ADVANCED BAYESIAN MODELING

Ship Damage Data and ANOVA

Ship Damage Data

McCullagh & Nelder (1989) compiled numbers of damaging accidents caused by waves to the forward section of cargo-carrying ships:

$$
\begin{aligned}
y_{ijk} &= \text{number of damage incidents to ships in category } (i,j,k) \\
t_{ijk} &= \text{total months of service represented by category } (i,j,k)
\end{aligned}
$$

i : type of ship $\qquad$ j : era of construction $\qquad$ k : period of operation

We expect the mean of y to be proportional to t (all else equal).

The categorical variables defining service category are:

- Ship type: A, B, C, D, or E
- Era of construction: 1960–64, 65–69, 70–74, 75–79
- Period of operation: 1960–74, 75–79

Note: Both era and period are categorical, despite being defined by underlying continuous variables.

```
> library(MASS)

> ships
   type year period service incidents
1     A   60     60     127         0
2     A   60     75      63         0
3     A   65     60    1095         3
4     A   65     75    1095         4
5     A   70     60    1512         6
6     A   70     75    3353        18
7     A   75     60       0         0
8     A   75     75    2244        11
9     B   60     60   44882        39
10    B   60     75   17176        29
11    B   65     60   28609        58
12    B   65     75   20370        53
13    B   70     60    7064        12
...
```

Note: `year` refers to era of construction.

There cannot be any damage incidents for categories with zero service.

We must use only categories for which there was service:

```
> shipssub <- subset(ships, service > 0)
```

We also note that `year` (era) and `period` need conversion to factor variables:

```
> class(shipssub$type)
[1] "factor"

> class(shipssub$year)
[1] "integer"

> class(shipssub$period)
[1] "integer"
```

Data Model

We consider a loglinear rate model for the data:

$$y_{ijk} \mid \beta, t_{ijk}, X_{ijk} \quad \sim \quad \text{indep. Poisson}(t_{ijk}\, r_{ijk})$$

$$\log r_{ijk} \;=\; X_{ijk}\beta \qquad\qquad r_{ijk} \;=\; e^{X_{ijk}\beta}$$

(Note: k is index of periods of operation, *not* number of columns of X)

Variables in X should apparently be indicator variables of type, era, and period, and possibly their interactions.

A full analysis of variance (ANOVA) model would have terms for type, era, and period, all of their two-way interactions, and their three-way interaction:

$$
\begin{aligned}
X_{ijk}\beta \;&=\; \beta^0 + \beta_i^t + \beta_j^e + \beta_k^p \\
&\quad + \beta_{ij}^{te} + \beta_{ik}^{tp} + \beta_{jk}^{ep} \\
&\quad + \beta_{ijk}^{tep}
\end{aligned}
$$

Because some service categories were not available (zero service), not all interactions can be estimated.

Therefore, we use a model without interactions ...

Bayesian Hierarchical ANOVA by Batch

$$X_{ijk}\beta \;=\; \beta^0 + \beta_i^t + \beta_j^e + \beta_k^p$$

As suggested in BDA3, Sec. 15.6, we treat coefficients as random effects, giving them variance components by "batch":

$$\begin{aligned}
\beta_i^t \mid \sigma_t^2 &\sim \text{iid N}(0, \sigma_t^2) \\
\beta_j^e \mid \sigma_e^2 &\sim \text{iid N}(0, \sigma_e^2) \\
\beta_k^p \mid \sigma_p^2 &\sim \text{iid N}(0, \sigma_p^2)
\end{aligned}$$

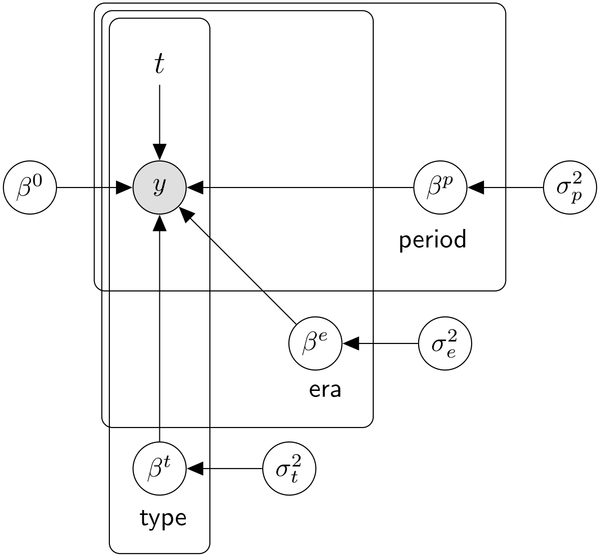
with all coefficients conditionally independent, given the variance components.

Ideally, we would give the top-level parameters $\beta^0, \sigma_t, \sigma_e, \sigma_p$ an improper flat prior:

$$p(\beta^0, \sigma_t, \sigma_e, \sigma_p) \quad \propto \quad 1 \qquad\qquad \sigma_t, \sigma_e, \sigma_p > 0$$

In practice (in JAGS), we try giving them independent diffuse but proper priors.

DAG Model



t
β^0
y
β^p
σ_p^2
period
β^e
σ_e^2
era
β^t
σ_t^2
type