

Introduction à la statistique bayésienne

Ladislav Nalborczyk

Aix-Marseille Univ, CNRS, LPC, LNC, Marseille, France

05.10.2022

 @lnalborczyk  Slides >> tinyurl.com/introbayes2022

Au programme

1. Contexte théorique

- Interprétations du concept de probabilité
- Qu'est-ce que l'inférence bayésienne ?
- Rappels : Théorie des probabilités
- Rappels : Test d'hypothèse fréquentiste
- Test d'hypothèse bayésien
- Analyse de la variance (ANOVA) bayésienne

2. Travaux pratiques avec jamovi

- Présentation du module d'analyse bayésienne
- Conduire une ANOVA bayésienne, interpréter, et rapporter les résultats

Contexte théorique

Interprétations du concept de probabilité

Axiomes des probabilités (Kolmogorov, 1933)

Une probabilité est une valeur numérique assignée à un événement A , compris comme une possibilité appartenant à l'univers Ω (l'ensemble de toutes les issues possibles). Les probabilités se conforment aux axiomes suivants :

- Non-negativity : $\Pr(A) \geq 0$
- Normalisation : $\Pr(\Omega) = 1$
- Additivity (pour des événements incompatibles) : $\Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2)$

Le dernier axiome est également connu comme la [règle de la somme](#), et peut se généraliser à des événements non mutuellement exclusifs : $\Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2)$.

Interprétations probabilistes

Quelle est la probabilité...

- D'obtenir un chiffre pair sur un lancer de dé ?
- Que j'apprenne quelque chose pendant ce cours ?

Est-ce qu'il s'agit, pour chaque exemple, de la même sorte de probabilité ?



Interprétation classique (ou théorique)

$$\Pr(\text{pair}) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{3}{6} = \frac{1}{2}$$

Problème : cette définition est uniquement applicable aux situations dans lesquelles il n'y a qu'un nombre fini de résultats possibles équiprobables...

Quelle est la probabilité qu'il pleuve demain ?

$$\Pr(\text{pluie}) = \frac{\text{pluie}}{\{\text{pluie, non-pluie}\}} = \frac{1}{2}$$

Interprétation fréquentiste (ou empirique)

$$\Pr(x) = \lim_{n_t \rightarrow \infty} \frac{n_x}{n_t}$$

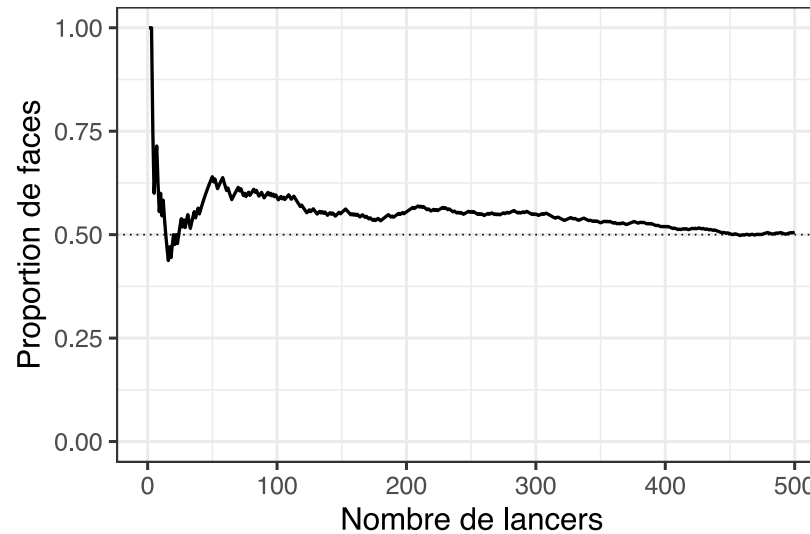
Où n_x est le nombre d'occurrences de l'événement x et n_t le nombre total d'essais. L'interprétation fréquentiste postule que, à long-terme (i.e., quand le nombre d'essais s'approche de l'infini), la fréquence relative va converger *exactement* vers ce qu'on appelle "probabilité".

Conséquence : le concept de probabilité s'applique uniquement aux collectifs, et non aux événements singuliers.

Interprétation fréquentiste (ou empirique)

```
library(tidyverse)

sample(c(0, 1), 500, replace = TRUE) %>%
  data.frame %>%
  mutate(x = seq_along(.), y = cumsum(.) / seq_along(.) ) %>%
  ggplot(aes(x = x, y = y), log = "y") +
  geom_line(lwd = 1) +
  geom_hline(yintercept = 0.5, lty = 3) +
  labs(x = "Nombre de lancers", y = "Proportion de faces") +
  ylim(0, 1) + theme_bw(base_size = 20)
```



Limites de l'interprétation fréquentiste...

Quelle classe de référence ? Quelle est la probabilité que je vive jusqu'à 80 ans ? En tant qu'homme ? En tant que Français ?

Quid des événements qui ne peuvent pas se répéter ? Quelle est la probabilité que j'apprenne quelque chose pendant ce cours ?

À partir de combien de lancers (d'une pièce par exemple) a-t-on une bonne approximation de la probabilité ? Une classe finie d'événements de taille n ne peut produire que des fréquences relatives de précision $1/n$...

Interprétation logique

Il y a 10 étudiants dans cette salle
9 portent un t-shirt **vert**
1 porte un t-shirt **rouge**
Une personne est tirée au sort...

Conclusion n°1 : l'étudiant tiré au sort porte un t-shirt ✓

Conclusion n°2 : l'étudiant tiré au sort porte un t-shirt **vert** ✗

Conclusion n°3 : l'étudiant tiré au sort porte un t-shirt **rouge** ✗

Interprétation logique

L'interprétation logique du concept de probabilité essaye de généraliser la logique (vrai / faux) au monde probabiliste. La probabilité représente donc le *degré de support logique* qu'une conclusion peut avoir, relativement à un ensemble de prémisses (Keynes, 1921; Carnap, 1950).

Conséquence : toute probabilité est conditionnelle.

Interprétation bayésienne

La probabilité est une mesure du degré d'incertitude. Un événement *certain* aura donc une probabilité de 1 et un événement *impossible* aura une probabilité de 0.

Pour parler de probabilités, dans ce cadre, nous n'avons donc plus besoin de nous référer à la limite d'occurrence d'un événement (fréquence).

So to assign equal probabilities to two events is not in any way an assertion that they must occur equally often in any random experiment [...], it is only a formal way of saying I don't know.

Edwin T. Jaynes (1986)

Interprétations probabilistes

- Interprétation classique (Laplace, Bernouilli, Leibniz)
- Interprétation fréquentiste (Venn, Reichenbach, von Mises)
- Interprétation propensionniste (Popper, Miller)
- Interprétation logique (Keynes, Carnap)
- Interprétation bayésienne (Jeffreys, de Finetti, Savage)

Plus de détails sur la Stanford Encyclopedia of Philosophy...

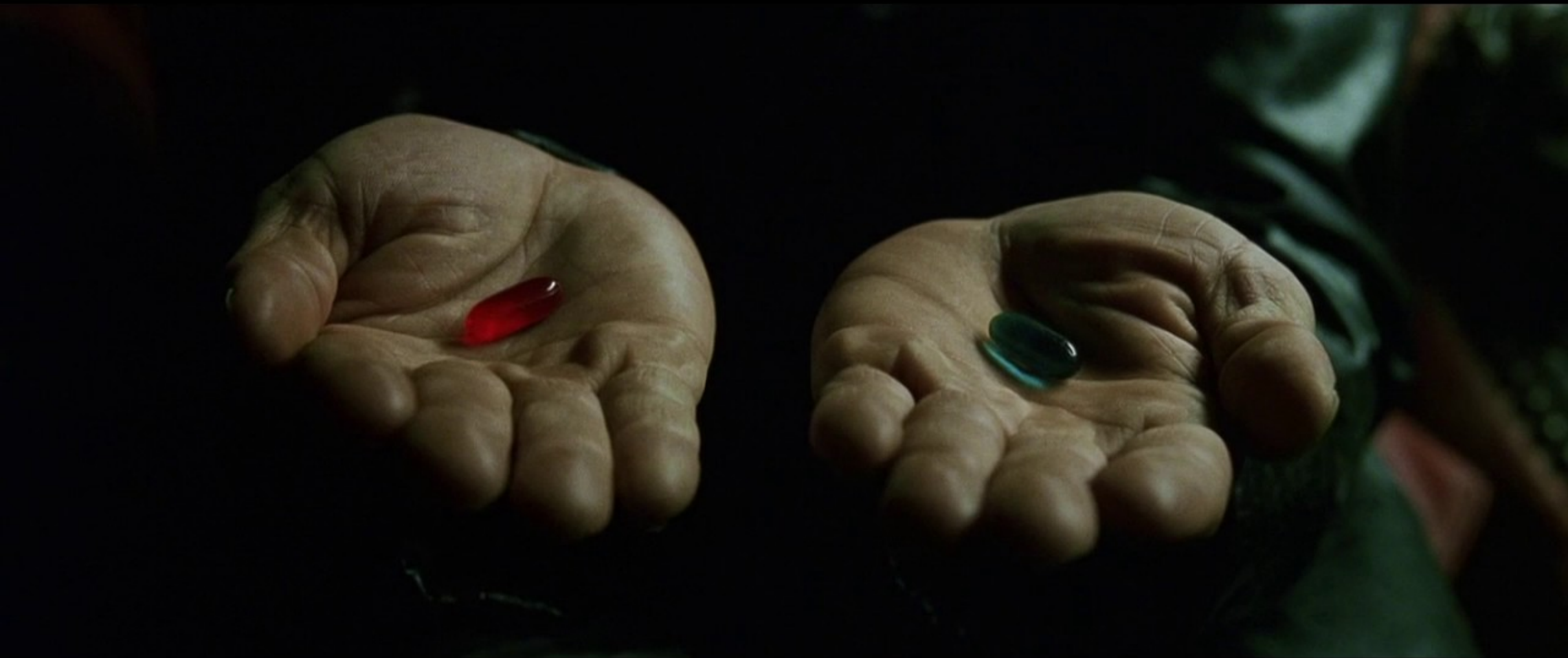
Interprétations probabilistes - résumé

Probabilité épistémique

- Toute probabilité est conditionnelle à de l'information disponible (e.g., prémisses ou données)
- La probabilité comme moyen de quantifier l'incertitude
- Probabilité logique, probabilité bayésienne

Probabilité physique

- Les probabilités dépendent d'un état du monde, de caractéristiques physiques, elles sont indépendantes de l'information disponible (ou de l'incertitude)
- Probabilité classique, probabilité fréquentiste



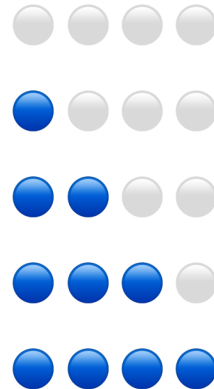
Contexte théorique

Qu'est-ce que l'inférence bayésienne ?

Exercice - Problème du sac de billes (McElreath, 2015)

Imaginons que nous disposions d'un sac contenant 4 billes. Ces billes peuvent être soit blanches, soit bleues. Nous savons qu'il y a précisément 4 billes, mais nous ne connaissons pas le nombre de billes de chaque couleur.

Nous savons cependant qu'il existe cinq possibilités (que nous considérons comme nos *hypothèses*) :

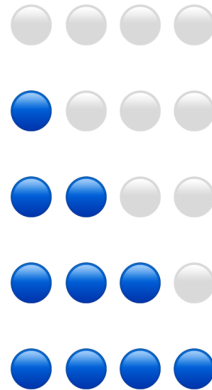


Exercice - Problème du sac de billes (McElreath, 2015)

Le but est de déterminer quelle combinaison serait la plus probable, sachant certaines observations. Imaginons que l'on tire trois billes à la suite, avec remise, et que l'on obtienne la séquence suivante :



Cette séquence représente nos données. À partir de là, quelle inférence peut-on faire sur le contenu du sac ? En d'autres termes, que peut-on dire de la probabilité de chaque hypothèse ?



02:00

Énumérer les possibilités

Hypothèse : ● ○ ○ ○

Données : ●

Énumérer les possibilités

Hypothèse : ☒ ☐ ☐ ☐

Données : ☒ ☐

Énumérer les possibilités

Hypothèse : ● ○ ○ ○

Données : ● ○ ●

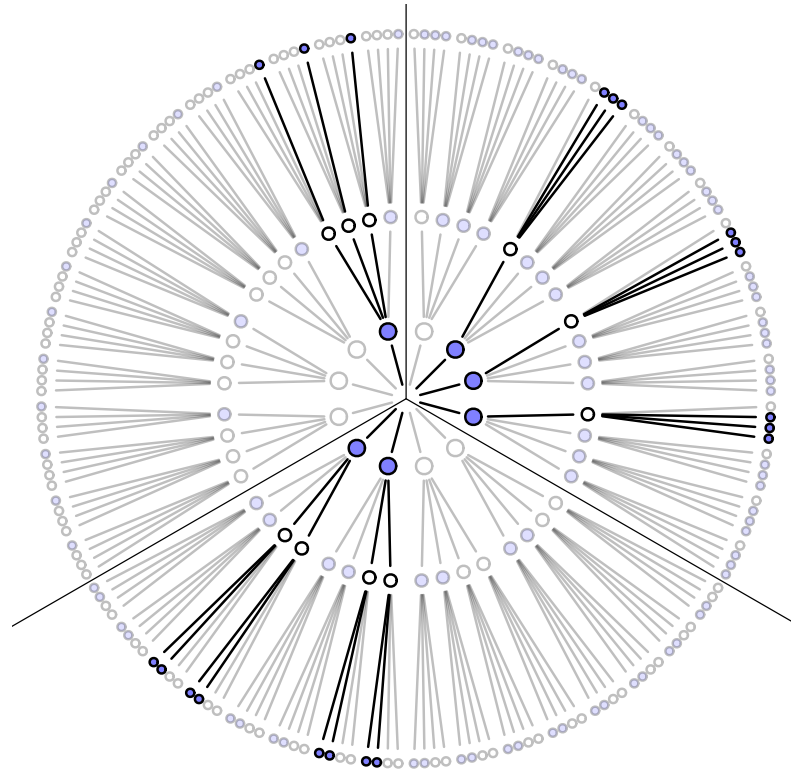
Énumérer les possibilités

Hypothèse : ● ○ ○ ○

Données : ● ○ ●






Énumérer les possibilités

Sous cette hypothèse, 3 chemins (sur $4^3 = 64$) conduisent au résultat obtenu. Qu'en est-il des autres hypothèses?



Comparer les hypothèses

Au vu des données, cette hypothèse est la plus *probable* car c'est l'hypothèse qui maximise le nombre de manières possibles d'obtenir les données obtenues.

Hypothèse	Façons d'obtenir les données
	$0 \times 4 \times 0 = 0$
	$1 \times 3 \times 1 = 3$
	$2 \times 2 \times 2 = 8$
	$3 \times 1 \times 3 = 9$
	$4 \times 0 \times 4 = 0$

Accumulation d'évidence






Dans le cas précédent, nous avons considéré que toutes les hypothèses étaient équiprobables a priori (suivant le principe d'indifférence).

Cependant, on pourrait avoir de l'information a priori, provenant de nos connaissances (des particularités des sacs de billes par exemple) ou de données antérieures.

Imaginons que nous tirions une nouvelle bille du sac, comment incorporer cette nouvelle donnée ?






Accumulation d'évidence

Il suffit d'appliquer la même stratégie que précédemment, et de mettre à jour le dernier compte en le multipliant par ces nouvelles données. *Yesterday's posterior is today's prior* (Lindley, 2000).

Hypothèse	Façons de produire ●	Compte précédent	Nouveau compte
	0	0	$0 \times 0 = 0$
	1	3	$3 \times 1 = 3$
	2	8	$8 \times 2 = 16$
	3	9	$9 \times 3 = 27$
	4	0	$0 \times 4 = 0$

Incorporer un prior

Un employé de l'usine de fabrication des billes nous dit que pour chaque sac contenant 3 billes bleues, ils fabriquent deux sacs en contenant seulement deux, et trois sacs en contenant seulement une. Il nous apprend également que tous les sacs contiennent au moins une bille bleue et une bille blanche...

Hypothèse	Comptes précédents	Prior usine	Nouveau compte
	0	0	$0 \times 0 = 0$
	3	3	$3 \times 3 = 9$
	16	2	$16 \times 2 = 32$
	27	1	$27 \times 1 = 27$
	0	0	$0 \times 0 = 0$

Des énumérations aux probabilités

La probabilité d'une hypothèse après avoir observé certaines données est proportionnelle au nombre de façons qu'a cette hypothèse de produire les données observées, multiplié par sa probabilité a priori.






$$\Pr(\text{hypothèse} \mid \text{données}) \propto \Pr(\text{données} \mid \text{hypothèse}) \times \Pr(\text{hypothèse})$$

Pour calculer la probabilité d'une hypothèse, il suffit de standardiser ce produit (i.e., nombre de façons de produire les données multiplié par la probabilité a priori de l'hypothèse) de manière à ce que la somme des probabilités de toutes les hypothèses possibles soit égale à 1.

$$\Pr(\text{hypothèse} \mid \text{données}) = \frac{\Pr(\text{données} \mid \text{hypothèse}) \times \Pr(\text{hypothèse})}{\text{somme des produits}}$$

Des énumérations aux probabilités

Par exemple, définissons p comme la proportion de billes bleues.

Hypothèse	p	Manières de produire les données	Probabilité
	0	0	0
	0.25	3	0.15
	0.5	8	0.40
	0.75	9	0.45
	1	0	0

```
chemins <- c(0, 3, 8, 9, 0)  
(probs <- chemins / sum(chemins) )
```

```
## [1] 0.00 0.15 0.40 0.45 0.00
```

Notations, terminologie

- θ est un paramètre ou vecteur de paramètres (e.g., la proportion de billes bleues).
- $p(x|\theta)$ la probabilité conditionnelle des données x sachant le paramètre θ [$p(x|\theta = \theta)$].
- $p(x|\theta)$ une fois que la valeur de x est connue, est vue comme la fonction de vraisemblance (*likelihood*) du paramètre θ . Attention, il ne s'agit pas d'une distribution de probabilité [$p(x = x|\theta)$].
- $p(\theta)$ la probabilité a priori de θ .
- $p(\theta|x)$ la probabilité a posteriori de θ (sachant x).
- $p(x)$ la probabilité marginale de x (sur θ) ou "vraisemblance marginale", "vraisemblance intégrée".

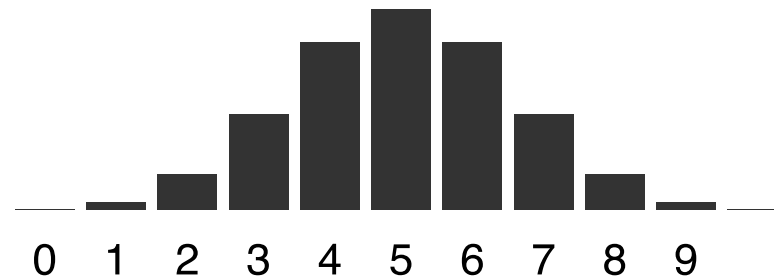
$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\sum_{\theta} p(x|\theta)p(\theta)} = \frac{p(x|\theta)p(\theta)}{\int_{\theta} p(x|\theta)p(\theta)d\theta} \propto p(x|\theta)p(\theta)$$

Contexte théorique

Rappels : Théorie des probabilités

Loi de probabilité, cas discret

Une fonction de masse (*probability mass function*, ou *PMF*) est une fonction qui attribue une probabilité à chaque valeur d'une variable aléatoire. Exemple de la distribution binomiale pour une pièce non biaisée, c'est à dire lorsque $\theta = 0.5$, qui nous indique la probabilité d'obtenir N faces sur 10 lancers.

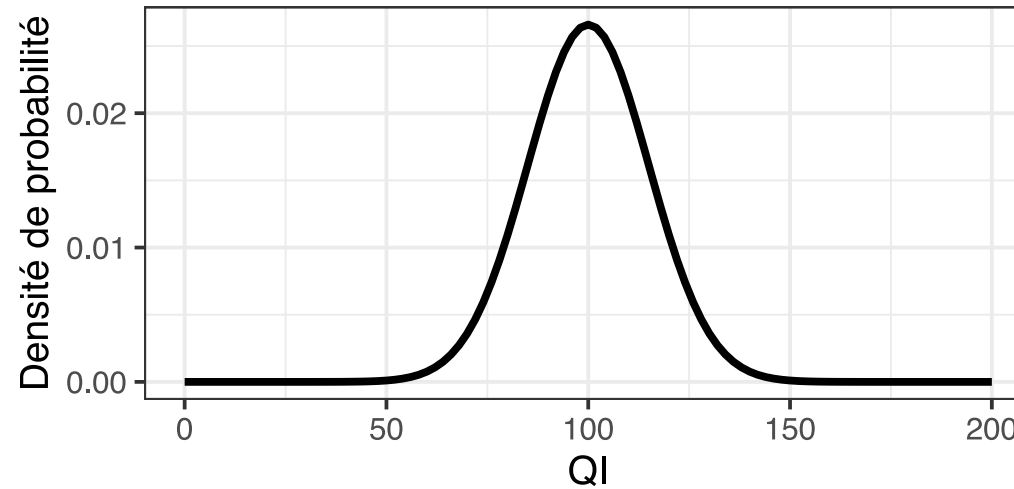


```
# PMFs sum to 1
dbinom(x = 0:10, size = 10, prob = 0.5) %>% sum
```

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


Loi de probabilité, cas continu

Une (fonction de) densité de probabilité (*probability density function*, ou *PDF*), est une fonction qui permet de représenter une loi de probabilité sous forme d'intégrales (l'équivalent de la PMF pour des variables aléatoires strictement continues).

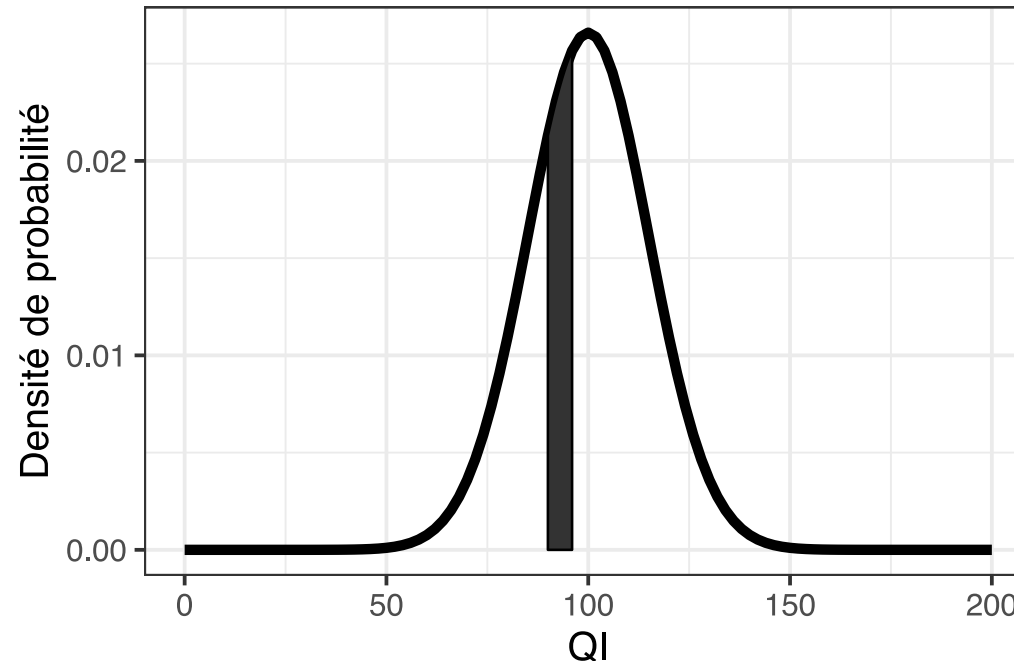


```
# PDFs integrate to 1
integrate(dnorm, -Inf, Inf, mean = 100, sd = 15)
```

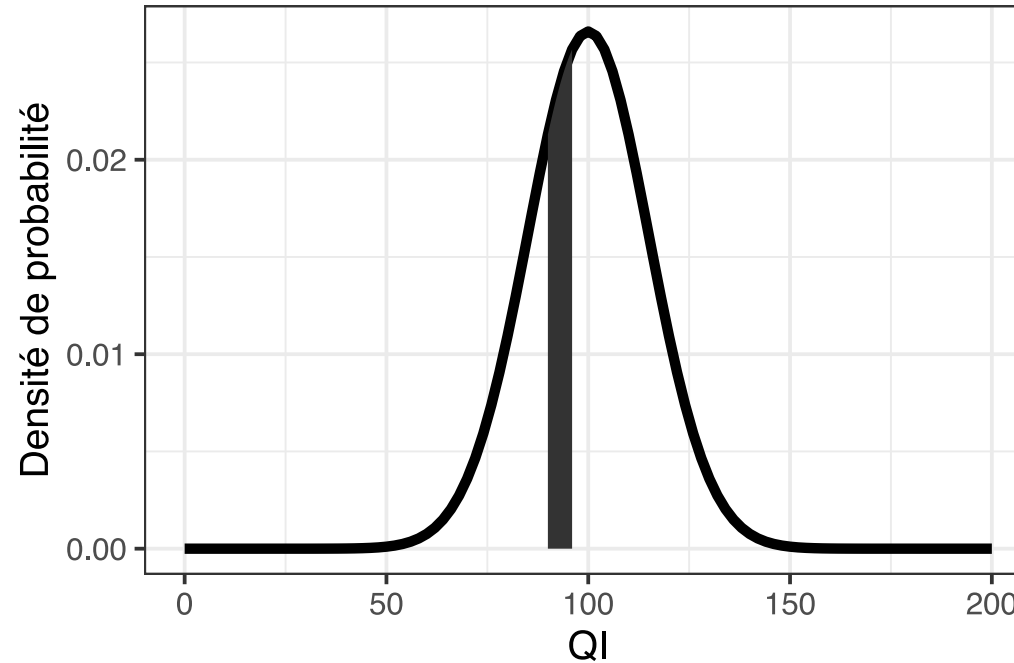
```
## 1 with absolute error < 1.3e-06
```

Aparté, qu'est-ce qu'une intégrale ?

Une intégrale correspond à la *surface* (aire géométrique) délimitée par la représentation graphique d'une fonction, *l'aire sous la courbe*. Une distribution est dite *impropre* si son intégrale n'est pas égale à un nombre fini (e.g., $+\infty$) et *normalisée* si son intégrale est égale à 1.



Aparté, qu'est-ce qu'une intégrale ?



L'intégrale de $f(x)$ sur l'intervalle $[90 ; 96]$ vaut : $p(90 < x < 96) = \int_{90}^{96} f(x) dx = 0.142$.

```
integrate(dnorm, 90, 96, mean = 100, sd = 15)
```

```
## 0.1423704 with absolute error < 1.6e-15
```

Probabilité conjointe

```
library(tidyverse)
data(HairEyeColor) # data adapted from Snee (1974)

cont <- apply(HairEyeColor, c(1, 2), sum) %>% t
cont <- round(cont / sum(cont), 2)
cont
```

```
##           Hair
## Eye      Black Brown  Red  Blond
## Brown  0.11  0.20 0.04  0.01
## Blue   0.03  0.14 0.03  0.16
## Hazel  0.03  0.09 0.02  0.02
## Green  0.01  0.05 0.02  0.03
```

Dans chaque cellule, on trouve la probabilité conjointe d'avoir telle couleur de cheveux ET telle couleur d'yeux, qui s'écrit $p(c, y) = p(y, c)$.

Probabilité marginale

```
cont2 <- cont %>% as.data.frame %>% mutate(marginal_eye = rowSums(cont) )
rownames(cont2) <- row.names(cont)
cont2
```

```
##      Black Brown  Red Blond marginal_eye
## Brown  0.11  0.20 0.04  0.01          0.36
## Blue   0.03  0.14 0.03  0.16          0.36
## Hazel  0.03  0.09 0.02  0.02          0.16
## Green  0.01  0.05 0.02  0.03          0.11
```

On peut aussi s'intéresser à la probabilité d'avoir des yeux bleus, de manière générale. Il s'agit de la probabilité *marginale* de l'événement *yeux bleus*, qui s'obtient par la somme de toutes les probabilités jointes impliquant l'événement *yeux bleus*. Elle s'écrit $p(y) = \sum_c p(y|c)p(c)$.

Probabilité marginale

```
cont3 <- rbind(cont2, colSums(cont2) )  
rownames(cont3) <- c(rownames(cont2), "marginal_hair")  
cont3
```

##	Black	Brown	Red	Blond	marginal_eye
## Brown	0.11	0.20	0.04	0.01	0.36
## Blue	0.03	0.14	0.03	0.16	0.36
## Hazel	0.03	0.09	0.02	0.02	0.16
## Green	0.01	0.05	0.02	0.03	0.11
## marginal_hair	0.18	0.48	0.11	0.22	0.99

On peut bien entendu aussi s'intéresser aux probabilités des couleurs de cheveux, de manière générale. Elle s'écrit $p(c) = \sum_y p(c|y)p(y)$.

Probabilité conditionnelle

On pourrait aussi s'intéresser à la probabilité qu'une personne ait les cheveux blonds, sachant qu'elle a les yeux bleus. Il s'agit d'une probabilité conditionnelle, et s'écrit $p(c|y)$. Cette probabilité conditionnelle peut se ré-écrire : $p(c|y) = \frac{p(c,y)}{p(y)}$.

```
##      Black Brown  Red Blond marginal_eye
## Blue  0.03  0.14 0.03  0.16          0.36
```

Par exemple, quelle est la probabilité d'avoir des yeux bleus lorsqu'on a les cheveux blonds ?

```
cont3["Blue", "Blond"] / cont3["Blue", "marginal_eye"]
```

```
##      Blue
## 0.4444444
```

Probabilité conditionnelle, règle du produit

On remarque dans le cas précédent que $p(\text{blonds}|\text{bleus})$ n'est pas nécessairement égal à $p(\text{bleus}|\text{blonds})$.

Autre exemple : la probabilité de mourir sachant qu'on a été attaqué par un requin n'est pas la même que la probabilité d'avoir été attaqué par un requin, sachant qu'on est mort (*confusion of the inverse*). De la même manière, $p(\text{data}|H_0) \neq p(H_0|\text{data})$.

À partir des axiomes de Kolmogorov (cf. début du cours), et des définitions précédentes des probabilités conjointes, marginales, et conditionnelles, découle la règle du produit :

$$p(a, b) = p(b) \times p(a|b) = p(a) \times p(b|a)$$

Twitframe



Bill Gates ✓
@BillGates · Follow



Why I would rather encounter a shark in the wild than a mosquito: b-gat.es/2XelyuL #MosquitoWeek

Mosquitoes kill more people in one day than sharks kill in a year.

gates
notes

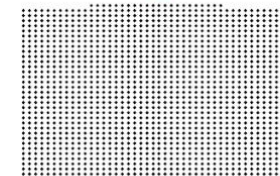


1 YEAR

7 deaths (2017)



1 DAY



1.474 deaths (2017)

Dérivation du théorème de Bayes

$$p(x, y) = p(x|y) \times p(y) = p(y|x) \times p(x)$$

$$p(y|x) \times p(x) = p(x|y) \times p(y)$$

$$p(y|x) = \frac{p(x|y) \times p(y)}{p(x)}$$

$$p(x|y) = \frac{p(y|x) \times p(x)}{p(y)}$$

$$\text{Pr(hypothèse | données)} = \frac{\text{Pr(données | hypothèse)} \times \text{Pr(hypothèse)}}{\text{somme des produits}}$$

Contexte théorique

Rappels : Test d'hypothèse fréquentiste

Null Hypothesis Significance Testing (NHST)

On s'intéresse aux différences de taille entre hommes et femmes. On va mesurer 100 femmes et 100 hommes.

```
men <- rnorm(100, 175, 10) # 100 tailles d'hommes  
women <- rnorm(100, 170, 10) # 100 tailles de femmes
```

```
t.test(x = men, y = women)
```

```
##  
##      Welch Two Sample t-test  
##  
## data:  men and women  
## t = 2.5199, df = 196.16, p-value = 0.01254  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
##  0.754446 6.186651  
## sample estimates:  
## mean of x mean of y  
## 173.4120 169.9414
```

Null Hypothesis Significance Testing (NHST)

On va simuler des t-valeurs issues de données générées sous l'hypothèse d'une absence de différence entre hommes et femmes.

```
nsims <- 1e4 # nombre de simulations
t <- rep(NA, nsims) # initialisation d'un vecteur vide

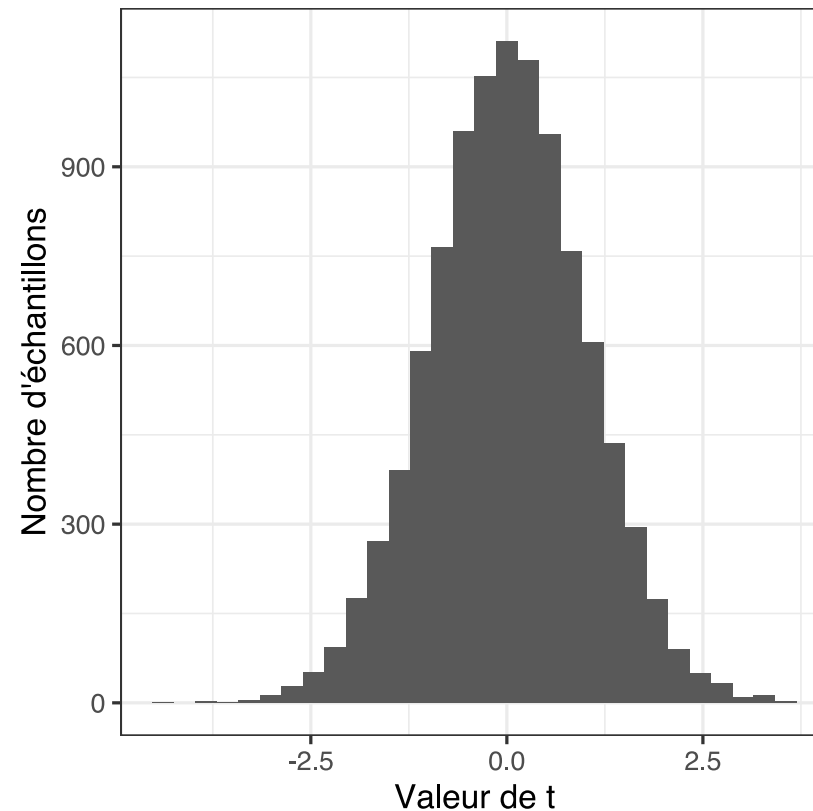
for (i in 1:nsims) {
  men2 <- rnorm(100, 170, 10) # 100 tailles d'hommes
  women2 <- rnorm(100, 170, 10) # 100 tailles de femmes
  t[i] <- t.test(men2, women2)$statistic # on conserve la t-valeur
}
```

Ou sans boucle **for**.

```
t <- replicate(n = nsims, expr = t.test(x = rnorm(100, 170, 10), y = rnorm(100, 170, 10) )$statistic)
```

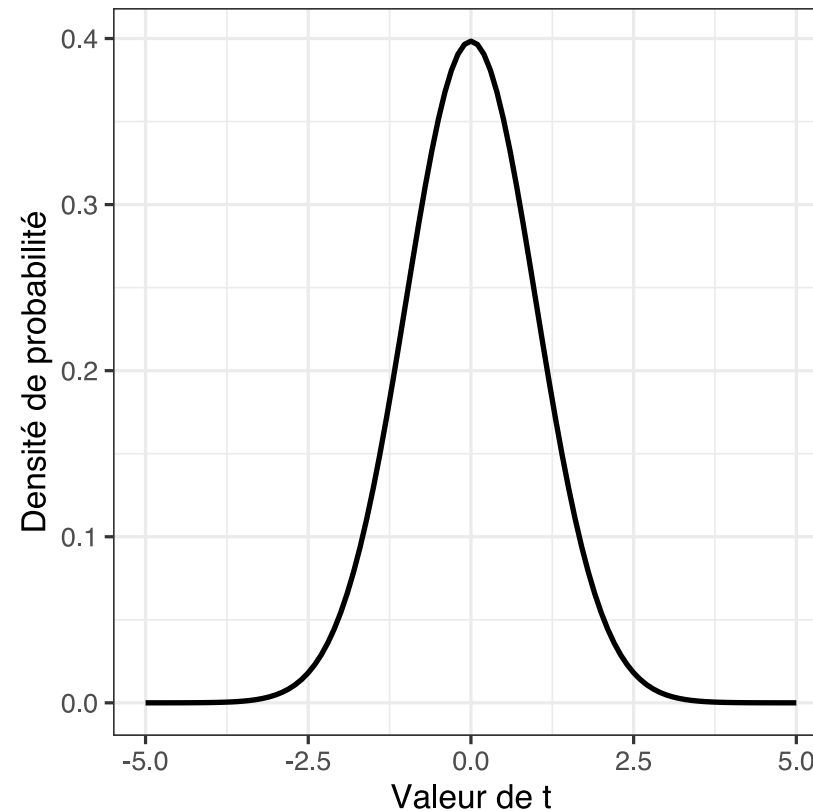
Null Hypothesis Significance Testing (NHST)

```
data.frame(t = t) %>%  
  ggplot(aes(x = t) ) +  
  geom_histogram() +  
  theme_bw(base_size = 20) +  
  labs(x = "Valeur de t", y = "Nombre d'échantillons")
```



Null Hypothesis Significance Testing (NHST)

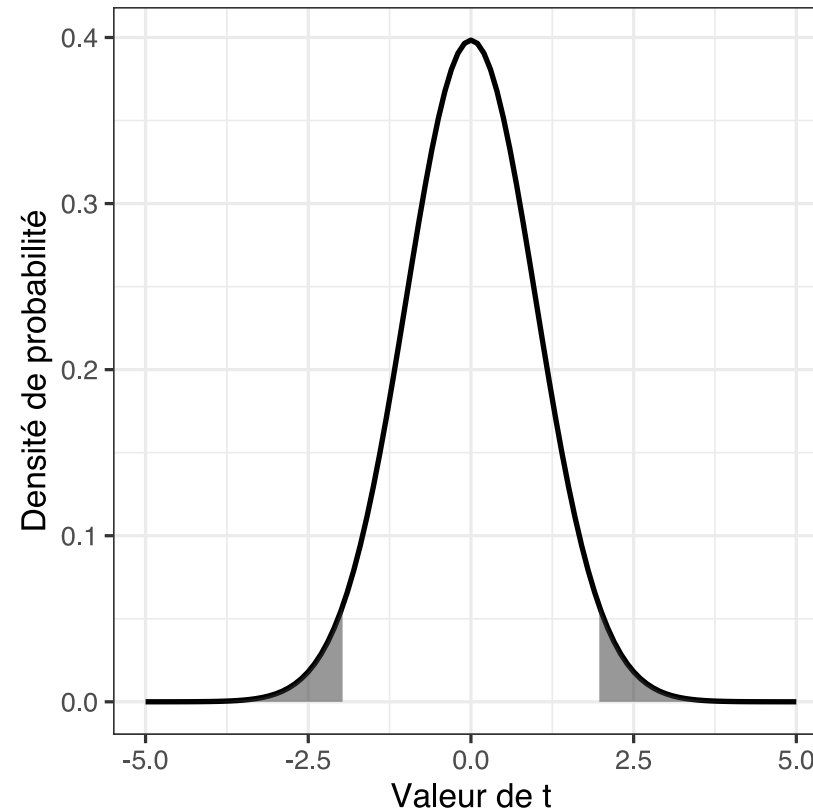
```
data.frame(t = c(-5, 5) ) %>%  
  ggplot(aes(x = t) ) +  
    stat_function(fun = dt, args = list(df = t.test(men, women)$parameter), size = 1.5) +  
    theme_bw(base_size = 20) +  
    labs(x = "Valeur de t", y = "Densité de probabilité")
```



Null Hypothesis Significance Testing (NHST)

```
alpha <- .05  
abs(qt(alpha / 2, df = t.test(x = men, y = women)$parameter) ) # valeur de t critique
```

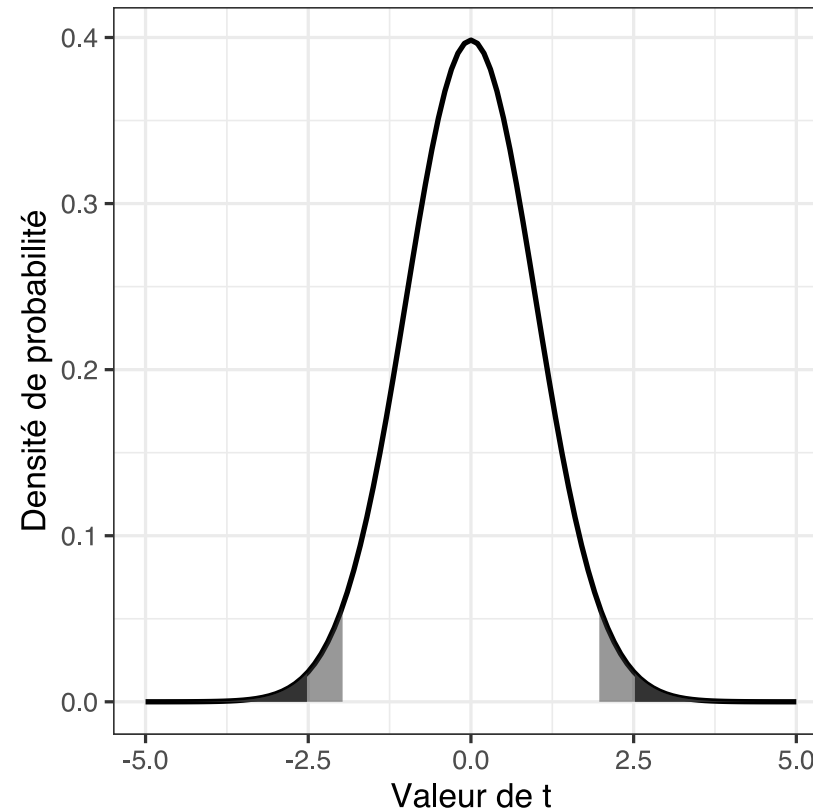
```
## [1] 1.972131
```



Null Hypothesis Significance Testing (NHST)

```
tobs <- t.test(x = men, y = women)$statistic # valeur de t observée  
tobs %>% as.numeric
```

```
## [1] 2.519926
```



P-values

La p-valeur (p-value) est une aire sous la courbe, une intégrale, sous la distribution de statistiques de test sous l'hypothèse nulle (i.e., étant admis que l'hypothèse nulle est vraie). La p-valeur indique la probabilité d'observer la valeur de la statistique de test observée, ou une valeur plus extrême, sous l'hypothèse nulle.

$$p[\mathbf{t}(\mathbf{x}^{\text{rep}} | H_0) \geq t(x)]$$

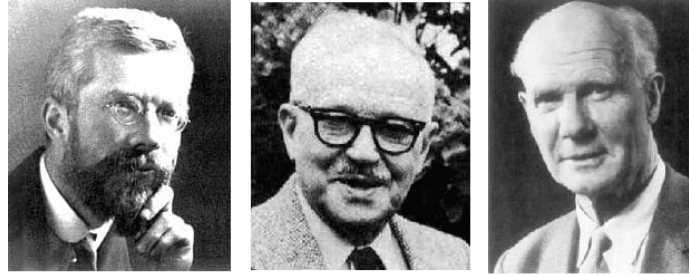
```
t.test(x = men, y = women)$p.value
```

```
## [1] 0.01253526
```

```
tvalue <- abs(t.test(x = men, y = women)$statistic)
df <- t.test(x = men, y = women)$parameter
2 * integrate(f = dt, lower = tvalue, upper = Inf, df = df)$value
```

```
## [1] 0.01253526
```

Fisher versus Neyman & Pearson



Fisher utilisait la p-valeur pour quantifier l'évidence *contre* l'hypothèse nulle. Plus petite est la p-valeur, plus forte est l'évidence contre l'hypothèse nulle. Cette interprétation "continue" de la p-valeur a été très largement critiquée (cf. slide suivante et [Wagenmakers, 2007](#)).

Neyman & Pearson proposèrent d'utiliser la p-valeur de manière à *contrôler* les erreurs de type 1 (faux positifs) et de type 2 (faux négatifs) sur le long terme (in the long run). Avec cette méthode, on n'interprète pas la p-valeur de manière continue, mais on classifie simplement les résultats comme "significatifs" ou "non-significatifs" en fonction d'un seuil pré-défini. Cette procédure a aussi de nombreux problèmes. Un problème majeur est celui du *domain problem* (e.g., [Trafimow & Earp, 2017](#)).

Logique, fréquentisme, et raisonnement probabiliste

Le *modus tollens* est un des raisonnements logiques les plus importants et les plus performants. Dans le cadre de l'inférence statistique, il s'applique parfaitement au cas suivant : "Si H_0 est vraie, alors x ne devrait pas se produire. On observe x . Alors H_0 est fausse".

MAIS nous avons le plus souvent affaire à des hypothèses "continues", probabilistes.

L'inférence fréquentiste (Fishérienne) est elle aussi probabiliste, de la forme "Si H_0 est vraie, alors x est peu probable. On observe x . Alors H_0 est peu probable."

Or cet argument est invalide, le *modus tollens* ne s'applique pas au monde probabiliste (e.g., Pollard & Richardson, 1987; Rouder, Morey, Verhagen, Province, & Wagenmakers, 2016).

Par exemple : *Si un individu est un homme, alors il est peu probable qu'il soit pape. François est pape. François n'est donc certainement pas un homme...*

Contexte théorique

Test d'hypothèse bayésien

Facteur de Bayes

Au lieu de tester une seule hypothèse (i.e., l'hypothèse nulle), on compare deux modèles :

- $H_0 : \mu_1 = \mu_2 \rightarrow \delta = 0$
- $H_1 : \mu_1 \neq \mu_2 \rightarrow \delta \neq 0$

$$\underbrace{\frac{p(H_0|D)}{p(H_1|D)}}_{\text{Posterior model odds}} = \underbrace{\frac{p(D|H_0)}{p(D|H_1)}}_{\text{Bayes factor BF}_{01}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior model odds}}$$

$$\text{evidence} = p(D|H) = \int p(\theta|H)p(D|\theta, H)d\theta$$

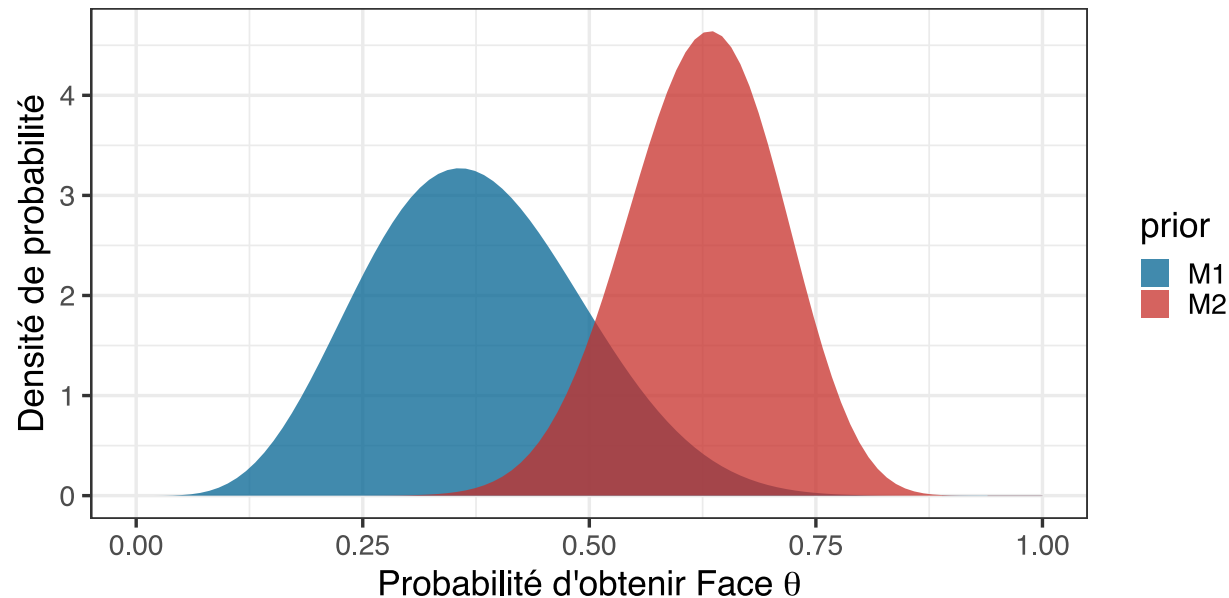
L'évidence en faveur d'un modèle correspond à la *vraisemblance marginale* d'un modèle (le dénominateur du théorème de Bayes), c'est à dire à la vraisemblance moyennée sur le prior... Ce qui fait du facteur de Bayes un ratio de vraisemblances, pondéré par (ou moyenné sur) le prior.

Facteur de Bayes, exemple d'application

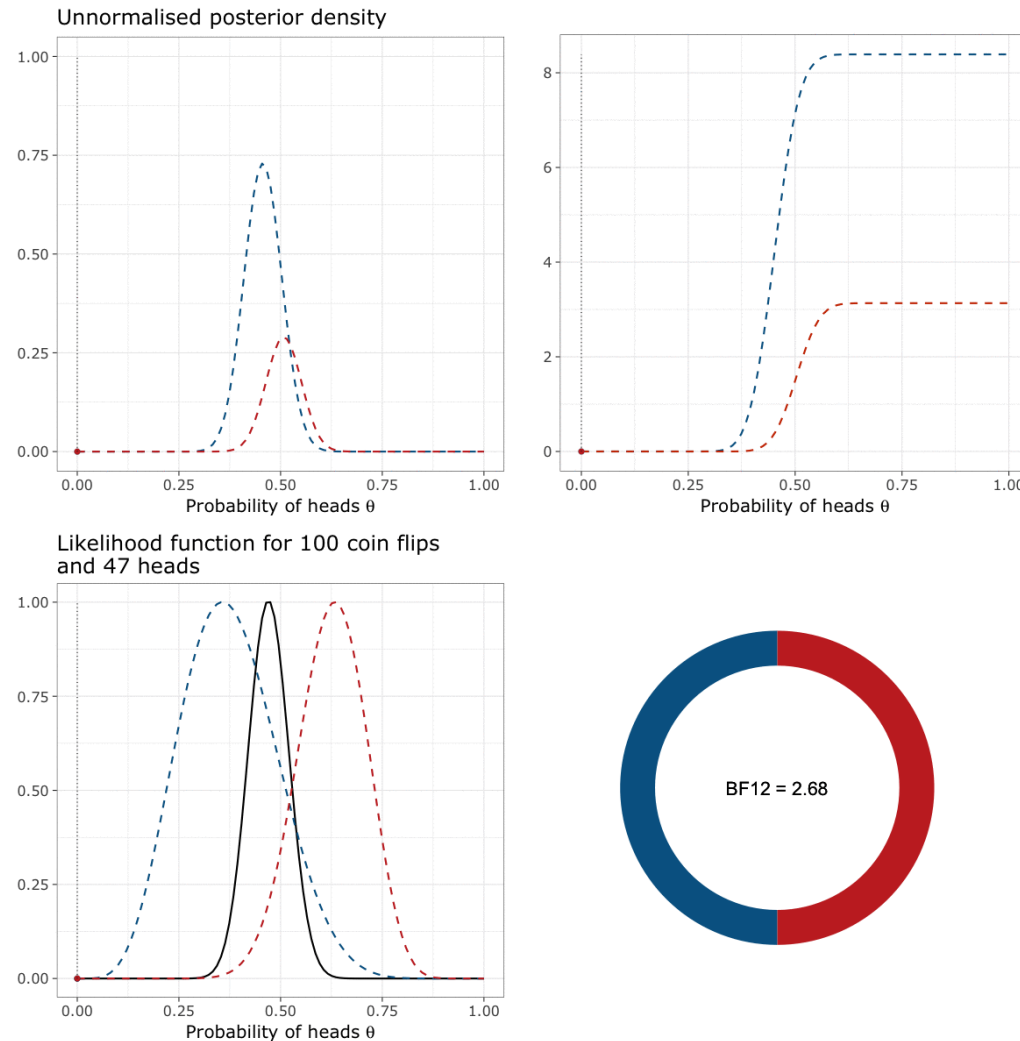
On lance une pièce 100 fois et on essaye d'estimer la probabilité θ (le biais de la pièce) d'obtenir Face. On compare deux modèles qui diffèrent par leur prior sur θ .

$$\begin{aligned}\mathcal{M}_1 : y_i &\sim \text{Binomial}(n, \theta) \\ \theta &\sim \text{Beta}(6, 10)\end{aligned}$$

$$\begin{aligned}\mathcal{M}_2 : y_i &\sim \text{Binomial}(n, \theta) \\ \theta &\sim \text{Beta}(20, 12)\end{aligned}$$



Facteur de Bayes, exemple d'application



Le facteur de Bayes est la nouvelle p-valeur

Attention à ne pas interpréter le BF comme un *posterior odds*...

Le BF est un facteur qui nous indique de combien nos *prior odds* doivent changer, au vue des données. Il ne nous dit pas quelle est l'hypothèse la plus probable, sachant les données (sauf si les *prior odds* sont 1:1).

Considérons un autre exemple :

- H_0 : la précognition n'existe pas !
- H_1 : la précognition existe

On fait une expérience et on calcule un $BF_{10} = 27$. Quelle est la probabilité a posteriori de H_1 ?

$$\underbrace{\frac{p(H_1|D)}{p(H_0|D)}}_{\text{Posterior model odds}} = \underbrace{\frac{27}{1}}_{\text{Bayes factor } BF_{10}} \times \underbrace{\frac{1}{1000}}_{\text{Prior model odds}} = \frac{27}{1000} = 0.027$$

Grille de lecture

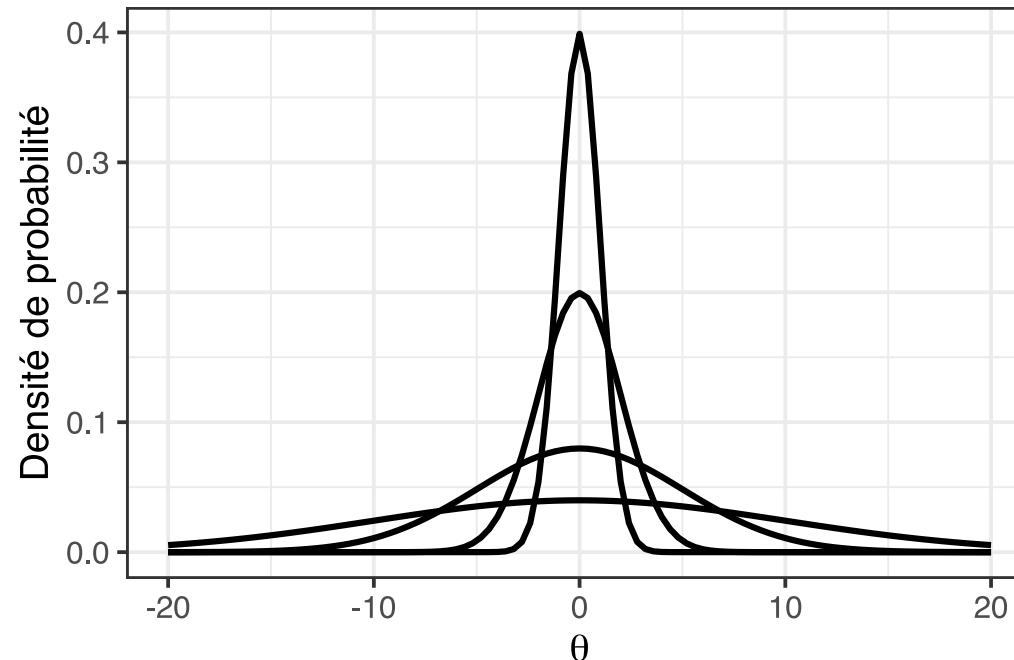
Bien que le BF s'interprète de manière continue, des grilles d'interprétation ont été proposées.

Bayes factor BF_{12}			Interpretation
	>	100	Extreme evidence for M_1
30	-	100	Very Strong evidence for M_1
10	-	30	Strong evidence for M_1
3	-	10	Moderate evidence for M_1
1	-	3	Anecdotal evidence for M_1
	1		No evidence
1/3	-	1	Anecdotal evidence for M_2
1/10	-	1/3	Moderate evidence for M_2
1/30	-	1/10	Strong evidence for M_2
1/100	-	1/30	Very Strong evidence for M_2
	<	1/100	Extreme evidence for M_2

Attention au double sens

Dans le cadre bayésien, le terme de *prior* peut faire référence soit :

- À la probabilité *a priori* ou *a posteriori* d'un modèle (par rapport à un autre modèle), c'est à dire $\Pr(M_i)$. Voir cet [article de blog](#).
- Aux prédictions d'un modèle, par exemple $\theta \sim \text{Normal}(0, 10)$. Voir cet [article de blog](#).



Contexte théorique

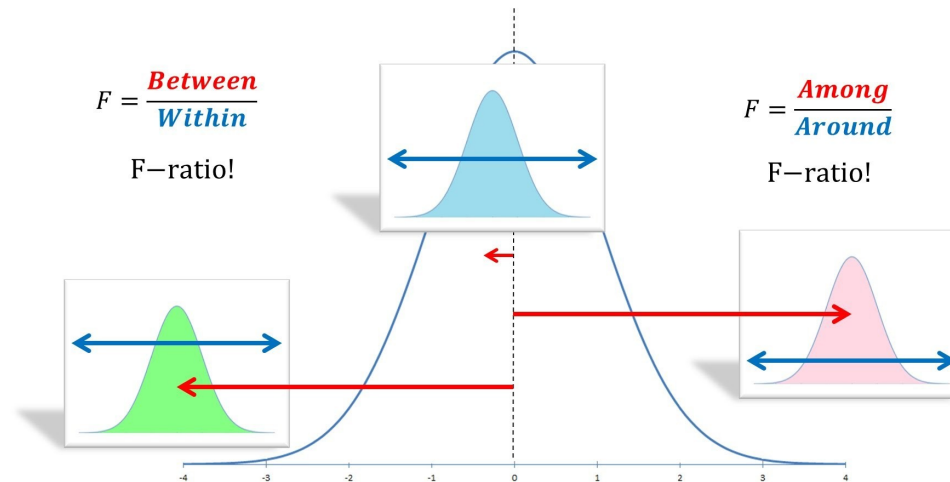
Analyse de la variance (ANOVA) bayésienne

Analyse de la variance (ANOVA) fréquentist

L'ANOVA fréquentiste implique une comparaison de variances. On compare la variance *entre* les différents niveaux d'un prédicteur (variance inter) à la variance *au sein* de ces niveaux (variance intra). Lorsqu'un prédicteur catégoriel n'a pas d'effet, la variance (dans la population) entre les niveaux est égale à la variance (dans la population) au sein des niveaux. Le ratio de ces variances est distribué selon une loi de Fisher-Snedecor. Sous l'hypothèse nulle, on peut calculer (comme nous l'avons fait pour le t-test), la probabilité d'obtenir un ratio de variances au moins aussi large que celui que nous avons observé (i.e., on peut calculer une p-value). Figure tirée de [cet article de blog](#).

ANOVA: Analysis of Variance is a *variability ratio*

Variance Between + Variance Within = Total Variance



Analyse de la variance (ANOVA) bayésienne

L'ANOVA bayésienne compare les capacités de prédiction de différents modèles. Les capacités de prédiction de différents modèles sont évaluées en confrontant leurs prédictions a priori avec les données (cf. exemple précédent sur le biais de la pièce). On va comparer deux modèles et quantifier le support relatif fourni par un jeu de données à ces deux modèles en calculant un facteur de Bayes. On peut comparer n'importe quelle paire de modèles (pas seulement l'hypothèse nulle et l'hypothèse alternative).

$$BF_{01} = \frac{p(D|H_0)}{p(D|H_1)} = \frac{\int p(\theta|H_0) \cdot p(D|\theta, H_0) d\theta}{\int p(\theta|H_1) \cdot p(D|\theta, H_1) d\theta}$$

Approche par comparaison de modèles

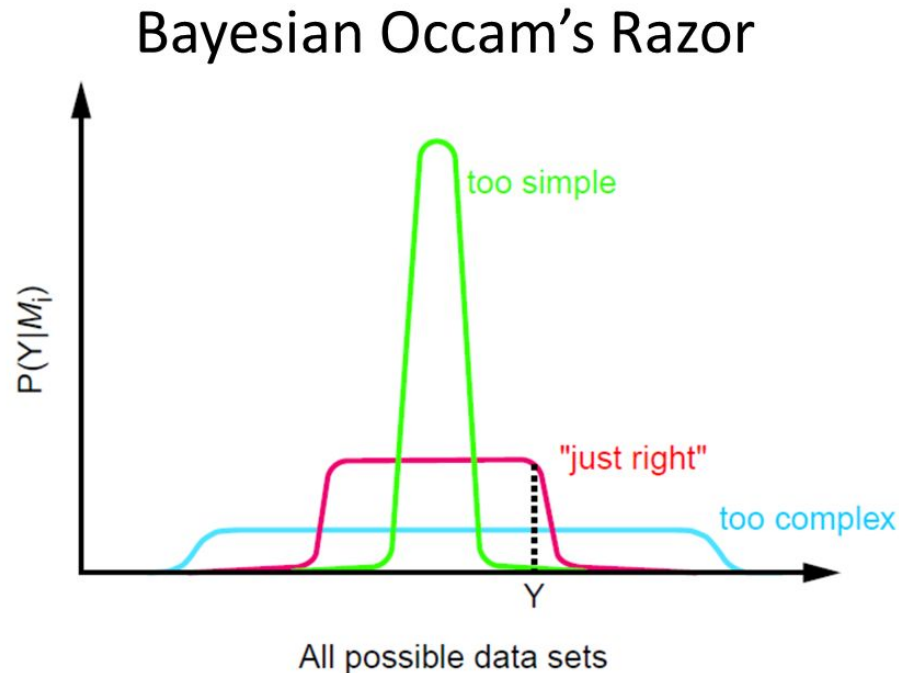
Afin de tester l'effet d'un prédicteur (sur une variable dépendante), on adopte donc une approche par comparaison de modèles. Par exemple, dans une situation avec deux variables indépendantes (VIs), on peut estimer l'effet de la première VI en comparant le modèle complet (e.g., $VD \sim VI1 + VI2 + VI1 * VI2$) à un modèle similaire mais sans l'effet de la première VI (i.e., $VD \sim VI2 + VI1 * VI2$). La comparaison de ces deux modèles nous renseigne sur l'effet (i.e., le "gain prédictif" apporté par $VI1$).

À noter que le facteur de Bayes implémente une sorte de rasoir d'Ockham automatique. En effet, comme le facteur de Bayes pondère la vraisemblance par les prédictions a priori du modèle, les modèles simples (qui font des prédictions plus précises) sont favorisés.

Rasoir d'Ockham automatique

$$\text{BF}_{01} = \frac{p(D|H_0)}{p(D|H_1)} = \frac{\int p(\theta|H_0) \cdot p(D|\theta, H_0) d\theta}{\int p(\theta|H_1) \cdot p(D|\theta, H_1) d\theta}$$

On voit que les modèles complexes, qui prédisent une plus grande variété de données possibles (en abscisse) sont "récompensés" par moins d'évidence (en ordonnée).



Travaux pratiques avec jamovi

Example d'application

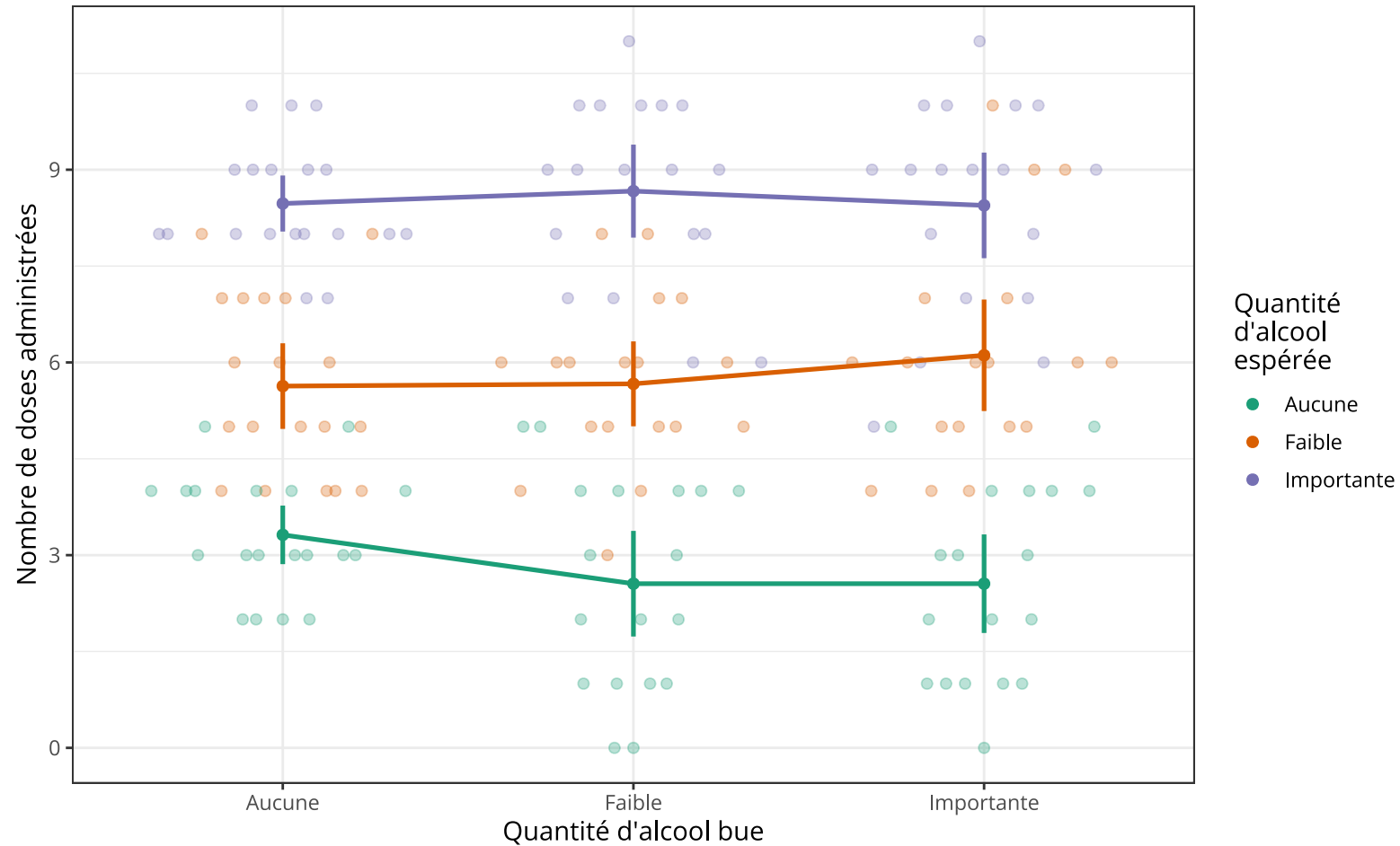
On a mesuré le nombre de doses d'épices administrées (mesure indirecte de l'agressivité) en fonction de la quantité d'alcool espérée **QAE** (aucune, faible, importante) et de la quantité d'alcool bue **QAB** (aucune, faible, importante).

```
df <- read.csv("data/alcool_data.csv", sep = ";")
str(df)
```

```
## 'data.frame':    165 obs. of  5 variables:
## $ sujet: int  1 2 3 4 5 6 7 8 9 10 ...
## $ Genre: chr  "H" "F" "H" "F" ...
## $ QAB  : int  0 0 0 1 1 1 2 2 2 0 ...
## $ QAE  : int  0 1 2 0 1 2 0 1 2 0 ...
## $ QE   : int  3 6 9 5 6 8 4 4 9 4 ...
```

Exploration graphique des données

Effet de la quantité d'alcool bue et espérée sur le nombre de doses d'épices administrées



ANOVA bayésienne

Model Comparison - QE

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Null model	0.200	2.25e-44	9.01e-44	1.0000	
QAB	0.200	1.44e-45	5.76e-45	0.0640	0.0240
QAE	0.200	0.9187	45.2290	4.08e+43	2.11e-6
QAB + QAE	0.200	0.0686	0.2946	3.05e+42	3.3799
QAB + QAE + QAB * QAE	0.200	0.0127	0.0513	5.62e+41	1.4928

On compare tous les modèles possibles incluant ces deux prédicteurs et leur interaction. La colonne **P(M)** indique la probabilité a priori de chaque modèle, par défaut fixée à 1 divisé par le nombre de modèles. La colonne **P(M | data)** indique la probabilité a posteriori de chaque modèle (sachant les données et les priors de chaque modèle).

ANOVA bayésienne

Model Comparison - QE

Models	P(M)	P(M data)	BF_M	BF_{10}	error %
Null model	0.200	2.25e-44	9.01e-44	1.0000	
QAB	0.200	1.44e-45	5.76e-45	0.0640	0.0240
QAE	0.200	0.9187	45.2290	4.08e+43	2.11e-6
QAB + QAE	0.200	0.0686	0.2946	3.05e+42	3.3799
QAB + QAE + QAB * QAE	0.200	0.0127	0.0513	5.62e+41	1.4928

La colonne BF_M indique le changement des prior odds vers les posterior odds pour chaque modèle. Par exemple, pour le meilleur modèle avec seulement l'effet de **QAE**, $BF_M = 0.9187 / (1 - 0.9187) \times (1 - 0.2) / 0.2 \approx 45.2$. La colonne BF_{10} indique le BF en faveur du modèle (vs. le modèle nul). Enfin, la dernière colonne rapporte l'erreur dans l'approximation de la valeur du BF. En général, et selon la valeur du BF, un taux d'erreur jusqu'à 20% est acceptable. Notons également que nous pouvons toujours passer de BF_{01} à BF_{10} en prenant la réciproque du premier, $BF_{10} = 1 / BF_{01}$.

ANOVA bayésienne

Model Comparison - QE

Models	P(M)	P(M data)	BF _M	BF ₁₀	error %
Null model	0.200	2.25e-44	9.01e-44	1.0000	
QAB	0.200	1.44e-45	5.76e-45	0.0640	0.0240
QAE	0.200	0.9187	45.2290	4.08e+43	2.11e-6
QAB + QAE	0.200	0.0686	0.2946	3.05e+42	3.3799
QAB + QAE + QAB * QAE	0.200	0.0127	0.0513	5.62e+41	1.4928

Les facteurs de Bayes sont *transitifs*, ce qui veut dire que si le modèle avec seulement **QAE** surpasse le modèle nul d'un facteur a (i.e., si $\text{BF}_{10} = a$) et que le modèle avec seulement **QAB** surpasse le modèle avec seulement **QAE** d'un facteur b , alors le modèle avec seulement **QAB** surpasse le modèle nul d'un facteur $a \times b$. Cette propriété de transitivité nous permet de calculer le facteur de Bayes pour des comparaisons non présentées dans la table. Par exemple, pour évaluer le BF associé à l'effet d'interaction, nous pouvons diviser le BF_{10} du modèle complet par le BF_{10} du même modèle sans l'effet d'interaction, ce qui nous donne $3.05e + 42 / 5.62e + 41 \approx 5.43$ (en faveur du modèle sans l'effet d'interaction).

Analyses post-hoc

Post Hoc Comparisons - QAE

		Prior Odds	Posterior Odds	BF _{10, U}	error %
0	1	0.587	2.17e+15	3.69e+15	4.06e-19
	2	0.587	1.23e+37	2.10e+37	9.77e-42
1	2	0.587	8.55e+13	1.46e+14	8.71e-18

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, $r = 1/\sqrt{2}$) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Une analyse approfondie, utilisant l'équivalent bayésien du t-test, nous révèle que toutes les conditions (tous les niveaux de **QAE**) considérées par paire diffèrent l'une de l'autre.

Rapporter les résultats

Methods

We conducted a Bayesian two-way ANOVA with default priors (Rouder et al., 2017) using a medium scale for the alternative hypothesis with $r = 0.5$. All analyses were conducted in jamovi (The jamovi project, 2021).

Results

This analysis revealed that the best model was the model including only an effect of the perceived quantity of alcohol, $\text{BF}_{10} = 4.08e + 43$. However, we found no effect of the actual consumption of alcohol on the number of doses of spices administered by the participant, with $\text{BF}_{10} = 0.06$, indicating that the data were $1 / 0.06 \approx 16.67$ times more likely under the null hypothesis than under the alternative hypothesis.

References

The jamovi project (2021). jamovi. (Version 1.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356-374.

Pour aller plus loin

Le matériel de la formation doctorale (20h) que j'enseigne : "Introduction à la modélisation statistique bayésienne en R" : <https://www.barelysignificant.com/IMSB2022>.

- McElreath, R. (2016, 2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.
- Kurz, S. (2019). *Statistical Rethinking with brms, ggplot2, and the tidyverse*. Available [online](#).
- Noël, Y. (2015). *Psychologie statistique avec R*. Coll. PratiqueR, Paris: EDP Sciences.
- Lambert, B. (2018). *A Student's Guide to Bayesian Statistics*. SAGE Publications Ltd.
- Kruschke, J. K. (2015). *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*. Academic Press / Elsevier.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B. (2013). *Bayesian Data Analysis, third edition*. London: CRC Press.

Take-home messages

- **Inférence bayésienne** : Un cadre général de l'estimation de paramètres reposant sur l'utilisation du théorème de Bayes. L'inférence bayésienne peut être vue comme une extension de la logique, où les conclusions (distribution a posteriori) dérivent des prémisses (distribution a priori) et de l'information contenue dans les données (fonction de vraisemblance).
- **Facteur de Bayes (*Bayes factor*)** : Mesure du support relatif fourni par un jeu de données pour deux hypothèses alternatives. Si on considère deux hypothèses \mathcal{H}_1 et \mathcal{H}_2 et certaines données x , alors le facteur de Bayes en faveur de \mathcal{H}_1 est $\mathbf{BF}_{12} = p(x \mid \mathcal{H}_1) / p(x \mid \mathcal{H}_2)$. Ce dernier peut-être interprété comme un "facteur de mise à jour" du rapport des chances a priori (i.e., avant de prendre connaissance des données) vers le rapport des chances a posteriori (i.e., après avoir pris connaissance des données).

 [Inalborczyk](#)  [Inalborczyk](#)  <https://osf.io/ba8xt>  www.barelysignificant.com