Bayes Computation Chapter 18, ARM

18.1 is a quick recap of frequentist linear models.
Assumed likelihood (under independence):

$$p(\mathbf{y}|\boldsymbol{\beta}, \sigma, \mathbf{X}) = \prod_{i=1}^{n} N(y_i|\mathbf{X}\boldsymbol{\beta},\ \sigma^2)$$

or with non-constant variances and/or correlations

$$p(\mathbf{y}|\beta, \Sigma, \mathbf{X}) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp(-(\mathbf{y}-\mathbf{X}\beta)^{\mathsf{T}}\Sigma^{-1}(\mathbf{y}-\mathbf{X}\beta)/2)$$

With weighted regression a special case.

GLM's

Logistic:

$$p(\mathbf{y}|\mathbf{X},\ \boldsymbol{\beta}) = \prod_{i=1}^{n}[\text{invlogit}(X_i\boldsymbol{\beta})]^{y_i}[1-\text{invlogit}(X_i\boldsymbol{\beta})]^{1-y_i}$$

Poisson:

$$p(\mathbf{y}|\mathbf{X},\ \boldsymbol{\beta}) = \prod_{i=1}^{n}\exp(\ln(u_i) - \ln(y_i) + y_i\ln(X_i\boldsymbol{\beta}) - X_i\boldsymbol{\beta})$$

With non-constant variance. Find MLE's via iteratively re–weighted least squares.

18.2 Uncertainty and Likelihood

One parameter case:
With one parameter, we can plot the likelihood as a curve. MLE is position of peak. 2nd derivative is negative at top, and magnitude indicates the SE of the MLE.
Two parameters β_0, β_1.
Plot likelihood surface as a hill. MLE is position of peak. 2nd derivative matrix is negative definite, its negative estimates $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}$.

Simulation

Interpret output from classical regression as likelihood times flat prior = posterior. To sample from posterior,

- draw a random χ^2_{n-k} and divide it into $\hat{\sigma}^2(n-k)$ to get an inverse scaled χ^2.
- Sample a $\boldsymbol{\beta}$ vector from $N(\hat{\boldsymbol{\beta}}, \sigma_s^2(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1})$
- Sample predicted y's as needed.

Informed Prior

Prior can act as additional data point.
Prior: $\beta_2 \sim N(5, .25)$
Tack on equation: $5 = \beta_2$ by appending 5 to $\mathbf{y}$, 0 to intercept column, and 1 to $\mathbf{x}_2$. Use weights: 1 for data (n times) and $\sigma_y^2/.25$ for the added point.
In normal (prior) – normal (likelihood) cases, the posterior is also normal with precision = sum of prior and likelihood precisions, and mean the weighted average of prior mean and sample mean.
If it's hard to invert $\mathbf{X}^T\mathbf{X}$, then adding $\mu_0^2\sigma_y^2/sigma_{prior}^2$ to bottom right corner avoids singularity.

Multilevel Models

Log radon in county j: $y_i \sim N(\alpha_{j[i]}, \sigma_y^2), i = 1, \ldots, n$
county mean: $\alpha_j \sim (\mu_a, \sigma_\alpha^2), j = 1 \ldots, J$.
Complete pooling: all $\hat{\alpha}_j = \hat{\mu_\alpha}$
No pooling: $\hat{\alpha}_j = \overline{y}_j$
Augmented Equations:

$$\mathbf{y}_* = \begin{bmatrix} \mathbf{y} \\ \mu_\beta \end{bmatrix}, \quad \mathbf{X}_* = \begin{bmatrix} \mathbf{X} \\ \mathbf{I}_k \end{bmatrix}, \quad \mathbf{W}_* = \begin{bmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}$$