Homework 4

Noah Johnson

March 7, 2018

Bolstad 5.8

Let Y have the Poisson($\mu = 3$) distribution

(a) Calculate P(Y=2).

$$P(Y = 2|\mu = 3) = \frac{3^2 e^{-3}}{2!} \approx 0.224$$

Or alternatively, we can use R.

```
dpois(x = 2, lambda = 3)
```

```
## [1] 0.2240418
```

(b) Calculate P(Y $\leq$ 2).

```
ppois(2, 3)
```

```
## [1] 0.4231901
```

(c) Calculate P(1 $\leq$ Y $<$ 4).

This is the same as $P(1 \leq Y \leq 3) = P(Y \leq 3) - P(Y = 0)$.

```
ppois(3, 3) - dpois(0, 3)
```

```
## [1] 0.5974448
```

BYSH 4.15.1 Conceptual Exercise 1

What are features of inferential OLS models that make them less suitable for count data?

We could try to fit a simple linear model to the expected mean, such as $\lambda_{X_i} = \beta_0 + \beta_1 X_i$, and then fit this model by minimizing the residual sum of squares $\sum_i (Y_i - \lambda_{X_i})^2$.

But count data can never be negative, which this simple linear model allows.

OLS assumes equal variance over all random variable responses. However, Poisson random variables have their variance increase as the mean rate λ increases.

BYSH 4.15.1 Conceptual Exercise 2

Models of the form $Y_i = \beta_0 + \beta_1 X_i, \epsilon \sim iidN(0, \sigma)$ are fit using the method of least squares [note that least squares estimates for linear regression are also MLEs]. What method is used to fit Poisson regression models?

Poisson regression models of the form $log(\lambda_{X_i}) = \beta_0 + \beta_1 X_i$ are fit using the method of maximum likelihood estimation. The joint probability of the observed data occurring is simply the product of each Poisson random variable, since the observations are assumed to be unique. The log of this likelihood is taken, and numerical methods for convex optimization such as gradient descent are used to maximize this function with respect to the beta coefficients.

This is equivalent to choosing the generating model which maximizes the probability of observing the data we did.

BYSH 4.15.2 Guided Exercise 1 Elephant Mating

How does age affect male elephant mating patterns? An article by Poole (1989) investigated whether mating success in male elephants increases with age and whether there is a peak age for mating success. To address this question, the research team followed 41 elephants for one year and recorded both their ages and their number of matings. The data is found in elephant.csv, and relevant R code can be found under elephantMating.R.

The variables are:
MATINGS: the number of matings in a given year
AGE: the age of the elephant in years.

(a) Create a histogram of MATINGS. Is there preliminary evidence that number of matings could be modeled as a Poisson response? Explain.
(b) Plot MATINGS by AGE. Add a least squares line. Is there evidence that modeling matings using a linear regression with age might not be appropriate? Explain.
(c) For each age, calculate the mean number of matings. Take the log of each mean and plot it by AGE.
i. What assumption can be assessed with this plot?
ii. Is there evidence of a quadratic trend on this plot?
(d) Fit a Poisson regression model with a linear term for AGE. Interpret the coefficient for AGE. Exponentiate and interpret the result.
(e) Construct a 95% confidence interval for the slope and interpret in context (you may want to exponentiate endpoints).