Problem 5.1

Recall that a distribution \( f(x; \theta) \), where \( x \) is the argument and \( \theta \) is a scaler parameter, belongs to an exponential family if it can be represented in the following form:

\[
f(x; \theta) = h(x)exp[\eta(\theta)T(x) - a(\theta)] = exp[\eta(\theta)T(x) - a(\theta) + b(x)]
\]

Here, the function \( \eta(\theta) \) is called the natural parameter, it’s often chosen as the link in a glm with independent variable \( Y \) from the distribution \( f(y; \theta) \). Show that the following distributions belong to the exponential family, and find the corresponding natural parameters:

a. Bernoulli (parameter \( p \))

\[
f(x; p) = p^x(1-p)^{1-x} = e^{ln[p^x(1-p)^{1-x}]} = e^{[x \ln p + (1-x) \ln(1-p)]} = e^{[x \ln p - x \ln(1-p)]}
\]

which implies that the Bernoulli distribution belongs to the exponential family with natural parameter \( \eta(p) = \ln \frac{p}{1-p} \)

b. Binomial (parameter \( p \); \( n \) is known)

\[
f(x; n, p) = \binom{n}{x} p^x(1-p)^{n-x} = e^{ln[\binom{n}{x} p^x(1-p)^{n-x}]} = e^{\ln(\binom{n}{x}) + x \ln p + (n-x) \ln(1-p)} = \binom{n}{x} e^{[x \ln p + (n-x) \ln(1-p)]}
\]

which implies that the Binomial distribution belongs to the exponential family with natural parameter \( \eta(p) = \ln \frac{p}{1-p} \)

c. Poisson (parameter \( \lambda \))

\[
f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = e^{\ln[\frac{\lambda^x e^{-\lambda}}{x!}]} = e^{[x \ln \lambda + \ln e^{-\lambda} - \ln x!]} = e^{[x \ln \lambda - \lambda e^{-\lambda} - \ln x!]} = e^{[x \ln \lambda - \lambda - \ln x!]}
\]
which implies that the Poisson distribution belongs to the exponential family with natural parameter \( \eta(\lambda) = \ln \lambda \)

d. **Geometric (parameter \( p \))**

\[
\begin{align*}
  f(x; p) &= p(1 - p)^{x-1} \\
  &= e^{\ln[p(1-p)^{x-1}]} \\
  &= e^{[\ln p + (x-1)\ln(1-p)]} \\
  &= e^{[\ln p + x\ln(1-p) - \ln(1-p)]} \\
  &= e^{[x\ln(1-p) + \ln \frac{1}{1-p}]} 
\end{align*}
\]

which implies that the Geometric distribution belongs to the exponential family with natural parameter \( \eta(p) = \ln(1 - p) \)

e. **Negative Binomial (parameter \( p \); the other parameter is known)**

\[
\begin{align*}
  f(x; p, r) &= \binom{x + r - 1}{r - 1} p^r (1 - p)^x \\
  &= e^{\ln[\binom{x + r - 1}{r - 1} p^r (1 - p)^x]} \\
  &= \left( \frac{x + r - 1}{r - 1} \right) e^{[\ln p + x\ln(1-p)]} \\
  &= \left( \frac{x + r - 1}{r - 1} \right) e^{[x\ln(1-p) + \ln p^r]} 
\end{align*}
\]

which implies that the Negative Binomial distribution belongs to the exponential family with natural parameter \( \eta(p) = \ln(1 - p) \)

f. **Normal (parameter \( \mu \); \( \sigma \) is known)**

\[
\begin{align*}
  f(x; \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \\
  &= e^{\left[ -\frac{1}{2} \ln(2\pi\sigma^2) \right] - \frac{(x - \mu)^2}{2\sigma^2}} \\
  &= e^{\left[ -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x - \mu)^2}{2\sigma^2} \right]} \\
  &= e^{\left[ \frac{1}{2\sigma^2} \left( \frac{x^2}{\sigma^2} + \ln(2\pi\sigma^2) \right) \right]} 
\end{align*}
\]

which implies that the Normal distribution belongs to the exponential family with natural parameter \( \eta(\mu) = \frac{\mu}{\sigma^2} \)

**Problem 5.2**

Consider a Bernoulli rv \( Y \) with probability of success \( \pi \) and a Bernoulli rv \( X \) with probability of success \( p \) such that their joint distribution is determined by

\[
P(Y = 1 | X = 0) = \pi(0)
\]

and

\[
P(Y = 1 | X = 1) = \pi(1)
\]

Write and solve (find parameters) a glm with random component represented by \( Y \), a systematic component represented by \( X \), and the following link functions:
a. **Identity**
The model in this case is $\pi_i = \alpha + \beta X_i$. Thus, $\pi(0) = \alpha$ and $\pi(1) = \alpha + \beta$. Hence

$$\alpha = \pi(0)$$

and

$$\beta = \pi(1) - \pi(0)$$

b. **Logit**
The model in this case is $\ln \frac{\pi_i}{1-\pi_i} = \alpha + \beta X_i$. Thus, $\ln \frac{\pi(0)}{1-\pi(0)} = \alpha$ and $\ln \frac{\pi(1)}{1-\pi(1)} = \alpha + \beta$. Hence

$$\alpha = \ln \frac{\pi(0)}{1-\pi(0)}$$

and

$$\beta = \ln \frac{\pi(1)}{1-\pi(1)} - \ln \frac{\pi(0)}{1-\pi(0)} = \ln \frac{\pi(1)(1-\pi(0))}{\pi(0)(1-\pi(1))}$$

c. **Probit**
The model in this case is $\Phi^{-1}(\pi_i) = \alpha + \beta X_i$. Thus, $\Phi^{-1}(\pi(0)) = \alpha$ and $\Phi^{-1}(\pi(1)) = \alpha + \beta$. Hence

$$\alpha = \Phi^{-1}(\pi(0))$$

and

$$\beta = \Phi^{-1}(\pi(1)) - \Phi^{-1}(\pi(0))$$

**Problem 5.3**

Recall the snoring problem that we have discussed in the Lab. With snoring levels $s = [0, 2, 4, 5]$, the logistic ML fit is

$$\text{logit}(\hat{\pi}) = -3.866 + 0.397s$$

a. **Interpret the sign of the estimated effect of $s$?**

Since $\hat{\beta} = 0.397 > 0$, the estimated probability $\hat{\pi}$ is larger at larger snoring level.

b. **Estimate the probabilities and odds ratios of heart disease at snoring levels 2 and 5?**

$$\pi(2) = \frac{e^{-3.866+0.397(2)}}{1 + e^{-3.866+0.397(2)}} = 0.0443$$

$$OR(S_2, S_0) = \frac{e^\alpha e^{3\times2}}{e^\alpha e^{3\times0}} = e^{2\beta} = e^{2(0.397)} = 2.212$$

$$\pi(5) = \frac{e^{-3.866+0.397(5)}}{1 + e^{-3.866+0.397(5)}} = 0.1323$$

$$OR(S_5, S_0) = \frac{e^\alpha e^{5\times5}}{e^\alpha e^{5\times0}} = e^{5\beta} = e^{5(0.397)} = 7.28$$

c. **Describe the estimated effect of snoring on the odds of heart disease?**

Snoring level has a positive effect on the odds of heart disease. In particular, patients with snoring level of 2 are 2.2 times more likely to have a heart disease than patients with snoring level of 0. Moreover, patients with snoring level of 5 are 7.28 times more likely to have a heart disease than patients with snoring level of 0.
Problem 5.4

A sample of subjects were asked their opinion about current laws legalizing abortion (support, oppose). For the explanatory variables gender (female, male), religious affiliation (Protestant, Catholic, Jewish), and political party affiliation (Democrat, Republican, Independent), the model for the probability $\pi$ of supporting legalized abortion

$$\text{logit}(\pi) = \alpha + \beta^G + \beta^R + \beta^P$$

has reported parameter estimates (setting the parameter for the last category of a variable equal to 0.0)$\hat{\alpha} = -0.11, \beta^G_1 = 0.16, \beta^G_2 = 0.0, \beta^R_1 = -0.57, \beta^R_2 = -0.66, \beta^R_3 = 0.0, \beta^P_1 = 0.84, \beta^P_2 = -1.67, \beta^P_3 = 0.0$

a. Interpret how the odds of supporting legalized abortion depend on gender, party affiliation?

$$\text{OR}(\text{Female, Male}) = \frac{e^\alpha e^{\beta^G_1}}{e^\alpha e^{\beta^G_2}} = e^{\beta^G_1} = e^{0.16} \approx 1.17$$

That is, Females are 1.17 times more likely to support legalizing abortion than males.

$$\text{OR}(\text{Democrat, Independent}) = \frac{e^\alpha e^{\beta^P_1}}{e^\alpha e^{\beta^P_3}} = e^{\beta^P_1} = e^{0.84} \approx 2.32$$

That is, Democrats are 2.32 times more likely to support legalizing abortion than Independents.

$$\text{OR}(\text{Republican, Independent}) = \frac{e^\alpha e^{\beta^P_2}}{e^\alpha e^{\beta^P_3}} = e^{\beta^P_2} = e^{-1.67} \approx 0.188$$

That is, Republicans are 5.3 ($1/0.188$) times less likely to support legalizing abortion than Independents.

b. Find the estimated probability of supporting abortion for male Catholic Republicans and female Jewish Democrats?

Male Catholic Republican

$$\pi = \frac{e^\alpha e^{\beta^G_2 + \beta^R_2 + \beta^P}}{1 + e^\alpha e^{\beta^G_2 + \beta^R_2 + \beta^P}} = \frac{e^{-0.11 - 0.66 - 1.67}}{1 + e^{-0.11 - 0.66 - 1.67}} \approx 0.080$$

Female Jewish Democrats

$$\pi = \frac{e^\alpha e^{\beta^G_2 + \beta^R_2 + \beta^P}}{1 + e^\alpha e^{\beta^G_2 + \beta^R_2 + \beta^P}} = \frac{e^{-0.11 + 0.16 + 0.84}}{1 + e^{-0.11 + 0.16 + 0.84}} \approx 0.71$$

c. If we defined parameters such that the first category of a variable has value 0, then what would $\beta^G_2$ equal?

We know that the odds ratio is invariant under simultaneous permutation of rows and columns, however; sign will change. This is true because if the rows (or the columns) of the table are interchanged, the reciprocal of the odds ratio is obtained. In this case

$$e^{\beta^G_2} = \frac{1}{e^{\beta^G_1}} = e^{-\beta^G_1}$$

Hence $\beta^G_2 = -\beta^G_1 = -0.16$

d. Find the odds ratio that describes the conditional effect of gender?

$$\text{OR}_{new}(\text{Male, Female}) = \frac{e^\alpha e^{\beta^G_2}}{e^\alpha e^{\beta^G_1}} = e^{-0.16} \approx 0.852$$
\[
OR_{dd}(\text{Female, Male}) = \frac{e^{a} e^{\beta_1} G}{e^{a} e^{\beta_2} G} = e^{0.16} \approx 1.173
\]

The above is true since \( \frac{1}{0.852} = 1.173 \)

**Problem 5.5**

For data from Florida on \( Y = \) whether someone convicted of multiple murders recieves the death penalty (1=Yes, 0=No) the prediction equation is

\[
\text{logit}(\hat{\pi}) = -2.06 + 0.87d - 2.40v
\]

where \( d \) and \( v \) are defendant’s race and victim’s race (1=black, 0=white).  
**True or False:**

a. The estimated probability of the death penalty is lowest when the defendant is white and victims are black.

**True**

we have 4 possible probabilities in here

\[
\pi(d = 0, v = 1) = \frac{e^{-2.06 -2.40}}{1 + e^{-2.06 -2.40}} = 0.01143
\]

\[
\pi(d = 0, v = 0) = \frac{e^{-2.06}}{1 + e^{-2.06}} = 0.113
\]

\[
\pi(d = 1, v = 0) = \frac{e^{-2.06+0.87}}{1 + e^{-2.06+0.87}} = 0.233
\]

\[
\pi(d = 1, v = 1) = \frac{e^{-2.06+0.87-2.40}}{1 + e^{-2.06+0.87-2.40}} = 0.0268
\]

and clearly \( \pi(d = 0, v = 1) \) is the lowest probability.

b. Controlling for victim’s race, the estimated odds of the death penalty for white defendants equal 0.87 times the estimated odds for black defendants.

**False**

When controlling for victim’s race, the estimated odds of the death penalty for white defendants equal \( e^{0.87} \) times the estimated odds for black defendants.

c. If we will use \( d = 1 \) for white defendants and \( d = 0 \) for black defendants, then the estimated coefficient of \( d \) would be \( 1/0.87 \approx 1.15 \).

**False**

As shown in problem 5.4(c), the only thing we know due the inversion of the categories is that the new odds ratio will equal the reciprocal of the old odds ratio

\[
e^{\beta_{\text{new}}} = \frac{1}{e^{\beta_{\text{old}}}} = \frac{1}{e^{0.87}} = e^{-0.87}
\]

This implies that \( \beta_{\text{new}} = -0.87 \neq 1.15 \)

d. The intercept term -2.06 is the estimated probability of the death penalty when the defendant and victims were white.

**False**

Probability must be between 0 and 1 and therefore cannot be -2.06. The correct statement is "The intercept term -2.06 is the logit transformation of the death penalty’s probability when the defendant and victims were white".