Assignment 2: Multinomial Choices

Yongyi Lin

3/24/2021

```
setwd("/Users/yongyilin/Econ613/Assignments/A2")
data(margarine)
```

Exercise 1 Data Description

```
choiceprice <- margarine$choicePrice
demos <- margarine$demos
### Average and dispersion in product characteristics
prod_avg <- as.matrix(apply(choiceprice[,2:12], 2, mean))
prod_avg
```

```
##              [,1]
## choice   3.2429530
## PPk_Stk  0.5184362
## PBB_Stk  0.5432103
## PFl_Stk  1.0150201
## PHse_Stk 0.4371477
## PGen_Stk 0.3452819
## PImp_Stk 0.7807785
## PSS_Tub  0.8250895
## PPk_Tub  1.0774094
## PFl_Tub  1.1893758
## PHse_Tub 0.5686734
```

```
prod_disp <- as.matrix(apply(choiceprice[,2:12], 2, sd))
prod_disp
```

```
##               [,1]
## choice   2.58721892
## PPk_Stk  0.15051740
## PBB_Stk  0.12033186
## PFl_Stk  0.04289519
## PHse_Stk 0.11883123
## PGen_Stk 0.03516605
## PImp_Stk 0.11464607
## PSS_Tub  0.06121159
## PPk_Tub  0.02972613
## PFl_Tub  0.01405451
## PHse_Tub 0.07245500
```

```
### Market share and market share by product characteristics
prod_share = as.matrix(summary(as.factor(choiceprice[,2])))/nrow(choiceprice)
prod_share
```

```
##          [,1]
## 1  0.39507830
## 2  0.15637584
## 3  0.05436242
## 4  0.13266219
## 5  0.07046980
## 6  0.01655481
## 7  0.07136465
## 8  0.04541387
## 9  0.05033557
## 10 0.00738255
```

```r
mar <- merge(x = choiceprice, y = demos, by = "hhid", all.x = TRUE)
### Mapping between income and choice
table(mar[,c(2,13)])
```

```
##       Income
## choice 2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5  55 67.5 87.5 130
##     1   19 117  196  318  292  195  209  132  125   83  47   19    9   5
##     2    4  54  106  100  123   94   84   34   33   22  30    4   10   1
##     3    0  13   41   27   34    9   28   17   33   23  11    1    3   3
##     4    2  34   44  111  154   67   64   29   23   16  32    8    1   8
##     5    6  19   23   21  123   18   54   23    6    7   7    6    0   2
##     6    0   2    9    5    2    6    4    1   20   17   3    2    1   2
##     7   16  27   40   54   41   24   49   15   27    6  12    7    1   0
##     8    1   6    8   19   36   25   19   14   21    9  42    3    0   0
##     9    2  22   25   20   30   34   33    9   14    2  17    0   12   5
##     10   0   1    3    2    8    4    5    5    1    3   0    1    0   0
```

```r
### Mapping between family size and choice
table(mar[,c(2,14)])
```

```
##       Fs3_4
## choice   0   1
##     1  864 902
##     2  339 360
##     3  181  62
##     4  295 298
##     5  128 187
##     6   56  18
##     7  162 157
##     8   81 122
##     9  157  68
##     10  21  12
```

```r
table(mar[,c(2,15)])
```

```
##       Fs5.
## choice    0    1
##     1  1524  242
##     2   621   78
##     3   223   20
##     4   475  118
##     5   252   63
##     6    51   23
##     7   299   20
```

```
##     8   192   11
##     9   214   11
##     10   15   18
```

```r
table(mar[,c(2,16)])
```

```
##       Fam_Size
## choice   1   2   3   4   5   6   7   8
##     1  148 474 400 502 160  76   1   5
##     2   49 212 165 195  53  22   1   2
##     3   38 123  29  33  20   0   0   0
##     4   23 154 119 179  72  33   8   5
##     5   10  55  60 127  33  24   2   4
##     6    7  26  11   7  23   0   0   0
##     7   25 117  77  80   8  12   0   0
##     8   18  52  46  76   2   9   0   0
##     9   34 112  48  20  11   0   0   0
##     10   0   3   3   9  13   5   0   0
```

```r
### Mapping between education status and choice
table(mar[,c(2,17)])
```

```
##       college
## choice    0    1
##     1  1205  561
##     2   480  219
##     3   133  110
##     4   419  174
##     5   229   86
##     6    42   32
##     7   216  103
##     8   151   52
##     9   163   62
##     10   18   15
```

```r
### Mapping between job status and choice
table(mar[,c(2,18)])
```

```
##       whtcollar
## choice    0    1
##     1   759 1007
##     2   319  380
##     3   111  132
##     4   242  351
##     5    90  225
##     6    32   42
##     7   135  184
##     8    87  116
##     9    95  130
##     10    2   31
```

```r
### Mapping between retirement status and choice
table(mar[,c(2,19)])
```

```
##       retired
## choice    0    1
##     1  1414  352
```

```
##     2   531  168
##     3   114  129
##     4   502   91
##     5   269   46
##     6    46   28
##     7   272   47
##     8   183   20
##     9   144   81
##     10   29    4
```

```
choice <- 1:10
names(choice) <- 1:10
y <- as.matrix(map_df(choice, function(x) as.integer(choiceprice$choice == x)))
```

Exercise 2 First Model

Conditional Logit Model

```
# Exercise 2 First Model
# Conditional logit model
x_1 <- mar[,3:12]
conl_p <- function(x,b) {
  pn <- exp(matrix(rep(c(0,b[1:9]),nrow(x)),byrow = TRUE,nrow(x))+x*b[10])
  pd <- apply(pn,1,sum)
  return(pn/pd)
}
conl_ll <- function(y,x,b) {
  l <- -sum(y*log(conl_p(x,b)))
  return(l)
}
conl <- optim(function(b) conl_ll(y=y,x=x_1,b=b),par=rep(0,10),method="BFGS")
conl$par
```

```
##  [1] -0.9543259  1.2969965 -1.7173298 -2.9040264 -1.5153021  0.2517927
##  [7]  1.4648942  2.3575437 -3.8966267 -6.6566340
```

The 10th parameter refers to the effect of price. The negative coefficient (-6.6566340) indicates that individual would be less likely to purchase the product as the price increases.

Exercise 3 Second Model

Multinomial Logit Model

```
# Exercise 3 Second Model
# Multinomial logit model
x_2 <- as.matrix(mar[,13],ncol=1)
colnames(x_2)[1] <- "income"
multil_p <- function(x,b) {
  pn <- exp(matrix(rep(c(0,b[1:9]),nrow(x)),byrow=TRUE,nrow(x))
                   +t(apply(x,1,function(x)x*c(0,b[10:18]))))
  pd <- matrix(apply(pn,1,sum)) %*% t(rep(1,10))
  return(pn/pd)
}
multil_ll <- function(y,x,b) {
  l <- -sum(y*log(multil_p(x,b)))
```

```r
  return(l)
}
multil <- optim(function(b) multil_ll(y=y,x=x_2,b=b),par=rep(0,18),method="BFGS")
multil$par
```

```
##  [1] -0.843545649 -2.397656003 -1.199428121 -1.688616844 -4.137055731
##  [6] -1.529169108 -2.846055103 -2.573291074 -4.279712750 -0.003156338
## [11]  0.014507166  0.003980338 -0.001328126  0.030527384 -0.007002723
## [16]  0.022807121  0.017661767  0.010698254
```

The last 9 parameters are the income effects of products 2~10. Thus, an individual would be more likely to purchase product 3, 4, 6, 8, 9, and 10, and less likely to purchase product 2, 5, and 7.

Exercise 4 Marginal Effects

Conditional Logit Model

```r
# Marginal effect for the first model
p_conl <- conl_p(x_1,conl$par)
ind <- array(0, dim=c(nrow(x_1),ncol(x_1),ncol(x_1)))
for (i in 1:nrow(x_1)) {
  diag(ind[i,,]) <- 1
}
me_conl <- array(0, dim=c(nrow(x_1),ncol(x_1),ncol(x_1)))
for (i in 1:nrow(x_1)) {
  for (j in 1:ncol(x_1)) {
    for (k in 1:ncol(x_1)) {
      me_conl[i,j,k] <- p_conl[i,j]*(ind[i,j,k]-p_conl[i,k])*conl$par[10]
    }
  }
}
apply(me_conl,c(2,3),mean)
```

```
##              [,1]         [,2]         [,3]         [,4]         [,5]
##  [1,] -1.28527511  0.295367502  0.120712274  0.295088219  0.156226788
##  [2,]  0.29536750 -0.745423961  0.055079343  0.133452998  0.072823346
##  [3,]  0.12071227  0.055079343 -0.337455325  0.050544822  0.030281111
##  [4,]  0.29508822  0.133452998  0.050544822 -0.712674221  0.064016299
##  [5,]  0.15622679  0.072823346  0.030281111  0.064016299 -0.428081154
##  [6,]  0.03732070  0.016725736  0.007104706  0.016551237  0.008748617
##  [7,]  0.15359916  0.069271384  0.029269291  0.063745696  0.037948441
##  [8,]  0.09929526  0.045206138  0.019664755  0.039262524  0.025089941
##  [9,]  0.11082216  0.050699567  0.021754646  0.044154905  0.028519973
## [10,]  0.01684304  0.006797947  0.003044377  0.005857523  0.004426639
##               [,6]         [,7]         [,8]         [,9]         [,10]
##  [1,]  0.0373206965  0.153599163  0.099295264  0.110822159  0.0168430432
##  [2,]  0.0167257356  0.069271384  0.045206138  0.050699567  0.0067979474
##  [3,]  0.0071047062  0.029269291  0.019664755  0.021754646  0.0030443769
##  [4,]  0.0165512365  0.063745696  0.039262524  0.044154905  0.0058575226
##  [5,]  0.0087486168  0.037948441  0.025089941  0.028519973  0.0044266387
##  [6,] -0.1073228959  0.008537950  0.005430207  0.006113638  0.0007901087
##  [7,]  0.0085379501 -0.420301907  0.025793628  0.027922401  0.0042139545
##  [8,]  0.0054302072  0.025793628 -0.282465600  0.019789652  0.0029334908
```

```
##  [9,]  0.0061136383  0.027922401  0.019789652 -0.313059015  0.0032820730
## [10,]  0.0007901087  0.004213955  0.002933491  0.003282073 -0.0481891557
```

As one can see from the matrix, only diagonal elements are negative whereas all the other elements are positive. This indicates that an individual would turn to other products, if the price of a given product increased, which is very intuitive and makes perfect sense.

Multinomial Logit Model

```
# Marginal effect for the second model
p_multil <- multil_p(x_2,multil$par)
b_multil <- c(0,multil$par[10:18])
me_multil <- array(0,dim=c(nrow(x_2),10))
for (i in 1:nrow(x_2)) {
  b_bar <- sum(p_multil[i,]*b_multil)
  for (j in 1:10) {
    me_multil[i,j] <- p_multil[i,j]*(b_multil[j]-b_bar)
  }
}
for (i in 1:nrow(x_2)) {
  b_bar <- sum(p_multil[i,]*b_multil)
  me_multil[i,] <- p_multil[i,]*(b_multil-b_bar)
}
apply(me_multil,2,mean)
```

```
##  [1] -0.0010504137 -0.0009016311  0.0006266867  0.0001660472 -0.0002794477
##  [6]  0.0004431356 -0.0006821378  0.0008861440  0.0007338590  0.0000577577
```

In this model, an individual would purchase product 1, 2, 5, and 7 more if his or her income increased.

Exercise 5 IIA

```
mix_ll <- function(y,x,b,p_mix) {
  return(-sum(y*log(p_mix(x,b))))
}
# Mixed logit with all choices
X_1 <- as.matrix(mar[,3:13])
p_mix_1 <- function(x,b) {
  pn <- exp(
    matrix(rep(c(0,b[1:9]),nrow(x)),
           byrow=TRUE,
           nrow(x))
    +x[,1:10]*b[10]
    +t(apply(matrix(x[,11],ncol=1),1,function(x) x*c(0,b[11:19])))
  )
  pd <- apply(pn,1,sum)
  return(pn/pd)
}
mixl_1 <- optim(function(b) mix_ll(y=y,x=X_1,b=b,p_mix=p_mix_1),par=rep(0,19),method="BFGS")
mixl_1$par
```

```
##  [1] -0.838705945  0.891148169 -1.826370582 -2.871247434 -2.454001559
##  [6]  0.498968897  0.805453868  1.866785193 -4.140083624 -6.659699884
```

```
## [11] -0.004333800  0.014258958  0.004025557 -0.001264787  0.029710007
## [16] -0.009327126  0.021914644  0.016902350  0.008674428
```

```
# Alternative specification: remove the fifth choice
X_2 <- X_1[,-5]
p_mix_2 <- function(x,b) {
  pn <- exp(
    matrix(rep(c(0,b[1:8]),nrow(x)),byrow=TRUE,nrow(x))
    +x[,1:9]*b[9]
    +t(apply(matrix(x[,10],ncol=1),1,function(x) x*c(0,b[10:17])))
  )
  pd <- apply(pn,1,sum)
  return(pn/pd)
}
mixl_2 <- optim(function(b) mix_ll(y=y[,-5],x=X_2,b=b,p_mix=p_mix_2),par=rep(0,17),method="BFGS")
mixl_2$par
```

```
##  [1] -0.833351265  0.835949948 -1.817194069 -2.479963339  0.470511500
##  [6]  0.759399878  1.796558890 -4.124827885 -6.526243552 -0.004466964
## [11]  0.014151365  0.003920641  0.029651615 -0.009542146  0.021621567
## [16]  0.016716707  0.009048957
```

```
# Compute test statistics
L_all <- mix_ll(y=y,x=X_1,b=mixl_1$par,p_mix=p_mix_1)
L_alter <- mix_ll(y=y[,-2],x=X_2,b=mixl_2$par,p_mix=p_mix_2)
MTT <- 2*(L_all-L_alter)
csq95 <- qchisq(.95, length(mixl_2$par))
MTT > csq95
```

```
## [1] TRUE
```

MTT > csq95. IIA is violated.