STAT 550 Homework 7

Yinsen Miao [*]

November 16, 2015

Problem 1

$$\pi_3(x) = \frac{\exp(\alpha_3 + \beta_3 x)}{1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)}$$

whose derivative with respective to x is

$$\begin{aligned} \frac{\partial \pi_3(x)}{\partial x} &= \frac{\beta_3 \exp(\alpha_3 + \beta_3 x)\left[1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)\right]}{\left[1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)\right]^2} - \\ &\quad \frac{\exp(\alpha_3 + \beta_3 x)\left[\beta_1 \exp(\alpha_1 + \beta_1 x) + \beta_2 \exp(\alpha_2 + \beta_2 x)\right]}{\left[1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)\right]^2} \\ &= \frac{-\beta_1 \exp(\alpha_1 + \beta_1 x) - \beta_2 \exp(\alpha_2 + \beta_2 x)}{\left[1 + \exp(\alpha_1 + \beta_1 x) + \exp(\alpha_2 + \beta_2 x)\right]^2} \end{aligned} \tag{1}$$

Note the parameters $\alpha_3 = \beta_3 = 0$ for baseline category 3 because of identifiability reasons.

Therefore,

1. $\pi_3(x)$ is decreasing if $\beta_1 > 0$ and $\beta_2 > 0$.
2. $\pi_3(x)$ is increasing if $\beta_1 < 0$ and $\beta_2 < 0$.
3. $\pi_3(x)$ is nonmonotone if β_1 and β_2 have different signs, since the sign of $\pi^{'}(x)$ depends on x in this case.

Problem 2

1. Use the definition of Cumulative Logits model, for $j < i$

   $$\text{logit}\left[P(Y \le j \mid X = x)\right] - \text{logit}\left[P(Y \le i \mid X = x)\right] = (a_j - a_i) + (\beta_j - \beta_i)x \tag{2}$$

   Since the logit is an increasing function of $P(Y \le j \mid x)$, Equation 2 cannot be positive. However, it is positive if $\beta_j > \beta_i$ and x is positive or if $\beta_j < \beta_i$ and x is negative. Therefore with x taking over the $\mathbb{R}$, the cumulative probabilities are misordered for some range of x.

---

[*] Department of Statistics, Rice University, E-mail: ym11@rice.edu

2. When $x \in \{0, 1\}$, Equation 2 becomes one of the following case

$$\begin{aligned} \text{logit}\left[P\left(Y \leq j \mid X = 0\right)\right] - \text{logit}\left[P\left(Y \leq i \mid X = 0\right)\right] &= \left(a_j - a_i\right) \quad (3) \\ \text{logit}\left[P\left(Y \leq j \mid X = 1\right)\right] - \text{logit}\left[P\left(Y \leq i \mid X = 1\right)\right] &= \left(a_j + \beta_j\right) - \left(a_i + \beta_i\right) \quad (4) \end{aligned}$$

   where Equation 3 is negative because of the usual ordering constraint on $\{a_j\}$ or $a_j < a_i$ for $j < i$. To make Equation 4 negative, we should constraint $\{\alpha_j + \beta_j\}$ be increasing in j. However, with all those constraints the model is equivalent to a saturated model or

$$\begin{aligned} \text{logit}\left[P\left(Y \leq j \mid X = 0\right)\right] &= \alpha_j \\ \text{logit}\left[P\left(Y \leq j \mid X = 1\right)\right] &= \alpha_j + \beta_j \end{aligned}$$

Problem 3

1. We here use the log link for Poisson GLM. According to the outputs of R below, the fitted model is

$$\log \hat{\mu} = -0.42841 + 0.5893x \quad (5)$$

```
# Call:
# glm(formula = satell ~ weight_kg, family = poisson(link = "log"),
#     data = df1)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -2.9307  -1.9981  -0.5627   0.9298   4.9992
#
# Coefficients:
#             Estimate Std. Error z value Pr(>|z|)
# (Intercept) -0.42841    0.17893  -2.394   0.0167 *
# weight_kg    0.58930    0.06502   9.064   <2e-16 ***
# ---
# Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
#
# (Dispersion parameter for poisson family taken to be 1)
#
#     Null deviance: 632.79  on 172  degrees of freedom
# Residual deviance: 560.87  on 171  degrees of freedom
# AIC: 920.16
#
# Number of Fisher Scoring iterations: 5
```

   Based on the outputs, p value is quite low rejecting null hypothesis of a zero-valued coefficient on weight. For model given in Equation 5, we interpret it as one unit increase in weight has a multiplicative impact of $\exp(0.5893)$ on μ which means that a 1 kg increase in weight yields a 80.2% increase in the estimated mean.

2. Yes, because the variance of satellites (9.912) is much larger than the mean of satellites (2.919), while Poisson distribution should have identical mean and variance.

3. After adjustment for over-dispersion, the new Poisson model is

$$\log(\hat{\mu}) = -0.4284 + 0.5893x \tag{6}$$

```
# Call:
# glm(formula = satell ~ weight_kg, family = quasipoisson(link = "log"),
#     data = df1)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -2.9307  -1.9981  -0.5627   0.9298   4.9992
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  -0.4284     0.3168  -1.352    0.178
# weight_kg     0.5893     0.1151   5.120 8.17e-07 ***
# ---
# Signif. codes:  0 ?***?0.001 ?**?0.01 ??0.05 ??0.1 ??1
#
# (Dispersion parameter for quasipoisson family taken to be 3.13414)
#
#     Null deviance: 632.79  on 172  degrees of freedom
# Residual deviance: 560.87  on 171  degrees of freedom
# AIC: NA
#
# Number of Fisher Scoring iterations: 5
```

whose results are not largely deviated from non-adjusted model. However, from the estimate of the dispersion parameter (sum of squared Pearson residuals divided by the residual degrees of freedom in Page 150) given in the R outputs, the variance of our random component (the number of satellites for each weight_kg) is roughly three times the size of its mean which largely confirms our evidence in 2. Also notice that the parameters' standard errors are larger for the over-dispersion adjusted (when $scale = 3.134$) compared to the non-adjusted (when $scale = 1$).

4. The R outputs for negative binomial model is given below

$$\log\hat{\mu} = -0.8577 + 0.7575x \tag{7}$$

```
# Call:
# glm(formula = satell ~ weight_kg, family = neg.bin(theta = 1),
#     data = df1)
#
# Deviance Residuals:
#     Min       1Q   Median       3Q      Max
# -1.8741  -1.4323  -0.3331   0.4902   2.1886
```

```
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  -0.8577     0.3759  -2.282   0.0238 *
# weight_kg     0.7575     0.1464   5.175 6.35e-07 ***
# ---
# Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1
#
# (Dispersion parameter for Negative Binomial family taken to be 0.9089647)
#
#     Null deviance: 224.93  on 172  degrees of freedom
# Residual deviance: 203.61  on 171  degrees of freedom
# AIC: 752.8
#
# Number of Fisher Scoring iterations: 6
```

We can use cross validation or randomly divide the dataset into training and testing data set. Use the training set to fit the model, predict on the testing and then compute its MSE. In Figure 1, we observed the MSE of the negative binomial model is slightly smaller than that of the Poisson.

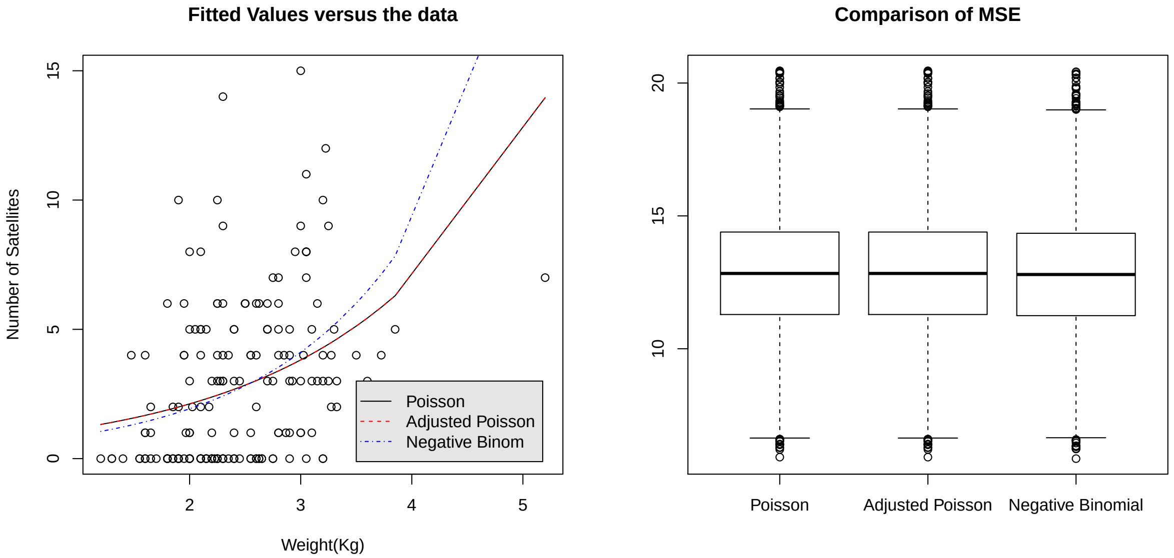

Figure 1: The fitted model versus the data (LEFT), MSE in the Cross Validation (RIGHT).

Problem 4

1. Based on the parameterization given in Textbook page 304, the R outputs for the cumulative logit model is given below. The estimated effect $\hat{\beta}_1 = -0.406$ and $\hat{\beta}_2 = -2.036$ suggests that the cumulative probability for very happy ordinal decreases when the traumatic score increases and is lower for the black than the white.

```
# Call:
# polr(formula = happy ~ race + trauma, data = df2)
#
# Coefficients:
#            Value Std. Error t value
# raceblack 2.0361    0.6859   2.968
# trauma    0.4056    0.1830   2.216
#
# Intercepts:
#                            Value   Std. Error t value
# very happy|pretty happy    -0.5181  0.3400    -1.5238
# pretty happy|not too happy  3.4006  0.5680     5.9872
#
# Residual Deviance: 148.407
# AIC: 156.407
```

2. Based on the outputs, there are two intercepts in the model. Since when $\underline{x} = 0$, the model becomes
$$\begin{aligned} \text{logit}\left[\hat{P}\left(Y \leq j \mid \underline{x}\right)\right] &= \alpha_j \\ \hat{P}\left(Y \leq j \mid \underline{x}\right) &= \frac{\exp\left(a_j + \underline{\beta}^T \underline{x}\right)}{1 + \exp\left(\alpha_j + \underline{\beta}^T \underline{x}\right)} \end{aligned} \tag{8}$$
where if we fixed $\underline{x}$, the intercept can be interpreted as the category separators. We consider the categorical outcomes as being driven by the replacement of α_is or the sequence of separating constants.

3. Let Y^* denote the latent variable whose has cdf $G\left(y^* - \underline{\beta}^T \underline{x}\right)$. Suppose thresholds $-\infty = \alpha_0 < \alpha_1 < \cdots < \alpha_J < \infty$ are cutpoints of the continuous scale where an observed response y satisfies
$$y = j \text{ iff } \alpha_{j-1} \leq y^* \leq \alpha_i$$
Therefore, y falls into j category j when the latent variable y^* falls into the jth intervals.
$$P\left(Y \leq j \mid \underline{x}\right) = P\left(Y^* \leq \alpha_i \mid \underline{x}\right) = G\left(\alpha_j - \underline{\beta}^T x\right)$$
where $\{a_j\}$ and $\underline{\beta}$ are the parameters of our interest in the latent model.

4. Since $P\left(Y = 2 \mid \underline{x}\right)$ and $P\left(Y \leq 1 \mid \underline{x}\right)$ are given as
$$\begin{aligned} P\left(Y \leq 2 \mid \underline{x}\right) &= \frac{\exp\left(3.4006 - 0.4056x_1 - 2.0361x_2\right)}{1 + \exp\left(3.4006 - 0.4056x_1 - 2.0361x_2\right)} \\ P\left(Y \leq 1 \mid \underline{x}\right) &= \frac{\exp\left(-0.5181 - 0.4056x_1 - 2.0361x_2\right)}{1 + \exp\left(-0.5181 - 0.4056x_1 - 2.0361x_2\right)} \end{aligned}$$
Therefore, the predict probabilities for the three levels of happy are shown in Figure 2.
$$\begin{aligned} P\left(Y = 1\right) &= P\left(Y \leq 1\right) \\ P\left(Y = 2\right) &= P\left(Y \leq 2\right) - P\left(Y \leq 1\right) \\ P\left(Y = 3\right) &= 1 - P\left(Y \leq 2\right) - P\left(Y \leq 1\right) \end{aligned}$$

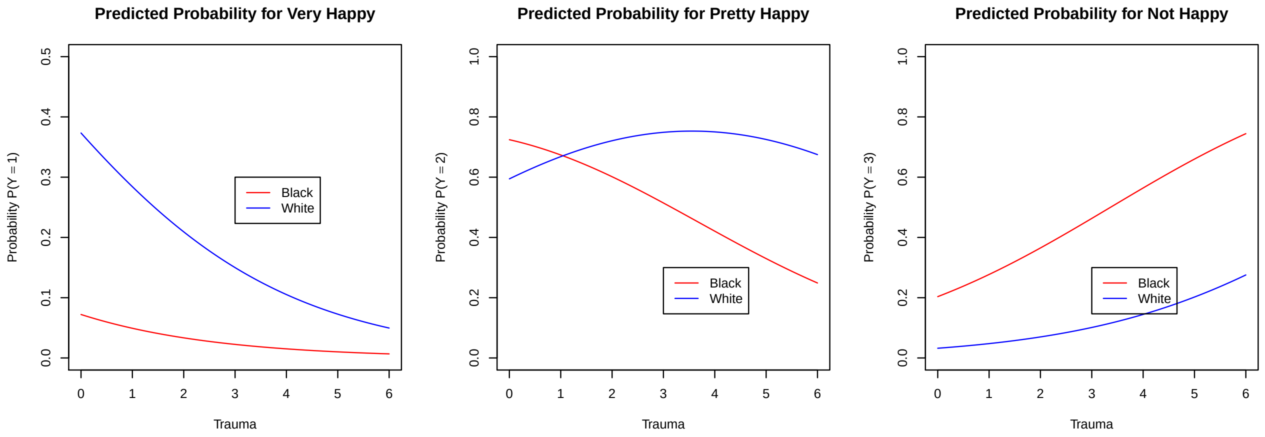

Figure 2: Estimated values of $P(Y=j)$ for $j \in \{1,2,3\}$ by $x_1 =$ traumatice scale,$x_2 =$ race.

Code

```
setwd("C:/Users/Yinsen/Desktop/Fall2015/STAT545/HW07/")
# Problem 3

www = "http://yinsenm.github.io/stat545/homework/horseshoecrab.csv"
df1 = read.csv(www)
model1 = glm(satell ~ weight_kg, family = poisson(link = "log"), data = df1)
summary(model1)

model2 = glm(satell ~ weight_kg, family = quasipoisson(link = "log"), data = df1)
summary(model2)

library(MASS)
model3 = glm(satell ~ weight_kg, family = neg.bin(theta = 1), data = df1)
summary(model3)

RMSE = function(model, test) {
  mean((df1$satell[test] - predict(model, df1[test,]))^2)
}


R = lapply(1:1e4, function(i) {
  test = sample(1:nrow(df1), size = nrow(df1)/2)
  modell1 = glm(satell ~ weight_kg, family = poisson(link = "log"), data = df1[-test,])
  modell2 = glm(satell ~ weight_kg, family = quasipoisson(link = "log"), data = df1[-test,])
  modell3 = glm(satell ~ weight_kg, family = neg.bin(theta = 1), data = df1[-test,])
```

```
  list(P = RMSE(modell1, test),
       AP = RMSE(modell2, test),
       NG = RMSE(modell3, test))
})

pdf("1.pdf", width = 12, height = 6)
par(mfrow = c(1,2))
idx = order(df1$weight_kg)
plot(df1$weight_kg, df1$satell, xlab = "Weight(Kg)",
     ylab = "Number of Satellites",
     main = "Fitted Values versus the data")
lines(df1$weight_kg[idx], fitted(model1)[idx], col = "black", lty = 1)
lines(df1$weight_kg[idx], fitted(model2)[idx], col = "red", lty = 2)
lines(df1$weight_kg[idx], fitted(model3)[idx], col = "blue", lty = 4)
legend(3.5, 3, c("Poisson", "Adjusted Poisson", "Negative Binom"),
       col = c("black", "red", "blue"),
       text.col = "black", lty = c(1, 2, 4),
       merge = TRUE, bg = "gray90")

RL = matrix(unlist(R), ncol = 3, byrow = T)
colnames(RL) = c("Poisson", "Adjusted Poisson", "Negative Binomial")
boxplot(RL, main = "Comparison of MSE")
par(mfrow = c(1,1))
dev.off()

# Problem 4
library(MASS)
www = "http://yinsenm.github.io/stat545/homework/GSS.csv"
df2 = read.csv(www)
df2$happy = ordered(df2$happy, levels = 1:3,
                    labels = c("very happy", "pretty happy", "not too happy"))
df2$race = factor(df2$race, levels = 0:1,
                  labels = c("white", "black"))
model4 = polr(happy ~ race + trauma,data = df2)
summary(model4)

(ctable <- coef(summary(model4)))
p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2
## combined table
(ctable <- cbind(ctable, "p value" = p))

P1 = function(x1, x2) {
  exp(-0.5181 - 0.4056*x1 - 2.0361*x2) /
    (1 + exp(-0.5181 - 0.4056*x1 - 2.0361*x2))
}

P2 = function(x1, x2) {
```

```
  exp(3.4006 - 0.4056*x1 - 2.0361*x2) /
    (1 + exp(3.4006 - 0.4056*x1 - 2.0361*x2))
}


truma = seq(0,6, by = 0.01)
p11 = sapply(truma, function(x) P1(x,0))
p21 = sapply(truma, function(x) P1(x,1))
p12 = sapply(truma, function(x) P2(x,0)) - p11
p22 = sapply(truma, function(x) P2(x,1)) - p21
p13 = 1 - p12 - p11
p23 = 1 - p22 - p21

library(latex2exp)
pdf(file = "2.pdf", height = 4, width = 11)
par(mfrow = c(1,3))
plot(truma, p21, type = "l", col = "red", ylim = c(0, 0.5),
     main = "Predicted Probability for Very Happy",
     ylab = latex2exp("Probability $P(Y = 1)$"), xlab = "Trauma")
lines(truma, p11, col = "blue")
legend(3, .3, c("Black", "White"),
       col = c("red", "blue"), lty = c(1,1))

plot(truma, p22, type = "l", col = "red", ylim = c(0, 1),
     main = "Predicted Probability for Pretty Happy",
     ylab = latex2exp("Probability $P(Y = 2)$"), xlab = "Trauma")
lines(truma, p12, col = "blue")
legend(3, .3, c("Black", "White"),
       col = c("red", "blue"), lty = c(1,1))

plot(truma, p23, type = "l", col = "red", ylim = c(0, 1),
     main = "Predicted Probability for Not Happy",
     ylab = latex2exp("Probability $P(Y = 3)$"), xlab = "Trauma")
lines(truma, p13, col = "blue")
legend(3, .3, c("Black", "White"),
       col = c("red", "blue"), lty = c(1,1))
par(mfrow = c(1,1))
truma = seq(0,6, by = 0.01)
dev.off()
```