

Bayesian hierarchical modeling

Jingchen (Monika) Hu

Vassar College

MATH 347 Bayesian Statistics

Outline

- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

Outline

- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

Korean Drama Ratings

- K-drama popularity in Asia and other parts of the world, “Hallyu” = “Korean wave”.
- K-dramas streaming in the US: Hulu, DramaFever, and Netflix.

Korean Drama Ratings

- K-drama popularity in Asia and other parts of the world, “Hallyu” = “Korean wave”.
- K-dramas streaming in the US: Hulu, DramaFever, and Netflix.
- How are K-dramas being rated in Korea?
- There are three main producers/companies, and each drama is broadcasted in one of four different times of the week.
- The drama ratings are collected from the AGB Nielsen Media Research Group (the national AGB TV ratings).
- In one study (a previous student's course project), ratings of 101 K-dramas were collected from 2014 to 2016.

Korean Drama Ratings cont'd

Let's focus on $n = 33$ KBS dramas (KBS is one of the 3 producers, "Korea Broadcasting System"):

Drama Title	Schedule	Rating	Date
All About My Mom	4	0.2750	8/15/15
Assembly	2	0.0531	7/15/15
Blood	1	0.0474	2/16/15
Cheer up!	1	0.0341	10/5/15
Descendants of the Sun	2	0.2858	2/24/16
Discovery of Love	1	0.0703	8/18/14
Five Enough	4	0.2720	2/2/16
Golden Cross	2	0.0800	4/9/14
Gunman in Joseon	2	0.1070	6/25/14

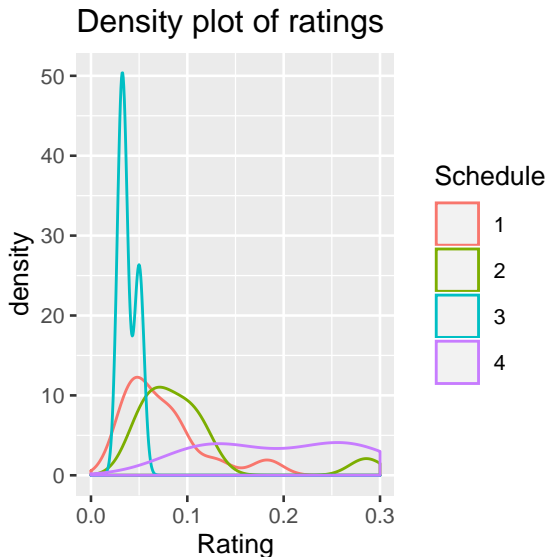
Table 1: The drama title, the broadcasting schedule, the rating, and the broadcasting date of 9 KBS dramas. The labeling of schedule is: 1 = Mondays and Tuesdays, 2 = Wednesdays and Thursdays, 3 = Fridays, and 4 = Saturdays and Sundays.

Ratings by Schedule

```
dramadata = read.csv("KDramaData.csv", header=T)

KBSdrama = dramadata[dramadata$Producer==2,]
KBSdrama$Schedule = as.factor(KBSdrama$Schedule)
```

Ratings by Schedule cont'd



Ratings by Schedule cont'd

```
table(KBSdrama$Schedule)
tapply(KBSdrama$Rating, KBSdrama$Schedule, summary)
tapply(KBSdrama$Rating, KBSdrama$Schedule, sd)
```

Schedule	Min	Mean	Max	sd	sample size
1 (M&Tu)	0.0331	0.0740	0.1829	0.0426	13
2 (W&Th)	0.0483	0.0998	0.2858	0.0661	11
3 (F)	0.0312	0.0384	0.0500	0.0101	3
4 (Sa&Su)	0.112	0.195	0.275	0.0757	6

- It seems to make sense to develop 4 schedule-specific Normal models, and label data as Y_{ij} .

Outline

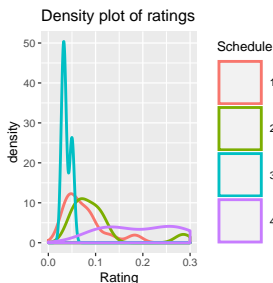
- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

Observations in Groups

- Observations can arise in groups. Examples:
 - ▶ Students' test scores from multiple schools
 - ▶ Ratings of movies of different genres
 - ▶ Ratings of dramas of different schedules
- A random variable can be labeled as

$$Y_{ij}, \quad (1)$$

where j is a group index, and i is the observation index within group j , and $i = 1, \dots, n_j$.



Ratings by Schedule

- Outcome is continuous, use a Normal model.
- Approach #1: separate estimates, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$

$$Y_{ij} \overset{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j) \quad (2)$$

- Approach #2: combined estimates, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$

$$Y_{ij} \overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \quad (3)$$

- Comments on pros and cons of each approach?

Ratings by Schedule

- Outcome is continuous, use a Normal model.
- Approach #1: separate estimates, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j) \quad (2)$$

- Approach #2: combined estimates, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \quad (3)$$

- Comments on pros and cons of each approach?
 - ▶ Approach #1: no connection between groups; groups with small sample size might suffer (extreme case: $n_j = 1$).
 - ▶ Approach #2: differences between groups are ignored.

Ratings by Schedule

- Outcome is continuous, use a Normal model.
- Approach #1: separate estimates, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$

$y_{\{ij\}}$ is within group j , so
we use i.i.d.;
if across groups, only ind

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j) \quad (2)$$

- Approach #2: combined estimates, for $i = 1, \dots, n_j$ and $j = 1, \dots, J$

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \quad (3)$$

- Comments on pros and cons of each approach?
 - ▶ Approach #1: no connection between groups; groups with small sample size might suffer (extreme case: $n_j = 1$).
 - ▶ Approach #2: differences between groups are ignored.
- Something in between? Hierarchical modeling!

Outline

- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

Review of Normal Model Inference

- Both parameters μ and σ are unknown.
- The sampling density:

$$Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \quad (4)$$

Review of Normal Model Inference

- Both parameters μ and σ are unknown.
- The sampling density:

$$Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \quad (4)$$

- The prior distributions $\pi(\mu, \sigma) = \pi_1(\mu)\pi_2(\sigma)$ take the form:

$$\mu \sim \text{Normal}(\mu_0, \sigma_0), \quad 1/\sigma^2 \sim \text{Gamma}(\alpha, \beta). \quad (5)$$

Review of Normal Model Inference

- Both parameters μ and σ are unknown.
- The sampling density:

$$Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Normal}(\mu, \sigma) \quad (4)$$

- The prior distributions $\pi(\mu, \sigma) = \pi_1(\mu)\pi_2(\sigma)$ take the form:

$$\mu \sim \text{Normal}(\mu_0, \sigma_0), \quad 1/\sigma^2 \sim \text{Gamma}(\alpha, \beta). \quad (5)$$

- Bayes' rule will produce two **conditional posterior distributions**:

$$\mu \mid y_1, \dots, y_n, \phi \sim \text{Normal} \left(\frac{\phi_0 \mu_0 + n \bar{y} \phi}{\phi_0 + n \phi}, \sqrt{\frac{1}{\phi_0 + n \phi}} \right), \quad (6)$$

$$1/\sigma^2 = \phi \mid y_1, \dots, y_n, \mu \sim \text{Gamma} \left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2 \right). \quad (7)$$

Group-specific Normal models

- A group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j), \quad (8)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

Group-specific Normal models

- A group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j), \quad (8)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- For separate estimates:
 - ① Put a Normal prior for μ_j and a Gamma prior for σ_j (independent priors).
 - ② Use Gibbs sampler to generate posterior samples of μ_j and σ_j (your own Gibbs sampler or JAGS).
 - ③ Perform MCMC diagnostics.
 - ④ Summarize the results.
- How can we link μ_j 's and σ_j 's in some way, if we believe they are related?

Group-specific Normal Models with Shared σ

- Without loss of generality, assume a group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (9)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j . **A commonly shared σ .** Reasonable?

Group-specific Normal Models with Shared σ

- Without loss of generality, assume a group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (9)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j . **A commonly shared σ .** Reasonable?

Schedule	Min	Mean	Max	sd	sample size
1 (M&Tu)	0.0331	0.0740	0.1829	0.0426	13
2 (W&Th)	0.0483	0.0998	0.2858	0.0661	11
3 (F)	0.0312	0.0384	0.0500	0.0101	3
4 (Sa&Su)	0.112	0.195	0.275	0.0757	6

Group-specific Normal Models with Shared σ

- Without loss of generality, assume a group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (9)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j . **A commonly shared σ .** Reasonable?

Schedule	Min	Mean	Max	sd	sample size
1 (M&Tu)	0.0331	0.0740	0.1829	0.0426	13
2 (W&Th)	0.0483	0.0998	0.2858	0.0661	11
3 (F)	0.0312	0.0384	0.0500	0.0101	3
4 (Sa&Su)	0.112	0.195	0.275	0.0757	6

Probably not... Schedule specific μ_j and σ_j as an exercise.

Group-specific Normal Models with Shared σ cont'd

- Without loss of generality, assume a group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (10)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j . **A commonly shared σ .**

- Specify a shared Normal prior distribution for μ_j (conjugacy):

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (11)$$

- What does this prior mean?

Group-specific Normal Models with Shared σ cont'd

- Without loss of generality, assume a group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (10)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j . **A commonly shared σ .**

- Specify a shared Normal prior distribution for μ_j (conjugacy):

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (11)$$

- What does this prior mean?

use small tau if you believe μ_j should be similar

- Assume all μ_j 's come from the same prior distribution, therefore they are related, and they are not the same.
- μ is the mean (controls the location), and τ is the standard deviation (controls the spread; large τ means large variability among the μ_j 's).
- μ and τ are called hyperparameters, i.e. parameters of distributions of parameters.

The Hierarchical Normal Model

- The sampling density for group j , and $j = 1, \dots, J$:

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (12)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- The stage 1 prior distribution for μ_j :

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (13)$$

- The stage 2 prior distribution for μ, τ :

$$\mu, \tau \sim g(\mu, \tau). \quad (14)$$

The Hierarchical Normal Model

- The sampling density for group j , and $j = 1, \dots, J$:

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma), \quad (12)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- The stage 1 prior distribution for μ_j :

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (13)$$

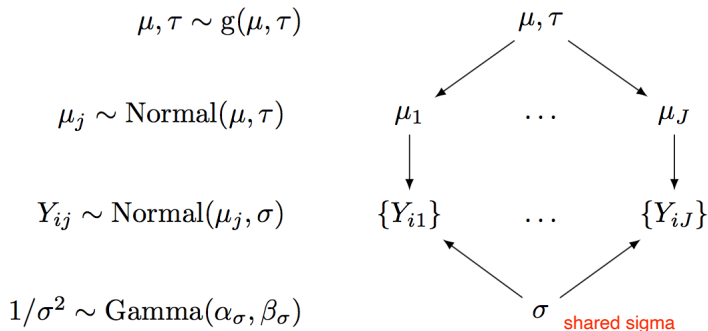
- The stage 2 prior distribution for μ_j :

$$\mu, \tau \sim g(\mu, \tau). \quad (14)$$

- The prior distribution for σ :

$$1/\sigma^2 \sim \text{Gamma}(\alpha_\sigma, \beta_\sigma). \quad (15)$$

Graphical Representation of the Hierarchical Model



- Arrows pointing
 - from parameters to random variable
 - from hyperparameters to parameters

Prior and Hyperprior Specifications

- The stage 1 prior distribution for μ_j :

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (16)$$

- The stage 2 prior distribution for μ_j :

$$\mu, \tau \sim g(\mu, \tau). \quad (17)$$

- Hyperpriors:

Prior and Hyperprior Specifications

- The stage 1 prior distribution for μ_j :

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (16)$$

- The stage 2 prior distribution for μ_j :

$$\mu, \tau \sim g(\mu, \tau). \quad (17)$$

- Hyperpriors:

$$\mu \mid \mu_0, \gamma_0 \sim \text{Normal}(\mu_0, \gamma_0), \quad (18)$$

$$1/\tau^2 \mid \alpha_\tau, \beta_\tau \sim \text{Gamma}(\alpha_\tau, \beta_\tau). \quad (19)$$

Prior and Hyperprior Specifications

- The stage 1 prior distribution for μ_j :

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (16)$$

- The stage 2 prior distribution for μ_j :

$$\mu, \tau \sim g(\mu, \tau). \quad (17)$$

- Hyperpriors:

$$\mu \mid \mu_0, \gamma_0 \sim \text{Normal}(\mu_0, \gamma_0), \quad (18)$$

$$1/\tau^2 \mid \alpha_\tau, \beta_\tau \sim \text{Gamma}(\alpha_\tau, \beta_\tau). \quad (19)$$

- The prior distribution for σ :

$$1/\sigma^2 \sim \text{Gamma}(\alpha_\sigma, \beta_\sigma). \quad (20)$$

Outline

- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS**
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

Recap: Prior and Hyperprior Specifications

- The stage 1 prior distribution for μ_j :

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (21)$$

- The stage 2 prior distribution for μ_j :

$$\mu, \tau \sim g(\mu, \tau). \quad (22)$$

- Hyperpriors:

$$\mu \mid \mu_0, \gamma_0 \sim \text{Normal}(0.1, 0.5), \quad (23)$$

$$1/\tau^2 \mid \alpha_\tau, \beta_\tau \sim \text{Gamma}(1, 1). \quad (24)$$

- The prior distribution for σ :

$$1/\sigma^2 \sim \text{Gamma}(1, 1). \quad (25)$$

JAGS Script for the Hierarchical Model

```

modelString <-"
model {
  ## likelihood
  for (i in 1:N){
    y[i] ~ dnorm(mu_j[schedule[i]], invsigma2)
  }

  ## priors
  for (j in 1:J){
    mu_j[j] ~ dnorm(mu, invtau2)
  }
  invsigma2 ~ dgamma(a_g, b_g)
  sigma <- sqrt(pow(invsigma2, -1))  recover sigma

  ## hyperpriors
  mu ~ dnorm(mu0, 1/g0^2)
  invtau2 ~ dgamma(a_t, b_t)
  tau <- sqrt(pow(invttau2, -1))  recover tau
}
"

```

JAGS Script for the Hierarchical Model cont'd

- Notes about the `modelString`
 - ① Need a vector of `mu_j`, of length `J`.
 - ② Need a vector of `schedule`, of length `N`. to avoid nested loop
 - ③ `dnorm` takes mean and **precision**. $\text{phi} = 1/\text{sig}$
 - ④ Work with `invsigma2`, can return `sigma`.
 - ⑤ Work with `invtau2`, can return `tau`.

use `~` in JAGS for distribution; use `<-` in JAGS for deterministic

JAGS Script for the Hierarchical Model cont'd

- Pass the data and hyperparameter values to JAGS:

```

y = KBSdrama$Rating
schedule = KBSdrama$Schedule
N = length(y)
J = length(unique(schedule))

initsfunction <- function(chain){
  .RNG.seed <- c(1,2)[chain]
  .RNG.name <- c("base::Super-Duper",
                 "base::Wichmann-Hill")[chain]
  return(list(.RNG.seed=.RNG.seed,
              .RNG.name=.RNG.name))
}

the_data <- list("y" = y, "schedule" = schedule, "N" = N, "J" = J,
                 "mu0" = 0.1, "g0" = 0.5,
                 "a_t" = 1, "b_t" = 1,
                 "a_g" = 1, "b_g" = 1)

```

JAGS Script for the Hierarchical Model cont'd

- Run the JAGS code for this model:

```
posterior <- run.jags(modelString,  
                      n.chains = 1,  
                      data = the_data,  
                      monitor = c("mu", "tau", "mu_j", "sigma"),  
                      adapt = 1000,  
                      burnin = 5000,  
                      sample = 5000,  
                      thin = 1,  
                      inits = initsfunction)
```

JAGS Output of the Hierarchical Model

- Obtain posterior summaries of all parameters:

```
summary(posterior)
```

##	Lower95	Median	Upper95	Mean	SD	Mode
## mu	-0.4904750	0.10801400	0.668230	0.10472246	0.28838371	NA
## tau	0.3526710	0.65852600	1.249940	0.72061723	0.27488876	NA
## mu_j[1]	-0.0719655	0.07132615	0.216751	0.07266281	0.07238920	NA
## mu_j[2]	-0.0568824	0.09944590	0.253260	0.09928626	0.07995014	NA
## mu_j[3]	-0.2414660	0.04479395	0.349108	0.04271370	0.15109887	NA
## mu_j[4]	-0.0248182	0.19140950	0.399113	0.19242820	0.10745634	NA
## sigma	0.2010700	0.26158400	0.332643	0.26503794	0.03460243	NA

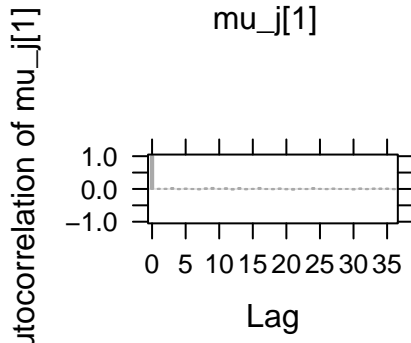
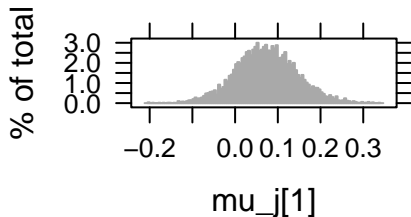
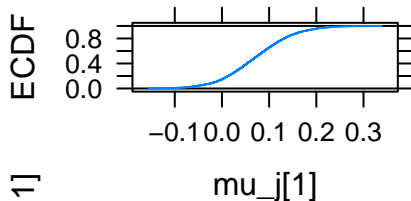
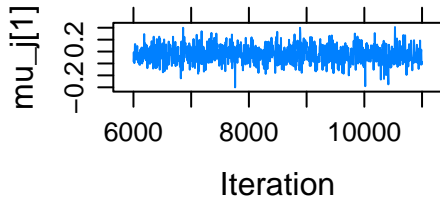
##	MCerr	MC%ofSD	SSeff	AC.10	psrf
## mu	0.0041809869	1.4	4758	-0.010518447	NA
## tau	0.0041977729	1.5	4288	-0.015113474	NA
## mu_j[1]	0.0010237378	1.4	5000	0.002230454	NA
## mu_j[2]	0.0011306657	1.4	5000	0.018813934	NA
## mu_j[3]	0.0020734838	1.4	5310	-0.007974236	NA
## mu_j[4]	0.0015196622	1.4	5000	-0.017268736	NA
## sigma	0.0005535074	1.6	3908	0.009879833	NA

negative autocorr

JAGS Output of the Hierarchical Model cont'd

```
plot(posterior, vars = "mu_j[1]")
```

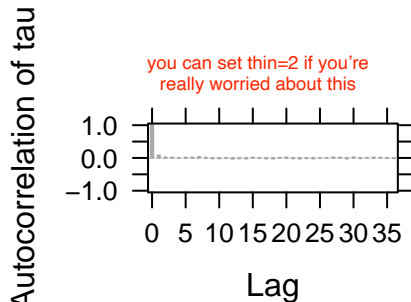
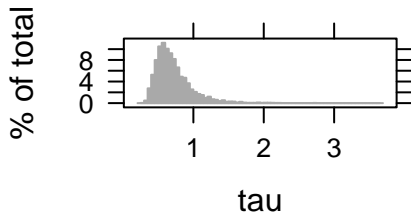
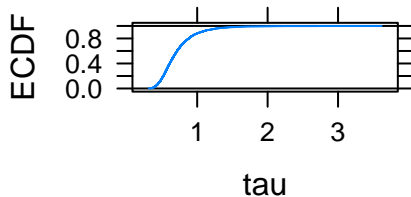
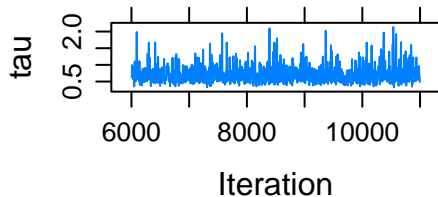
```
## Generating plots...
```



JAGS Output of the Hierarchical Model cont'd

```
plot(posterior, vars = "tau")
```

```
## Generating plots...
```



Shrinkage/Pooling Effects

```

Ind_Stats = as.data.frame(matrix(NA, J, 2))
names(Ind_Stats) = c("mean", "sd")
for (j in 1:J){
  Ind_Stats[j, ] = c(mean(KBSdrama$Rating[KBSdrama$Schedule == j]),
                     sd(KBSdrama$Rating[KBSdrama$Schedule == j]))
}

Post_Means <- summary(posterior)[, 4]

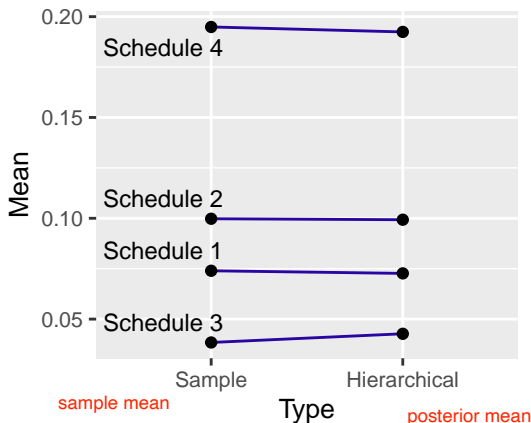
Means1 <- data.frame(Type = "Sample", Mean = Ind_Stats$mean)
Means2 <- data.frame(Type = "Hierarchical", Mean =
                     Post_Means[3:(4 + J - 2)])

Means1$Title <- c("Schedule 1", "Schedule 2", "Schedule 3",
                  "Schedule 4")
Means2$Title <- c("Schedule 1", "Schedule 2", "Schedule 3",
                  "Schedule 4")

```

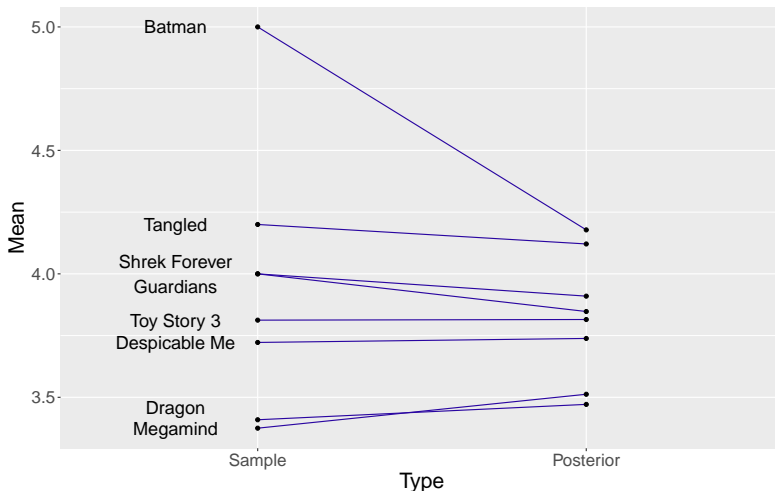
Shrinkage/Pooling Effects cont'd

```
ggplot(rbind(Means1, Means2), aes(Type, Mean, group=Title)) +
  geom_line(color = crcblue) + geom_point() +
  annotate(geom = "text", x = 0.75,
    y = Means1$Mean + c(0.01, 0.01, 0.01, -0.01),
    size = 3, label = Means1$Title) + increasefont(Size = 10)
```



Shrinkage/Pooling Effects cont'd

An example of movie ratings:



Sources of Variability

- Two sources of variability in Y_{ij} :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma) \text{ [within-group variability]} \quad (26)$$

$$\mu_j \mid \mu, \tau \sim \text{Normal}(\mu, \tau) \text{ [between-group variability]} \quad (27)$$

Sources of Variability

- Two sources of variability in Y_{ij} :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma) \text{ [within-group variability]} \quad (26)$$

$$\mu_j \mid \mu, \tau \sim \text{Normal}(\mu, \tau) \text{ [between-group variability]} \quad (27)$$

- To compare these two sources of variability, one can compute the fraction

$$R = \frac{\tau^2}{\tau^2 + \sigma^2}, \quad (28)$$

from the posterior draws of τ and σ .

Sources of Variability

- Two sources of variability in Y_{ij} :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma) \quad [\text{within-group variability}] \quad (26)$$

$$\mu_j \mid \mu, \tau \sim \text{Normal}(\mu, \tau) \quad [\text{between-group variability}] \quad (27)$$

- To compare these two sources of variability, one can compute the fraction

$$R = \frac{\tau^2}{\tau^2 + \sigma^2}, \quad (28)$$

from the posterior draws of τ and σ .

- The closer the value of R to 1, the higher the between-group variability.

Compute and Graph Sources of Variability

- We need the coda R package

```
install.packages("coda")
```

```
require(coda)
```

```
tau_draws <- as.mcmc(posterior, vars = "tau")
```

```
sigma_draws <- as.mcmc(posterior, vars = "sigma")
```

```
R <- tau_draws^2/(tau_draws^2 + sigma_draws^2)
```

```
df <- as.data.frame(R)
```

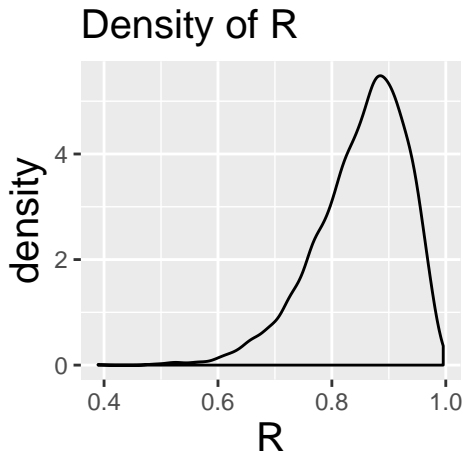
```
quantile(R, c(0.025, 0.975))
```

```
##          2.5%          97.5%
```

```
## 0.6620552 0.9687804
```

Compute and Graph Sources of Variability cont'd

```
ggplot(df, aes(x=R)) + geom_density() +  
  labs(title="Density of R") +  
  theme(plot.title = element_text(size=15)) +  
  theme(axis.title = element_text(size=15))
```



Outline

- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

A More Flexible Hierarchical Model

- A group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j), \quad (29)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

A More Flexible Hierarchical Model

- A group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j), \quad (29)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- This is probably a more suitable hierarchical model for the K-drama ratings dataset.

Schedule	Min	Mean	Max	sd	sample size
1 (M&Tu)	0.0331	0.0740	0.1829	0.0426	13
2 (W&Th)	0.0483	0.0998	0.2858	0.0661	11
3 (F)	0.0312	0.0384	0.0500	0.0101	3
4 (Sa&Su)	0.112	0.195	0.275	0.0757	6

A More Flexible Hierarchical Model cont'd

- A group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j), \quad (30)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- Specify a shared Normal prior distribution for μ_j (conjugacy):

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (31)$$

- In addition, specify a shared Gamma prior distribution for σ_j (conjugacy):

A More Flexible Hierarchical Model cont'd

- A group-specific Normal model for group j :

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j), \quad (30)$$

where $i = 1, \dots, n_j$ and n_j is the number of observations in group j .

- Specify a shared Normal prior distribution for μ_j (conjugacy):

$$\mu_j \sim \text{Normal}(\mu, \tau). \quad (31)$$

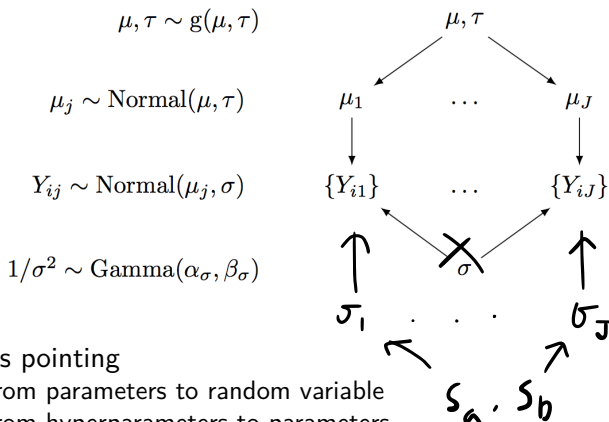
- In addition, specify a shared Gamma prior distribution for σ_j (conjugacy):

$$1/\sigma_j^2 \sim \text{Gamma}(s_a, s_b). \quad (32)$$

- What do these two prior mean?

A More Flexible Hierarchical Model cont'd

- Previously we have a graphical representation of the hierarchical model (group-specific μ_j and shared σ):



- Arrows pointing
 - from parameters to random variable
 - from hyperparameters to parameters

A More Flexible Hierarchical Model cont'd

The complete specification model specification:

- The sampling density for group j , and $j = 1, \dots, J$:

$$Y_{ij} \stackrel{i.i.d.}{\sim} \text{Normal}(\mu_j, \sigma_j). \quad (32)$$

- The two-stage prior distribution for μ_j :

- The two-stage prior distribution for σ_j :

A More Flexible Hierarchical Model cont'd

Create JAGS script for this more flexible hierarchical model. Previously...

```
modelString <-"
model {
  ## likelihood
  for (i in 1:N){
    y[i] ~ dnorm(mu_j[schedule[i]], invsigma2)
  }
  ## priors
  for (j in 1:J){
    mu_j[j] ~ dnorm(mu, invtau2)
  }
  invsigma2 ~ dgamma(a_g, b_g)
  sigma <- sqrt(pow(invsigma2, -1))
  ## hyperpriors
  mu ~ dnorm(mu0, 1/g0^2)
  invtau2 ~ dgamma(a_t, b_t)
  tau <- sqrt(pow(invttau2, -1))
}
```

need to adjust appropriately

Outline

- 1 Example: Korean Drama Ratings
- 2 Observations in groups: approaches to modeling
- 3 A two-stage prior in a hierarchical model
- 4 MCMC simulation by JAGS
- 5 Exercise: Hierarchical model with schedule-specific μ_j and σ_j
- 6 Recap

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of unknown parameter(s) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about unknown parameter(s) given the information from the sample (posterior).

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of unknown parameter(s) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about unknown parameter(s) given the information from the sample (posterior).
- For continuous outcome in groups, consider a hierarchical Normal model.
- A two-stage prior for μ . Also possible to do a two-stage prior for σ .

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of unknown parameter(s) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about unknown parameter(s) given the information from the sample (posterior).
- For continuous outcome in groups, consider a hierarchical Normal model.
- A two-stage prior for μ . Also possible to do a two-stage prior for σ .
- Use JAGS for MCMC and make sure to perform MCMC diagnostics.

Recap

- Bayesian inference procedure:
 - ▶ Step 1: express an opinion about the location of unknown parameter(s) before sampling (prior).
 - ▶ Step 2: take the sample (data/likelihood).
 - ▶ Step 3: use Bayes' rule to sharpen and update the previous opinion about unknown parameter(s) given the information from the sample (posterior).
- For continuous outcome in groups, consider a hierarchical Normal model.
- A two-stage prior for μ . Also possible to do a two-stage prior for σ .
- Use JAGS for MCMC and make sure to perform MCMC diagnostics.
- Additional topics:
 - 1 Shrinkage/pooling effects.
 - 2 Sources of variability.