STAT 431 — Applied Bayesian Analysis — Course Notes

Hierarchical Models

Spring 2019

Models may have more than one “level” of unobserved random quantities:

e.g. random effects (with priors on random effect parameters)

e.g. random **hyperparameters**: parameters of the priors that themselves have **hyperpriors**

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. (HW?)

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

$$Y_1, \ldots, Y_n \mid X \;\sim\; \text{i.i.d.}$$

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

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

Practical implication:

If we believe some random variables in a model are exchangeable, we should try to model them as $\mathrm{i.i.d.}$, conditional on some *latent* (unobserved) variable X.

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

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

(See Cowles, Sec. 9.1–9.4 for a complete example involving a softball player's batting average.)

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

$$Y_{ij} \;=\; \mu + \alpha_i + \varepsilon_{ij}$$

where α_is are the random effects and ε_{ij}s the errors:

$$\left.\begin{array}{rcl} \alpha_i \mid \sigma_B^2 & \sim & \text{i.i.d. } \mathrm{N}(0, \sigma_B^2) \\ \varepsilon_{ij} \mid \sigma_W^2 & \sim & \text{i.i.d. } \mathrm{N}(0, \sigma_W^2) \end{array}\right\} \text{ conditionally indep.}$$

Hierarchical representation:

Let

$$\mu_i \;=\; \mu + \alpha_i$$

Then

$$Y_{ij} \mid \mu_i, \sigma^2_W \;\sim\; \text{indep. } \mathrm{N}(\mu_i, \sigma^2_W)$$

$$\mu_i \mid \mu, \sigma^2_B \;\sim\; \text{i.i.d. } \mathrm{N}(\mu, \sigma^2_B)$$

Note:

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

[Graph densities in hierarchy ...]

For priors, we can take, for example,

$$\left.\begin{array}{rcl} \mu & \sim & \mathrm{N}(0, 1000000) \\ \sigma_W^2 & \sim & \mathrm{IG}(0.001, 0.001) \\ \sigma_B^2 & \sim & \mathrm{IG}(0.001, 0.001) \end{array}\right\} \text{independent}$$

which are reasonably “vague” but proper.

Since JAGS prefers precisions to variances, define

$$\tau_W^2 = 1/\sigma_W^2 \qquad \tau_B^2 = 1/\sigma_B^2$$

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 ...]

R/JAGS Example 9.1:

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.
- Even if it were possible to specify improper priors for the variance components (σ_B^2 and σ_W^2), we would not do so, because that would risk creating an improper posterior.

Example: Airliner Fatalities

Y_i = number of passenger-fatal events for airliners of type i

T_i = millions of flights (total) by airliners of type i

By the Poisson approximation to the binomial,

$$Y_i \mid \lambda_i \quad \sim \quad \text{indep. Poisson}(\lambda_i T_i)$$

where

λ_i = passenger-fatal event rate per million flights for airliners of type i

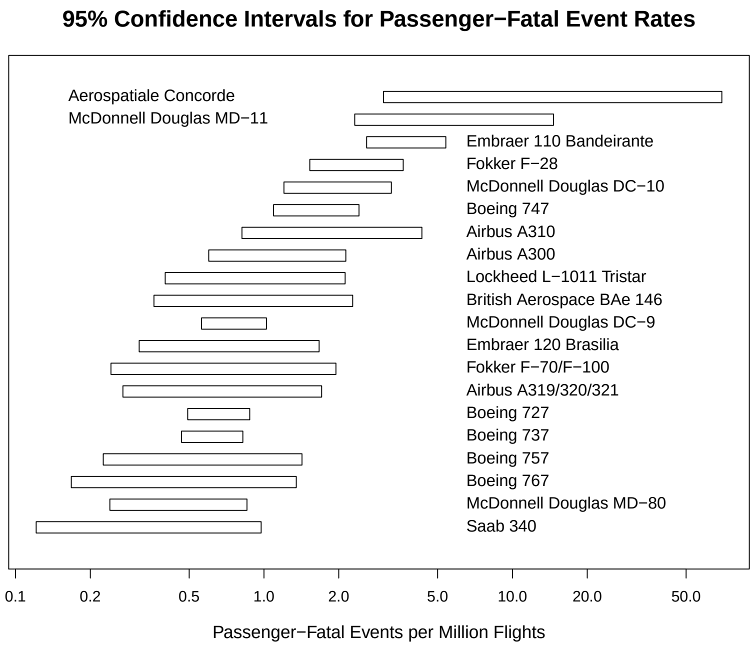
Since

$$\mathrm{E}(Y_i \mid \lambda_i) = \lambda_i T_i$$

a natural unbiased frequentist estimator is

$$\hat{\lambda}_i = \frac{Y_i}{T_i}$$

It is also possible to form individual 95% confidence intervals for the λ_is ...


95% Confidence Intervals for Passenger−Fatal Event Rates
Aerospatiale Concorde
McDonnell Douglas MD−11
Embraer 110 Bandeirante
Fokker F−28
McDonnell Douglas DC−10
Boeing 747
Airbus A310
Airbus A300
Lockheed L−1011 Tristar
British Aerospace BAe 146
McDonnell Douglas DC−9
Embraer 120 Brasilia
Fokker F−70/F−100
Airbus A319/320/321
Boeing 727
Boeing 737
Boeing 757
Boeing 767
McDonnell Douglas MD−80
Saab 340
0.1
0.2
0.5
1.0
2.0
5.0
10.0
20.0
50.0
Passenger−Fatal Events per Million Flights

Some possible research questions:

- What is the "average" passenger-fatal event rate, after taking uncertainties into account?
- How would "pooling" information across different types of aircraft affect estimates of their individual passenger-fatal event rates λ_i?
- If these are representative of a "population" of airliner types, what range of passenger-fatal event rates might we expect for a "new" type of airliner?

Should we just analyze different airliner types separately, with separate "noninformative" priors on the λ_is?

No, for at least these reasons:

- A Bayesian framework that links the λ_is might allow better estimation ("shrinkage") of poorly-estimated cases.
- Some research questions require consideration of a "population" of λ_is.

Without prior information about differences between the types of airliner, we choose to model them as exchangeable.

Suppose we give the λ_is a gamma prior:

$$\lambda_i \mid \alpha, \beta \sim \text{i.i.d. gamma}(\alpha, \beta)$$

(Why gamma? Can show it is semi-conjugate.)

A gamma (hyper)prior would be semi-conjugate for β. (Show.)

There is no semi-conjugate (hyper)prior for α, but we will choose it to be gamma, as well.

Here's a possibly informative choice:

$$\alpha, \beta \ \sim \ \text{indep. gamma}(1, 1) \ = \ \text{exponential}(1)$$

Here's a possibly "vague" choice:

$$\alpha, \beta \ \sim \ \text{indep. gamma}(0.001, 0.001)$$

[Draw model graph (preliminary) ...]

For the informative case, for example:

```
model {
  for(i in 1:length(y)) {
    y[i] ~ dpois(ymean[i])
    ymean[i] <- lambda[i] * t[i]
    lambda[i] ~ dgamma(alpha,beta)
  }

  alpha ~ dexp(1)
  beta ~ dexp(1)

  lambdamean <- alpha / beta
  lambdavar <- alpha / beta^2

  lambdanew ~ dgamma(alpha,beta)
}
```

[Draw revised model graph ...]

R/JAGS Example 9.2:

Poisson Hierarchical Model

Remarks:

- Note "shrinkage" phenomenon: $\mathrm{E}(\lambda_i \mid \boldsymbol{y})$ is generally less extreme (closer to the population center) than its frequentist counterpart $\hat{\lambda}_i$

- Initial values of α and β that are "too extreme" may cause the process to fail

Model Graphs

A graph has **nodes** (variables) and **edges** (relationships).

A model graph should be a **directed acyclic graph (DAG)**:

- All edges have one-way arrows
- There are no **cycles**, such as ...

  [Draw graph with cycle ...]

Edges represent "parent/child" relationships:

[Draw relationships ...]

A "founder" node has no parents.

Kinds of nodes:

- constant (rectangle): do not change in value, have no distribution, usually have no parents
- stochastic (oval): have an assigned distribution
- logical/deterministic (usually oval): computed purely from other nodes, and get their distribution from those nodes only

Kinds of edges:

- stochastic ($\rightarrow$): each parent is a parameter in the assigned distribution of the child
- logical/deterministic ($\Rightarrow$): the parents are used to compute the child (deterministically)

Plates (big rectangles) indicate replication.

They can be nested (one inside another).