Homework 7

This problem is about generalized linear models involving the gamma distribution, which is useful for modeling non-negative random variables. There are several equivalent parametrizations of the gamma density. One of the common ones is

\[
f(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y}
\]

Then

\[
E(Y) = \frac{\alpha}{\lambda} \quad Var(Y) = \frac{\alpha}{\lambda^2}
\]

Sometimes it is more convenient to parametrize the distribution in terms of its mean \(\mu = \alpha/\lambda\). Then

\[
f(y) = \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\alpha y/\mu}
\]

Then \(E(Y) = \mu\) and \(Var(Y) = \mu^2/\alpha\). The ratio of the standard deviation to the mean is called the “coefficient of variation.” In this case the coefficient of variation is \(\alpha^{-1/2}\).

For the classical linear model the standard assumption is that the variance is constant and the mean changes as a function of covariates. The analogue for a generalized linear model based on a gamma is that the coefficient of variation is constant while the mean changes; that is, the standard deviation is proportional to the mean.

Suppose then that the mean, \(\mu\), is modelled as a function of \(x \beta\), \(g(\mu) = x \beta\) and it is assumed that \(\alpha\) is constant.

1. Show that the gamma distribution belongs to the exponential family. Identify \(\theta\), \(b(\theta)\), etc.

\[
\log f(y) = -\log \Gamma(\alpha) + \alpha \log \alpha - \alpha \log \mu + (\alpha - 1) \log y - \alpha \log \mu
\]

\[
= -\alpha(y/\mu + \log \mu) + (\alpha - 1) \log y - \alpha \log \alpha - \log \Gamma(\alpha)
\]

\[
= \frac{y\theta - \log \theta}{a(\phi)} + C(y, \phi)
\]

where \(\theta = \mu^{-1}\) and \(a(\phi) = -\alpha^{-1}\). This is in the canonical form of the exponential family, where \(b(\theta) = \log \theta\). Also, note that

\[
b'(\theta) = \theta^{-1} = \mu
\]

\[
b''(\theta) a(\phi) = \alpha^{-1} \mu^2 = Var(Y)
\]
2. Consider the canonical link function, \( g(\mu) = \mu^{-1} \). Suppose that the observations are \((Y_i, x_i), i = 1, \ldots, n\), and the \( Y_i \) are independent.


\[
\ell(\beta) = \sum_{i=1}^{n} \ell_i(\beta) = -\alpha n \sum_{i=1}^{n} (Y_i x_i \beta - \log x_i \beta) + \sum_{i=1}^{n} c(Y_i, \alpha)
\]

4. Using (3) show that the maximizing \( \beta \) does not depend on \( \alpha \), and that the maximum likelihood estimate of \( \beta \) satisfies a system of equations of the form, \( X^T(Y - \mu(\beta)) = 0 \).

\[
\frac{\partial \ell(\beta)}{\partial \beta_j} = -\alpha \sum_{i=1}^{n} (x_{ij}Y_i - x_{ij}/x_i \beta)
\]

\[
= -\alpha \sum_{i=1}^{n} x_{ij}(Y_i - \mu_i(\beta))
\]

where \( \mu_i(\beta) = E(Y_i) \). Writing this in vector form we have

\[
\nabla \ell(\beta) = -\alpha X^T(Y - \mu(\beta))
\]

so the maximum likelihood estimate of \( \beta \) satisfies \( X^T(Y - \mu(\beta)) = 0 \) and does not depend on \( \alpha \).

5. Derive the form of the update \( h^{(k)} = \beta^{(k+1)} - \beta^{(k)} \).

\[
\frac{\partial^2 \ell(\beta)}{\partial \beta_k \partial \beta_j} = -\alpha \sum_{i=1}^{n} \frac{x_{ij}x_{ij}}{(x_i \beta)^2}
\]

\[
= -\alpha \sum_{i=1}^{n} x_{ji}x_{ik}V^{-1}_i
\]

In matrix form \( \nabla^2 \ell(\beta) = -\alpha X^T V^{-1} X \)

\[
h^{(k)} = -[\nabla^2 \ell(\beta)]^{-1} \nabla \ell(\beta)
\]

\[
= -(X^T V^{-1} X)^{-1} X^T (Y - \mu(\beta))
\]
6. Give the form of the IRLS algorithm and identify the adjusted dependent variable. Show that it is equivalent to (5).

\[ \beta^{(k+1)} = \beta^{(k)} + h^{(k)} \]

\[ = (X^T V^{-1} X)^{-1} X^T V^{-1} (X \beta^{(k)} - V(Y - \mu)) \]

Thus the adjusted variables are

\[ Z = X \beta - V(Y - \mu) \]

or

\[ Z_i = x_i \beta - V_i (Y_i - \mu_i) \]

This can be interpreted as a Taylor series approximation to \( g(Y_i) \) where \( g(\mu) = \mu^{-1} = x_i \beta \) is the link function,

\[ g(Y_i) \approx x_i \beta - (Y_i - \mu)(x_i \beta)^2 \]

Then, since \( \text{Var}(Y_i) = \alpha^{-1}(x_i \beta)^{-2} \)

\[ \text{Var}(g(Y_i)) \approx \alpha^{-1} V_i \]

and the IRLS form above can be interpreted as weighted least squares on the adjusted variables.

7. What equation does the maximum likelihood estimate of \( \alpha \) satisfy? Since the maximum likelihood estimate of \( \beta \) does not depend on \( \alpha \), the log likelihood of \( \alpha \) is

\[ \ell(\alpha) = -\alpha \sum_{i=1}^{n} (Y_i x_i \hat{\beta} - \log x_i \hat{\beta}) + \sum_{i=1}^{n} C(Y_i, \alpha) \]

This would have to be maximized by an iterative process.

8. An un-natural aspect of the link function above is that \( x \beta \) must be positive. For this and other reasons it may be desirable to consider a link function which automatically guarantees that \( g(x \beta) > 0 \). For the choice \( g(\mu) = \exp(x \beta) \), explain how IRLS can be used to estimate \( \beta \).

There was a typo. I intended to write \( \mu = \exp(x \beta) \), so \( g(\mu) = \log(\mu) \).
Dropping the subscript $i$, the contribution to the total log likelihood from the $i$th observation is (neglecting the constant $C(Y, \alpha)$)

$$\ell(\beta) = -\alpha(Y_i\theta(\beta) - b(\theta(\beta)))$$

The derivative with respect to $\beta_j$ is most easily expressed by using the chain rule:

$$\frac{\partial \ell}{\partial \beta_j} = \frac{\partial \ell}{\partial \theta} \frac{\partial \theta}{\partial \mu} \frac{\partial \mu}{\partial \eta} \frac{\partial \eta}{\partial \beta_j}$$

where $\eta = g(\mu) = x\beta$. Then

$$\frac{\partial \ell}{\partial \theta} = -\alpha(Y - b'(\theta))$$
$$\frac{\partial \theta}{\partial \mu} = \frac{1}{\partial \mu/\partial \theta} = \frac{1}{b''(\theta)}$$
$$\frac{\partial \mu}{\partial \eta} = e^\eta$$
$$\frac{\partial \eta}{\partial \beta_j} = x_j$$

Since

$$b''(\theta) = -\frac{1}{\theta^2}$$
$$= -\mu^2$$
$$= -e^{2\eta}$$

$$\frac{\partial \ell}{\partial \beta_j} = \alpha(Y - \mu)e^{-x\beta_j}x_j$$

Thus

$$\frac{\partial \ell}{\partial \beta_j} = \alpha \sum_{i=1}^n (Y_i - \mu_i)e^{-x_i\beta_j}x_{ij}$$

And the maximum likelihood estimate of $\beta$ satisfies

$$\nabla(\beta) = X^T D(Y - \mu(\beta)) = 0$$

where $D = \text{diag}(e^{-x_i\beta})$. To solve for the maximum likelihood estimate using IRLS, form the adjusted variables

$$Z_i = g(\mu_i) + (Y_i - \mu_i)g'(\mu_i)$$
$$= \log \mu_i + (Y_i - \mu_i)/\mu_i$$
which have variance

\[ \text{Var}(Z_i) = \frac{\text{Var}(Y_i)}{\mu_i^2} = \alpha^{-1} \]

Since this is constant, the iteration is

\[ \beta^{(k+1)} = (X^T X)^{-1} X^T Z^{(k)} \]