

Bayesian Analysis for Political Science Workshop

Sarah Hunter

University of Colorado, Boulder

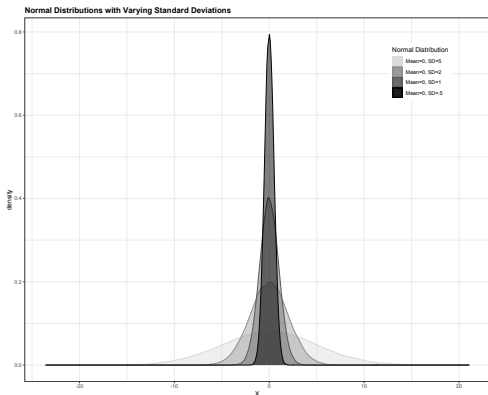
sarah.hunter@colorado.edu

Day 3

Overview

- 1 Review
- 2 Today's Agenda
- 3 Under the Hood
- 4 Types of Sampling
- 5 Convergence
- 6 Convergence Diagnostics
- 7 Solving Slow Convergence

Factors that Control the Impact of the Prior



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

Back to Hypothesis Testing

How do you know your hypothesis is supported?

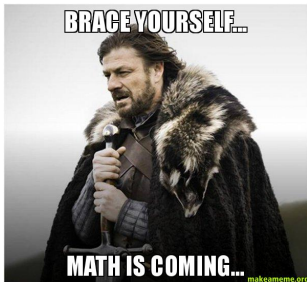
- Look for zero in the credible interval (not the best idea)
- How much of the posterior distribution is above or below zero?
- Highest Posterior Density (HPD) Regions (don't have to be contiguous)

Today's Agenda

- What is going on under the hood?
- Types of Sampling
- Convergence
- No Evidence for Non-Convergence



The Math of Bayes



The Simple Bayes Rule

$$\pi(\theta|X) = \frac{L(\theta|X) * p(\theta)}{\int p(\theta)L(\theta|X)d\theta}$$

However, the reality of models are quite complicated when adding the marginal likelihoods for each parameter

The Messy Posterior Distributions

Integrate this:

$$\pi(\gamma, \theta) \propto L(\gamma, \theta | \mathbf{X}, \mathbf{Y}) p(\theta) p(\gamma) p(b | \tau) p(\tau)$$

which becomes:

$$\propto \prod_{i=1}^n \prod_{j=1}^{C-1} \prod_{k=1}^p [\Lambda(\theta_j - \mathbf{X}_i' \gamma + \mathbf{b}_i) - \Lambda \theta_{j-1} - \mathbf{X}_i' \gamma + \mathbf{b}_i)]^{z_{ij}} * \\ \exp\left(-\frac{(\gamma_k - \mu_{\gamma k})^2}{2\sigma_\gamma^2} - \frac{\theta_j^2}{2\sigma_\theta^2} - \frac{b_i^2}{2\tau^2} - \frac{\delta_2}{\tau}\right) \tau^{-(\delta_1+1)}$$

So, how do we deal with that?

- Think about MLE methods.
- The basic principle of likelihood theory (according Fisher): what value of the unknown parameter is “most likely” to have generated the observed data
- The Goal of MLE: To find the parameters that maximizes the likelihood function of the observed data
- Mount Likelihood and Iterative processes

Markov Chain Monte Carlo (MCMC)

Monte Carlo:

- Remember the random draws of distributions that I did yesterday?
- AKA Simulations

Markov Chains:

- Colloquially: “A Markov chain wanders around the state space remembering only where it has been in the last period.”
(Stolen from Ryan Bakker's Bayes Class)
- Formally: A stochastic process where the value of π at time t is only dependent on the value of π at time $t-1$

A Simple Markov Chain Example

Say we have a choice between two things. People that normally choose X_1 have an 80% chance of selecting X_1 again, while those who normally select X_2 have a 40% chance of choosing X_2 again.

We can then get the transition matrix:

$$\text{current period} \begin{cases} \theta_1 \\ \theta_2 \end{cases} \begin{matrix} \overbrace{\theta_1 \ \theta_2}^{\text{next period}} \\ \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{matrix}.$$

A Simple Markov Chain Example

- 1 Assign a starting point:

$$S_0 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$$

- 2 To get to the first state, we multiply the initial state by the transition matrix:

$$S_1 = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix}$$

- 3 The next few iterations:

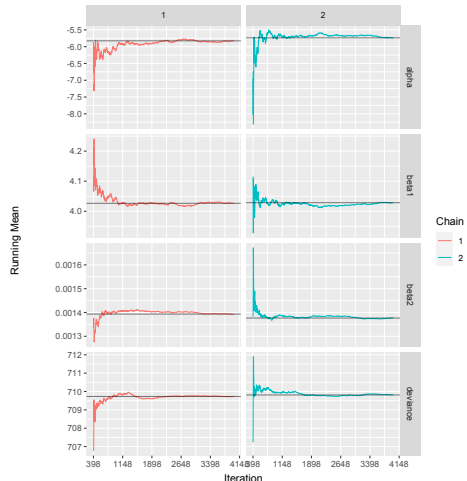
$$\text{Second state: } S_2 = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.74 & 0.26 \end{bmatrix}$$

$$\text{Third state: } S_3 = \begin{bmatrix} 0.74 & 0.26 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.748 & 0.252 \end{bmatrix}$$

$$\text{Fourth state: } S_4 = \begin{bmatrix} 0.748 & 0.252 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.7496 & 0.2504 \end{bmatrix}$$

MCMC Methods

- Seek out areas of highest density
- Only have memory of the previous iteration
- Keep running under they reach a **stable state**



Potential Problems with MCMC



According to Gelman:

- The assumed model may not be realistic
- Errors in calculation or programming
- Slow convergence

Types of Sampling

- Inversion Sampling
- Rejection Sampling
- Metropolis-Hastings

Inversion Sample

Procedure

- Draw a uniform random number u between 0 and 1. This represents the area under $f(x)$ to that point.
- Then $z=F^{-1}(u)$ is a draw from $f(x)$.

End Goal: to find a value that is the integral of $f(x)dx$

But, Inversion Sampling has its faults:

- Some inverse functions are impossible to derive (think the normal distribution)
- Does not work with multivariate distributions as the inverse is not unique beyond one dimension.

Rejection Sampling

Procedure

- Sample a value z from a distribution $g(x)$ which is easy to sample from and for which all values of $m * g(x)$ (aka the envelope function) are greater than $f(x)$ at all points.
- Compute the ratio: $R = \frac{f(z)}{mg(z)}$
- Sample $u \sim U(0, 1)$. If $R > u$, then accept z as a draw from $f(x)$. Otherwise return to step 1.
- Then we compute the ratio of densities of $f(x)$ to $m * g(x)$ for a given value of x .
- Finally, we compare this ratio to a random uniform draw. This ratio is then the probability we accept a draw at a given value x as coming from $f(x)$.

Rejection Sampling

Advantages of Rejection Sampling

- Can be used more most distributions

Limitations

- Finding the correct envelope function can be impossible
- Inefficient

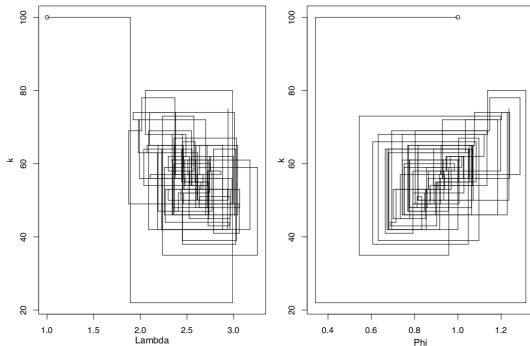
Gibbs Sampler

- Adaptive Rejection Sampling
- The algorithm used by JAGS



Gibbs Sampler

Simple Gibbs Sampler Example



Metropolis-Hastings

Procedure

- Establish starting values for parameters
- Draw a 'candidate' value from a proposal density
- Assess probability this is from the target distribution—like rejection

Advantages

- Algorithm that generates samples from the full joint density
- Works on multivariate distributions
- Doesn't require an envelope

Defining Convergence

How do you know your MCMC chain is done?

Definition

When our sampler reaches a stationary distribution (i.e. does not move outside the distribution)

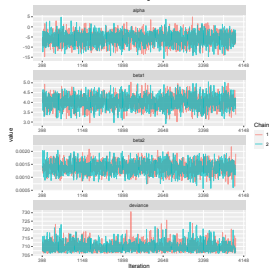
- You cannot “prove” convergence
- You can provide evidence against nonconvergence
- Convergence is important because if your model is not converged, it is worthless

Convergence Diagnostics

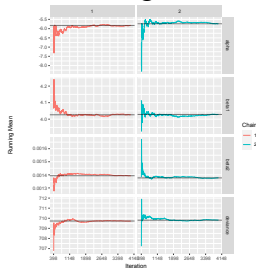
- Visual Diagnostics
- Geweke Test
- Gelman and Rubin
- Heidelberger-Welch
- Raftery-Lewis

Visual Diagnostics

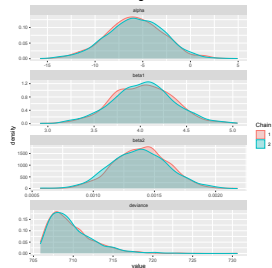
Traceplots



Running Plots



Density Plots



Geweke Test

Compares the posterior mean of the first half of the Markov Chain with the second half

Interpretation:

H_0 : Convergence

H_A : Non-Convergence

Example

R Code

```
geweke.diag(angellfit.mcmc)
```

Gelman and Rubin Test

Compares within and between chain variances for each variable

Interpretation:

The larger the difference, the more evidence of non-convergence ($\hat{R} > 2$ tends to be cause for alarm)

Example

```
gelman.plot(angellfit.mcmc)
```

Heidelberger-Welch

Tests if each variable is from a stationary distribution

Interpretation:

H_0 : Variable is from a stationary distribution

H_A : Variable is not from a stationary distribution

Example

```
heidel.diag(angellfit.mcmc)
```

Raftery-Lewis

Tells the researcher the minimum number of iterations that would be needed to estimate the specified quantile to the desired precision if the samples in the chain were independent.

Interpretation:

If you need more than 100,000 iterations, you should be concerned.

Example

```
raftery.diag(angellfit.mcmc)
```

A Note on Convergence Diagnostics

- Convergence does not necessarily mean you have a good model.
- Always run more than one convergence diagnostic
- Don't be surprised/sad to see that your convergence diagnostics give you conflicting information.
- Statistics is an art as much as a science.

So your JAGS model is taking forever...



Tips to Speed Convergence

- Good initial values
- Standardized variable
- Try a multivariate normal distribution
- Change sampling algorithms
- Just wait

Until Next Time