Bayesian hypothesis testing for hierarchical models using transdimensional Markov chain Monte Carlo methods

Tom Lodewyckx (University of Amsterdam)
Michael Lee (University of California, Irvine)
Eric-Jan Wagenmakers (University of Amsterdam)

July 28, 2008

OUTLINE

OUTLINE

Heads or tails?

Question

When tossing a coin frequently, the true probabilities of heads and tails are equal

→ **Is the coin truly fair?**

Hypothesis testing

N: Number of tosses
K: Frequency of tails

Hypothesis testing

N: Number of tosses
K: Frequency of tails

Null model (M_0)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta = .5$
- Equal probabilities

Full model (M_1)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta \neq .5$
- Heads or tails more probable

Hypothesis testing

N: Number of tosses
K: Frequency of tails

Null model (M_0)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta = .5$
- Equal probabilities

Full model (M_1)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta \neq .5$
- Heads or tails more probable

$\rightarrow$ **Which is the most plausible assumption for θ?**
$\rightarrow$ **Bayes factor**

The Bayes factor

What is a Bayes factor?

- Model selection tool in Bayesian framework
- Compares the "evidences" of both models
- Model with highest evidence is supported
- Quantification of how strong that support is

The Bayes factor

What is a Bayes factor?

- Model selection tool in Bayesian framework
- Compares the "evidences" of both models
- Model with highest evidence is supported
- Quantification of how strong that support is

Notation

B_{10} is the Bayes factor in favor of M_1 (full model)

The Bayes factor

Formal definition

$$B_{10} = \frac{\text{Marginal LL }(M_1)}{\text{Marginal LL }(M_0)} = \frac{f(y \mid M_1)}{f(y \mid M_0)}$$

The Bayes factor

Formal definition

$$B_{10} = \frac{\text{Marginal LL } (M_1)}{\text{Marginal LL } (M_0)} = \frac{f(y \mid M_1)}{f(y \mid M_0)}$$

$$= \frac{\text{Posterior model odds}}{\text{Prior model odds}} = \frac{P(M_1 \mid y)/P(M_0 \mid y)}{P(M_1)/P(M_0)}$$

The Bayes factor

Interpretation scheme Raftery (1995)

$\log(B_{10})$	Evidence?
$<$ -5	Very strong evidence for M_0
-5 to -3	Strong evidence for M_0
-3 to -1	Positive evidence for M_0
-1 to 0	Weak evidence for M_0
0	No evidence
0 to 1	Weak evidence for M_1
1 to 3	Positive evidence for M_1
3 to 5	Strong evidence for M_1
$>$ 5	Very strong evidence for M_1

The Bayes factor

Advantages

- *Intuitive*
- *Model averaging*
- *Model complexity*

The Bayes factor

Advantages

- *Intuitive*
- *Model averaging*
- *Model complexity*

Problems

- *Depends on prior distribution*
- *Computational*

Research goal

Research goal: Estimating Bayes factors should be..

1. *easy to implement*
2. *precise*
3. *flexible*

OUTLINE

Transdimensional MCMC methods

Markov chain Monte Carlo (MCMC) methods

- *What?* Simulation techniques to simulate values from posterior distribution
- *Why?* Facilitate Bayesian parameter estimation
- *Where?* Parameter space $\Omega = [\Theta] = [\{\alpha, \beta, \gamma, \ldots\}]$

Transdimensional MCMC methods

Markov chain Monte Carlo (MCMC) methods

- *What?* Simulation techniques to simulate values from posterior distribution
- *Why?* Facilitate Bayesian parameter estimation
- *Where?* Parameter space $\Omega = [\Theta] = [\{\alpha, \beta, \gamma, \ldots\}]$

Transdimensional MCMC methods

- *What?* MCMC methods that operates on at least 2 models
- *Why?* Simultaneous estimation of Bayesian models, hypothesis testing, model selection
- *Where?* Parameter space $\Omega = [M, \Theta_A, \Theta_B, \Theta_C, \Theta_D, \ldots]$

Transdimensional MCMC methods

Transdimensional MCMC methods of interest

- *What?* MCMC methods that operates on M_0 and M_1
- *Why?* hypothesis testing
- *Where?* Parameter space $\Omega = [M, \Theta_0, \Theta_1]$

Hypothesis testing

Specify prior distribution: $\Omega = [M, \Theta_0, \Theta_1]$

Hypothesis testing

Specify prior distribution: $\Omega = [M, \Theta_0, \Theta_1]$

Transdimensional MCMC sampling: Simulate values from posterior distribution of M

- $M = 0 \rightarrow$ Simulate values posterior Θ_0
- $M = 1 \rightarrow$ Simulate values posterior Θ_1

Hypothesis testing

Specify prior distribution: $\Omega = [M, \Theta_0, \Theta_1]$

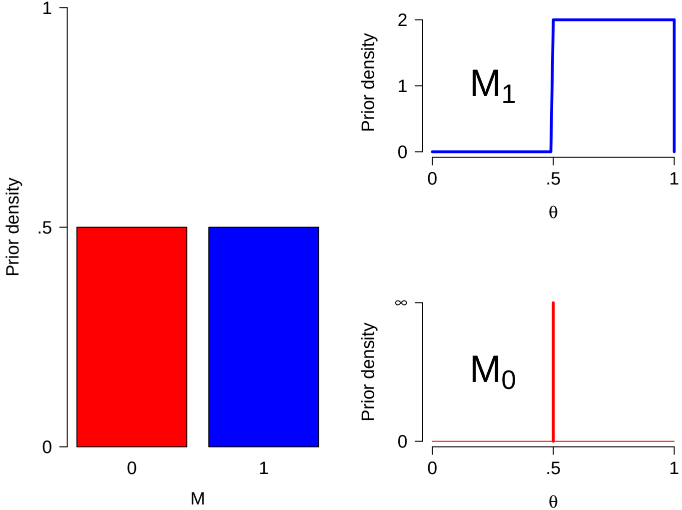
Transdimensional MCMC sampling: Simulate values from posterior distribution of M

- $M = 0 \rightarrow$ Simulate values posterior Θ_0
- $M = 1 \rightarrow$ Simulate values posterior Θ_1

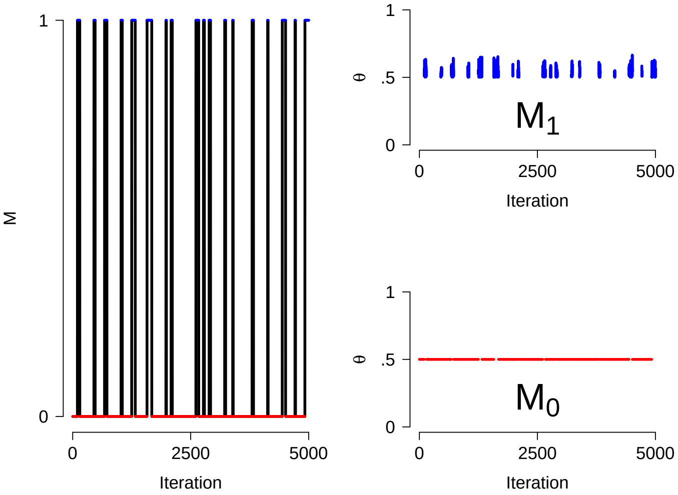
Estimate Bayes factor: Use prior and posterior chances of M

$$B_{10} = \frac{P(M = 1 \mid y)/P(M = 0 \mid y)}{P(M = 1)/P(M = 0)}$$

Hypothesis testing: Prior distribution

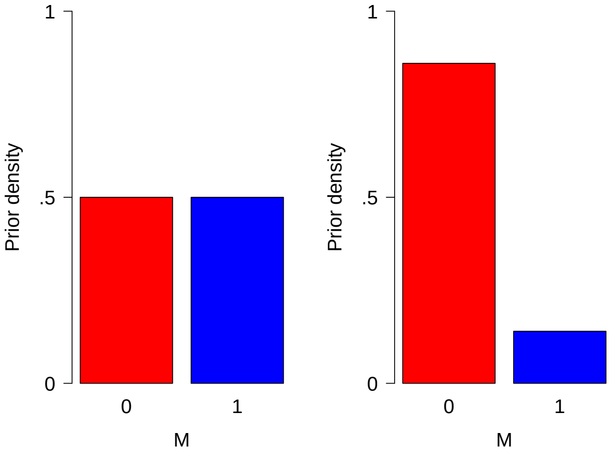


Hypothesis testing: Transdimensional MCMC



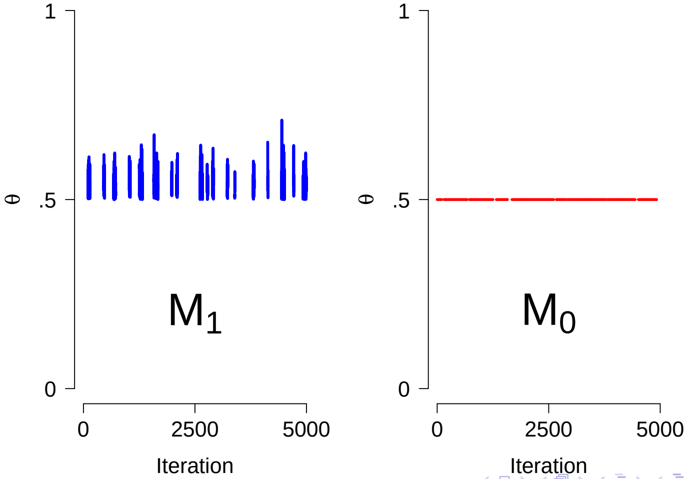
Hypothesis testing: Bayes factor

$\log(BF_{10}) = -1.81$



Hypothesis testing

Problem: No continual sampling of the parameter vectors

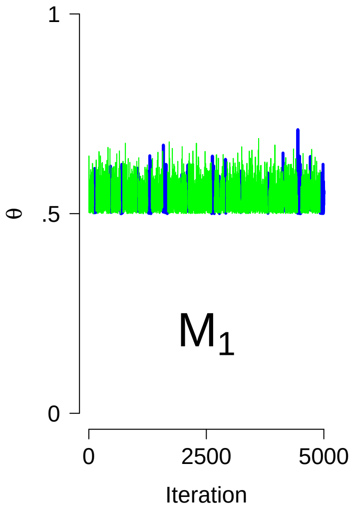


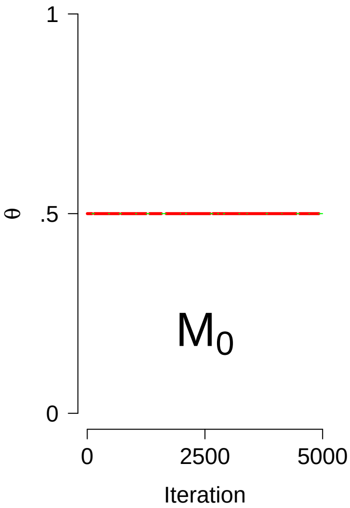
C^3 method

Combined Carlin & Chib (C^3) method

- Pseudopriors are used for sampling from the parameter vector when the model is deactivated
- Recommended choice: posterior distribution
- Combination of three sampling paths:
  1. $\Omega = [M, \Theta_0, \Theta_1]$
  2. $\Omega = [\Theta_0] \rightarrow$ Pseudoprior Θ_0
  3. $\Omega = [\Theta_1] \rightarrow$ Pseudoprior Θ_1

C^3 method





OUTLINE

1. Bayesian hypothesis testing
2. Hypothesis testing with transdimensional MCMC
3. **Applications**
4. Conclusion

Is priming truly subliminal?

Question

Subliminal priming studies assume that the prime stimulus is perceived on a subliminal level

→ **Assumption plausible?**

Is priming truly subliminal?

Study by Rouder, Morey, Speckman & Pratte (2007)

- Visual stimuli $[2, 3, 4, 6, 7, 8]$
- In each trial, participant was presented a 22 ms prime stimulus, followed by a 200 ms target stimulus
- Indicate whether prime stimulus was higher than 5 ("Yes" / "No")
- Results in K correct identifications out of N trials

Is priming truly subliminal?

Null model (M_0)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta = .5$ (at chance)
- Subliminal perception of prime stimulus

Is priming truly subliminal?

Null model (M_0)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta = .5$ (at chance)
- Subliminal perception of prime stimulus

Full model (M_1)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta > .5$ (above chance)
- Supraliminal perception of prime stimulus

Is priming truly subliminal?

Null model (M_0)

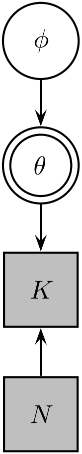
- $K \sim \text{Binomial}(\theta, N)$ and $\theta = .5$ (at chance)
- Subliminal perception of prime stimulus

Full model (M_1)

- $K \sim \text{Binomial}(\theta, N)$ and $\theta > .5$ (above chance)
- Supraliminal perception of prime stimulus

→ Estimate log Bayes factor with Combined Carlin & Chib method for each subject (non-hierarchical) and for the group (hierarchical)

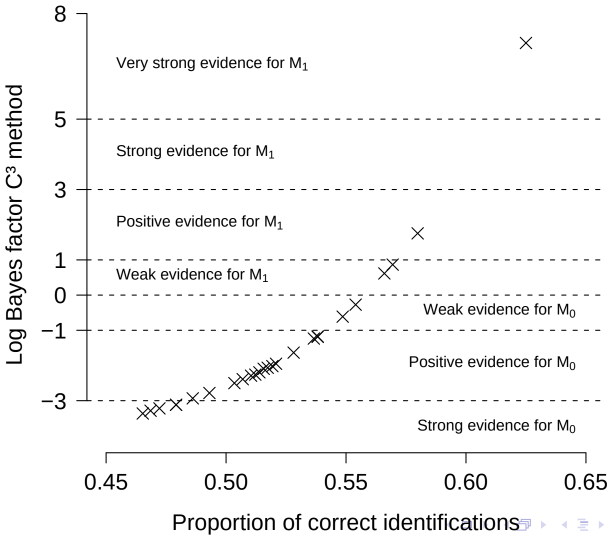
Non-hierarchical application: Graphical model



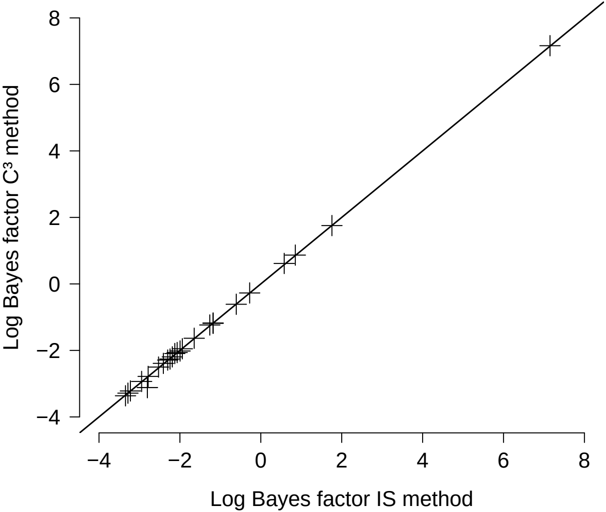
$$K \sim \text{Binomial}(\theta, N)$$
$$\theta = \Phi(\phi)$$

$$M_0 : \phi = 0$$
$$M_1 : \phi \sim \text{Normal}_{(0,+\infty)}(0, 1)$$

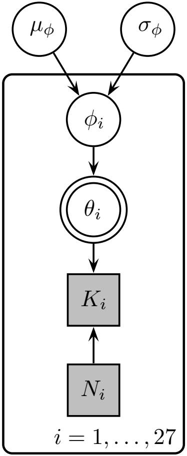
Non-hierarchical application: Results



Non-hierarchical application: Validation



Hierarchical application: Graphical model



$$K_i \sim \text{Binomial}(\theta_i, N_i)$$
$$\theta_i = \Phi(\phi_i)$$
$$\phi_i \sim \text{Normal}_{(0,+\infty)}(\mu_\phi, \sigma_\phi)$$
$$\sigma_\phi \sim \text{Uniform}(0, 1.5)$$

$$M_0 : \mu_\phi = 0$$
$$M_1 : \mu_\phi \sim \text{Normal}_{(0,+\infty)}(0, 1)$$

Hierarchical application: Results & Validation

Log Bayes factor?

- C^3 method: $log(BF_{10}) \approx -3.6 \rightarrow$ Strong evidence M_0
- Consistent with importance sampling method

OUTLINE

1. Bayesian hypothesis testing
2. Hypothesis testing with transdimensional MCMC
3. Applications
4. **Conclusion**

Conclusion

Estimating Bayes factors with C^3 method is..

1. *easy to implement*
2. *precise*
3. *flexible*

Conclusion

Estimating Bayes factors with C^3 method is..

1. *easy to implement*
2. *precise*
3. *flexible*

$\rightarrow$ C^3 method seems a good candidate for Bayesian hypothesis testing in experimental psychology

Questions