TP4 stats

Romain PEREIRA

27 Avril 2018

1) Test de Student

1) Test sur une distribution $N(\mu = 1, \sigma = 2)$ (en supposant σ inconnue)

(a) Risque de 1ère espèce

H_0 : $\mu = \mu_0$

H_1 : $\mu = \mu_1$

Soit $\boxed{\alpha = \mathbb{P}(choisir \quad H_1 \mid H_0 \quad est \quad vraie)}$ est l'erreur de 1ère espèce.

(b) Lemme de Neyman-Pearson

La région critique optimale au seuil α ('zone de rejet') est W, et peut se mettre sous la forme:

$$\boxed{W = \{\Lambda_n > K_\alpha\}} \quad avec \quad \boxed{\Lambda_n = \frac{\sqrt{n}(\bar{X_n} - \mu_0)}{S_n}} \quad et \quad \boxed{K_\alpha = \mu_0 + \frac{\sigma_0}{\sqrt{n}} F^{-1}_{T_{n-1}}(1-\alpha)}$$

où $F_{T_{n-1}}$ est la fonction de répartition de T_{n-1} (Loi de Student à $n-1$ degrés de liberté).

(c) Programmation de la règle de décision.

```
# Règle de décision : 'S_n' suit elle une loi N('u0', *) ou N('u1', *) ?
# sous un risque de 1ère espèce 'a'
# la fonction renvoie True si l'hypothèse H0 : u = u0 est rejeté, False sinon
delta <- function(S_n, a, u0, u1) {
  n        <- length(S_n)
  X_n_bar  <- 1.0 / n * sum(S_n)
  sigma_2  <- 1.0 / (n - 1) * sum((S_n - X_n_bar) * (S_n - X_n_bar))
  lambda_n <- sqrt(n / sigma_2) * (X_n_bar - u0)
  K_a      <- qt(p=1.0 - a, df=n-1)
  return (lambda_n > K_a)
}

if (delta(rnorm(n=20, mean=1, sd=2), 0.05, 1.0, 1.5)) {
  print("H0 est rejeté: u=u1")
} else {
  print("H0 non rejeté: u=u0")
}
```

```
## [1] "H0 non rejeté: u=u0"
```

2) Test sur $N = 100$ distributions $N(\mu = 1, \sigma = 2)$ (en supposant σ inconnue)

(a)

```
sum <- 0
for (i in 1:100) {
  if (delta(rnorm(n=20, mean=1, sd=2), 0.05, 1.0, 1.5)) {
    sum <- sum + 1
  }
}
print(paste("nombre de fois que H0 a été rejeté: ", sum, sep=""))
```

```
## [1] "nombre de fois que H0 a été rejeté: 4"
```

On observe que pour certains echantillons, le test de Student rejette l'hypothèse $H_0 : \mu = \mu_0$ (alors qu'on sait que cette hypothèse est vraie)

(b) Si α augmente, alors K_α diminue, la zone de rejet augmente. (on aura plus (+) tendance à choisir H_1 alors que H_0 est vrai)

(c)

```
for (a in c(0.2, 0.1, 0.05, 0.01)) {
  sum <- 0
  for (i in 1:100) {
    if (delta(rnorm(n=20, mean=1, sd=2), a, 1.0, 1.5)) {
      sum <- sum + 1
    }
  }
  print(paste("alpha=", a, " ; nombre de fois que H0 a été rejeté: ", sum, sep=""))
}
```

```
## [1] "alpha=0.2 ; nombre de fois que H0 a été rejeté: 20"
## [1] "alpha=0.1 ; nombre de fois que H0 a été rejeté: 8"
## [1] "alpha=0.05 ; nombre de fois que H0 a été rejeté: 9"
## [1] "alpha=0.01 ; nombre de fois que H0 a été rejeté: 0"
```

3) Test sur $N = 100$ distributions $N(\mu = 1.5, \sigma = 2)$ (en supposant σ inconnue)

(a)

```
for (a in c(0.2, 0.1, 0.05, 0.01)) {
  sum <- 0
  for (i in 1:100) {
    # si vrai, alors H0 est rejeté
    if (delta(rnorm(n=20, mean=1.5, sd=2), a, 1.0, 1.5)) {
      sum <- sum + 1
    }
  }
  print(paste("nombre de fois que H0 a été rejeté: ", sum, sep=""))
}
```

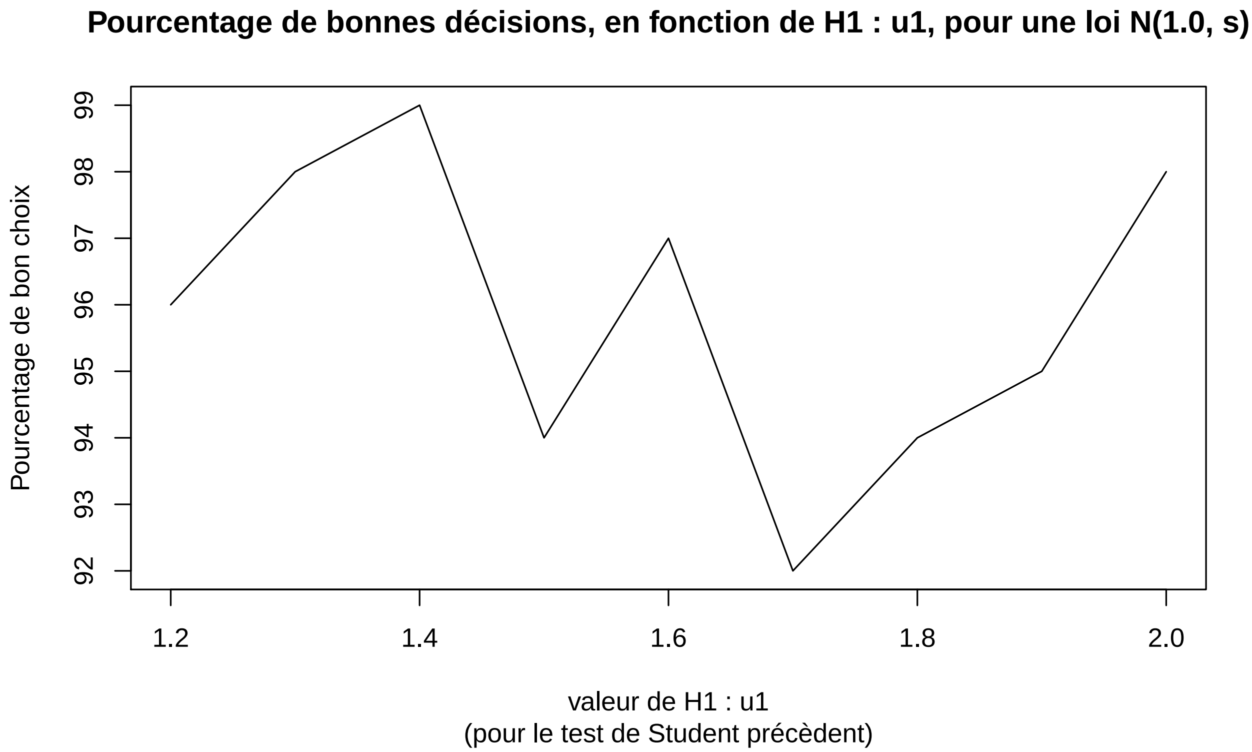
```
## [1] "nombre de fois que H0 a été rejeté: 63"
## [1] "nombre de fois que H0 a été rejeté: 51"
## [1] "nombre de fois que H0 a été rejeté: 25"
## [1] "nombre de fois que H0 a été rejeté: 12"
```

On constate que plus α est grand, moins l'on rejète H0, et donc plus l'on se trompe si l'on suit le test.

(b) La puissance d'un test, notée P est la probabilité de rejeter H_0 lorsque H_1 est vraie.

(c)

```
u1s <- c(1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0)
y   <- c()
for (u1 in u1s) {
  sum <- 0 # nombre de fois ou on rejete H0 (alors que H0 est vrai)
  for (i in 1:100) {
    if (delta(rnorm(n=20, mean=1.0, sd=2), 0.05, 1.0, u1)) {
      sum <- sum + 1
    }
  }
  y <- c(y, sum)
}
plot(u1s, 100 - y, xlab="valeur de H1 : u1", ylab="Pourcentage de bon choix", main="Pourcentage de bonn
```



4) Utilisation de 't.test'

```r
Sn <- rnorm(n=20, mean=1.0, sd=2)
t.test(Sn, mu=1.0)
```

```
##
##  One Sample t-test
##
## data:  Sn
## t = 0.62063, df = 19, p-value = 0.5422
## alternative hypothesis: true mean is not equal to 1
## 95 percent confidence interval:
##  0.3665811 2.1674044
## sample estimates:
## mean of x
##  1.266993
```

t est la valeur qui correspond au Λ_n précèdent. C'est le t du test de Student. (https://en.wikipedia.org/wiki/Student's_t-test)

df (degree of freedom) est le nombre de degree de liberté dans le test de Student.