36-463/663: Multilevel and Hierarchical Models

Multilevel Models in lmer and jags

Brian Junker
132E Baker Hall
brian@stat.cmu.edu

Outline

- Quick review: Bayesian Statistics, MCMC and JAGS
- *Example 1*: Minnesota Radon – Intercept Only
  - What's new?
  - What is $\hat{R}$?
- *Example 2*: Mn Radon: Level 1 predictor "floor", Level 2 predictor "log(uranium)"
- *Example 3*: CD4 levels in HIV-positive youth

Bayesian Statistics and MCMC

- Our slogan(s)
  - (posterior)$\propto$(likelihood)$\times$(prior)
  - (posterior)$\propto$(level 1)$\times$(level 2)
  - (posterior)$\propto$(level 1)$\times$(level 2)$\times$(level 3)
  - etc.
- Often there are no formulas for posterior, so we have to simulate draws from the posterior
- MCMC is a general method for doing posterior draws
  - Calculate complete conditionals
  - Invent methods for sampling from complete conditionals
  - Stitch together samples into a Markov Chain
  - Summarize Markov Chain with histograms, medians, credible intervals, etc.

Bayesian Statistics and JAGS

- JAGS[1] automates MCMC
  - Describe “slogan” to JAGS in R-like language
  - JAGS figures out complete conditionals
  - JAGS invents MCMC sampling methods
  - JAGS generates the random draws for you
- JAGS is just the latest implementation of the BUGS[2] platform (declarative language and AI system for constructing MCMC). Others:
  - WinBUGS
  - OpenBUGS

1. **J**ust **A**nother **G**ibbs **S**ampler
2. **B**ayesian inference **U**sing **G**ibbs **S**ampling

Multilevel Models in JAGS

- MLM form

$$y_i = \alpha_{0j[i]} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$
$$\alpha_{0j} = \beta_0 + \eta_j, \quad \eta_j \sim N(0, \tau^2)$$

- Hierarchical form

Level 1: $y_i \sim N(\alpha_{0j[i]}, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_0, \tau^2)$

- BUGS/rube form

```
model {
  # LEVEL 1
  for (i in 1:N) {
    log.radon[i] ~
       dnorm(a0[county[i]],prec.y)
  }
  # LEVEL 2
  for (j in 1:MAX.COUNTY=85) {
    a0[j] ~ dnorm(b0,prec.cty)
  }

  # PRIORS/LEVEL 3:
  b0 ~ dnorm(0,PREC.b0=1e-6)
  prec.y ~ dgamma(ALPHA=0.001,BETA=0.001)
  prec.cty ~ dgamma(ALPHA=0.001,BETA=0.001)

  # CONVERT PRECISION TO VARIANCES
  var.y <- 1/prec.y
  var.cty <- 1/prec.cty
}
```


Multilevel Models in JAGS

- MLM form

$$y_i = \alpha_{0j[i]} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$
$$\alpha_{0j} = \beta_0 + \eta_j, \quad \eta_j \sim N(0, \tau^2)$$

- Hierarchical form

Level 1: $y_i \sim N(\alpha_{0j[i]}, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_0, \tau^2)$

- BUGS/rube form

```
model {
  # LEVEL 1
  for (i in 1:N) {
    log.radon[i] ~
       dnorm(a0[county[i]],prec.y)
  }
  # LEVEL 2
  for (j in 1:MAX.COUNTY=85) {
    a0[j] ~ dnorm(b0,prec.cty)
  }

  # PRIORS/LEVEL 3:
  b0 ~ dnorm(0,PREC.b0=1e-6)
  prec.y ~ dgamma(ALPHA=0.001,BETA=0.001)
  prec.cty ~ dgamma(ALPHA=0.001,BETA=0.001)

  # CONVERT PRECISION TO VARIANCES
  var.y <- 1/prec.y
  var.cty <- 1/prec.cty
}
```

Multilevel Models in JAGS

- MLM form

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \epsilon_i, \\ &\quad \epsilon_i \sim N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \eta_j, \\ &\quad \eta_j \sim N(0, \tau^2) \end{aligned}$$

- Hierarchical form

Level 1: $y_i \sim N(\alpha_{0j[i]}, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_0, \tau^2)$

- BUGS/rube form

```
model {
  # LEVEL 1
  for (i in 1:N) {
    log.radon[i] ~
       dnorm(a0[county[i]],prec.y)
  }
  # LEVEL 2
  for (j in 1:MAX.COUNTY=85) {
    a0[j] ~ dnorm(b0,prec.cty)
  }

  # PRIORS/LEVEL 3:
  b0 ~ dnorm(0,PREC.b0=1e-6)
  prec.y ~ dgamma(ALPHA=0.001,BETA=0.001)
  prec.cty ~ dgamma(ALPHA=0.001,BETA=0.001)

  # CONVERT PRECISION TO VARIANCES
  var.y <- 1/prec.y
  var.cty <- 1/prec.cty
}
```


Multilevel Models in JAGS

- MLM form

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \epsilon_i, \\ &\quad \epsilon_i \sim N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \eta_j, \\ &\quad \eta_j \sim N(0, \tau^2) \end{aligned}$$

- Hierarchical form

Level 1: $y_i \sim N(\alpha_{0j[i]}, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_0, \tau^2)$

- BUGS/rube form

```
model {
  # LEVEL 1
  for (i in 1:N) {
    log.radon[i] ~
       dnorm(a0[county[i]],prec.y)
  }
  # LEVEL 2
  for (j in 1:MAX.COUNTY=85) {
    a0[j] ~ dnorm(b0,prec.cty)
  }

  # PRIORS/LEVEL 3:
  b0 ~ dnorm(0,PREC.b0=1e-6)
  prec.y ~ dgamma(ALPHA=0.001,BETA=0.001)
  prec.cty ~ dgamma(ALPHA=0.001,BETA=0.001)

  # CONVERT PRECISION TO VARIANCES
  var.y <- 1/prec.y
  var.cty <- 1/prec.cty
}
```

Look at the Distributions section (pp 56ff in manual14.pdf)

Multilevel Models in JAGS

- MLM form

$$y_i = \alpha_{0j[i]} + \epsilon_i,$$
$$\epsilon_i \sim N(0, \sigma^2)$$
$$\alpha_{0j} = \beta_0 + \eta_j,$$
$$\eta_j \sim N(0, \tau^2)$$

- Hierarchical form

Level 1: $y_i \sim N(\alpha_{0j[i]}, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_0, \tau^2)$

- BUGS/rube form

```
model {
  # LEVEL 1
  for (i in 1:N) {
    log.radon[i] ~
       dnorm(a0[county[i]],prec.y)
  }
  # LEVEL 2
  for (j in 1:MAX.COUNTY=85) {
    a0[j] ~ dnorm(b0,prec.cty)
  }

  # PRIORS/LEVEL 3:
  b0 ~ dnorm(0,PREC.b0=1e-6)
  prec.y ~ dgamma(ALPHA=0.001,BETA=0.001)
  prec.cty ~ dgamma(ALPHA=0.001,BETA=0.001)

  # CONVERT PRECISION TO VARIANCES
  var.y <- 1/prec.y
  var.cty <- 1/prec.cty
}
```


Multilevel Models in JAGS

- MLM form

$$y_i = \alpha_{0j[i]} + \epsilon_i,$$
$$\epsilon_i \sim N(0, \sigma^2)$$
$$\alpha_{0j} = \beta_0 + \eta_j,$$
$$\eta_j \sim N(0, \tau^2)$$

- Hierarchical form

Level 1: $y_i \sim N(\alpha_{0j[i]}, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_0, \tau^2)$

- BUGS/rube form

```
model {
  # LEVEL 1
  for (i in 1:N) {
    log.radon[i] ~
       dnorm(a0[county[i]],prec.y)
  }
  # LEVEL 2
  for (j in 1:MAX.COUNTY=85) {
    a0[j] ~ dnorm(b0,prec.cty)
  }

  # PRIORS/LEVEL 3:
  b0 ~ dnorm(0,PREC.b0=1e-6)
  prec.y ~ dgamma(ALPHA=0.001,BETA=0.001)
  prec.cty ~ dgamma(ALPHA=0.001,BETA=0.001)

  # CONVERT PRECISION TO VARIANCES
  var.y <- 1/prec.y
  var.cty <- 1/prec.cty
}
```

Have to add priors
to all free parameters

Example 1: Minnesota Radon – Intercept Only

- MLM:

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \epsilon_i, \\ & \quad \epsilon_i \sim N(0, \sigma^2) \\ \alpha_{0j} &= \beta_0 + \eta_j, \\ & \quad \eta_j \sim N(0, \tau^2) \end{aligned}$$

- Hierarchical:

$$\textbf{Level 1:}\ y_i \sim N(\alpha_{0j[i]}, \sigma^2)$$

$$\textbf{Level 2:}\ \alpha_{0j} \sim N(\beta_0, \tau^2)$$

- Demonstration in R and rube()/JAGS…
- (comparison with lmer also)

What's new?

- summary(rube.object): point estimates and CI's for "some" parameters. Others available in
  - rube.object$mean
  - rube.object$sd
  - rube.object$median
  - rube.object$sims.list
  - etc
- p3(rube.object): interactive graphical summaries

What's new?

- rube()/WinBUGS/JAGS automatically
  - Runs 3 separate MCMC chains
  - Runs each MCMC chain for 2000 steps, and throws away the first 1000 steps as "burn-in"
  - Thins each chain by 1/3 to reduce autocorrelation
  - *You can change this when you run rube(); see pp. 25ff. of the "rube.pdf" documentation.*
- rube()/WinBUGS/JAGS reports an "Rhat" statistic for each parameter estimated
  - Rhat is a ratio of between-chain to within-chain variation
  - *When the chain is converged, Rhat < 1.2. Otherwise, the chain hasn't run long enough yet.*

What is $\hat{R}$?

- Suspose we have M chains:

Chains				Means	Variances
$\theta^{(1;1)}$,	$\theta^{(1;2)}$,	$\ldots$,	$\theta^{(1;N)}$	$\bar{\theta}_1$	W_1
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
$\theta^{(M;1)}$,	$\theta^{(M;2)}$,	$\ldots$,	$\theta^{(M;N)}$	$\bar{\theta}_M$	W_M
			Grand mean	$\bar{\theta}$	

- Define

$$W = \frac{1}{M(N-1)} \sum_{m=1}^{M} \sum_{n=1}^{N} (\theta^{(m;n)} - \bar{\theta}_n)^2 = \frac{1}{M} \sum_{m=1}^{M} W_m$$
$$= \text{Average within-chain variance}$$

$$B = \frac{M}{M-1} \sum_{m=1}^{M} (\bar{\theta}_m - \bar{\theta})^2$$
$$= \text{Between-chain variance, inflated for sample size}$$

$$V = \frac{M-1}{M} W + \frac{1}{M} B$$
$$= \text{Pooled variance estimate,}$$

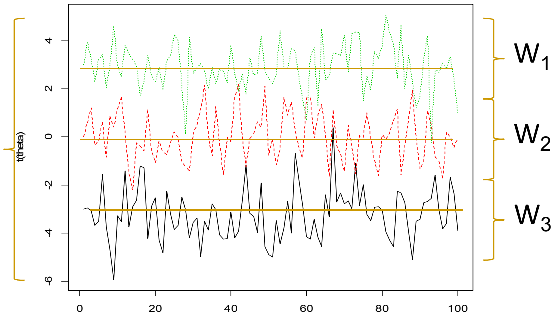
What is $\hat{R}$?

- Separated chains:

$$\begin{aligned} W &= \frac{1}{N}\sum_{n=1}^{N} W_n &= 1.02 \\ B &= \frac{M}{N-1}\sum_{n=1}^{N}(\bar{\theta}_n - \bar{\theta})^2 &= 938.53 \\ V &= \frac{M-1}{M}W + \frac{1}{M}B &= 10.40 \end{aligned}$$

$$\hat{R} = \sqrt{V/W} = 3.19$$

B

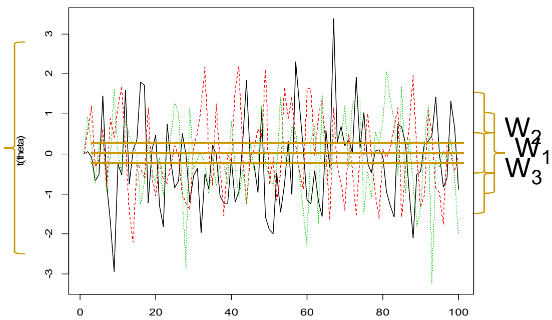


- Converged chains:

$$\begin{aligned} W &= \frac{1}{N}\sum_{n=1}^{N} W_n &= 1.02 \\ B &= \frac{M}{N-1}\sum_{n=1}^{N}(\bar{\theta}_n - \bar{\theta})^2 &= 1.32 \\ V &= \frac{M-1}{M}W + \frac{1}{M}B &= 1.03 \end{aligned}$$

$$\hat{R} = \sqrt{V/W} = 1.00$$

B



Example 2: Mn Radon: Level 1 predictor "floor", Level 2 predictor "log(uranium)"

- MLM:

$$\begin{aligned} y_i &= \alpha_{0j[i]} + \alpha_{1j[i]}(\text{floor})_i + \epsilon_{ij[i]},\ \epsilon_{ij} \sim N(0,\sigma^2) \\ \alpha_{0j} &= \beta_{00} + \beta_{01}\log(\text{uranium}_j) + \eta_{0j},\ \eta_{0j} \sim N(0,\tau_0^2) \\ \alpha_{1j} &= \beta_{10} + \eta_{1j},\ \eta_{1j} \sim N(0,\tau_1^2) \end{aligned}$$

- Hierarchical:

Level 1: $y_i \sim N(\alpha_{0j[i]} + \alpha_{1j[i]}(\text{floor})_i, \sigma^2)$

Level 2: $\alpha_{0j} \sim N(\beta_{00} + \beta_{01}\log(\text{uranium}_j), \tau_0^2)$

$\alpha_{1j} \sim N(\beta_{10}, \tau_1^2)$

- Demonstration in R and rube()/JAGS…

Example 3: CD4 in HIV-positive youth

- See R handout, and demonstration in class

Summary

- JAGS automates MCMC
  - Specify (posterior)$\propto$(level 1)$\times$(level 2)$\times$…
    in an R-like language
  - JAGS designs and runs the MCMC for you
- rube() makes JAGS less painful
  - Gelman & Hill use R2WinBUGS(), we will use rube()
- Partial summaries of parameter estimates, and good initial graphs, come easy in rube().
  - Rhat < 1.2 is a handy "convergence diagnostic"