Exercises: Week 2

Prof. Conlon

Due: 2/8/21

1. Let's load the Boston HMDA data.

The function should take the following arguments:

- dir: debt to income ratio
- hir : housing to income ratio
- single : dummy for single borrower
- self : dummy for self-employed

```
library("Ecdat")
```

```
## Loading required package: Ecfun

##
## Attaching package: 'Ecfun'

## The following object is masked from 'package:base':
##
##     sign

##
## Attaching package: 'Ecdat'

## The following object is masked from 'package:datasets':
##
##     Orange
```

```
data("Hmda")

probit <- glm(deny ~ dir + hir + single + self, data = Hmda, family = binomial(link = "probit"))
logit <- glm(deny ~ dir + hir + single + self, data = Hmda, family = binomial(link = "logit"))

summary(logit)
```

```
##
## Call:
## glm(formula = deny ~ dir + hir + single + self, family = binomial(link = "logit"),
##     data = Hmda)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.6537  -0.5238  -0.4531  -0.3694   2.7437
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.1539     0.2845 -14.598  < 2e-16 ***
## dir           6.1184     0.9174   6.670 2.57e-11 ***
```

```
## hir          -0.7501     1.0418  -0.720 0.471516
## singleyes     0.4503     0.1306   3.447 0.000567 ***
## selfyes       0.3609     0.1885   1.915 0.055524 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1744.2  on 2379  degrees of freedom
## Residual deviance: 1644.8  on 2375  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 1654.8
##
## Number of Fisher Scoring iterations: 5
```

2. Consider the regression model of the logit regression:

$$deny_i = F(\beta_1 \cdot dir_i + \beta_2 \cdot hir_i + \beta_3 \cdot single_i + \beta_4 \cdot self_i)$$

   For a single observation compute the contribution to the log-likelihood (analytically)

3. For a single observation compute the Score (analytically).
4. Compute the Hessian Matrix and Fisher information (analytically).
5. Code up the Fisher Information for the logit model above $I(\widehat{\beta})$ using the Hessian Matrix.
6. Code up the Fisher Information for the logit model above $I(\widehat{\beta})$ using the score method.
7. Compute the standard errors from the Fisher information and compare them to the standard errors reported from the regression. How do they compare?
8. Generate $n = 100$ observations where $\lambda = 15$ from a poisson model:

$$Y_i \sim Pois(\lambda)$$

9. The poisson distribution is a discrete distribution for count data where the p.m.f. is given by:

$$Pr(Y_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}a$$

10. Write the log-likelihood $\ell(y_1, \ldots, y_n; \lambda)$ (analytically).
11. Write the Score contribution $\mathcal{S}_i(y_i; \lambda)$ (analytically).
12. Write the Hessian Contribution $\mathcal{H}_i(y_i; \lambda)$ (analytically).
13. Code up the log-likelihood function

```r
pois_log_lik <- function(lambda,y){

  return(ll)
}
```

14. Find the value of λ that maximizes your log likelihood using `optim` in R.
15. Write a function that returns the standard error of $\hat{\lambda}$:

```r
pois_se <- function(lambda_hat,y){

  return(se)
}
```