesian**

SPM: Stage 2 Model Specification





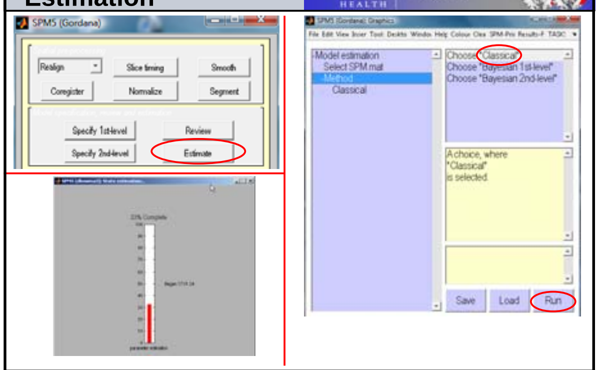


Design:
Conditions: 2

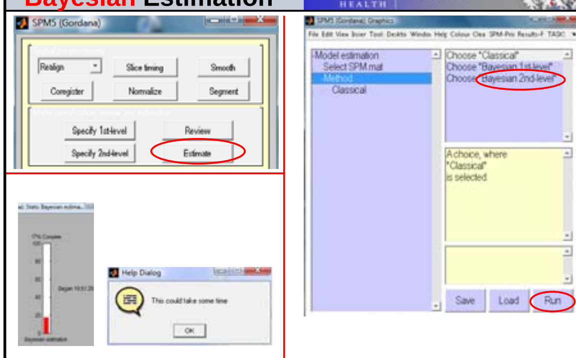
25 d.f.
[27 images – 2 parameters]

Same for **classical** and **Bayesian**

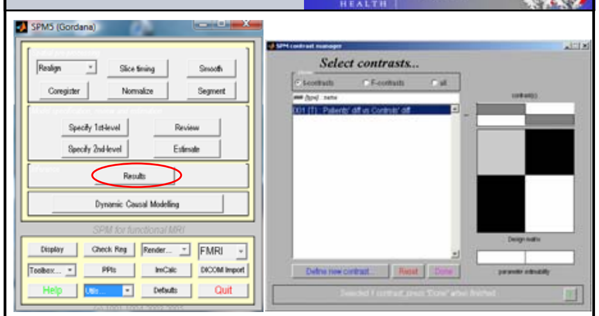
SPM: Stage 2 **Classical** Estimation



SPM: Stage 2 **Bayesian** Estimation

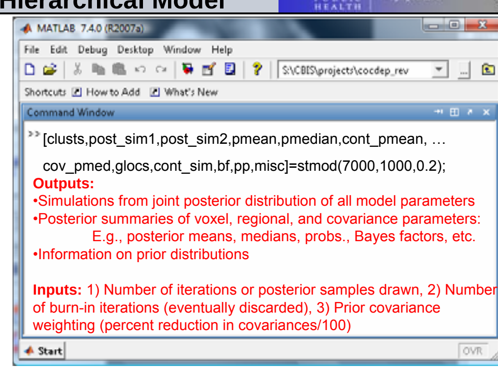


SPM: Stage 2 Results

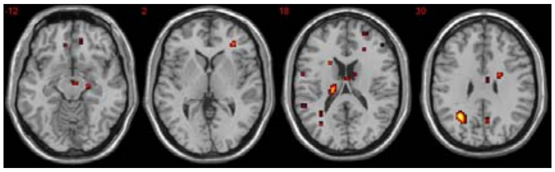


Similar for **classical** and **Bayesian**

Spatial Bayesian Hierarchical Model



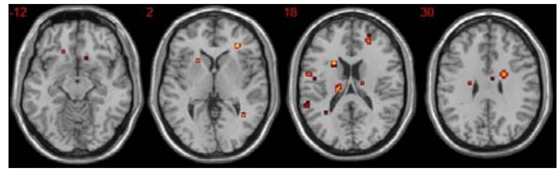
Results: SPM



1. Classical/Classical

- Voxel-level t-statistic maps
- p=0.005 (uncorrected)

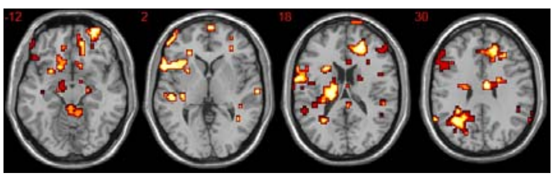
Results: SPM



2. Classical/Bayesian

- Voxel-level PPM's
- Activation threshold (γ)=0
- α=0.85

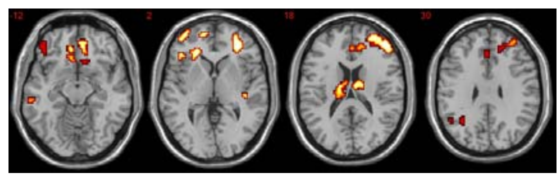
Results: SPM



3. Bayesian/Bayesian

- Voxel-level PPM
- Activation threshold (γ)=0
- α=0.85

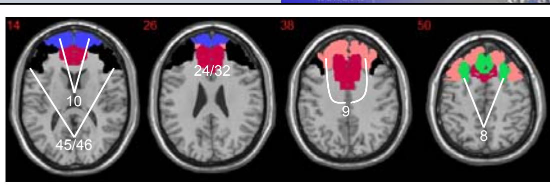
Results: Spatial BHM



4. Classical/Spatial BHM

- Voxel-level PPM
- Likelihood of voxel-level activations
- α=0.85

Results: Spatial BHM



4. Classical/Spatial BHM

- Regional-level PPM
- Likelihood of regional-level activations
- α=0.50

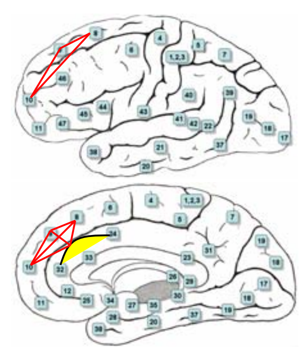
BA 10:	p=0.50 (Blue)
BA 45/46:	p=0.52 (Black)
BA 8:	p=0.51 (Green)
BA 9:	p=0.55 (Pink)
BA 24/32:	p=0.53 (Red)

Results: Spatial BHM

Functional Connectivity

- Inter-regional task-related functional connections in cocaine addicts following treatment
- FC>0.3

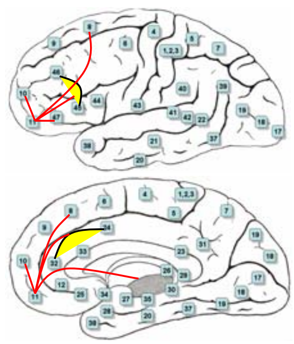
- BA 24/32: Anterior cingulate



Results: Spatial BHM

Functional Connectivity

- Posterior probs. of FC **changes** (or differences)
  - Δ_s=Group difference in connectivity during session *s*
  - $\Pr(\Delta_{pre} > \Delta_{post}) > 0.9$
- BA 24/32: Anterior cingulate



Bayesian Analysis:

IV. Summary

Summary: Strengths

Bayesian			Classical
SPM 2: C/B	**SPM 3: B/B**	**4. C/SBHM**	**SPM 1: C/C**
PPM's to quantify evidence (also Bayes Factors)			Conventional testing methodology
Flexible inferences from joint posterior distribution			
May circumvent (mitigate) multiple testing issues			
Fast computations		Fast computations	Fast computations
SPM	SPM		SPM
		Regional inferences	
		FC inferences	
	Local *spatial corrs.*	Intra/inter-regional *spatial corrs.*	
	No smoothing in pre-process.	No or minimal smoothing	

Summary: Weaknesses

Bayesian			Classical
SPM 2: C/B	**SPM 3: B/B**	**4. C/SBHM**	**SPM 1: C/C**
Must specify prior distributions (somewhat subjectively)			
	Long computations		
		Personal implementation	
		Small sample constraint on no. of regions	
Unconventional thresholding methods	Unconventional thresholding methods		

References

1. Bowman, F. D., Caffo, B. A, Bassett, S., and Kilts, C. (2008). Bayesian Hierarchical Framework for Spatial Modeling of fMRI Data. *NeuroImage* 39, 146–156.
2. Guillaume Flandin and William D. Penny (2007): Bayesian fMRI data analysis with sparse spatial basis function priors, *NeuroImage* 34(3), 1108-1125.
3. Friston K.J., Glaser D.E., Henson R.N.A., Kiebel S., Phillips C., Ashburner J. (2002). Classical and Bayesian Inference in Neuroimaging: applications. NeuroImage 16(2): 484-512.
4. Friston K, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. (2002). Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16(2): 465-483
5. Penny,W., Kiebel,S., and Friston,K. (2003). Variational Bayesian inference for fMRI time series. *NeuroImage 19*, 727-741.
6. Penny,W., Trujillo-Barreto,N.J., and Friston,K. (2005) Bayesian fMRI time series analysis with spatial priors. *NeuroImage* 24(2), 350-362.

Website: http://www.sph.emory.edu/bios/CBIS