MSc: Statistical Analysis (Spring Term)

Generalized Linear Models: Exercises 1

1. The random variable Y is said to be from the Exponential distribution with parameter $\lambda > 0$, i.e. $Exp(\lambda)$, if it has probability density function (p.d.f.)

$$f(y) = \lambda e^{-\lambda y}, \quad y > 0.$$

   (a) By working directly with the above p.d.f., show that the $Exp(\lambda)$ distribution is a member of the *generalized exponential family of distributions.*

   (b) Further show that the canonical link function can be taken to be $g(\mu) = -1/\mu$.

2. Consider the r.v. Y to have a Gamma distribution, with 'familiar' parameters λ and ν. i.e. $Y \sim G(\lambda, \nu)$, with p.d.f.

$$f(y) = \frac{1}{\Gamma(\nu)} \lambda^{\nu} y^{\nu-1} \exp(-\lambda y), \quad y \geq 0; \quad \lambda, \nu > 0.$$

   In this form

$$E[Y] = \frac{\nu}{\lambda} \quad \text{and} \quad \text{Var}[Y] = \frac{\nu}{\lambda^2}$$

   [Notes: If $\nu = 1$ this reduces to $f(y) = \lambda e^{-\lambda y}$, which is the exponential distribution.

   Recall that the time between events in a Poisson process $P(\lambda)$ follows an exponential distribution with $\alpha = \lambda$. If $\nu = n$, then the time to the nth event in a Poisson process $P(\lambda)$ follows the Gamma distribution $G(\lambda, n)$.

   The Gamma distribution also has the chi-squared distribution ($\chi^2(k)$) as a special case, by setting $\alpha = \frac{1}{2}$ and $\nu = \frac{k}{2}$.]

   (a) By considering the alternative parametrization of $Y \sim G(\mu, \nu)$, where $\mu = \nu/\lambda$, show that the Gamma distribution, with p.d.f.

$$f(y) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} y^{\nu-1} \exp\left(-\frac{\nu y}{\mu}\right), \quad y \geq 0; \quad \mu, \nu > 0.$$

   is a member of the *generalized exponential family of distributions.*

   (b) What are the mean and variance of Y in this alternative parametrization? What is the variance-mean relationship? Show that as the mean changes (perhaps as a function of explanatory variables) the coefficient of variation remains constant.

   (c) Further show that the canonical link function can be taken to be $g(\mu) = -1/\mu$.

3. Suppose that we have observations on a continuous non-negative response variable Y with support $[0, \infty)$ and a continuous non-negative explanatory variable X. In this exercise we explore some of the non-linear relationships between $E[Y] = \mu$ and values x of X that can be modelled using an appropriate combination of link function g and linear predictor η. In each case either sketch the relationship, or use R to plot it (or a special case of it). Note that these relationships involve asymptotes.

   (a) Suppose that $g(\mu) = 1/\mu$ and consider the following linear predictors:

      (i) $\eta_i = \beta_0 + \beta_1 x_i$, where $\beta_0, \beta_1 > 0$. Note that this gives an inverse linear model for μ_i, allowing an asymptote as $x_i \to \infty$.

      (ii) $\eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$, where $\beta_0, \beta_2 > 0$, $\beta_1 < 0$ and $\beta_1^2 < 4\beta_0\beta_2$. This is an inverse quadratic model for μ_i, also with an asymptote as $x_i \to \infty$.

      (iii) $\eta_i = \beta_0 x_i + \beta_1 + \beta_2 \dfrac{1}{x_i}$, where $\beta_0, \beta_2 > 0$, $\beta_1 < 0$ and $\beta_1^2 < 4\beta_0\beta_2$.

      (iv) $\eta_i = \beta_0 + \beta_1 \dfrac{1}{x_i}$, where $\beta_0, \beta_1 > 0$.

   (b) Now let $g(\mu) = \log(\mu)$ and consider a linear predictor of the following general form:

   $$\eta_i = \beta_0 + \beta_1 x_i + \beta_2 \frac{1}{x_i}.$$

   Set $\beta_0 = 1$ and plot the relationship between μ_i and x_i corresponding to the following combinations of (β_1, β_2):

   (i) $(\beta_1, \beta_2) = (1, 1)$, (ii) $(\beta_1, \beta_2) = (-1, -1)$, (iii) $(\beta_1, \beta_2) = (1, -1)$, (iv) $(\beta_1, \beta_2) = (-1, 1)$.