STAT 431 — Applied Bayesian Analysis — Course Notes

Generalized Linear Models

- Introduction
- Predictor Centering
- GLM Regression
  - i) Logistic "
  - ii) Poisson Loglinear regression
  - . Example : Shark
  - . JAGS code
  - . Notes

Fall 2022

Introduction

Consider regression of a response Y (random) on a predictor X (fixed).

Linear regression is well suited to cases where Y (conditionally on X) has a normal distribution (perhaps after transformation).

What if Y is clearly not normal (nor transformable to normal)?

For example Y may be a relatively small count.

Predictor Centering

It is customary to **center** the predictor, i.e., to use

$$X_i^{\mathsf{cent}} \;=\; X_i - \bar{X}$$

where $\bar{X}$ is its sample mean.

One advantage: May improve Gibbs sampler mixing (because regression coefficients are less correlated).

GLM Regression

Idea: Express a mean-related parameter of the model distribution of Y as a (transformed) linear regression on X.

Eg: Logistic Regression

$$Y_i \mid \theta_i \quad \sim \quad \text{Binomial}(n_i, \theta_i)$$

$$\ln\left(\frac{\theta_i}{1-\theta_i}\right) \quad = \quad \text{logit}(\theta_i) \quad = \quad \beta_1 \, + \, (X_i - \bar{X})\beta_2$$

Eg: Poisson Loglinear Regression

$$
\begin{aligned}
Y_i \mid \lambda_i &\sim \text{Poisson}(\lambda_i) \\
\ln(\lambda_i) &= \beta_1 + (X_i - \bar{X})\beta_2
\end{aligned}
$$

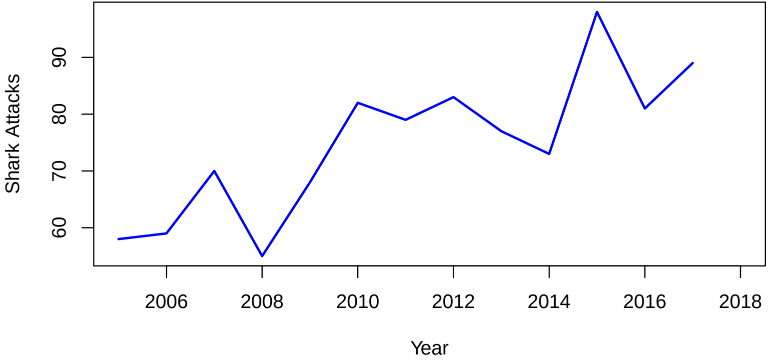
Priors on β_1 and β_2 can be specified similarly to linear regression, e.g.

$$
\beta_1, \beta_2 \sim 1\, d\beta_1\, d\beta_2
$$

Example: Shark Attacks

$$Y_i \quad = \quad \text{number of shark attacks (worldwide)}$$

$$X_i \quad = \quad \text{year (2005–2017)}$$



- Are shark attacks becoming more frequent?
- How many were predicted for 2018? (Actual: 68)

Since attacks are "rare" and usually unrelated, suppose

$$Y_i \mid \lambda_i \;\sim\; indep\ \mathrm{Poisson}(\lambda_i)$$

$$\ln(\lambda_i) \;=\; \beta_1 + (X_i - \bar{X})\beta_2$$

We will choose “vague” but proper priors:

$$\beta_1, \beta_2 \quad \sim \quad indep \ \mathrm{Normal}(0, 100^2)$$

[Draw preliminary model graph ...]

The data:

```
x     y
2005  58
2006  59
2007  70
2008  55
2009  68
2010  82
2011  79
2012  83
2013  77
2014  73
2015  98
2016  81
2017  89
2018  NA
```

Note the response of `NA` for the year 2018.

The "missing" Y value for 2018 will be sampled as an unobserved random node, to give its posterior predictive distribution.

The JAGS code:

```
data {
  xmean <- mean(x[1:(length(x)-1)])
  for(i in 1:length(x)) {
    xcent[i] <- x[i] - xmean
  }
}
model {
  for(i in 1:length(y)) {
    y[i] ~ dpois(lambda[i])
    log(lambda[i]) <- beta1 + beta2 * xcent[i]
  }

  beta1 ~ dnorm(0, 0.0001)
  beta2 ~ dnorm(0, 0.0001)

  beta2.gt.0 <- beta2 > 0
}
```

Notes:

- To get the centered version of X (`xcent`), we subtract `xmean`.

  Only the observed cases (`1:(length(x)-1)`) are used in the mean.

- We define `beta2.gt.0` so we can approximate the posterior probability that $\beta_2 > 0$.

R/JAGS Example 4.1:

Poisson Regression