MAKAROFF_NICOLAS_TPSTAT4

Nicolas Makaroff

19/04/2019

Test de Student

Question 1

(a)

Sous les hypothèses H_0: $\mu = \mu_0$ et H_1: $\mu = \mu_1$, α correspond à la probabilité de l'Erreur de première espèce appellé aussi **seuil** ou encore **niveau de signification du test**. Cette valeur représente la probabilité de rejeter l'hypothèse h_0 à tord. Elle s'exprime mathématiquement par : $\mathbb{P}[\text{choisir } H_1 | H_0 \text{vraie}]$

(b)

Afin de trouver la statistique de test, on applique le test de Nieman-Pearson : $\Lambda(x_1, x_2, ..., x_n) = \frac{L(x_1,...,x_n;\mu_1)}{L(x_1,...,x_n;\mu_0)}$ et on en déduira la zone de rejet : $W = \{(x_1, ..., x_n) | \Lambda(x_1, x_2, ..., x_n) = \frac{L(x_1,...,x_n;\mu_1)}{L(x_1,...,x_n;\mu_0)} > K_\alpha$ avec K_α choisi tel que $\mathbb{P}[\Lambda > K_\alpha] = \alpha$ On obtient $\Lambda_n = \frac{\sqrt{n}(\bar{X_n} - \mu_0)}{S_n}$ Sous l'hypothèse H_0, Λ_n suit une loi de Student à $n-1$ degré de liberté : T_{n-1}. Ainsi, $P_{H0}(W) = P_{H0}(\Lambda_n > K_\alpha) = 1 - F_{T_{n-1}}(K_\alpha) = \alpha$ où $F_{T_{n-1}}$ est la fonction de répartition de T_{n-1}. On en déduit que $K_\alpha = F^{-1}_{T_{n-1}}(1-\alpha)$ Finalement, on rejette l'hypothèse H_0 si $\bar{X_n} > \mu_0 + \frac{S_n}{\sqrt{n}} F_{T_{n-1}}(1-\alpha)$.

(c)

La rêgle de décision ici est de choisir entre $S_n \rightsquigarrow \mathcal{N}(\mu_0, ?)$ ou $\rightsquigarrow \mathcal{N}(\mu_1, ?)$

```
#alpha : risque de première espèce
Delta <- function(Sn,alpha,mu0,mu1){
  n<- length(Sn)
  emp_Xn <- 1.0 / n *sum(Sn)
  sigma <- 1.0 / (n-1) * sum((Sn - emp_Xn)**2)
  lambda <- sqrt(n / sigma) * (emp_Xn - mu0)
  Ka <- qt(p=1.0 - alpha, df=n-1)
  return (lambda >  Ka)
}

if (Delta(rnorm(20,1,sqrt(2)),0.05,1,1.5)){
  print("H0 est rejeté : mu = mu1")
}else{
  print("H0 n'est pas rejeté : mu = mu0")
}
```

```
## [1] "H0 n'est pas rejeté : mu = mu0"
```

Question 2

(a)

```
acc <- 0
for (i in 1:100){
  if (Delta(rnorm(20,1,sqrt(2)),0.05,1,1.5)){
   acc <- acc + 1
  }
}
print(paste("nombre de rejet du test = ",acc,sep = ""))
```

```
## [1] "nombre de rejet du test = 2"
```

On remarque que le nombre de rejet varie accès peu et qu'il vaut généralement 4.

(b)

```
alpha_list <- c(0.2,0.1,0.05,0.01)
for (i in alpha_list){
  Ka <- qt(p=1.0 - i, df=20-1)
  print(Ka)
}
```

```
## [1] 0.8609506
## [1] 1.327728
## [1] 1.729133
## [1] 2.539483
```

La valeur de α influe directement sur la zone de rejet du test. Ainsi si α augmente K_α va diminuer.

(c)

```
for(i in alpha_list){
    acc <- 0
    for(j in 1:100){
      if(Delta(rnorm(20,1,sqrt(2)),i,1,1.5))
        acc <- acc + 1
    }
    print(paste("alpha = ",i," nombre de rejet du test = ",acc,sep = ""))
}
```

```
## [1] "alpha = 0.2 nombre de rejet du test = 20"
## [1] "alpha = 0.1 nombre de rejet du test = 14"
## [1] "alpha = 0.05 nombre de rejet du test = 5"
## [1] "alpha = 0.01 nombre de rejet du test = 0"
```

Question 3

(a)

```r
for(i in alpha_list){
    acc <- 0
    for(j in 1:100){
      if(Delta(rnorm(20,1.5,sqrt(2)),i,1,1.5))
        acc <- acc + 1
    }
    print(paste("alpha = ",i," nombre de rejet du test = ",acc,sep = ""))
}
```

```
## [1] "alpha = 0.2 nombre de rejet du test = 80"
## [1] "alpha = 0.1 nombre de rejet du test = 60"
## [1] "alpha = 0.05 nombre de rejet du test = 44"
## [1] "alpha = 0.01 nombre de rejet du test = 28"
```

On remarque dans un premier que logiquement que rejeté beaucoup plus que précédemment. D'autre part, α influe encore car plus α est grand plus on se trompe sur le testde décision.

(b)

La puissance d'un test, notée β, est la probabilité de rejeter H_0 lorsque H_1 est vraie ($\beta = \mathbb{P}[\text{choisir } H_1 | H_0 \text{vraie}]$

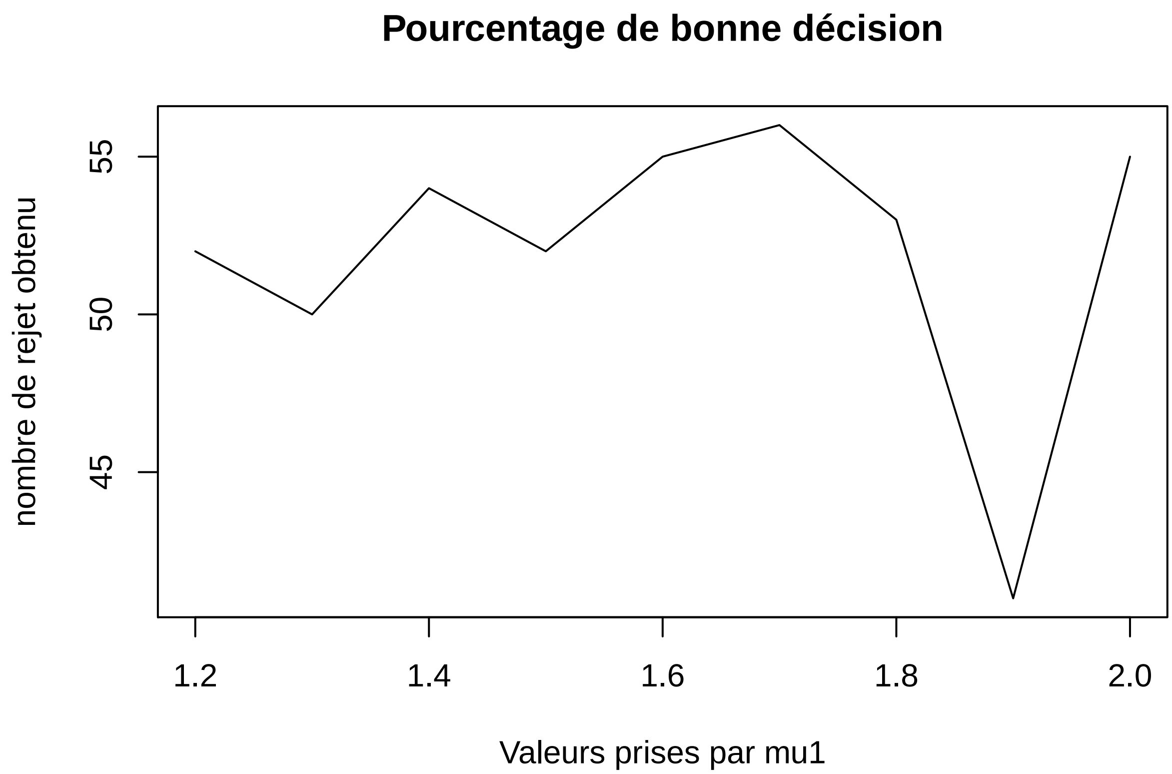
Dans le cas présent, on a : $\beta = P_{H_1}(W) = P_{H_1}(\Lambda_n > K_\alpha) = 1 - F_{T_{n-1}} \frac{\sqrt{n}(\mu - \mu_0)}{S_n} + F^{-1}_{T_{n-1}}(1-\alpha)$

(c)

```r
mu_liste <- c(1.2,1.3,1.4,1.5,1.6,1.7,1.8,1.9,2.0)
test <- c()
for(i in mu_liste){
    acc <- 0
    for(j in 1:100){
      if(Delta(rnorm(20,1.5,sqrt(2)),0.05,1,i))
        acc <- acc +1 #nombre de mauvaise décision
    }
    test <- c(test,acc)
    print(paste("alpha = ",i," nombre de rejet du test = ",acc,sep = ""))
}
```

```
## [1] "alpha = 1.2 nombre de rejet du test = 48"
## [1] "alpha = 1.3 nombre de rejet du test = 50"
## [1] "alpha = 1.4 nombre de rejet du test = 46"
## [1] "alpha = 1.5 nombre de rejet du test = 48"
## [1] "alpha = 1.6 nombre de rejet du test = 45"
## [1] "alpha = 1.7 nombre de rejet du test = 44"
## [1] "alpha = 1.8 nombre de rejet du test = 47"
## [1] "alpha = 1.9 nombre de rejet du test = 59"
## [1] "alpha = 2 nombre de rejet du test = 45"
```

```r
plot(mu_liste,100 - test,xlab = "Valeurs prises par mu1",ylab = "nombre de rejet obtenu",main = "Pource
```



Question 4

(a)

```
t.test(rnorm(20,1,sqrt(2)),mu = 1,var.equal = FALSE)
```

```
##
##  One Sample t-test
##
## data:  rnorm(20, 1, sqrt(2))
## t = -0.54598, df = 19, p-value = 0.5914
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.1123378 1.5203694
## sample estimates:
## mean of x
## 0.8163536
```

t est la valeur correspondant à la statistique donnée par Neyman-Pearson. *df* est le degré de liberté de la loi de Student.

(b)

On sait que la zone de rejet est de la forme : $\bar{X}_n > \mu_0 + \frac{S_n}{\sqrt{n}} F_{n-1}^{-1}$ D'autre part la $p - value = 1 - F_{n-1}^{T}(t)$. On reconnait alors dans la p-value la valeur habituel notée $1 - alpha$ correspondant à la précision.

(c)

```r
for (i in alpha_list){
  test <- t.test(rnorm(20,1,sqrt(2)),mu = 1, conf.level = 1 - i)
  print(test)
}
```

```
##
##  One Sample t-test
##
## data:  rnorm(20, 1, sqrt(2))
## t = -0.29936, df = 19, p-value = 0.7679
## alternative hypothesis: true mean is not equal to 1
## 80 percent confidence interval:
##  0.3774785 1.3934519
## sample estimates:
## mean of x
## 0.8854652
##
##
##  One Sample t-test
##
## data:  rnorm(20, 1, sqrt(2))
## t = -0.70433, df = 19, p-value = 0.4898
## alternative hypothesis: true mean is not equal to 1
## 90 percent confidence interval:
##  0.2504002 1.3156767
## sample estimates:
## mean of x
## 0.7830384
##
##
##  One Sample t-test
##
## data:  rnorm(20, 1, sqrt(2))
## t = 0.99787, df = 19, p-value = 0.3309
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.7221269 1.7842527
## sample estimates:
## mean of x
##   1.25319
##
##
##  One Sample t-test
##
## data:  rnorm(20, 1, sqrt(2))
## t = 1.6843, df = 19, p-value = 0.1085
## alternative hypothesis: true mean is not equal to 1
## 99 percent confidence interval:
##  0.6214422 2.4623569
## sample estimates:
## mean of x
##    1.5419
```

(d)

Calculons l'intervalle de confiance : IC : En utilisant le théorème centrale limite pour se ramener à une loi normale centrée réduite, on obtient alors : $1 - \alpha = \mathbb{P}[\bar{x}_n \in [t^{n-1}_{1-\frac{\alpha}{2}} \times \frac{\sqrt{S_n}}{\sqrt{n-1}}]]$

En récupérant les valeurs de la question, on remarque 1 appartient un nombre de fois qui correspond à la précision du test voulue selon α.