Biost/Stat 533
Emerson, Spr 09
Homework #6 Key
June 1, 2009

Written problems to be handed in Friday, June 5 at 9 am.

All problems consider the general parametric regression model in which we observe pairs $(Y_i, \vec{X}_i)$ for $i = 1, \ldots, n$ in which

$$Y_i \mid \vec{X}_i \;\sim\; f_Y(y; \theta_i) \qquad \text{with} \qquad g(\theta_i) = \vec{X}_i^T \vec{\beta},$$

with the Y_i's mutually independent, $\vec{X}_i = (1, X_{i1}, X_{i2}, \ldots, X_{ip}$ known covariates, and $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ a $p+1$ vector to be estimated and/or tested.

1. Suppose the link function g is the identity function $g(x) = x$. Find the score equations for the following choices of f_Y and θ_i.

Ans: Some preliminaries common to all parts of this problem:

For notational convenience, I define $X_{i0} = 1$.

Because I am concerned only with cases in which the Y_i's are independent, I note that the likelihood can be written as

$$L(\vec{\beta} \,|\, \vec{Y}) = \prod_{i=1}^{n} f(Y_i \,;\, \theta_i),$$

the log likelihood can be written as

$$\mathcal{L}(\vec{\beta}) = \log(L(\vec{\beta} \,|\, \vec{Y})) = \sum_{i=1}^{n} \log(f(Y_i \,;\, \theta_i)),$$

and the score functions can be written as

$$\mathcal{U}_j(\vec{\beta}) = \frac{\partial}{\partial \beta_j} \mathcal{L}(\vec{\beta}) = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_j} \log(f(Y_i \,;\, \theta_i)) = \sum_{i=1}^{n} \left[\frac{\partial}{\partial \theta_i} \log(f(Y_i \,;\, \theta_i)) \; \frac{\partial}{\partial \beta_j} \theta_i \right].$$

Now our regression model has $g(\theta_i) = \eta_i = \vec{X}_i^T \vec{\beta}$, so

$$\frac{\partial}{\partial \beta_j} \theta_i = \frac{d}{dx} g^{-}1(x) \Big|_{x = \vec{X}_i^T \vec{\beta}} X_{ij}.$$

In problem 1, we use the identity link, $g(x) = x$, so $g^{-1}(x) = x$, and

$$\frac{\partial}{\partial \beta_j} \theta_i = \frac{d}{dx} g^{-}1(x) \Big|_{x = \vec{X}_i^T \vec{\beta}} X_{ij} = X_{ij}.$$

Hence

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[X_{ij} \; \frac{\partial}{\partial \theta_i} \log(f(Y_i \,;\, \theta_i)) \right].$$

a. Bernoulli: $Y_i \;\sim\; \mathcal{B}(1, p_i)$ and $\theta_i = p_i$.

Ans: In order to use the general results defined above, I find

$$\begin{aligned} \log(f(Y_i \;\; \theta_i = p_i) &= \log\left(\; p_i^{Y_i} (1 - p_i)^{1 - Y_i} \; \right) \\ &= Y_i \log(p_i) + (1 - Y_i) \log(1 - p_i) \\ \frac{\partial}{\partial \theta_i} \log(f(Y_i \;\; \theta_i = p_i) &= \frac{Y_i}{p_i} - \frac{1 - Y_i}{1 - p_i} = \frac{Y_i - p_i}{p_i(1 - p_i)}, \end{aligned}$$

so

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - p_i)}{p_i(1-p_i)} \; X_{ij} \right].$$

Note that in this model, $E(Y_i) = p_i$ and $Var(Y_i) = p_i(1-p_i)$, so the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - E[Y_i])}{Var(Y_i)} \; X_{ij} \right].$$

b. Poisson: $Y_i \sim \mathcal{P}(\lambda_i)$ and $\theta_i = \lambda_i$.

Ans: In order to use the general results defined above, I find

$$\begin{aligned}
\log(f(Y_i \;\; \theta_i = \lambda_i) &= \log\left(\frac{e^{-\lambda_i}\lambda_i^{Y_i}}{Y_i!} \right) \\
&= -\lambda_i + Y_i \log(\lambda_i) - \log(Y_i!) \\
\frac{\partial}{\partial\theta_i}\log(f(Y_i \;\; \theta_i = \lambda_i) &= -1 + \frac{Y_i}{\lambda_i} = \frac{Y_i - \lambda_i}{\lambda_i},
\end{aligned}$$

so

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - \lambda_i)}{\lambda_i} \; X_{ij} \right].$$

Note that in this model, $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i$, so the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - E[Y_i])}{Var(Y_i)} \; X_{ij} \right].$$

c. Exponential: $Y_i \sim \mathcal{E}(\lambda_i)$ and $\theta_i = \lambda_i$, where $E(Y_i) = \lambda_i$.

Ans: In order to use the general results defined above, I find

$$\begin{aligned}
\log(f(Y_i \;\; \theta_i = \lambda_i) &= \log\left(\frac{1}{\lambda_i} e^{-\frac{Y_i}{\lambda_i}} \right) \\
&= -\log(\lambda_i) - \frac{Y_i}{\lambda_i} \\
\frac{\partial}{\partial\theta_i}\log(f(Y_i \;\; \theta_i = \lambda_i) &= -\frac{1}{\lambda_i} + \frac{Y_i}{\lambda_i^2} = \frac{Y_i - \lambda_i}{\lambda_i^2},
\end{aligned}$$

so

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - \lambda_i)}{\lambda_i^2} \; X_{ij} \right].$$

Note that in this model, $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i^2$, so the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - E[Y_i])}{Var(Y_i)} \; X_{ij} \right].$$

d. Exponential: $Y_i \sim \mathcal{E}(\lambda_i)$ and $\theta_i = \lambda_i$, where $E(Y_i) = 1/\lambda_i$.

Ans: In order to use the general results defined above, I find

$$\begin{aligned}
\log(f(Y_i \;\; \theta_i = \lambda_i) &= \log\left(\lambda_i e^{-Y_i\lambda_i} \right) \\
&= \log(\lambda_i) - Y_i\lambda_i \\
\frac{\partial}{\partial\theta_i}\log(f(Y_i \;\; \theta_i = \lambda_i) &= \frac{1}{\lambda_i} - Y_i,
\end{aligned}$$

so

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\ -(Y_i - \frac{1}{\lambda_i}) \ X_{ij} \ \right].$$

Note that in this model, $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i^2$, so the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} [\ -(Y_i - E[Y_i]) \ X_{ij} \ ].$$

2. Repeat problem 1 with the specified link functions.

a. Bernoulli: $Y_i \sim \mathcal{B}(1, p_i)$ and $\theta_i = p_i$. Use link $g(x) = \text{logit}(x)$.

Ans: We have inverse link function

$$g^{-1}(x) = \text{expit}(x) = \frac{e^x}{1+e^x},$$

hence

$$\frac{d}{dx} g^{-1}(x) = \frac{e^x}{1+e^x} - \left(\frac{e^x}{1+e^x} \right)^2,$$

and

$$\frac{d}{dx} g^{-1}(x) \Big|_{x=\vec{X}_i^T \vec{\beta}} = p_i(1-p_i).$$

Using results from the preliminaries and part a of problem 1, we thus obtain

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - p_i)}{p_i(1-p_i)} \ p_i(1-p_i) X_{ij} \right] = \sum_{i=1}^{n} [\ (Y_i - p_i) \ X_{ij} \ ],$$

and the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} [\ (Y_i - E[Y_i]) \ X_{ij} \ ].$$

b. Poisson: $Y_i \sim \mathcal{P}(\lambda_i)$and $\theta_i = \lambda_i$. Use link $g(x) = \log(x)$.

Ans: We have inverse link function

$$g^{-1}(x) = \exp(x),$$

hence

$$\frac{d}{dx} g^{-1}(x) = \exp(x),$$

and

$$\frac{d}{dx} g^{-1}(x) \Big|_{x=\vec{X}_i^T \vec{\beta}} = \lambda_i.$$

Using results from the preliminaries and part b of problem 1, we thus obtain

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - \lambda_i)}{\lambda_i} \ \lambda_i X_{ij} \right] = \sum_{i=1}^{n} [\ (Y_i - \lambda_i) \ X_{ij} \ ],$$

and the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} [\ (Y_i - E[Y_i]) \ X_{ij} \ ].$$

c. Exponential: $Y_i \sim \mathcal{E}(\lambda_i)$ and $\theta_i = \lambda_i$, where $E(Y_i) = \lambda_i$. Use link $g(x) = \log(x)$.

Ans: We have inverse link function

$$g^{-1}(x) = \exp(x),$$

hence

$$\frac{d}{dx} g^{-1}(x) = \exp(x),$$

and

$$\frac{d}{dx} g^{-1}(x)\Big|_{x=\vec{X}_i^T\vec{\beta}} = \lambda_i.$$

Using results from the preliminaries and part c of problem 1, we thus obtain

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - \lambda_i)}{\lambda_i^2} \lambda_i X_{ij} \right] = \sum_{i=1}^{n} \left[\frac{(Y_i - \lambda_i)}{\lambda_i} X_{ij} \right].$$

and the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - E[Y_i])}{\sqrt{Var(Y_i)}} X_{ij} \right].$$

d. Exponential: $Y_i \sim \mathcal{E}(\lambda_i)$ and $\theta_i = \lambda_i$, where $E(Y_i) = 1/\lambda_i$. Use link $g(x) = \log(x)$.

Ans: We have inverse link function

$$g^{-1}(x) = \exp(x),$$

hence

$$\frac{d}{dx} g^{-1}(x) = \exp(x),$$

and

$$\frac{d}{dx} g^{-1}(x)\Big|_{x=\vec{X}_i^T\vec{\beta}} = \lambda_i.$$

Using results from the preliminaries and part d of problem 1, we thus obtain

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[-(Y_i - \frac{1}{\lambda_i}) \lambda_i X_{ij} \right] = \sum_{i=1}^{n} \left[-\frac{(Y_i - \frac{1}{\lambda_i})}{\frac{1}{\lambda_i}} X_{ij} \right],$$

and the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[-\frac{(Y_i - E[Y_i])}{\sqrt{Var(Y_i)}} X_{ij} \right].$$

e. Exponential: $Y_i \sim \mathcal{E}(\lambda_i)$ and $\theta_i = \lambda_i$, where $E(Y_i) = \lambda_i$. Use link $g(x) = 1/x$.

Ans: We have inverse link function

$$g^{-1}(x) = \frac{1}{x},$$

hence

$$\frac{d}{dx} g^{-1}(x) = -\frac{1}{x^2},$$

and

$$\frac{d}{dx} g^{-1}(x)\Big|_{x=\vec{X}_i^T\vec{\beta}} = \lambda_i^2.$$

Using results from the preliminaries and part c of problem 1, we thus obtain

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[\frac{(Y_i - \lambda_i)}{\lambda_i^2} \lambda_i^2 X_{ij} \right] = \sum_{i=1}^{n} [\, (Y_i - \lambda_i) \, X_{ij} \,] .$$

and the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} [\, (Y_i - E[Y_i]) \, X_{ij} \,] .$$

f. Exponential: $Y_i \sim \mathcal{E}(\lambda_i)$ and $\theta_i = \lambda_i$, where $E(Y_i) = 1/\lambda_i$. Use link $g(x) = 1/x$.

Ans: We have inverse link function

$$g^{-1}(x) = \frac{1}{x},$$

hence

$$\frac{d}{dx} g^{-1}(x) = -\frac{1}{x^2},$$

and

$$\frac{d}{dx} g^{-1}(x)\Big|_{x=\vec{X}_i^T \vec{\beta}} = \lambda_i^2 .$$

Using results from the preliminaries and part d of problem 1, we thus obtain

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[-(Y_i - \frac{1}{\lambda_i}) \, \lambda_i^2 X_{ij} \right] = \sum_{i=1}^{n} \left[-\frac{(Y_i - \frac{1}{\lambda_i})}{\frac{1}{\lambda_i^2}} \, X_{ij} \right],$$

and the score function is of the form

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^{n} \left[-\frac{(Y_i - E[Y_i])}{Var(Y_i)} \, X_{ij} \right] .$$

3. Comment on the similarity of the forms of these score equations for exponential family models.

Ans: In these three exponential family distributions, when our regression model involved an identity link for the mean, we had a score function equal to that of WLSE.

In these same settings, when our regression model involved the canonical link (so parts a, b, and e of problem 2) we had a score function similar to that of OLSE.

These results can be generalized for the "Generalized Linear Model". Suppose $Y_i \,|\, \vec{X}_i$ has a density (probability mass function) in the exponential family written in "canonical form" as

$$f_{Y\,|\,\vec{X}}(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right),$$

and canonical parameter $\theta = \theta(\vec{X}, \vec{\beta})$ is some function of the covariate vector $\vec{X}_i$ and an unknown regression parameter vector $\vec{\beta}$.

The moment generating function for a distribution of this class is found to be

$$\begin{aligned} M_Y(t) = E[e^{ty}] & \int e^{ty} f_Y(t; \theta, \phi) \, dy \\ &= \int \exp\left(\frac{yta(\phi) + y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right) dy \\ &= \exp\left(\frac{b(ta(\phi) + \theta) - b(\theta)}{a(\phi)} \right) \int \exp\left(\frac{y(ta(\phi) + \theta) - b(ta(\phi) + \theta)}{a(\phi)} + c(y, \phi) \right) dy \, , \\ &= \frac{b(ta(\phi) + \theta) - b(\theta)}{a(\phi)} \end{aligned}$$

from which we find that

$$\begin{aligned}
E(Y|\vec{X}) &= \frac{d}{dt}M_Y(t)\Big|_{t=0} = \frac{a(\phi)b'(ta(\phi)+\theta)}{a(\phi)}M_Y(t)\Big|_{t=0} = b'(\theta) \\
E(Y^2|\vec{X}) &= \frac{d^2}{dt^2}M_Y(t)\Big|_{t=0} = \left(a(\phi)b''(ta(\phi)+\theta) + (b'(ta(\phi)+\theta))^2\right)M_Y(t)\Big|_{t=0} \\
&= a(\phi)b''(\theta) + (b'(\theta))^2 \\
Var(Y\ \vec{X}) &= a(\phi)b''(\theta).
\end{aligned}$$

Now, a generalized linear regression model considers

$$g(E[Y|\vec{X}]) = \vec{X}^T\vec{\beta},$$

so in terms of θ, we have a regression model

$$g^*(\theta_i) = \vec{X}_i^T\vec{\beta},$$

where $g^*(\cdot) = g \circ b'(\cdot)$. A canonical link function would have g^* be the identity, so the canonical link is $g(\cdot) = [b'(\cdot)]^{-1}$.

The score function for our generalized linear model is found as

$$\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \frac{\partial}{\partial\beta_j}\mathcal{L}(\vec{\beta}) = \sum_{i=1}^n \frac{\partial}{\partial\beta_j}\log(f(Y_i\,;\,\theta_i)) \\
&= \sum_{i=1}^n \left[\ \frac{\partial}{\partial\theta_i}\left(\frac{Y_i\theta_i - b(\theta_i)}{a(\phi_i)} + c(Y_i,\phi_i)\right)\ \frac{\partial}{\partial\beta_j}\theta_i\ \right] \\
&= \sum_{i=1}^n \left[\ \left(\frac{Y_i - b'(\theta_i)}{a(\phi_i)}\right))\ \frac{\partial}{\partial\beta_j}\theta_i\ \right] \\
&= \sum_{i=1}^n \left[\ \left(\frac{Y_i - E[Y_i]}{Var(Y_i)}\right)\ b''(\theta_i)\frac{\partial}{\partial\beta_j}\theta_i\ \right]
\end{aligned}$$

.

Now when we use the identity link, then $b'(\theta_i) = \vec{X}_i^T\vec{\beta}$, and

$$\frac{\partial}{\partial\beta_j}\theta_i\ = \frac{1}{b''(\theta_i)}X_{ij},$$

so

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^n \left[\ \left(\frac{Y_i - E[Y_i]}{Var(Y_i)}\right)\ X_{ij}\right].$$

When we use the canonical link $\theta_i = \vec{X}_i^T\vec{\beta}$, and

$$\frac{\partial}{\partial\beta_j}\theta_i\ = X_{ij},$$

so

$$\mathcal{U}_j(\vec{\beta}) = \sum_{i=1}^n \left[\ \left(\frac{Y_i - E[Y_i]}{a(\phi)}\right)\ X_{ij}\right].$$

I note that in the Bernoulli, Poisson, and exponential models, $a_(\phi) = 1$.

With other link functions, the score function may be of a markedly different form.

For what it is worth, we do not always use regression models of the mean, nor do we only consider parametric regression models involving exponential family:

- When $Y \mid \vec{X}$ has a log normal distribution, we most often model the geometric mean using a log link. This is generally effected using linear regression on $\log(Y)$.
- When $Y \mid \vec{X}$ has a distribution according to some parametric accelerated failure time (AFT) distribution, we most often model the quantiles using a log link. In the AFT model, we assume that

$$Y_i \mid \vec{X}_i \;\sim\; F_i(y) = F_0(ye^{\vec{X}^T\vec{\beta}})$$

  for some known distribution function F_0. Such a model forces constant ratios between F_i and F_0 for all quantiles, so it is immaterial which quantile you use as θ_i. This is generally effected using likelihood theory for the distribution of $\log(Y)$, rather than the likelihood for Y.
- When $Y \mid \vec{X}$ has a distribution according to some semiparametric proportional hazards (PH) distribution, we most often model the hazard function using a log link. In the PH model, we assume that

$$Y_i \mid \vec{X}_i \;\sim\; F_i(y) = 1 - [1 - F_0(y)]^{e^{\vec{X}^T\vec{\beta}}}$$

  for some unspecified distribution function F_0.
- The parametric Weibull distribution has $F(y) = 1 - \exp((\lambda y)^p)$. This family encompasses all distributions that are both AFT and PH. The exponential is a special case, but not all Weibull distributions are exponential family. Parametric regression models for this distribution could be formulated by either considering the hazard or the quantiles.