BIOSTAT 651
Homework #2
due: Monday, Feb 6

- turn in at the start of class

- each sub-question=2 points; total 20 points

1. Suppose that Y follows the following density (Inverse Gaussian distribution):

$$f(Y;\mu,\lambda) = \left[\frac{\lambda}{2\pi Y^3}\right]^{1/2} \exp\frac{-\lambda(Y-\mu)^2}{2\mu^2 Y} \qquad \lambda > 0, \mu > 0, Y > 0$$

where λ is a nuisance parameter.

(a) Show that f is an exponential family in canonical form. Show $t(Y)$, $a(\phi)$, θ, $b(\theta)$ and $c(Y,\phi)$. Keep in mind that $a(\phi)$ should be positive (i.e. $a(\phi) > 0$).

(b) Using the properties of the exponential family, find $E(Y)$ and $Var(Y)$.

2. Suppose that Y_i has the following density function (Gamma distribution),

$$f(Y_i;\lambda,\nu) \quad = \quad \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\lambda}\right)^{\nu} Y_i^{\nu-1} \exp\left\{\frac{-\nu}{\lambda}Y_i\right\} \qquad Y_i > 0, \lambda > 0, \nu > 0$$

for $i = 1,\ldots,n$, where ν is a (known) nuisance parameter.

(a) Suppose that data $(\mathbf{x}_i, Y_i), i = 1,\ldots,n$, are observed. Researchers want to use GLM (Gamma regression) with the canonical link $g(.)$, i.e., $g(\mu_i) = \mathbf{x}_i^T\boldsymbol{\beta}$, where $\mathbf{x}_i$ is the vector of covariates. Derive the canonical link $g(\cdot)$, score function, $U(\boldsymbol{\beta})$, and expected information, $I(\boldsymbol{\beta})$.

(b) Sketch out an Iteratively Re-Weighted Least Squares (IRWLS) algorithm for estimating $\boldsymbol{\beta}$.

(c) Suppose we observed the data in HW2.xls and decided to use the following model:

$$g(\mu_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2,$$

where $g(\cdot)$ is the canonical link function obtained in (a). Using IRWLS, estimate $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$. Please use the initial value $\boldsymbol{\beta}_0 = (-1, -1, -1)$. The nuisance parameter $\nu = 3$. (Keep in mind $a(\phi) > 0$)

(d) Obtain the 95% confidence intervals for β_1 and β_2

(e) Carry out the likelihood ratio test for $H_0 : \beta_1 = \beta_2 = 0$.

3. Solve 4.1 (a) (b) on page 69 of the textbook (Dobson) and the following question (c). You can use either SAS or R for plots and GLM fitting.

(c) Fit the following generalized linear model to these data using the Poisson distribution,

$$\log(\lambda_i) = \beta_0 + \beta_1 T_i,$$

where $T_i = \log(i) - \log(10)$. Estimate and interpret β_0 and β_1.