36-463/663: Multilevel & Hierarchical Models

(P)review: In-Class Final Exam

Brian Junker

132E Baker Hall

brian@stat.cmu.edu

Overall

- Go back and review lecture 14, week 08 (review for midterm exam)
  - The final will focus on material after the midterm

    <u>BUT</u>
  - I will assume you can handle anything from the first part of the course, if it comes up in some way on the final
- Final exam will cover material from week 06 through week 14 of the course:
  - Multi-level models
  - Model-checking using AIC/BIC/DIC, residuals, and/or simulation
  - Basic facts about MLE's, Fisher Information & SE's, CI's, etc.
  - Bayesian statistics, the slogan, computation with conjugate priors
  - Random simulation and MCMC as a way to "estimate" a model, CI's
  - Fitting and interpreting models using JAGS and rube()
  - Using JAGS for simulation tests
- Selected topics from G&H Ch's 11-15 and Lynch Ch's 2, 3, 4 & 9
  - G&H Ch's 16-18, 20 and 24 have material similar to what we covered in class, as well, though I didn't really assign them as reading.

Reading

- All class notes, R handouts, HW's, solutions, etc.
- Gelman & Hill, Ch's 11-15, (16-18, 20 and 24)
  - I did not do anything in as great a detail as G&H. The level of material I expect you to "get" is somewhere between my class notes and G&H, but closer to my class notes.
  - One topic that I avoided and will not ask you about is problems involving the Wishart distribution. Nothing wrong with it, just not enough time to do everything.
- Lynch Ch's 2, 3, 4, 9
  - I will only ask things about Ch 9 that are related to lectures or hw's that you've had, fitting multilevel models with JAGS/rube (same as WinBUGS)
  - I didn't talk about the dirichlet, inverse-gamma, or wishart distributions, and I won't ask anything about them either.
  - I will feel free to refer to any distributions or models that appeared in class notes, R notes, or hw problems & solutions.

Multi-level models

- Dealing with groups:
  - *Totally pooled* and *totally unpooled* models use only fixed effects
  - *Partially pooled* models use random effects for variation among groups (and may use fixed effects for other things, like overall intercepts, slopes, etc.)
- lmer() notation, lmer() fits, lmer() output
- Regression to the mean and "shrinkage"

Multi-level models

- Hierarchical Bayes Model

$$\text{Level 2: } \alpha_i \overset{iid}{\sim} N(\beta_0, \tau^2)$$

$$\text{Level 1: } y_{ij} \overset{indep}{\sim} N(\alpha_i, \sigma^2)$$

- Multi-Level Model

$$\begin{aligned} y_{ij} &= \alpha_i + \epsilon_{ij}, \ \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \\ \alpha_i &= \beta_0 + \eta_i, \ \eta_i \overset{iid}{\sim} N(0, \tau^2) \end{aligned}$$

- Variance-Components Model

$$\begin{aligned} y_{ij} &= \beta_0 + \eta_i + \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \\ & \qquad\qquad\qquad\quad \eta_i \overset{iid}{\sim} N(0, \tau^2) \end{aligned}$$

- lmer() notation

y ~ 1 + (1|group)

Model-checking using AIC/BIC/DIC and/or simulation

- Appropriate use of likelihood ratio tests, vs AIC/BIC/DIC
- Definitions and anticipated effects of AIC/BIC/DIC
  - Greater penalties lead to simpler models
  - When does the random effect variance lead to higher or lower estimates of number of parameters for DIC?
- Idea of simulation tests: H_0: real data comes from same distribution as data simulated from the fitted model.

Basic facts about MLE's, Fisher Information & SE's, CI's, etc.

- For MLE
  - Need a function *proportional* to L(θ)
  - Calculate MLE by setting 0 = L'(θ)
  - Calculate $SE = 1/\sqrt{I(\theta)}$ where I(θ) = E[-LL''(θ)]
- Confidence Interval (CI)

$$(\hat{\theta} - 2SE,\ \hat{\theta} + 2SE)$$

"In 95% of analyses this interval will cover the true θ"

Bayesian statistics, the slogan, computation with conjugate priors

- For Bayes
  - Need a function proportional to L(θ)
  - Need a prior distribution
  - ***Slogan:*** (posterior) $\propto$ (likelihood)$\times$(prior)
  - Calculate posterior mean, SE,
    - Use formula if you have one
    - Use simulation if you don't!
- Credible Interval (CI)

$$(\hat{\theta} - 2SE,\ \hat{\theta} + 2SE) \text{ or } (\theta^{post}_{0.025},\ \theta^{post}_{0.975})$$

"The probability that θ is in the CI is 95%."

Bayesian statistics, the slogan, computation with conjugate priors

- A prior distribution is ***conjugate*** to the likelihood, if the posterior distribution comes from the same family as the prior; e.g.:
  - If the prior is beta and the likelihood is binomial, the posterior will again be a beta
  - If the prior is gamma and the likelihood is poisson, the posterior will again be gamma
  - If the prior and the likelihood are both normal (and the variances are known) then the posterior will be normal
- Problems with conjugate priors generally have simple formulas for their solutions.
- Problems without conjugate priors cause us to resort to simulation methods to get an answer.

Bayesian statistics, the slogan, computation with conjugate priors

- Bayesian shrinkage: the posterior is always "between" the likelihood and the prior
- In the normal prior / normal likelihood model, the posterior mean is exactly a weighted average of the prior mean and the data mean:

$$\mu_i^{post} = \left(\frac{\tau_0^2}{\tau_0^2 + \sigma^2/n_i}\right)\bar{y}_i + \left(\frac{\sigma^2/n_i}{\tau_0^2 + \sigma^2/n_i}\right)\mu_0$$

- This is "really what is going on" with shrinkage of random effects in multilevel models

Random simulation and MCMC as a way to “estimate” a model, CI’s

- If we do not have formulas for the posterior, we can usually simulate
- *For one-parameter problems* we can usually cook up a simulation using
  - Inverse CDF sampling (we talked about)
  - Rejection sampling & other methods (we didn’t cover)
- Estimating a CI from the simulated output:

$(\overline{\theta}_{sim} - 2SE_{sim},\ \overline{\theta}_{sim} + 2SE_{sim})$ or $(\theta^{sim}_{0.025},\ \theta^{sim}_{0.975})$

Random simulation and MCMC as a way to “estimate” a model, CI’s

- Bayesian models with many parameters can often be separated into “levels”:

  (posterior) $\propto$ (level 1)$\times$(level 2)$\times$(level 3)…
- Simulation for many parameters can be carried out with Markov Chain Monte Carlo (MCMC):
  - Compute the “complete conditional densities”, e.g.

$$\theta_3 \sim f(\theta_3 | \theta_1, \theta_2, \theta_4, \ldots, \theta_K)$$

  - Sample successively from each complete conditional
  - *(This is what JAGS does; we did not do this by hand!)*

Random simulation and MCMC as a way to "estimate" a model, CI's

- Throw away first part of MCMC sample as "burn-in"
- Look at autocorrelation plot, and time series ("random walk") plot to see that there are no serious problems
- Check R-hat for "convergence" to posterior distribution
- Use the MCMC sample to calculate CI's, etc., just as with other simulation samples.

Fitting and interpreting models using JAGS and rube()

- *lmer:* lmer(y ~ 1 + (1 | group)
- *JAGS:*

```
model {
        for (i in 1:n) {
                y[i] ~ dnorm(mu[i],sig2inv)
                mu[i] <- b0 + a0[group[i]]
        }
        for (j in 1:n.groups) {
                a0[j] ~ dnorm(0,tau2inv)
        }
        b0 ~ dnorm(0,0.000001)
        tau2inv <- pow(tau,-2);  tau ~ dunif(0,100)
        sig2inv <- pow(sig,-2);  sig ~ dunif(0,100)
}
```

lmer() requires only a general description of structure

- You describe the level 1, level 2, level 3 parts of the model;
- JAGS applies the slogan and figures out the complete conditionals;
- JAGS produces MCMC sample

Use p3() to inspect time series plot, auto-correlation plot, R-hat

Using JAGS for simulation tests

```
model {
        for (i in 1:n) {
                y[i] ~ dnorm(mu[i],sig2inv)
                mu[i] <- b0 + a0[group[i]]
        }
        for (j in 1:n.groups) {
                a0[j] ~ dnorm(0,tau2inv)
        }
        b0 ~ dnorm(0,0.000001)
        tau2inv <- pow(tau,-2);  tau ~ dunif(0,1000)
        sig2inv <- pow(sig,-2);  sig ~ dunif(0,1000)

        for (i in 1:n) {
                ynew[i] ~ dnorm(mu[i],sig2inv)
        }

  }
```

In the JAGS output, ynew will be a n.iter x n matrix.

Each row is a replication of the full data set.

Use to construct H_0's for simulation checks of the model.

Summary

- Final exam will cover material from week 06 through week 14 of the course:
  - Multi-level models
  - Model-checking using AIC/BIC/DIC, residuals, and/or simulation
  - Basic facts about MLE's, Fisher Information & SE's, CI's, etc.
  - Bayesian statistics, the slogan, computation with conjugate priors
  - Random simulation and MCMC as a way to "estimate" a model, CI's
  - Fitting and interpreting models using JAGS and rube()
  - Using JAGS for simulation tests
- Reading:
  - Class notes and other materials I've handed out (hw, hw sols…)
  - Selected topics from G&H Ch's 11-15 and Lynch Ch's 2, 3, 4 & 9
    - G&H Ch's 16-18, 20 and 24 have material similar to what we covered in class, as well, though I didn't really assign them as reading.