TP_2

EL_Hadrami_N'DOYE_Florence

22/01/2021

```r
source("TP2Etudiant.R")
```

```
## Loading required package: xtable
```

EXERCICE

1. Soit $N = 6$ et $n = 3$. Afin d'étudier les distributions des estimateurs, nous supposons connues les valeurs pour toute la population :

$$y_1 = 98 \quad y_3 = 154 \quad y_5 = 190$$

$$y_2 = 102 \quad y_4 = 133 \quad y_6 = 175$$

Creations de deux fonctions un_dedansp et deux_dedansp pour determiner les probabilités d'inclusion du premier et second ordre sur tous les plans.

```r
un_dedansp <- function(p,k,prob){
  sizek <- length(k)
  # nombre d'echantillons
  sizeech <- nrow(p)
  appartenancepik <- matrix(rep(NA,(sizeech * sizek)),nrow = sizek ,ncol=sizeech)
  rownames(appartenancepik) <- paste("k=",k)
  colnames(appartenancepik) <- paste("ech",1:sizeech)
  pik <-  rep(NA,sizek)
  #colnames(pik) <- paste("k=",k)
  for (i in 1:sizeech){
    for(j in 1:sizek){
      appartenancepik[j,i] <- un_dedans(p[i,],k[j])
      appartenancepik[appartenancepik==TRUE] <- 1
      appartenancepik[appartenancepik==FALSE] <- 0
      pik[j] <-  sum(appartenancepik[j,]*prob[i])
    }

  }
  return(list(appartenancepik,pik))
}
```

```r
deux_dedansp <- function(p,k,prob){
  sizek <- length(k)
  # le nombre de couple i j
  sizecoupleij <- choose(sizek,2)
  sizeech <- nrow(p)
  # couple i,j
  coupleij <- t(combn(k,2))
```

```
  appartenancepikl <- matrix(rep(NA,(sizecoupleij * sizeech)),nrow = sizecoupleij ,ncol=sizeech)
  rownames(appartenancepikl) <- paste(coupleij[,1],coupleij[,2],sep = ",")
  colnames(appartenancepikl) <- paste("ech",1:sizeech)
  #rownames(pikl) <- paste(coupleij[,1],coupleij[,2],sep = ",")
  #colnames(pikl) <- paste("ech",1:sizeech)
  pikl <-  rep(NA,sizecoupleij)
  for (i in 1:sizecoupleij){
    for(j in 1:sizeech){
      appartenancepikl[i,j] <- deux_dedans(p[j,],coupleij[i,1],coupleij[i,2])
      appartenancepikl[appartenancepikl==TRUE] <- 1
      appartenancepikl[appartenancepikl==FALSE] <- 0
      pikl[i] <-  sum(appartenancepikl[i,]*prob[j])
    }


  }
  return(list(appartenancepikl,pikl))
}
```

(a) : Déterminons les probabilités d'inclusion du premier et du second ordre pour ces deux plans de sondages.

<u>Plan1</u>

- Probabilité d'inclusion du premier ordre

```
n <- 3
N <- 6
p <- plan1[,1:n]
sizeech <- nrow(p)
k <- c(1:N)
probp1 <- plan1[,4]
rownames(p) <- paste("echantillon",1:sizeech)
appartenancepikp1 <- un_dedansp(p,k,probp1)[[1]]
pikp1 <- un_dedansp(p,k,probp1)[[2]]
p
```

```
##               i1 i2 i3
## echantillon 1  1  3  5
## echantillon 2  1  3  6
## echantillon 3  1  4  5
## echantillon 4  1  4  6
## echantillon 5  2  3  5
## echantillon 6  2  3  6
## echantillon 7  2  4  5
## echantillon 8  2  4  6
```

```
head(appartenancepikp1,6)
```

```
##      ech 1 ech 2 ech 3 ech 4 ech 5 ech 6 ech 7 ech 8
## k= 1     1     1     1     1     0     0     0     0
## k= 2     0     0     0     0     1     1     1     1
## k= 3     1     1     0     0     1     1     0     0
## k= 4     0     0     1     1     0     0     1     1
## k= 5     1     0     1     0     1     0     1     0
## k= 6     0     1     0     1     0     1     0     1
```

```r
head(pikp1,6)
```

```
## [1] 0,5 0,5 0,5 0,5 0,5 0,5
```

- Probabilité d'inclusion du second ordre

```r
n <- 3
N <- 6
p <- plan1[,1:n]
sizeech <- nrow(p)
k <- c(1:N)
rownames(p) <- paste("echantillon",1:sizeech)
appartenancepiklp1 <- deux_dedansp(p,k,probp1)[[1]]
piklp1 <- deux_dedansp(p,k,probp1)[[2]]
p
```

```
##               i1 i2 i3
## echantillon 1  1  3  5
## echantillon 2  1  3  6
## echantillon 3  1  4  5
## echantillon 4  1  4  6
## echantillon 5  2  3  5
## echantillon 6  2  3  6
## echantillon 7  2  4  5
## echantillon 8  2  4  6
```

```r
head(appartenancepiklp1,N)
```

```
##     ech 1 ech 2 ech 3 ech 4 ech 5 ech 6 ech 7 ech 8
## 1,2     0     0     0     0     0     0     0     0
## 1,3     1     1     0     0     0     0     0     0
## 1,4     0     0     1     1     0     0     0     0
## 1,5     1     0     1     0     0     0     0     0
## 1,6     0     1     0     1     0     0     0     0
## 2,3     0     0     0     0     1     1     0     0
```

```r
head(piklp1,N)
```

```
## [1] 0,00 0,25 0,25 0,25 0,25 0,25
```

Plan2

- Probabilité d'inclusion du premier ordre

```r
n <- 3
p <- plan2[,1:n]
k <- c(1:6)
probp2 <- plan2[,4]
sizeech <- nrow(p)
rownames(p) <- paste0("echantillon",1:sizeech)
appartenancepikp2 <- un_dedansp(p,k,probp2)[[1]]
pikp2 <- un_dedansp(p,k,probp2)[[2]]
p
```

```
##              i1 i2 i3
## echantillon1  1  4  6
## echantillon2  2  3  6
## echantillon3  1  3  5
```

```r
head(appartenancepikp2,6)
```

```
##      ech 1 ech 2 ech 3
## k= 1     1     0     1
## k= 2     0     1     0
## k= 3     0     1     1
## k= 4     1     0     0
## k= 5     0     0     1
## k= 6     1     1     0
```

```r
head(pikp2,6)
```

```
## [1] 0,50 0,25 0,50 0,25 0,25 0,50
```

- Probabilité d'inclusion du second ordre

```r
n <- 3
p <- plan2[,1:n]
sizeech <- nrow(p)
k <- c(1:6)
rownames(p) <- paste("echantillon",1:sizeech)
appartenancepiklp2 <- deux_dedansp(p,k,probp2)[[1]]
piklp2 <- deux_dedansp(p,k,probp2)[[2]]
p
```

```
##               i1 i2 i3
## echantillon 1  1  4  6
## echantillon 2  2  3  6
## echantillon 3  1  3  5
```

```r
head(appartenancepiklp2,6)
```

```
##     ech 1 ech 2 ech 3
## 1,2     0     0     0
## 1,3     0     0     1
## 1,4     1     0     0
## 1,5     0     0     1
## 1,6     1     0     0
## 2,3     0     1     0
```

```r
head(piklp2,6)
```

```
## [1] 0,00 0,25 0,25 0,25 0,25 0,25
```

(b): La valeur de $\bar{y_u}$

```r
y <- c(98,102,154,133,190,175)
ybarmu <- mean(y)
ybarmu
```

```
## [1] 142
```

(c) Soit $\bar{y}$ la moyenne des valeurs de l'échantillon. Pour chacun des plans trouvons

Plan1

i. $\mathbf{E}[\bar{y}]$:

```r
# Moyenne de chaque echantillon du plan p1
x <- plan1[,1:3]
```

```r
meanechtp1 <- rep(NA,sizeech)
for (i in 1:sizeech){
  meanechtp1[i] <- mean(y[x[i,]])
}
meanechtp1
```

```
## [1] 147,3333 142,3333 140,3333
```

```r
# Moyenne de chaque echantillon du plan p1
ind <- plan2[,1:3]
sizeind <- nrow(ind)
meanechtp2 <- rep(NA,sizeind)
for (i in 1:sizeind){
  meanechtp2[i] <- mean(y[ind[i,]])
}
meanechtp2
```

```
## [1] 135,3333 143,6667 147,3333
```

```r
probp1 <- plan1[,4]
espybarp1 <- sum(probp1 * meanechtp1)
```

```
## Warning in probp1 * meanechtp1: la taille d'un objet plus long n'est pas
## multiple de la taille d'un objet plus court
```

```r
espybarp1
```

```
## [1] 143,7083
```

ii. $\mathbf{V}[\bar{\boldsymbol{y}}]$:

Formule de la variance

$\mathrm{V}[\bar{y}] = E[\bar{y^2}] - \mathrm{E}[\bar{y}]^2$

```r
varybarp1 <- sum(meanechtp1^2*probp1) - espybarp1^2
```

```
## Warning in meanechtp1^2 * probp1: la taille d'un objet plus long n'est pas
## multiple de la taille d'un objet plus court
```

```r
varybarp1
```

```
## [1] 8,484375
```

iii.Biais$[\bar{\boldsymbol{y}}]$:

$\mathrm{Biais}[\bar{y}] = \mathrm{E}[\bar{y}] - \bar{y}_{\mu}$

```r
biaisybarp1 <- espybarp1 - ybarmu
biaisybarp1
```

```
## [1] 1,708333
```

iiv.EQM$[\bar{\boldsymbol{y}}]$:

Comme le biais est nul donc $\mathrm{EQM}[\bar{y}] = \mathrm{V}[\bar{y}]$

```r
eqmybarp1 <- varybarp1
eqmybarp1
```

```
## [1] 8,484375
```

Plan2

i. $\mathbf{E}[\bar{y}]$:

```
probp2 <- plan2[,4]
espybarp2 <- sum(probp2 * meanechtp2)
espybarp2
```

```
## [1] 142,5
```

ii. $\mathbf{V}[\bar{y}]$:

```
varybarp2 <- sum(meanechtp2^2*probp2) - espybarp2^2
varybarp2
```

```
## [1] 19,36111
```

iii.Biais$[\bar{y}]$:

```
biaisybarp2 <- espybarp2 - ybarmu
biaisybarp2
```

```
## [1] 0,5
```

iiv.EQM$[\bar{y}]$:

```
eqmybarp2 <- varybarp2 - biaisybarp2^2
eqmybarp2
```

```
## [1] 19,11111
```

(d)Lequel des plans est le meilleur ? Pourquoi ?

Le plan2 est le meilleur car il minimise le EQM

2. Pour la population utilisée comme exemple en classe, formée de 8 individus.

$$U = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

(a) Déterminons la probabilité de sélection π_i , pour chaque unité i.

Le nombre total des individus est de 8. Donc la probabilité de selectionnée chaque individu est de $\frac{1}{8}$

$$\boxed{\pi_i = \frac{1}{8}}$$

(b) Quelle est dans ce contexte la distribution de $\hat{t} = 8\bar{y}$?

```
Pop <- c (1,2,4,4,7,7, 7,8)
# extraction des echantillons
echtp3 <- plan[,1:4]
meanechtp3 <- rep(NA,nrow(echtp3))
for (i in 1:nrow(echtp3)){
  meanechtp3[i] <- mean(Pop[echtp3[i,]])
}
tchap <- 8 * meanechtp3
#shapiro.test(tchap)
```

3. Estimation d'Horvitz-Thompson

(a) Déterminons les probabilités d'inclusion du premier et du second ordre de ce plan de sondage.

Probabilités d'inclusion du premier ordre

```r
N <- length(Pop)
n <- 4
p <- plan4[,1:n]
sizeech <- nrow(p)
k <- c(1:N)
probp4 <- plan4[,5]
appartenancepikp4 <- un_dedansp(p,k,probp4)[[1]]
pikp4 <- un_dedansp(p,k,probp4)[[2]]
p
```

```
##    i1 i2 i3 i4
## 1   3  4  5  7
## 2   2  4  5  6
## 3   3  4  5  8
## 4   3  4  7  8
## 5   3  4  6  8
## 6   3  4  6  7
## 7   2  5  7  8
## 8   1  6  7  8
## 9   1  2  3  6
## 10  1  5  6  7
## 11  1  2  4  8
## 12  4  5  6  8
## 13  3  5  6  8
## 14  4  5  7  8
## 15  1  5  7  8
## 16  2  5  6  7
## 17  1  4  7  8
## 18  3  5  7  8
## 19  2  4  6  7
## 20  1  2  6  8
```

```r
head(appartenancepikp4,N)
```

```
##       ech 1 ech 2 ech 3 ech 4 ech 5 ech 6 ech 7 ech 8 ech 9 ech 10 ech 11 ech 12
## k= 1     0     0     0     0     0     0     0     1     1      1      1      0
## k= 2     0     1     0     0     0     0     1     0     1      0      1      0
## k= 3     1     0     1     1     1     1     0     0     1      0      0      0
## k= 4     1     1     1     1     1     1     0     0     0      0      1      1
## k= 5     1     1     1     0     0     0     1     0     0      1      0      1
## k= 6     0     1     0     0     1     1     0     1     1      1      0      1
## k= 7     1     0     0     1     0     1     1     1     0      1      0      0
## k= 8     0     0     1     1     1     0     1     1     0      0      1      1
##       ech 13 ech 14 ech 15 ech 16 ech 17 ech 18 ech 19 ech 20
## k= 1      0      0      1      0      1      0      0      1
## k= 2      0      0      0      1      0      0      1      1
## k= 3      1      0      0      0      0      1      0      0
## k= 4      0      1      0      0      1      0      1      0
## k= 5      1      1      1      1      0      1      0      0
## k= 6      1      0      0      1      0      0      1      1
## k= 7      0      1      1      1      1      1      1      0
## k= 8      1      1      1      0      1      1      0      1
```

```r
head(pikp4,N)
```

```
## [1] 0,35 0,35 0,40 0,55 0,55 0,55 0,60 0,65
```

Probabilités d'inclusion du seconde ordre

```r
N <- length(Pop)
n <- 4
p <- plan4[,1:n]
sizeech <- nrow(p)
k <- c(1:N)
appartenancepiklp4 <- deux_dedansp(p,k,probp4)[[1]]
pikp4 <- un_dedansp(p,k,probp4)[[2]]
p
```

```
##    i1 i2 i3 i4
## 1   3  4  5  7
## 2   2  4  5  6
## 3   3  4  5  8
## 4   3  4  7  8
## 5   3  4  6  8
## 6   3  4  6  7
## 7   2  5  7  8
## 8   1  6  7  8
## 9   1  2  3  6
## 10  1  5  6  7
## 11  1  2  4  8
## 12  4  5  6  8
## 13  3  5  6  8
## 14  4  5  7  8
## 15  1  5  7  8
## 16  2  5  6  7
## 17  1  4  7  8
## 18  3  5  7  8
## 19  2  4  6  7
## 20  1  2  6  8
```

```r
head(appartenancepiklp4,N)
```

```
##     ech 1 ech 2 ech 3 ech 4 ech 5 ech 6 ech 7 ech 8 ech 9 ech 10 ech 11 ech 12
## 1,2     0     0     0     0     0     0     0     0     1      0      1      0
## 1,3     0     0     0     0     0     0     0     0     1      0      0      0
## 1,4     0     0     0     0     0     0     0     0     0      0      1      0
## 1,5     0     0     0     0     0     0     0     0     0      1      0      0
## 1,6     0     0     0     0     0     0     0     1     1      1      0      0
## 1,7     0     0     0     0     0     0     0     1     0      1      0      0
## 1,8     0     0     0     0     0     0     0     1     0      0      1      0
## 2,3     0     0     0     0     0     0     0     0     1      0      0      0
##     ech 13 ech 14 ech 15 ech 16 ech 17 ech 18 ech 19 ech 20
## 1,2      0      0      0      0      0      0      0      1
## 1,3      0      0      0      0      0      0      0      0
## 1,4      0      0      0      0      1      0      0      0
## 1,5      0      0      1      0      0      0      0      0
## 1,6      0      0      0      0      0      0      0      1
## 1,7      0      0      1      0      1      0      0      0
## 1,8      0      0      1      0      1      0      0      1
## 2,3      0      0      0      0      0      0      0      0
```

```r
head(pikp4,N)
```

```
## [1] 0,35 0,35 0,40 0,55 0,55 0,55 0,60 0,65
```

(b) Ajoutons au tableau du tableau plan4 une colonne donnant l'estimation du total calculé par la formule (1)

L'estimation total est :

$$\widehat{\tau} = \sum_{k \in \omega} \frac{y_k}{\pi_k}$$

```r
echp4 <- plan4[,1:4]
sizeechp4 <- nrow(echp4)
# estimation total
tauchap <- rep(NA,sizeechp4)
# probabilité d'inclusion d'un individu
pind <- 0.25
for (i in 1: sizeechp4){
  tauchap[i] <- sum(echp4[i,] / 0.25)
}
tauchap
```

```
##  [1] 76 68 80 88 84 80 88 88 48 76 60 92 88 96 84 80 80 92 76 68
```

```r
# ajout d'une colonne tauchapeaux
plan4tauchapp <- cbind(plan4,tauchap)
```

(c) Évaluons l'espérance mathématique de cette variable aléatoire à l'aide du tableau obtenu. L'estimateur est-il biaisé ?

```r
tauchap <- plan4tauchapp[,6]
pr_ech <- plan4tauchapp[,5]
esptauchap <- sum(pr_ech * tauchap)
esptauchap
```

```
## [1] 79,6
```

```r
esptauchap - mean(esptauchap)
```

```
## [1] 0
```

soit $E[\widehat{\tau}] =$ esptauchap et $\widehat{\widehat{\tau}}$ (moyenne de tauchap), on a montrer que $E[\widehat{\tau}] - \widehat{\widehat{\tau}} = 0$ donc l'esperance n'est pas biaisé.

(d) Évaluons la variance de cette variable aléatoire toujours à l'aide du tableau obtenu.

```r
vartauchap <- sum(tauchap^2 * pr_ech) - (esptauchap^2)
vartauchap
```

```
## [1] 129,44
```

(e) Calculons la variance avec la formule (2), page2.

$$\text{Var}[\widehat{\tau}] = \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} (\pi_k \pi_l - \pi_{kl})(\frac{t_k}{\pi_k} - \frac{t_l}{\pi_l})$$

```r
for (k in 1:N-1){
  for(l in k +1 :N){
    vartau <- sum((pind^2 - pind) * (((echp4[k,]) / pind) - ((echp4[l,] / pind)^2)))
  }
}
```