Stat 582
Emerson, Win 12
Homework #5 Key
February 22, 2012

Written problems to be handed in Wednesday, February 29, 2012.

Problems 1-3 consider the general parametric regression model in which we observe pairs $(Y_i, \vec{X}_i)$ for $i = 1, \ldots, n$ in which

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

with the $Y_i$'s mutually independent, $\vec{X}_i = (1, X_{i1}, X_{i2}, \ldots, X_{ip})^T$ known covariates, and $\vec{\beta}$ 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 and information matrix for the following choices of $f_Y$ and $\theta_i$.

**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

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

The components of the information matrix are found as

$$\mathcal{I}_{jk}(\vec{\beta}) = -E\left[ \frac{\partial}{\partial \beta_j} \mathcal{U}_k(\vec{\beta}) \right] = -\sum_{i=1}^{n} E\left[ \frac{\partial}{\partial \theta_i} \mathcal{U}_k(\vec{\beta}) \ \frac{\partial \theta_i}{\partial \beta_j} \right].$$

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

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

$$\mathcal{I}_{jk}(\vec{\beta}) = -\sum_{i=1}^{n} \ X_{ij} X_{ik} \ E\left[ \frac{\partial^2}{\partial \theta_i^2} \log(f(Y_i \,;\, \theta_i)) \right].$$

a. Normal: $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ and $\theta_i = \mu_i$.

**Ans:** In this part (unlike the remaining parts of this problem), I must consider the parameter of interest $\theta_i$ and the "nuisance" parameter $\sigma^2$. In order to use the general results defined above, I find

$$\begin{aligned}
\log(f(Y_i \ \ \theta_i = \mu_i,\ \sigma^2) &= \log\left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(Y_i-\mu_i)^2}{2\sigma^2} \right\} \right) \\
&= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(Y_i-\mu_i)^2}{2\sigma^2} \\
\frac{\partial}{\partial\theta_i}\log(f(Y_i \ \ \theta_i = \mu_i,\ \sigma^2) &= \frac{(Y_i-\mu_i)}{\sigma^2} \\
\frac{\partial}{\partial\sigma^2}\log(f(Y_i \ \ \theta_i = \mu_i,\ \sigma^2) &= -\frac{1}{2\sigma^2} + \frac{(Y_i-\mu_i)^2}{2\sigma^4} \\
\frac{\partial^2}{\partial\theta_i^2}\log(f(Y_i \ \ \theta_i = \mu_i,\ \sigma^2) &= -\frac{1}{\sigma^2}, \\
\frac{\partial^2}{\partial(\sigma^2)^2}\log(f(Y_i \ \ \theta_i = \mu_i,\ \sigma^2) &= \frac{1}{2\sigma^2} - \frac{(Y_i-\mu_i)^2}{\sigma^6}, \\
\frac{\partial^2}{\partial\theta_i\partial\sigma^2}\log(f(Y_i \ \ \theta_i = \mu_i,\ \sigma^2) &= -\frac{(Y_i-\mu_i)}{\sigma^4},
\end{aligned}$$

so using $E[Y_i] = \mu_i$ and $E[(Y_i-\mu_i)^2] = \sigma^2$ we find for $0 \le j, k \le p$ and $\vec{\phi} = (\beta_T\ \sigma^2)^T$

$$\begin{aligned}
\mathcal{U}_j(\vec{\phi}) &= \sum_{i=1}^{n} \left[ \frac{(Y_i-\mu_i)}{\sigma^2} X_{ij} \right] \\
\mathcal{U}_{p+1}(\vec{\phi}) &= -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \left[ \frac{(Y_i-\mu_i)^2}{2\sigma^4} X_{ij} \right] \\
\mathcal{I}_{jk}(\vec{\phi}) &= \sum_{i=1}^{n} \frac{1}{\sigma^2} X_{ij} X_{ik} \\
\mathcal{I}_{j(p+1)}(\vec{\phi}) &= 0 \\
\mathcal{I}_{(p+1)(p+1)}(\vec{\phi}) &= \sum_{i=1}^{n} \frac{1}{2\sigma^4}
\end{aligned}$$

Note that in this model, $E(Y_i) = \mu_i$ and $Var(Y_i) = \sigma^2$, so the score function and information matrix related to $\vec{\beta}$ are of the forms

$$\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{(Y_i - E[Y_i])}{Var(Y_i)} X_{ij} \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{1}{Var(Y_i)} X_{ij} X_{ik} \right].
\end{aligned}$$

*(For what it is worth, we rarely use this regression model by estimating the MLE for $\phi$ and deriving the asymptotic distribution for the $T$ statistic based on the delta method. Instead, we tend to work the problem as if $\sigma^2$ were known, and then use the unbiased estimate of $\sigma^2$ with Slutsky's theorem (if we are in a distribution free model) or with normal based theory for a $T$ statistic if we really regard that our data are normally distributed.)*

b. 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

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

so using $E[Y_i] = p_i$ we find

$$
\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \ \frac{(Y_i - p_i)}{p_i(1-p_i)} \ X_{ij} \ \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \frac{1}{p_i(1-p_i)} X_{ij} X_{ik}.
\end{aligned}
$$

Note that in this model, $E(Y_i) = p_i$ and $Var(Y_i) = p_i(1-p_i)$, so the score function and information matrix are of the forms

$$
\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \ \frac{(Y_i - E[Y_i])}{Var(Y_i)} \ X_{ij} \ \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{1}{Var(Y_i)} X_{ij} X_{ik} \right].
\end{aligned}
$$

c. 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

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

so using $E[[Y_i] = \lambda_i$ we find

$$
\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \ \frac{(Y_i - \lambda_i)}{\lambda_i} \ X_{ij} \ \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \frac{1}{\lambda_i} X_{ij} X_{ik}.
\end{aligned}
$$

Note that in this model, $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i$, so the score function and information matrix are of the forms

$$
\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \ \frac{(Y_i - E[Y_i])}{Var(Y_i)} \ X_{ij} \ \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{1}{Var(Y_i)} X_{ij} X_{ik} \right].
\end{aligned}
$$

d. 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

$$\begin{aligned}
\log(f(Y_i \;\; \theta_i = \lambda_i) &= \log\left( \frac{1}{\lambda_i} e^{-\frac{Y_i}{\lambda_i}} \right) \\
&= -\log(\lambda_i) - \frac{Y_i}{\lambda_i} \\
\frac{\partial}{\partial \theta_i} \log(f(Y_i \;\; \theta_i = \lambda_i) &= -\frac{1}{\lambda_i} + \frac{Y_i}{\lambda_i^2} = \frac{Y_i - \lambda_i}{\lambda_i^2}, \\
\frac{\partial^2}{\partial \theta_i^2} \log(f(Y_i \;\; \theta_i = \lambda_i) &= \frac{-2Y_i + \lambda_i}{\lambda_i^3},
\end{aligned}$$

so using $E[[Y_i] = \lambda_i$ we find

$$\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{(Y_i - \lambda_i)}{\lambda_i^2} \; X_{ij} \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \frac{1}{\lambda_i^2} X_{ij} X_{ik}.
\end{aligned}$$

Note that in this model, $E(Y_i) = \lambda_i$ and $Var(Y_i) = \lambda_i^2$, so the score function and information matrix are of the forms

$$\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{(Y_i - E[Y_i])}{Var(Y_i)} \; X_{ij} \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \left[ \frac{1}{Var(Y_i)} X_{ij} X_{ik} \right].
\end{aligned}$$

e. 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

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

so

$$\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} \left[ -(Y_i - \frac{1}{\lambda_i}) \; X_{ij} \right] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} \frac{1}{\lambda_i^2} X_{ij} X_{ik}.
\end{aligned}$$

Note that in this model, $E(Y_i) = 1/\lambda_i$ and $Var(Y_i) = 1/\lambda_i^2$, so the score function and information matrix are of the forms

$$\begin{aligned}
\mathcal{U}_j(\vec{\beta}) &= \sum_{i=1}^{n} [\; -(Y_i - E[Y_i]) \; X_{ij} \;] \\
\mathcal{I}_{jk}(\vec{\beta}) &= \sum_{i=1}^{n} [Var(Y_i) X_{ij} X_{ik}].
\end{aligned}$$

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 $g(x) = logit(x)$.

**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 b of problem 1, we thus obtain

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} \frac{1}{p_i(1-p_i)} \; [p_i(1-p_i)]^2 X_{ij}X_{ik} = \sum_{i=1}^{n} p_i(1-p_i)X_{ij}X_{ik}.$$

and the score function and information matrix are of the forms

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} Var(Y_i)X_{ij}X_{ik}.$$

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 c of problem 1, we thus obtain

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} \frac{1}{\lambda_i} \; \lambda_i^2 X_{ij}X_{ik} = \sum_{i=1}^{n} \lambda_i X_{ij}X_{ik}.$$

and the score function and information matrix are of the forms

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} Var(Y_i)X_{ij}X_{ik}.$$

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 d of problem 1, we thus obtain

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

and the score function and information matrix are of the forms

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

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 e of problem 1, we thus obtain

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

and the score function and information matrix are of the forms

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

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 d of problem 1, we thus obtain

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} \frac{1}{\lambda_i^2} \; \lambda_i^4 X_{ij} X_{ik} = \sum_{i=1}^{n} \lambda_i^2 X_{ij} X_{ik},$$

and the score function and information matrix are of the forms

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} Var(Y_i) X_{ij} X_{ik}.$$

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

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} \frac{1}{\lambda_i^2} \; \lambda_i^4 X_{ij} X_{ik} = \sum_{i=1}^{n} \lambda_i^2 X_{ij} X_{ik},$$

and the score function and information matrix are of the forms

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

$$\mathcal{I}_{jk}(\vec{\beta}) = \sum_{i=1}^{n} \frac{1}{Var(Y_i)} X_{ij} X_{ik}.$$

3. Comment on the similarity of the forms of these score equations for exponential family models. What would be the asymptotic distribution of the score statistics under departures from the parametric families specified? (Do not solve this in detail. Just provide general properties of the distribution.)

**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 $\theta$, 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

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

An important result of the above is that regardless of the true distribution of the data (so long as it has a mean) and regardless of the link function, the score function has expectation 0, and it represents the sum of independent random variables (transformations of the original random variables). Hence, under suitable "regularity conditions" (having to do with the distribution of predictors and the weights assigned to each observation), by appealing to the Lindeberg-Feller central limit theorem we might reasonably expect the score function to be approximately normally distributed.

We cannot, however, expect that the information matrix as computed using the second partial of the log likelihood will be the variance of the score function *unless* the data is truly distributed according to the density used in the log likelihood. Under suitable regularity conditions, however, we might consider that we can estimate the variance of the score function using the average squares of the contributions of each observation to the score:

$$\widehat{Var}\left(\vec{\mathcal{U}}(\vec{\beta}\,|\,\vec{Y})\right) = \sum_{i=1}^{n} \vec{\mathcal{U}}(\vec{\beta}\,|\,Y_i)\vec{\mathcal{U}}^T(\vec{\beta}\,|\,Y_i).$$

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, though in all the cases considered the score function has expectation 0 for arbitrary distributions. Hence, we would expect any of these "estimating equations" could be used for relatively robust inference about the means. They would not, however, be equally efficient, because they correspond to different weightings of the observation.

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 $Y\,|\,\vec{X}$ 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 $Y\,|\,\vec{X}$ 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

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

  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 $\theta_i$. 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.

4. Consider the censored data setting in which independent, identically distributed random variables $T_{ij} \sim \mathcal{E}(\lambda_i)$ for $i = 0, 1$ and $j = 1, \ldots, n_i$ (with $ET_{ij} = \frac{1}{\lambda_i}$) are subject to censoring by known $C_{ij}$. Hence, we observe only $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\delta_{ij} = 1_{[T_{ij}=Y_{ij}]}$. Find the maximum likelihood estimator of $\theta = \lambda_1/\lambda_0$ along with its asymptotic distribution.

**Ans:** Two facets of this problem that must be addressed are the handling of the censored observations (i.e., what is the contribution to the likelihood from each case) and the restrictions that must be placed on the problem by the sampling scheme.

For an observed event, $\delta_{ij} = 1$, and the contribution to the likelihood should be

$$f(Y_{ij}) = \lambda_i e^{-\lambda_i Y_{ij}}.$$

For a censored observation, we only know that $T_{ij} > Y_{ij}$, so the contribution to the likelihood should be

$$P(T_{ij} > Y_{ij}) = e^{-\lambda_i Y_{ij}}.$$

The likelihood function for $\vec{\lambda}$ is thus

$$\begin{aligned} L(\vec{\lambda} \mid \vec{Y}) &= \prod_{i=0}^{1} \prod_{j=1}^{n_i} \left[\lambda_i e^{-\lambda_i Y_{ij}}\right]^{\delta_{ij}} \left[e^{-\lambda_i Y_{ij}}\right]^{1-\delta_{ij}} \\ &= \prod_{i=0}^{1} \prod_{j=1}^{n_i} \lambda_i^{\delta_{ij}} e^{-\lambda_i Y_{ij}}, \end{aligned}$$

and the log likelihood function is

$$\mathcal{L}(\vec{\lambda} \mid \vec{Y}) = \sum_{i=0}^{1} \sum_{j=1}^{n_i} \left[\delta_{ij} \log(\lambda_i) - \lambda_i Y_{ij}\right].$$

For notational convenience, define $d_i = \sum_{j=1}^{n_i} \delta_{ij}$ and $Y_i = \sum_{j=1}^{n_i} Y_{ij}$. Then we have

$$\begin{aligned}
\mathcal{L}(\vec{\lambda}\,|\,\vec{Y}) &= d_0 \log(\lambda_0) - Y_0\lambda_0 + d_1 \log(\lambda_1) - Y_1\lambda_1 \\
\mathcal{U}_{n:0}(\vec{\lambda}\,|\,\vec{Y}) &= \frac{\partial}{\partial \lambda_0}\mathcal{L}(\vec{\lambda}\,|\,\vec{Y}) = \frac{d_0}{\lambda_0} - Y_0 \\
\mathcal{U}_{n:1}(\vec{\lambda}\,|\,\vec{Y}) &= \frac{\partial}{\partial \lambda_1}\mathcal{L}(\vec{\lambda}\,|\,\vec{Y}) = \frac{d_1}{\lambda_1} - Y_1 \\
\mathcal{I}_{n:00}(\vec{\lambda}\,|\,\vec{Y}) &= -E\left[\frac{\partial}{\partial \lambda_0}\mathcal{U}_{n:0}(\vec{\lambda}\,|\,\vec{Y})\right] = \frac{E(d_0)}{\lambda_0^2} \\
\mathcal{I}_{n:11}(\vec{\lambda}\,|\,\vec{Y}) &= -E\left[\frac{\partial}{\partial \lambda_1}\mathcal{U}_{n:1}(\vec{\lambda}\,|\,\vec{Y})\right] = \frac{E(d_1)}{\lambda_1^2} \\
\mathcal{I}_{n:01}(\vec{\lambda}\,|\,\vec{Y}) &= -E\left[\frac{\partial}{\partial \lambda_0}\mathcal{U}_{n:1}(\vec{\lambda}\,|\,\vec{Y})\right] = 0 \\
\mathcal{I}_{n:10}(\vec{\lambda}\,|\,\vec{Y}) &= -E\left[\frac{\partial}{\partial \lambda_1}\mathcal{U}_{n:0}(\vec{\lambda}\,|\,\vec{Y})\right] = 0
\end{aligned}$$

The estimates derived as roots of the score equations are thus $\tilde{\lambda}_i = d_i/Y_i$.

Regular asymptotic theory (if it applies) would thus suggest that for consistent estimator $\hat{\mathcal{I}}(\vec{\lambda})$ of information matrix $\mathcal{I}(\vec{\lambda})$

$$\hat{\mathcal{I}}_n^{1/2}(\tilde{\vec{\lambda}}_n - \vec{\lambda}) = \begin{pmatrix} \sqrt{\hat{\mathcal{I}}_{n:00}(\vec{\lambda})}\;(\tilde{\lambda}_0 - \lambda_0) \\ \sqrt{\hat{\mathcal{I}}_{n:11}(\vec{\lambda})}\;(\tilde{\lambda}_1 - \lambda_1) \end{pmatrix} \to_d \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

Consistent estimators for the information matrix can be based on

- $E[d_i]$ as can be calculated from the known $C_{ij}$ and hypothesized $\lambda_i$.
- Using $d_i$ as an estimate of $E[d_i]$ and $\tilde{\lambda}_i$ as an estimate of $\lambda_i$.

Now the validity of the above asymptotic results follows from either of the approaches used in problem 4 of homework 2.

- By treating each sample independently, we have a regular problem and under suitable restrictions on the $C_{ij}$'s (merely that they must be strictly positive infinitely often), we have the existence and consistency of the individual estimates. Then we can easily argue the asymptotic joint normality from independence, with some suitable definition placed on the accrual of information from each sample.
- We can use the Lindeberg-Feller central limit theorem on the score functions derived above, and use the Lindeberg condition to place restrictions on the accrual of information about each of the $\lambda_i$'s.

Below I condition on $\mathcal{I}_{n:11}/\mathcal{I}_{n:00} \to r > 0$ as $n \to \infty$. We then use Slutsky's theorem to find that

$$\begin{pmatrix} \sqrt{\frac{\mathcal{I}_{n:11}}{\mathcal{I}_{n:00}}} & 1 \end{pmatrix} \begin{pmatrix} \sqrt{\hat{\mathcal{I}}_{n:00}(\vec{\lambda})}\;(\tilde{\lambda}_0 - \lambda_0) \\ \sqrt{\hat{\mathcal{I}}_{n:11}(\vec{\lambda})}\;(\tilde{\lambda}_1 - \lambda_1) \end{pmatrix} = \sqrt{\mathcal{I}_{n:11}} \begin{pmatrix} (\tilde{\lambda}_0 - \lambda_0) \\ (\tilde{\lambda}_1 - \lambda_1) \end{pmatrix} \to_d \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \right).$$

Then, providing that $\mathcal{I}_{n:11} \to \infty$ as $n \to \infty$, we can use the delta method to find that $\tilde{\theta}_n = \tilde{\lambda}_1/\tilde{\lambda}_0$ has asymptotic distribution

$$\sqrt{\mathcal{I}_{n:11}}\left(\tilde{\theta}_n - \theta\right) \to_d \mathcal{N}\left(0, \frac{r}{\lambda_1^2} + \frac{\lambda_0^2}{\lambda_1^4}\right).$$