Copyright 2019, Jeffrey Stanton

IST772 Bayesian Inference (Week 5)
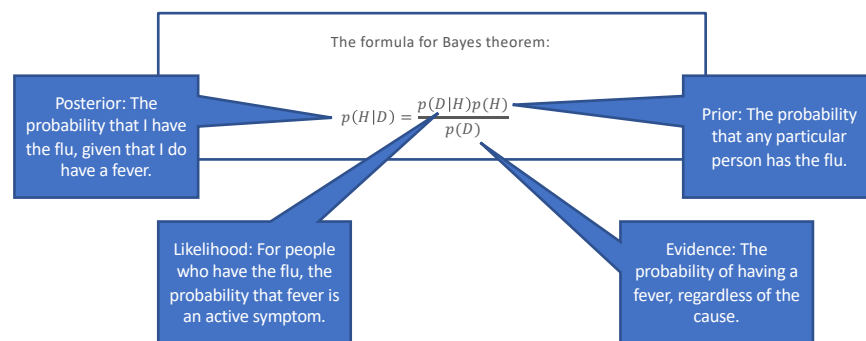
Pre-class activity:

The following problem is modified from Haller and Krauss (2002): "Suppose you have a drug that you suspect may alter performance on a certain task. You compare the means of control and experimental groups (n=20 in each group). Further, suppose you use the independent means t-test and your result is (t = 2.7, df=18, p=0.0112). Consider these interpretations of your results:

1. You have disproved the null hypothesis (that is, the hypothesis of no difference between the population means).
2. You have calculated the probability of the null hypothesis being true.
3. You have proved your experimental hypothesis (there is a difference between the population means).
4. You can deduce/calculate the probability of the experimental hypothesis being true from the provided results.
5. Assuming you decide to reject the null hypothesis, you know the probability that you are making the wrong decision.
6. You have a reliable experimental finding in the sense that if, hypothetically, you repeated the experiment many times, you would obtain a significant result on 99% of the replications.

Send me a private chat message indicating which, if any, of these statements is false. It is possible that more than one statement is false.
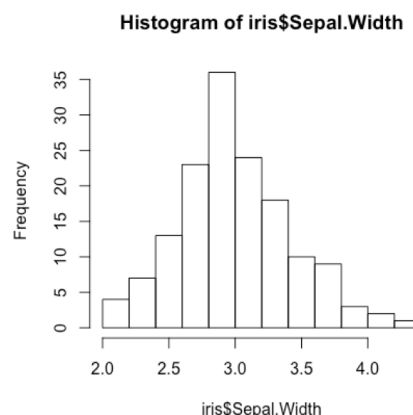
Copyright 2019, Jeffrey Stanton and Jeffrey Saltz

Bayes Theorem: I have a fever. . . Do I have the flu?

The formula for Bayes theorem:

Posterior: The probability that I have the flu, given that I do have a fever.

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

Prior: The probability that any particular person has the flu.

Likelihood: For people who have the flu, the probability that fever is an active symptom.

Evidence: The probability of having a fever, regardless of the cause.

In practice, using multivariate/metric data instead of simple event counts, the denominator cannot be directly calculated, as its definition contains an integral with no closed form expression. We therefore must use an iterative computational technique to estimate p(D). This is the purpose of MCMC estimation.

MCMC Estimation – Metropolis-Hastings Demo

- A Markov Chain is a series of linked values where the value of t+1 depends only on the value of t
- A Markov "kernel" defines the transition between value t and value t+1
- The Metropolis-Hastings algorithm is one such kernel: We can model an arbitrary distribution using a simple rule. . .

Histogram of iris$Sepal.Width

Mean: 3.057333; SD: 0.4358663

Accept or Reject a Candidate Value Based on its Likelihood in the Proposal Distribution

target <- rnorm(1, 0, sd=sigma) # Draw one value from a proposal normal distribution
can <- x + target # Candidate value links to previous value: Makes it a Markov Chain

Acceptance probability as the min of 1 or the ratio of density values of candidate and x.
aprob <- min(1, **dnorm(can, mean=mu, sd=sigma)/dnorm(x, mean=mu, sd=sigma)**)

# Accept/reject the candidate value by
# comparing to a random draw from uniform(0, 1)
if (aprob >= runif(1)) { x <- can }

The relative probability in favor of the candidate instead of the previous value (x).

vec[i] <- x # Either keep the old x or use the new value depending upon the previous test

Results of 10,000 Candidates (less 1000 burnin)

Trace Plot of Post Burnin Values

Histogram of Post Burnin Values
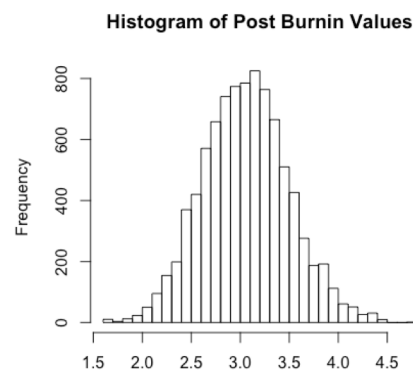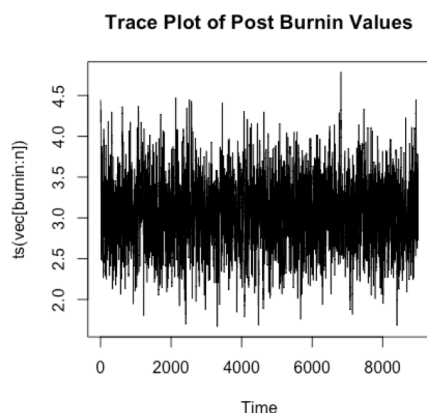
median	mean	var	std.dev
3.057871002	3.057546233	0.192490420	0.438737302

Original data: Mean: 3.057333; SD: 0.4358663

Breakout 1 –

- Compare MPG for automatic and manual transmission groups using the best procedure:
  - If you have installed JAGS, the rjags package, and the BEST package, you will be able to do this in your version of R
  - If you have not installed these, you can use a website: http://www.sumsar.net/best_online/
- Produce some diagnostics and interpret your results
- Share your code on https://codeshare.io/5MYNZ3

All handouts for this class: **https://tinyurl.com/IST777summer19**

Fun With rjags

- rjags is the interface between R and JAGS
- JAGS – Just another Gibbs sampler – is a variant on the BUGS language that was developed to facilitate MCMC analysis of unknown distributions
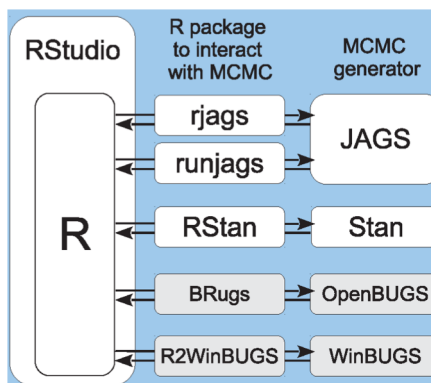- John Kruschke uses BUGS and JAGS in his detailed book on Bayesian analysis

Image credit: J. Kruschke

School of Information Studies | Syracuse University   7

JAGS is a Model Building Language

In this simple model, we want to find a posterior distribution for mu and sigma squared based on a normal prior for mu, a uniform prior for sigma-squared, and observed data modeled as normal.

```
model{

for (i in 1:n){
    y[i] ~ dnorm(beta.0, tau.sq)
}

    beta.0 ~ dnorm(0 , 0.0001 )
    tau.sq <- 1/sigma.sq
    sigma.sq ~ dunif (0 , 100)
}
```

Here we are modeling the observations as normal.

This sets the distribution for the priors on mu.

This sets the distribution for the priors on sigma squared.

School of Information Studies | Syracuse University   8

4

The JAGS Object Returned to R

A complex, nested data object consisting of operational parameters and R code to conduct MCMC sampling.

Name	Type	Value
jags.m	list [8] (S3: jags)	List of length 8
ptr	function	function() { ... }
data	function	function() { ... }
model	function	function() { ... }
state	function	function(internal = FALSE) { ... }
nchain	function	function() { ... }
iter	function	function() { ... }
sync	function	function() { ... }
recompile	function	function() { ... }

School of Information Studies | Syracuse University 9
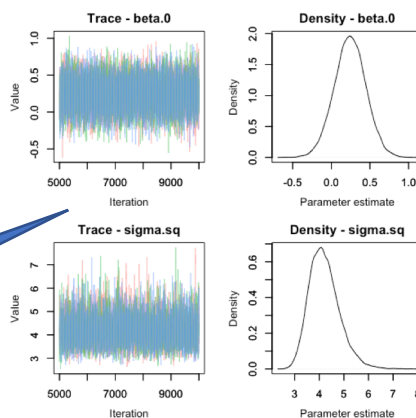
MCMC results from n=100 observations of y

```
y <- rnorm(n=100, mean=0, sd=5)

mean(y); sd(y); var(y)

[1] 0.2434796
[1] 2.008435
[1] 4.03381
```

Beta.0 is our model for mu and sigma.sq is our model for sigma squared.

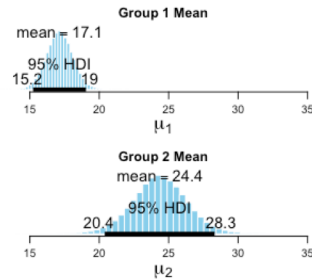School of Information Studies | Syracuse University 10

Breakout 2 –

- Examine a JAGS model for comparing two independent means
- Set up a model run with data from mtcars
- Run the model and examine the results
- Share your code on https://codeshare.io/5MYNZ3

All handouts for this class: **https://tinyurl.com/IST777summer19**

Plot Explorations

- In the BEST package the plotAll() command draws diagnostic plots for a variety of aspects of the Bayesian estimation
- The following slides explain some of that output
- The whole display as produced by R appears in the next slide, followed by several slides that focus on particular parameters

Group 1 Mean
mean = 17.1
95% HDI
15.2   19
μ_1

Group 2 Mean
mean = 24.4
95% HDI
20.4   28.3
μ_2

The histograms to the left show the means from each of the 100,000 MCMC simulations for each of the two groups. The X-axis is calibrated identically between the two histograms to simplify the visual comparison.
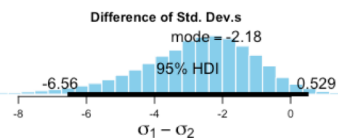
Difference of Means
mean = -7.24
99.9% < 0 < 0.1%
95% HDI
-11.6   -2.85
$\mu_1 - \mu_2$

For all of these histograms, the dark line on the X-axis shows the extent of the "highest density interval." Each HDI represents the 95% of the possibilities for the posterior distribution of the given parameter, in this case the means.
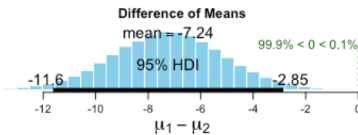
The histogram above shows the difference between the group means for each of the 100,000 MCMC simulations.
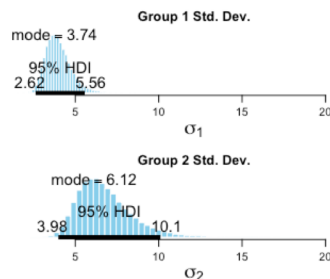

The BEST procedure also models the standard deviations for each of the 100,000 MCMC simulation runs. As a result, we can also compare SDs between the two groups. The difference in SDs is modeled in the histogram below.

Group 1 Std. Dev.
mode = 3.74
95% HDI
2.62   5.56
σ_1

Does the HDI for the difference in SDs overlap with zero? Is there a credible difference between the SDs of the two groups?

Group 2 Std. Dev.
mode = 6.12
95% HDI
3.98   10.1
σ_2

Difference of Std. Dev.s
mode = -2.18
95% HDI
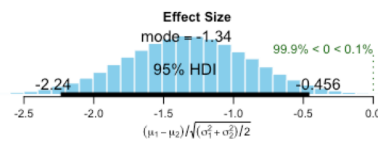-6.56   0.529
$\sigma_1 - \sigma_2$

For each of the 100,000 MCMC simulation runs, the BEST procedure calculates an index analogous to df, notated below as the log (base 10) of "nu."

What we are looking for is values of log10(nu) that are in the region of 1.5. To the extent that the HDI of this display does NOT overlap with 1.5, there may be severe non-normality. The BEST procedure is robust against non-normality whereas the classical t-test is not.
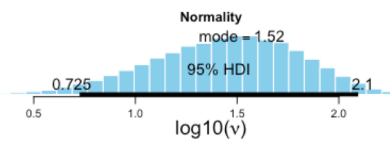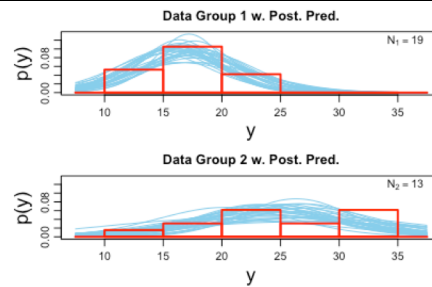
The histogram below models the "effect size" of each of the simulation runs. Effect size is calculated as Cohen's D, which is simply the difference in means divided by the pooled standard deviation (a kind of average SD across the two groups).

If the effect size HDI does not overlap with zero, this is suggestive of a credible difference between means. Remembering that Cohen's D is calibrated in SDs, anything with an absolute value > 0.8 is considered a large effect.

Normality
mode = 1.52

95% HDI

0.725   2.1

0.5   1.0   1.5   2.0

$$\log 10(\nu)$$

Effect Size
mode = -1.34

99.9% < 0 < 0.1%

95% HDI

-2.24   -1.5   -1.0   -0.456   0.0

-2.5   -2.0   -0.5

$$(\mu_1 - \mu_2)/\sqrt{(\sigma_1^2 + \sigma_2^2)/2}$$

These displays provide a diagnostic showing whether it is appropriate to be using the BEST procedure. The blue curves show a random sample of plots of the posterior t distribution for each group, using the mean, SD, and normality coefficient generated by a particular run of the MCMC process. The red histogram is the actual sample data.

To the extent that data are highly multi-modal or weirdly distributed, the t-distribution may be a bad fit.

Data Group 1 w. Post. Pred.

$N_1 = 19$

p(y)   0.08
        0.00

10   15   20   25   30   35
                y

Data Group 2 w. Post. Pred.

$N_2 = 13$

p(y)   0.08
        0.00

10   15   20   25   30   35
                y

For example, the rightmost category (30-35) in the Data Group 2 histogram could be a second mode for these data. With only n=13 observations, however, there is not enough data to really say for sure. In the absence of other anomalies, this is not a severe enough problem to raise a concern.
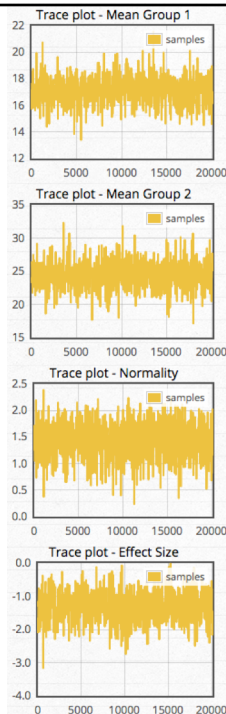
These trace plots do not appear in the plotAll() output from BESTmcmc(), but they do appear on Rasmus Baath's webpage. We will also see examples of these trace plots later in the course.

Each trace plot represents the same data that you saw in the HDI histograms on previous pages, but rather than displaying them as a histogram, they show as a time series.

RB's webpage defaults to 20,000 MCMC simulations. For each simulation run, the various parameters generated for that step in the robot's path are displayed as a new dot just to the right of the previous dot. You can watch these unfold in real time as the Javascript processes your data.

RB says that these should look like a hairy caterpillar. We want to see roughly equal variation around a reasonably stable mean (for the topmost plot about 17 mpg) throughout the whole run (post burn-in), with just a few small spikes here and there.

Paper of the Week – Plummer 2017

- The JAGS language is not difficult to understand if you know R
- Martin Plummer worked on the implementation of JAGS that we use and write the manual
- You can explore the manual to learn more about the JAGS models we explored in this class and to develop your own

JAGS Version 4.3.0 user manual

Martyn Plummer

28 June 2017

Homework and Practice Exam

- Make sure you are using the updated syllabus that I distributed at the beginning of the semester (on the wall and in the handouts folder).
- The second practice exam can be found in the handouts folder. I recommend that you try to complete the practice exam in an hour.
- I will post the key on Thursday.
- Feel free to submit your exam file to the LMS if you are unsure about any of your answers.
- The homework for week five is exercises 6 through 10 on pages 86 and 87.