Bayesian Analysis and MCMC

Scott A. Baldwin
My Goals

1. Introduce Bayesian ideas and why you might care

2. Demonstration
Goals of Statistical Analysis

1. Estimate Parameters $\mu, \sigma, r, \beta, \lambda$

2. Predictions What will new data look like?
Approaches

Frequentist
Maximum Likelihood

Bayesian
MCMC
Maximum Likelihood

What parameters most likely produced the data?
Flip a coin 10 times:
6 Heads; 4 Tails

What's the probability of heads?

\[ p(6) = \binom{10}{6} p^6 (1 - p)^{10 - 6} \]

unknown
\[ p(6) = \binom{10}{6} p^6 (1 - p)^{10-6} \]

\( p = 0.6 \) is the ML estimate
Goals of Statistical Analysis

1. Estimate Parameters $\mu, \sigma, r, \beta, \lambda$

2. Predictions What will new data look like?
Predicted Data

\texttt{rbinom(10000, 10, .6)}
Goals of Statistical Analysis

1. Estimate Parameters \( \mu, \sigma, r, \beta, \lambda \)

2. Predictions

What will new data look like?
Bayesian Analysis

Estimate unknowns via Bayes theorem
Estimate unknowns via Bayes theorem

\[ p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory}) p(\text{theory})}{p(\text{data})} \]
Estimate unknowns via Bayes theorem

\[ p(\text{theory} | \text{data}) = \frac{p(\text{data} | \text{theory}) p(\text{theory})}{p(\text{data})} \]
Priors

Represent Prior Beliefs

Uncertainty about the unknowns prior to seeing the data
Prior for an intraclass correlation

\[ \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \]

Prior research suggests most ICCs for group therapy studies range from 0-0.30
Prior for an intraclass correlation

\[ \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \]
“More generally, though, I think we should avoid the temptation to think that, when a Bayesian inference goes wrong, it has to be a problem with the prior. That’s old-fashioned thinking, the idea that the likelihood is God-given and known perfectly, leaving us all to fight over our priors. In many cases, the model matters (for example, in our discussion above about natural-seeming but flawed discrete models). Even if the data model generally makes sense, its details can matter: as I point out to my students, the prior only counts once in the posterior, but the likelihood comes in over and over again, once for each data point.”

Estimate unknowns via Bayes theorem

\[ p(\text{theory} \mid \text{data}) = \frac{p(\text{data} \mid \text{theory}) p(\text{theory})}{p(\text{data})} \]
Posterior Distributions

What is a posterior distribution?

Combination of information from:
the data and the prior

A probability distribution for a parameter
Distribution not a point estimate
Distribution not a point estimate
Posterior simulations

If the posterior distribution conforms to a known probability distribution:

• we just use what we know about the probability distribution

If we don’t, we use simulation:

Simon Jackman (paraphrasing): Anything we want to know about an unknown parameter can be found by simulating from the distribution of that parameter.
tempdata <- rnorm(n=20000,mean=?,sd=?)

What’s the mean? What’s the sd?

Use the 20,000 draws to learn about the distribution
tempdata <- rnorm(n=20000,mean=?,sd=?)

> mean(tempdata)  > sd(tempdata)

> quantile(tempdata, c(.025,.975))
   2.5%   97.5%
30.45645 69.79223
tempdata <- rnorm(n=20000, mean=?, sd=?)
Posterior of Unknown Form

But we have functions for normal distributions, so simulating is easy....What do we do if we don’t know the actual form of the posterior distribution?
Markov Chain Monte Carlo
MCMC

All you (or the computer program) need to know is the form of the posterior up to a constant

MCMC will produce random draws from the posterior
MCMC will produce random draws from the posterior 

\[ \rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \]

Can easily obtain posterior distributions for combination of parameters
Can easily obtain posterior distributions for combination of parameters

\[ GC_t = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_{p,t,e}^2 / m} \]

Baldwin, Imel, and Atkins, 2012
<table>
<thead>
<tr>
<th>Group</th>
<th>Anx</th>
<th>Control</th>
<th>Dep</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRN ROI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERN ROI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERN</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Baldwin, Clayson, & Larson, under review
Priors help stabilize estimates

Baldwin & Fellingham, 2013
Priors help give more realistic estimates

Bayesian Estimate of the Cluster Variance when the REML Estimate is 0

Baldwin & Fellingham, 2013
MCMC can be used when ML has difficulty

Therapist Fidelity in MI - MITI

Baldwin, Imel, Braithwaite, & Atkins, in progress
MCMC can be used when ML has difficulty

Swiss Cheese Problem

Lee, Baldwin, & Atkins, in progress
But I want the data, and only the data, to speak!

“More generally, though, I think we should avoid the temptation to think that, when a Bayesian inference goes wrong, it has to be a problem with the prior. That’s old-fashioned thinking, the idea that the likelihood is God-given and known perfectly, leaving us all to fight over our priors. In many cases, the model matters (for example, in our discussion above about natural-seeming but flawed discrete models). Even if the data model generally makes sense, its details can matter: as I point out to my students, the prior only counts once in the posterior, but the likelihood comes in over and over again, once for each data point.”

Prediction - Likelihood Matters

Day 1

Minutes of Moderate-to-Vigorous Physical Activity

Baldwin, Baldwin, & Fellingham, in progress
Goals of Statistical Analysis

1. Estimate Parameters $\mu, \sigma, r, \beta, \lambda$

2. Predictions What will new data look like?
Demo
Software

Options

- General purpose MCMC software
  - WinBUGS
  - JAGS
  - STAN
  - PyMC
  - Proc MCMC

```R
model {
  for (i in 1:mpath1) {
    fcz1_path[i] ~ dnorm(upath[pathid[i]], ppath)
  }
  for (j in 1:gpath1) {
    upath[j] ~ dnorm(bpath, pupath)
  }

  bpath ~ dnorm(0, .0001)
  varpath ~ dunif(0, 100)
  ppath <- 1/varpath
  errpath ~ dunif(0, 200)
  ppath <- 1/errpath

  for (i in 1:mcont1) {
    fcz1_cont[i] ~ dnorm(ucont[contid[i]], pcont)
  }
  for (j in 1:gcont1) {
    ucont[j] ~ dnorm(bcont, pucont)
  }

  bcont ~ dnorm(0, .0001)
  varcont ~ dunif(0, 100)
  pucont <- 1/varcont
  errcont ~ dunif(0, 200)
  pcont <- 1/errcont

  icccont <- varcont/(varcont+errcont)
  icccpath <- varpath/(varpath+errpath)
}
```
PROC MCMC

proc mcmc data=one nbi=20000 nmc=400000 thin=50 diag=(autocorr ess geweke raftery)
   propcov=quanew monitor=(_parms_) simreport=10 outpost=normal dic;
ods select Parameters PostSummaries PostIntervals tadpanel dic ess AutoCorr Geweke Raftery;
parms beta0 su;
parms se;

   prior beta0  ~ normal(0, var=1000);
prior su  ~ gamma(shape=12, scale=10.5);
prior se  ~ gamma(shape=12, scale=10.5);

   random u ~ normal(beta0, sd=su) subject=newid monitor=(u_1);

   model mvpa ~ normal(u, sd=se);
run;
Software

Software with Bayes options

• Mplus
• MLWin
• R packages
  • MCMCglmm
• SAS
Software

Write your own code

• Any general purpose programming language
  • R
  • SAS IML
  • Python
  • C, C++
  • Fortran
  • JAVA
  • etc.
Thanks! Questions?