Problem Set 3

Munan Hou

November 8, 2016

```
> library(dplyr)
```

1

```
> hipp <- read.csv("C:/YUKIHO/BU/MA 615/Classes/9.12.2016/class 9.12.2016/ma_568/Homework
> plot(hipp$xN,hipp$yN,type='l',col='blue', xlab="X", ylab="Y")
> hipp1 <- subset(hipp,spikes==1)
> hipp2 <- subset(hipp,spikes2==1)
> points(hipp1$xN,hipp1$yN,col='red',pch=19)
> points(hipp2$xN,hipp2$yN,col='green',pch=19)
```

2

```
> model1 <- glm(spikes~xN+yN,family = poisson,data=hipp)
> summary(model1)

Call:
glm(formula = spikes ~ xN + yN, family = poisson, data = hipp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5501  -0.3718  -0.2621  -0.1928   2.6004

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.24918    0.02760 -117.74   <2e-16 ***
xN          -0.50623    0.04724  -10.72   <2e-16 ***
yN          -1.18010    0.04477  -26.36   <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: 12311  on 41327  degrees of freedom
Residual deviance: 11388  on 41325  degrees of freedom
AIC: 15492

Number of Fisher Scoring iterations: 6
```

```
> lambda <- function(x,y) {
+   exp(model1$coefficients[1]+model1$coefficients[2]*x+model1$coefficients[3]*y)
+ }
> x_new <- seq(-1,1,.1)
> y_new <- seq(-1,1,.1)
> z<- outer(x_new,y_new,lambda)
> persp(x_new, y_new, z, theta = 30, phi = 30, expand = 0.5, col = "lightblue", xlab="X",
> phi <- seq(0, 2*pi, len = 201)
> xr <- cos(phi)
> yr <- sin(phi)
> lines(trans3d(xr,yr,lambda(xr,yr),res), col = "pink", lwd = 2)
```

3

```
> hipp$xN2 <- (hipp$xN)^2
> hipp$yN2 <- (hipp$yN)^2
> model2 <- glm(spikes~xN2+yN2+xN+yN,family = poisson,data=hipp)
> summary(model2)

Call:
glm(formula = spikes ~ xN2 + yN2 + xN + yN, family = poisson,
    data = hipp)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7260  -0.3359  -0.1564  -0.0580   3.5332

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.82026    0.03723  -48.89   <2e-16 ***
xN2         -5.81075    0.19772  -29.39   <2e-16 ***
yN2         -3.66241    0.13449  -27.23   <2e-16 ***
xN          -1.45219    0.09165  -15.85   <2e-16 ***
yN          -2.40928    0.09536  -25.27   <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: 12311.2  on 41327  degrees of freedom
Residual deviance:  9141.5  on 41323  degrees of freedom
AIC: 13249

Number of Fisher Scoring iterations: 7

> lambda2 <- function(x,y) {
+   exp(model2$coefficients[1]+model2$coefficients[2]*(x^2)+model2$coefficients[3]*(y^2)+
+   }
> z2<- outer(x_new,y_new,lambda2)
> persp(x_new, y_new, z2, theta = 30, phi = 30, expand = 0.5, col = "lightblue", xlab="X",
```

```
> phi <- seq(0, 2*pi, len = 201)
> xr <- cos(phi)
> yr <- sin(phi)
> lines(trans3d(xr,yr,lambda(xr,yr),res), col = "pink", lwd = 2)

> model2$aic

[1] 13249.46

> model1$aic

[1] 15492.16
```

4

```
> hipp_none <- select(hipp,spikes,xN,yN,xN2:yN2,spikes_hist1:spikes_hist20)
> p <- c()
> AIC <- c()
> for (i in 5:25) {
+   hipp_haha <- hipp_none[,1:i]
+   aa <- glm(spikes~.,family = poisson,data = hipp_haha)
+   AIC <- c(AIC,aa$aic)
+   p <- c(p,i-5)
+   }
> record <- data.frame("p"=p,"AIC"=AIC)
> plot(record$p,record$AIC,type="l",xlab="p",ylab="AIC")
> points(record$p,record$AIC)

> subset(record, AIC==min(AIC))

  p      AIC
9 8 12222.27
```

5

```
> hipp_none2 <- select(hipp,spikes,xN,yN,xN2,yN2,spikes_hist1:spikes_hist8, spikes2_hist1:
> p2 <- c()
> AIC2 <- c()
> for (i in 13:33) {
+   hipp_haha <- hipp_none2[,1:i]
+   aa <- glm(spikes~.,family = poisson,data = hipp_haha)
+   AIC2 <- c(AIC2,aa$aic)
+   p2 <- c(p2,i-13)
+   }
> record2 <- data.frame("p1"=8,"p2"=p2,"AIC2"=AIC2)
> subset(record2, AIC2==min(AIC2))

  p1 p2     AIC2
1  8  0 12222.27

>
```

6

```
> hipp_none3 <- select(hipp,spikes,xN,yN,xN2,yN2,spikes_hist1:spikes_hist20, spikes2_hist1
> p41 <- c()
> p42 <- c()
> AIC4 <- c()
> for (k in 5:25){
+   hipp_haha <- hipp_none3[,1:k] %>% select(-xN,-yN,-xN2,-yN2)
+   aa <- glm(spikes~.,family = poisson,data = hipp_haha)
+   AIC4 <- c(AIC4,aa$aic)
+   p41 <- c(p41,k-5)
+   p42 <- c(p42,0)
+   for (i in 26:45){
+     hipp_haha <- hipp_none3[,c(1:k,26:i)] %>% select(-xN,-yN,-xN2,-yN2)
+     aa <- glm(spikes~.,family = poisson,data = hipp_haha)
+     AIC4 <- c(AIC4,aa$aic)
+     p41 <- c(p41,k-5)
+     p42 <- c(p42,i-25)
+     }
+   }
> record4 <- data.frame("p1"=p41,"p2"=p42,"AIC"=AIC4)
> subset(record4,AIC==min(AIC))

    p1 p2      AIC
441 20 20 13534.88

> hipp_none4 <- hipp_none3 %>% select(-xN,-yN,-xN2,-yN2)
> summary(glm(spikes~.,family = poisson,data=hipp_none4))

Call:
glm(formula = spikes ~ ., family = poisson, data = hipp_none4)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0625  -0.2489  -0.1972  -0.1972   3.1930

Coefficients:
              Estimate Std. Error  z value Pr(>|z|)
(Intercept)    -3.94051    0.03742 -105.301  < 2e-16 ***
spikes_hist1   -2.55106    0.18268  -13.965  < 2e-16 ***
spikes_hist2   -1.42312    0.11547  -12.324  < 2e-16 ***
spikes_hist3   -0.45895    0.08818   -5.205 1.94e-07 ***
spikes_hist4    0.59444    0.06985    8.510  < 2e-16 ***
spikes_hist5    1.24482    0.06522   19.087  < 2e-16 ***
spikes_hist6    1.25323    0.06858   18.274  < 2e-16 ***
spikes_hist7    1.00806    0.07442   13.546  < 2e-16 ***
spikes_hist8    0.63389    0.07856    8.068 7.12e-16 ***
spikes_hist9    0.39560    0.07784    5.082 3.73e-07 ***
spikes_hist10   0.33893    0.07565    4.480 7.45e-06 ***
spikes_hist11   0.28919    0.07621    3.795 0.000148 ***
spikes_hist12   0.31453    0.07596    4.141 3.46e-05 ***
```

```
spikes_hist13   0.39616    0.07504    5.280 1.30e-07 ***
spikes_hist14   0.31213    0.07841    3.981 6.87e-05 ***
spikes_hist15   0.38548    0.07560    5.099 3.42e-07 ***
spikes_hist16   0.28811    0.07638    3.772 0.000162 ***
spikes_hist17   0.32802    0.07433    4.413 1.02e-05 ***
spikes_hist18   0.24784    0.07586    3.267 0.001087 **
spikes_hist19   0.20233    0.07453    2.715 0.006635 **
spikes_hist20   0.16042    0.07262    2.209 0.027186 *
spikes2_hist1   0.22766    0.07776    2.928 0.003417 **
spikes2_hist2   0.24561    0.08016    3.064 0.002184 **
spikes2_hist3   0.39466    0.07968    4.953 7.31e-07 ***
spikes2_hist4   0.46794    0.07878    5.940 2.85e-09 ***
spikes2_hist5   0.41320    0.08068    5.121 3.03e-07 ***
spikes2_hist6   0.36760    0.08117    4.529 5.93e-06 ***
spikes2_hist7   0.27017    0.08237    3.280 0.001038 **
spikes2_hist8   0.32439    0.07973    4.069 4.73e-05 ***
spikes2_hist9   0.22958    0.08322    2.759 0.005800 **
spikes2_hist10  0.19785    0.08286    2.388 0.016949 *
spikes2_hist11  0.17897    0.08314    2.153 0.031344 *
spikes2_hist12  0.24048    0.08084    2.975 0.002931 **
spikes2_hist13  0.25011    0.08106    3.086 0.002031 **
spikes2_hist14  0.14923    0.08450    1.766 0.077414 .
spikes2_hist15  0.13958    0.08404    1.661 0.096743 .
spikes2_hist16  0.21924    0.08027    2.731 0.006308 **
spikes2_hist17  0.17816    0.08212    2.170 0.030035 *
spikes2_hist18  0.28880    0.07794    3.706 0.000211 ***
spikes2_hist19  0.08885    0.08510    1.044 0.296475
spikes2_hist20  0.22993    0.07722    2.978 0.002904 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 12311.2  on 41327  degrees of freedom
Residual deviance:  9354.9  on 41287  degrees of freedom
AIC: 13535

Number of Fisher Scoring iterations: 6
```

7

```
> finalmodel <- glm(spikes~xN2+yN2+xN+yN+spikes_hist1+spikes_hist2+spikes_hist3+spikes_his
> hipp[,"num"] <- 1:nrow(hipp)
> rehipp <- subset(hipp,spikes==1)
> Xi <- rehipp[,"num"]
> rescaled <- exp(finalmodel$coefficients[1]+finalmodel$coefficients[2]*hipp[,"xN2"]+ fina
> rescaled <- (c(0,rescaled)+c(rescaled,0))/2
> rescaled <- rescaled[2:(length(rescaled)-1)]
> rescaled1 <- c()
```

```
> for (i in 1:(length(Xi)-1)) {
+   huhu <- sum(rescaled[Xi[i]:(Xi[i+1]-1)])
+   rescaled1 <- c(rescaled1,huhu)
+   }
> w <- seq(0,max(rescaled1),length.out = 1000)
> Femp <- numeric(length(w))
> Qs <- rescaled1[order(rescaled1)]
> for (i in 1:length(w)){
+   Femp[i] <- sum(Qs<=w[i])/length(Qs)
+   }
> plot(Femp,1-exp(-w),type = "l",lwd=2,ylab="Model_CDF", xlab="Empirical_CDF",main="Rescal
> abline(0,1,col="red")
> abline(1.36/sqrt(nrow(rehipp)),1,lty=2,col="red")
> abline(-1.36/sqrt(nrow(rehipp)),1,lty=2,col="red")
```