TP 10 : Révisions : Intervalles de confiance, normalité asymptotique, tests

Exercice 1. Taux d'incidence dans deux sous-populations.

1. Réponse : On a deux variables X_A (resp. X_B) suivant une loi binomiale de paramètre p_A (resp. p_B). On veut tester $H_0 :=$ "$p_A = p_B$" contre $H_1 :=$ "$p_A \neq p_B$".

2. Réponse : le TCL donne seulement la convergence en loi suivante :

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \to \mathcal{N}(0,1).$$

Problème : p est inconnu est est au dénominateur. On remplace donc les p au dénominateur en utilisant la loi des grands nombres : $\bar{X}_n \to p$ presque surement. Pour agréger ces deux résultats, on fait appel au Lemme de Slutsky, qui donne au final :

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \to \mathcal{N}(0,1).$$

On a une normalité asymptotique, on utilise donc les quatiles de la loi normale. On a

$$\mathbb{P}\left(q_{\alpha/2} \leq \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leq q_{1-\alpha/2}\right) \to 1-\alpha,$$

ce qui donne, par symmétrie de la gaussienne ($q_{\alpha/2} = -q_{1-\alpha/2}$) :

$$\mathbb{P}\left(\bar{X}_n - \frac{q_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}} \leq p \leq \bar{X}_n + \frac{q_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}\right) \to 1-\alpha,$$

d'où l'intervalle de confiance EXACT, ASYMPTOTIQUE suivant :

$$IC_{1-\alpha} := \left[\bar{X}_n \pm \frac{q_{1-\alpha/2}\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}\right].$$

3.

Voici le code :

```
alpha = 0.05 #le niveau de confiance donné
## premier intervalle pour p_A
nA = 730
XA = 12
fA = XA/nA
```

```
IC0 = c(fA - qnorm(1-alpha/2)*sqrt(fA*(1-fA))/sqrt(nA),fA + qnorm(1-alpha/2)*sqrt(fA*(1-fA))/sqrt(nA))

## second intervalle pour p_B
nB = 1256
XB = 3
fB = XB/nB
IC1 = c(fB - qnorm(1-alpha/2)*sqrt(fB*(1-fB))/sqrt(nB),fB + qnorm(1-alpha/2)*sqrt(fB*(1-fB))/sqrt(nB))

IC0
```

```
## [1] 0.007214412 0.025662300
```

```
IC1
```

```
## [1] -0.0003110673  0.0050881374
```

Conclusion : Les intervalles sont disjoints : on peut donc rejeter H_0, et p_B est significativement inférieur à p_A (autrement dit, les pla population B soint moins malades que celles de la population A).

Exercice 2. Maximum de vraisemblance et intervalle de confiance

1. C'est exactement les mêmes arguments que précédemment : le lemme de Slutsky permet de remplacer le a du dénominateur par son estimée. Avec les quantiles de la loi gaussienne, l'intervalle de confiance obtenu est

$$IC_{1-\alpha} := \left[\hat{a} \pm \frac{q_{1-\alpha/2}(\hat{a}-1)}{\sqrt{n}}\right].$$

Cet intervalle est exact asymptotique.

2. Voici le code :

```
n=1000
a=2.2
alpha=0.01

U=runif(n)
X=U^(-1/(a-1))

hata = 1 + 1/(mean(log(X)))
IC = c(hata - qnorm(1-alpha/2)*(hata-1)/sqrt(n),hata + qnorm(1-alpha/2)*(hata-1)/sqrt(n))
IC
```

```
## [1] 2.097442 2.292080
```

Réponse : on constate qu'à chaque fois (quasiment), le paramètre a tombe bien dans l'intervalle.

3. Voici le code :

```
data = read.csv("moby_dick.csv")
X=data$occurrences

alpha=0.05
hata = 1 + 1/(mean(log(X)))
IC = c(hata - qnorm(1-alpha/2)*(hata-1)/sqrt(n),hata + qnorm(1-alpha/2)*(hata-1)/sqrt(n))
IC
```

```
## [1] 2.125723 2.274487
```

Exercice 3. Le Titanic : à vous de jouer (non corrigé)