Discrete Response Model Lecture 5

Models for Count Response, Discrete Response Model Evaluation, and Model Selection

datascience@berkeley

Parameter Estimation and Inference

Maximum Likelihood

Maximum likelihood estimation is used again to find the MLEs. Suppose my sample is denoted as (y_i, x_{i1}, …, x_{ip}) with i = 1, …, n. The likelihood function is

$$L(\beta_0, \quad , \beta_p \mid y_1, \quad , y_n) = \prod_{i=1}^{n} \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

where $\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \quad + \beta_p x_{ip})$.

For most situations, the likelihood function needs to be maximized using iterative numerical procedures. The glm() function in R completes this maximization where the family argument needs to be given as poisson(link = log).

- The covariance matrix for the parameter estimators follows from using standard likelihood procedures as outlined in the book's Appendix B.

- Wald and LR-based inference methods are performed in the same ways as for likelihood procedures in earlier weeks

Example: Horseshoe Crabs

The purpose of this example is to determine if the shell width of a female (x) is related to the number of satellites (Y) she has around her.

$$\log(\mu) = \beta_0 + \beta_1 x$$

where

Y = Number of satellites
x = Shell width of female (measured in cm)

can be used to estimate the mean number of satellites given a shell width.

```
> crab <- read.csv(file  = "HorseshoeCrabs.csv")
> str(crab)
'data.frame':   173 obs. of  5 variables:
 $ Color : int  2 3 3 4 2 1 4 2 2 2 ...
 $ Spine : int  3 3 3 2 3 2 3 3 1 3 ...
 $ Width : num  28.3 26 25.6 21 29 25 26.2 24.9 25.7 27.5 ...
 $ Weight: num  3.05 2.6 2.15 1.85 3 2.3 1.3 2.1 2 3.15 ...
 $ Sat   : int  8 4 0 0 1 3 0 0 8 6 ...
> head(crab)
  Color Spine Width Weight Sat
1     2     3  28.3   3.05   8
2     3     3  26.0   2.60   4
3     3     3  25.6   2.15   0
4     4     2  21.0   1.85   0
5     2     3  29.0   3.00   1
6     1     2  25.0   2.30   3
```

Model Estimation and Estimation Results

```
> mod.fit<-glm(formula = Sat ~ Width, data = crab,
+              family = poisson(link = log))
> summary(mod.fit)

Call:
glm(formula = Sat ~ Width, family = poisson(link = log), data = crab)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8526  -1.9884  -0.4933   1.0970   4.9221

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.30476    0.54224  -6.095  1.1e-09 ***
Width        0.16405    0.01997   8.216  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 632.79  on 172  degrees of freedom
Residual deviance: 567.88  on 171  degrees of freedom
AIC: 927.18

Number of Fisher Scoring iterations: 6
```

The estimated Poisson regression model is

$$\hat{\mu} = \exp(-3.3048 + 0.1640x)$$

The model could also be written as:

$$\log(\hat{\mu}) = -3.3048 + 0.1640x$$

Berkeley SCHOOL OF INFORMATION