Stat 616 Generalized Linear Models, HW #1 **Spring 2019**

Due: Thursday, January 31

- *I put a few hints on page 3, in case you really need them. Use them at your own risk.*
- *Review the concepts of basic statistical inference: confidence interval (the meaning of CI, confidence level, common form of CIs, etc.) and test of significance (set up of null and alternative hypothesis, the meaning of test statistic, p-values, etc.)*

1. Dobson & Barnett, Exercise 3.1. (p. 55)

   3.1 The following relationships can be described by generalized linear models. For each one, identify the response variable and the explanatory variables, select a probability distribution for the response (justifying your choice) and write down the linear component.

   (a) The effect of age, sex, height, mean daily food intake and mean daily energy expenditure on a person's weight.

   (b) The proportions of laboratory mice that became infected after exposure to bacteria when five different exposure levels are used and 20 mice are exposed at each level.

   (c) The relationship between the number of trips per week to the supermarket for a household and the number of people in the household, the household income and the distance to the supermarket.

2. Dobson & Barnett, Exercise 3.2. (p. 56)

   Skip the "find E(Y) and var(Y)" part. Also, note that β is the *rate parameter* in this pdf.

   3.2 If the random variable Y has the **Gamma distribution** with a scale parameter β, which is the parameter of interest, and a known shape parameter α, then its probability density function is

$$f(y;\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta}.$$

   Show that this distribution belongs to the exponential family and find the natural parameter. ~~Also using results in this chapter, find $E(Y)$ and $var(Y)$~~.

3. Dobson & Barnett, Exercise 3.7.(a, b, c, d) (p. 56)

   In part (d), let $\mathbf{x}^T\boldsymbol{\beta} = \beta_1 + \beta_2 x$, if you are not comfortable with the matrix notation.

   3.7 Consider N independent binary random variables $Y_1, \ldots, Y_N$ with

$$P(Y_i = 1) = \pi_i \text{ and } P(Y_i = 0) = 1 - \pi_i.$$

   The probability function of Y_i, the Bernoulli distribution $\mathrm{B}(\pi)$, can be written as

$$\pi_i^{y_i} (1 - \pi_i)^{1-y_i},$$

   where $y_i = 0$ or 1.

   (a) Show that this probability function belongs to the exponential family of distributions.

   (b) Show that the natural parameter is

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right).$$

   This function, the logarithm of the **odds** $\pi_i/(1 - \pi_i)$, is called the **logit** function.

   (c) Show that $\mathrm{E}(Y_i) = \pi_i$.

   (d) If the link function is

$$g(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = \mathbf{x}^T\boldsymbol{\beta},$$

   show that this is equivalent to modelling the probability π as

$$\pi = \frac{e^{\mathbf{x}^T\boldsymbol{\beta}}}{1 + e^{\mathbf{x}^T\boldsymbol{\beta}}}.$$

4. Dobson & Barnett, Exercise 3.9. (p. 56)

   Refer to Exercise 3.3(a) for Pareto distribution.

   3.9 Suppose $Y_1, \ldots, Y_N$ are independent random variables each with the Pareto distribution and

$$\mathrm{E}(Y_i) = (\beta_0 + \beta_1 x_i)^2.$$

   Is this a generalized linear model? Give reasons for your answer.

   3.3 Show that the following probability density functions belong to the exponential family:

   (a) Pareto distribution $f(y; \theta) = \theta y^{-\theta-1}$.

This is the end of HW 1

Hints:

1. Dobson & Barnett, Exercise 3.1. (p. 55)

   Choose from the following probability distributions for the response variables in this set of problems: Binomial, Normal, and Poisson. If you are unfamiliar with these distributions, read Section 3.2 from the textbook (p.46). You can also find more information about these distributions online.

2. Dobson & Barnett, Exercise 3.2. (p. 56)

   If you did not write it down in your notes, you can find the meaning of “natural parameter” in Section 3.2 from the textbook (p.46).

3. Dobson & Barnett, Exercise 3.7. (p. 56)

   No hint.

4. Dobson & Barnett, Exercise 3.9. (p. 56)

   To show a GLM, verify the distribution belongs to the Exponential Family with canonical form, state the link function, and show the linear function.