HW2

2023-03-26

1.4 Simulation of MLE

2. Draw N = 2, use optim to get estimates.

```
library ( stats )
library ( tidyverse )
```

```
## Warning: 套件 'tidyverse' 是用 R 版本 4.2.3 來建造的
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.0     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.1     ✓ tibble    3.1.8
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to
become errors
```

```
set.seed(1)
q14 <- function ( param ) {
sig1 <-param [1]
sig2 <-param [2]
ep1 = rnorm (2 ,0 , sig1 ^2) #N=2
ep2 = rnorm (2 ,0 , sig2 ^2)
y = ep1 + ep2

log( sig1 ^2+ sig2 ^2) +log (2*pi ) +(1 /(2*( sig1 ^2+ sig2 ^2) ) )*sum ( y ^2)
# since ‘optim ‘ minimizes the function , we have to minimize the negative of the log -likeli
hood to get the maximization .
}

param <- c(2 ,1) # initial values
q14.result <- optim ( param , q14 , method ="L-BFGS-B", hessian = TRUE )
q14.result $par #  2.029604 0.668017
```

```
## [1] 2.029604 0.668017
```

3.2 Propensity Score

10. Add control variables. Can you recover the parameters?

```
library(MASS)
```

```
##
## 載入套件：'MASS'
```

```
## 下列物件被遮斷自 'package:dplyr':
##
##     select
```

```
library(data.table)
```

```
## Warning: 套件 'data.table' 是用 R 版本 4.2.3 來建造的
```

```
##
## 載入套件：'data.table'
```

```
## 下列物件被遮斷自 'package:lubridate':
##
##     hour, isoweek, mday, minute, month, quarter, second, wday, week,
##     yday, year
```

```
## 下列物件被遮斷自 'package:dplyr':
##
##     between, first, last
```

```
## 下列物件被遮斷自 'package:purrr':
##
##     transpose
```

```
#set parameters
mu0 =13
mu1 =15
sigma0 =1
sigma1 =1.5
sigma01 =1.1
beta1 =0.5
beta2 =0.005

sigma.nu <-sqrt ( sigma0 ^2+ sigma1 ^2 -2* sigma01 )
rho <- sigma01 /( sigma0 * sigma1 )

# generate multivariate normal
sigma <-rbind (c( sigma0 ^2 , sigma01 ) ,c( sigma01 , sigma1 ^2) )
mu <-c(0 ,0)
q4df <-as.data.frame ( mvrnorm ( n =10000 , mu = mu , Sigma = sigma ) )
q4df <-data.table (q4df)
colnames ( q4df ) <-c("ep0","ep1")
X1 <- sample (seq (0 ,40 ,1) ,10000 , replace = TRUE )
X2 <- sample (c(0 ,6 ,9 ,12 ,16 ,18 ,24) ,10000 , replace = TRUE )
q4df <-cbind ( q4df , X1 , X2 )
q4df <-q4df %>%mutate ( w0 = mu0 + beta1 *X1 + ep0,w1 = mu1 + beta1 *X1 + beta2 *X2 + ep1,I=i
f_else ( w1-w0 >0 ,1 ,0) ,prop_formula =1 - pnorm (( mu0 - mu1 - beta2 *X2 )/ sigma.nu ,0 ,1)
)
#the last line : estimate propensity score using the formula

# estimate propensity score using logit
logit <-glm(I~X2,data = q4df,family = binomial( link ="logit") )
q4df$prop_logit = predict ( logit,type ="response")

cor( q4df$prop_formula , q4df$prop_logit ) # correlation = 0.9999974
```

```
## [1] 0.999951
```

```
#IPW
q4df <-q4df %>%
mutate ( w.formula =if_else (I==1 ,1/ prop_formula ,1/(1 - prop_formula ) ) ,
w.logit =if_else (I==1 ,1/ prop_logit ,1/(1 - prop_logit ) ) ,
wage =if_else (I==1 , w1 , w0 ) )
ipw <-lm( wage~I, weights = w.formula , data = q4df )
summary (ipw)
```

```
##
## Call:
## lm(formula = wage ~ I, data = q4df, weights = w.formula)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -77.063  -5.235   0.012   5.267  80.505
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.77119    0.08495  268.04   <2e-16 ***
## I            2.30438    0.11882   19.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.311 on 9998 degrees of freedom
## Multiple R-squared:  0.03626,    Adjusted R-squared:  0.03616
## F-statistic: 376.1 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```
ipw$coefficients
```

```
## (Intercept)           I
##    22.77119     2.30438
```

```
ipw.logit <-lm( wage~I, weights = w.logit , data = q4df )
summary (ipw.logit)
```

```
##
## Call:
## lm(formula = wage ~ I, data = q4df, weights = w.logit)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -78.914  -5.232   0.012   5.268  81.224
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.78276    0.08398  271.28   <2e-16 ***
## I            2.29285    0.11876   19.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.398 on 9998 degrees of freedom
## Multiple R-squared:  0.03594,    Adjusted R-squared:  0.03584
## F-statistic: 372.7 on 1 and 9998 DF,  p-value: < 2.2e-16
```

```
ipw.logit$coefficients
```

```
## (Intercept)           I
##   22.782758    2.292855
```

```
# regress w on I
q9 <-lm( wage ~I, data = q4df )
summary ( q9 )
```

```
##
## Call:
## lm(formula = wage ~ I, data = q4df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.6409  -5.0575   0.0114   5.0855  13.7962
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  22.8038     0.4176  54.612  < 2e-16 ***
## I             2.2719     0.4221   5.383  7.5e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.08 on 9998 degrees of freedom
## Multiple R-squared:  0.00289,    Adjusted R-squared:  0.00279
## F-statistic: 28.98 on 1 and 9998 DF,  p-value: 7.496e-08
```

```
q9$ coefficients
```

```
## (Intercept)           I
##   22.803768    2.271906
```

```
#add covariates
q10 <-lm( wage ~I+ X2 , data = q4df )
summary ( q10 )
```

```
##
## Call:
## lm(formula = wage ~ I + X2, data = q4df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.571  -5.067   0.011   5.087  13.730
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.736606   0.428731  53.032  < 2e-16 ***
## I            2.269544   0.422086   5.377 7.75e-08 ***
## X2           0.005664   0.008194   0.691    0.489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.08 on 9997 degrees of freedom
## Multiple R-squared:  0.002937,   Adjusted R-squared:  0.002738
## F-statistic: 14.73 on 2 and 9997 DF,  p-value: 4.112e-07
```

```
q10 $ coefficients
```

```
##  (Intercept)            I           X2
## 22.736606337  2.269544196  0.005663619
```