HW2 Answer Key

Jan 28 2017

Problem 1

(a) The log-likelihood is

$$l = \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_i-\mu_i)^2\right] = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2$$

(b) Given $\mu_i = \mu$, we choose $\hat{\mu}$ to maximize the log-likelihood, i.e.,

$$\max_{\hat{\mu}} l = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\hat{\mu})^2$$

First-order condition implies that

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\hat{\mu}) = 0$$

which simplifies to

$$n\hat{\mu} - \sum_{i=1}^{n} y_i = 0$$

Therefore, the maximum likelihood estimate for μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

i.e., the arithmetic mean of outcomes.

Problem 2

(a) First, note that $f' = 3(e^x-2)^2e^x$, therefore, we use

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{(e^{x_i}-2)^3}{3(e^{x_i}-2)^2e^{x_i}} = x_i - \frac{e^{x_i}-2}{3e^{x_i}} = x_i + \frac{2}{3}e^{-x_i} - \frac{1}{3}$$

to iterate for solution.[1] The code for function is:

```
NR<-function(x){
  counter <- 1
  while( abs(-1/3+2*(exp(-x))/3)>1e-10 & counter<=300000 ){
    x <- x-1/3+2*(exp(-x))/3
    counter <- counter+1
  }
  x
}
```

Note that, the condition `abs(-1/3+2*(exp(-x))/3)>1e-10` represents $|x_{i+1} - x_i| > 0$, which indicates that an additional iteration does not refine the value of x.

Then, simply by typing `NR(0)` we will get the answer 0.6931472.

[1] Of course you may use the original expression. However, using simplified equation will accelerate the computation and make it more precise.

(b) Now, $f(x) = (e^x - 2)^n$, so $f' = n(e^x - 2)^{n-1}e^x$. Similar to Part (a), the NR iteration is

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{(e^{x_i} - 2)^n}{n(e^{x_i} - 2)^{n-1}e^{x_i}} = x_i - \frac{e^{x_i} - 2}{3e^{x_i}} = x_i + \frac{2}{n}e^{-x_i} - \frac{1}{n}$$

We modify the interation function in R so that n enters as an argument. Further, the returning value now (`c(x,counter)`) is a vector that includes both the solution and the number of iteration.

```
NR<-function(x,n){
  counter <- 1
  while( abs(-1/n+2*(exp(-x))/n)>1e-10 & counter<=30000){
    x <- x-1/n+2*(exp(-x))/n
    counter <- counter+1
  }
  c(x,counter)
}
```

By entering from `NR(0,2)` to `NR(0,8)`, you will see

n	Solution	Number of Iterations
2	0.6931472	32
3	0.6931472	53
4	0.6931472	74
5	0.6931472	94
6	0.6931472	114
7	0.6931472	133
8	0.6931472	153

Notes: depending on your criteria, you may get different number of iterations.

Problem 3

(a) We use Poisson regression to specify the relation between real wage and our attributes of interest. By running the following code

```
library(foreign)
d<-read.dta("D:/org_example.dta")
ds<-subset(d,state=="CA" & year==2013)
poissonreg<-glm(rw~educ+female+age+wbho,ds,family="poisson"(link="log"))
summary(poissonreg)
```

You will get the result

```
Call:
glm(formula = rw ~ educ + female + age + wbho, family = poisson(link = "log"),
    data = ds)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-9.9092  -1.8621  -0.6145   0.9679  25.9487

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)       2.2051037  0.0151662  145.40   <2e-16 ***
educHS            0.3397354  0.0125034   27.17   <2e-16 ***
```

```
educSome college  0.4720058  0.0122888   38.41   <2e-16 ***
educCollege       0.9353610  0.0123980   75.44   <2e-16 ***
educAdvanced      1.1797471  0.0126980   92.91   <2e-16 ***
female           -0.1918608  0.0050182  -38.23   <2e-16 ***
age               0.0111549  0.0001857   60.08   <2e-16 ***
wbhoBlack        -0.2713582  0.0129974  -20.88   <2e-16 ***
wbhoHispanic     -0.1710364  0.0067647  -25.28   <2e-16 ***
wbhoOther        -0.0875031  0.0066407  -13.18   <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: 83415  on 6840  degrees of freedom
Residual deviance: 50183  on 6831  degrees of freedom
  (6948 observations deleted due to missingness)
AIC: Inf

Number of Fisher Scoring iterations: 5
```

Given other variables fixed, a female earns 19.2% less than a male, given other factors fixed.

Notes: some students replace NA in outcomes with 0 and get a different result (around 30%), or directly calculate $\mathrm{E}[Y|female = 1] - \mathrm{E}[Y|female = 0]$ rather than the marginal effect. They are also correct.

(b) Here is the LM test. Note that the variable wbho includes four choices, thus the degree of freedom is 3.

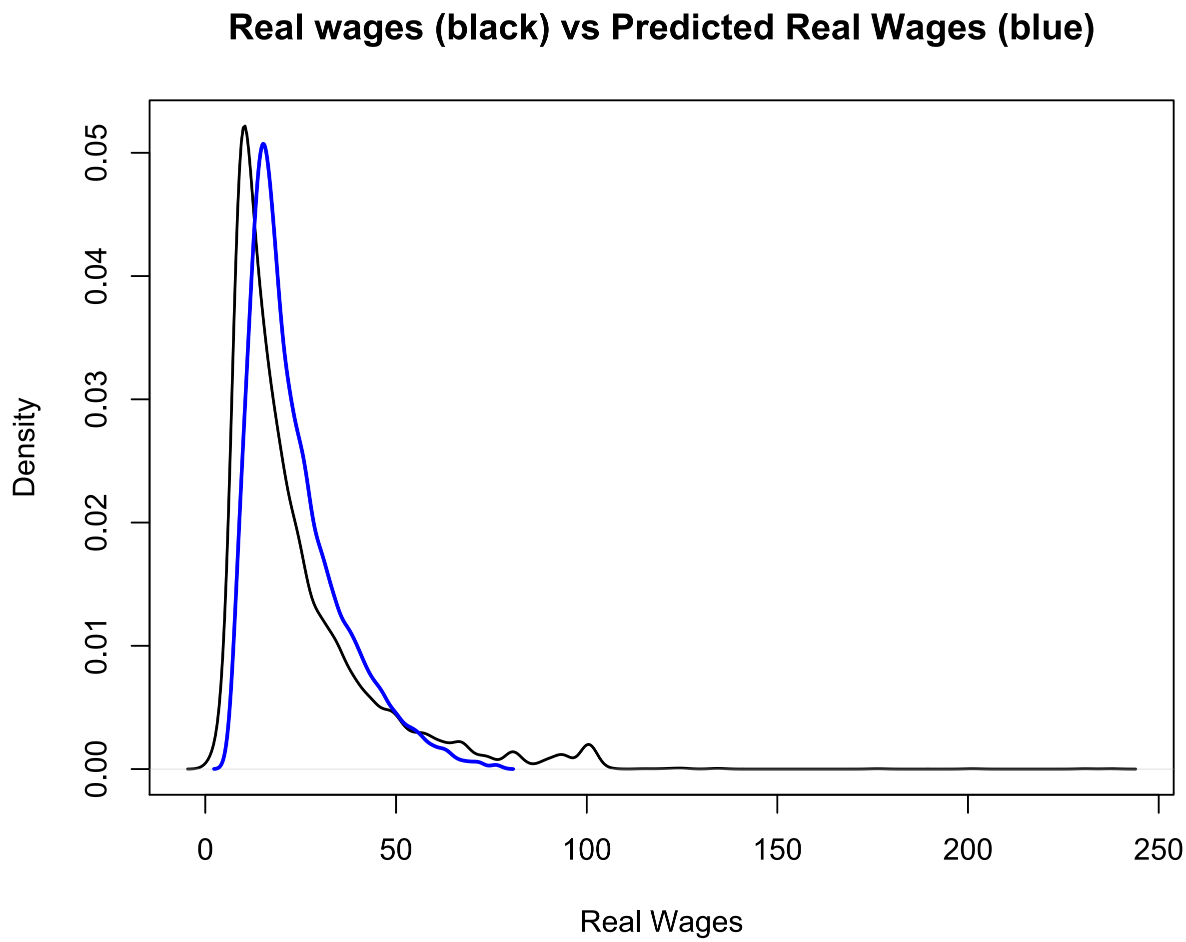
```
poissonreg2<-glm(rw~educ+female+age,ds,family="poisson"(link="log"))
LR<-(poissonreg2$deviance-poissonreg$deviance)
chi_crit<-qchisq(.95, df=3)
ifelse(LR>chi_crit,"Reject the restrictions", "Fail to reject the restrictions")
```

The null hypothesis is rejected at 95% significance level, i.e., ethnicity does play a role in wage outcomes.

(c) Using the regression from part (a), we use the following code:

```
ds$glm_predict<-predict(poissonreg,newdata=ds,type="response")
plot(density(ds$rw,na.rm=TRUE),lwd=1.5,main="Real wages (black) vs Predicted Real Wages (blue)",xlab="Re
lines(density(ds$glm_predict,na.rm=TRUE),lwd=2,col="blue")
```

The output is in the figure below. Here, we see that while the distributions of actual values and predicted values are pretty similar, the mode of the predicted values is skewed to the right a bit. This is because there are some very high wages in the data, which pull up the average from zero. However, the independent variables do not have sufficient variation to explain these high values, so the regression compensates by increasing the predicted value of other observations so that the means are the same for both predicted and actual values.


Real wages (black) vs Predicted Real Wages (blue)
Density
0.00
0.01
0.02
0.03
0.04
0.05
0
50
100
150
200
250
Real Wages