Econ 613 Assignment 2

Xiaolin Ding

03/17/21

```
library(bayesm)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v stringr 1.4.0
## v tidyr   1.1.3     v forcats 0.5.1
## v readr   1.4.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data(margarine)
choiceprice = margarine$choicePrice
demos = margarine$demo
```

Exercise 1 Data Description

Average prices of each product

```
print(apply(as.matrix(choiceprice[,3:12]), 2, mean))
```

```
##   PPk_Stk   PBB_Stk   PFl_Stk  PHse_Stk  PGen_Stk  PImp_Stk   PSS_Tub   PPk_Tub
## 0.5184362 0.5432103 1.0150201 0.4371477 0.3452819 0.7807785 0.8250895 1.0774094
##   PFl_Tub  PHse_Tub
## 1.1893758 0.5686734
```

Dispersion (variance) in prices of each product

```r
print(apply(as.matrix(choiceprice[,3:12]), 2, var))
```

```
##      PPk_Stk      PBB_Stk      PFl_Stk     PHse_Stk     PGen_Stk     PImp_Stk
## 0.0226554865 0.0144797566 0.0018399974 0.0141208621 0.0012366513 0.0131437214
##      PSS_Tub      PPk_Tub      PFl_Tub     PHse_Tub
## 0.0037468593 0.0008836431 0.0001975293 0.0052497277
```

Average price of all products

```r
average = mean(unlist(choiceprice[,3:12]))
print(average)
```

```
## [1] 0.7300423
```

Dispersion (variance) in prices of all products

```r
dispersion = var(unlist(choiceprice[,3:12]))
print(dispersion)
```

```
## [1] 0.08415785
```

Market share of each product

```r
marketshare1 = data.frame(matrix(ncol = 10, nrow = 0))
colnames(marketshare1) = c(names(choiceprice[,3:12]))
for (i in 1:10){
  count = 0
  for (j in 1:nrow(choiceprice)){
    if(choiceprice[j,2]==i){
      count = count+1
    }
    marketshare1[1,i] = count/nrow(choiceprice)
  }
}
print(as.matrix(marketshare1))
```

```
##     PPk_Stk   PBB_Stk    PFl_Stk  PHse_Stk  PGen_Stk   PImp_Stk    PSS_Tub
## 1 0.3950783 0.1563758 0.05436242 0.1326622 0.0704698 0.01655481 0.07136465
##      PPk_Tub    PFl_Tub   PHse_Tub
## 1 0.04541387 0.05033557 0.00738255
```

Market share by price bins

```r
marketshare2 = data.frame(matrix(ncol = 2, nrow = 0))
colnames(marketshare2) = c('Below average', 'Above average')
count = 0
for (i in 1:nrow(choiceprice)){
```

```r
  if (choiceprice[i,choiceprice$choice[i]+2]<=average){
    count = count + 1
  }
}
marketshare2[1,1] = count/nrow(choiceprice)
marketshare2[1,2] = 1 - marketshare2[1,1]
print(as.matrix(marketshare2))
```

```
##   Below average Above average
## 1     0.7863535     0.2136465
```

Most popular choice for different types of households

```r
# Extract hhid for different types of households
incomeba = c()
incomeaa = c()
for (i in 1:nrow(demos)){
  if (demos$Income[i]<=mean(demos$Income)){
    incomeba = append(incomeba, demos$hhid[i])
  }
  else {
    incomeaa = append(incomeaa, demos$hhid[i])
  }
}
fs3orless = c()
fsmorethan3 = c()
for (i in 1:nrow(demos)){
  if (demos$Fam_Size[i]<=3){
    fs3orless = append(fs3orless, demos$hhid[i])
  }
  else {
    fsmorethan3 = append(fsmorethan3, demos$hhid[i])
  }
}
college = c()
nocollege = c()
for (i in 1:nrow(demos)){
  if (demos$college[i]==1){
    college = append(college, demos$hhid[i])
  }
  else {
    nocollege = append(nocollege, demos$hhid[i])
  }
}
whtcollar = c()
notwhtcollar = c()
for (i in 1:nrow(demos)){
  if (demos$whtcollar[i]==1){
    whtcollar = append(whtcollar, demos$hhid[i])
  }
  else {
    notwhtcollar = append(notwhtcollar, demos$hhid[i])
```

```
  }
}
retired = c()
notretired = c()
for (i in 1:nrow(demos)){
  if (demos$retired[i]==1){
    retired = append(retired, demos$hhid[i])
  }
  else {
    notretired = append(notretired, demos$hhid[i])
  }
}

# Define the function mode()
mode = function(x) {
  ux = unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}

# Most popular choice by household types
popularchoice = data.frame(matrix(ncol = 10, nrow = 0))
colnames(popularchoice) = c('Income below average', 'Income above average',
                            'Family size 3 or less', 'Family size more than 3',
                            'College', 'No college', 'White collar',
                            'Not white collar', 'Retired', 'Not retired')
popularchoice[1,1] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,incomeba))$choice)
popularchoice[1,2] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,incomeaa))$choice)
popularchoice[1,3] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,fs3orless))$choice)
popularchoice[1,4] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,fsmorethan3))$choice)
popularchoice[1,5] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,college))$choice)
popularchoice[1,6] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,nocollege))$choice)
popularchoice[1,7] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,whtcollar))$choice)
popularchoice[1,8] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,notwhtcollar))$choice)
popularchoice[1,9] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,retired))$choice)
popularchoice[1,10] =
  mode(subset(choiceprice,is.element(choiceprice$hhid,notretired))$choice)
print(as.matrix(popularchoice))
```

```
##   Income below average Income above average Family size 3 or less
## 1                    1                    1                     1
##   Family size more than 3 College No college White collar Not white collar
## 1                       1       1          1            1                1
##   Retired Not retired
## 1       1           1
```

Exercise 2 First Model

```
# Create the data "cp" by adding the variable "price" and "type" to the data
# "choiceprice", where "price" is the price of the product chosen and "type"
# is the type of the product chosen
cp = choiceprice
cp$price = NA
cp$type = NA
for (i in 1:nrow(cp)){
  cp$price[i] = cp[,cp$choice[i]+2][i]
  cp$type[i] = colnames(cp[cp$choice[i]+2])
}
cp[1:10,]
```

```
##       hhid choice PPk_Stk PBB_Stk PFl_Stk PHse_Stk PGen_Stk PImp_Stk PSS_Tub
## 1  2100016      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
## 2  2100016      1    0.63    0.67    0.99     0.57     0.36     1.03    0.85
## 3  2100016      1    0.29    0.50    0.99     0.57     0.36     0.69    0.79
## 4  2100016      1    0.62    0.61    0.99     0.57     0.36     0.75    0.85
## 5  2100016      1    0.50    0.58    0.99     0.45     0.33     0.72    0.85
## 6  2100016      4    0.58    0.45    0.99     0.45     0.33     0.72    0.85
## 7  2100016      1    0.29    0.51    0.99     0.29     0.33     0.72    0.85
## 8  2100024      1    0.66    0.45    1.08     0.57     0.36     0.93    0.85
## 9  2100024      4    0.66    0.59    1.08     0.57     0.36     0.93    0.85
## 10 2100024      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
##    PPk_Tub PFl_Tub PHse_Tub price     type
## 1     1.09    1.19     0.33  0.66  PPk_Stk
## 2     1.09    1.19     0.37  0.63  PPk_Stk
## 3     1.09    1.19     0.59  0.29  PPk_Stk
## 4     1.09    1.19     0.59  0.62  PPk_Stk
## 5     1.07    1.19     0.59  0.50  PPk_Stk
## 6     1.07    1.19     0.59  0.45 PHse_Stk
## 7     1.07    1.19     0.59  0.29  PPk_Stk
## 8     1.09    1.19     0.33  0.66  PPk_Stk
## 9     1.09    1.34     0.33  0.57 PHse_Stk
## 10    1.09    1.19     0.33  0.66  PPk_Stk
```

Mixed logit model

- Choice probability

$$p_i(j) = \frac{e^{c_j + \alpha * price_{ij}}}{\sum_l e^{c_l + \alpha * price_{il}}}$$

- Log likelihood function

$$L(c, \alpha | Data = price) = \sum_{i=1}^{n} \sum_{l} log(pi(l))$$

- Maximum likelihood estimation

$$\max_{c,\alpha} L(c, \alpha | Data)$$

```
# Log likelihood function for the data "cp"
likelihood1 = function(par){
  cp$constant = 0
  cp$constant[cp$type=='PPk_Stk'] = par[1]
```

```r
  cp$constant[cp$type=='PBB_Stk'] = par[2]
  cp$constant[cp$type=='PFl_Stk'] = par[3]
  cp$constant[cp$type=='PHse_Stk'] = par[4]
  cp$constant[cp$type=='PGen_Stk'] = par[5]
  cp$constant[cp$type=='PImp_Stk'] = par[6]
  cp$constant[cp$type=='PSS_Tub'] = par[7]
  cp$constant[cp$type=='PPk_Tub'] = par[8]
  cp$constant[cp$type=='PFl_Tub'] = par[9]
  pr = exp(cp$constant+par[10]*cp$price)/(exp(par[1]+par[10]*cp$PPk_Stk)
                                          +exp(par[2]+par[10]*cp$PBB_Stk)
                                          +exp(par[3]+par[10]*cp$PFl_Stk)
                                          +exp(par[4]+par[10]*cp$PHse_Stk)
                                          +exp(par[5]+par[10]*cp$PGen_Stk)
                                          +exp(par[6]+par[10]*cp$PImp_Stk)
                                          +exp(par[7]+par[10]*cp$PSS_Tub)
                                          +exp(par[8]+par[10]*cp$PPk_Tub)
                                          +exp(par[9]+par[10]*cp$PFl_Tub)
                                          +exp(par[10]*cp$PHse_Tub))
  return(-sum(log(pr)))
}

# Maximum likelihood estimation
est1 = optim(runif(10),likelihood1,method='BFGS')
```

The coefficient on price is -6.6565787. Therefore, an increase in the price of product j will decrease the probability that household i buys product j.

Exercise 3 Second Model

```r
# Create the data "ci" by changing the variable "price" to "income"
ci = cp
colnames(ci)[13] = 'income'
ci$income = demos$Income[match(ci$hhid, demos$hhid)]
ci[1:10,]
```

```
##       hhid choice PPk_Stk PBB_Stk PFl_Stk PHse_Stk PGen_Stk PImp_Stk PSS_Tub
## 1  2100016      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
## 2  2100016      1    0.63    0.67    0.99     0.57     0.36     1.03    0.85
## 3  2100016      1    0.29    0.50    0.99     0.57     0.36     0.69    0.79
## 4  2100016      1    0.62    0.61    0.99     0.57     0.36     0.75    0.85
## 5  2100016      1    0.50    0.58    0.99     0.45     0.33     0.72    0.85
## 6  2100016      4    0.58    0.45    0.99     0.45     0.33     0.72    0.85
## 7  2100016      1    0.29    0.51    0.99     0.29     0.33     0.72    0.85
## 8  2100024      1    0.66    0.45    1.08     0.57     0.36     0.93    0.85
## 9  2100024      4    0.66    0.59    1.08     0.57     0.36     0.93    0.85
## 10 2100024      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
##    PPk_Tub PFl_Tub PHse_Tub income    type
## 1     1.09    1.19     0.33   32.5 PPk_Stk
## 2     1.09    1.19     0.37   32.5 PPk_Stk
## 3     1.09    1.19     0.59   32.5 PPk_Stk
## 4     1.09    1.19     0.59   32.5 PPk_Stk
## 5     1.07    1.19     0.59   32.5 PPk_Stk
```

```
## 6      1.07     1.19      0.59    32.5 PHse_Stk
## 7      1.07     1.19      0.59    32.5  PPk_Stk
## 8      1.09     1.19      0.33    17.5  PPk_Stk
## 9      1.09     1.34      0.33    17.5 PHse_Stk
## 10     1.09     1.19      0.33    17.5  PPk_Stk
```

Mixed logit model

- Choice probability

$$p_i(j) = \frac{e^{c_j+\beta*income_i}}{\sum_l e^{c_l+\beta*income_i}}$$

- Log likelihood function

$$L(c, \alpha|Data = income) = \sum_{i=1}^{n} \sum_{l} log(pi(l))$$

- Maximum likelihood estimation

$$\max_{c,\beta} L(c, \beta|Data)$$

```
# Log likelihood function for the data "ci"
likelihood2 = function(par){
  ci$constant = 0
  ci$constant[ci$type=='PPk_Stk'] = par[1]
  ci$constant[ci$type=='PBB_Stk'] = par[2]
  ci$constant[ci$type=='PFl_Stk'] = par[3]
  ci$constant[ci$type=='PHse_Stk'] = par[4]
  ci$constant[ci$type=='PGen_Stk'] = par[5]
  ci$constant[ci$type=='PImp_Stk'] = par[6]
  ci$constant[ci$type=='PSS_Tub'] = par[7]
  ci$constant[ci$type=='PPk_Tub'] = par[8]
  ci$constant[ci$type=='PFl_Tub'] = par[9]
  pr = exp(ci$constant+par[10]*ci$income)/(exp(par[1]+par[10]*ci$income)
                                          +exp(par[2]+par[10]*ci$income)
                                          +exp(par[3]+par[10]*ci$income)
                                          +exp(par[4]+par[10]*ci$income)
                                          +exp(par[5]+par[10]*ci$income)
                                          +exp(par[6]+par[10]*ci$income)
                                          +exp(par[7]+par[10]*ci$income)
                                          +exp(par[8]+par[10]*ci$income)
                                          +exp(par[9]+par[10]*ci$income)
                                          +exp(par[10]*ci$income))
  return(-sum(log(pr)))
}

# Maximum likelihood estimation
est2 = optim(runif(10),likelihood2,method='BFGS')
```

The coefficient on income is 0.7343736. Therefore, an increase in household i's income will increase the probability that household i buys product j.

Exercise 4 Marginal Effects

- The marginal effect of the price of product k on household i's demand of product j is

$$\frac{\partial pij}{\partial price_{ik}} = p_{ij}(\delta_{ijk} - p_{ik})\alpha = p_{ij}(\delta_{ijk} - p_{ik})(-6.6565787)$$

  , where δ_{ijk} is an indicator variable equal to 1 if $j = k$ and equal to 0 otherwise.
- The marginal effect of household i's income on household i's demand of product j is

$$\frac{\partial p_{ij}}{\partial income_i} = p_{ij}(1-1)\beta = 0$$

  .

Exercise 5 IIA

Full set of alternatives

```r
# Create the data "cf" by adding he variable "income" to the data "cp"
cf = cp
cf$income = ci$income
cf[1:10,]
```

```
##       hhid choice PPk_Stk PBB_Stk PFl_Stk PHse_Stk PGen_Stk PImp_Stk PSS_Tub
## 1  2100016      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
## 2  2100016      1    0.63    0.67    0.99     0.57     0.36     1.03    0.85
## 3  2100016      1    0.29    0.50    0.99     0.57     0.36     0.69    0.79
## 4  2100016      1    0.62    0.61    0.99     0.57     0.36     0.75    0.85
## 5  2100016      1    0.50    0.58    0.99     0.45     0.33     0.72    0.85
## 6  2100016      4    0.58    0.45    0.99     0.45     0.33     0.72    0.85
## 7  2100016      1    0.29    0.51    0.99     0.29     0.33     0.72    0.85
## 8  2100024      1    0.66    0.45    1.08     0.57     0.36     0.93    0.85
## 9  2100024      4    0.66    0.59    1.08     0.57     0.36     0.93    0.85
## 10 2100024      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
##    PPk_Tub PFl_Tub PHse_Tub price     type income
## 1     1.09    1.19     0.33  0.66  PPk_Stk   32.5
## 2     1.09    1.19     0.37  0.63  PPk_Stk   32.5
## 3     1.09    1.19     0.59  0.29  PPk_Stk   32.5
## 4     1.09    1.19     0.59  0.62  PPk_Stk   32.5
## 5     1.07    1.19     0.59  0.50  PPk_Stk   32.5
## 6     1.07    1.19     0.59  0.45 PHse_Stk   32.5
## 7     1.07    1.19     0.59  0.29  PPk_Stk   32.5
## 8     1.09    1.19     0.33  0.66  PPk_Stk   17.5
## 9     1.09    1.34     0.33  0.57 PHse_Stk   17.5
## 10    1.09    1.19     0.33  0.66  PPk_Stk   17.5
```

Mixed logit model

- Choice probability

$$p_i(j) = \frac{e^{c_j + \alpha * price_{ij} + \beta * income_i}}{\sum_l e^{c_l + \alpha * price_{ij} + \beta * income_i}}$$

- Log likelihood function

$$L(c, \alpha | Data = price, income) = \sum_{i=1}^{n} \sum_{l} log(pi(l))$$

- Maximum likelihood estimation

$$\max_{c,\alpha,\beta} L(c, \alpha, \beta | Data)$$

```
# Log likelihood function for the data "cf"
likelihood3 = function(par){
  cf$constant = 0
  cf$constant[cf$type=='PPk_Stk'] = par[1]
  cf$constant[cf$type=='PBB_Stk'] = par[2]
  cf$constant[cf$type=='PFl_Stk'] = par[3]
  cf$constant[cf$type=='PHse_Stk'] = par[4]
  cf$constant[cf$type=='PGen_Stk'] = par[5]
  cf$constant[cf$type=='PImp_Stk'] = par[6]
  cf$constant[cf$type=='PSS_Tub'] = par[7]
  cf$constant[cf$type=='PPk_Tub'] = par[8]
  cf$constant[cf$type=='PFl_Tub'] = par[9]
  pr = exp(cf$constant+par[10]*cf$price+par[11]*cf$income)/
    (exp(par[1]+par[10]*cf$PPk_Stk+par[11]*cf$income)
    +exp(par[2]+par[10]*cf$PBB_Stk+par[11]*cf$income)
    +exp(par[3]+par[10]*cf$PFl_Stk+par[11]*cf$income)
    +exp(par[4]+par[10]*cf$PHse_Stk+par[11]*cf$income)
    +exp(par[5]+par[10]*cf$PGen_Stk+par[11]*cf$income)
    +exp(par[6]+par[10]*cf$PImp_Stk+par[11]*cf$income)
    +exp(par[7]+par[10]*cf$PSS_Tub+par[11]*cf$income)
    +exp(par[8]+par[10]*cf$PPk_Tub+par[11]*cf$income)
    +exp(par[9]+par[10]*cf$PFl_Tub+par[11]*cf$income)
    +exp(par[10]*cf$PHse_Tub+par[11]*cf$income))
  return(-sum(log(pr)))
}

# Maximum likelihood estimation
est3 = optim(runif(11),likelihood3,method='BFGS')
```

The estimated coefficients on price and income are -6.6564485 and 0.3516711, respectively.

Subset of alternatives

```
# Create the data "cr" by removing the choice "PImp_Stk" from the data "cf"
cr = cf[!(cf$type=='PImp_Stk'),]
cr[1:10,]
```

```
##       hhid choice PPk_Stk PBB_Stk PFl_Stk PHse_Stk PGen_Stk PImp_Stk PSS_Tub
## 1  2100016      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
## 2  2100016      1    0.63    0.67    0.99     0.57     0.36     1.03    0.85
## 3  2100016      1    0.29    0.50    0.99     0.57     0.36     0.69    0.79
## 4  2100016      1    0.62    0.61    0.99     0.57     0.36     0.75    0.85
## 5  2100016      1    0.50    0.58    0.99     0.45     0.33     0.72    0.85
## 6  2100016      4    0.58    0.45    0.99     0.45     0.33     0.72    0.85
## 7  2100016      1    0.29    0.51    0.99     0.29     0.33     0.72    0.85
## 8  2100024      1    0.66    0.45    1.08     0.57     0.36     0.93    0.85
```

```
## 9  2100024      4    0.66    0.59    1.08     0.57     0.36     0.93    0.85
## 10 2100024      1    0.66    0.67    1.09     0.57     0.36     0.93    0.85
##    PPk_Tub PFl_Tub PHse_Tub price     type income
## 1     1.09    1.19     0.33  0.66  PPk_Stk   32.5
## 2     1.09    1.19     0.37  0.63  PPk_Stk   32.5
## 3     1.09    1.19     0.59  0.29  PPk_Stk   32.5
## 4     1.09    1.19     0.59  0.62  PPk_Stk   32.5
## 5     1.07    1.19     0.59  0.50  PPk_Stk   32.5
## 6     1.07    1.19     0.59  0.45 PHse_Stk   32.5
## 7     1.07    1.19     0.59  0.29  PPk_Stk   32.5
## 8     1.09    1.19     0.33  0.66  PPk_Stk   17.5
## 9     1.09    1.34     0.33  0.57 PHse_Stk   17.5
## 10    1.09    1.19     0.33  0.66  PPk_Stk   17.5
```

Mixed logit model

```r
# Log likelihood function for the data "cr"
likelihood4 = function(par){
  cr$constant = 0
  cr$constant[cr$type=='PPk_Stk'] = par[1]
  cr$constant[cr$type=='PBB_Stk'] = par[2]
  cr$constant[cr$type=='PFl_Stk'] = par[3]
  cr$constant[cr$type=='PHse_Stk'] = par[4]
  cr$constant[cr$type=='PGen_Stk'] = par[5]
  cr$constant[cr$type=='PSS_Tub'] = par[6]
  cr$constant[cr$type=='PPk_Tub'] = par[7]
  cr$constant[cr$type=='PFl_Tub'] = par[8]
  pr = exp(cr$constant+par[9]*cr$price+par[10]*cr$income)/
    (exp(par[1]+par[9]*cr$PPk_Stk+par[10]*cr$income)
    +exp(par[2]+par[9]*cr$PBB_Stk+par[10]*cr$income)
    +exp(par[3]+par[9]*cr$PFl_Stk+par[10]*cr$income)
    +exp(par[4]+par[9]*cr$PHse_Stk+par[10]*cr$income)
    +exp(par[5]+par[9]*cr$PGen_Stk+par[10]*cr$income)
    +exp(par[6]+par[9]*cr$PSS_Tub+par[10]*cr$income)
    +exp(par[7]+par[9]*cr$PPk_Tub+par[10]*cr$income)
    +exp(par[8]+par[9]*cr$PFl_Tub+par[10]*cr$income)
    +exp(par[9]*cr$PHse_Tub+par[10]*cr$income))
  return(-sum(log(pr)))
}

# Maximum likelihood estimation
est4 = optim(runif(10),likelihood4,method='BFGS')
```

The estimated coefficients on price and income are -6.7207599 and 0.3297127, respectively.

Compute the test statistic:

```r
MTT = -2*(-likelihood4(est3$par[-6])+likelihood4(est4$par))
```

$MTT = 0.147831$

Conclude on IIA

```
p = pchisq(MTT,df=10,lower.tail=FALSE)
```

$P(X > MTT|X \sim \chi^2(10)) = 1$, so IIA is not violated.