Stat 201 - Generalized Linear Models

Mandatory Assignment - due on October 8th 2021

1. A sample of n i.i.d. observations has been drawn from a Bernoulli distribution with parameter p:

$$y_i \sim p^y(1-p)^{1-y}.$$

Let $\eta = \log \frac{p}{1-p}$ be the parameter of interest.

(a) Obtain the MLE $\hat{\eta}$ of η ?

(b) Obtain the (asymptotic) sampling distribution of $\hat{\eta}$?

(c) Obtain a confidence interval for $\hat{\eta}$, with level $1-\alpha$

2. Load the dataset `gamma regression data.txt` from the course web page. Assuming the following model Almost same as group, OA3

$$y_i \sim f(y_i;\beta_i) = \beta_i^2 y_i e^{-\beta_i y_i}$$
$$\log \mu_i = \log \mathbb{E} y_i = \beta_1 + \beta_2 x_i$$

(a) explain why this is a GLM

(b) obtain the IWLS iteration for estimating the regression coefficients

(c) provide an R code that implements the iteration

(d) initialize the iteration at $\beta_1 = \beta_2 = 0.5$ and provide the sequence of the first 10 estimates

(e) draw a plot where the estimated regression function in overlapped on the data

(f) obtain the standard errors of both regression coefficients

3. Load the dataset `pollution.txt` from the course web page. It includes daily airquality measurements in New York, May to September 1973: ozone concentrations (ppb), solar radiation (lang), temperature (degrees F) and classes of wind speed (mph).

(a) Fit a linear model that estimates the effect of solar radiation, temperature and wind speed on ozone concentration and interpret the results (do not include interactions).

(b) Check the normality assumption

(c) Check the presence of outliers

(d) Using the estimated model, predict ozone concentration at solar radiation = 100, temperature = 70 and wind speed between 1.7 and 7.4 mph and provide a 95% confidence interval for such prevision

(e) Can you improve the model by including significant interaction effects?