MATH523 Assignment 1

Kabilan Sriranjan

February 6, 2018

Question 1.

a)

$$
\begin{aligned}
f(y;\mu,\lambda) &= \Big(\frac{\lambda}{2\pi y^3}\Big)^{\frac{1}{2}} exp\Big(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\Big) \\
&= exp\Big(\frac{log(\lambda)}{2} - \frac{log(2\pi y^3)}{2} - \frac{\lambda y^2 - 2\lambda y\mu + \lambda\mu^2}{2\mu^2 y}\Big) \\
&= exp\Big(-\frac{\lambda y}{2\mu^2} + \frac{\lambda}{\mu} + \frac{\lambda}{2y} + \frac{log(\lambda)}{2} - \frac{log(2\pi y^3)}{2}\Big) \\
&= exp\Big(-\frac{\lambda}{2}\Big(\frac{y}{\mu^2} - \frac{2}{\mu}\Big) + \frac{\lambda}{2y} + \frac{log(\lambda)}{2} - \frac{log(2\pi y^3)}{2}\Big) \\
&= exp\Big(\frac{y\frac{1}{\mu^2} - 2\sqrt{\frac{1}{\mu^2}}}{-\frac{2}{\lambda}} + \frac{\lambda}{2y} + \frac{log(\lambda)}{2} - \frac{log(2\pi y^3)}{2}\Big) \\
&= exp\Big(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\Big)
\end{aligned}
$$

with the canonical and dispersion parameters

$$
\begin{aligned}
\theta &= \frac{1}{\mu^2} \\
\phi &\equiv \lambda
\end{aligned}
$$

and the functions b and c are given by

$$
\begin{aligned}
b(\theta) &= 2\sqrt{\theta} \\
c(y,\phi) &= \frac{\phi}{2y} + \frac{log(\phi)}{2} - \frac{log(2\pi y^3)}{2}
\end{aligned}
$$

Since ϕ is unknown we know that the family of inverse Gaussian distributions is an exponential dispersion family.

b)

$$
\begin{aligned}
b'(\theta) &= 2\Big(\frac{1}{2\theta^{\frac{1}{2}}}\Big) \\
&= \frac{1}{\sqrt{\theta}} \\
&= \frac{1}{\sqrt{\frac{1}{\mu^2}}} \\
&= \mu \\
b''(\theta) &= -\frac{1}{2\theta^{\frac{3}{2}}} \\
&= -\frac{1}{2(\sqrt{\theta})^3} \\
&= -\frac{1}{2(\sqrt{\frac{1}{\mu^2}})^3} \\
&= -\frac{\mu^3}{2} \\
a(\phi) &= -\frac{2}{\lambda}
\end{aligned}
$$

Let Y be an inverse Gaussian random variable

$$
\begin{aligned}
\mathbb{E}[Y] &= b'(\theta) \\
&= \mu \\
var(Y) &= b''(\theta)a(\phi) \\
&= \frac{\mu^3}{\lambda} \\
V(\mu) &= b''(\theta) \\
&= -\mu^3
\end{aligned}
$$

c)

$$
\begin{aligned}
\theta &= \frac{1}{\mu^2} \\
\implies g(x) &= \frac{1}{x^2}
\end{aligned}
$$

The canonical link I've given seems to be sensible. It is smooth and monotone on the support of the inverse Gaussian distribution.

d)

$$
\begin{aligned}
g(\mu_i) &= X\beta \\
\frac{1}{\mu_i^2} &= X\beta \\
\implies \mu_i &= \frac{1}{\sqrt{X\beta}} \\
l(\beta) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi)} x_{ij} &\equiv 0 \\
\implies \sum_{i=1}^{n} \frac{y_i - \frac{1}{\sqrt{X\beta}}}{-\frac{2}{\lambda}} x_{ij} &= 0 \\
\implies \sum_{i=1}^{n} (y_i - \frac{1}{\sqrt{X\beta}}) x_{ij} &= 0
\end{aligned}
$$

e)

Score equation for $j = 1$:

$$
\begin{aligned}
\sum_{i=1}^{n} (y_i - \frac{1}{\sqrt{\beta_0 + \beta_1 x_i}}) &= 0 \\
\implies \sum_{i=1}^{n_A} (y_i - \frac{1}{\sqrt{\beta_0 + \beta_1}}) + \sum_{i=n_A}^{n_A+n_B} (y_i - \frac{1}{\sqrt{\beta_0}}) &= 0
\end{aligned}
$$

Score equation for $j = 2$:

$$
\sum_{i=1}^{n_A} (y_i - \frac{1}{\sqrt{\beta_0 + \beta_1}}) = 0
$$

From these two equations we get the following:

$$
\begin{aligned}
\frac{1}{n_A} \sum_{i=1}^{n_A} y_i &= \frac{1}{\sqrt{\beta_0 + \beta_1}} \\
\frac{1}{n_B} \sum_{i=1}^{n_A+n_B} y_i &= \frac{1}{\sqrt{\beta_0}}
\end{aligned}
$$

Which are exactly the group means

Question 2

We can follow the same steps as for Q1.e but with a generic link g(x) Score equation for $j = 1$:

$$\sum_{i=1}^{n}(y_i - (g)^{-1}(\beta_0 + \beta_1 x_i) = 0$$
$$\implies \sum_{i=1}^{n_A}(y_i - (g)^{-1}(\beta_0 + \beta_1)) + \sum_{i=n_A}^{n_A+n_B}(y_i - (g)^{-1}(\beta_0)) = 0$$

Score equation for $j = 2$:

$$\sum_{i=1}^{n_A}(y_i - (g)^{-1}(\beta_0 + \beta_1) = 0$$

Combining the two we get:

$$\frac{1}{n_A}\sum_{i=1}^{n_A} y_i = (g)^{-1}(\beta_0 + \beta_1)$$
$$\frac{1}{n_B}\sum_{i=1}^{n_A+n_B} y_i = (g)^{-1}(\beta_0)$$

Question 3

a)

$$S(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i(\beta))^2}{var(y_i)}$$
$$\widehat{\beta} = argmax(S(\beta))$$

We can minimize S(β) by differentiating with respect to β_j for j = 1, . . . , p and setting the derivative to 0

$$\frac{\partial S}{\partial \beta_j} S(\beta) = 0$$
$$\sum_{i=1}^{n} \frac{2(y_i - \mu_i(\beta))}{var(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0$$
$$\sum_{i=1}^{n} \frac{(y_i - \mu_i(\beta))}{var(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0$$

Thus the score equations can be obtained by deriving the least squares estimates

b)

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{(y_i - \mu_i)}{var(y_i)} \frac{\partial \mu_i}{\partial \beta_j} = 0 \\
\implies &\sum_{i=1}^{n} \frac{y_i - \mu_i}{a(\phi) b''(\theta_i)} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = 0 \\
\implies &\frac{1}{a(\phi)} \sum_{i=1}^{n} \frac{y_i - \mu_i}{b''(\theta_i)} \frac{1}{g'(\mu_i)} x_{ij} = 0 \\
\implies &\sum_{i=1}^{n} \frac{y_i - \mu_i}{b''(\theta_i)} b''(\theta_i) x_{ij} = 0 \\
\implies &\sum_{i=1}^{n} (y_i - \mu_i) x_{ij} = 0 \\
\implies &(\mathbf{y} - \mu) \cdot \mathbf{x}_j = 0
\end{aligned}
$$

For any j, $\mathbf{y} - \mu$ is perpendicular to $\mathbf{x}_j$ so the residual vector is orthogonal to the entire column space of $\mathbf{X}$

Question 4

a)

```
beetles <- read.table("Beetles2.dat", header=TRUE)
attach(beetles)
logitmod <- glm(cbind(dead, n-dead)~logdose, family=binomial)
summary(logitmod)
```

```
##
## Call:
## glm(formula = cbind(dead, n - dead) ~ logdose, family = binomial)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5878  -0.4085   0.8442   1.2455   1.5860
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -60.740      5.182  -11.72   <2e-16 ***
## logdose       34.286      2.913   11.77   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.202  on 7  degrees of freedom
## Residual deviance:  11.116  on 6  degrees of freedom
```

```
## AIC: 41.314
##
## Number of Fisher Scoring iterations: 4
```

```
logitmod$coefficients
```

```
## (Intercept)     logdose
##   -60.74013    34.28593
```

The most obvious distribution is the binomial distribution as there are n beetles that either died or survived. Using the canonical link we can see that logdose is significant.

b)

$$\frac{\frac{\widehat{\pi}(x)}{1-\widehat{\pi}(x)}}{\frac{\widehat{\pi}(x+1)}{1-\widehat{\pi}(x+1)}} = e^{-\widehat{\beta}_1}$$
$$\frac{\widehat{\pi}(x)}{1-\widehat{\pi}(x)} = e^{-\widehat{\beta}_1}\frac{\widehat{\pi}(x+1)}{1-\widehat{\pi}(x+1)}$$

$e^{-0.1\widehat{\beta}_1} \approx 0.324$ so increasing temperature by 0.1 multiplies the odds of a beetle dying by 30. Wow! #c)

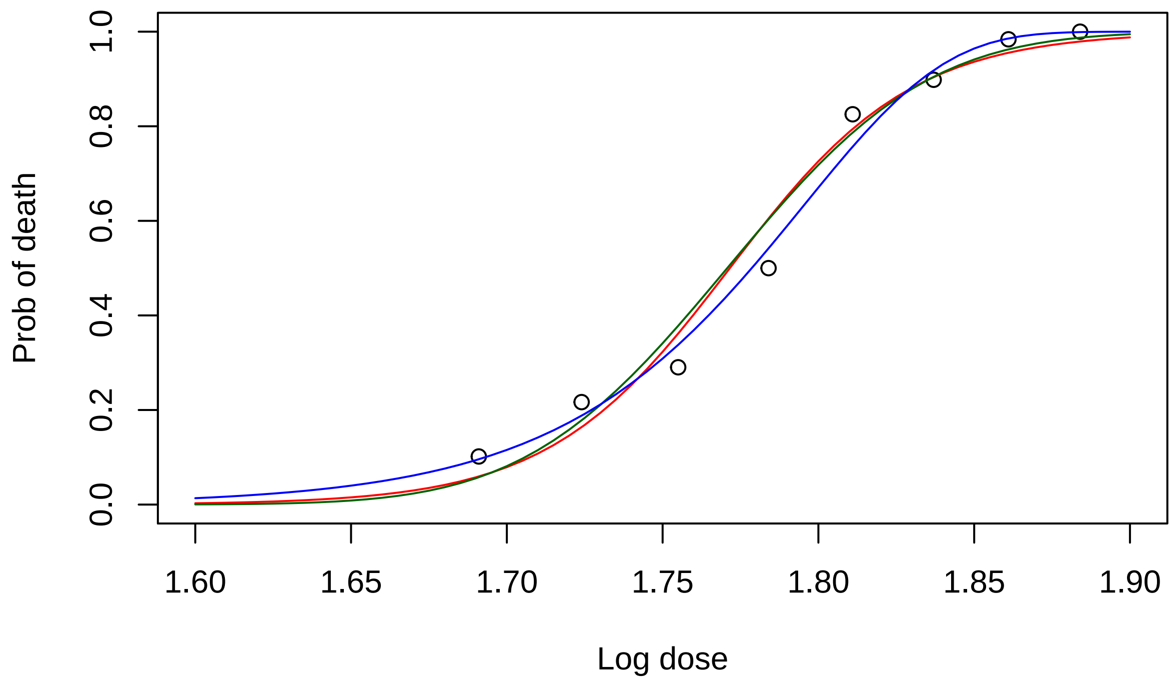
```
probitmod <- glm(cbind(dead,n-dead)~logdose,family=binomial(link=probit))
loglogmod <- glm(cbind(dead,n-dead)~logdose,family=binomial(link=cloglog))
```

The probit link and the log-log link are both other possible links we can use.

d)

```
plot(dead/n~logdose,xlim=c(1.6,1.9),ylim=c(0,1),xlab="Log dose",ylab="Prob of death")
x <- seq(from=1.6,to=1.9,by=0.005)

lines(x,exp(coef(logitmod)[1]+coef(logitmod)[2]*x)/(1+exp(coef(logitmod)[1]+coef(logitmod)[2]*x)),col="
lines(x,pnorm(coef(probitmod)[1]+coef(probitmod)[2]*x),col="darkgreen")
lines(x,1-exp(-exp(coef(loglogmod)[1]+coef(loglogmod)[2]*x)),col="blue")
```



Logit given in red, probit in green, and log-log in blue

e)

```
exp(coef(logitmod)[1]+coef(logitmod)[2]*logdose)/(1+exp(coef(logitmod)[1]+coef(logitmod)[2]*logdose))
```

```
## [1] 0.05937747 0.16366723 0.36162283 0.60490961 0.79440490 0.90405532
## [7] 0.95546748 0.97925643
```

```
pnorm(coef(probitmod)[1]+coef(probitmod)[2]*logdose)
```

```
## [1] 0.0577367 0.1781060 0.3780390 0.6032833 0.7866532 0.9045852 0.9626183
## [8] 0.9873227
```

```
1-exp(-exp(coef(loglogmod)[1]+coef(loglogmod)[2]*logdose))
```

```
## [1] 0.09582195 0.18802653 0.33777217 0.54177644 0.75683967 0.91843509
## [7] 0.98575181 0.99913561
```

Finding the fitted values for each model

f)

```
logitmod$aic
```

```
## [1] 41.31361
```

```
probitmod$aic
```

```
## [1] 40.18499
```

```r
loglogmod$aic
```

```
## [1] 33.71237
```

The log-log model has the smallest AIC so it appears to be the best. The model with the minimal AIC is the one that we expect to have lost the least amount of information from the true model.