Hierarchical Model Specification in Quantitative Research.

Chapters 14-15, Multilevel Logistic Regression and More

JEFF GILL
Distinguished Professor
Departments of Government and Mathematics & Statistics
American University

Bayesian Normal Models

- Why Be Normal?
  - A great deal of standard theory is based on normal assumptions.
  - Nature loves the normal: CLT.
  - Even non-normal data can often be modeled with normals.
  - Mixtures of normals are extremely flexible.
- Bayesian Normal Models
  - Easy.
  - Have good frequentist properties.
  - Lead directly to the Bayesian linear regression model (Lindley & Smith 1972).
  - Today: lots of conjugacy.

Bayesian Normal Models, Mean Unknown, Variance Known

- Assume that the data are iid with unknown mean μ and known variance σ_0^2:

$$X|\mu,\sigma_0^2 \sim \mathcal{N}(\mu,\sigma_0^2) = (2\pi\sigma_0^2)^{-\frac{1}{2}}\exp\left[-\frac{1}{2\sigma_0^2}(X-\mu)^2\right]$$
$$-\infty < \mu < \infty, \sigma_0^2 \text{ known.}$$

- Then specify a normal prior distribution for μ:

$$\mu|m,s^2 \sim \mathcal{N}(m,s^2)$$
$$= (2\pi s^2)^{-\frac{1}{2}}\exp\left[-\frac{1}{2s^2}(\mu-m)^2\right]$$
$$m, s \text{ given.}$$

- This is another case of *conjugacy*, where a normal prior distribution and a normal likelihood lead to a normal posterior distribution.

Bayesian Normal Models, Mean Unknown, Variance Known

- Posterior Calculation:

$$\pi(\mu|\mathbf{x}) \propto p(\mathbf{x}|\mu)p(\mu)$$

$$\propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2\sigma_0^2}(x_i-\mu)^2\right] \exp\left[-\frac{1}{2s^2}(\mu-m)^2\right]$$

$$= \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2 + \frac{1}{s^2}(\mu-m)^2\right)\right].$$

- Now expand the two squares.

$$\pi(\mu|\mathbf{x}) \propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i^2-2x_i\mu+\mu^2) + \frac{1}{s^2}(\mu^2-2\mu m+m^2)\right)\right]$$

Bayesian Normal Models, Mean Unknown, Variance Known

- Continue with the expansion...

$$\pi(\mu|\mathbf{x}) \propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_0^2}\frac{s^2}{s^2}\sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2) + \frac{1}{s^2}\frac{\sigma_0^2}{\sigma_0^2}(\mu^2 - 2\mu m + m^2)\right)\right]$$

$$= \exp\left[-\frac{1}{2}\frac{1}{\sigma_0^2 s^2}\left(s^2\sum_{i=1}^{n}x_i^2 - 2s^2\mu n\bar{x} + n\mu^2 s^2 + \sigma_0^2\mu^2 - 2\sigma_0^2\mu m + \sigma_0^2 m^2\right)\right]$$

and gather by order of μ...

$$= \exp\left[-\frac{1}{2}\frac{1}{\sigma_0^2 s^2}\left(\mu^2(\sigma_0^2 + ns^2) - 2\mu(m\sigma_0^2 + s^2 n\bar{x}) + \underbrace{(m^2\sigma_0^2 + s^2\sum_{i=1}^{n}x_i^2)}_{k}\right)\right].$$

- Here the blue terms are for μ^2 and the red terms are for μ.
- The last term in the expansion can be treated as part of the normalizing constant, and just labeled k.

Bayesian Normal Models, Mean Unknown, Variance Known

- Rearrange into a Normal Form:

$$\pi(\mu|\mathbf{x}) \propto \exp\left[-\frac{1}{2}\left(\mu^2\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)-2\mu\left(\frac{m}{s^2}+\frac{n\bar{x}}{\sigma_0^2}\right)+k\right)\right]$$

$$= \exp\left[-\frac{1}{2}\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)\left(\mu^2\frac{\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)}{\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)}-2\mu\frac{\left(\frac{m}{s^2}+\frac{n\bar{x}}{\sigma_0^2}\right)}{\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)}+k\right)\right]$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)\left(\mu-\frac{\left(\frac{m}{s^2}+\frac{n\bar{x}}{\sigma_0^2}\right)}{\left(\frac{1}{s^2}+\frac{n}{\sigma_0^2}\right)}\right)^2\right].$$

Bayesian Normal Models, Mean Unknown, Variance Known, Results

- Therefore the point estimate of the mean is:

$$\hat{\mu} = \left(\frac{m}{s^2} + \frac{n\bar{x}}{\sigma_0^2}\right) \Big/ \left(\frac{1}{s^2} + \frac{n}{\sigma_0^2}\right),$$

- And the variance of $\hat{\mu}$ is:

$$\hat{\sigma}_\mu^2 = \left(\frac{1}{s^2} + \frac{n}{\sigma_0^2}\right)^{-1}.$$

- Notice that the posterior mean depends on the data only through $\bar{x}$ (the *sufficient statistic*).
- Proportionality and later normalizing with k made things much easier.

Bayesian Normal Models, Mean Unknown, Variance Known, Precisions

- In the variance term $\frac{1}{s^2}$ is the *prior precision*
- And $\frac{n}{\sigma_0^2}$ is the *data precision*
- The *posterior precision* is the sum of these:

$$\frac{1}{\hat{\sigma}_\mu^2} = \frac{1}{s^2} + \frac{n}{\sigma_0^2}$$

from:

$$\hat{\sigma}_\mu^2 = \left(\frac{1}{s^2} + \frac{n}{\sigma_0^2}\right)^{-1}.$$

- Note what happens as the data size increases for fixed σ_0^2 (this is why precisions are convenient for Bayesians).

Bayesian Normal Models, Mean Unknown, Variance Known, Asymptotics

- The posterior mean estimate:

$$\lim_{n\to\infty} \hat{\mu} = \lim_{n\to\infty} \frac{\frac{m}{s^2} + \frac{n\bar{x}}{\sigma_0^2}}{\frac{1}{s^2} + \frac{n}{\sigma_0^2}} = \lim_{n\to\infty} \frac{\frac{m\sigma_0^2}{ns^2} + \bar{x}}{\frac{\sigma_0^2}{ns^2} + 1} = \bar{x},$$

- The posterior variance of the mean estimate (not the variance of the data):

$$\lim_{n\to\infty} \hat{\sigma_\mu^2} = \lim_{n\to\infty} \frac{1}{\frac{1}{s^2} + \frac{n}{\sigma_0^2}} = \lim_{n\to\infty} \frac{\sigma_0^2}{\frac{\sigma_0^2}{s^2} + n} = \frac{\sigma_0^2}{n}.$$

- Keep in mind that $\hat{\sigma_\mu^2}$ is the variance of the posterior of μ not the posterior of σ^2.

Bayesian Normal Models, Mean Known, Variance Unknown

- Now assume:

$$p(X|\mu_0, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}(X-\mu_0)^2\right].$$

- The corresponding likelihood function is:

$$L(\sigma^2|\mathbf{x}) \propto (\sigma^2)^{-\frac{n}{2}} \exp\Bigg[-\frac{n}{2\sigma^2}\underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2\right)}_{\text{sufficient statistic}}\Bigg].$$

- Relabel the sufficient statistic for σ^2 as a convenience to $\tilde{x}$:

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_0)^2$$

giving the simplified form:

$$L(\sigma^2|\mathbf{x}) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{n}{2\sigma^2}\tilde{x}\right].$$

Bayesian Normal Models, Mean Known, Variance Unknown

- Assign an inverse gamma prior for σ^2:

$$\mathcal{IG}(\sigma^2|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)}\exp[-\beta/\sigma^2]$$

where: $\sigma^2 > 0, \alpha > 0, \beta > 0.$

- This has some moment limitations as well:

$$E[\sigma^2] = \frac{\beta}{\alpha-1}, \qquad \alpha > 1$$

$$\text{Var}[\sigma^2] = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \qquad \alpha > 2.$$

Bayesian Normal Models, Mean Known, Variance Unknown

- Posterior calculation:

$$\pi(\sigma^2|\mathbf{x}) \propto L(\sigma^2|\mathbf{x})p(\sigma^2|\alpha,\beta)$$

$$= (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{n}{2\sigma^2}\tilde{x}\right] \frac{\beta^\alpha}{\Gamma(\alpha)}(\sigma^2)^{-(\alpha+1)} \exp[-\beta/\sigma^2]$$

$$\propto (\sigma^2)^{-((\alpha+\frac{n}{2})+1)} \exp\left[-\left(\beta+\frac{n}{2}\tilde{x}\right)/\sigma^2\right].$$

- Which actually gives the kernel of a different inverse gamma PDF:

$$\sigma^2|\mathbf{x} \sim \mathcal{IG}(\alpha+\frac{n}{2}, \beta+\frac{n}{2}\tilde{x}).$$

Multivariate Normal Model, Mean and Variance Both Unknown

- The most realistic assumption in this family and therefore worthy of considerable attention here.
- The conjugate prior specification for the mean has the same added complexity as before: it must be specified with a dependency the variance: $p(\mu|\sigma^2)$.
- If this is unrealistic, then a nonconjugate prior should be specified.
- For the multivariate case assume:
  - for the n rows of $\mathbf{X}$, this is a k-dimensional vector representing a single case,
  - so now $\boldsymbol{\mu}$ is a vector and $\boldsymbol{\Sigma}$ is a matrix, both to be estimated,
  - from the PDF of the multivariate normal, the likelihood function can be expressed and manipulated as follows...

Both Unknown, Looking at the Likelihood Function

- Since:

$$\left(\sum_{i=1}^{n}(\mathbf{x}_i'\mathbf{x}_i) - n\bar{\mathbf{x}}'\bar{\mathbf{x}}\right) = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}}) \equiv S^2,$$

then $L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X})]$ is a function of the data only through the two-component sufficient statistic: $[\bar{\mathbf{x}}, S^2]$, simplifying the likelihood:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{X}) = |\boldsymbol{\Sigma}|^{-n/2}\exp\left[-\frac{1}{2}\left(\text{tr}(\boldsymbol{\Sigma}^{-1})S^2 + n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})\right)\right]$$

(notice the Gauss-Markov assumptions).

- The conjugate priors for this setup are:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathcal{N}_k\left(\mathbf{m}, \frac{\boldsymbol{\Sigma}}{n_0}\right), \qquad \boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\alpha, \boldsymbol{\beta}),$$

where $\mathcal{W}()$ denotes the Wishart distribution, which is a multivariate generalization of the gamma PDF (an obvious choice for modeling multivariate variances).

Both Unknown (cont.)

- Wishart Form:

$$\mathcal{W}(\boldsymbol{\Sigma}^{-1}|\alpha,\boldsymbol{\beta}) = \frac{|\boldsymbol{\Sigma}^{-1}|^{(\alpha-(k+1))/2}}{\Gamma_k(\alpha)|\boldsymbol{\beta}|^{\alpha/2}}\exp[-\text{tr}(\boldsymbol{\beta}^{-1}\boldsymbol{\Sigma}^{-1})/2]$$

$$\text{where: } \Gamma_k(\alpha) = 2^{\alpha k/2}\pi^{k(k-1)/4}\prod_{i=1}^{k}\Gamma\left(\frac{\alpha+1-i}{2}\right), \quad 2\alpha > k-1, \quad \text{and } \boldsymbol{\beta} \text{ nonsingular.}$$

  where the term $\Gamma_k(\alpha)$ is the k-dimensional generalized gamma function, and is ignorable except for normalizing considerations.

- The parameter n_0 here is not a prior sample size; it is intended to be a reflection of prior precision relative to the sample size that is tunable by the researcher to reflect prior confidence in representability.

- The smaller the ratio n_0/n, the less weight on the prior, and therefore the closer the results will be closer to classical results.

- For additional mathematical details, see:
  `http://www.tc.umn.edu/~nydic001/docs/unpubs/Wishart_Distribution.pdf`.

Both Unknown (cont.)

- The resulting marginal posteriors are produced by taking integrals (reasonable agony involved):

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathcal{N}_k\left(\frac{n_0\mathbf{m} + n\bar{\mathbf{x}}}{n_0+n}, \frac{\boldsymbol{\Sigma}}{n_0+n}\right)$$

$$\boldsymbol{\Sigma}^{-1} \sim \mathcal{W}_k\left(\alpha + n, \boldsymbol{\beta}^{-1} + S^2 + \frac{n_0 n}{n_0+n}(\bar{x}-\mathbf{m})(\bar{x}-\mathbf{m})'\right).$$

- Note that the dependency of the posterior distribution for $\boldsymbol{\mu}$ on $\boldsymbol{\Sigma}$ exists here in the multivariate case as well.

Example: Variance Estimation with Public Health Data

- Consider data from the 2000 U.S. census and North Carolina public records (North Carolina Division of Public Health, Women's and Children's Health Section in Conjunction with State Center for Health Statistics).
- Each case is one of 100 North Carolina counties, and we will use only the following subset of the variables.
- `Substantiated.Abuse`: within family documented abuse for the county.
- `Percent.Poverty`: percent within the county living in poverty, U.S. definition (`http://www.census.gov/hhes/www/poverty/threshld/thresh98.html`).
- `Total.Population`: county population/1000.
- Each $\mathbf{X}$ row is a k-dimensional (3 here) vector representing a single case, distributed $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Example: Variance Estimation with Public Health Data (cont.)

- Relatively uninformed, $\alpha = 3$, $\mathbf{m} = (250, 16, 88)$, $n_0 = 0.01$, $\boldsymbol{\beta}$ a diagonal matrix $w/100$:

$\boldsymbol{\mu}$ Quantile	Abuse	%Poverty	Population
0.01	195.8976	14.2399	77.9827
0.25	199.6618	14.3123	79.7873
0.50	201.2110	14.3409	80.5230
0.75	202.7294	14.3698	81.2590
0.99	206.4080	14.4400	83.0124

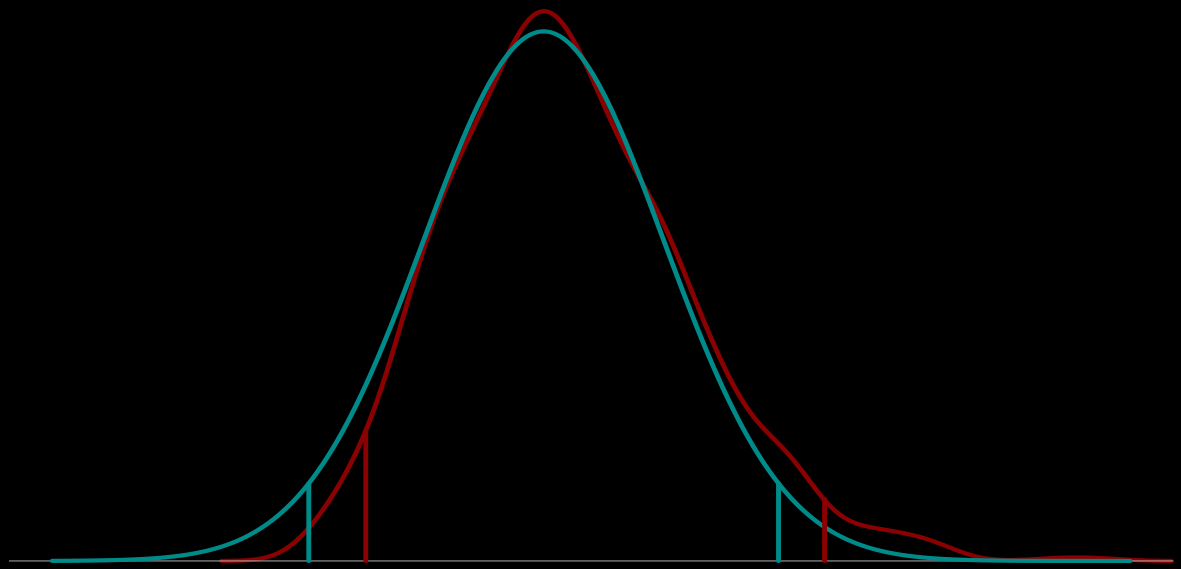
$$\bar{\boldsymbol{\Sigma}} = \begin{bmatrix} 531.553969 & -3.2723672 & 200.207935 \\ -3.272367 & 0.1870651 & -1.672702 \\ 200.207935 & -1.6727021 & 117.901661 \end{bmatrix}$$

- Now add strong priors, $\alpha = 3$, $\mathbf{m} = (100, 6, 88)$, $n_0 = 99$, $\boldsymbol{\beta}$ a diagonal matrix $w/10$:

$\boldsymbol{\mu}$ Quantile	Abuse	%Poverty	Population
0.01	138.4181	9.190786	82.33058
0.25	147.1816	9.902820	83.64427
0.50	150.7495	10.187523	84.19891
0.75	154.3365	10.478351	84.79181
0.99	163.0994	11.159200	86.25384

$$\bar{\boldsymbol{\Sigma}} = \begin{bmatrix} 5678.6595 & 421.23489 & -181.05113 \\ 421.2349 & 35.20976 & -33.15966 \\ -181.0511 & -33.15966 & 146.30970 \end{bmatrix} \quad (1)$$

Comparing the Posterior Distributions for the $\boldsymbol{\Sigma}[1,1]$ Parameter



Note: green line for the likelihood, and red line for the posterior with uninformed prior parameter values.

R Code for the Example

```
nc.sub.df <- read.table("http://jeffgill.org/files/jeffgill/files/nc.sub_.dat_.txt",
    header=TRUE)
library(bayesm); library(BaM)    # FOR THE rwishart AND rmultinorm FUNCTION

Alpha <- 3 + nrow(nc.sub.df)
Beta.inv <- solve(diag(3)*100)
m <- c(250,16,88)
n0 <- 0.01
x.bar <- apply(nc.sub.df,2,mean)
S.sq <- var(nc.sub.df)

k <- (n0 * nrow(nc.sub.df))/(n0 + nrow(nc.sub.df))
p.Beta <- solve( Beta.inv + S.sq + k * round((x.bar-m) %*% t(x.bar-m),2) )
Sigma <- array(NA,dim=c(3,3,1))
for (i in 1:10000) Sigma <- array(c(Sigma,rwishart(Alpha,p.Beta)$IW),dim=c(3,3,(i+1)))
Sigma <- Sigma[,,-1]
```

R Code for the Example

```
Sigma.Mean <- apply(Sigma,c(1,2),mean)
           [,1]       [,2]       [,3]
[1,] 531.553969 -3.2723672 200.207935
[2,]  -3.272367  0.1870651  -1.672702
[3,] 200.207935 -1.6727021 117.901661

# ANALYTICAL MEAN OF THE INVERSE WISHART:
( (Alpha-ncol(nc.sub.df)-1)^(-1) )*solve(p.Beta)
                    Substantiated.Abuse Percent.Poverty
Substantiated.Abuse          531.736988     -3.2745226
Percent.Poverty               -3.274523      0.1872254
Total.Population             200.408410     -1.6760040
                    Total.Population
Substantiated.Abuse       200.408410
Percent.Poverty            -1.676004
Total.Population          118.023667
```

R Code for the Example

```
Sigma.SD <- apply(Sigma,c(1,2),sd)
          [,1]       [,2]       [,3]
[1,] 75.510375 1.04926933 32.2891887
[2,]  1.049269 0.02689763  0.5020452
[3,] 32.289189 0.50204522 16.8975802

# VECTOR MEAN BY SIMULATION
Mu <- rmultinorm(5000,(n0*m + nrow(nc.sub.df)*x.bar)/(n0 + nrow(nc.sub.df)),
    Sigma.Mean/(n0+nrow(nc.sub.df)))
apply(Mu,2,quantile, probs = c(0.01,0.25,0.50,0.75,0.99))
        [,1]     [,2]     [,3]
1%  195.8976 14.23990 77.98269
25% 199.6618 14.31230 79.78725
50% 201.2110 14.34090 80.52302
75% 202.7294 14.36977 81.25900
99% 206.4080 14.44002 83.01237
```

General Comments on Uninformative Priors (more later)

- Somewhat antithetical to the Bayesian principle, but popular.
- Uninformative priors are never really totally "uninformative" since every specified prior has information.
- Usually mathematically more difficult, but an easier "sell."
- Current trends: mildly informed priors, nonparametric priors.
- Warning #1: it is possible to specify a uninformative prior such that posterior credible regions end up with pathological properties such as $P(C|\mathbf{X})$ being dissimilar than $P(C|\theta)$ for all θ (Bernardo and Smith 1994).
- Warning #2: it is possible to specify an improper prior such that the posterior distribution is also improper (Hobert and Casella 1998).

Bayesian Normal Models, Uninformative Priors

- The posterior distribution of the mean parameter is:

$$\pi(\mu|\mathbf{x}) \propto \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{((\pi(n-1))^{\frac{1}{2}}} \left(\frac{n}{s^2}\right)^{\frac{1}{2}} \left(1 + \frac{1}{n-1}\left(\frac{\mu-\bar{x}}{s/\sqrt{n}}\right)^2\right)^{-\frac{1}{2}n}$$

- Therefore the marginal posterior of $\frac{\mu-\bar{x}}{s/\sqrt{n}}$ is student's-t with $\theta = n-1$ degrees of freedom, so the marginal posterior of μ is also student's-t with non-centrality parameter $\bar{x}$.

- Now obtain the marginal posterior of σ by dividing the joint posterior by the conditional distribution of μ assuming that σ is known.

$$\pi(\sigma|\mathbf{x}) = \frac{\pi(\mu,\sigma|\mathbf{x})}{p(\mu|\sigma,\mathbf{x})} = \frac{\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \frac{\left(\frac{(n-1)s^2}{2}\right)^{\frac{n-1}{2}}}{\frac{1}{2}\Gamma\left(\frac{n-1}{2}\right)} \sigma^{-(n+1)} \exp\left[-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\mu-\bar{x})^2\right)\right]}{\sqrt{n}(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right]}$$

$$\propto \sigma^{-((n-1)+1)} \exp\left[-\frac{1}{2}(n-1)s^2/\sigma^2\right].$$

- So the marginal posterior of σ^2 is distributed $\mathcal{IG}((n-2)/2, (n-1)s^2/2)$.

Bayesian Normal Models, IQ Example

- IQ tests are purported to be biased towards Western Europeans and North Americans given their wording and structure.
- Question: is there evidence of economic and cultural biases in national level aggregation of IQ scores.
- The test is designed to have a mean response of 100 with a standard deviation of 15 (the Stanford-Binet version has a standard deviation of 16).

Bayesian Normal Models, IQ Example (cont.)

- Consider collected IQ data (Lynn & Vanhanen 2001) for 81 countries.

Argentina	96	Australia	98	Austria	102	Barbados	78
Belgium	100	Brazil	87	Bulgaria	93	Canada	97
China	100	Congo (Br.)	73	Congo (Zr.)	65	Croatia	90
Cuba	85	Czech Repub.	97	Denmark	98	Ecuador	80
Egypt	83	Eq. Guinea	59	Ethiopia	63	Fiji	84
Finland	97	France	98	Germany	102	Ghana	71
Greece	92	Guatemala	79	Guinea	66	Hong Kong	107
Hungary	99	India	81	Indonesia	89	Iran	84
Iraq	87	Ireland	93	Israel	94	Italy	102
Jamaica	72	Japan	105	Kenya	72	Korea (S.)	106
Lebanon	86	Malaysia	92	Marshall I.	84	Mexico	87
Morocco	85	Nepal	78	Netherlands	102	New Zealand	100
Nigeria	67	Norway	98	Peru	90	Phillipines	86
Poland	99	Portugal	95	Puerto Rico	84	Qatar	78
Romania	94	Russia	96	Samoa	87	Sierra Leone	64
Singapore	103	Slovakia	96	Slovenia	95	South.Africa	72
Spain	97	Sudan	72	Surname	89	Sweden	101
Switzerland	101	Taiwan	104	Tanzania	72	Thailand	91
Tonga	87	Turkey	90	Uganda	73	U.K.	100
U.S.	98	Uruguay	96	Zambia	77	Zimbabwe	66

Bayesian Normal Models, IQ Example (cont.)

- Using the priors: $p(\mu) \propto c$, $p(\sigma) \propto \sigma^{-1}$, we get the posterior summary:

Quantile:	0.01	0.10	0.25	0.50	0.75	0.90	0.99
μ	85.05	86.48	87.30	88.21	89.11	89.93	91.38
σ^2	56.74	63.42	67.71	72.97	78.81	84.61	96.12

- Note that the distribution of μ is centered at 88 rather than 100, and the mode of the posterior variance implies a standard error of roughly 8.5.

New Example in Chapter 14

- Estimating state-level opinions from national polls correcting for non-response at the group (state) level.
- Data come from a 1988 CBS News poll with random digit dialing (RDD) across 51 groups.
- Two steps: fit the multilevel model for all groups, then fit group-level predictions: MRP = Multilevel Regression + Poststratification (Gelman and Little 1997, Gelman, Park, Bafumi 2004, 2006).
- Data Details:
  - Outcome is the binary choice: $y = 1$ for Republican vote, $y = 0$ for Democratic vote.
  - It is assumed that there is no binomial overdispersion.
  - $\theta_\ell =$ average response for each cross-classification of state and categorical demographics, sex (male, female), race (African American, other), age (4 categories), education (4 categories).
  - Therefore $\ell = 1, \ldots, L = 2 \times 2 \times 4 \times 4 \times 51 = 3264$, but we can reduce this by keeping states separated.
  - Cross-classification example: male, African American, age group 2, education group 4, in New York.

Primary Quantity of Interest

- Define N_ℓ as the number of survey respondents in category ℓ, from national demographics (census data).
- Note that this *not* the number of respondents in each category from the survey, which we could label n_ℓ.
- The estimated population average of y in state j is:

$$\hat{\theta}_j = \frac{\sum_{\ell \in j} N_\ell \theta_\ell}{\sum_{\ell \in j} N_\ell}$$

  where the summation is over the $\ell = 2 \times 2 \times 4 \times 4 = 64$ demographic categories in state j.
- This weighting by population average is currently called poststratification.
- We do this because some categories in some states will be small or empty and we want to weight accordingly.
- Another cross-classification example: male, African American, age group 2, education group 4, in Wyoming.

Primary Quantity of Interest

- A non-multilevel model cannot handle this issue directly, because the state-level effects are not treated at a different level of hierarchy.
- The key quantity of interest is the average value of y within each of the ℓ poststratification categories, which should be labeled y_ℓ.
- The combination of poststratification and the use of multilevel model is called "multilevel regression with poststratification or "Mister P."
- MRP cannot fix a bad model.
- If you randomly choose a survey question from a survey and randomly choose any state-level predictor, there is a good possibility that your predictions will be wrong: you have to have a *reason* for picking that predictor.
- Note also that large sample sizes, typical of national surveys, are necessary to be unbiased.
- The standard alternative is disaggregating national survey data to the state level and computing means within.

The Simplest Model

- Specify:

$$p(y_i = 1) = \text{logit}^{-1}(\mathbf{X}_i\boldsymbol{\beta}) = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}$$

- Which for two covariates means:

$$p(y_i = 1) = \text{logit}^{-1}(\alpha_{j[i]} + \beta^{\text{female}} \cdot \text{female} + \beta^{\text{black}} \cdot \text{black}), \qquad i = 1, \ldots, n$$

$$\alpha_j \sim N(\mu_\alpha, \sigma^2_{\text{state}}), \qquad j = 1, \ldots, 51.$$

- Data setup (Gelman):

```
lapply(c("lme4","arm"),library, character.only=TRUE)
data (state)                    # "state" IS AN R DATA FILE
state.abbr <- c (state.abb[1:8], "DC", state.abb[9:50])
dc <- 9
not.dc <- c(1:8,10:51)
region <- c(3,4,4,3,4,4,1,1,5,3,3,4,4,2,2,2,2,3,3,1,1,1,2,2,3,2,4,2,4,
            1,1,4,1,3,2,2,3,4,1,1,3,2,3,3,4,1,3,4,1,2,4)
```

More Data Setup

- Get the file from: JeffGill.org. Right click it on the webpage since it is a Stata file.

```
library(foreign)
polls <- read.dta("CLASSES/Class.Multilevel/examples/election88/polls.dta")
attach(polls)

# SELECT JUST THE DATA FROM THE LAST SURVEY (#9158)
table (survey)                  # look at the survey id's

survey
 9152  9153  9154  9155 9156a 9156b  9157  9158
 1611  1653  1833  1943   684  1478  2149  2193

ok <- survey==9158              # define the condition
polls.subset <- polls[ok,]      # select the subset of interest
detach(polls)
```

More Data Setup

```
print (polls.subset[1:5,])
         org year survey bush state edu age female black weight
11352 cbsnyt    7   9158   NA     7   3   1      1     0    923
11353 cbsnyt    7   9158    1    39   4   2      1     0    558
11354 cbsnyt    7   9158    0    31   2   4      1     0    448
11355 cbsnyt    7   9158    0     7   3   1      1     0    923
11356 cbsnyt    7   9158    1    33   2   2      1     0    403

# CREATE CASEWISE DELETED DATASET(!)
polls.subset.delete <- NULL
for (i in 1:nrow(polls.subset))
    if ( sum(is.na(polls.subset[i,])) == 0 )
        polls.subset.delete <- rbind(polls.subset.delete,polls.subset[i,])
y <- polls.subset.delete$bush
```

The Model

```
M1 <- lmer(y ~ black + female + (1 | state), family=binomial(link="logit"),
      data=polls.subset.delete)
summary(M1)

Formula: y ~ black + female + (1 | state)
   Data: polls.subset
  AIC  BIC logLik deviance
 2667 2689  -1329     2659
Random effects:
 Groups Name        Variance Std.Dev.
 state  (Intercept) 0.169    0.411
Number of obs: 2015, groups: state, 49

Fixed effects:                                    Correlation of Fixed Effects:
            Estimate Std. Error z value Pr(>|z|)        (Intr) black
(Intercept)   0.4452     0.1009    4.41  1.0e-05     black  -0.114
black        -1.7416     0.2080   -8.37  < 2e-16     female -0.551 -0.006
female       -0.0970     0.0946   -1.03     0.31
```

Notes On the Shorter Model

- We do not get σ_y^2 because it is fixed in the logit model at 1.6 (roughly $\pi/\sqrt{3}$) to identify the scale, whereas probit is fixed at 1.0 (see `http://andrewgelman.com/2006/06/take_logit_coef/`).
- We can explicitly create an analogous variance component once we're modeling in `bugs` /`jags` .
- The 51 coefficients estimate vectors by state are given by:

```
coef(M1)
$state
   (Intercept)   black    female    (Intercept)   black    female    (Intercept)   black    female
1    0.9905732 -1.7416 -0.097046 3    0.6861961 -1.7416 -0.097046 4    0.3149191 -1.7416 -0.097046
5    0.3064689 -1.7416 -0.097046 6    0.4050408 -1.7416 -0.097046 7    0.5254054 -1.7416 -0.097046
8    0.2079747 -1.7416 -0.097046 9    0.3516474 -1.7416 -0.097046 10   0.5550147 -1.7416 -0.097046
11   0.6803185 -1.7416 -0.097046 13   0.2466995 -1.7416 -0.097046 14   0.1273424 -1.7416 -0.097046
15   0.6035602 -1.7416 -0.097046 16  -0.0026701 -1.7416 -0.097046 17   0.7726262 -1.7416 -0.097046
18   0.5872641 -1.7416 -0.097046 19   0.5910221 -1.7416 -0.097046 20   0.2515000 -1.7416 -0.097046
21  -0.1121011 -1.7416 -0.097046 22  -0.0427777 -1.7416 -0.097046 23   0.2749340 -1.7416 -0.097046
24  -0.0275582 -1.7416 -0.097046 25   0.8466771 -1.7416 -0.097046 26   0.3433893 -1.7416 -0.097046
27   0.3080935 -1.7416 -0.097046 28   0.4347462 -1.7416 -0.097046 29   0.5339232 -1.7416 -0.097046
30   0.4746324 -1.7416 -0.097046 31   0.4785591 -1.7416 -0.097046 32   0.2224769 -1.7416 -0.097046
33  -0.0155143 -1.7416 -0.097046 34   0.8088724 -1.7416 -0.097046 35   0.4100830 -1.7416 -0.097046
36   0.7300007 -1.7416 -0.097046 37   0.5490738 -1.7416 -0.097046 38  -0.0097895 -1.7416 -0.097046
39   0.2640775 -1.7416 -0.097046 40   0.1995630 -1.7416 -0.097046 41   0.8028206 -1.7416 -0.097046
42   0.3687074 -1.7416 -0.097046 43   1.0934871 -1.7416 -0.097046 44   0.5629053 -1.7416 -0.097046
45   0.9725737 -1.7416 -0.097046 46   0.5676579 -1.7416 -0.097046 47   0.9125061 -1.7416 -0.097046
48   0.4002012 -1.7416 -0.097046 49   0.4215526 -1.7416 -0.097046 50   0.1547878 -1.7416 -0.097046
51   0.5676579 -1.7416 -0.097046
```

Notes On the Shorter Model

- The output from `display` and `summary` gives *model*, but not *null*, deviance:

```
AIC = 2666.7, DIC = 2658.7
deviance = 2658.7
```

```
     AIC  BIC logLik deviance
    2667 2689  -1329     2659
```

- This is rectified by running a null model:

```
M0 <- lmer(y ~ (1 | state), family=binomial(link="logit"),
           data=polls.subset.delete)
summary(M0)
  AIC  BIC logLik deviance
 2749 2760  -1373     2745
Random effects:
 Groups Name        Variance Std.Dev.
 state  (Intercept) 0.13599  0.36877
Number of obs: 2015, groups: state, 49

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.26229    0.07746   3.386 0.000709
```

More On the Shorter Model

- So the deviance comparison comes from comparing the model:

```
 AIC  BIC logLik deviance
2667 2689  -1329     2659
```

to the null model:

```
 AIC  BIC logLik deviance
2749 2760  -1373     2745
```

- This is mechanically done by:

```
pchisq(2745-2659, df=2, lower.tail=FALSE)
[1] 2.1151e-19
```

- Note also that the output of `coef()` is a list so the matrix needs to be pulled out with:

```
dim(coef(M1)[[1]])
[1] 49  3
```

(we lost Alaska and Idaho in the casewise deleting process).

A Fuller Model with Non-Nested Factors

- Now include hierarchies for four factors and some interactions:

$$p(y_i = 1) = \text{logit}^{-1}\Big(\beta^0 + \beta^{\text{female}} \cdot \text{female}_i + \beta^{\text{black}} \cdot \text{black}_i + \beta^{\text{female.black}} \cdot \text{female}_i \cdot \text{black}_i + \alpha^{\text{age}}_{k[i]} + \alpha^{\text{edu}}_{\ell[i]} + \alpha^{\text{age.edu}}_{k[i],\ell[i]} + \alpha^{\text{state}}_{j[i]}\Big)$$

$$\alpha^{\text{state}}_j \sim N(\alpha^{\text{region}}_{m[j]} + \beta^{\text{v.prev}} \cdot \text{v.prev}_j, \sigma^2_{\text{state}})$$

- Notice that these are 4 *non-nested* hierarchies.
- The remaining coefficients are modeled as:

$$\alpha^{\text{age}}_k \sim N(0, \sigma^2_{\text{age}}), \qquad \text{for } k = 1, \ldots, 4$$

$$\alpha^{\text{edu}}_\ell \sim N(0, \sigma^2_{\text{edu}}), \qquad \text{for } \ell = 1, \ldots, 4$$

$$\alpha^{\text{age.edu}}_{k,\ell} \sim N(0, \sigma^2_{\text{age.edu}}), \qquad \text{for } k = 1, \ldots, 4, \ell = 1, \ldots, 4$$

$$\alpha^{\text{region}}_m \sim N(0, \sigma^2_{\text{region}}), \qquad \text{for } m = 1, \ldots, 5$$

A Fuller Model with Non-Nested Factors, Data Setup

```
v.prev=c(0.57,0.63,0.61,0.54,0.54,0.58,0.57,0.53,0.15,0.61,0.52,
         0.52,0.66,0.54,0.58,0.52,0.60,0.53,0.58,0.55,0.50,0.49,
         0.55,0.51,0.58,0.53,0.57,0.67,0.60,0.63,0.56,0.55,0.52,
         0.57,0.61,0.55,0.60,0.54,0.53,0.48,0.57,0.60,0.55,0.59,
         0.69,0.56,0.61,0.55,0.49,0.53,0.63)

attach(polls.subset.delete)
n.edu <- 4                          # 4 CATEGORIES OF EDUCATION
age.edu <- n.edu*(age-1) + edu      # 4 times (age-1) + edu
region.full <- region[state]        # 2193 LONG DATA-LEVEL VECTOR FOR REGIONS
v.prev.full <- v.prev[state]        # 2193 LONG DATA-LEVEL VECTOR FOR PREVIOUS VOTE
black.female <- black*female       # CREATES AN INTERACTION VARIABLE
detach(polls.subset.delete)
```

A Fuller Model with Non-Nested Factors, Estimation

```
M2 <- lmer(y ~ black + female + black.female + v.prev.full + (1 | age) + (1 | edu)
              + (1 | age.edu) + (1 | state) + (1 | region.full),
           data=polls.subset.delete,family=binomial(link="logit"))
summary(M2)
```

```
Formula: y ~ black + female + black.female + v.prev.full + (1 | age) +
             (1 | edu) + (1 | age.edu) + (1 | state) + (1 | region.full)
   Data: polls.subset.delete
  AIC  BIC logLik deviance
 2650 2706  -1315     2630
Random effects:
 Groups      Name        Variance   Std.Dev.
 state       (Intercept) 3.9330e-02 1.9832e-01
 age.edu     (Intercept) 2.2414e-02 1.4971e-01
 region.full (Intercept) 3.1180e-02 1.7658e-01
 edu         (Intercept) 1.1169e-02 1.0568e-01
 age         (Intercept) 1.0243e-09 3.2004e-05
Number of obs: 2015, groups: state, 49; age.edu, 16; region.full, 5; edu, 4; age, 4
```

A Fuller Model with Non-Nested Factors, Estimation

```
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.45146    0.98249  -3.513 0.000443
black        -1.63303    0.32445  -5.033 4.82e-07
female       -0.09002    0.09784  -0.920 0.357503
black.female -0.17916    0.41956  -0.427 0.669369
v.prev.full   6.96836    1.75620   3.968 7.25e-05

Correlation of Fixed Effects:
            (Intr) black  female blck.f
black       -0.026
female      -0.053  0.181
black.femal  0.023 -0.764 -0.233
v.prev.full -0.990  0.009 -0.006 -0.009
```

These Results Can Be Hard To Summarize In Detail

```
coef(M2)
$state
   (Intercept)  black    female black.female v.prev.full
1      -3.2988 -1.633 -0.089996     -0.17918      6.9682
3      -3.4214 -1.633 -0.089996     -0.17918      6.9682
4      -3.5024 -1.633 -0.089996     -0.17918      6.9682
5      -3.4129 -1.633 -0.089996     -0.17918      6.9682
6      -3.4898 -1.633 -0.089996     -0.17918      6.9682
7      -3.4195 -1.633 -0.089996     -0.17918      6.9682
8      -3.5007 -1.633 -0.089996     -0.17918      6.9682
9      -3.4531 -1.633 -0.089996     -0.17918      6.9682
10     -3.6953 -1.633 -0.089996     -0.17918      6.9682
11     -3.3341 -1.633 -0.089996     -0.17918      6.9682
13     -3.5220 -1.633 -0.089996     -0.17918      6.9682
14     -3.5321 -1.633 -0.089996     -0.17918      6.9682
15     -3.3935 -1.633 -0.089996     -0.17918      6.9682
16     -3.5560 -1.633 -0.089996     -0.17918      6.9682
17     -3.3592 -1.633 -0.089996     -0.17918      6.9682
18     -3.4007 -1.633 -0.089996     -0.17918      6.9682
```

19	-3.4759	-1.633	-0.089996	-0.17918	6.9682
20	-3.4970	-1.633	-0.089996	-0.17918	6.9682
21	-3.5688	-1.633	-0.089996	-0.17918	6.9682
22	-3.4978	-1.633	-0.089996	-0.17918	6.9682
23	-3.4909	-1.633	-0.089996	-0.17918	6.9682
24	-3.5443	-1.633	-0.089996	-0.17918	6.9682
25	-3.3692	-1.633	-0.089996	-0.17918	6.9682
26	-3.4208	-1.633	-0.089996	-0.17918	6.9682
27	-3.4853	-1.633	-0.089996	-0.17918	6.9682
28	-3.5423	-1.633	-0.089996	-0.17918	6.9682
29	-3.4384	-1.633	-0.089996	-0.17918	6.9682
30	-3.4526	-1.633	-0.089996	-0.17918	6.9682
31	-3.4040	-1.633	-0.089996	-0.17918	6.9682
32	-3.5080	-1.633	-0.089996	-0.17918	6.9682
33	-3.5556	-1.633	-0.089996	-0.17918	6.9682
34	-3.3694	-1.633	-0.089996	-0.17918	6.9682
35	-3.4761	-1.633	-0.089996	-0.17918	6.9682
36	-3.2177	-1.633	-0.089996	-0.17918	6.9682
37	-3.4895	-1.633	-0.089996	-0.17918	6.9682
38	-3.5627	-1.633	-0.089996	-0.17918	6.9682
39	-3.4199	-1.633	-0.089996	-0.17918	6.9682

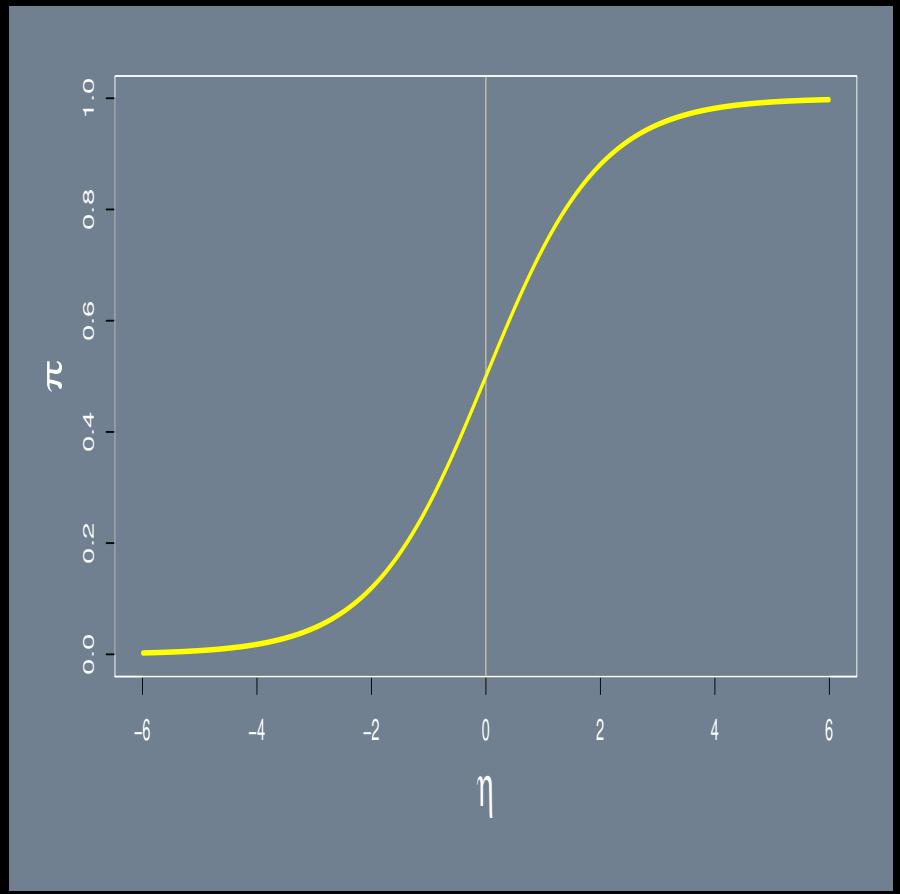
```
40     -3.4622 -1.633 -0.089996     -0.17918      6.9682
41     -3.3786 -1.633 -0.089996     -0.17918      6.9682
42     -3.4786 -1.633 -0.089996     -0.17918      6.9682
43     -3.2179 -1.633 -0.089996     -0.17918      6.9682
44     -3.6143 -1.633 -0.089996     -0.17918      6.9682
45     -3.3620 -1.633 -0.089996     -0.17918      6.9682
46     -3.4178 -1.633 -0.089996     -0.17918      6.9682
47     -3.4213 -1.633 -0.089996     -0.17918      6.9682
48     -3.4367 -1.633 -0.089996     -0.17918      6.9682
49     -3.3320 -1.633 -0.089996     -0.17918      6.9682
50     -3.5036 -1.633 -0.089996     -0.17918      6.9682
51     -3.4297 -1.633 -0.089996     -0.17918      6.9682

$age.edu
   (Intercept)  black    female black.female v.prev.full
1      -3.4929 -1.633 -0.089996     -0.17918      6.9682
2      -3.3408 -1.633 -0.089996     -0.17918      6.9682
3      -3.3491 -1.633 -0.089996     -0.17918      6.9682
4      -3.4286 -1.633 -0.089996     -0.17918      6.9682
5      -3.3884 -1.633 -0.089996     -0.17918      6.9682
6      -3.6042 -1.633 -0.089996     -0.17918      6.9682
```

```
7      -3.4658 -1.633 -0.089996     -0.17918      6.9682
8      -3.5248 -1.633 -0.089996     -0.17918      6.9682
9      -3.4256 -1.633 -0.089996     -0.17918      6.9682
10     -3.4184 -1.633 -0.089996     -0.17918      6.9682
11     -3.4118 -1.633 -0.089996     -0.17918      6.9682
12     -3.4401 -1.633 -0.089996     -0.17918      6.9682
13     -3.6582 -1.633 -0.089996     -0.17918      6.9682
14     -3.4803 -1.633 -0.089996     -0.17918      6.9682
15     -3.3920 -1.633 -0.089996     -0.17918      6.9682
16     -3.4076 -1.633 -0.089996     -0.17918      6.9682

$region.full
  (Intercept)  black    female black.female v.prev.full
1     -3.5383 -1.633 -0.089996     -0.17918      6.9682
2     -3.5283 -1.633 -0.089996     -0.17918      6.9682
3     -3.2117 -1.633 -0.089996     -0.17918      6.9682
4     -3.5324 -1.633 -0.089996     -0.17918      6.9682
5     -3.4528 -1.633 -0.089996     -0.17918      6.9682

$edu
  (Intercept)  black    female black.female v.prev.full
```

```
1     -3.5308 -1.633 -0.089996     -0.17918      6.9682
2     -3.4703 -1.633 -0.089996     -0.17918      6.9682
3     -3.3581 -1.633 -0.089996     -0.17918      6.9682
4     -3.4490 -1.633 -0.089996     -0.17918      6.9682

$age
  (Intercept)  black    female black.female v.prev.full
1     -3.4515 -1.633 -0.089996     -0.17918      6.9682
2     -3.4515 -1.633 -0.089996     -0.17918      6.9682
3     -3.4515 -1.633 -0.089996     -0.17918      6.9682
4     -3.4515 -1.633 -0.089996     -0.17918      6.9682
```

The Divide-By-4 Rule

- The logistic curve is steepest at its center, where $\alpha + \beta x = 0$, and $\text{logit}^{-1}(\alpha + \beta x) = 0.5$ (to generalize, see Nagler's Scobit model, AJPS 1994).



The Divide-By-4 Rule

- The slope at this inflection point is the biggest of anywhere on the curve, solving:

$$\begin{aligned}\frac{d}{dx}\left[1+\exp(x\beta)\right]^{-1} &= \left[1+\exp(x\beta)\right]^{-2}(-1)\exp(x\beta)\beta \\ &= \left.\frac{-\exp(x\beta)\beta}{(1+\exp(x\beta))^2}\right|_{x=0} \\ &= -\frac{\beta}{4}.\end{aligned}$$

- Therefore 4 is the maximum change in $p(Y=1)$ for a one-unit change in x, since this change cannot exceed 1 in absolute value (the sign isn't important here).
- So we can take logistic regression coefficients (other than the constant term) and divide them by 4 to get an upper bound of the predictive difference corresponding to a unit difference in x.
- Assumptions: we are near the middle on the x-axis, and there are no large covariances between model coefficients.

Gelman & Hill Observations

- The intercept,

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.45146    0.98249  -3.513 0.000443
```

  is not easily interpretable since it corresponds to a case in which black, female, and v.prev are all 0, but `v.prev` typically takes on values near 0.5 and is never 0.

- The coefficient for black is:

```
             Estimate Std. Error z value Pr(>|z|)
black        -1.63303    0.32445  -5.033 4.82e-07
```

  Dividing by 4 (see page 82) yields a rough estimate that African-American men were 40% less likely than other men to support Bush, after controlling for age, education, and state.

- The coefficient for female is:

```
             Estimate Std. Error z value Pr(>|z|)
female       -0.09002    0.09784  -0.920 0.357503
```

  Dividing by 4 shows that non-African-American women were very slightly less likely than non-African-American men to support Bush, after controlling for age, education, and state, although the standard error does not let us make this claim.

Gelman & Hill Observations

- The large standard error on the coefficient for black:female,

```
             Estimate Std. Error z value Pr(>|z|)
black.female -0.17916    0.41956  -0.427 0.669369
```

  indicates that the sample size is too small to estimate this interaction precisely.

- The coefficient for v.prev.full is:

```
             Estimate Std. Error z value Pr(>|z|)
v.prev.full   6.96836    1.75620   3.968 7.25e-05
```

  which, when divided by 4, is 1.7, suggesting that a 1% increase in a state's support for Republican candidates in the previous election mapped to a predicted 1.7% difference in support for Bush in 1988.

Gelman & Hill Observations

- The state-level errors have an estimated standard deviation of roughly 0.2 on the logit scale:

```
Groups       Name        Variance   Std.Dev.
state        (Intercept) 3.9330e-02 1.9832e-01
```

  Dividing by 4 tells us that the states differed by approximately $\pm 5\%$ on the probability scale (over and above the differences explained by demographic factors).

- The differences among age-education groups and regions are also approximately $\pm 5\%$ on the probability scale:

```
Groups       Name        Variance   Std.Dev.
age.edu      (Intercept) 2.2414e-02 1.4971e-01
region.full (Intercept) 3.1180e-02 1.7658e-01
```

- Very little variation is found among age groups or education groups after controlling for the other predictors in the model.

```
Groups       Name        Variance   Std.Dev.
edu          (Intercept) 1.1169e-02 1.0568e-01
age          (Intercept) 1.0243e-09 3.2004e-05
```

Using the Model Inferences to Estimate Average Opinion For Each State

- The model gives the probability that any adult will prefer Bush, given the ethnicity, age, education level, and state of the person.
- We can now compute weighted averages of these probabilities to represent the proportion of Bush supporters in any specified subset of the population: *poststratification*.
- We first extract from the U.S. Census the counts N_ℓ in each of the 3264 cross-classification cells and create a 3264×6 data frame, census, indicating the sex ethnicity, age, education, state, and number of people.
- This corresponds to each cell according to our data:

```
         org year survey bush state edu age female black weight
11353 cbsnyt    7   9158    1    39   4   2      1     0    558
11354 cbsnyt    7   9158    0    31   2   4      1     0    448
11355 cbsnyt    7   9158    0     7   3   1      1     0    923
11356 cbsnyt    7   9158    1    33   2   2      1     0    403
11357 cbsnyt    7   9158    1    33   4   4      1     0    317
11358 cbsnyt    7   9158    1    39   2   2      0     0   1532
11359 cbsnyt    7   9158    1    20   2   4      1     0    896
:
```

Using the Model Inferences to Estimate Average Opinion For Each State

- The estimated population average of y in state j is:

$$\hat{\theta}_j = \frac{\sum_{\ell \in j} N_\ell \theta_\ell}{\sum_{\ell \in j} N_\ell}$$

  where the summation is over the $\ell = 2 \times 2 \times 4 \times 4 = 64$ demographic categories in state

- Then compute y^{pred} for each category.

- Our output will be $p(y_i) = 1$, meaning the probability that one of these distinct cases votes for Bush.

Using the Model Inferences to Estimate Average Opinion For Each State

```
# CREATE A GRID OF ALL POSSIBLE CROSSES
polls.X <- expand.grid(state=1:51,edu=1:4,age=1:4,female=0:1,black=0:1)
rbind(polls.X[1:7,],polls.X[(nrow(polls.X)-6):nrow(polls.X),]) #ILLUSTRATION
     state edu age female black
1        1   1   1      0     0
2        2   1   1      0     0
3        3   1   1      0     0
4        4   1   1      0     0
5        5   1   1      0     0
6        6   1   1      0     0
7        7   1   1      0     0
3258    45   4   4      1     1
3259    46   4   4      1     1
3260    47   4   4      1     1
3261    48   4   4      1     1
3262    49   4   4      1     1
3263    50   4   4      1     1
3264    51   4   4      1     1
```

Using the Model Inferences to Estimate Average Opinion For Each State

```
polls.X[1,]           state edu age female black
                  1     1   1   1      0     0

# LOOK AT ALABAMA
coef(M2)$state[1,]
  (Intercept)  black    female black.female v.prev.full
1     -3.2988 -1.633 -0.089996     -0.17918      6.9682

# THIS IS THE OFFSET FROM THE US MEAN, BUT WE'LL DIRECTLY USE THE INTERCEPT
ranef(M1)$state[1,]          [1] 0.54534

# MRP PREDICTION OF THE CASE DESCRIBED BY polls.X[1,]
( y1.pred <- inv.logit( coef(M2)$state[1,1]
                + polls.X[1,5]*coef(M2)$state[1,]$black
                + polls.X[1,4]*coef(M2)$state[1,]$female
                + polls.X[1,4]*polls.X[1,5]*coef(M2)$state[1,]$black.female
                + v.prev[1]*coef(M2)$state[1,]$v.prev.full ) )
[1] 0.72225
```

Item-Response and Ideal-Point Models, Rasch

- The idea is to develop a model for success or failure over testing items: the probability that individual i gets question j correct, giving the outcome: $y_{jk} = 1$.
- A standard model in this literature is the Rasch Model, which just an *logit item response model*.
- Assume $j = 1, 2, \ldots, J$ test-takers and $k = 1, 2, \ldots, K$ items/questions.
- The model is:

$$p(y_{jk} = 1) = \text{logit}^{-1}(\alpha_j - \beta_k)$$

where:

- α_j is the ability/knowledge/skill of person j
- β_k is the difficult of question k.

Item-Response and Ideal-Point Models, Rasch

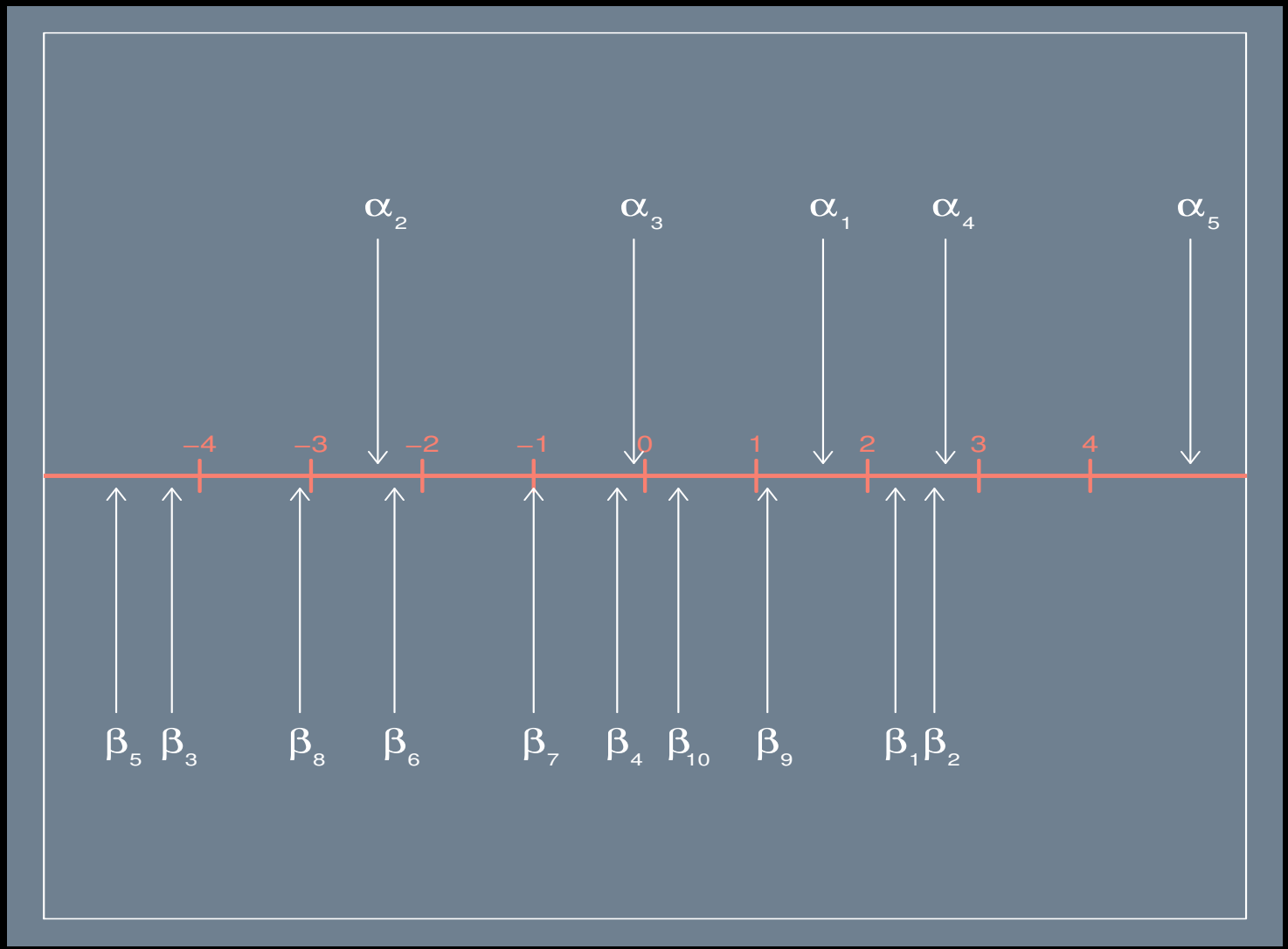
- If not every person answers every question then it is better to put this into a multilevel context.
- If the subset of answered questions is indexed by i then
  - $\alpha_{j[i]}$ is the ability/knowledge/skill of person j
  - $\beta_{k[i]}$ is the difficult of question k.

  for $i < K$.
- The model is now:

$$p(y_{jk} = 1) = \text{logit}^{-1}(\alpha_{j[i]} - \beta_{k[i]}).$$

- In the following illustrated example, with 5 persons and 10 items, if the person's α is greater than a specific β, then there is better than a 0.5 probability of getting the answer right (an assumed normal distribution around both terms).

Item Response Example, Page 315



Identifiability Problem

- The figure also shows the non identifiability with the model so far: the probabilities depend only on the *relative* positions of the ability and difficulty parameters.
- This means that any a constant could be added to all the α and β and all the inferences would remain the same.
- Common solutions:
  - Make the α and β vectors have mean or sum zero.
  - Give the α and β terms a distribution with mean zero.
  - Make $\alpha_1 = 0$, OR $\beta_1 = 0$, to serve as comparison points.

Multilevel Item Response Model

- Start with the normal assumptions for the ability and difficulty parameters:

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \quad \text{for } j = 1, 2, \ldots, J$$
$$\beta_k \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \quad \text{for } k = 1, 2, \ldots, K.$$

- This model is identified if we either set $\mu_\alpha = 0$ OR $\mu_\beta = 0$.

- Now include group level covariates

$$\alpha_j \sim \mathcal{N}(\mathbf{X}_j^\alpha \boldsymbol{\gamma}_\alpha, \sigma_\alpha^2), \quad \text{for } j = 1, 2, \ldots, J$$
$$\beta_k \sim \mathcal{N}(\mathbf{X}_k^\beta \boldsymbol{\gamma}_\beta, \sigma_\beta^2), \quad \text{for } k = 1, 2, \ldots, K$$

  (there are two typos on that last line in my copy of the book).

- G&H P.316: "In an educational testing example, the person-level predictors X^α could include age, sex, and previous test scores, and the item-level predictors X^β could include a prior measure of item difficulty (perhaps the average score for that item from a previous administration of the test)."

Another Approach to Identifiability

- Rescaling both parameters by the mean of the ability parameters works:

$$\alpha_j^{\text{adj}} = \alpha_j - \bar{\alpha}, \quad \text{for} j = 1, 2, \ldots, J$$
$$\beta_k^{\text{adj}} = \beta_k - \bar{\alpha}, \quad \text{for} k = 1, 2, \ldots, K.$$

- Note that both quantities subtract $\bar{\alpha}$, or the identification process would not align.
- These LHS quantities are identified replacing the original definitions such that:

$$p(y_i = 1) = \text{logit}^{1-}(\alpha_{j[i]}^{\text{adj}} - \beta_{k[i]}^{\text{adj}}).$$

Adding a Discrimination Parameter

- Sometimes we want to modify the trajectory of the logit curve to better discriminate the subjects taking the test (see Figure 14.14 on P.317 in G&H).
- This is done by adding a new term:

$$p(y_i = 1) = \text{logit}^{1-}(\gamma_{k[i]}(\alpha^{\text{adj}}_{j[i]} - \beta^{\text{adj}}_{k[i]}))$$

  where $\gamma_{k[i]}$ is the "discrimination" of test item k.
- If $\gamma_{k[i]} = 0$ then there is no discrimination and $p(y_i = 1) = 0.5$ for everybody.
- Conversely high values of $\gamma_{k[i]}$ means that there is a strong relation between ability and the probability of getting a correct answer.
- So discrimination is a measure of quality of the question, but we need an additional identifiability constraint which G&H get back to in Chapter 20 (P.416).
- In educational testing (ACT, GMAT, MCAT, LSAT, etc.) higher discrimination is important and the goal is to have this plus a wide range of β values to more accurately place test-takers.
- Note also that negative values of $\gamma_{k[i]}$ are construction mistakes that reward poorer test-takers.

Application to the SCOTUS

- A common application of item-response models replaces correct test answers with "correct" voting in some predetermined direction, in this case "correct" = "conservative" in votes but the ideology direction is totally arbitrary.
- This assumption sets the ideology scale as getting more conservative going rightward on the x-axis.
- The G&H data come from decisions 1965 to 2006 where each vote i is associated with justice $j[i]$ and case $k[i]$.
- The coding is 1 for a yes vote and 0 for a no vote in the conservative direction.
- Specify a logit model with the probability of voting yes based on an "ideal point" for each justice, α_j, and the positive for each case before the court, β_k.
- Also specify a discrimination parameter $\gamma_{k[i]}$.

Application to the SCOTUS

- For justice j and case k, the value of $\alpha_j - \beta_k$ gives the judges relative position (likelihood of voting yes) to the case.
- If β_k is near the justice's ideal point (distribution mean) then he/she has an even chance of voting yes or no.
- But if this distance is large (one way or another) the probability of a particular vote is very high.
- If the discrimination parameter is near zero then votes look fairly uniformly random.
- But if the discrimination parameter is large (in either the positive or negative direction) then the α_j is wholly determinate.

Application to the SCOTUS

- There are three sources of non-identifiability here. Two are:
  - the same situation from before where adding a constant to all of the α and β terms does not change the inferences.
  - multiplying the γ terms by some constant and dividing α and β terms by the same constant does not change the inferences.
- Both problems are solved here by giving the α_j terms a $\mathcal{N}(0,1)$ distribution.
- The other terms remain unconstrained:

$$\beta_k \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \quad \text{for } k = 1, 2, \ldots, K$$
$$\gamma_k \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \quad \text{for } k = 1, 2, \ldots, K$$

Application to the SCOTUS

- The third source of non identification is the result of multiplying all of the terms (α, β, and γ) by -1.
- This will produce a *bimodal* likelihood and posterior distribution, and the question will be which mode do we want (two MLEs or Bayesian posterior modes for each parameter).
- Solution #1: constrain the γ parameters to be positive. This relies a lot on how votes pre-coded as being in the liberal or conservative direction, and may be a poor choice for moderate judges.
- Solution #2: choose one of the 3 parameters and restrict its sign as positive or negative. G&H example: constrain α_j to be negative for the extremely liberal William Douglas, or constrain α_j to be positive for the extremely conservative Antonin Scalia, or just constrain Douglas to be less than Scalia.

Application to the SCOTUS

- Solution #3: add a *group level* regression covariate, whose coefficient is constrained to be positive.
- In this example G&H stipulate an indicator variable in the hierarchy that equals 1 for Scalia, -1 for Douglas, and 0 for all other justices.
- This identifies the model since aligning the positive direction with the difference between these two extremes.
- Note that this requires defensible *prior information* about the coefficients.

Non-Nested Overdispersed Model for Death Sentence Reversals

- We can modify models in this chapter to account for outcomes that are proportions.
- So y_i is the probability of a success, not the observation of a success.
- Or y_i is the number of successes out of n_i attempts.
- Section 6.3 describes data where the outcome variable is *the number of death penalty sentences per state that are reversed by a higher court.*
- G&H create a non-nested model for state, $j = 1, \ldots, J$, and year, $t = 1, \ldots, T$, coefficients.

Non-Nested Overdispersed Model for Death Sentence Reversals

- Explanatory variables: frequency that the death sentence was imposed, the backlog of capital cases in the appeals courts, the level of political pressure on judges, indicators for the years from 1973 to 1995 for the 34 states (all of those in this time span that had death penalty laws).

- The regression model with all these predictors is:

$$y_i \sim \text{Bin}(n_i, p_i)$$
$$p_i = \text{logit}^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \alpha_{j[i]} + \gamma_{t[i]})$$

  where j indexes states and t indexes years.

- The necessary distributions for the state and year coefficients are:

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$
$$\gamma_t \sim N(a + bt, \sigma_\gamma^2)$$

- The coefficients for year are include as a linear time trend to capture the overall increase in reversal rates over the time of the study.

Non-Nested Overdispersed Model for Death Sentence Reversals

- The model for the γ_t hierarchy includes an intercept, and so we do not need to include a constant term in the hierarchy for α_j or at the individual level.
- The multilevel structure here is just to take care of data heterogeneity and get better estimates for the $\boldsymbol{\beta}$ terms.
- There is an overdispersion problem for these data, with this model:
  - Create the standardized residuals:

$$z_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)/n_i}}$$

    where $\hat{p}_i = \text{logit}^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \alpha_{j[i]} + \gamma_{t[i]})$.

  - Under the binomial assumptions $z_i \sim N(0,1)$, but a better test comes from:

$$\sum_{i=1}^{n} z_i^2 \sim \chi^2_{df=(T\times J)-k}$$

    where $T \times J = 520$, and k is the number of explanatory variables including the constant ($T \times J$ is less than $23 \times 34 = 782$ because not all states had death penalty laws in all years).

Non-Nested Overdispersed Model for Death Sentence Reversals

- Fix #1: use the beta-binomial distribution instead of the binomial distribution:

$$y_i \sim \text{beta-binomial}(n_i, p_i, \omega),$$

  where $\omega \geq 1$ is the overdispersion parameter and the model with $\omega = 1$ reduces to the binomial.

- In `R`, use the `glm` function with `quasibinomial(link="logit")`

- Fix #2: use the binomial-normal model instead of the binomial distribution by adding normally distributed errors on the logistic scale:

$$p_i = \text{logit}^{-1}(\mathbf{X}_i\boldsymbol{\beta} + \alpha_{j[i]} + \gamma_{t[i]} + \xi_i)$$
$$\xi_i = N(0, \sigma_\xi^2)$$

  where the model reduces to the binomial when $\sigma_\xi^2 = 0$.

- Generally this has to be done with `bugs` /`jags` .

Non-Nested Overdispersed Model for Death Sentence Reversals

- With moderate sample sizes, it is typically difficult to distinguish between the beta-binomial and binomial-normal models.
- The beta-binomial model adds only one new parameter and so it can be easier to fit
- The binomial-normal model has the advantage that the new error term, ξ_i, is on the same scale as the group-level predictors, $\alpha_{j[i]}$ and $\gamma_{t[i]}$ which can make the fitted model easier to explain.

Overdispersed Poisson Regression

- Data that are fit with a GLM that have greater variance than assumed by the model used are called overdispersed.
- The binomial and Poisson regression models lack a natural parameter to deal with overdispersion so one must be added.
- Chapter 15 focuses on the *police stops data* from chapter 6.
- Instead consider the data from: Koch, M. T. Cranmer, S. (2007). "Testing the 'Dick Cheney' Hypothesis: Do Governments of the Left Attract More than Governments of the Right?" *Conflict Management and Peace Science* **24**, 311-326.

Data Description

- In this example we look at terrorist activity in 22 Asian democracies over 8 years (1990-1997) with data subsetted from Koch and Cranmer (2007), $n = 150$.
- The outcome of interest a count of violent terrorists attacks in a country/year pair, `ATT`.
- `DEM` for these countries is the Polity IV 21-point democracy scale ranging from -10 indicating a hereditary monarchy to +10 indicating a fully consolidated democracy.
- The variable `FED` is assigned zero if sub-national governments do not have substantial taxing, spending, and regulatory authority, and one otherwise.
- Governmental systems, `SYS`, coded as: (0) for direct presidential elections, (1) for strong president elected by assembly, and (2) dominant parliamentary government.
- `AUT` is a dichotomous variable indicating whether or not there are autonomous regions not directly controlled by central government.
- `LEF` is 1 if the government is coded left-of-center, 0 otherwise.
- `CID` is the country-identifying code.
- `SUM` is the total number of attacks by country over the period of study.

Poisson Regression with an Offset

- Generalize from the standard Poisson GLM:

$$y_i \sim \text{Poisson}(\theta_i), \qquad \theta_i = \exp(\mathbf{X}_i\boldsymbol{\beta}),$$

  to:

$$y_i \sim \text{Poisson}(u_i\theta_i), \qquad \theta_i = \exp(\mathbf{X}_i\boldsymbol{\beta}).$$

- Here u_i is the exposure, and $\log(u_i)$ is the offset.
- The idea is to scale the effect of the rate.
- Typical strategy: put the log of the exposure into the model as an offset which forces its regression coefficient to be 1.
- Here we will scale the count per case by the sum total.

Using an Offset

- We just modeled these as counts independent of the amount of exposure.
- But the counts are actually out of a number of cases exposed.
- This is called a rate model in the count literature: events per unit of exposed.
- Thus we want to put exposure on the RHS of the model, being careful about logs:

$$\log\left(\frac{E[Y|\boldsymbol{\beta},\mathbf{X}]}{\text{exposure}}\right) = \mathbf{X}\boldsymbol{\beta}$$

$$\log(E[Y|\boldsymbol{\beta},\mathbf{X}]) - \log(\text{exposure}) = \mathbf{X}\boldsymbol{\beta}$$

$$\log(E[Y|\boldsymbol{\beta},\mathbf{X}]) = \mathbf{X}\boldsymbol{\beta} + \log(\text{exposure})$$

$$E[Y|\boldsymbol{\beta},\mathbf{X}] = \exp\left[\mathbf{X}\boldsymbol{\beta} + \log(\text{exposure})\right]$$

  which justifies putting a log-constant on the RHS to reflect the number exposed in each case.

- In `R` this is done with the `offset()` specification.

Terrorism Count Data

```
library(foreign)
# THE SYLLABUS HAS THE CONDENSED DATA SET (1990 TO 1997, ASIA ONLY, CASEWISE DELETED):
asia.sub.df <-
    read.table("CLASSES/Class.Multilevel/Data/cheney.asia.sub.txt",
    header=FALSE)
names(asia.sub.df) <- c("ATT","DEM","FED","SYS","AUT","LEF","CID","SUM")

table(asia.sub.df$ATT)
 0  1  2  3  4  5  6  7  8  9 10 11 12 13 16 38
83 14  8 12 10  4  3  4  3  3  1  1  1  1  1  1

table(asia.sub.df$LEF)
 0  1
99 51
```

Terrorism Count Data, Simple Poisson GLM

```
asia.1 <- glm(ATT ~  DEM + FED + SYS + AUT + LEF, family=poisson, data=asia.sub.df)
summary(asia.1)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.7523     0.1126    6.68  2.4e-11
DEM           0.0600     0.0116    5.16  2.5e-07
FED          -0.5194     0.1944   -2.67   0.0075
SYS          -0.3213     0.0717   -4.48  7.4e-06
AUT           0.0027     0.3046    0.01   0.9929
LEF           0.6795     0.1399    4.86  1.2e-06
Null deviance: 763.32  on 149  degrees of freedom
Residual deviance: 707.82  on 144  degrees of freedom
AIC: 928.6

# ESTIMATED OVERDISPERSION
z <- (asia.sub.df$ATT - asia.1$fitted.values)/sd(asia.1$fitted.values)
pchisq(sum(z^2), df=asia.1$df.residual, lower.tail=FALSE)
[1] 0
```

Terrorism Count Data, Model With Log-Exposure

```
asia.2 <- glm(ATT ~  DEM + FED + SYS + AUT + LEF, offset=log1p(SUM), family=poisson,
              data=asia.sub.df)
summary(asia.2)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.024354   0.106984  -18.92  < 2e-16
DEM          0.000402   0.011437    0.04    0.972
FED         -0.436426   0.197849   -2.21    0.027
SYS         -0.100306   0.067730   -1.48    0.139
AUT         -0.572688   0.314415   -1.82    0.069
LEF          0.685259   0.135114    5.07  3.9e-07
Null deviance: 359.64  on 149  degrees of freedom
Residual deviance: 334.22  on 144  degrees of freedom
AIC: 555.1

# ESTIMATED OVERDISPERSION
z <- (asia.sub.df$ATT - asia.2$fitted.values)/sd(asia.2$fitted.values)
pchisq(sum(z^2), df=asia.2$df.residual, lower.tail=FALSE)
[1] 2.0001e-05
```

Terrorism Count Data, Modeling With Overdispersion

```
asia.3 <- glm(ATT ~  DEM + FED + SYS + AUT + LEF, offset=log1p(SUM),
              family=quasipoisson, data=asia.sub.df)
summary(asia.3)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.024354   0.169716  -11.93   <2e-16
DEM          0.000402   0.018143    0.02   0.9823
FED         -0.436426   0.313862   -1.39   0.1665
SYS         -0.100306   0.107444   -0.93   0.3521
AUT         -0.572688   0.498779   -1.15   0.2528
LEF          0.685259   0.214340    3.20   0.0017

(Dispersion parameter for quasipoisson family taken to be 2.5166)

Null deviance: 359.64  on 149  degrees of freedom
Residual deviance: 334.22  on 144  degrees of freedom
```

A Multilevel Poisson Regression Model

- Add a hyperparameter σ_ϵ that measures the amount of overdispersion accounting for country as a group:

$$y_i \sim \text{Poisson}(u_i\theta_i), \qquad \theta_i = \exp(\mathbf{X}_i\boldsymbol{\beta} + \epsilon_j[i]), \qquad \epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2).$$

- We can use `CID` to create a country-level in the model

```
asia.4 <- lmer(ATT ~  DEM + FED + SYS + AUT + LEF + (1|CID), offset=log1p(SUM),
               family=poisson, data=asia.sub.df)
summary(asia.4)
```

- Model Summaries:

```
 AIC BIC logLik deviance
 345 366   -165      331
```

- Random Effects:

```
 Groups Name        Variance Std.Dev.
 CID    (Intercept) 0.0434   0.208
Number of obs: 150, groups: CID, 21
```

A Multilevel Poisson Regression Model

- Fixed Effects:

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.0933     0.1498  -13.97  < 2e-16
DEM           0.0160     0.0152    1.05    0.292
FED          -0.3973     0.2786   -1.43    0.154
SYS          -0.1755     0.0979   -1.79    0.073
AUT          -0.3795     0.4026   -0.94    0.346
LEF           0.7318     0.1500    4.88  1.1e-06
```

- Correlation of Fixed Effects:

```
(Intr) DEM    FED    SYS    AUT
DEM -0.343
FED  0.237 -0.210
SYS -0.620 -0.103 -0.395
AUT -0.170  0.402  0.102 -0.094
LEF -0.382  0.038 -0.401  0.147 -0.256
```

Observations

- This is a good example of why we need to move to `bugs` /`jags` .
- `lmer` gives (decent) approximations for hierarchical parameters.
- It also assumes that σ_y is constant in groups.
- We also do not get uncertainty estimates for terms like σ_α.
- `lmer` is a great learning tool and a good place to get rough estimates for more complicated models.
- Moving to `bugs` /`jags` also gives us more flexibility in defining hierarchies: third levels and beyond, non-normal distributional assumptions, and more.

Getting Started with the `bugs` Language

- Models in the `bugs` language are specified more theoretically than in other packages.
- We will follow G&H and use vague priors everywhere, but this is an important issue in some instances.
- Example from the assigned reading for next week: "Bayesian Analytical Methods: A Methodological Prescription for Public Administration."
- Data from the 1998 and 2004 rounds of the American State Administrator's Project (ASAP) survey.
- Our outcome variable of interest `grp.influence` is an index of the respondents' (senior executives') perceptions of the influence that clientele groups have on the total agency budget, specialized program budgets, and agency policies. Each of these questions is a seven-point scale and they have been summed to create a single outcome variable (ranging from 3 to 21).

Getting Started with the bugs Language

- Explanatory variables: contracting, gov.influence, leg.influence, elect.board, years.tenure, education, party.ID, category2, category3, category4, category5, category6, category7, category8, category9, category10, category11, category12, med.time, medt.contr, gov.ideology, lobbyists, nonprofits.
- Data for jags needs to be in list form:

```
asap.jags.list <- list(STATES <- 50, SUBJECTS <- 713,
    state.id <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...),
    contracting <- c(6, 2, 0, 0, 0, 1, 0, 3, 3, 6, 0, 1, 5, 1, 1, 1, 0,...),
    :
    nonprofits <- c(1.9783, 0.509, 2.0701, 1.3639, 15.6682, 2.7968,...)
)
```

- Or read this in from a file that you've constructed.

Getting Started with the `bugs` Language

- Markov chains require starting points, so create the following to modify:

```
"tau" <- 1
"tau.alpha" <- 1
"alpha" <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
             0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
             0,0,0,0,0,0,0,0,0,0,0,0)
"beta" <- c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
"gamma" <- c(0,0,0)
```

- Then make them different.

The First Part of the jags Code

```
model  {
  for (i in 1:SUBJECTS) {
    mu[i] <- alpha[state.id[i]]
      + beta[1]*contracting[i] + beta[2]*gov.influence[i] + beta[3]*leg.influence[i]
      + beta[4]*elect.board[i] + beta[5]*years.tenure[i]  + beta[6]*education[i]
      + beta[7]*party.ID[i]    + beta[8]*category2[i]     + beta[9]*category3[i]
      + beta[10]*category4[i]  + beta[11]*category5[i]    + beta[12]*category6[i]
      + beta[13]*category7[i]  + beta[14]*category8[i]    + beta[15]*category9[i]
      + beta[16]*category10[i] + beta[17]*category11[i]   + beta[18]*category12[i]
      + beta[19]*med.time[i]   + beta[20]*medt.contr[i]
    grp.influence[i] ~ dnorm(mu[i],tau)
  }
  for (j in 1:STATES) {
    eta[j]  <- gamma[1]*gov.ideology[j] + gamma[2]*lobbyists[j]
             + gamma[3]*nonprofits[j]
    alpha[j] ~ dnorm(eta[j],tau.alpha)
  }
```

The Second Part of the jags Code

```
  beta[1]    ~ dnorm(0.070,1)      # PRIOR MEANS FROM KELLEHER AND YACKEE 2009, MODEL 3
  beta[2]    ~ dnorm(-0.054,1)
  beta[3]    ~ dnorm(0.139,1)
  beta[4]    ~ dnorm(0.051,1)
  beta[5]    ~ dnorm(0.017,1)
  beta[6]    ~ dnorm(0.056,1)
  beta[7]    ~ dnorm(0.039,1)
  beta[8]    ~ dnorm(0.0,1)        # DIFFUSE PRIORS
  :
  beta[18]   ~ dnorm(0.0,1)
  beta[19]   ~ dnorm(0.184,1)      # PRIOR MEANS FROM KELLEHER AND YACKEE 2009, MODEL 3
  beta[20]   ~ dnorm(0.156,1)
  gamma[1]   ~ dnorm(0.0,1)        # DIFFUSE PRIORS
  gamma[2]   ~ dnorm(0.0,1)
  gamma[3]   ~ dnorm(0.0,1)
  tau        ~ dgamma(1.0,1)
  tau.alpha  ~ dgamma(1.0,1)
}
```

Running jags From R

```
# LOAD LIBRARY AND SOURCE FILES
library(rjags); library(arm); library(coda); library(superdiag)

# DEFINE THE MODEL
asap.model2.rjags  <- function()  {
  for (i in 1:SUBJECTS) {
  :
}

# SAVE MODEL TO A FILE
write.model(asap.model2.rjags, "Article.JPART/asap.model2.rjags")
```

Running jags From R

```
# RUN THE SAMPLER AND COLLECT coda SAMPLES
asap2.model <- jags.model(file="Article.JPART/asap.model2.rjags",
    inits=asap.inits, data=asap.jags.list, n.chains=3,  n.adapt=5000)
update(asap2.model, n.iter=2500)
asap2.mcmc <- coda.samples(model=asap2.model, variable.names=names(asap.jags.list),
    n.iter=2500)
summary(asap2.mcmc)

# CHECK CONVERGENCE
superdiag(as.mcmc.list(asap2.mcmc), burnin=0)

# GET THE DEVIANCE AND THE DIC
asap2.dic <- dic.samples(asap2.model, n.iter=25000, type="pD")
```