Midterm Exam Answer Key

Jan 28 2016

Problem 1

Since

$$\log f(y;\theta) = \log\left(\frac{\theta^y e^{-\theta}}{y!}\right) = y\log\theta - \theta - \log(y!)$$

we have

$$b(\theta) = \log\theta \qquad c(\theta) = -\theta \qquad d(y) = -\log(y!)$$

Note that Poisson distribution belongs to canonical exponential distribution family, we can directly calculate the expectation and variance by:

$$\mathrm{E}[Y] = -\frac{c'(\theta)}{b'(\theta)} = -\frac{-1}{\left(\frac{1}{\theta}\right)} = \theta$$

$$\mathrm{Var}[Y] = -\frac{b''(\theta)\mathrm{E}[Y] + c''(\theta)}{(b'(\theta))^2} = -\frac{\left(-\frac{1}{\theta^2}\right)\theta + 0}{\left(\frac{1}{\theta}\right)^2} = \theta$$

2 for rewriting pdf, 1 for each part of $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$, 2 for mean and 3 for variance. You get the points if you calculate the mean and variance in different way but get the correct answer.

Problem 2

From the link function, we know that $\mathrm{E}[Y] = \mu = e^\beta$. Note that $\mathrm{E}[Y] = \theta$ (as we have derived above in Problem 1), substituting θ by e^β gives the density function:

$$f(y_i;\beta) = \frac{(e^\beta)^{y_i} e^{-e^\beta}}{y_i!} = \frac{e^{\beta y_i} e^{-e^\beta}}{y_i!}$$

Assuming that y_i is i.i,d, the log-likelihood function is

$$l = \sum_{i=1}^N \log\left(\frac{e^{\beta y_i} e^{-e^\beta}}{y_i!}\right) = \sum_{i=1}^N \left(\beta y_i - e^\beta - \log(y_i!)\right) = \beta\sum_{i=1}^N y_i - Ne^\beta - \sum_{i=1}^N \log(y_i!)$$

Taking derivative with respect to β and we get the score function

$$U = \frac{dl}{d\beta} = \sum_{i=1}^N y_i - Ne^\beta$$

Let $U = 0$ we can solve for the value of β that maximizes this log-likelihood

$$\hat\beta_{MLE} = \log\left(\frac{1}{N}\sum_{i=1}^N y_i\right) = \log\bar y$$

4 for density function/relation between β and θ, 3 for likelihood, 2 for score/FOC and 1 for result of β. You may also find the MLE estimate for θ first and then find β. Partial credit is given if you correctly estimate θ.

Problem 3

```
ds<-subset(d,year==2008)
```

This line trims the data by keeping observations from year of 2008 and save it as a sub dataset called `ds`.

```
poissonreg<-glm(hourslw~educ+female+age,ds,family="poisson"(link="log"))
```

This line estimates a poisson regression with log link, regressing `hourslw` on `educ`, `female`, and `age`, using the sample `ds`. The estimate is saved as `poissonreg`.

```
summary(poissonreg)
```

It summaries the result `poissonreg`, showing the basic info of model, each coefficient estimation and its std.err, p-values, etc.

```
ds$glm_predict<-predict(poissonreg)
```

It makes in-sample prediction **on the scale of the linear predictor, not the outcome** and save the fitted values into the column `ds$glm_predict`.

Since the outcomes variables might be not reported, **we should use the original dataset as "new-data" to generate predictions:**

```
ds$glm_predict<-predict(poissonreg,newdata=ds)
```

2 for each command explanation and 2 for error finding.

Problem 4

a) Having an advanced degree reduces the probability of being out of labor force by 37.9%

$$\frac{e^{0.305-1.713}}{1+e^{0.305-1.713}} - \frac{e^{0.305}}{1+e^{0.305}} = -0.379$$

Or: Having an advanced degree reduces the log of odds ratio of being out of labor force by 1.71 points.

5 points.

b) The idea here is to run two logit regressions, similar to how we build up the log-odds ratios for the multi-nomial logit. One intuitive way of doing this would be comparing `nilf` and employment to unemployment. For example:

```
d1<-subset(d, nilf==1|unem==1)
logitreg1<-glm(nilf~educ,d1,family=binomial(link="logit"))

d2<-subset(d, nilf==0)
d2$emp<-1-d$unem
logitreg2<-glm(emp~educ,d2,family=binomial(link="logit"))
```

By running these regressions, we're basically comparing unemployment to other outcomes, viewing un-employment like the base group. This is essentially like multi-nomial logit.

5 points. Using multinomial logit/nested logit is also acceptable.

Problem 5

```
# calculate LR ,restricted deviance minus unrestricted deviance
LR<-(logitreg1$deviance-logitreg2$deviance)
# get critical value for chi square at quantile 95%, df=4-1=3
chi_critval <- qchisq(.95, df=3)
ifelse(LR>chi_critval,"Reject Null Hypothesis", "Fail to Reject Null Hypothesis")
```

5 for specifying the deviance difference in and 5 for make a test.

Problem 6

$$\begin{aligned}
\frac{d\pi_{ij}}{dp_j} &= \frac{e^{\beta_j+\alpha p_j}\alpha\left(1+\sum_{s\neq 1} e^{\beta_s+\alpha p_s}\right) - e^{\beta_j+\alpha p_j}\left(e^{\beta_j+\alpha p_j}\alpha\right)}{\left(1+\sum_{s\neq 1} e^{\beta_s+\alpha p_s}\right)^2} \\
&= \frac{e^{\beta_j+\alpha p_j}}{1+\sum_{s\neq 1} e^{\beta_s+\alpha p_s}} \frac{\alpha\left(1+\sum_{s\neq 1, s\neq j} e^{\beta_j+\alpha p_j}\right)}{1+\sum_{s\neq 1} e^{\beta_s+\alpha p_s}} \\
&= \pi_{ij}(1-\pi_{ij})\alpha
\end{aligned}$$

Alternatively, writing in the form of

$$\frac{d\pi_{ij}/\pi_{ij}}{dp_j} = (1-\pi_{ij})\alpha$$

or in the form elasticity,

$$\frac{d\pi_{ij}/\pi_{ij}}{dp_j/p_j} = (1-\pi_{ij})\alpha p_j$$

Therefore, in order to know the price effect, we only need to estimate α but not β_js.

10 max.