Assignment 4

Ko-Cheng Chang

January 2021

1 Introduction

In assignment 4, we try to estimate optimal stock investment and Medical age assessment.

2 Q1(a)

2.1 Problem

Here we will use collected data to estimate more sample. Which is V_q

2.2 Theory and implementation

Here I input the dataset, and transform into log-form. After that calculate different log-matrix between n and $n-1$ day, where $n = 2, 3, 4..., 100$. With different log-matrix, I have mean and covariance-matrix for 7 stocks. With such information, I estimate sample V_q, with sample size $S = 1000$.

2.3 Results and discussion

Result is V_q

3 Q1(b)

3.1 Problem

Here we try to calculate expectation outcome with two possibilities. 1. Each stock receives equal investment. 2. The stock with the best expected performance receives all the investment.

3.2 Theory and implementation

I first trans V_q in log, define function capture expectation of stock with variable k. Question ask to use two strategy, invest equally or invest best expected performance. In Q1(b), I calculate mean different in log, shows that S4 is the best expected performance stock. Where expected performance, $\mathrm{E}[u(T)]$ is given by:

$$\mathrm{E}[u(T)] \approx \frac{1}{S} \sum_{q=1}^{S} \frac{1}{k} (1 - (\sum_{i=1}^{m} w_i \exp(V_q i))^{-k})$$

3.3 Results and discussion

In invest equally:

$$k = -0.5, \mathrm{E}[u(T)] = 0.05282092$$
$$k = 0.5, \mathrm{E}[u(T)] = 0.04199535$$
$$k = 1.5, \mathrm{E}[u(T)] = 0.03191279$$

In S4 only:

$$k = -0.5, \mathrm{E}[u(T)] = 0.0686841$$
$$k = 0.5, \mathrm{E}[u(T)] = 0.0395914$$
$$k = 1.5, \mathrm{E}[u(T)] = 0.01209601$$

Where I find out that, although, S4 only strategy have higher expected outcome, it suffer more drop than invest equally when k increasing

4 Q1(c)

4.1 Problem

In question, we want to find best w and $1-w$, which return best outcome in $\mathrm{E}[u(T)]$ when limit to invest only S3 and S4 stock.

4.2 Theory and implementation

I define function only take S3 and S4 into consider, and find optimal outcome when $k = -0.5$ and $k = 1.5$

4.3 Results and discussion

When $k = -0.5$, best $w = 0.6537733$

$$\mathrm{E}[u(T)] = 0.08534997$$

When $k = 1.5$, best $w = 0.2957975$

$$\mathrm{E}[u(T)] = 0.02917283$$

Where we find that, when k increasing, $\mathrm{E}[u(T)]$ decreasing. Comparing with equally investment strategy in Q1(b), we obtain huge better $\mathrm{E}[u(T)]$ when $k = -0.5$ without loss many when $k = 1.5$

5 Q1(d)

5.1 Problem

Discuss problems and weaknesses with using the approach we used to reach a decision about how to invest.

5.2 Results and discussion

6 Q2(a)

6.1 Problem

Find the MLE estimator a and b by given data. After that, compute logit regression with a and b and plot the result.

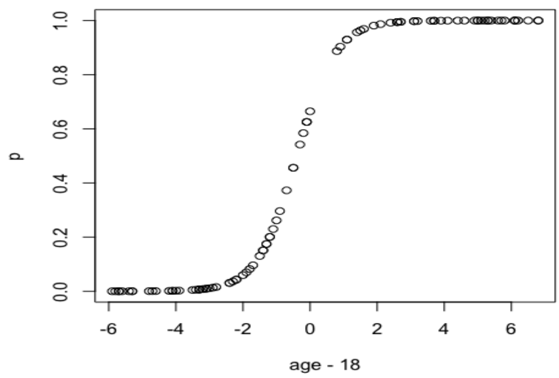

Figure 1: Plot showing possibility of having mature knee with given x, which is $age - 18$.

6.2 Theory and implementation

First of all, I combine two dataset and assign mature with variable $y = 1$, immature with $y = 0$. Second, I create function to compute MLE and optimal a and b. Where MLE function is given by :

$$\mathrm{L}(a,b) = \prod_{i=1}^{n}(\frac{\exp(a+bx_i)}{1-\exp(a+bx_i)})^{y_i}(1-\frac{\exp(a+bx_i)}{1-\exp(a+bx_i)})^{1-y_i}$$

6.3 Results and discussion

Where we find that with $a = 0.6848532$ and $b = 1.7217080$, provide max $\mathrm{L}(a,b) = 1.614533e-07$ Plot is shown in Figure 1. The result is reasonable, from plot we know, p is range between 0 and 1, with higher/lower age the chance having mature knee is higher/lower.

7 Q2(b)

7.1 Problem

Make a 21x21 grid plot to illustrate relation of $a \in (-0.5, 2)$,$b \in (0.5, 3)$ and posterior, where posterior is $\mathrm{L}(a,b)$

7.2 Theory and implementation

Here I create vector of a and b with 21 length, put $\mathrm{L}(a,b)$ into matrix form, and use **image** function to plot a,b and $\mathrm{L}(a,b)$

7.3 Results and discussion

In Figure 2, we find out that value of $\mathrm{L}(a,b)$ is concentrate in the range of given a and b, which provide same result in Q2(a), when $a = 0.6848532$ and $b = 1.7217080$ exist max $\mathrm{L}(a,b)$ is $1.614533e-07$

8 Q2(c)

8.1 Problem

Compare outcome of different cost function c_1 and c_2 with (i),(ii) and (iii) situation.

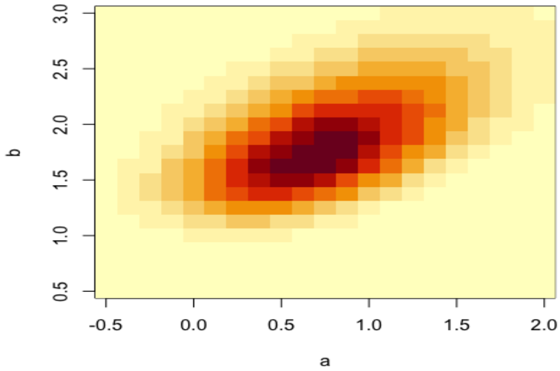

Figure 2: The L(a, b) over a and b in the logistic regression

8.2 Theory and implementation

Cost functions, c_1 is that cost of classify a child as an adult is B times the cost to classify an adult as a child, c_2 is another form of cost function, the cost is based on the difference between the actual age and 18.

Here I define C_a and C_c function with c_1 and c_2, after that, make outcome list to store outcome as $y = 1$ if $C_a > C_c$ and $y = 0$ if $C_a <= C_c$

8.3 Results and discussion

The result is in outcome_i,outcome_ii and outcome_iii, each represent different situation set up. From outcomes, we know that with c_1 cost function, no any situation that "mature knee" should result in a classification as an adult. In other hand, with c_2 cost function, all situation imply that "mature knee" should result in a classification as an adult. The different might result in different assumed in cost function.

9 Q2(d)

9.1 Problem

Try to combine result in (b) and (c), to generate the average result in uncertainly a and b.

9.2 Theory and implementation

9.3 Results and discussion

Appendix - code

```
setwd("~/        /Stochastic data processing and simulation")

install.packages("LearnBayes")
install.packages("dplyr")
library("LearnBayes")
library("dplyr")

set.seed(321)
# Q1(a)
stock <- as.matrix(read.csv("stockvalues.txt"))

log_stock <- log(stock) # gen dif-log matrix
df_stock <- matrix(0, nrow = 1005, ncol = 7)
for (m in 1:7) {
  for (i in 1:1005){
    df_stock[i,m] <- log_stock[i+1,m] - log_stock[i,m]
  }
}
cov_sto <- cov(df_stock) #cov-martix
mean_sto <- colMeans(df_stock) # mean vector

V_q <- rmnorm(1000,100*mean_sto,100*cov_sto) # gen Vq

#Q1(b)
exp_Vq <- exp(V_q)
# stock recived equal investment
fun_1 <- function(k){
  mean((1/k)*(1 - (rowSums(1/7 * exp_Vq))^-k))
}

fun_1(-0.5)
# k = -0.5 ans 0.05282092
fun_1(0.5)
# k = 0.5 ans 0.04199535
fun_1(1.5)
# k = 1.5 ans 0.03191279

# all invest into best stock, which is forth stock
fun_2 <- function(k){
  mean((1/k)*(1 - exp_Vq[,4]^-k))
}
fun_2(-0.5)
# k = -0.5 ans 0.0686841
fun_2(0.5)
# k = 0.5 ans 0.0395914
fun_2(1.5)
# k = 1.5 ans 0.01209601

#Q1(c)
#gen function
fun_3 <- function(w){
 -1 * mean((1/k)*(1 - (w * exp_Vq[,3] + (1-w)*exp_Vq[,4])^-k))
}

# k = -0.5
k <- -0.5
```

```
optimize(fun_3, c(0,1), lower = 0, upper = 1)
# ans 0.6537733 with max 0.08534997

# k= 1.5
k <- 1.5
optimize(fun_3, c(0,1), lower = 0, upper = 1)
# ans 0.2957975 with max 0.02917283

#Q1(d)
# write in pdf

#Q2(a)
ma_kn <- read.table("matureKnee.txt") # read data
im_kn <- read.table("immatureKnee.txt")

ma_kn$new_col <- 1 # adding y
im_kn$new_col <- 0

knee <- rbind(ma_kn,im_kn)
knee <- rename(knee, "x" = "V1" , "y" = "new_col")

x <- knee$x - 18 # age and mature-knee
y <- knee$y

# MLE function
fun_4 <- function(a,b){
  p <- exp(a + b*x )/(1 + exp(a + b*x))
  out_1 <- prod(p^y * (1-p)^(1-y))
  return(out_1)
  }

# transform for max
fun_5 <- function(x) {
  a <- x[1]
  b <- x[2]
  -fun_4(a, b)
}
optim(c(2,2),fun_5) # find max number, max L is -1.614533e-07

#  input a and b
a <- 0.6848532
b <- 1.7217080
p <- exp(a + b*x)/(1 + exp(a + b*x)) # p value

plot(x,p, xlab = "age - 18") # plot of p and x

#Q2(b)
a <- seq(-0.5,2, length.out = 21) # gen interval of a and b
b <- seq(0.5,3, length.out = 21)

out_2 <- matrix(0,21,21) # gen matrix for outcome

# gen matrix to store outcome
for (i in 1:21){
  for (m in 1:21){
    out_2[i,m] <- outer(a[i],b[m], FUN = "fun_4")
  }
}
```

```
image(a,b,out_2)# plot

#Q2(c)
a <- 0.6848532
b <- 1.7217080
# for simplization , B = 10
B <- 10

# gen function , we assumed that age at least 14
cost_1 <- function(age){
  pi <- dgamma(age-14, shape = shape, rate = rate) # distribution of age
  c_1 <- ifelse(age <= 18, B , 1) # cost_1 function
  fx <- exp(a + b*(age-18))/(1+exp(a + b* (age-18))) # p-value
  out <- pi * fx * c_1
  return(out)
}

cost_2 <- function(age){
  pi <- dgamma(age-14, shape = shape, rate = rate) # distribution of age
  c_2 <- ifelse(age <= 18, B*(18-age) , age - 18) # cost_2 function
  fx <- exp(a + b*(age-18))/(1+exp(a + b* (age-18))) # p-value
  out <- pi * fx * c_2
  return(out)
}

# in i
shape <- 3 # set up alpha and mu
mu <- 18.5
rate <- shape/(mu-14)
outcome_i <- c(0,0)
outcome_i[1] <- ifelse (integrate(cost_1,18,100)$value > integrate(cost_1
    ,0,18)$value, 1,0)
outcome_i[2] <- ifelse (integrate(cost_2,18,100)$value > integrate(cost_2
    ,0,18)$value, 1,0)
# with c_1, c_a > c_c
# with c_2, c_a < c_c


# in ii
shape <- 6 # set up alpha and mu
mu <- 19.5
rate <- shape/(mu-14)
outcome_ii <- c(0,0)
outcome_ii[1] <- ifelse (integrate(cost_1,18,100)$value > integrate(cost_1
    ,0,18)$value, 1,0)
outcome_ii[2] <- ifelse (integrate(cost_2,18,100)$value > integrate(cost_2
    ,0,18)$value, 1,0)
# with c_1, c_a > c_c
# with c_2, c_a < c_c


# in iii
shape <- 3 # set up alpha and mu
mu <- 20.58
rate <- shape/(mu-14)
outcome_iii <- c(0,0)
```

```r
outcome_iii[1] <- ifelse (integrate(cost_1,18,100)$value > integrate(cost_1
    ,0,18)$value, 1,0)
outcome_iii[2] <- ifelse (integrate(cost_2,18,100)$value > integrate(cost_2
    ,0,18)$value, 1,0)
# with c_1, c_a > c_c
# with c_2, c_a < c_c

#Q2(d)
# I don't where to start :(((
```