

Science des données I : module 78



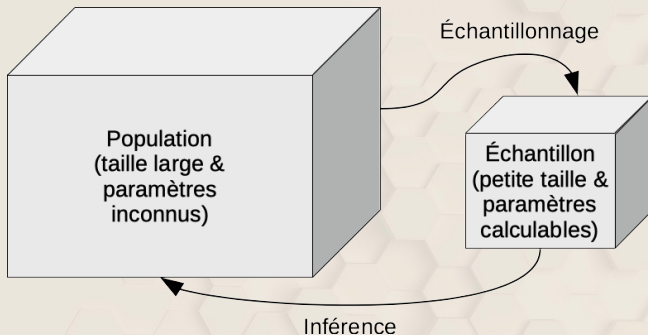
Reprise et calculs probabilités - quantiles

Philippe Grosjean & Guyliann Engels

Université de Mons, Belgique
Laboratoire d'Écologie numérique



<https://wp.sciviews.org>
sdd@sciviews.org



Le test d'hypothèse

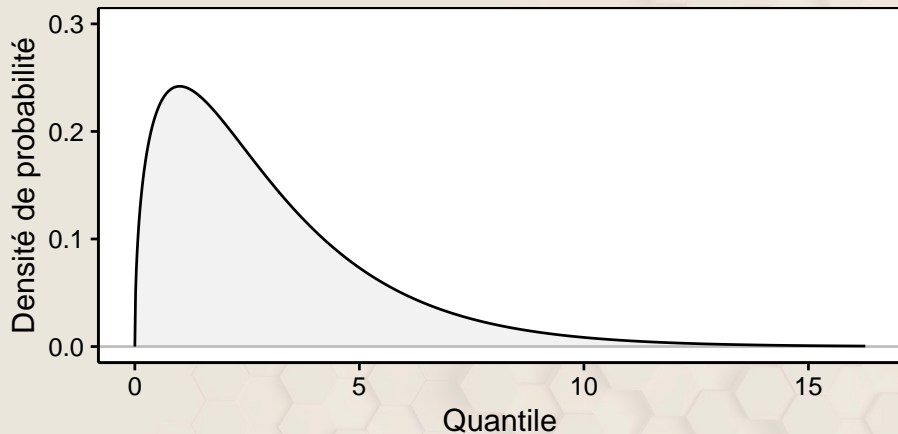
Le test d'hypothèse est l'outil le plus simple pour répondre à une question via l'inférence.

Il faut déterminer

- H_0 , l'hypothèse nulle \Rightarrow l'affirmation à réfuter
- H_1 , l'hypothèse alternative

Distribution du χ^2

Cette distribution admet un paramètre qui sont les degrés de liberté (valeur de 3 pour la distribution ci-dessous)



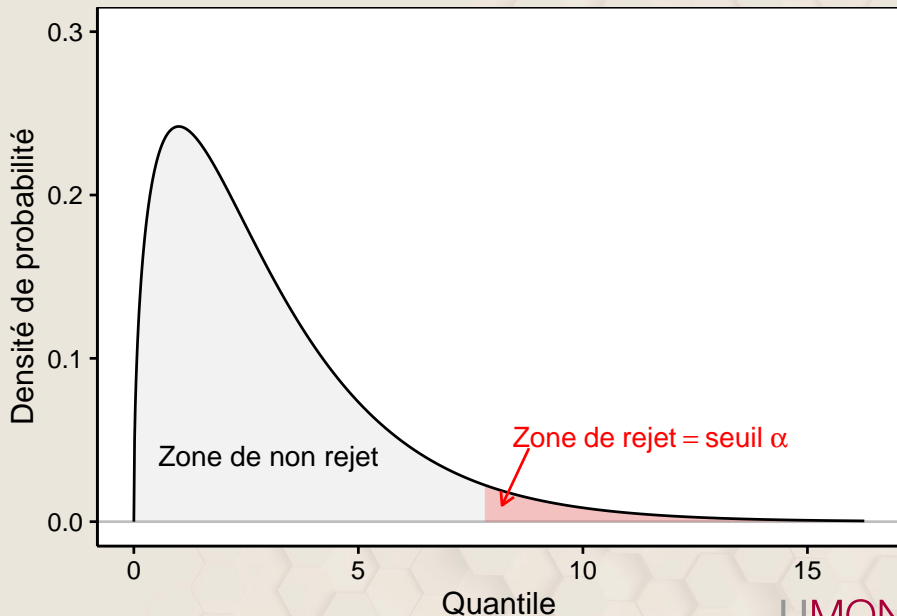
Zone de rejet

Cette zone de rejet est déterminée par une probabilité nommé le seuil α déterminé avant de réaliser un test d'hypothèse.

Nous fixons le seuil alpha à 5%

```
qchisq(0.05, 3, lower.tail = FALSE)
```

```
## [1] 7.814728
```



Situation fictive

Notre hypothèse de départ est

H_0 : il est équiprobable d'être du groupe A, B, C ou D

Nos observations :

```
(tab <- as.table(c(A = 5, B = 10, C = 20, D = 10)))
```

```
##  A  B  C  D
```

```
##  5 10 20 10
```

```
chisq.test(tab, p = c(1/4, 1/4, 1/4, 1/4))
```

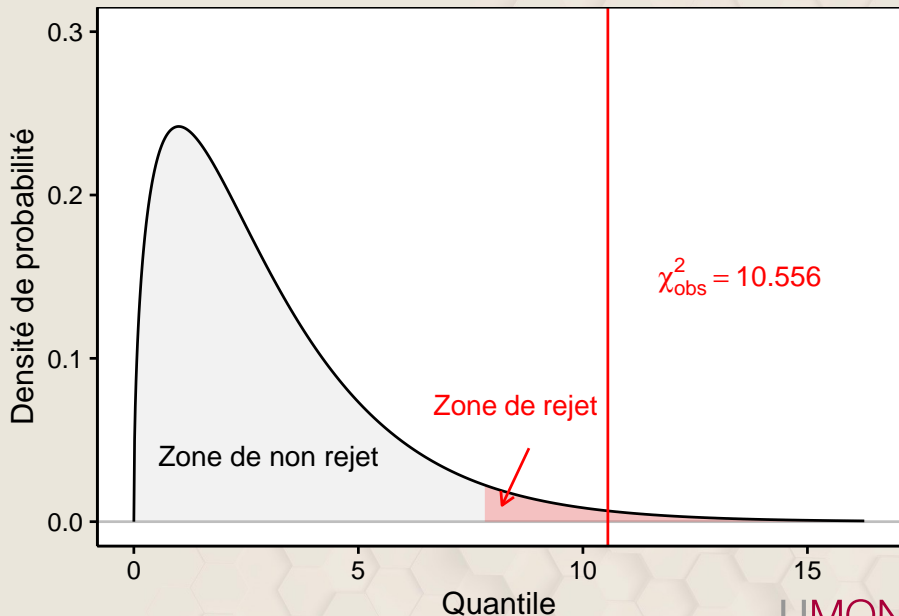
```
##
```

```
## Chi-squared test for given probabilities
```

```
##
```

```
## data:  tab
```

```
## X-squared = 10.556, df = 3, p-value = 0.01439
```



Les conditions d'application

Assurez vous que les conditions d'application soient respectées

- échantillonnage aléatoire et observations indépendantes,
- aucun effectif théorique (ou probabilité) sous H_0 nul,
- aucun effectif observé, si possible, inférieur à 5 (ceci n'est **pas** une condition stricte ; le test sera “approximativement” bon dans le cas contraire).