Stat 201 - Generalized Linear Models

Additional exercises

1. A sample of n i.i.d. observations has been drawn from a Poisson distribution with parameter λ:
$$y_i \sim \frac{\lambda^y e^{-\lambda}}{y!}$$
Let $\eta = \log \lambda$ 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. A sample of n i.i.d. observations has been drawn from a Normal distribution with known mean μ and unknown variance σ^2 Let $\eta = \sigma^2$ 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 $1 - \alpha$

3. Do exercise 5.3 of the DB book.

4. Load the dataset `binary.txt` from the course web page. It includes a binary variable that indicates whether an undergraduate is admitted to a graduate school and three covariates: GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution. Assuming the following model
$$\begin{aligned} y_i \sim & f(y_i; p_i) = p_i^{y_i}(1-p_i)^{1-y_i} \\ \log \frac{\mu_i}{1-\mu_i} = & \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + \beta_4 x_{3i} \end{aligned}$$
where $\mu_i = \mathbb{E} y_i$,
   (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) run the R code and find the MLE of the parameters β, by appropriately initializing the iteration
   (e) provide the standard errors of the the parameters

5. Load the data `PLOS` from the `Dobson` library. Consider the family of 8 linear models
$$y = \beta_0 + \beta_1 f(x)$$
where
$$f(x) = \begin{cases} x^p & p = -2, -1, -0.5, 0.5, 1, 2, 3 \\ log(x) \end{cases}$$
where y is the title lenght and x is the number of authors.
   (a) Select the best model and explain your selection
   (b) Overlap the predicted model on a plot of the data and interpret the results