STAT 431 — Applied Bayesian Analysis — Course Notes

Random Effects and Hierarchical Models

Fall 2022

Recall that models may have more than one "level" of unobserved random quantities, e.g., the airliner fatalities example, with hyperparameters (having hyperpriors) defining the distribution of fatal accident rate parameters of different airliners.

Models with more than one level are often called **hierarchical**.

Exchangeability

Random variables $Y_1, \ldots, Y_n$ are **exchangeable** if any (non-random) permutation of their indices results in the same joint distribution.

Eg: Y_1, Y_2 are exchangeable if (Y_1, Y_2) has the same distribution as (Y_2, Y_1)

Note: Exchangeable random variables all have the same marginal distribution.

Fact: $Y_1, \ldots, Y_n$ are exchangeable if there exists a random variable (or vector) V such that

$$Y_1, \ldots, Y_n \mid V \quad \sim \quad iid$$

that is, if they are conditionally independent and identically distributed, given V.

Note: $Y_1, \ldots, Y_n$ need not be *unconditionally* (marginally) independent.

Practical implication:

If we believe some random variables in a model are exchangeable, we should try to model them as iid, conditional on some *latent* (unobserved) variable V.

We would then have to choose the nature of the dependence on V and the distribution of V.

The distribution of V could depend on parameters. We could give priors to those parameters, creating a hierarchical model.

Example: Dye Yield

$$Y_{ij} \;=\; \text{yield of dye in } j\text{th preparation made from } i\text{th batch of raw material}$$

Goal: Distinguish within-batch variation from between-batch variation.

The classical *variance-components model* is

$$Y_{ij} \;=\; \mu + \tilde{\alpha}_i + \varepsilon_{ij}$$

where $\tilde{\alpha}_i$s are **random effects** and ε_{ij}s are errors:

$$\left.\begin{array}{rcl} \tilde{\alpha}_i \mid \sigma_B^2 & \sim & iid\ \mathrm{Normal}(0, \sigma_B^2) \\ \varepsilon_{ij} \mid \sigma_W^2 & \sim & iid\ \mathrm{Normal}(0, \sigma_W^2) \end{array}\right\} \text{ conditionally indep.}$$

Hierarchical representation:

Let

$$\alpha_i \;=\; \mu + \tilde{\alpha}_i$$

Then

$$\begin{aligned} Y_{ij} \mid \alpha_i, \sigma_W^2 &\sim \textit{indep}\ \mathrm{Normal}(\alpha_i, \sigma_W^2) \\ \alpha_i \mid \mu, \sigma_B^2 &\sim \textit{iid}\ \mathrm{Normal}(\mu, \sigma_B^2) \end{aligned}$$

Note:

- For each i, the Y_{ij}s are exchangeable. (Why?)
- The α_is are exchangeable.

[Graph densities in hierarchy ...]

For priors, we must be careful.

We will take

$$\left.\begin{array}{rcl} \mu & \sim & \mathrm{Normal}(0, 1000000) \\[1ex] \sigma_W^2 & \sim & \mathrm{InvGamma}(0.001, 0.001) \\[1ex] \sigma_B & \sim & \mathrm{Exponential}(0.001) \end{array}\right\} \text{independent}$$

where $\sigma_B = \sqrt{\sigma_B^2}$.

The prior on σ_B is related to *penalized complexity priors* (see BSM, Sec. 2.3.5, if interested).

Why not just take a semi-conjugate prior

$$\sigma_B^2 \sim \mathrm{InvGamma}(0.001, 0.001)$$

similarly to the recommendation in BSM, Sec. 4.4?

This would approximate the improper prior

$$\pi(\sigma_B^2) \propto \frac{1}{\sigma_B^2} \qquad (\sigma_B^2 > 0)$$

but it turns out that this would lead to an improper posterior!

Since JAGS prefers precisions to variances, define

$$\tau_W^2 = 1/\sigma_W^2$$

We might also be interested in the *intra-class correlation*:

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

This is the (frequentist) correlation between responses from samples from the same batch (same i).

[Draw model graph ...]

```
data {
  dimy <- dim(y)
  batches <- dimy[1]
  samples <- dimy[2]
}

model {
  for (i in 1:batches) {

    for (j in 1:samples) {
      y[i,j] ~ dnorm(alpha[i], tausqW)
    }

    alpha[i] ~ dnorm(mu, 1/sigmasqB)
  }

  mu ~ dnorm(0, 0.000001)

  tausqW ~ dgamma(0.001, 0.001)
  sigmaB ~ dexp(0.001)

  sigmasqW <- 1 / tausqW
  sigmasqB <- sigmaB^2

  rho <- sigmasqB / (sigmasqB + sigmasqW)
}
```

R/JAGS Example 4.2:

Normal Random-Effects Model

Notes:

- The `dim` function is allowed only in the `data` block.
- We can specify initial values for top-level nodes (no incoming arrows), and then let the rest be auto-generated.

Hierarchical Normal Regression

Consider a one-predictor regression (Y versus X), but suppose that, in addition to Y and X, there is a grouping variable.

Let

$$\begin{aligned} Y_{ij} &= \text{response of } j\text{th observation in group } i \\ X_{ij} &= \text{its predictor value} \end{aligned}$$

Let

$$\bar{X} = \text{average of } \textit{all } X_{ij} \text{ values}$$

(We will use the same covariate centering for all groups.)

Each group can have its own regression line:

$$
\begin{aligned}
Y_{ij} &= \alpha_{i1} + \alpha_{i2}(X_{ij} - \bar{X}) + \varepsilon_{ij} \\
\varepsilon_{ij} &\sim iid \text{ Normal}(0, \sigma_y^2)
\end{aligned}
$$

The data model becomes

$$
Y_{ij} \mid \alpha_{i1}, \alpha_{i2}, \sigma_y^2 \quad \sim \quad indep \text{ Normal}\big(\alpha_{i1} + \alpha_{i2}(X_{ij} - \bar{X}),\, \sigma_y^2\big)
$$

A semi-conjugate prior for the variance:

$$\sigma_y^2 \sim \mathrm{InvGamma}(a_y, b_y)$$

We will assume it is independent of the other parameters.

Two potential prior formulations for α_{i1} and α_{i2}:

- Univariate: assumes α_{i1} and α_{i2} are (a priori) independent
- Bivariate: allows (conditional) prior correlations between α_{i1} and α_{i2}

Correlations between α_{i1} and α_{i2} are frequently encountered ...

[Illustrate with regression lines ...]

Univariate Formulation

$$\left.\begin{aligned} \alpha_{i1} \mid \beta_1, \sigma^2_{\alpha_1} &\sim \mathrm{Normal}(\beta_1, \sigma^2_{\alpha_1}) \\ \alpha_{i2} \mid \beta_2, \sigma^2_{\alpha_2} &\sim \mathrm{Normal}(\beta_2, \sigma^2_{\alpha_2}) \end{aligned}\right\} \text{ all conditionally independent}$$

$$\left.\begin{aligned} \beta_1 &\sim \mathrm{Normal}(\mu_1, \sigma^2_1) \\ \beta_2 &\sim \mathrm{Normal}(\mu_2, \sigma^2_2) \\ \sigma_{\alpha_1} &\sim \mathrm{Exponential}(b_{\alpha_1}) \\ \sigma_{\alpha_2} &\sim \mathrm{Exponential}(b_{\alpha_2}) \end{aligned}\right\} \text{ independent}$$

[Draw model graph ...]

Example: Baby Rat Weights

$$
\begin{aligned}
Y_{ij} &= \text{mass of rat } i \text{ (g?) at } j\text{th measurement} \\
X_{ij} &= \text{age of rat } i \text{ (days) at } j\text{th measurement}
\end{aligned}
$$

The measurements were synchronous:

$$X_{ij} = X_j \qquad (8,\ 15,\ 22,\ 29,\ \text{or } 36)$$

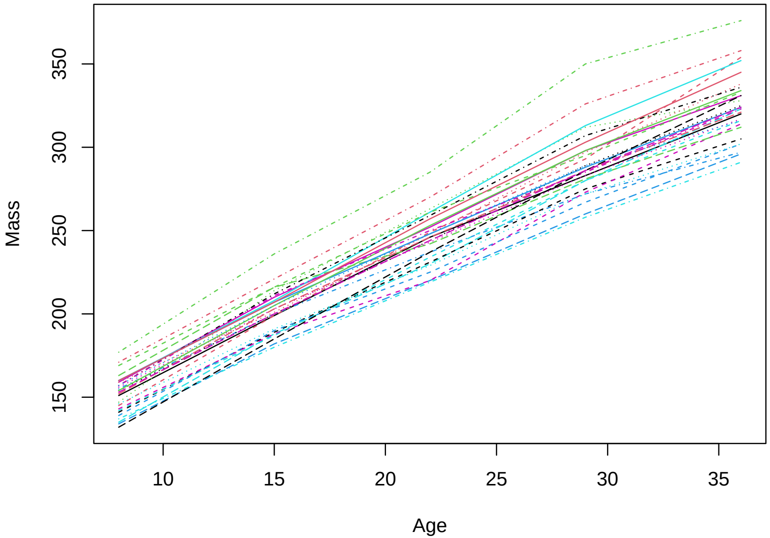
so $\bar{X} = 22$.

Rat mass values in file `ex4.3data.txt` (truncated):

```
151  199  246  283  320
145  199  249  293  354
147  214  263  312  328
155  200  237  272  297
135  188  230  280  323
159  210  252  298  331
...
```

Each row is a different rat, and each column is a different age.

Plot of the "growth curves":



We can create the necessary data objects in R:

```
ages <- c(8, 15, 22, 29, 36)

d <- list(Y = read.table("ex4.3data.txt"),
          X = ages,
          Xbar = mean(ages))
```

`Y` is a data frame, but it can be indexed like a matrix.

The JAGS code:

```
data {
  dim.Y <- dim(Y)
}

model {
  for(i in 1:dim.Y[1]) {

    for(j in 1:dim.Y[2]) {
      Y[i,j] ~ dnorm(mu[i,j], tausq.y)
      mu[i,j] <- alpha[i,1] + alpha[i,2] * (X[j] - Xbar)
    }

    alpha[i,1] ~ dnorm(beta1, 1 / sigma.alpha1^2)
    alpha[i,2] ~ dnorm(beta2, 1 / sigma.alpha2^2)
  }

  tausq.y ~ dgamma(0.001, 0.001)
  sigma.y <- 1 / sqrt(tausq.y)

  beta1 ~ dnorm(0.0, 1.0E-6)
  beta2 ~ dnorm(0.0, 1.0E-6)
  sigma.alpha1 ~ dexp(0.001)
  sigma.alpha2 ~ dexp(0.001)
}
```

R/JAGS Example 4.3:

Hierarchical Normal Regression: Univariate Formulation

Bivariate Formulation

$$\boldsymbol{\alpha}_i \;=\; \begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \end{bmatrix} \;\Bigg|\; \boldsymbol{\beta}, \boldsymbol{\Omega} \quad \sim \quad iid \;\; \mathrm{Normal}(\boldsymbol{\beta}, \boldsymbol{\Omega})$$

where

$$\boldsymbol{\beta} \;=\; \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \qquad\qquad \boldsymbol{\Omega} \;=\; \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{12} & \Omega_{22} \end{bmatrix}$$

But we probably want to let $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ be chosen by the data, rather than arbitrarily specified, so we add another prior level ...

A semi-conjugate hyperprior specification:

$$\left.\begin{array}{lcl} \boldsymbol{\beta} & \sim & \mathrm{Normal}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\[1em] \boldsymbol{\Omega} & \sim & \mathrm{InvWishart}(\nu,\, \nu\, \boldsymbol{\Omega}_0) \end{array}\right\} \text{ independent}$$

where

$$\boldsymbol{\mu}_0 \quad \text{is a } 2 \times 1 \text{ vector}$$

$$\boldsymbol{\Sigma}_0 \quad \text{and} \quad \boldsymbol{\Omega}_0 \quad \text{are } 2 \times 2 \text{ invertible covariance-type matrices}$$

$$\text{and} \quad \nu > 1 \quad \text{is a scalar.}$$

The $\mathrm{InvWishart}$ distribution generalizes the inverse gamma distribution to covariance matrices (see BSM, Sec. 2.1.7).

Remarks:

- Need $\nu > p - 1$ for the $p \times p$ $\mathrm{InvWishart}$ distribution to exist.

  This suggests $\nu = p$ might be a good choice — not quite "vague," but at least has relatively little information.

- If a matrix is $\mathrm{InvWishart}$, then its inverse has a **Wishart distribution**.

  JAGS provides the Wishart distribution, rather than $\mathrm{InvWishart}$, and with a different parameterization than in BSM.

[Draw model graph ...]

Example: Baby Rat Weights (continued)

As before, we define the data in R.

We also add to the data some objects to help specify the prior:

```
ages <- c(8, 15, 22, 29, 36)

d <- list(Y = read.table("ex4.4data.txt"),
          X = ages,
          Xbar = mean(ages),
          Omega0 = rbind(c(100, 0),
                         c(0, 0.1)),
          mu0 = c(0,0),
          Sigma0.inv = rbind(c(1.0E-6, 0),
                             c(0, 1.0E-6)))
```

(File `ex4.4data.txt` contains the mass data, as before.)

The JAGS code:

```
data {
  dim.Y <- dim(Y)
}

model {
  for(i in 1:dim.Y[1]) {

    for(j in 1:dim.Y[2]) {
      Y[i,j] ~ dnorm(mu[i,j], tausq.y)
      mu[i,j] <- alpha[i,1] + alpha[i,2] * (X[j] - Xbar)
    }

    alpha[i,1:2] ~ dmnorm(beta, Omega.inv)
  }

  tausq.y ~ dgamma(0.001, 0.001)
  sigma.y <- 1 / sqrt(tausq.y)

  beta ~ dmnorm(mu0, Sigma0.inv)
  Omega.inv ~ dwish(2*Omega0, 2)
  Omega <- inverse(Omega.inv)

  rho <- Omega[1,2] / sqrt(Omega[1,1] * Omega[2,2])
}
```

R/JAGS Example 4.4:

Hierarchical Normal Regression: Bivariate Formulation