Bayesian Analysis for Political Science Workshop

Sarah Hunter

University of Colorado, Boulder

sarah.hunter@colorado.edu

Day 2

Overview

Review

- Bayesian inference is based on a subjective definition of probability.
- Bayesian methods reverse the conditionality of frequentist arguments to give you the probability of the parameters, given the data.
- Our Bayes Mantra: $\pi(\theta|X) \propto L(\theta|X) * p(\theta)$
  - Where the posterior distribution is a combination of the prior information and the data

Today's Agenda

- Priors: AKA The Controversy of Bayesian Analysis
- Types of Priors and how to use them
- Do priors actually have much impact?
- The Bayesian Linear Model

Priors

$$\underbrace{\pi(\theta|X)}_{\text{Posterior}} \propto \underbrace{L(\theta|X)}_{\text{Likelihood}} * \underbrace{p(\theta)}_{\text{Prior}}$$

Defined Probability statements about unknown quantities of interest.

- Otherwise known as the prior belief about the probability of the unknown quantity (i.e. parameter)
- A way to include prior research and expectations directly into a model.
- Also necessary for Bayes' Law.

What are Priors in a Model?

The Frequentist

A way for the research to inject subjective beliefs into a regression

The Bayes-or-Bust Bayesian

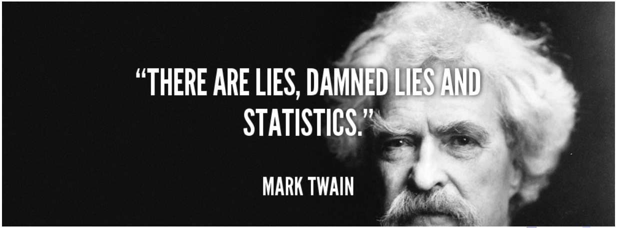
A way to include prior information AND uncertainty into the model; a way to not pretend ignorance

The Practical Bayesian

A mathematical necessity to do inference without NHST

The Prior Controversy

- Frequentist see priors as adding personal-subjective beliefs to statistical analysis.
- However, all statistical models are created using subjective choices.
- "Everyone is a Bayesian, some of us know it."



Types of Priors

Point Priors

- Point Estimate as a prior
- Not often a good solution (i.e. no uncertainty)

Distributions

- Distributions (with mean and standard deviations) as priors
- Most commonly used
- A way to inject both information and unvertainty into the model.

Types of Priors

Uninformative Priors

- Intentionally add little new information about the unknown parameter.
- Useful when little is known, or to satisfy frequentist criticism of subjectivity

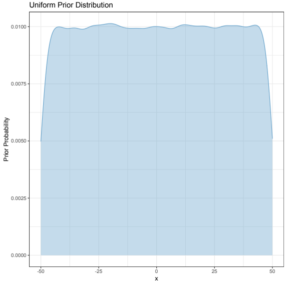
Informative/ Elicited Priors

- Adds information based on previous research directly into the model.
- These priors give more weight to certain values than others.

Uninformative Priors

Any prior distribution can be uninformative, but the usual choice is the **Uniform Prior**

- A prior based on a uniform distribution
- Every value within the range of the uniform distribution has an equal probability.
- Two types:
  - Proper: Integrates to 1
  - Improper: Does not integrate to 1



Informative Priors

Informative Priors (AKA Elicited Priors) fall into several categories:

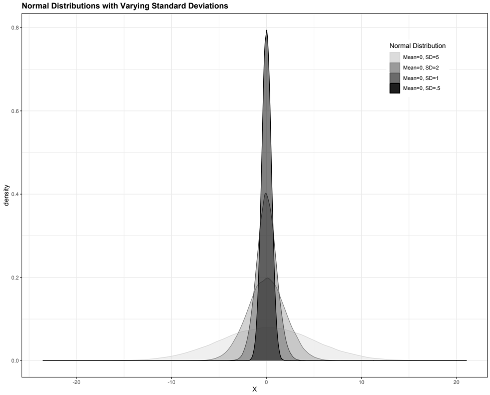
- **Clinical Priors**: use information form experts working on the project
- **Skeptical Priors**: Assumes the hypothesized effect does not exist
- **Enthusiastic Priors**: Opposite of the skeptical prior

Choosing an Appropriate Prior

Keep in mind:

- The function of the prior
- Some priors give weird results
- Some priors make estimation easier
- Publication
- Some priors actually make estimation impossible

Factors that Control the Impact of the Prior



- Sample size
- Strength of relationship being tested
- **Standard Deviation** of the prior
- Prior distribution

The Posterior Distribution

Instead of point estimates of parameters (e.g. β), Bayesian models estimate a **posterior distribution**.

Advantages of Posteriors:

- No Central Limit Theorem
- No assumptions of normality
- Inference
- Inference in small samples

The Posterior Distribution

Can you know the shape of the posterior distribution?

Yes!

How? Conjugacy.

Form	Conjugate Prior Distribution	Hyperparameters
Bernoulli	Beta	$\alpha > 0,\ \beta > 0$
Binomial	Beta	$\alpha > 0,\ \beta > 0$
Multinomial	Dirichlet	$\theta_j > 0,\ \Sigma\theta_j = \theta_0$
Negative Binomial	Beta	$\alpha > 0,\ \beta > 0$
Poisson	Gamma	$\alpha > 0,\ \beta > 0$
Exponential	Gamma	$\alpha > 0,\ \beta > 0$
Gamma (incl. χ^2)	Gamma	$\alpha > 0,\ \beta > 0$
Normal for μ	Normal	$\mu \in \mathbb{R},\ \sigma^2 > 0$
Normal for σ^2	Inverse Gamma	$\alpha > 0,\ \beta > 0$
Pareto for α	Gamma	$\alpha > 0,\ \beta > 0$
Pareto for β	Pareto	$\alpha > 0,\ \beta > 0$
Uniform	Pareto	$\alpha > 0,\ \beta > 0$

Implementing Bayesian Models

1. Write your model function.
2. Choose your priors.
3. Convert your data to list format.
4. Choose the parameters you want to monitor.
5. Estimate the model.
6. Conduct convergence diagnostics.

Model Specification

$$y_i = \beta_0 + \beta_1 * x1_i + \beta_2 * x2_i$$

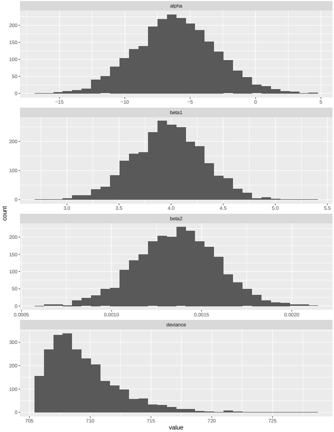
where:

$$y \sim \mathcal{N}(\mu, \tau)$$
$$\beta_0 \sim \mathcal{N}(0, 1)$$
$$\beta_1 \sim \mathcal{N}(0, 1)$$
$$\beta_2 \sim \mathcal{N}(0, 1)$$
$$tau \sim Gamma(.1, .1)$$

Example

```
jags.model<-function(){
for(i in 1:N){
prestige[i]~dnorm(mu[i], tau)
mu[i]<-alpha + beta1*education[i]
+ beta2*income[i]
}
alpha~dnorm(0, 1)
beta1~dnorm(0, 1)
beta2~dnorm(0, 1)
tau~dgamma(.1,.1)
}
```

Model Estimation



Example

```
fit<-jags(data=jagsdata,
inits=NULL, params,
n.chains=2, n.iter=4000,
n.burnin=400, model.file=
prestige.model.jags)
print(fit)
```

Model Presentation and Interpretation

What R Gives You

```
##          mu.vect sd.vect    2.5%     25%     50%     75%   97.5% Rhat n.eff
## alpha     -0.792   3.181  -6.932  -2.954  -0.804   1.429   5.263 1.000  2000
## beta1      7.302   3.366   1.136   4.976   7.260   9.574  14.043 1.001  2000
## beta2      0.942   0.172   0.602   0.830   0.944   1.051   1.270 1.003  1000
## deviance 125.741   2.620 122.427 123.907 125.148 126.996 132.582 1.001  2000
##
```

How to read the results:

- "mu.vect" = mean of the posterior distribution
- "sd.vect" = sd or the posterior distribution
- "2.5 %" : "97.5%" = The point at which that much of the distribution is at or below.

Model Presentation and Interpretation

What R Gives You

```
##          mu.vect sd.vect    2.5%     25%     50%     75%   97.5%  Rhat n.eff
## alpha     -0.792   3.181  -6.932  -2.954  -0.804   1.429   5.263 1.000  2000
## beta1      7.302   3.366   1.136   4.976   7.260   9.574  14.043 1.001  2000
## beta2      0.942   0.172   0.602   0.830   0.944   1.051   1.270 1.003  1000
## deviance 125.741   2.620 122.427 123.907 125.148 126.996 132.582 1.001  2000
##
```

What do you need from that?

- The names of the variables
- The posterior mean
- The deviance
- Either:
  - 2.5% and 97.5% (what makes the 95% Credible Interval)
  - The standard deviation

Model Presentation and Interpretation

	mean	sd	2.5%	97.5%
Constant	-5.84	3.14	-11.93	0.45
Education	4.03	0.34	3.36	4.71
Income	0.001	0.001	0.001	0.002
deviance	709.81	2.82	706.17	716.71

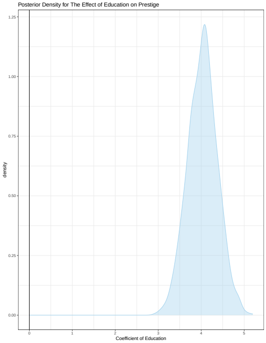
Credible Intervals

95% of the posterior distribution for β_1 falls between 1.136 and 14.043.

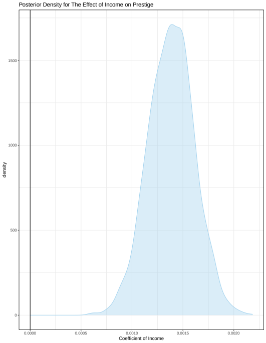
You know the shape of the posterior, just plot it.

Model Presentation and Interpretation

How much of the distribution is above/below 0?



Proportion of density above 0: .9999999999999



Proportion of density above 0: .9999999999999

Back to Hypothesis Testing

How do you know your hypothesis is supported?

- Look for zero in the credible interval (not the best idea)
- How much of the posterior distribution is above or below zero?
- Highest Posterior Density (HPD) Regions (don't have to be contiguous)

A Note on Directional Hypotheses

Perfect to test with Bayes!

- What are your thoughts on directional hypotheses?
- How is Bayes ideal for testing directional hypotheses?

Until Next Time