

# Introduction to JAGS and Problem #1



Nora Mitchell  
January 24, 2017

# Goals for Today's Lab

- ▶ Grasp basic concepts behind Bayesian inference
- ▶ Learn the components to coding a Bayesian model
- ▶ Understand how to use R2jags to run a Bayesian model
- ▶ Set-up the model for Problem #1

# Bayesian Inference

Conceptions and misconceptions!

Let's see where we stand.

# Bayesian Inference: Why?

Why be Bayesian? Traditional, frequentist statistics ask: how likely are our observations given a certain hypothesis.

Bayesian statistics ask: how likely is our hypothesis given the observations and our knowledge, the question we really want to ask!

# Bayesian Inference: Why?

In addition, Bayesian statistics:

- ▶ most obviously, allows us to take into account prior knowledge
- ▶ gives increased **flexibility**
- ▶ gives a full posterior distribution of parameter estimates instead of point estimates
- ▶ compute **credible intervals** taken from the posterior that are easily interpretable: 95% cred. interval means that there's a 95% chance the true value lies in the interval!

# Bayesian Inference: Why?

Continued! Bayesian statistics:

- ▶ allows us to do model comparisons using Bayes Factors
- ▶ allows for estimation of **uncertainty**
- ▶ make comparisons of parameter estimates by drawing from the posterior distributions
- ▶ MCMC algorithms allow us to quickly explore parameter space
- ▶ Bayesian statistics can easily be used in a **hierarchical fashion**, with multiple levels of parameters nested inside each other

# A Brief Intro to Bayesian Inference

What is the *posterior probability* given a *prior probability* and *likelihood*?

What is the probability of our hypothesis given our data and previous expectations?

$$p(\phi|x) = \frac{p(x|\phi) \cdot p(\phi)}{p(x)}$$

$x = \text{data}$

$\phi = \text{parameters}$

# Bayesian Inference

What is the *posterior probability* given a *prior probability* and *likelihood*?

$$\begin{array}{c} \text{posterior} \\ | \\ p(\phi|x) = \frac{\begin{array}{c} \text{likelihood} \\ | \\ p(x|\phi) \end{array} \cdot \begin{array}{c} \text{prior} \\ | \\ p(\phi) \end{array}}{\begin{array}{c} p(x) \\ | \\ \text{marginal} \\ \text{probability} \end{array}} \end{array}$$

$x = \text{data}$

$\phi = \text{parameters}$



# Steps in Bayesian Inference

How do we do this?

1. Create a full **probability model**, incorporating observed data and unobserved parameters
2. Calculate the **posterior distribution** (conditional probability) of parameters given the observed data
3. Perform model diagnostics, interpret the results

# Bayesian Models

Example of a simple linear regression model in a Bayesian context:

$$y_i = \alpha + \beta x_i + e_i$$
$$e_i \sim N(0, \sigma^2)$$

# Bayesian Models

Example of a simple linear regression model in a Bayesian context:

$$\begin{array}{ccccccc} & & & \text{regression} & & & \\ & & & \text{coeffi-} & & & \\ & & & \text{cient} & & & \\ & & & \text{times} & & & \\ & & & \text{covari-} & & & \\ & & & \text{ate} & & & \\ \text{obs.} & & \text{intercept} & & & & \text{error} \\ | & & | & & | & & | \\ \boxed{y_i} & = & \boxed{\alpha} & + & \boxed{\beta x_i} & + & \boxed{e_i} \\ & & & & e_i \sim N(0, \sigma^2) & & \end{array}$$

# Bayesian Models

$$y_i = \alpha + \beta x_i + e_i$$
$$e_i \sim N(0, \sigma^2)$$

- ▶ Observations:  $x, y$
- ▶ Parameters:  $\alpha, \beta, \sigma$

Parameters will need prior distributions!

# Bayesian Inference: Priors

The *prior* is a distribution that takes into account your knowledge.

Priors can be:

- ▶ flat:  $U(-\infty, \infty)$
- ▶ uninformative or loose:  $N(0, 10000)$ 
  - ▶ informative:  $N(0, 0.001)$

# Bayesian Inference: Priors

As Paul says, a prior is like a rubber band.

Strong evidence from the likelihood can “pull” the posterior away from the prior distribution, but weak evidence from the likelihood will result in a posterior that resembles the prior.

On the other hand, a weak prior means that we have no real expectations, and the likelihood drives most of the posterior, even if the evidence is not very strong.

The choice of the prior, in terms of distribution type and strength, is very important!

# Interactive Demonstration

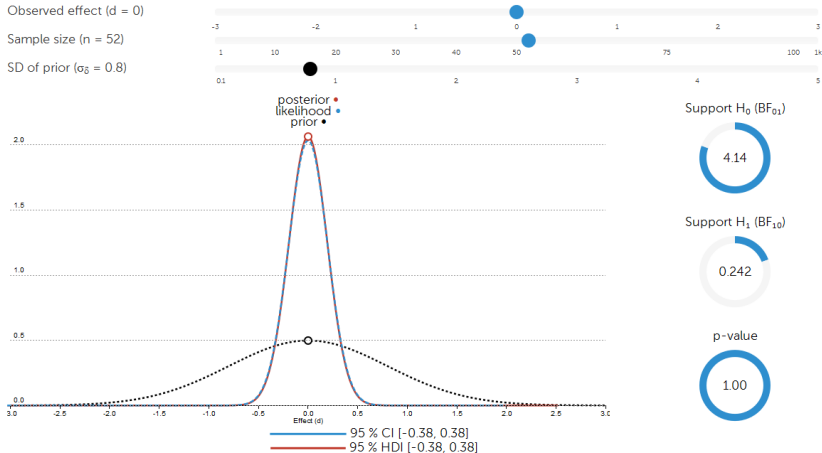
<http://rpsychologist.com/d3/bayes/>

## Settings

Observed effect ( $d = 0$ )

Sample size ( $n = 52$ )

SD of prior ( $\sigma_0 = 0.8$ )



# Hierarchical Modeling

“It’s complicated...”.

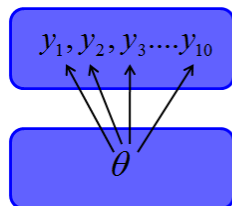
Includes models with lots of different set-ups, including:

- ▶ multi-level models
- ▶ random effect models
- ▶ variance-component models
- ▶ mixed effect models

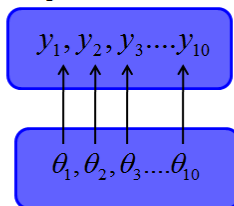


# Hierarchical Modeling

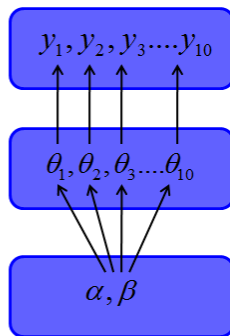
Common Parameter



Independent Parameters

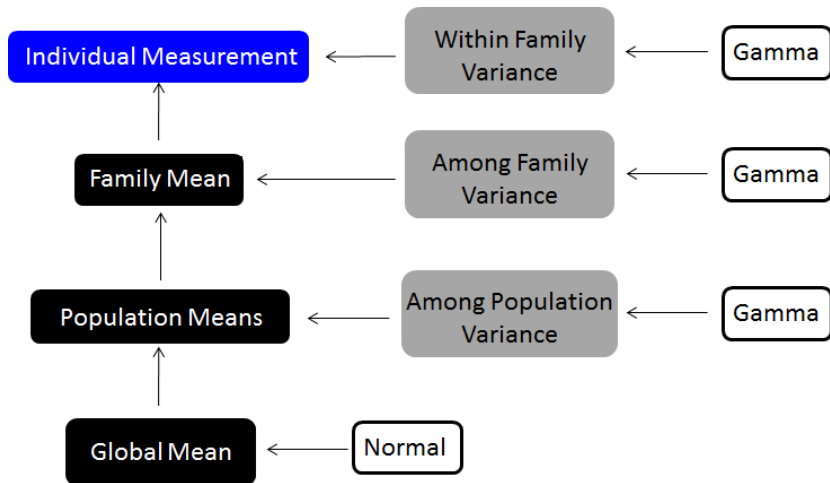


Hierarchical



# Hierarchical Modeling

A biological example, measuring variance components in a common garden.



# Analysis

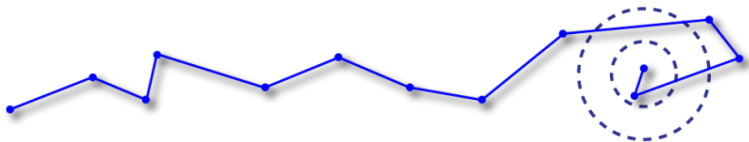
How do we actually do this?

Read in observations, write a model, connect them, and explore parameter space using MCMC sampling!

# Exploring parameter space: MCMC

We use a Markov-Chain Monte Carlo simulation to explore parameter space

Think of this as a robot climbing a hill...



# What Is JAGS?

## JAGS: **J**ust **A**nother **G**ibbs **S**ampler

Gibbs Sampling is an MCMC sampling algorithm often used in a Bayesian context to obtain a sequence of observations from a multivariate probability distribution.

BUGS (**B**ayesian Inference **U**sing **G**ibbs **S**ampling) is a similar algorithm, but we will use JAGS because it can interface easily with R.

# How do we use JAGS?

Steps to running a model in JAGS:

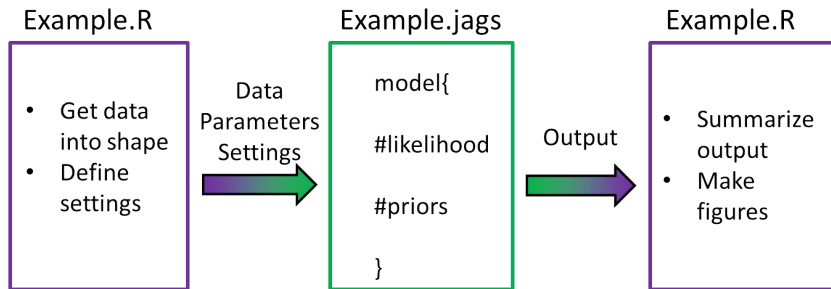
1. Definition of the model (including parameters!)
2. Compilation: program makes a graph depicting the model
3. Initialization: set initial values, set seed, choose sampler
4. Adaptation and burn-in: set MCMC settings
5. Monitoring: record sampled values

# R2jags

We will interface JAGS with R. To do this, we will use the package R2jags which we installed and played with last week.

1. Write code in R to import the data and pass it to JAGS
2. Add information re: our compilation/initiation/MCMC settings
3. Code the “model” into a separate .txt file, and then use R to pass this to JAGS as well.
4. Run the code in R, which talks to JAGS and the model file.
5. Get output and perform diagnostics, interpret results

## How do R and JAGS communicate?





# R and JAGS

R and JAGS are not the same program!

We run JAGS *through* R

We will always have **TWO** separate files when working between R and JAGS

1. R script written and executed in RStudio (Example.R)
2. Model file written in a text editor (Example.jags)

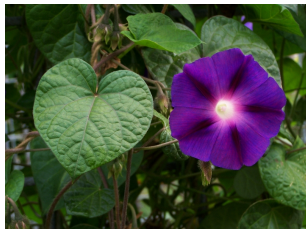
We **do not** ever open up JAGS directly!

# Problem #1

Kent assigned Problem #1 today, hurrah!

- ▶ Step 1: Get data set up in R (R2jags should already be installed). Write model file in a text editor.
- ▶ Step 2: Play with the code and model, run it.
- ▶ Step 3: Get output and analyze it!

# Problem #1



Data from Schoen and Clegg (1985)

Used fast and slow allelic forms (biallelic) of an esterase to analyze mating system in the common morning glory *Ipomoea purpurea*

Measured genotypes of mothers and their offspring (one per)

*I. purpurea* reproduces by a mixture of self-fertilization and outcrossing!

# Problem #1

1. What is the outcrossing rate in this experimental population?
2. Is there any evidence of an allele frequency difference between maternal plants and pollen?
3. What is the inbreeding coefficient in the maternal plants?

# Problem #1

Kent's hints:

- ▶ Treat each maternal genotype separately
- ▶ Treat the number of mothers belonging to each maternal genotype as a multinomial sample
- ▶ Collect genotype counts of mothers into a vector in R, send this to JAGS
- ▶ Use his priors

# Problem #1

The project, including write-up and code, will be due to me **electronically** next **Tuesday, January 31**, before lab starts (**9:30 am!**)

See expectations handout

[nora.mitchell@uconn.edu](mailto:nora.mitchell@uconn.edu)

Feel free to email me with questions during the week, or stop by office hours Thursday 9:30-10:30 in TLS 313.

## Works Cited

- ▶ Ogle, K. 2013. Introduction to Bayesian Modelling, Joint 2013 MBI-NIMBioS-CAMBAM Summer Graduate Workshop. Knoxville, TN.
- ▶ Plummer, M. 2003, JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003):20-22.
- ▶ Sanchez-Robles, J. M., J. L. Garcia-Castano, F. Balao, A. Terrab, L. Navarro-Sampedro, K. Tremetsberger, and S. Talavera. 2014. Schoen, Daniel J., and Michael T. Clegg. "The influence of flower color on outcrossing rate and male reproductive success in *Ipomoea purpurea*." *Evolution* (1985): 1242-1249.