Model Selection and Multi-Model Inference: Big Picture

- Model Selection
  - Controversial topic
  - Lots of possible approaches (we will look at one)
    - Bayes Factors/Posterior Model Probabilities
  - Multiple implementations (we will look at one)
    - Using RJMCMC
- Multi-model inference
  - Approach for summarizing information (and uncertainty) across multiple candidate models
    - Bayesian model averaging

Posterior Model Probabilities

- Consider k models: M_1, M_2, …, M_k
  - Model indicators
  - Associated with model i are (a vector of) parameters θ_i
- Of interest $p(M_1|y), \ldots, p(M_k|y)$
  - $p(M_j|y)$ is the posterior probability of model j being "true" conditional on the data
  - Simple specification
    - Like most/all of Bayesian inference the devil is in the details
  - Require $p(M_1), p(M_2), \ldots, p(M_k)$
    - Prior model probabilities
    - We will not talk a lot about these (but they are important)

Simple Example

- Collected data $y_1, \ldots, y_n$
- Hypothesize two possible models for the data:
  - Model 1: $y_1, \ldots, y_n \stackrel{iid}{\sim} \mathcal{N}(0, 1)$
  - Model 2: $y_1, \ldots, y_n \stackrel{iid}{\sim} \mathcal{N}(\mu, 1)$
    - μ is unknown with prior $\mu \sim \mathcal{N}(0, \kappa^2)$
- Comparing $p(M_1|y)$ and $p(M_2|y)$
  - Effectively comparing whether μ is zero or non-zero
- Look at this example in more detail later

Bayes Factors

- Bayes factors are an alternative way to present the posterior model probabilities
- The Bayes factor (between model i and model j) is

$$BF_{ij} = \frac{p(y|M_i)}{p(y|M_j)}$$

- What is this quantity?
- Define a marginal likelihood (for model i) as:

$$p(y|M_i) = \int p(y|\theta_i, M_i)p(\theta_i)d\theta_i$$

- *Marginal* likelihood ratio
  - Parameters have been integrated out
    - Prior distribution on parameter matters (more on this later)
    - cf traditional likelihood ratio where θ is set to some value $\tilde{\theta}$

Bayes Factors vs Posterior Model Probabilities

- Bayes factors have a direct relationship to posterior model probabilities

$$
\begin{aligned}
BF_{ij} &= \frac{p(y|M_i)}{p(y|M_j)} \\
&= \frac{p(y|M_i)p(M_i)}{p(y)} \frac{p(y)}{p(y|M_j)p(M_j)} \frac{p(M_j)}{p(M_i)} \\
&= \frac{p(M_i|y)}{p(M_j|y)} \frac{p(M_j)}{p(M_i)} \\
&= \frac{\text{Posterior odds}}{\text{Prior odds}}
\end{aligned}
$$

- i.e. Bayes factors are the mechanism that turn prior odds into posterior odds
  - Posterior odds = BF × prior odds

Bayes Factors

- Jeffrey's suggests the following scale for Bayes factors

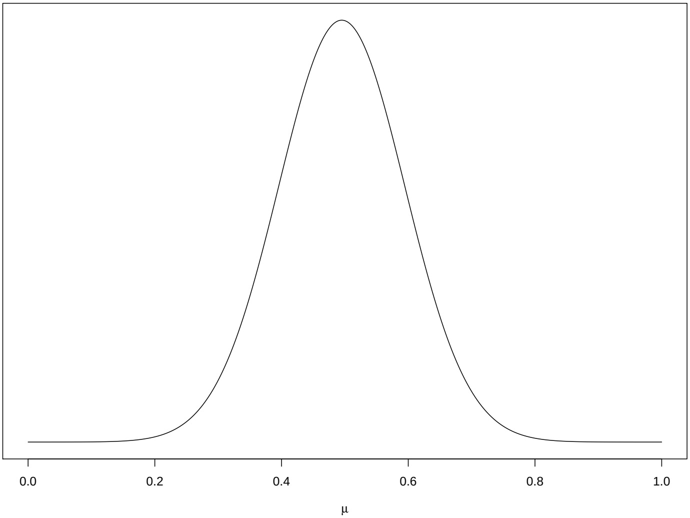
B_{10}	Evidence for M_1
< 1	Negative: support for M_0
1 to 3	Barely worth mentioning
3 to 12	Positive
12 to 150	Strong
> 150	Very strong

- The Bayes factor (posterior model probabilities) can give you evidence in support of a hypothesis/model

Return to example

- Data: $y_1, \ldots, y_n$
  - Model 1: $y_1, \ldots, y_n \overset{iid}{\sim} \mathcal{N}(0, 1)$
  - Model 2: $y_1, \ldots, y_n \overset{iid}{\sim} \mathcal{N}(\mu, 1)$
    - μ is unknown with prior $\mu \sim \mathcal{N}(0, \kappa^2)$
- We had $n = 100$ and $\kappa = 1$. We observed $\bar{y} = 0.5$
- Before we look at Bayes factor
  - First look at the posterior distribution of μ in model 2

Posterior of μ



BF for example

- Data: $y_1, \ldots, y_n$
  - Model 1: $y_1, \ldots, y_n \overset{iid}{\sim} \mathcal{N}(0, 1)$
  - Model 2: $y_1, \ldots, y_n \overset{iid}{\sim} \mathcal{N}(\mu, 1)$
    - μ is unknown with prior $\mu \sim \mathcal{N}(0, \kappa^2)$
- We had $n = 100$ and $\kappa = 1$. We observed $\bar{y} = 0.5$
- In this example we can evaluate the marginal likelihoods analytically (by hand):

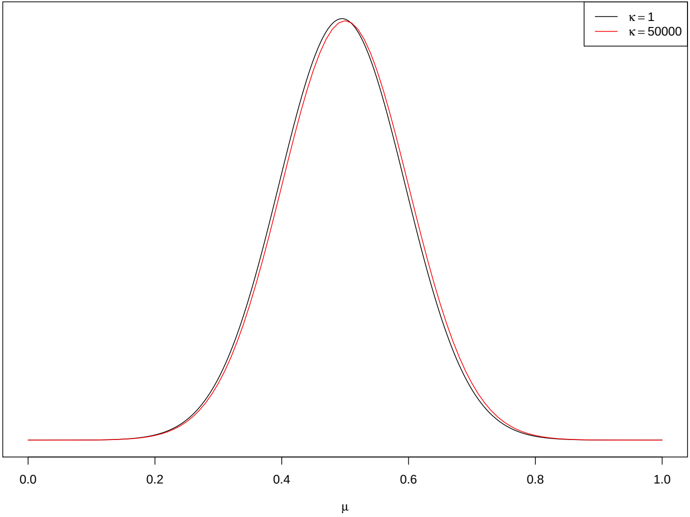
$$BF_{21} = (1 + n\kappa^2)^{-0.5} \exp\left(\frac{n^2\kappa^2}{2(1 + n\kappa^2)}\bar{y}^2\right)$$

- We need to plug-in some values!
  - $BF_{21} \approx 23600$
  - Strong support for model 2

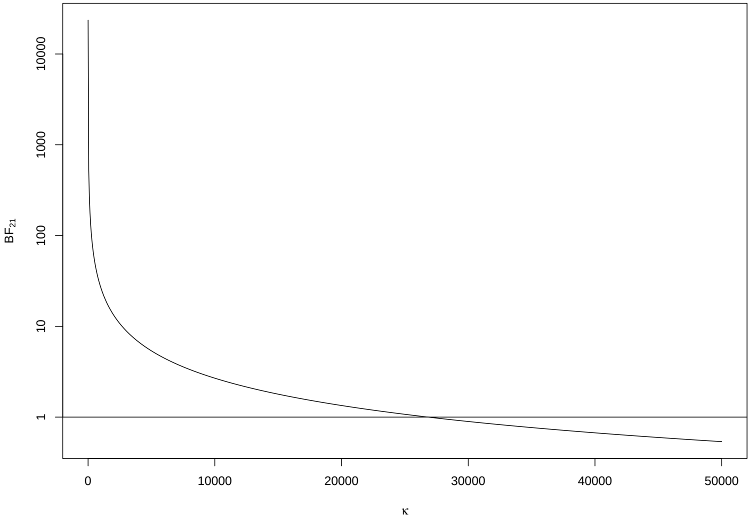
Caution I: Priors (on parameters) matter

- "Vague" / "non-informative" / "flat" priors can be problematic
- Plot posteriod distribution for μ and BF_{21} over a range of κ values from 1 to $50,000$
  - The prior for μ is becoming more and more flat

Posterior distribution for μ



Bayes factor



Caution I: Priors (on parameters) matter

- When $\kappa = 1$ we have strong support for model 2
- When $\kappa = 50,000$ we have support for model 1
- **Priors matter when using Bayes factors**
  - Even though the prior has little effect on the posterior distribution for μ

Caution II: Model probabilities vs p-values

- Even though Bayes factors share a lot in common with traditional hypothesis testing
  - Not the same
- $p(M_j|y)$ is not the same as a p-value
  - $p(M_j|y)$ is the probability of model j given the data y
  - A p-value is the probability of observing data as (or more) extreme than that observed assuming the null hypothesis is true.
- They are different quantities
- They often disagree
  - Referred to as Lindley's paradox

Problem

- It is easy to define a Bayes factor in terms of marginal likelihoods
  - Difficult to calculate it
- To find the marginal likelihood we need to evaluate the (nasty) integral that led us to use MCMC in the first place
- One approach is to once again avoid evaluating this interval using MCMC
  - Use a special flavor of MCMC called trans-dimensional MCMC
    - e.g. reversible jump MCMC

Trans-dimensional MCMC

- Include a model indicator
- Another unknown
  - Switch between models in different iterations
    - e.g. move from model 1 in iteration 1 to model 4 in iteration 2, etc
  - Find relative support for each model
    - Posterior model probability is estimated as the % of iterations in each model
- Why is it special/difficult?
  - Have to take into account differences in the dimension of parameters between different models

Approach of Carlin and Chib

- Complete parameter space
  - Make one “super” model that includes all parameters from every model
  - Model indicator that specifies which parameters are included in the likelihood function
    - Necessary to specify “pseudo-priors” for all parameters for when they are not included in likelihood
    - These can be chosen to “optimize” the algorithm (or chosen for convenience)

Reversible jump MCMC (Green)

- Consider moves between each pair of models separately
  - Have to specify how parameters in model i correspond to parameters in model j
  - Take care when the dimension of the parameters differs
    - Specify an "augmenting variable" that balances the dimension
- Various other approaches
  - Show that the two approaches mentioned are more similar than it appears
- Best seen with an example (in JAGS)

Example: Return to Lake Brunner[1]

- Return rates for brown trout in Lake Brunner, New Zealand
  - Tag and release trout. Observe which trout return one year later.
- Five candidate models:

1. $\text{logit}(\pi_i) = \beta_0$
2. $\text{logit}(\pi_i) = \beta_0 + \beta_1 S_i$
3. $\text{logit}(\pi_i) = \beta_0 \qquad\qquad + \beta_2 L_i$
4. $\text{logit}(\pi_i) = \beta_0 + \beta_1 S_i + \beta_2 L_i$
5. $\text{logit}(\pi_i) = \beta_0 + \beta_1 S_i + \beta_2 L_i + \beta_{12} S_i L_i$

- In JAGS

[1] Example from Link and Barker (2010)

JAGS code: part I

```
### Logistic regression
for (i in 1:n){
 returned[i] ~ dbern(p[i])
 logit(p[i]) <- beta0 + in.mod.sex*beta1*S[i] +
                               in.mod.len*beta2*L[i] +
                                  in.mod.int*beta12*SL[i]
}
```

JAGS code: part II

```
### Priors
beta0 ~ dt(0,0.04,3)
beta1 ~ dt(0,0.25,3)
beta2 ~ dt(0,0.25,3)
beta12 ~ dt(0,0.25,3)
```

JAGS code: part III

```
### Model indicator
mod ~ dcat(p.model[1:5])

### Determining whether terms are in the model
mod4 <- (mod==4)
mod5 <- (mod==5)
in.mod.sex <- (mod==2) + mod4 + mod5
in.mod.len <- (mod==3) + mod4 + mod5
in.mod.int <- mod5
```

Results

- $p(M_1|y) \approx 0.837$
- $p(M_2|y) \approx 0.045$
- $p(M_3|y) \approx 0.110$
- $p(M_4|y) \approx 0.006$
- $p(M_5|y) \approx 0.003$

Model Averaging

- Suppose we have K candidate models
  - e.g. linear regression with various possible predictor variables
- In all models a quantity of interest γ is well defined
  - e.g. prediction at a certain value
- We could find the best model
  - Make the prediction under that model
- Suboptimal
  - Not taking all uncertainty into account
  - Uncertainty in the model selection process
  - Interval estimate will be too precise
- Make the prediction averaging across the models

Model Averaging

- Suppose for each of K models we have $p(\gamma|y, M_i)$
  - Posterior distribution of γ under model i
- We want the "model averaged" posterior distribution

$$p(\gamma|y) = \sum_{i=1}^{K} p(\gamma|y, M_i)p(M_i|y)$$

- This distribution takes into account the model uncertainty
  - i.e. that we do not know the correct model M_i

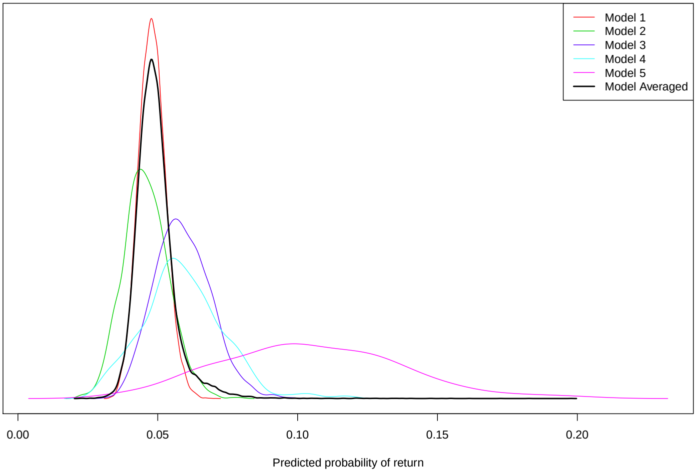
Example

- We can do this for the Lake Brunner trout example.
- Predict the return probability for a trout with sex 0 of (standardized) length 1.5.
- Either do this directly in JAGS (see model) or in R after model is fitted (if we have stored the appropriate parameter values)

JAGS code: part IV

```
### Predicting the observation
logit(pred.prob) <-
      beta0 + in.mod.sex*beta1*sexpred +
                  in.mod.len*beta2*lenpred +
                       in.mod.int*beta12*sexlenpred
```

Results


Model 1
Model 2
Model 3
Model 4
Model 5
Model Averaged
0.00
0.05
0.10
0.15
0.20
Predicted probability of return