Assignment 3

Xiaoshu Gui

Setup

```
library(dplyr)
library(purrr)
library(magrittr)
library(tidyr)
library(tibble)
library(stringr)
library(reshape2)
library(lme4)
library(mfx)
library(bayesm)
# load data
data(margarine)
# Create a dataframe that merges product characteristics with household demos by hhid.
choiceprice <- as.matrix(margarine$choicePrice)
demos <- as.matrix(margarine$demos)
marg <- merge(choiceprice, demos, by = "hhid")
```

Exercise 1 Data Description

- Average and dispersion in product characteristics.

```
# average:
apply(marg[, 3:12], 2, mean)
```

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

```
# dispersion
apply(marg[, 3:12], 2, sd)
```

```
##     PPk_Stk     PBB_Stk     PFl_Stk    PHse_Stk    PGen_Stk    PImp_Stk
## 0.15051740 0.12033186 0.04289519 0.11883123 0.03516605 0.11464607
##     PSS_Tub     PPk_Tub     PFl_Tub    PHse_Tub
## 0.06121159 0.02972613 0.01405451 0.07245500
```

- Market share, and market share by product characteristics.

```
# market share by product
ms_product <- table(marg$choice)/4470
names(ms_product) <- names(marg[,3:12])
print(ms_product)
```

```
##     PPk_Stk     PBB_Stk     PFl_Stk    PHse_Stk    PGen_Stk    PImp_Stk
## 0.39507830 0.15637584 0.05436242 0.13266219 0.07046980 0.01655481
##     PSS_Tub     PPk_Tub     PFl_Tub    PHse_Tub
## 0.07136465 0.04541387 0.05033557 0.00738255
```

It shows that the first, second and fourth product take the largest market share.

```
# market share by product characteristics: brand and type
brand_name <- names(marg[,3:12]) %>%
  str_replace_all("_Stk|_Tub", "")
```

```r
ms_brand <- cbind.data.frame(brand_name, ms_product) %>%
  group_by(brand_name) %>%
  summarise(market_share = sum(Freq)) %>%
  arrange(desc(market_share))
print(ms_brand)
```

```
## # A tibble: 7 x 2
##   brand_name market_share
##   <fct>             <dbl>
## 1 PPk              0.440
## 2 PBB              0.156
## 3 PHse             0.140
## 4 PFl              0.105
## 5 PSS              0.0714
## 6 PGen             0.0705
## 7 PImp             0.0166
```

```r
# by product type (stick and tub)
sum(ms_product[1:6]) # market share of stick
```

```
## [1] 0.8255034
```

```r
sum(ms_product[7:10]) # market share of tub
```

```
## [1] 0.1744966
```

- Mapping between observed attributes and choices. Create tables of choices by different household attributes:

```r
# income level & choices
t1 <- table(marg$Income, marg$choice) %>% print()
```

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

```r
# family size:
table(marg$Fam_Size, marg$choice)
```

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

```r
# education status & choices
table(marg$college, marg$choice)
```

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

```r
# job status & choicies
table(marg$whtcollar, marg$choice)
```

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

```r
# retirement status & choices
table(marg$retired, marg$choice)
```

```
##
##        1    2    3    4    5    6    7    8    9   10
##    0 1414  531  114  502  269   46  272  183  144   29
##    1  352  168  129   91   46   28   47   20   81    4
```

Recap Multinomial Models

There are m alternatives and the dependent variable y is defined to take value j if the jth alternative is taken, j = 1, . . . , m. Based on the random utility model, let U_{ij} denote the utility of individual i derive when choosing altertive *j*. j is chosen if and only if $U_{ij} > U_{ik}$ for all $k \neq j$. Although we can't observe U_{ij}, we can treat it as independent random variables with a systematic component V_{ij} and a random component ϵ_{ij} such that $U_{ij} = V_{ij} + \epsilon_{ij}$.

Define the probability that alternative j is chosen by individual i as:

$$P_{ij} = Pr[y_i = j] = \frac{V_{ij}}{\sum_{k=1}^{m} V_{ik}}, \quad j = 1, ..., m,$$

where $V_{ij} > 0$ can be general functions of regressors X_i and parameters β. This is a *universal logit model*. Different specifications for V_{ij} corresponds to specific models, such as multinomial logit and conditional logit models. In that sense, all these models are variants of the same model. They only differ in their parametrization of the systematic compotents V_{ij}. The log likelihood function of the universal logit model is (assuming independent realizations by summing up all N individual contributions):

$$L = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \; ln \; P_{ij}, \qquad j = 1, ..., m, \quad i = 1, ..., N,$$

where P_{ij} is defined above.

Exercise 2 First Model

- We are interested in the effect of price on demand. Propose a model specification.

Since the price of a product varies by different choices, a conditional logit model is chosen. Here $V_{ij} = X_{ij}$, specifying characteristics of the alternatives (price). The probability of the ith househould choosing product j is given by

$$P_{ij} = Pr[y_i = j] = \frac{\exp(X_{ij}\beta)}{\sum_{k=1}^{m} exp(X_{ik}\beta)}, \quad j = 1, ...m$$

where X denotes price, the substrcipt i denotes the ith household, subscript j or k denotes the alternative, and parameter β is contant across alternatives. Note that it is possible to go from alternative-varying regressors to alternative-invariant format. Let X_i be a $K * 1$ vector. Define X_{ij} to be a $Km * 1$ vector with zeros except that the jth block is X_i, that is $X_{ij} = [0'...0', X_i, 0', ...0']'$, and define

$$\beta = [0', \beta_2', ...\beta_m']',$$

where $\beta_1 = 0$ is a normalization. Then $X_i'\beta_j = X_{ij}'\beta$.

- The likelihood function of conditional multinomial model and its optimization

```
loglik_cl <- function(beta) {
  X = marg[, 3:12] # price takes dif values for dif alternatives (1*10)
  X_beta = X * beta # beta is a constant for each decision maker i. (1*10)
  X_beta_j = matrix(nrow = nrow(marg), ncol = 1) # coefficient for a given choice j (N*1)
  for (i in 1:nrow(marg)) {
    jstar = marg[i, "choice"]
    X_beta_j[i] = X_beta[i, jstar]
  }
  numerator = exp(X_beta_j)
  denominator = rowSums(exp(X_beta))
  pij = numerator/denominator
  ll = log(pij)
  loglik_cl <- -sum(ll)
}
# optimize the likelihood using optim()
fit_cl <- optim(0, loglik_cl, method = "BFGS",  hessian = TRUE)
fit_cl$par # using nlm() returns the same result: nlm(loglik_cl, 0)
```

```
## [1] -2.428201
```

Interpretation: Note that the estimated beta < 0, it suggests that an increase in the price of one alternative decreases the probability of choosing that alternative and increases the probability of choosing other alternatives.

Exercise 3 Second Model

- We are interested in the effect of family income on demand. Propose a model specification.

Since family income is a fixed constant for households and does not vary across product choices, a multinomial logit model is chosen to address alternative-invariant regressors. The probability of the ith household choosing product j is given by:

$$P_{ij} = Pr[y_i = j] = \frac{\exp(\alpha_j + X_i\beta_j)}{\sum_{k=1}^{m} exp(\alpha_k + X_i\beta_k)}, \quad j = 1, ...m,$$

where X denotes income. The likelihood function is:

```
loglik_mnl <- function(beta) {
  X = as.matrix(marg[, "Income"]) # income is a N*1 vector
  #beta = matrix(nrow = 1, ncol = 10) # alternative-specific constant (1*m)
  beta[1] = 0  # set beta_1 to 0-- use product 1 as reference group
  X_beta = X %*% beta # returns a matrix of N*m
  X_beta_j = matrix(nrow = nrow(marg), ncol = 1)

  for (i in 1: nrow(marg)) {
    jstar  = marg$choice[i]
    X_beta_j[i] = X_beta[i, jstar]
  }
  numerator = exp(X_beta_j)
  denominator = sum(exp(X_beta))
  pij = numerator / denominator
  ll = log(pij)
  return(-sum(ll))
}
# Use optim funtion to optimize the likelihood:
fit_mnl = optim(c(rep(0, 10)), loglik_mnl, method = "BFGS",  hessian = TRUE)
fit_mnl$par # nlm() does not work here...
```

```
##  [1]  0.00000000 -0.01162303 -0.04225309 -0.01490642 -0.03812994
##  [6] -0.08203845 -0.03927242 -0.04495720 -0.04373569 -0.13914935
```

Interpretation of MNL estimates is relative to the reference group. I set the first product as the base group, so the coefficients for the first product is normalized to zero. Compared to *PKK_stk*, higher income levels lead to reduced likelihood of buying all other products (since all other coefficients are negative), which seems to lack variation in cunsumption choice. Go to *t1* in *Ex1* to look at the relation between customers' choices and their income level. Then check the average price of each product:

```
marg[,3:12] %>% summarise_all(mean)
```

```
##     PPk_Stk   PBB_Stk PFl_Stk  PHse_Stk  PGen_Stk  PImp_Stk   PSS_Tub
## 1 0.5184362 0.5432103 1.01502 0.4371477 0.3452819 0.7807785 0.8250895
##    PPk_Tub  PFl_Tub  PHse_Tub
## 1 1.077409 1.189376 0.5686734
```

Indeed, the first product is the most popular one, it might be true that people make this decison independent of their income. Note that richer families do not always prefer more expensive products– many of them still choose inexpensive products. It suggests that family income might not lead to choices of more expensive products. Now we change a reference group to further test the effct family income on households' choice of margarine. The seventh product is chosen, because it is more expensive than the first one, yet still attracts a lot of customers.

```
loglik_mnl_2 <- function(beta) {
  X = as.matrix(marg[, "Income"])
  beta[7] = 0  # set beta_3 to 0-- use product 3 as reference group
  X_beta = X %*% beta
  X_beta_j = matrix(nrow = nrow(marg), ncol = 1)
  for (i in 1: nrow(marg)) {
    jstar  = marg$choice[i]
    X_beta_j[i] = X_beta[i, jstar]
  }
  numerator = exp(X_beta_j)
  denominator = sum(exp(X_beta))
  pij = numerator / denominator
  ll = log(pij)
  return(-sum(ll))
}
fit_mnl_2 = optim(c(rep(0, 10)), loglik_mnl_2, method = "BFGS",  hessian = TRUE)
fit_mnl_2$par
```

```
##  [1]  0.024163240  0.004520727 -0.021704674  0.001801872 -0.018099470
##  [6] -0.056817568  0.000000000 -0.024083784 -0.023004170 -0.107278541
```

Now we have some variation in consumption choice with regard to income. Compared to the seventh product, higher family income levels lead to greater likelihood of purchasing the first, second and fourth product (since their coefficients are positive), and reduced likelihood of all other products as their coefficients are negative. Changing different reference groups, MNL models show that households are consistently more likely to choose the first, second and fourth product, with an increase of family income. However, the prices of these three product are not very expensive, indicating that family income might not be a good discriminator for consumer packaged goods (CPG) choice such as margarine.

Exercise 4 Marginal Effects

Compute and interpret the marginal effects for the first and second models.

```
# Marginal effect of the conditional logit model
X = marg[, 3:12] # N*m
b = fit_cl$par # 1
X_beta = X * b # N*m
X_beta_j = matrix(nrow = nrow(marg), ncol = 1) # N*1
xbetak =  exp(X_beta)
denominator = rowSums(xbetak)
pr_ij = as.matrix(xbetak/denominator) # N*m
pij = t(pr_ij) %*% pr_ij * (-b) # m*m (10*10)
margin = matrix(rep(colSums(pr_ij) * b, 10), ncol=10 )
```

```r
margin = margin * diag(10)
me_cl = (pij + margin)/nrow(marg)
me_cl
```

```
##               PPk_Stk      PBB_Stk      PFl_Stk    PHse_Stk    PGen_Stk
## PPk_Stk  -0.286734523  0.042634266  0.013429853  0.05597838  0.06819125
## PBB_Stk   0.042634266 -0.271403038  0.012659776  0.05218752  0.06439140
## PFl_Stk   0.013429853  0.012659776 -0.093788131  0.01620342  0.02015687
## PHse_Stk  0.055978376  0.052187521  0.016203419 -0.33356490  0.08189012
## PGen_Stk  0.068191248  0.064391402  0.020156866  0.08189012 -0.39393032
## PImp_Stk  0.024224316  0.022641091  0.007120508  0.02947916  0.03583419
## PSS_Tub   0.021469284  0.020143158  0.006323770  0.02576285  0.03208617
## PPk_Tub   0.011550687  0.010948415  0.003444206  0.01387440  0.01745943
## PFl_Tub   0.008791944  0.008297717  0.002608263  0.01052612  0.01325992
## PHse_Tub  0.040464550  0.037499692  0.011841469  0.04766294  0.06066097
##              PImp_Stk      PSS_Tub      PPk_Tub      PFl_Tub      PHse_Tub
## PPk_Stk   0.024224316  0.021469284  0.011550687  0.008791944  0.040464550
## PBB_Stk   0.022641091  0.020143158  0.010948415  0.008297717  0.037499692
## PFl_Stk   0.007120508  0.006323770  0.003444206  0.002608263  0.011841469
## PHse_Stk  0.029479160  0.025762848  0.013874398  0.010526120  0.047662944
## PGen_Stk  0.035834194  0.032086165  0.017459431  0.013259921  0.060660968
## PImp_Stk -0.162091248  0.011260483  0.006079360  0.004614158  0.020837979
## PSS_Tub   0.011260483 -0.145753596  0.005506807  0.004165078  0.019036002
## PPk_Tub   0.006079360  0.005506807 -0.081560330  0.002282135  0.010414890
## PFl_Tub   0.004614158  0.004165078  0.002282135 -0.062433390  0.007888054
## PHse_Tub  0.020837979  0.019036002  0.010414890  0.007888054 -0.256306546
```

```r
# Marginal effect of the multinomial model
X = as.matrix(marg[, "Income"])
beta = matrix(fit_mnl$par, nrow = 1, byrow = T)
X_beta_j = X %*% beta
numerator = exp(X_beta_j)
pij = t((apply(numerator, 1, function(x) x / sum(x))))
beta_bar = pij %*% t(beta)
beta_hat = matrix(rep(beta_bar, 10), ncol = 10)
beta_j = matrix(rep(t(beta)), nrow(marg), byrow = T, ncol = 10)
me_mnl = colSums(pij * (beta_j - beta_hat))/nrow(marg)
me_mnl
```

```
##  [1]  0.005499779  0.002063292 -0.001211830  0.001416207 -0.001017888
##  [6] -0.001734755 -0.001076289 -0.001315883 -0.001270955 -0.001351677
```

Exercise 5 IIA

Now combine the above two models to estimate the effect of price and family income on choices of margarine. The mixed logit model is specified as:

$$P_{ij} = \frac{exp(X_{ij}\beta \ + W_i\gamma_j)}{\sum_{k=1}^{m} exp(X_{ik}\beta + W_i\gamma_k)}, \quad j = 1, ..., m$$

Its likelihood function and optimization:

```r
loglik_mixed = function(beta) {
  X = marg[, 3:12] - marg[, 3] # set price of the first product as reference
  b = beta[1] # alternative-variant coefficient
  gamma = beta[2:11] # alternative-invariant coefficient (1*m)
  gamma[1] = 0
  X_beta = X * b # N*m (alternative-variance component)
  gamma_choice = matrix(nrow = nrow(marg), ncol = 1) # N*1 (systematic component)
  X_beta_j = matrix(nrow = nrow(marg), ncol = 1)
  gamma_k = matrix(rep(t(gamma), times = nrow(marg)), ncol = ncol(t(gamma)), byrow = T)
```

```r
  for (i in 1: nrow(marg)) {
    jstar  = marg[i, "choice"]
    gamma_j = gamma[jstar]
    gamma_choice[i] = gamma_j
    X_beta_j[i] = X_beta[i, jstar]
  }
  numerator = exp(X_beta_j + gamma_choice)
  Xbeta_k =  exp(X_beta + gamma_k)
  denominator = rowSums(Xbeta_k)
  Pij = numerator / denominator
  ll = log(Pij)
  loglik_mixed = - sum(ll)
}
# optimize the likelihood:
fit_mixed = nlm(f = loglik_mixed, p = c(rep(0, 11))) # nlm() returns the same result but takes longer.
beta_f <- fit_mixed$estimate %>% print()
```

```
##  [1] -6.6565906  0.0000000 -0.9543061  1.2969786 -1.7173341 -2.9040074
##  [7] -1.5153149  0.2517637  1.4648579  2.3575174 -3.8965934
```

Recap: The IIA assumption

The ratio of logit probabilities of any two alternatives j and k is

$$\frac{Pr(y_i = i)}{Pr(y_i = k)} = \frac{exp(V_{ij})}{exp(V_{ik})} = exp(V_{ij} - V_{ik})$$

Note that the above ratio only depends on alternatives j and k. Because the ratio is independent of alternatives other than j and k, MNL logit models are said to be independent of irrelevant alternatives (IIA). IIA implies that presence or absence of another alternative should not alter the relative probabilities of any single decision-maker (conditional on the model's systematic component).

- Consider an alternative specification, where we remove one choice (the last one) from the data.

```r
sub_marg <- marg %>% # create a subset of data which remove the 10th choice.
  filter(choice < 10)
loglik_mixed_2 = function(beta) {
  X = sub_marg[, 3:11] - sub_marg[, 3] # do not include the price of 10th product
  b = beta[1] # alternative-variant coefficient
  gamma = beta[2:10] # alternative-invariant coefficient
  gamma[1] = 0 # set the first product as reference group
  X_beta = X * b
  gamma_choice = matrix(nrow = nrow(sub_marg), ncol = 1)
  X_beta_j = matrix(nrow = nrow(sub_marg), ncol = 1)
  gamma_k = matrix(rep(t(gamma), times = nrow(sub_marg)), ncol = ncol(t(gamma)), byrow = T)
  for (i in 1: nrow(sub_marg)) {
    jstar  = sub_marg[i, "choice"]
    gamma_choice[i] = gamma[jstar]
    X_beta_j[i] = X_beta[i, jstar]
  }
  numerator = exp(X_beta_j + gamma_choice)
  Xbeta_k =  exp(X_beta + gamma_k)
  denominator = rowSums(Xbeta_k)
  Pij = numerator / denominator
  ll = log(Pij)
  loglik_mixed_2 = - sum(ll)
}
# fit_mixed_2 = optim(c(rep(0, 10)), loglik_mixed_2, hessian = TRUE) # optim is too slow!
fit_mixed_2 <- nlm(loglik_mixed_2, c(rep(0, 10)))
beta_r <- fit_mixed_2$estimate %>% print()
```

```
##  [1] -6.6594376  0.0000000 -0.9559718  1.2965597 -1.7177884 -2.9059224
##  [7] -1.5169826  0.2511291  1.4647655  2.3582114
```

By dropping one alternative (further tests can drop more irrelavant alternatives at a time) and reestimating the model, we can see that the coefficients do not change, indicating that IIA might hold. * Compute the test statistics:

$$MTT = -2[L_r(\beta^r) - L_r(\beta^f)] = -2ln\frac{L(\beta^r)}{L(\beta_f)} = 2ln\frac{L(\beta^f)}{L(\beta_r)} = LR,$$

where $L(\beta^r)$is the likelihood evaluated at the MLE and $L(\beta^r)$ is the maximum of likelihood subject to the restriction (that r parameters unconstrained in the full likelihood analysis are assigned fixed values).

```r
# calculate the likelihood evaluated at MLE (beta_f)
X = marg[, 3:12] - marg[, 3]
b = beta_f[1]
gamma = as.matrix(beta_f[2:11]) # 10*1
X_beta = X * b
gamma_k = matrix(rep(t(gamma), times = nrow(marg)), ncol = ncol(t(gamma)), byrow = T) # 4470*10
Xbeta_k =  exp(X_beta + gamma_k) # 10*10
denominator = rowSums(Xbeta_k)
Pij = Xbeta_k / denominator
ll = log(Pij)
loglik_beta_f = sum(ll)
# ML subject to restriction (beta_r)
X = sub_marg[, 3:11] - sub_marg[, 3]
b = beta_r[1]
gamma = beta_r[2:10]
X_beta = X * b
gamma_k = matrix(rep(t(gamma), times = nrow(sub_marg)), ncol = ncol(t(gamma)), byrow = T)
Xbeta_k =  exp(X_beta + gamma_k)
denominator = rowSums(Xbeta_k)
Pij = Xbeta_k  / denominator
ll = log(Pij)
loglik_beta_r = sum(ll)
# compute the test statistics
mtt <- log(loglik_beta_f/loglik_beta_r)*2
mtt
```

```
## [1] 0.4032821
```

```r
# For sufficiently large sample size, the LR test statistic is chisqured distributed
# a chi-sqaure with r degrees of freedom
pchisq(mtt, df = length(beta_r))
```

```
## [1] 2.34917e-06
```

Another way to calculate the Hausman and McFadden Test by hand:

```r
beta_f1 <- beta_f[1:10] # beta_f has one more parameter than beta_r
beta_diff <- beta_r - beta_f1
hm <- beta_diff %*% solve(var(beta_r) - var(beta_f1)) %*% t(beta_diff)
pv <- pchisq(hm,df = 2*10) # the degrees of freedom of the Chi-Square distribution used to test the
# LR Chi-Square statistic is defined by the number of models estimated (2) times the number of
# predictors in the model (10).
```

The result of statistical test suggests that IIA holds.