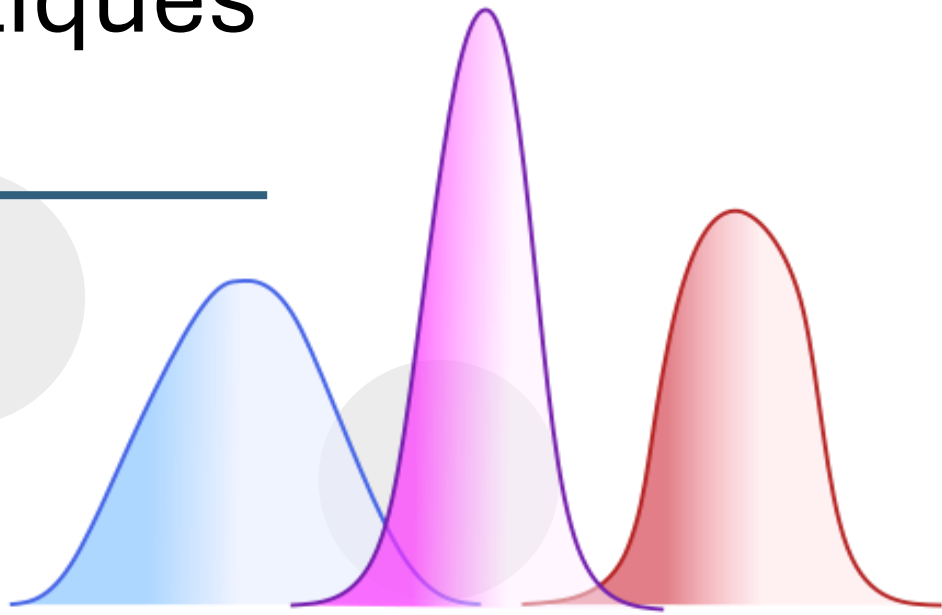
$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$\frac{\int_a^b \binom{n+m}{m} x^m (1-x)^n dx}{\int_0^1 \binom{n+m}{m} x^m (1-x)^n dx}$$

Introduction aux statistiques bayésiennes

Romain di Stasi, PhD

$$\frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{n} P(B|A_j) P(A_j)}$$

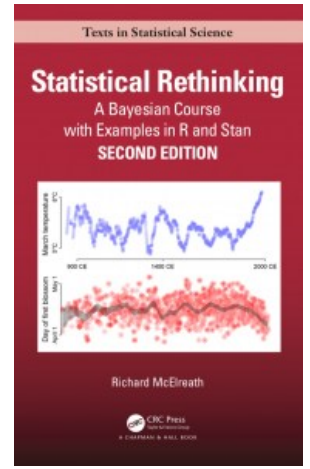


Belgian Association for Psychological Sciences

UNIVERSITÉ LIBRE DE BRUXELLES

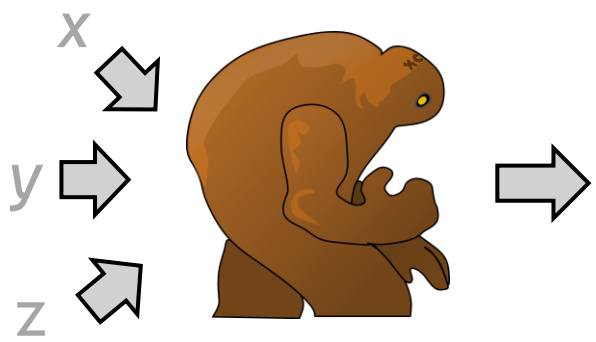
CENTER FOR RESEARCH IN COGNITION & NEUROSCIENCES /UNESCOG

Rappel qu'est-ce qu'un modèle



McElreath (2015)

Paramètres

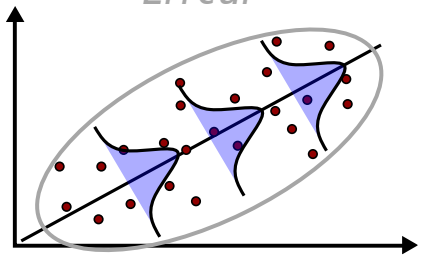


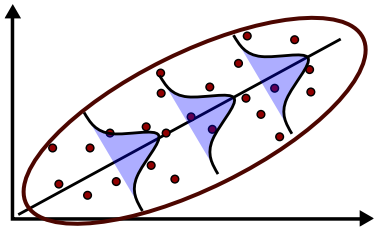
Golem de Prague

Il l'exécute.

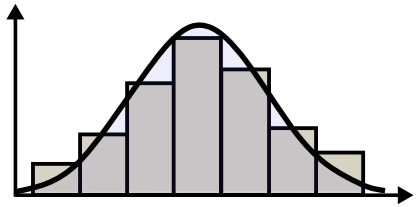
Un modèle ≠ *Réalité*

Erreur

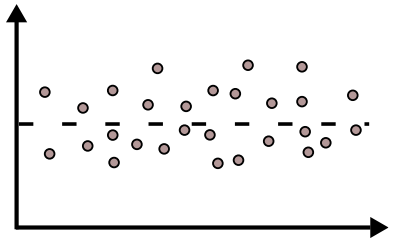




Résidu



Normal

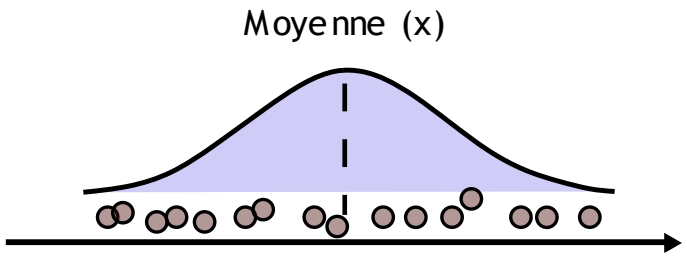


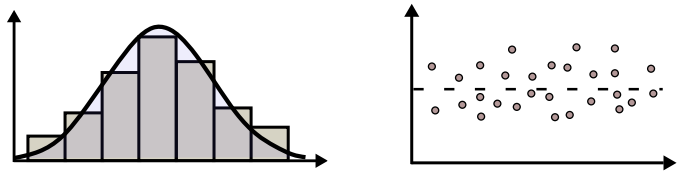
Homoscédasticité

Mais ces modèles ne sont pas toujours adaptées.

Rappel qu'est-ce qu'un modèle

Paramétrique



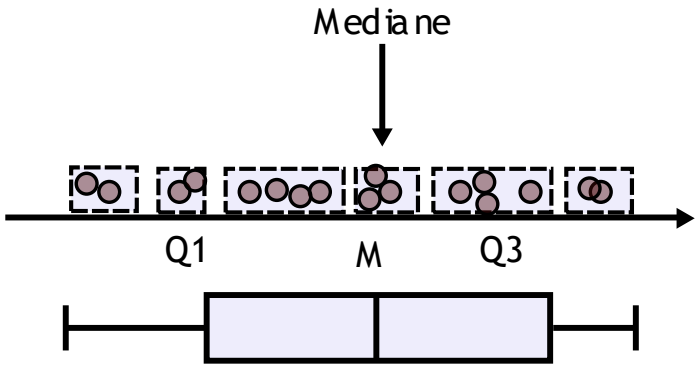


Normal

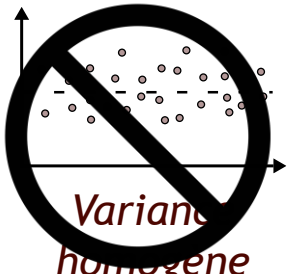
Variance homogène

(*e.g.*, t-test, ANOVA)

Non paramétrique



Normal

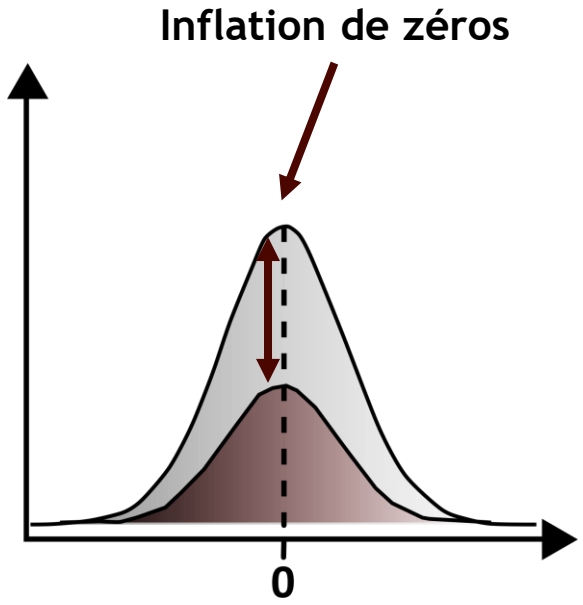


Variance homogène

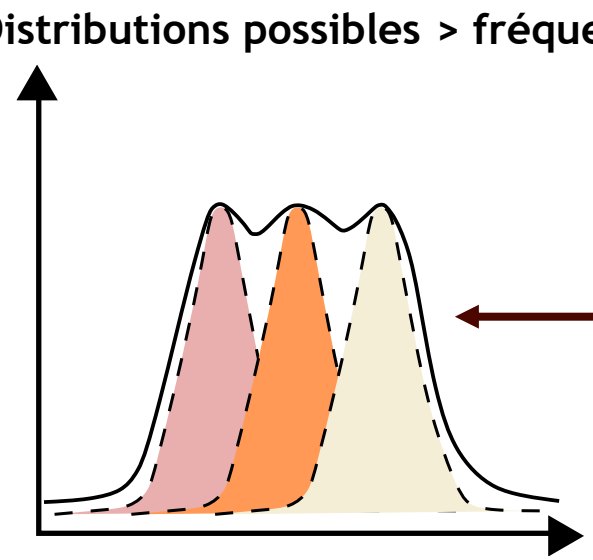
(*e.g., Mann-Whitney U test, Wilcoxon signed-rank test*)

Rappel qu'est-ce qu'un modèle

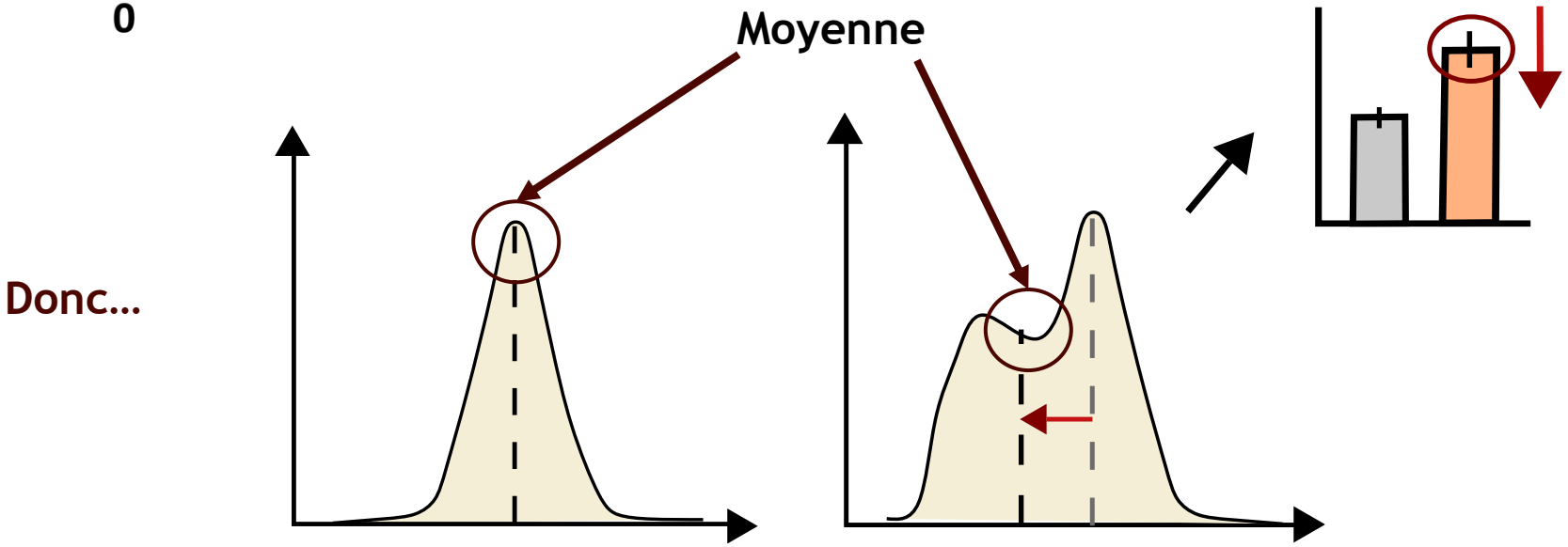
Une distribution mal spécifiée peut biaiser même les statistiques descriptives



&

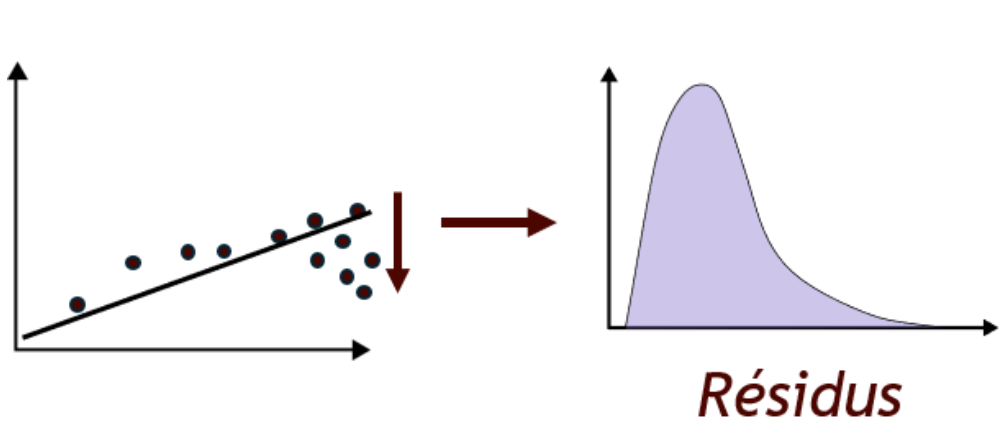


(Bürkner, 2017a, 2017b, 2020; Kruschke, 2015)



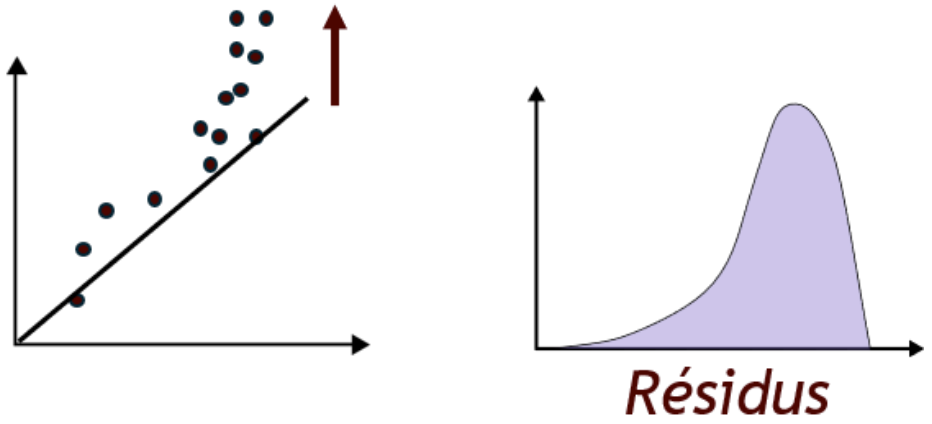
Rappel qu'est-ce qu'un modèle

Une distribution mal spécifiée peut biaiser même les statistiques descriptives



Résidus

ou



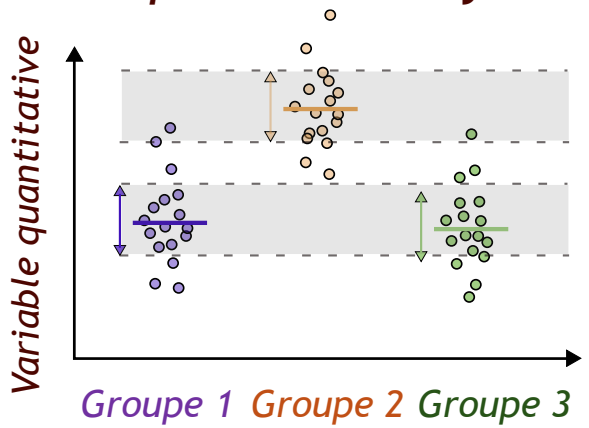
Résidus

Site GitHub

Rappel qu'est-ce qu'un modèle

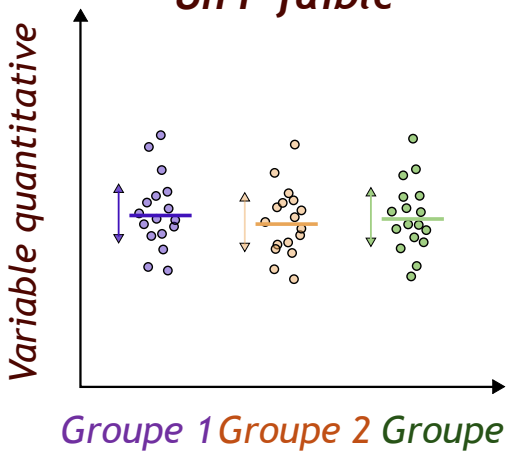
Cela peut biaiser l'analyse de variance

Comparaison de moyenne

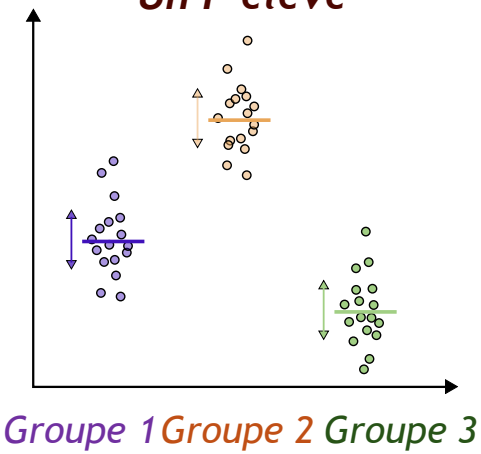


L'ANOVA est le rapport $\frac{Variance\ inter-groupe}{Variance\ intra-groupe}$

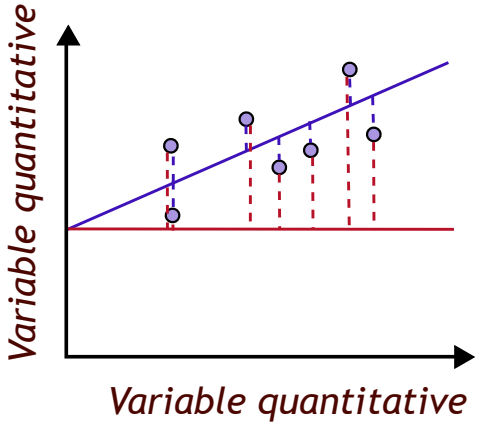
Un F faible



Un F élevé

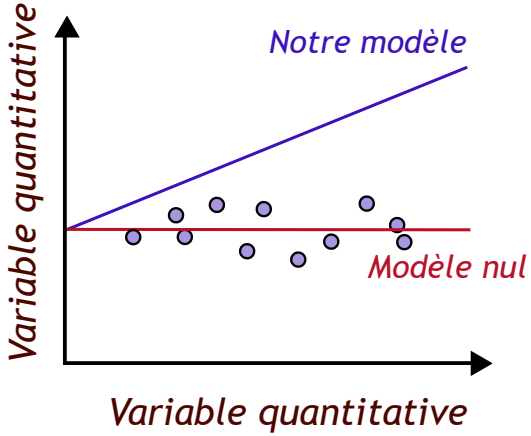


Corrélation

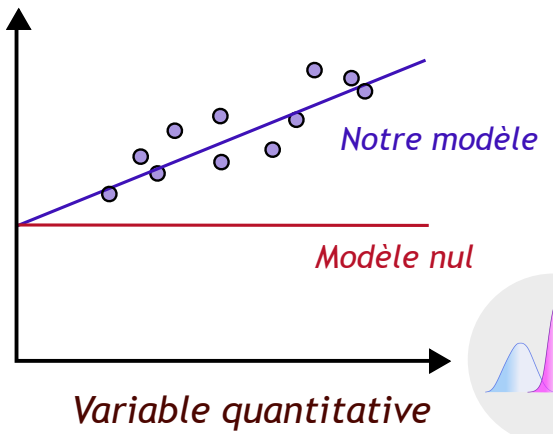


L'ANOVA est le rapport $\frac{Variance\ \beta_1 \neq 0}{Variance\ restante}$

Un F faible



Un F élevé



Site GitHub

Comprendre les statistiques bayésiennes

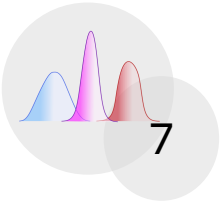
- C’est d’abord comprendre…

  → Le **théorème de Bayes**

  → La **loi binomiale**

  → Une **vraisemblance** et le **maximum de vraisemblance**

Mais pas de panique, nous allons revenir sur ces aspects ici.

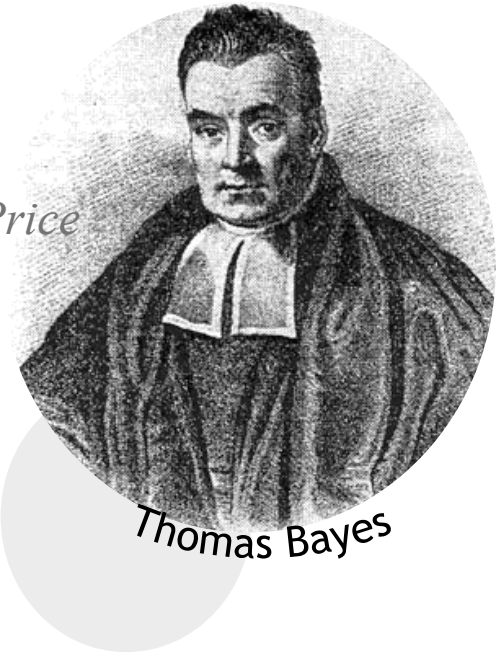


Le théorème de Bayes

- Contrairement aux statistiques fréquentistes, les statistiques bayésiennes ne suggèrent pas que la recherche part de zéro.

→ Tout part du mathématicien Britannique Thomas Bayes (1702-1761)

publié à titre posthume par Richard Price



Thomas Bayes

Probabilité de l'hypothèse A sachant les données B

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

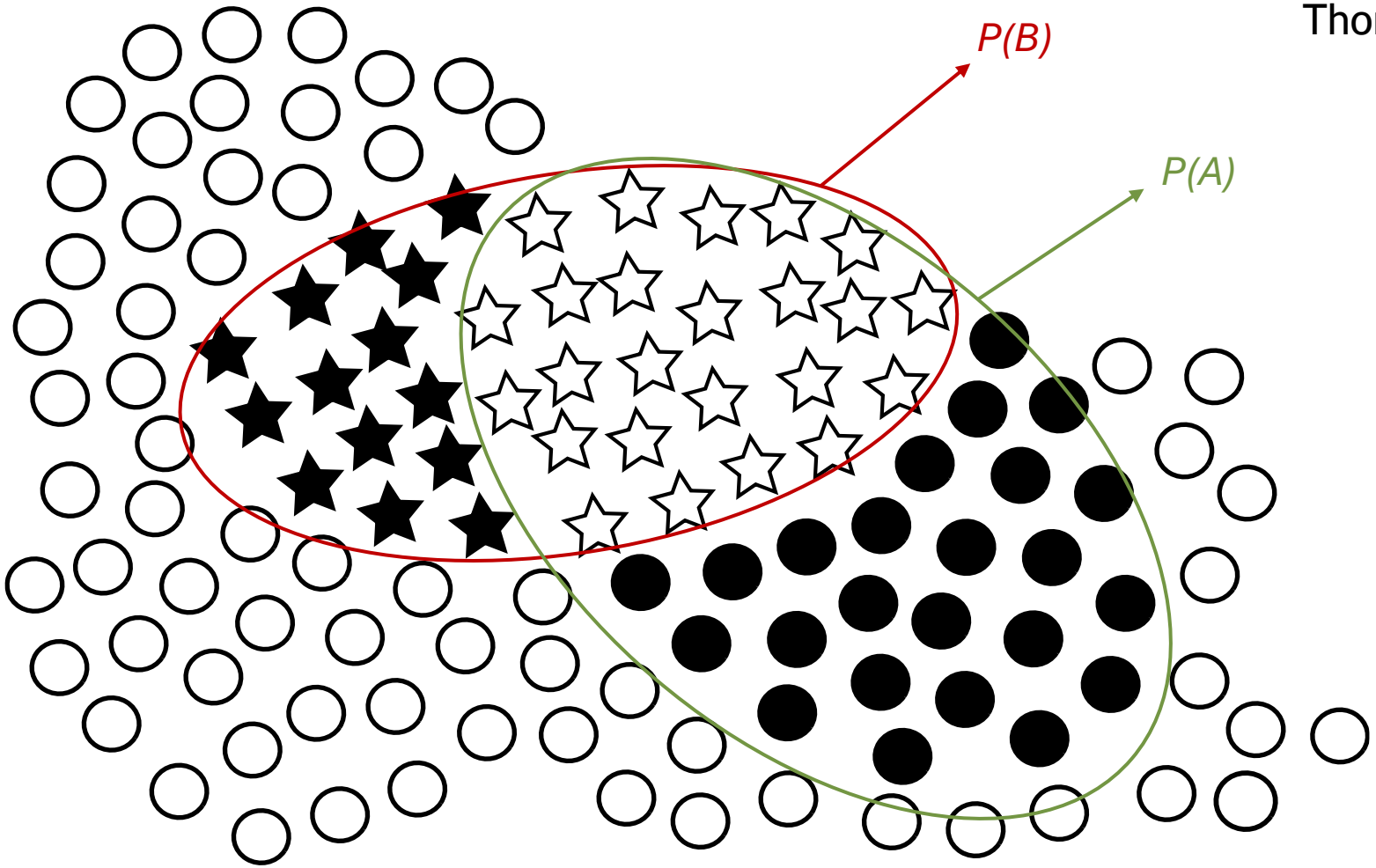
Probabilité de l'hypothèse A sachant les données B

Probabilité de l'hypothèse avant d'avoir vu les données → **probabilité a priori**.

Probabilité d'observer les données (toutes hypothèses confondues) → **évidence**.

(Kruschke 2015, p. 99-118)

Le théorème de Bayes – une brève histoire



Thomas Bayes montre ceci…

$$\mathrm{P}\,(A \mid B) = P(A).P(B|A)$$

Cela revient au même que faire l'inverse :

$$\mathrm{P}\,(B \mid A) = P(B).P(A|B)$$

Donc on obtient l'égalité suivante :

$$P(A).P(B|A) = P(B).P(A|B)$$

Le problème, c'est qu'écrit ainsi, on ne peut pas calculer la probabilité d'un événement en connaissant sa cause (l'autre probabilité). En effet, si l'on considère que B est la cause d'une partie de A, alors pour pouvoir estimer la probabilité de A étant donné B, il est indispensable de connaître la probabilité de B.

(Kruschke 2015, p. 99-118)

Le théorème de Bayes – une brève histoire

Thomas Bayes lui-même ne s'y est pas intéressé. C'est son ami Richard Price qui a publié ses travaux après sa mort.

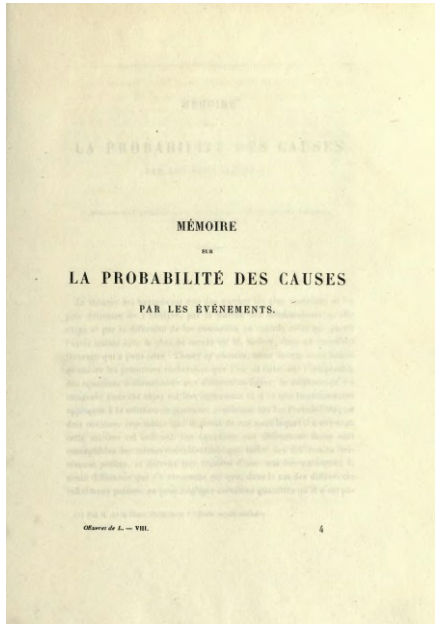
Puis Pierre-Simon Laplace a découvert l'intérêt de l'égalité mise en évidence par Thomas Bayes qu'il a reformulée ainsi:

Pierre-Simon Laplace

$$P(A)\,P(B|A) = P(B).P(A|B)$$

$$P(B|A) = \frac{P(B).P(A|B)}{P(A)}$$

Appelé probabilité inverse



MÉMOIRE

SUR

LA PROBABILITÉ DES CAUSES

PAR LES ÉVÉNEMENTS.

Pourquoi cela change toute notre approche ? Ici il faut revenir aux « vraisemblance » et à la manière dont les statistiques fréquentistes fonctionnent pour les différencier de l'approche bayésienne.

(Kruschke 2015, p. 99-118)

Le théorème de Bayes – une brève histoire

Fréquentiste

Vraisemblance des résultats x
dans le cadre d'une hypothèse H

$$P(x|Ho)$$

Quelle est la probabilité que j'aie un résultat x dans le cas où j'ai une hypothèse H. C'est « la vraisemblance du résultat ».

C'est tout ou rien, soit je conserve, soit je rejette Ho.

Bayésiens

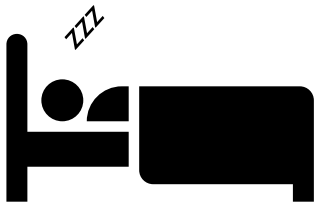
Plausibilité de l'hypothèse H
au vu des résultats x

$$P(Ho|x)$$

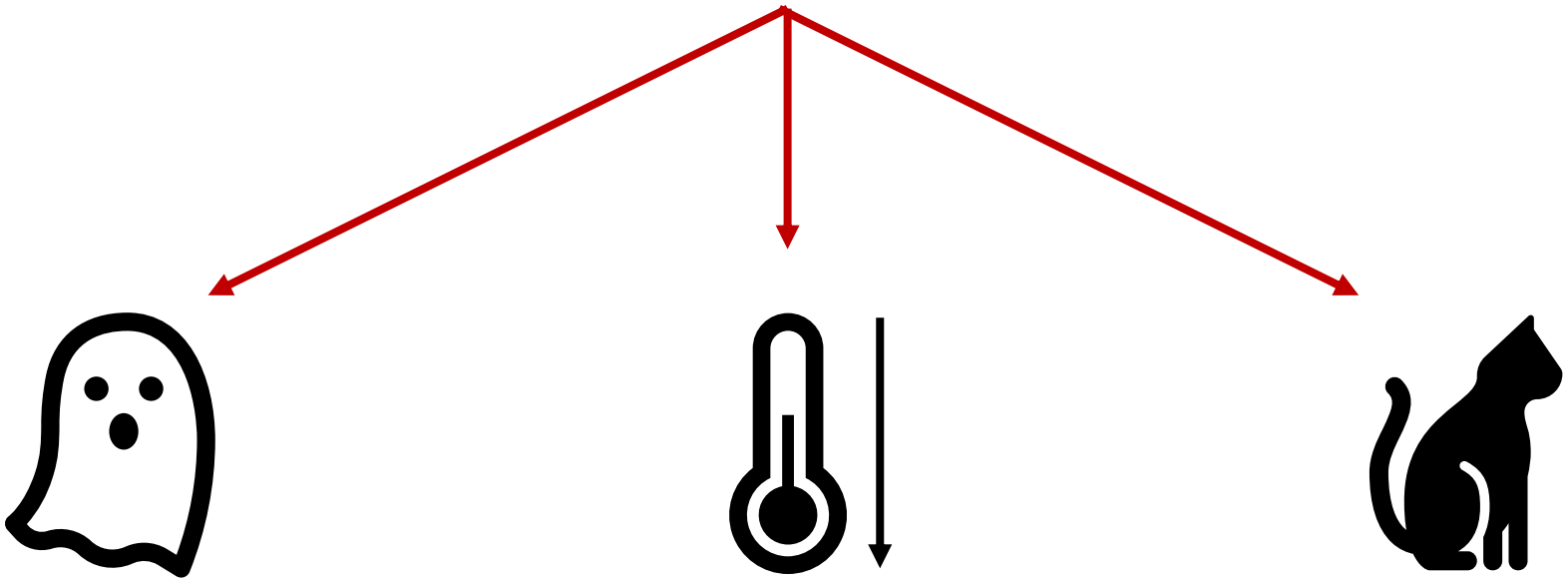
Quelle est la probabilité que mon hypothèse H soit vraie au vu des résultats obtenus x, c'est la plausibilité de mes hypothèses.

Elles permettent d'évaluer les niveaux de crédibilités de chaque hypothèse.

Le théorème de Bayes – une brève histoire



J’entends un bruit de grincement de parquet, c’est à cause de



H1: 2 %

H2: 32 %

H3 :66 %

Contrairement aux fréquentistes, les bayésiens n’éliminent pas les hypothèses : ils les évaluent.

Le théorème de Bayes – une brève histoire

Vraisemblance des résultats x
dans le cadre d'une hypothèse H

$$P(x|Ho)$$

La méthode fréquentiste

Soit H0

Si $P(x|Ho) < 1\ \%$

Alors on rejette H0, l'hypothèse doit être fausse

Si $P(x|Ho) \geq 1\ \%$

Alors on doit retenir H0

Ronald Fisher

Ronald Fisher a rejeté le concept de probabilité inverse pour deux raisons:

(1) Trop **complexe à calculer**

(2) Trop **subjectif**

Le théorème de Bayes – une brève histoire

***Ronald Fisher** a* rejeté le concept de probabilité inverse pour deux raisons:

(1) Trop **complexe à calculer**

$$P(H0\,|x) = \frac{P(x|H0).P(H0)}{P(x|H0).P(H0) + P(x|H1).P(H1)}$$

$$P(H1\,|x) = \frac{P(x|H1).P(H1)}{P(x|H0).P(H0) + P(x|H1).P(H1)}$$

Il faut calculer les vraisemblance de chaque hypothèse…

≠

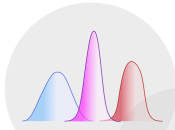
Fréquentiste

(2) Trop **subjective**

$$P(H1\,|x) = \frac{P(x|H1).P(H1)}{P(x)}$$

Prior

Il faut avoir un a priori sur l'hypothèse (prédictions) avant même de l'avoir testée.



Le théorème de Bayes – une brève histoire

***Ronald Fisher** a* rejeté le concept de probabilité inverse pour deux raisons:

(1) Trop **complexe à calculer**

Cette difficulté est avant tout **mathématique**, mais **le concept reste simple**.

Cela n'est plus un problème, maintenant que des ordinateurs peuvent réaliser ces calculs en quelques millionièmes de seconde.

(2) Trop **subjectif**

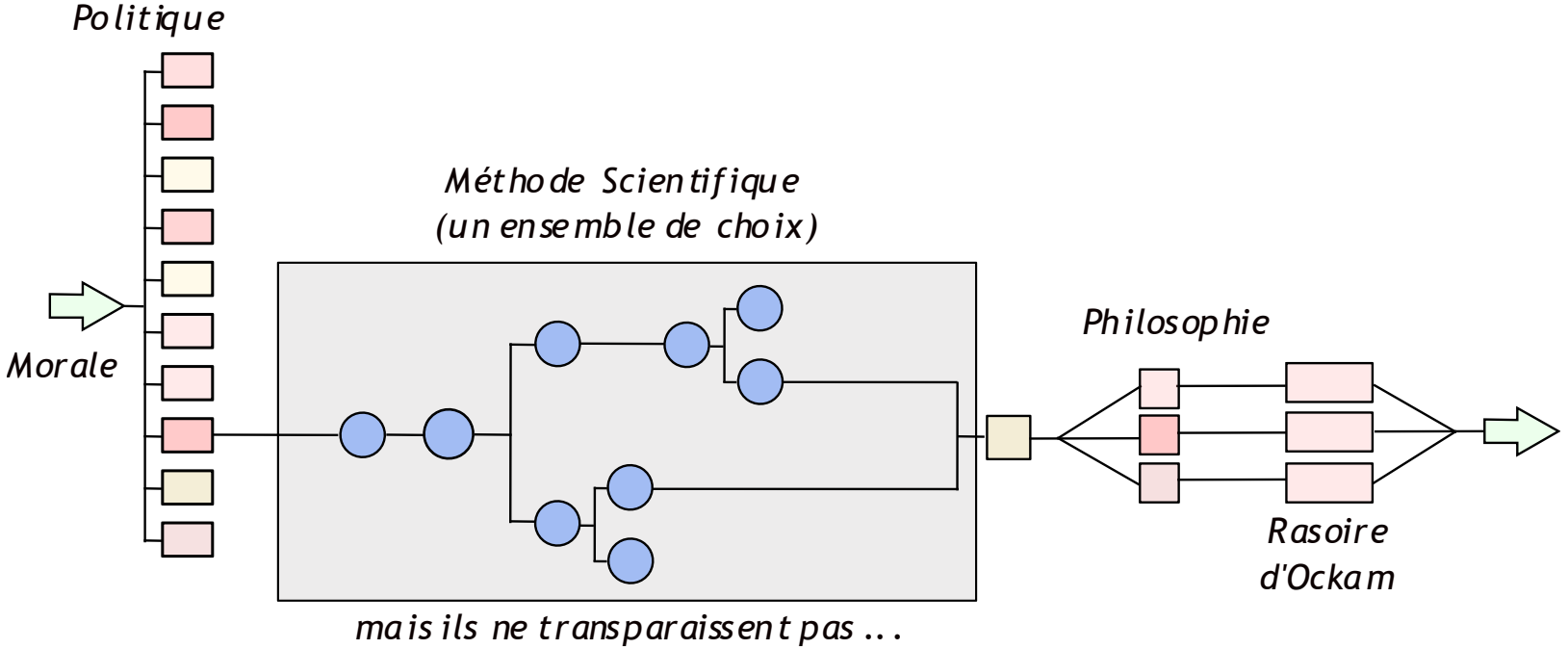
Les statistiques fréquentistes ne le sont pas moins puisqu'elles reposent sur des choix.

Le théorème de Bayes – une brève histoire

***Ronald Fisher** a* rejeté le concept de probabilité inverse pour deux raisons:

(2) Trop **subjectif**

Les statistiques fréquentistes ne sont pas moins subjectives, car elles reposent aussi sur des choix.



En fréquentiste, cette subjectivité est juste masquée artificiellement !

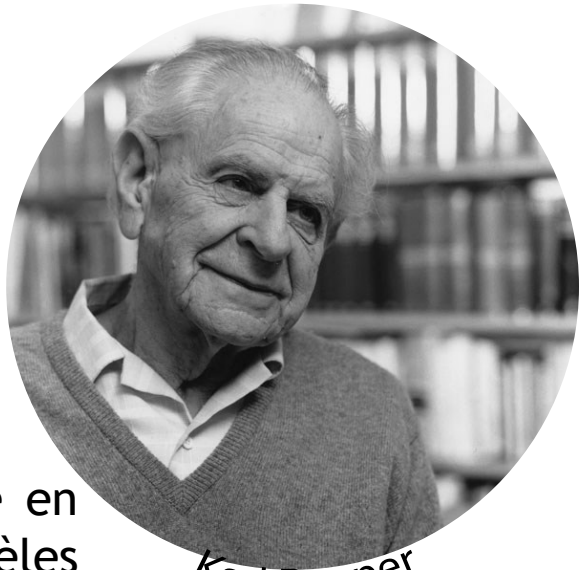
Les **statistiques bayésiennes** ont le mérite d'assumer cette subjectivité !

Le théorème de Bayes – une brève histoire

***Ronald Fisher** a* rejeté le concept de probabilité inverse pour deux raisons:

(2) Trop **subjectif**

> Mais… les statistiques bayésiennes sont parmi les plus proches de l'approche hypothético-déductive, puisqu'elles traduisent directement ce qu'on appelle la **falsification de Popper**.


Karl Popper

En fréquentiste on part de $H0$ est synonyme de pas d'effet
et on ne peut jamais l'accepter.

En Bayésien, on ne se contente pas de tester une hypothèse nulle H_0 comme en fréquentiste, mais on compare explicitement plusieurs hypothèses ou modèles concurrents (H_0 et H_1) à l'aide du facteur bayésien (BF). Cette logique rend plus visible l'idée poppérienne de falsification, puisque chaque hypothèse est réellement mise à l'épreuve par rapport à une autre. Il est donc tout à fait possible de conclure qu'il y a davantage de soutien pour H_0, plutôt que de la rejeter systématiquement.

(McElreath 2015, p. 5)

Le théorème de Bayes – une brève histoire

Pour rappel, la **falsification selon Popper** repose sur l'idée que, pour qu'une discipline soit considérée comme scientifique, ses hypothèses doivent être « falsifiables ». Il doit être possible de vérifier empiriquement si une hypothèse peut être réfutée ou non.

Il est impossible de montrer empiriquement que Dieu existe ou n'existe pas,
c'est pourquoi la religion n'est pas une science.

L'exemple des moutons

~~Hypothèse 1 – tous les moutons sont blanc ou noir~~

Hypothèse 2 – presque tous les moutons sont blancs ou noirs mais il en existe des oranges

Le théorème de Bayes – une brève histoire

Pour reprendre l'exemple des moutons du cours précédent.

~~Hypothèse 1 – tous les moutons sont blanc ou noir~~

Hypothèse 2 – presque tous les moutons sont blancs ou noirs mais il en existe des oranges

Les statistiques Bayésiennes c'est exactement ça. Une évolution des connaissances.

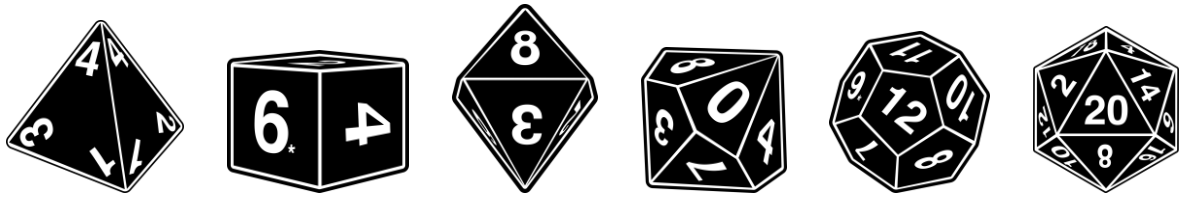
~~**Croyance A priori (Prior 1)** – tous les moutons sont blanc ou noir.~~

Réfuté

Croyance A posteriori (Posterior) – presque tous les moutons sont blancs ou noirs mais il en existe des oranges.

Qui deviendra le Prior de l'étude suivante.

Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.



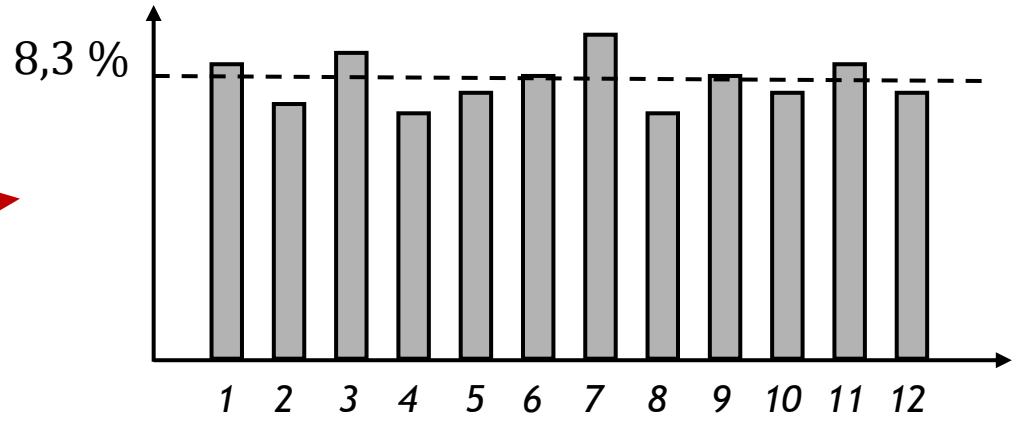
D4 *D6* *D8* *D10* *D12* *D20*

J'ai le D12, quelles sont les chances que j'aie chaque résultat possible ?

Il y a 12 faces numérotés donc 12 résultats possibles ?

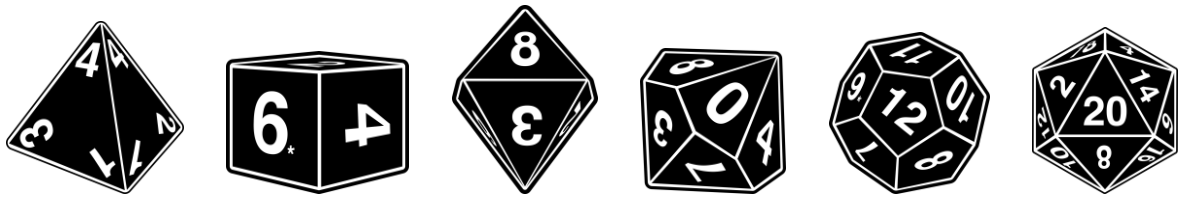
$\frac{100\ \%}{12} = 8{,}3\ \%$ pour chaque face.

Si les faces ne sont pas équiprobables, on peut effectuer 10 000 lancers et établir une table de distribution empirique.



Cela est vrai pour n'importe quel type de dés.

Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.



D4 *D6* *D8* *D10* *D12* *D20*

L'objectif sera de répondre à la question suivante :

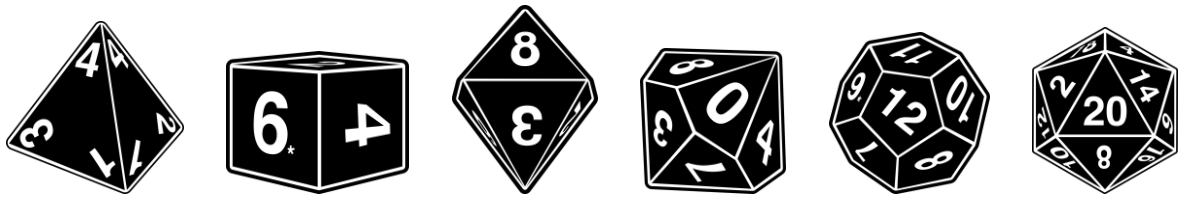
Le maître de jeu lance, il fait un 7. Est-ce que c'est un résultat extraordinaire ?

Le **statisticien fréquentiste** dira que cela dépend de ce qu'on appelle un résultat extraordinaire.

Il faudra définir un seuil pour différencier l'ordinaire de l'extraordinaire. Si on considère un seuil à 1 % (p_{seuil}) et que ma proba dans un D12 est de 8,3 %. Non ce n'est pas un résultat extraordinaire. Pas besoin de remettre en doute le maître de jeu.

Toutefois s'il tire 10 fois 7, là je pourrais me poser la question puisque là si j'ai 1/12 la probabilité de base qui est répétée 10 fois nous avons $\frac{1}{12^{10}} \approx 10^{-11}$ ce qui deçà de 1 % (p_{seuil}).

Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.



D4 *D6* *D8* *D10* *D12* *D20*

L'objectif sera de répondre à la question suivante :

Le maître de jeu lance, il fait un 7. Est-ce que c'est un résultat extraordinaire ?

Le statisticien bayésien se demandera : quel dé a été lancé ?

Ici, je ne cherche pas les chances d'obtenir ce résultat directement, mais davantage quel est le dé qui a été lancé.

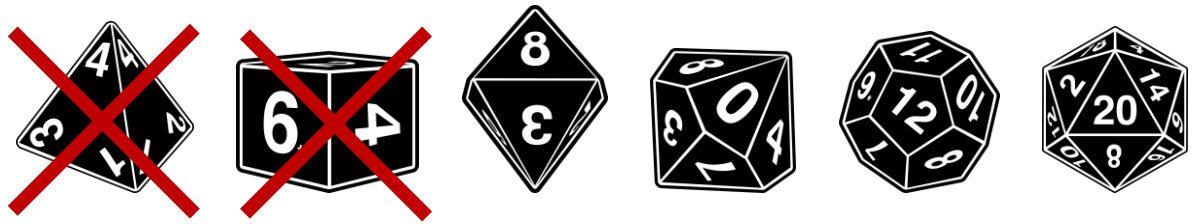
Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.

L'objectif sera de répondre à la question suivante :

Le maître de jeu lance, il fait un 7. Est-ce que c'est un résultat extraordinaire ?

Le statisticien bayésien se demandera : quel dé a été lancé ?

Ici, je ne cherche pas les chances d'obtenir ce résultat directement, mais davantage quel est le dé qui a été lancé.



D4	D6	D8	D10	D12	D20
~~16,7 %~~	~~16,7 %~~	~~16,7 %~~	~~16,7 %~~	~~16,7 %~~	~~16,7 %~~
		25 %	**25 %**	**25 %**	**25 %**

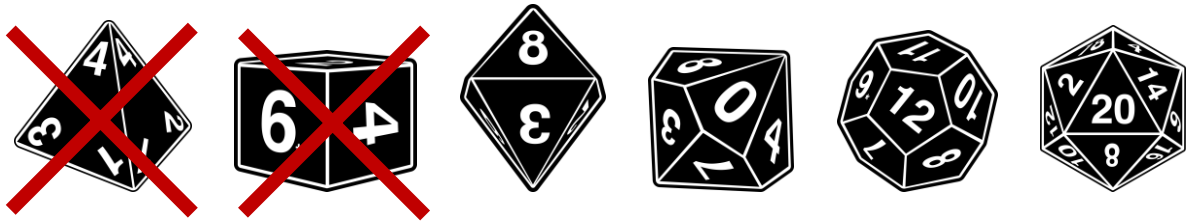
Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.

L'objectif sera de répondre à la question suivante :

Le maître de jeu lance, il fait un 7. Est-ce que c'est un résultat extraordinaire ?

Le statisticien bayésien se demandera : quel dé a été lancé ?

Ici, je ne cherche pas les chances d'obtenir ce résultat directement, mais davantage quel est le dé qui a été lancé.



D4 *D6* *D8* *D10* *D12* *D20*

~~*16,7 %*~~ ~~*16,7 %*~~ ~~*16,7 %*~~ ~~*16,7 %*~~ ~~*16,7 %*~~ ~~*16,7 %*~~

25 % ***25 %*** ***25 %*** ***25 %***

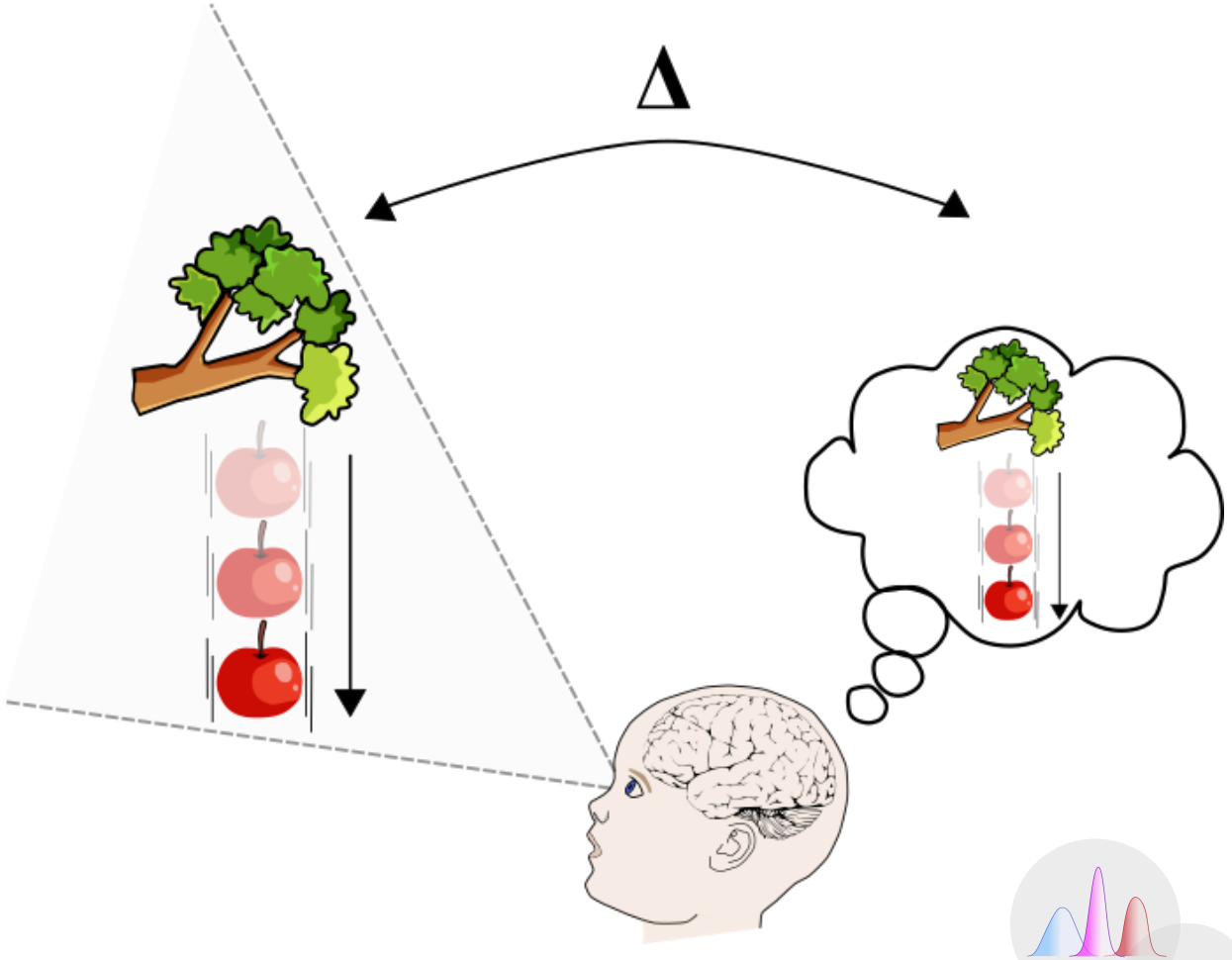
D'accord, mais si nous devions parier, quel est le dé ayant la plus forte probabilité d'avoir été utilisé ?

L'approche fréquentiste considère que les chances sont égales, ce qui est une erreur...

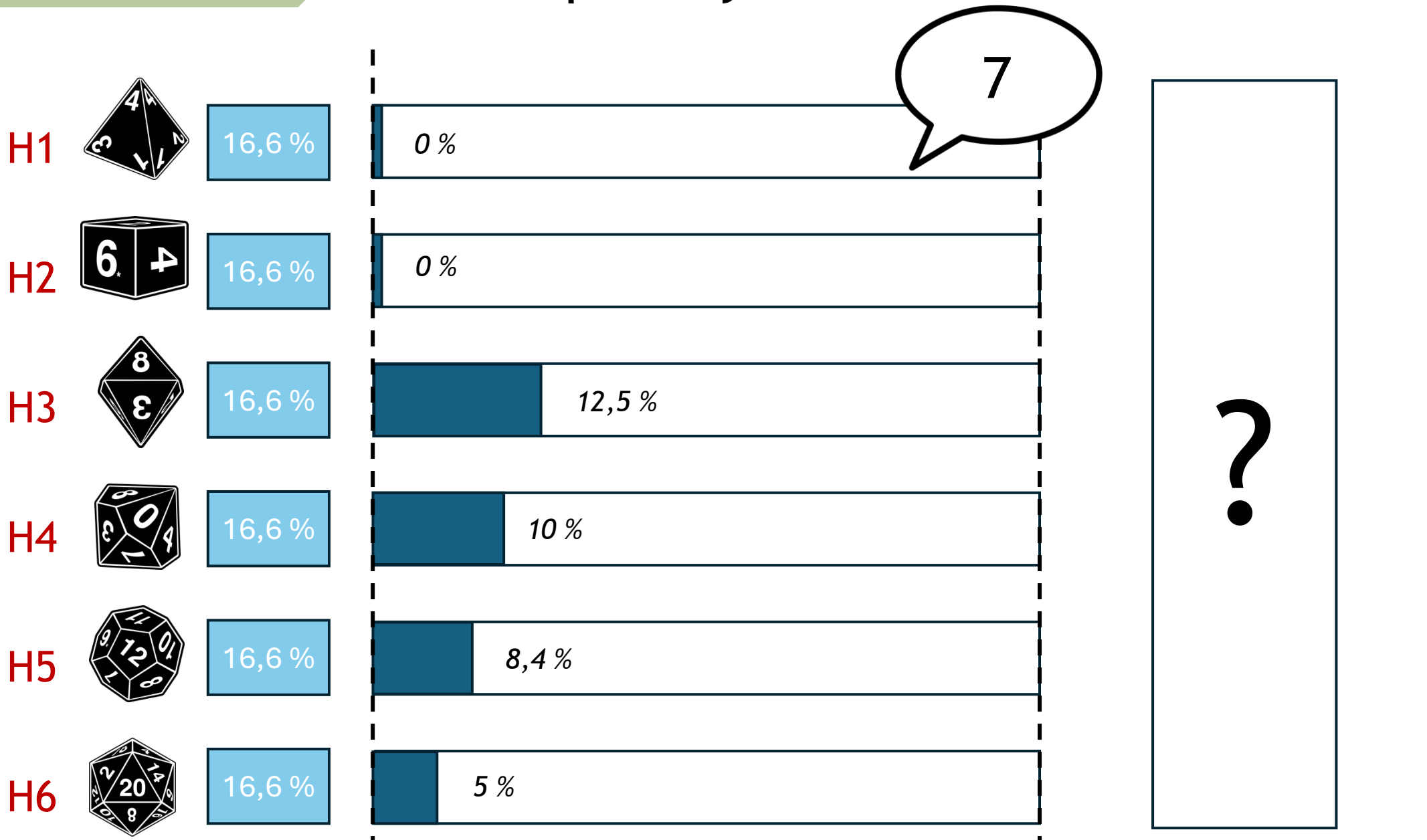
Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.

Différents dés sont les hypothèses pour modéliser le résultat obtenu.

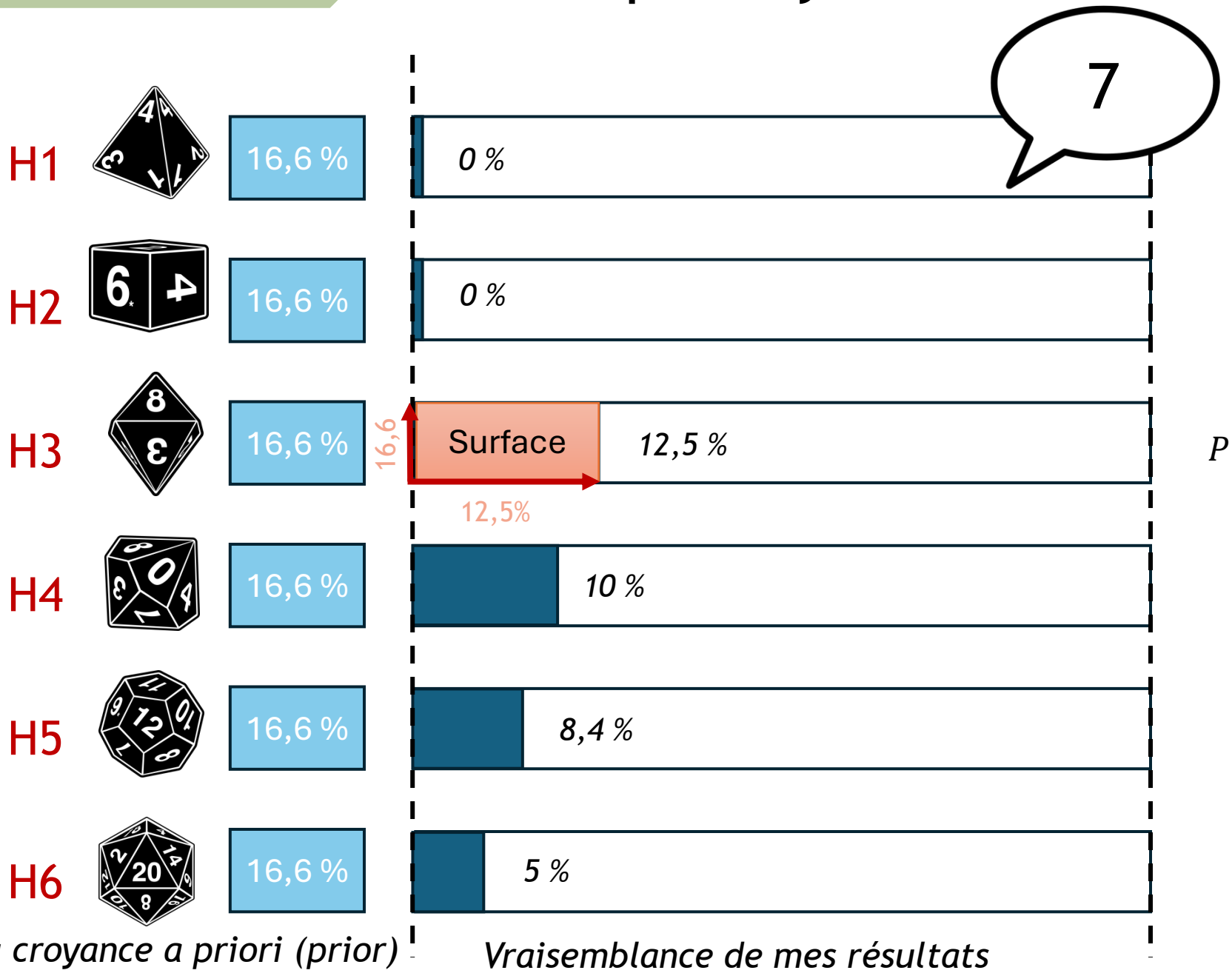
H1	H2	H3	H4	H5	H6
					
D4	*D6*	*D8*	*D10*	*D12*	*D20*



Statistique bayésienne – et lancé de dés



Statistique bayésienne - et lancé de dés



$$P(H3\,|x) = \frac{P(x\,|H3).\,P(H3)}{P(x)}$$

Statistique bayésienne – et lancé de dés

Ma croyance a priori (prior)

Vraisemblance de mes résultats

Le $P(x)$de cette équation est la réponses que les statisticiens fréquentistes cherchent, mais ils vont systématiquement s'arrêter là.

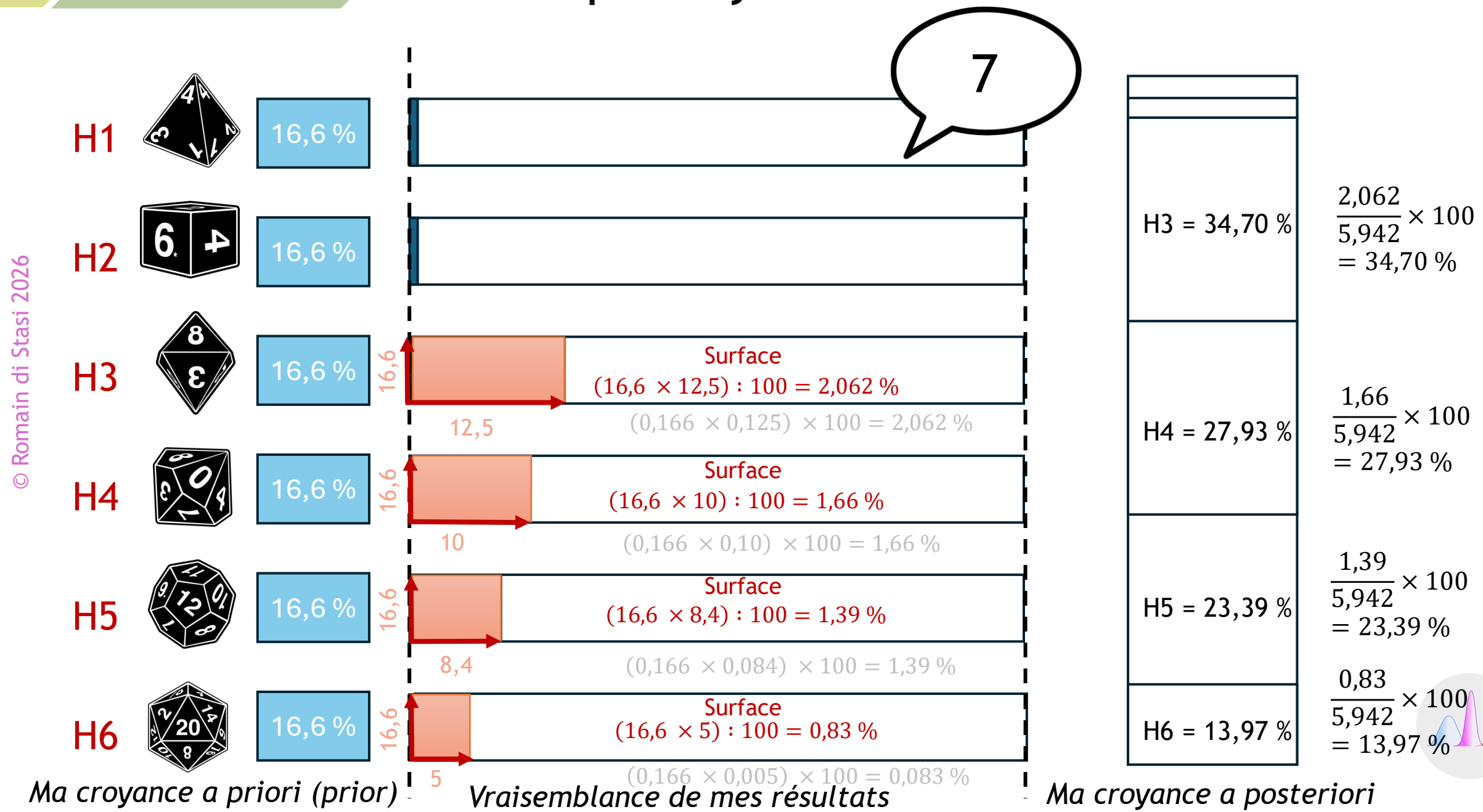
$$P(H3\,|x) = \frac{P(x\,|H3).P(H3)}{P(x)}$$

$$\sum Surfaces$$

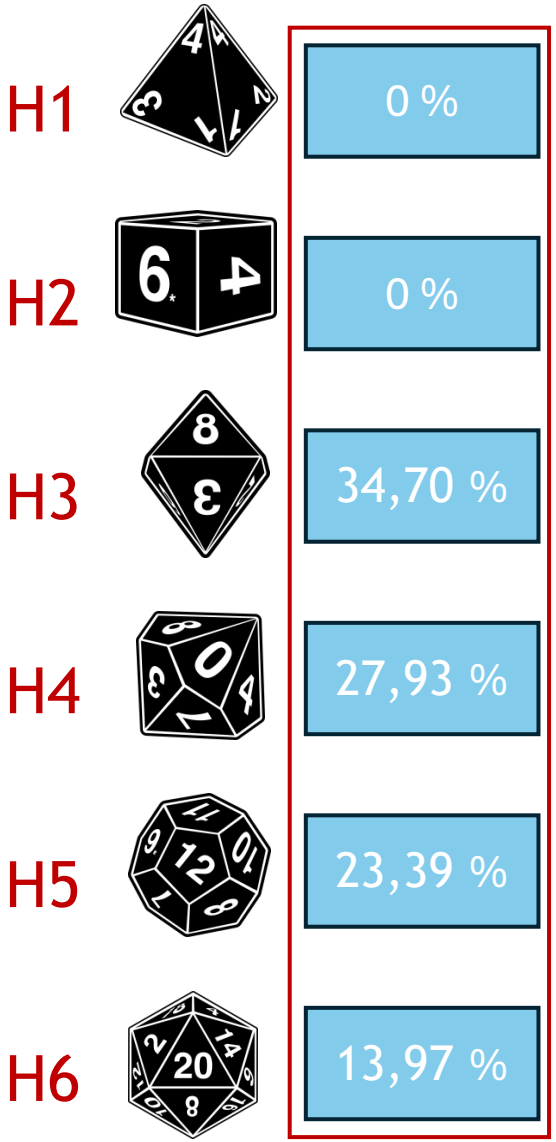
Ici c'est...

$$2{,}062 + 1{,}66 + 1{,}39 + 0{,}83 = 5{,}942\ \%$$

Statistique bayésienne - et lancé de dés



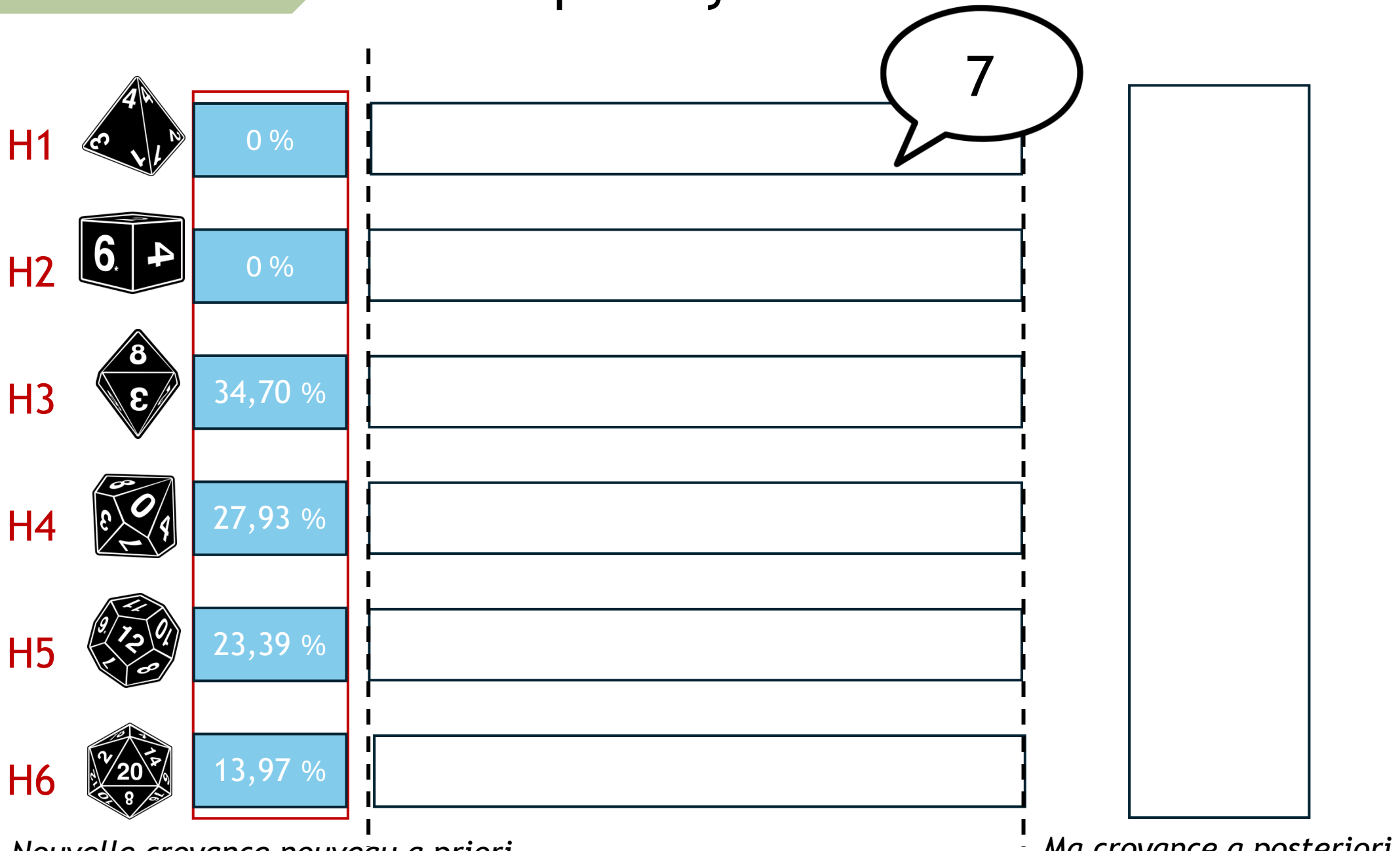
Statistique bayésienne – et lancé de dés



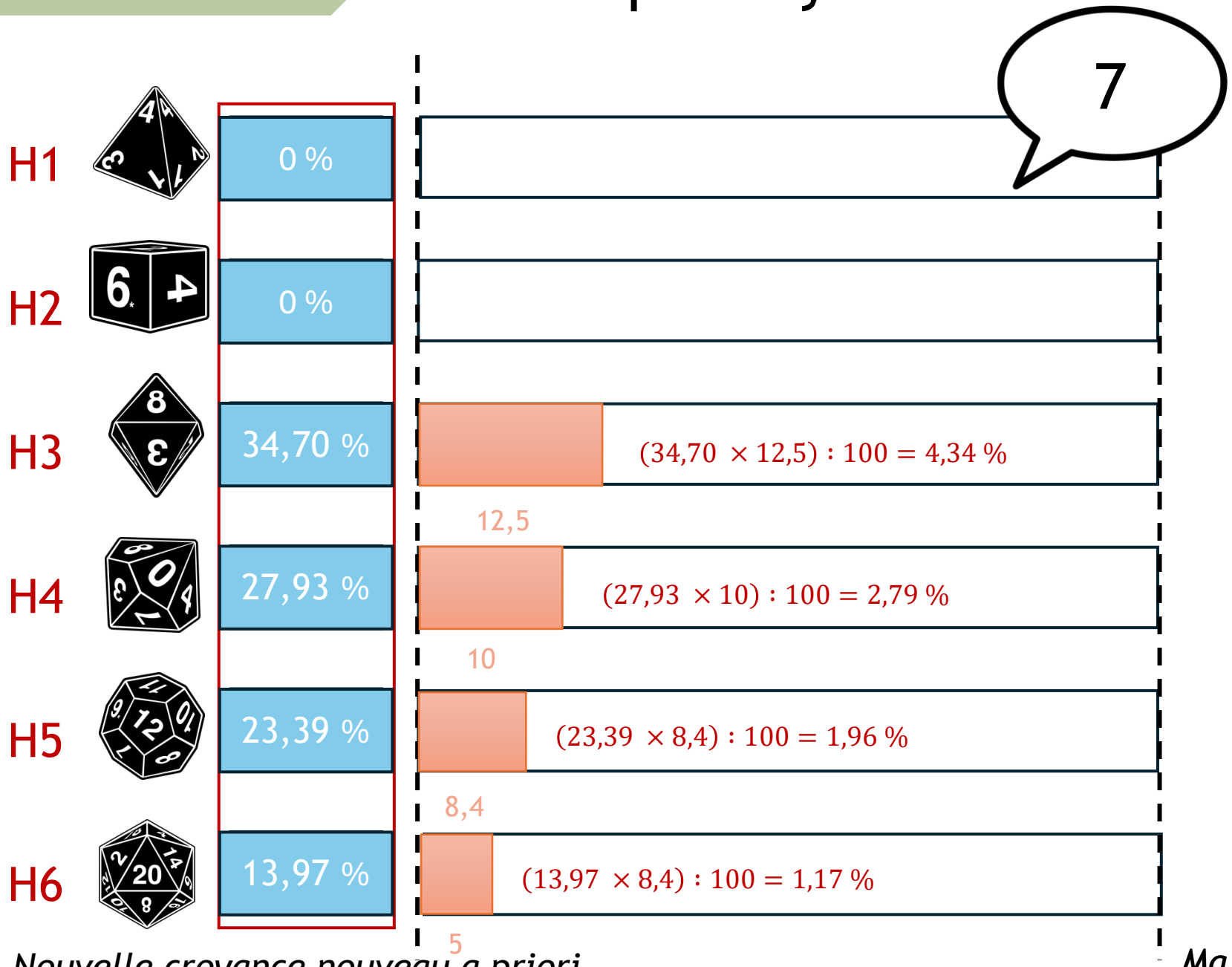
Nouvelle croyance nouveau a priori après le premier lancer (prior)

Puis il y aura une réactualisation des connaissance à chaque lancer. Si je relance un 7, on refait le calcul.

Statistique bayésienne – et lancé de dés



Statistique bayésienne - et lancé de dés

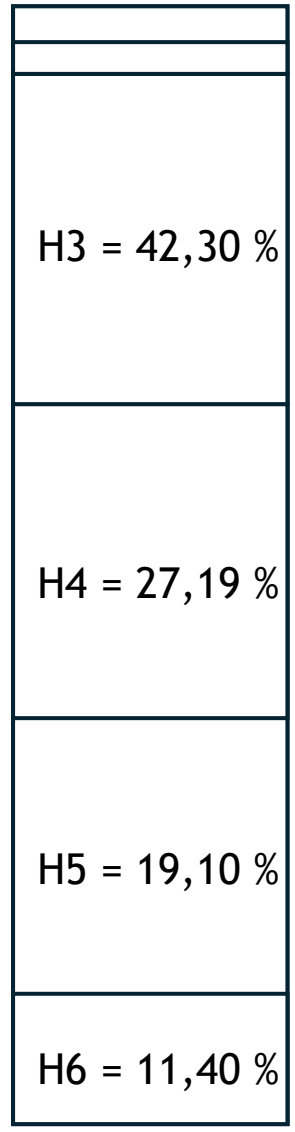


Nouvelle croyance nouveau a priori

$$P(x) = 4{,}34 + 2{,}79 + 1{,}96 + 1{,}17$$

$$= 10{,}26\ \%$$

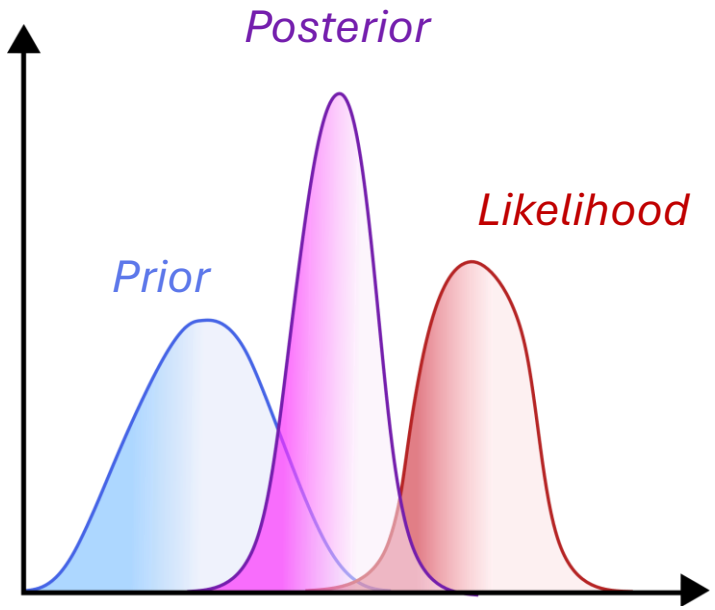
Attention, je l'exprime ici en pourcentage, mais ce n'est pas obligatoire.



Ma croyance a posteriori

Statistique bayésienne - et lancé de dés

$$P(B|A) = \frac{P(B).P(A|B)}{P(A)}$$



$$Posterior = \frac{Likelihood \times Prior}{Average\ Likelihood}$$

Le théorème de Bayes

Un autre exemple…

- Dans un cas de test de dépistage d'une maladie

  Tu es médecin et tu veux estimer la probabilité qu'un patient soit malade, sachant que son test est positif.

  Nous avons 3 probabilités :

  - Mal = « le patient a la maladie »
  - $\overline{Mal}$ = « le patient n'a pas la maladie »
  - Pos = le test est positif

  Nous avons 3 données connues (*priors)* :

  - $P(Mal)$ =prévalence de la maladie, 10% donc 0,1
  - Spécificité du test $P\left(\overline{Pos}\,|\overline{Mal}\right) = 0{,}05$ puisqu'il y a 5% de chance d'un faux positif
  - Sensibilité du test $P(Pos\,|Mal) = 0{,}99$ puisque le test détecte 99% des cas réel

(Adapté de Kruschke 2015, p. 104, table 5.4)

Le théorème de Bayes

Un autre exemple…

- Dans un cas de test de dépistage d'une maladie

  Quelle est la probabilité que le patient soit malade sachant qu'il a eu un test positif ?

$$P(A|B) = \frac{P(Pos|Mal).P(Mal)}{P(Pos|Mal).P(Mal) + P(Pos|\overline{Mal}).P(\overline{Mal})}$$

$$= \frac{0{,}99.0{,}1}{0{,}99.0{,}1+0{,}05.0{,}99} = \frac{0{,}099}{0{,}099+0{,}0495} = \frac{0{,}099}{0{,}1485} \approx 0{,}667 \approx 66{,}7\ \%$$

Ainsi, même si le test est fiable à 99 % en sensibilité et à 95 % en spécificité, si la maladie est rare (1 %), alors un test positif ne signifie pas nécessairement que la personne est malade.

(Adapté de Kruschke 2015, p. 104, table 5.4)

La notion de vraisemblance

- Dans cet exemple de dépistage de la maladie, il faut bien comprendre la **notion de vraisemblance…**

  Ici nous avons mesuré la probabilité que le test soit positif sachant que $P(A)$, la prévalence de la maladie est égale à 0,1…

La **vraisemblance** correspond à la probabilité d'observer certaines données *à supposer qu'une hypothèse soit vraie*. Nous l'avons déjà vue à travers l'exemple des dés. Par exemple, si je pense qu'il y a 40 % de chances que le dé utilisé soit un D10, **la vraisemblance est la probabilité que ce dé ait produit un 7 - autrement dit : *quelle est la compatibilité entre cette hypothèse et mon observation*.**

(McElreath 2015, p. 32-33)

Dans le contexte du dépistage, la **vraisemblance correspond à la probabilité que le test soit positif si la personne est réellement malade**.

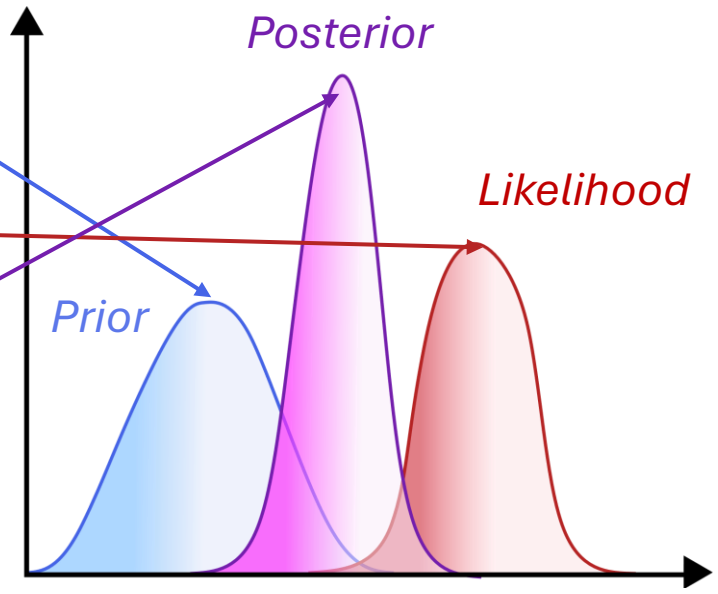
Ne pas confondre vraisemblance et *a posteriori*

Vraisemblance	*A posteriori*
$P(test\ positif\ \|malade)$	$P(malade\|test\ positif)$

Vraisemblance vs. Probabilité *a posteriori*

Cas du dépistage

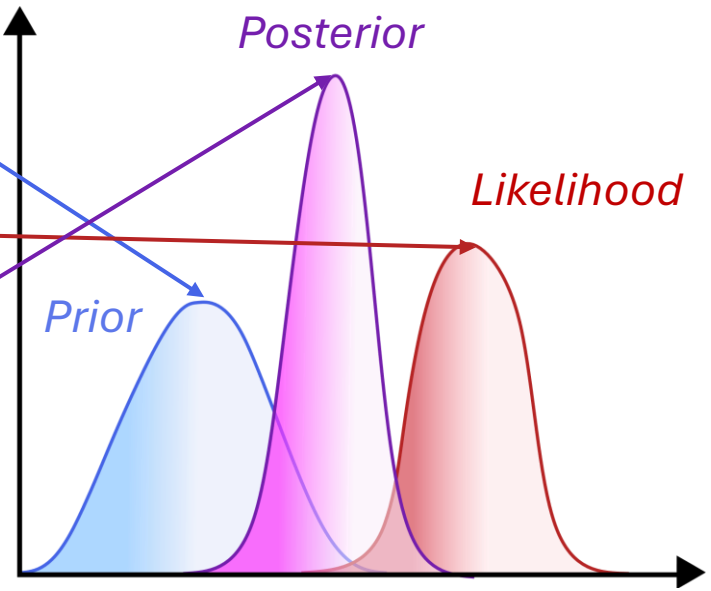
Terme	Représente quoi	Example dans le test
Croyance A priori (*prior*)	Ce que je pense avant de voir le test	$P(Mal) = 1\%$
Vraisemblance (*Likelihood*)	Probabilité d'observer le test positif *si la personne est malade*	$P(Pos)$
Croyance A poseriori (*posterior*)	Ce que je crois après avoir vu le test	$P(Mal) = 1$



Vraisemblance vs. Probabilité *a posteriori*

Cas des dé

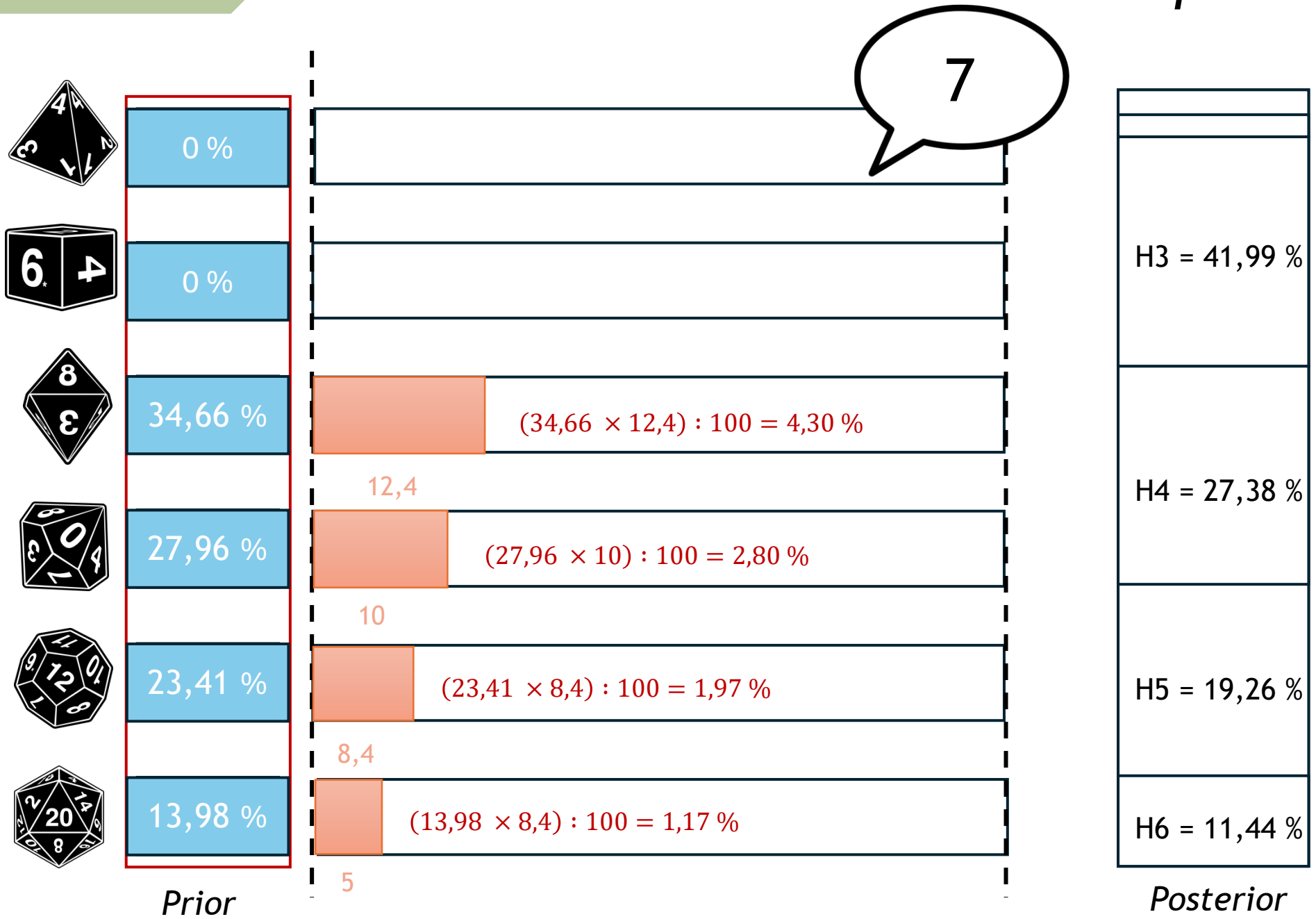
Terme	Représente quoi	Example dans le test
Croyance A priori (*prior*)	Ce que je pense avoir comme résultat	P(dé = D10) = 40%
Vraisemblance (*Likelihood*)	Probabilité d'obtenir le résultat *si* l'hypothèse est vraie	P(7)
Croyance A poseriori (*posterior*)	Ce que je crois après avoir vu le test	P (Dé = D10)



- Le croyance *A posteriori* (ou *posterior*) représente la combinaison de l'information qui provient de l'observation des données et nos croyances *a priori* (*prior*)

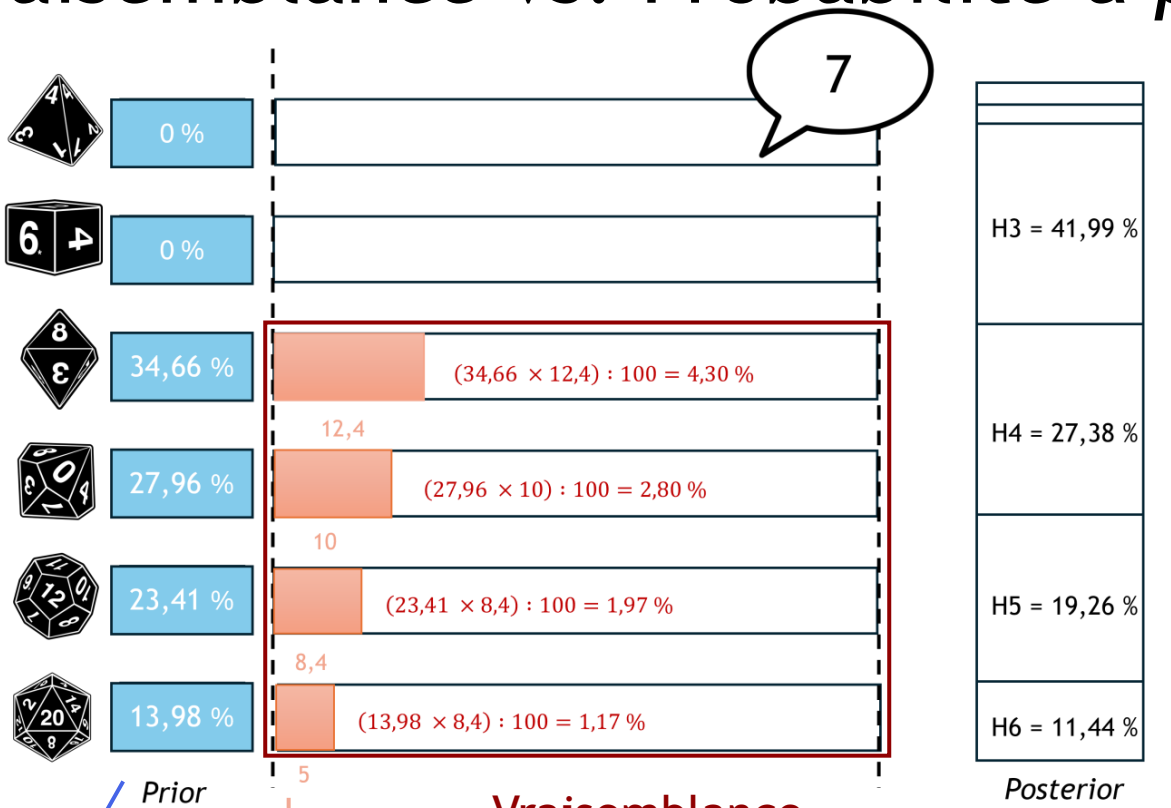
Vraisemblance vs. Probabilité *a posteriori*

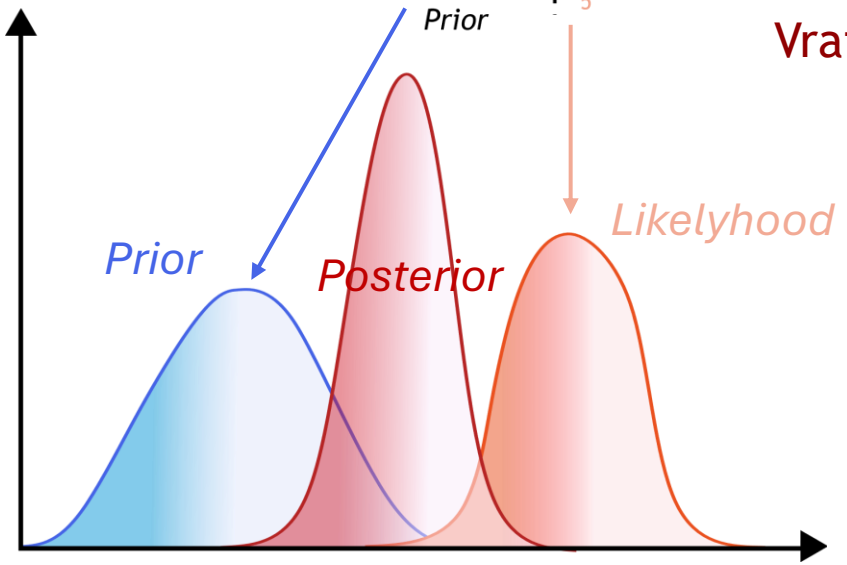
Rappel



Vraisemblance vs. Probabilité *a posteriori*

Rappel





Vraisemblance

La probabilité de nos données sachant notre ou nos hypothèses

Donc

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Prior et loi binomiale

- Afin de bien comprendre l'approche bayésienne et la vraisemblance il est assez intéressant de considérer une loi simple, **la loi binomiale**, pour plusieurs raisons.

→ Elle modélise des situations très courantes : succès/échec répétés.

→ Elle joue un rôle central en tant que **vraisemblance** dans l'approche bayésienne.

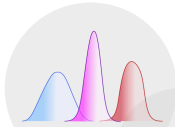
Puisque…

- Cette loi se base sur une probabilité *a priori* comme dans l'équation de Bayes

→ Elle est associée à une loi **conjuguée simple** : la loi bêta.

Elle est donc utilisable…

- Dans de nombreux cas qui nous intéressent en psychologie et neuroscience (e.g., différence de pourcentages).



Loi Binomiale

Rappel

- La loi binomiale consiste en la probabilité de constater une fréquence observée en fonction d'une fréquence attendue.

  Cette dernière est binaire : cela peut se traduire par le fait d'avoir réussi on non son exam, d'être malade ou non, *marquer à un panier de* basket *ou non, etc.*

© exemple tiré de Primer

- Admettons qu'un joueur de basket ait 60 % de chances de réussir un lancer franc.

  → Quelle est la probabilité qu'il ne marque aucun point en deux essais ?

  → Une fois sur les deux essais ?

  → Deux fois sur les deux essais ?

(Kruschke 2015, p. 72-97 & 123-141)

Loi Binomiale

Rappel

- Admettons qu'un joueur de basket ait 60 % de chances de réussir un lancer franc.

→ Quelle est la probabilité qu'il ne marque aucun point en deux essais ?

$p(fail, fail) = (1 - p(success)) \times (1 - p(success)) = 0{,}40 \times 0{,}40 = 0{,}16$ *donc* 16%

→ Une fois sur les deux essais ?

$p(\text{success}, \text{fail}) + p(\text{fail}, \text{sucess}) = 0{,}60 \times 0{,}40 + 0{,}40 \times 0{,}60 = 0{,}48$ donc 48%

→ Deux fois sur les deux essais ?

$p(sucess, succes) = p(\text{sucess}) \times p(\text{sucess}) = 0{,}60 \times 0{,}60 = 0{,}36$ *donc* 36%

© exemple tiré de Primer

(Kruschke 2015, p. 72-97 & 123-141)

Loi Binomiale

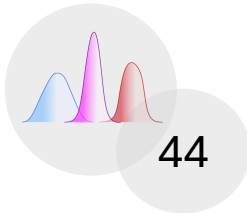
Rappel

- Cas d'une étude sur les déterminants des troubles du sommeil chez les enfants de 2 à 3 ans

  - Je sais que le pourcentage d'enfants atteint de trouble du sommeil dans la population d'étude est $p = 0{,}17$ ou 17%.

  - J'ai une taille d'échantillon de $n = 10$ et un nombre de personnesdans mon échantillon qui ont ce trouble de $k = 4$

La question que l'on se pose est : cet échantillon provient-il bien de la population étudiée ?



(Kruschke 2015, p. 72-97 & 123-141)

Loi Binomiale

Donc, quelle est la probabilité d'observer 4 malades dans un échantillon de 10 sujets pris au hasard, sachant que la prévalence de la maladie est de 17 % ?

(1) Soit cette probabilité est élevée, et dans ce cas, l'échantillon observé peut s'expliquer par une simple fluctuation aléatoire.

(2) Soit elle est faible et l'échantillon ne représente pas la population.

Ce qu'on va chercher à faire c'est de comprendre par étapes :

(1) Quelle est la probabilité d'observer k individus possédant une caractéristique donnée…

(2) Dans un échantillon de n individus

(3) Tirés dans une population où la proportion p de la caractéristique est connue.

(Kruschke 2015, p. 72-97 & 123-141)

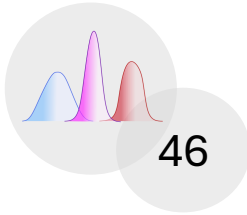
Loi Binomiale

Caractéristique = trouble du sommeil	$\boldsymbol{k}$	4
Taille d'échantillon	$\boldsymbol{n}$	10
Proportion de sujets porteurs de la caractéristique de la population	$\boldsymbol{p}$	0,17

Quelle est la probabilité de $\boldsymbol{k}$ succès au bout de n tentatives sachant que la probabilité $\boldsymbol{p}$ de gagner à chacune des tentatives.

$$\Pr(\mathrm{k}) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$\Pr(4) = \frac{10!}{4!\,(10-4)!} 0{,}17^4 (1-0{,}17)^{10-4} = 0{,}057 = 5{,}7\ \%$$



(Kruschke 2015, p. 72-97 & 123-141 ; Myung 2003)

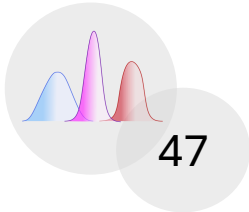
Loi Binomiale

Caractéristique = trouble du sommeil	k	0
Taille d'échantillon	n	10
Proportion de sujets porteurs de la caractéristique de la population	p	0,17

Quelle est la probabilité de k succès au bout de n tentatives sachant que la probabilité p de gagner à chacune des tentatives.

$$\Pr(k) = \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$\Pr(0) = \frac{10!}{0!\,(10-0)!} 0{,}17^0 (1-0{,}17)^{10-0} = 0{,}155 = 15{,}5\ \%$$



(Kruschke 2015, p. 72-97 & 123-141 ; Myung 2003)

Loi Binomiale - cumul des probabilités

$$\Pr(0) = \frac{10!}{0!\,(10-0)!}\,0{,}17^{0}(1-0{,}17)^{10-0} = 0{,}155 = 15{,}5\,\%$$

$$\Pr(1) = \frac{10!}{1!\,(10-1)!}\,0{,}17^{1}(1-0{,}17)^{10-1} = 0{,}318 = 31{,}8\,\%$$

$$\Pr(2) = \frac{10!}{2!\,(10-2)!}\,0{,}17^{2}(1-0{,}17)^{10-2} = 0{,}293 = 29{,}3\,\%$$

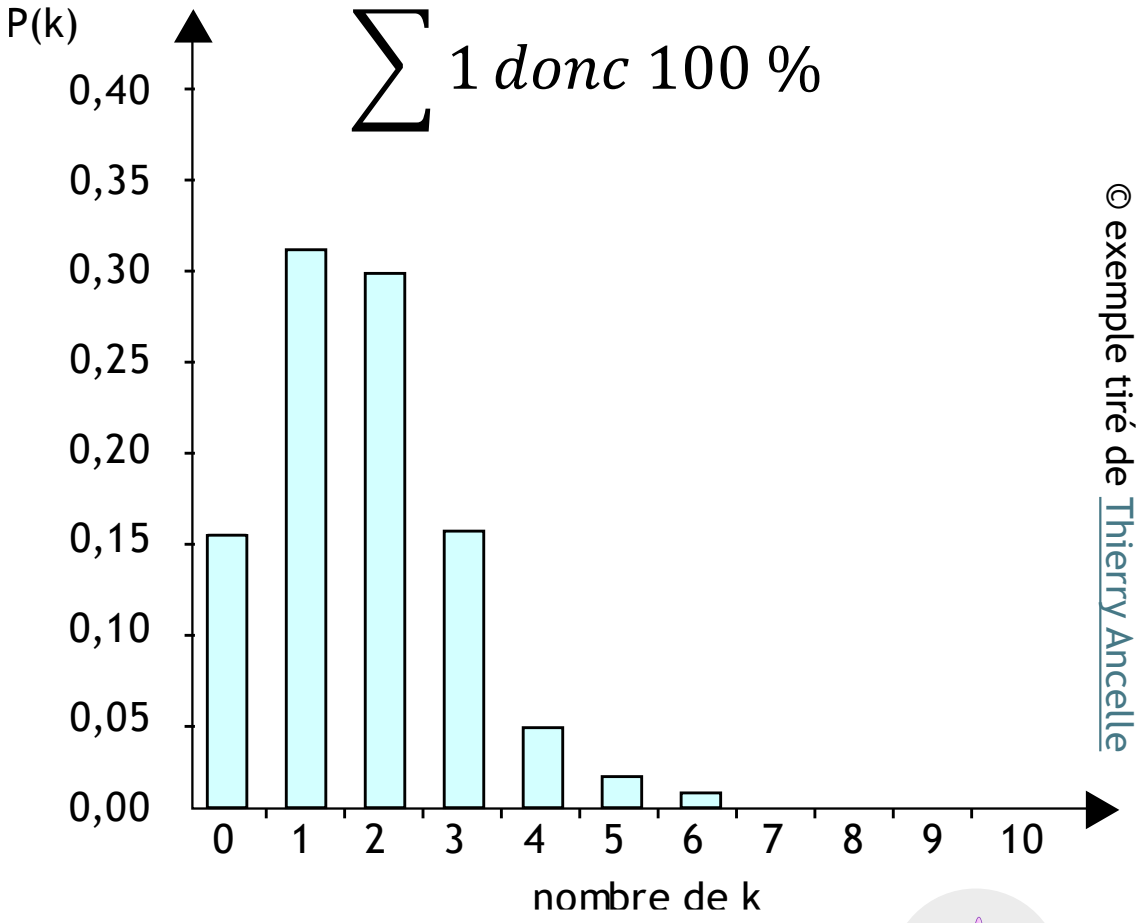
$$\Pr(3) = \frac{10!}{3!\,(10-3)!}\,0{,}17^{3}(1-0{,}17)^{10-3} = 0{,}160 = 16{,}6\,\%$$

$$\Pr(4) = \frac{10!}{4!\,(10-4)!}\,0{,}17^{4}(1-0{,}17)^{10-4} = 0{,}057 = 5{,}7\,\%$$

$$\Pr(5) = \frac{10!}{5!\,(10-5)!}\,0{,}17^{5}(1-0{,}17)^{10-5} = 0{,}014 = 1{,}4\,\%$$

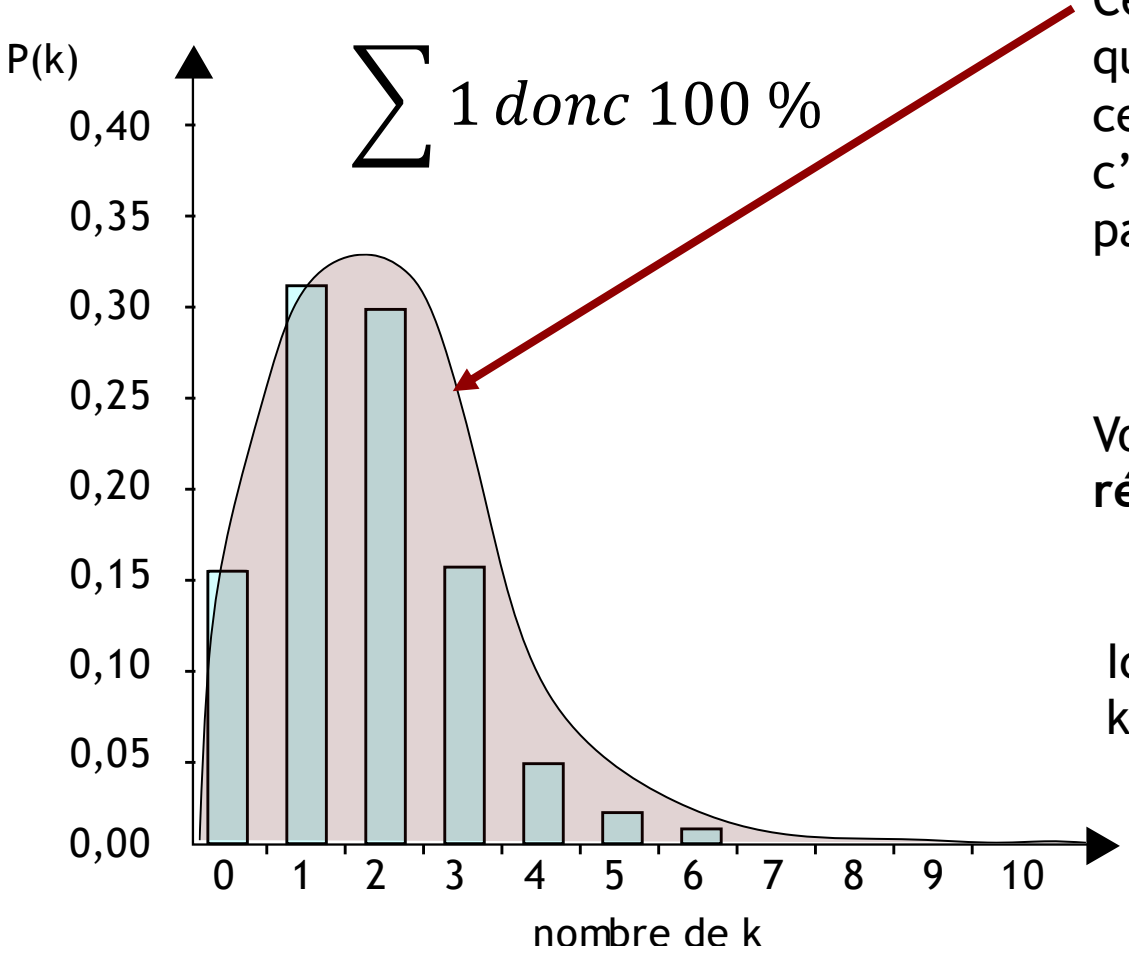
⋮

$$\Pr(10) = \frac{10!}{10!\,(10-10)!}\,0{,}17^{10}(1-0{,}17)^{10-10} = 0{,}00000002$$
$$= 0{,}000002\,\%$$



(Kruschke 2015, p. 72-97 & 123-141 ; Myung 2003)

Loi Binomiale – cumul des probabilités

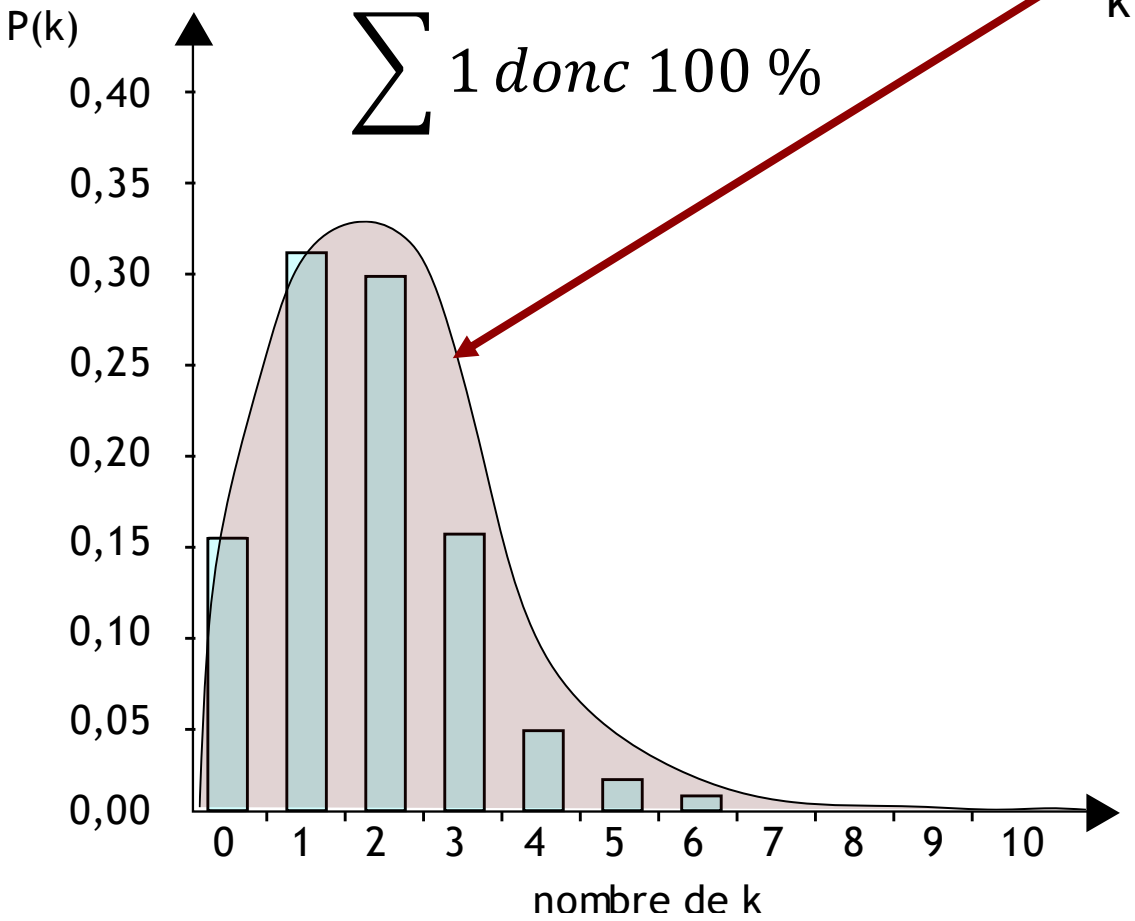


Ceci est une **fonction de densité.** La distribution théorique que vous êtes sensé obtenir si vous échantillonnez un certain nombre de participant. Elle est très importante car c'est elle qui va nous permettre de comprendre ce qui se passe statistiquement en traçant une droite de régression.

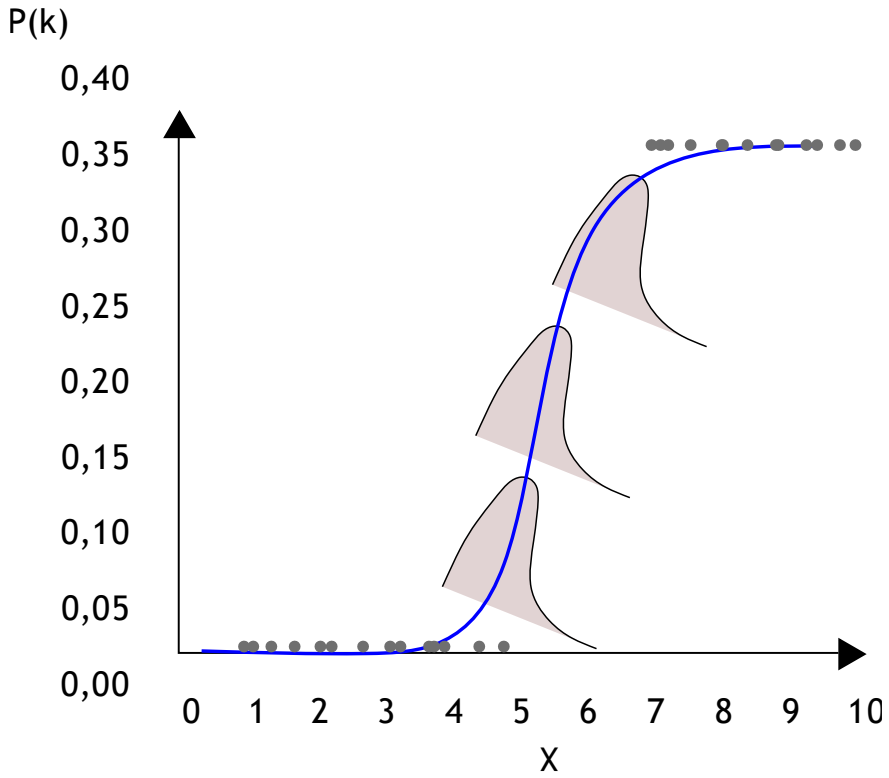
Vous avez sans doute toutes et tous entendu parler de **la régression linéaire**.

Ici ce serait une **régression logistique binomiale** puisque k ne peut prendre que deux modalités 0 ou 1.

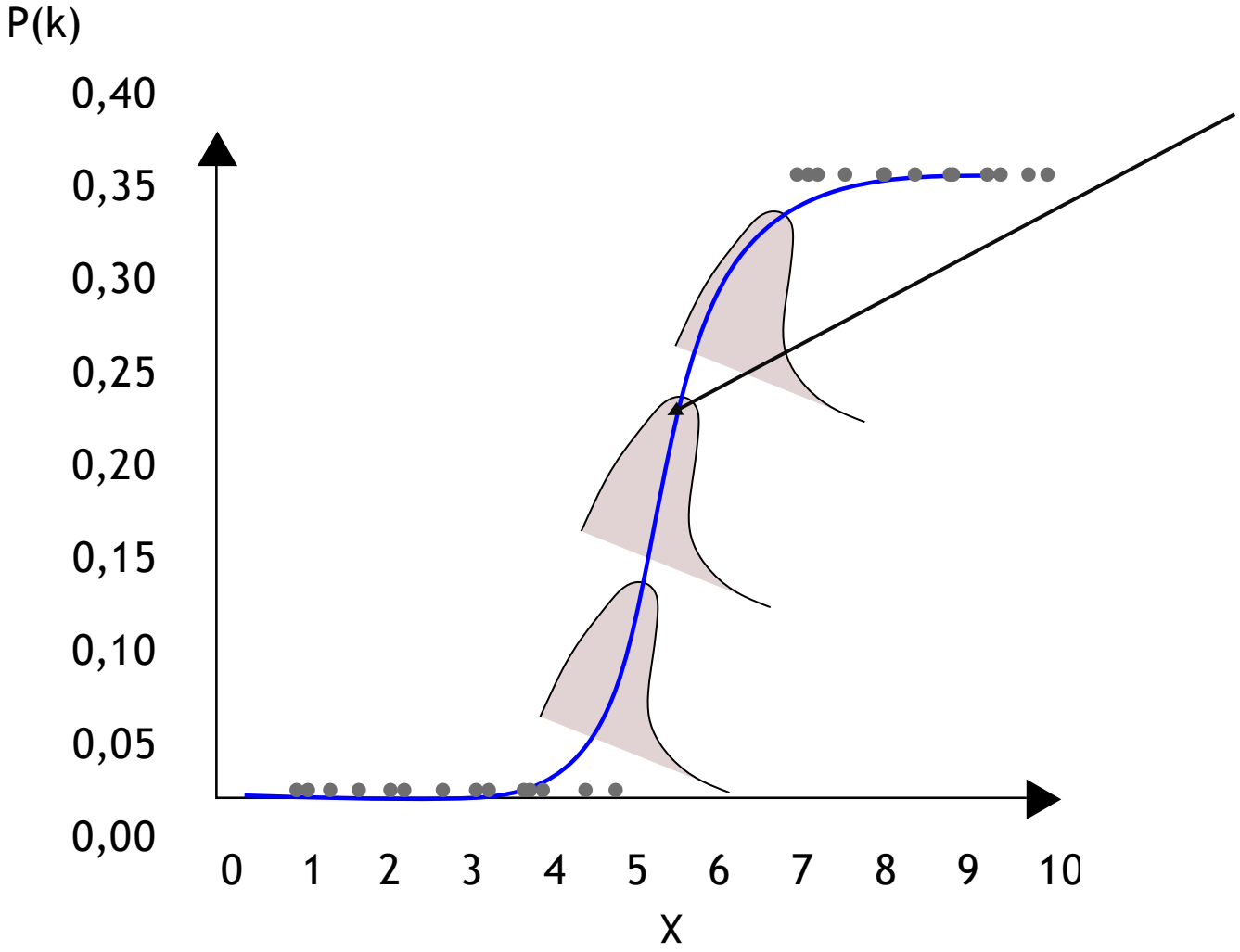
Loi Binomiale - cumul des probabilités



Ici ce serait une **régression logistique binomiale** puisque k ne peut prendre que deux modalités 0 ou 1.



Loi Binomiale – cumul des probabilités



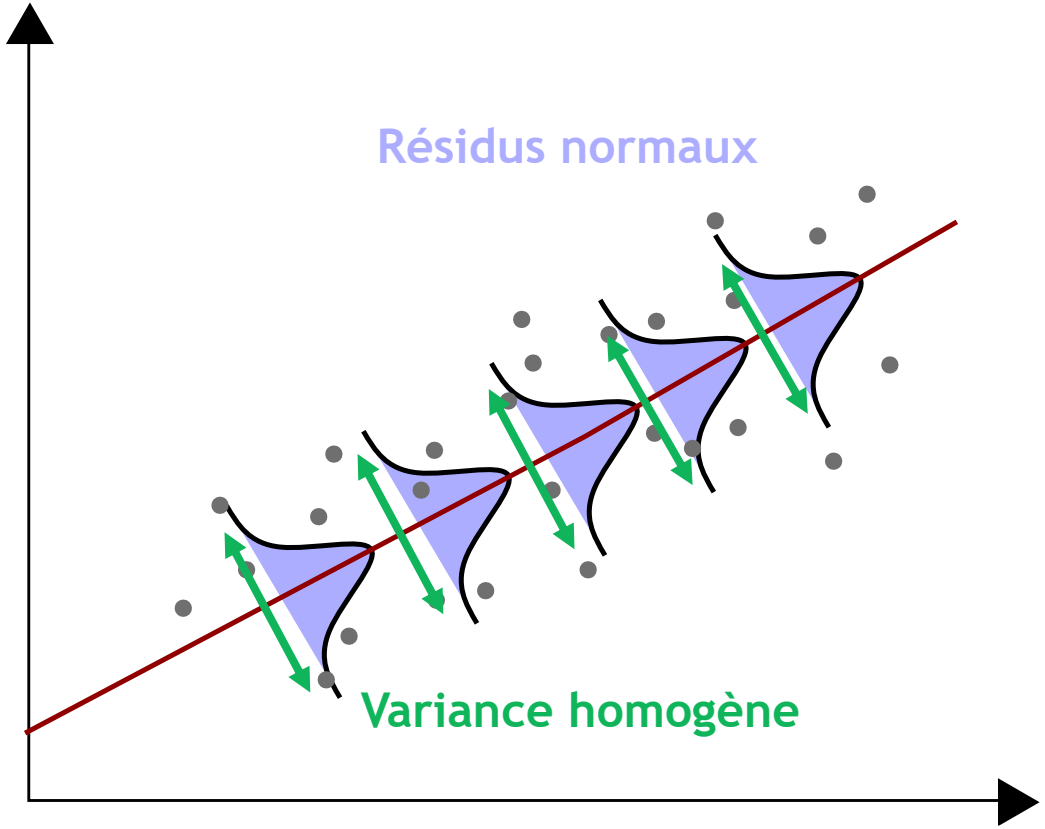
En traçant cette droite, nous pouvons comprendre la relation qui existe entre deux variables, qu'il s'agisse de deux variables quantitatives ou d'une variable quantitative et d'une variable qualitative (par exemple : groupe contrôle vs groupe test)

C'est le même principe que la régression linéaire mais avec X pouvant prendre que deux valeurs 0 ou 1.

Loi Binomiale – cumul des probabilités

Pour **placer la droite** correctement vis-à-vis de cette fonction de densité il existe deux méthodes.

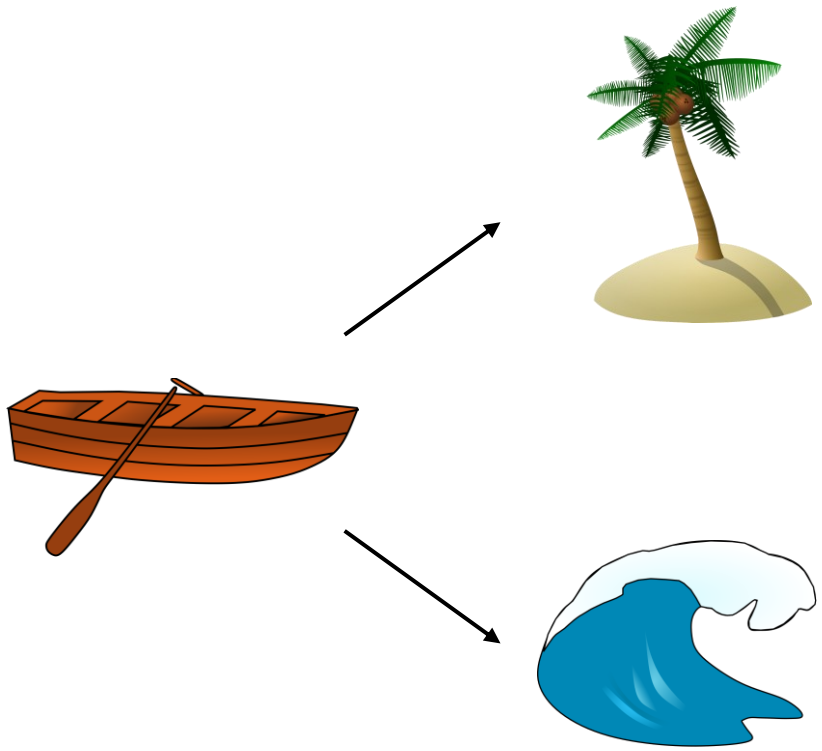
Les **moindres carrées** dans le cas d'un **régression linéaire classique**. Cela suppose que nous attendons à ce que les résidus de la pente (la fonction de densité) suivent une loi normale et la variance de ces derniers soit homogènes comme illustré ici.

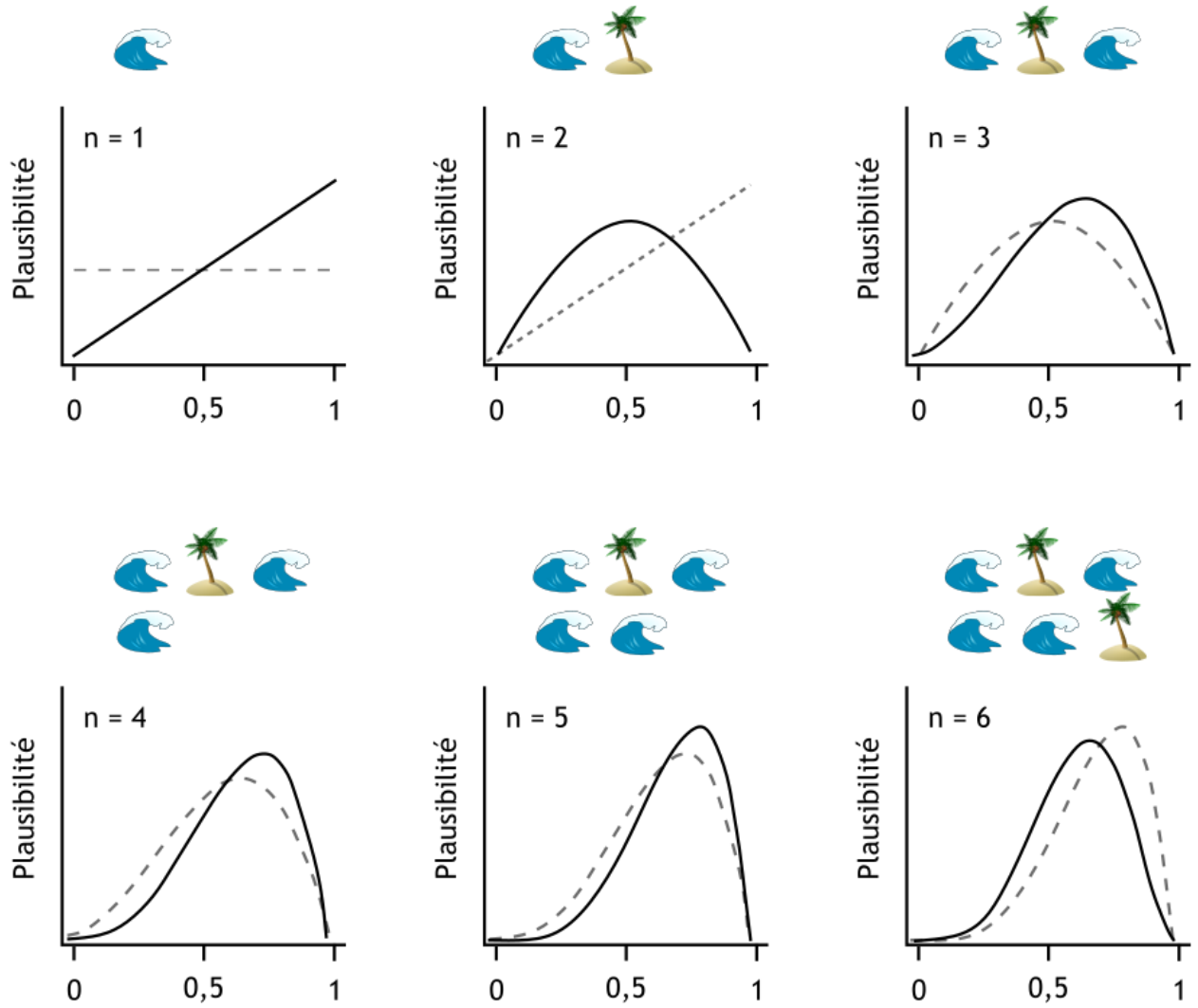


Pour tous les autres cas nous appliquons le **maximum de vraisemblance**.

(Kruschke 2015, p. 72-97 & 123-141 ; Myung 2003)

Actualisation du *prior* bayésien



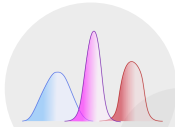


$$\Pr(e \mid n, p) = \frac{e!}{e!\,(n-w)!} p^{e} (1-p)^{n-e}$$

Adapté de McEareath 2015, p. 28-32

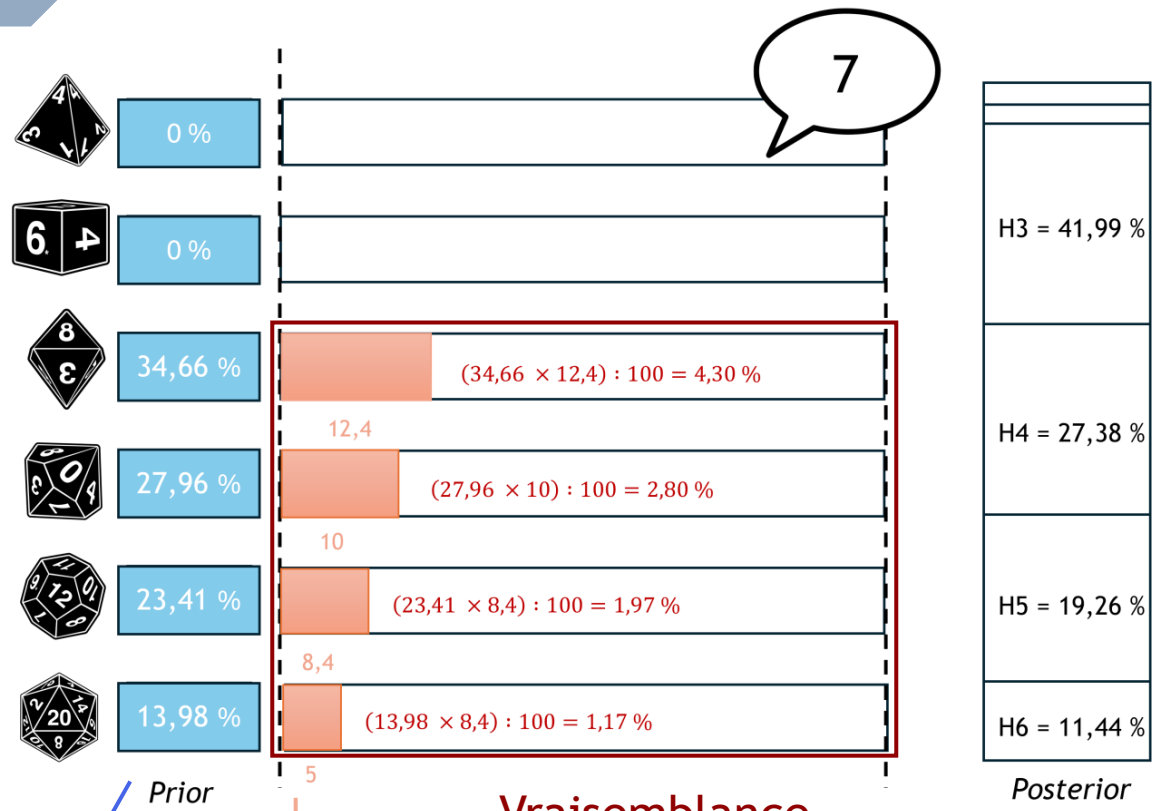
Maximum de vraisemblance

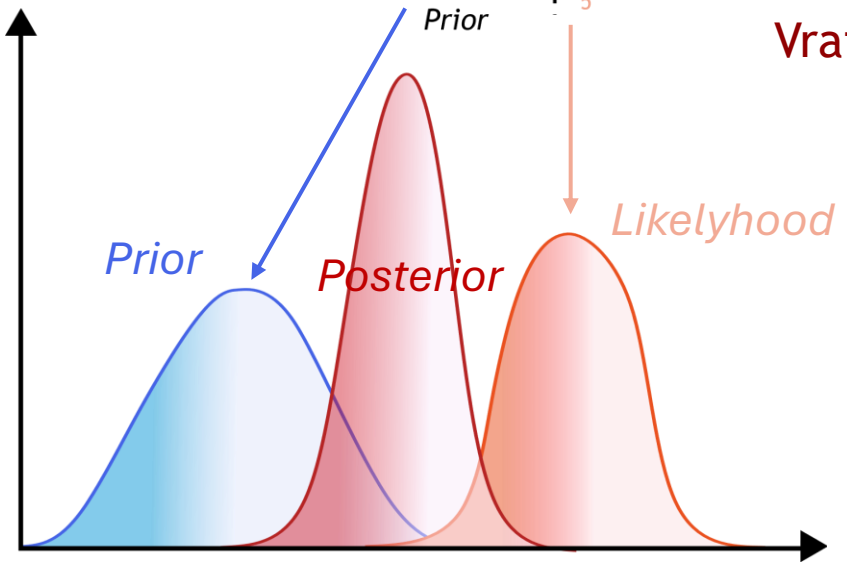
Appliquons maintenant le maximum de vraisemblance à la loi binomiale



Maximum de vraisemblance

Rappel





Vraisemblance

La probabilité de nos données sachant notre ou nos hypothèses

Donc

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Maximum de vraisemblance
Loi Binomiale

- La vraisemblance est une fonction notée :

Probabilité d'observer les données si le paramètre du modèle est θ

$$L(\theta) = P(données \mid \theta)$$

- Le paramètre θ c'est le ou les paramètres inconnus que l'on cherche à estimer, il peut suivre un ensemble de loi de probabilité comme la loi normale, la loi de poisson (qui ont tous deux plusieurs paramètres) mais ici nous commencerons par loi de Binomiale qui n'en possède qu'un $\theta = p$, dont p est une probabilité fixe.

- Ensuite, nous avons ce que nous avons mesuré qui nous l'espérons suit la même loi, indépendante et identiquement distribuée notée $\{x_1, x_2, \dots, x_n\}$, elle suivront une loi de probabilité $f(x \mid \theta)$

- La fonction de vraisemblance est donc :

Fonction de densité dans laquelle nous avons la probabilité d'observer x_i selon θ

$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

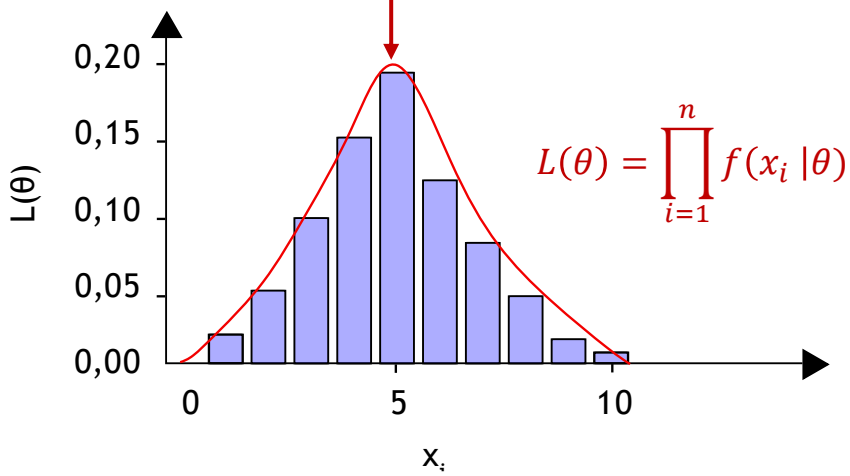
Le produit de toutes les probabilités en supposant qu'elles soient indépendantes

Maximum de vraisemblance
Loi Binomiale

- On cherche ensuite le **maximum de vraisemblance (MLE)**

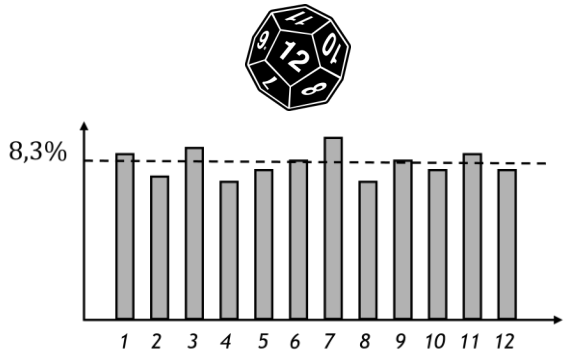
L'idée ici est de trouver la valeur de θ qui rend les données les plus probables. Formellement ça donne:

$$\arg\max L(\theta)$$



Rappel

Fonction de densité



Exemple

- Supposons que **je lance une pièce 10 fois**, donc $n = 10$ et que j'observe $k = 7$ faces. Je veux estimer p, la probabilité de tomber sur face. La vraisemblance de la loi binomiale est :

$$L(p) = P(k\ face \mid p) = \binom{10}{7} p^7 (1-p)^3$$

Puisque la loi binomial c'est $f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$

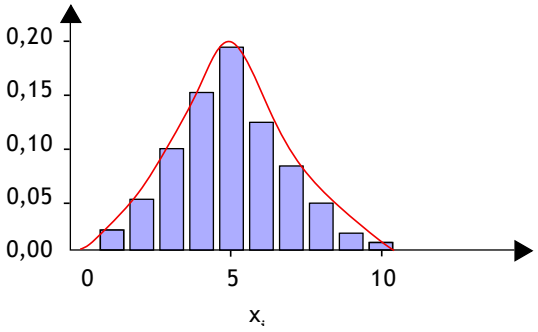
- Le maximum de vraisemblance correspond au p qui maximise cette expression

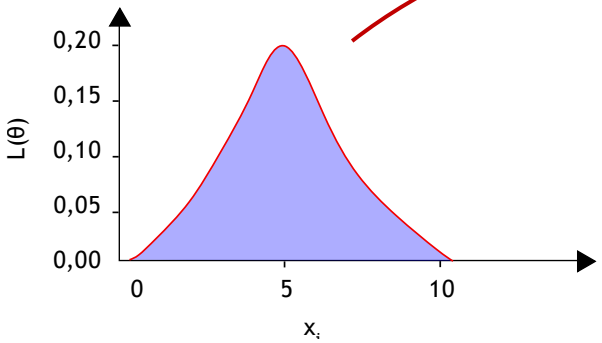
$$\hat{p}MLE = \frac{7}{10} = 0{,}7$$

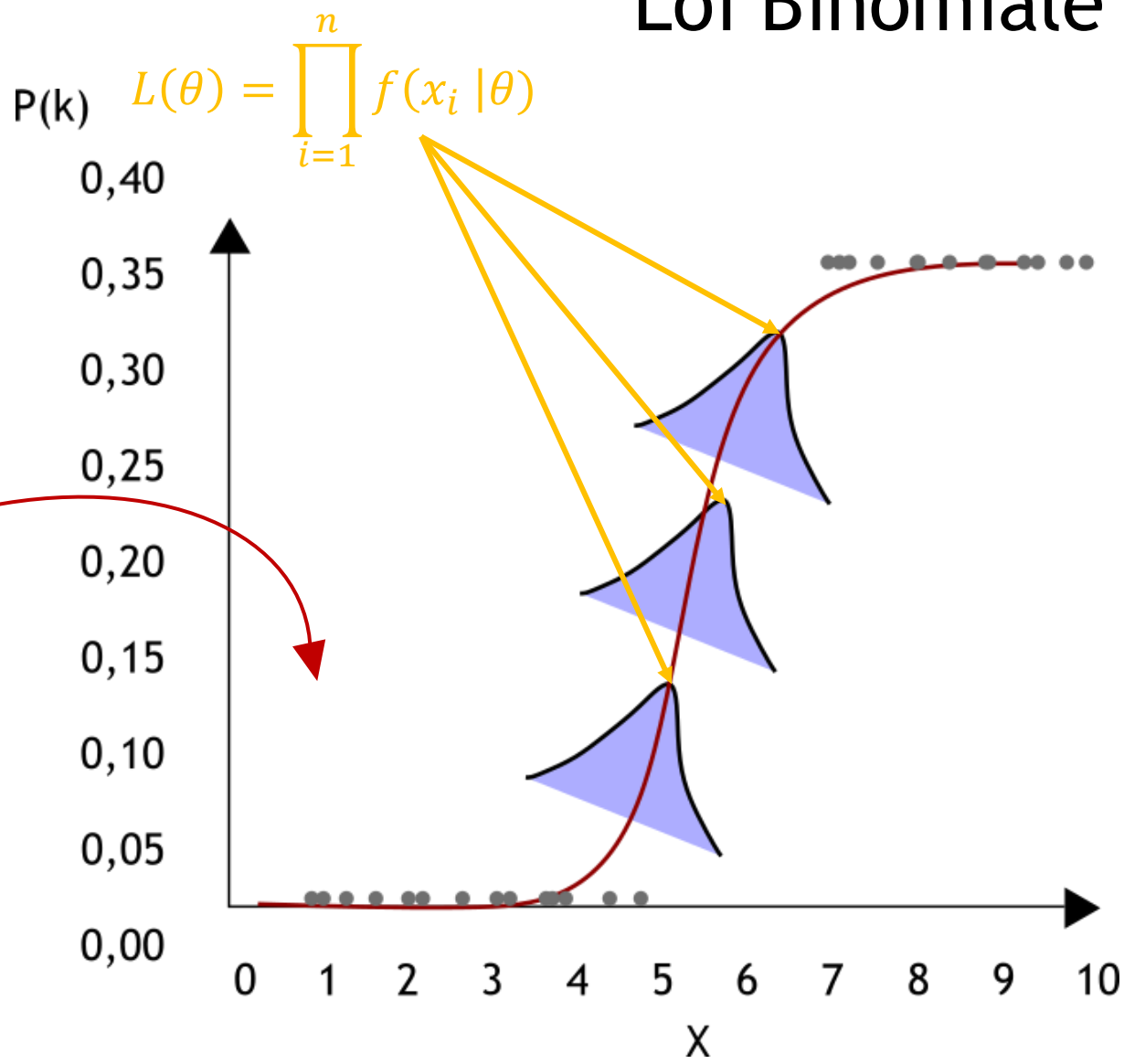
Maximum de vraisemblance

Loi Binomiale

$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

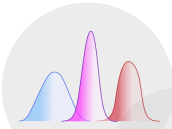






Maximum de vraisemblance

Cas particulier : en régression linéaire, la méthode des moindres carrés et le maximum de vraisemblance sont équivalents



Régression linéaire et moindres carrés

Rappel

La régression linéaire

- Qu'est-ce que la méthode des moindres carrés ?

  Qui dit « moindres carrés » dit un retour aux régressions.

- Une question que l'on m'a souvent posée ces dernier temps est… y a-t-il des similitudes entre régression linéaire et statistiques bayésiennes.

  En réalité, peu importe le type de statistiques que l'on fait, notre méthode repose toujours sur une formation linaire de la fonction suivante :

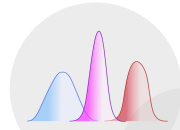
$$y = ax + b$$

Si nous remontons à l'équation de la droite au lycée

et

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} + \in_i$$

Si nous revenons à quelque chose de plus récent



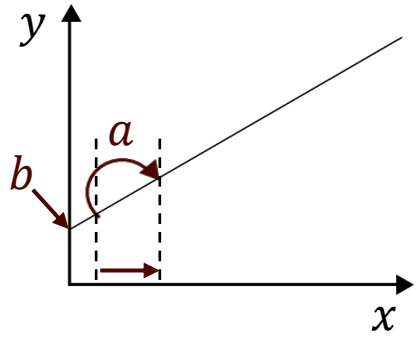
Régression linéaire et moindres carrés

Rappel

La régression linéaire

- La régression linéaire correspond à l'équation de la droite vue au lycée

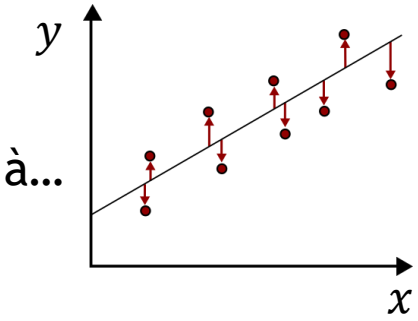
$$y = ax + b$$



- A laquelle on ajoute l'erreur $(\in_i)$

$$y = ax + b + (\in_i)$$

Qui correspond à…



Ce que la droite ne prédit pas

- Dans des livres de statistiques vous trouverez la formulation suivante

$$y_i = \beta_0 + \beta_1 x_{i,1} + \in_i$$

b $\quad$ ax

Régression linéaire et moindres carrés

Rappel

La régression linéaire

La durée de sommeil est-elle associée au nombre de mots mémorisés ?

- La variable mesurée dépendante est le ***nombre d'items retenus par le participant***.
- La variable indépendante est le ***nombre d'heures de sommeil que le participant a dormi la nuit dernière***.

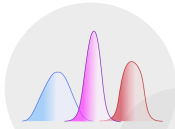
Reprenons la formule

$$y_i = \beta_0 + \beta_1 x_{i,1} + \in_i$$

β_0 Nombre de mots prédits avec **0 h de sommeil** $\beta_0 = 4$

β_1 Nombre de mots rappelés en plus pour chaque heure de sommeil en plus $\beta_1 = 1{,}2$

Donc à chaque heure dormie en plus le participant est censé se rappeler 1,2 mots en plus en moyenne.



Régression linéaire et moindres carrés

Rappel

La régression linéaire

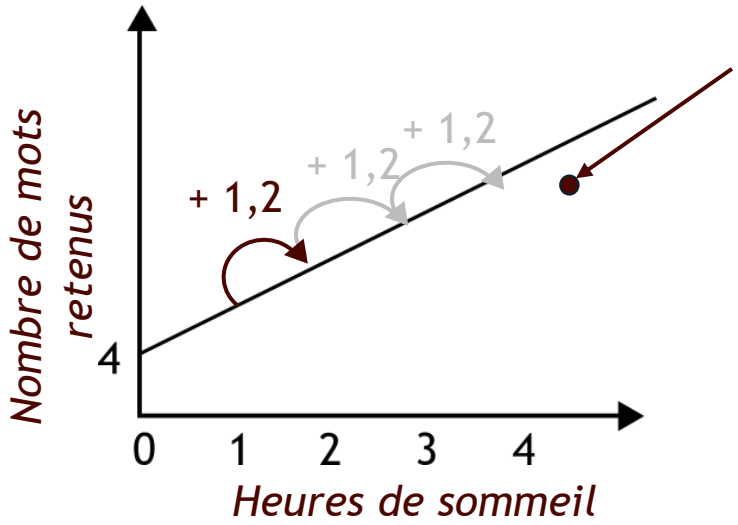
La durée de sommeil est-elle associée au nombre de mots mémorisés ?

$$y_i = \beta_0 + \beta_1 x_{i,1} + \in_i$$

β_0 Nombre de mots prédits avec **0 h de sommeil** $\beta_0 = 4$

β_1 Nombre de mots rappelés en plus pour chaque heure de sommeil en plus $\beta_1 = 1{,}2$

Ainsi le modèle prédit:



Mon premier participant a dormi 4 heures mais n'a retenu que 9 mots alors que mon modèle prévoyait :

$$y_i = 4 + 1{,}2 \times 4 = 9{,}6\ mots$$

L'erreur $\in$ correspond au décalage entre la valeur prédite par mon modèle et la valeur réelle. $\in_i$ correspond aux erreurs cumulées.

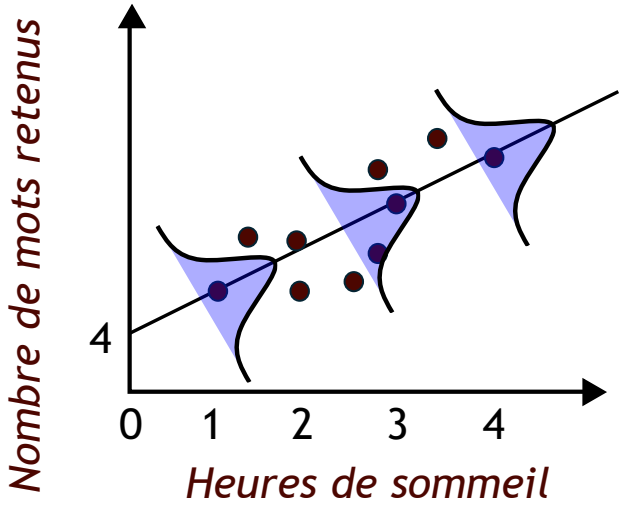
Régression linéaire et moindres carrés

Rappel

La régression linéaire

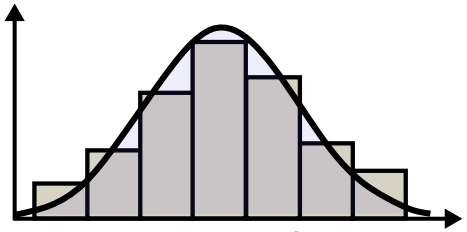
La durée de sommeil est-elle associée au nombre de mots mémorisés ?

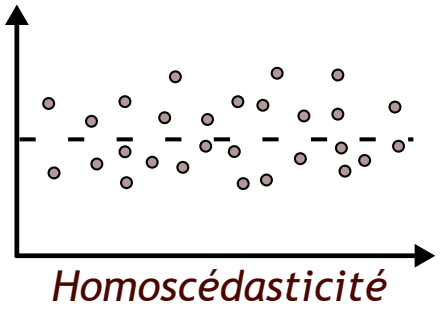
J'ai recruté 10 participants



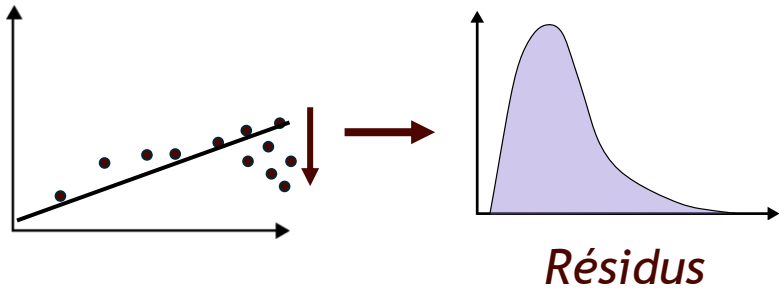
Tester si les données peuvent suivre une relation linéaire.

C'est identifier si l'ensemble des erreurs $\in_i$ respecte deux principes fondamentaux.

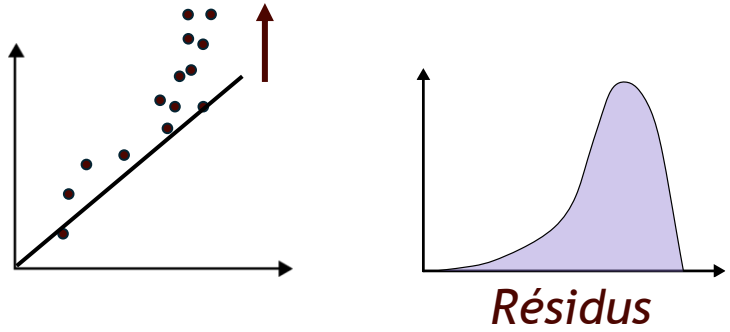




Pourquoi c'est important ?



ou



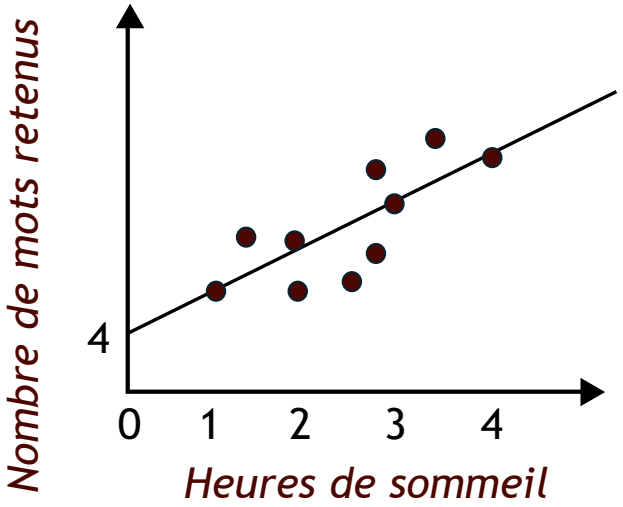
Régression linéaire et moindres carrés

Rappel

La régression linéaire

La durée de sommeil est-elle associée au nombre de mots mémorisés ?

J'ai recruté 10 participants



La relation observée reflète-t-elle une association réelle ou peut-elle s'expliquer par le hasard ?

- *Premièrement, je peux regarder le R^2, pour vérifier la* qualité de prédiction du modèle.

→ Proche de 1 les points sont proches de la droite

Le modèle explique bien les données

→ Proche de 0 les points sont très dispersés autour de la droite *Le modèle explique peu les données*

- Deuxièmement, je peux tester si mon modèle explique significativement mieux les données qu'un modèle sans prédicteur, grâce à une analyse de variance du modèle.

Mais comment fonctionne concrètement une ANOVA ?

Régression linéaire et moindres carrés

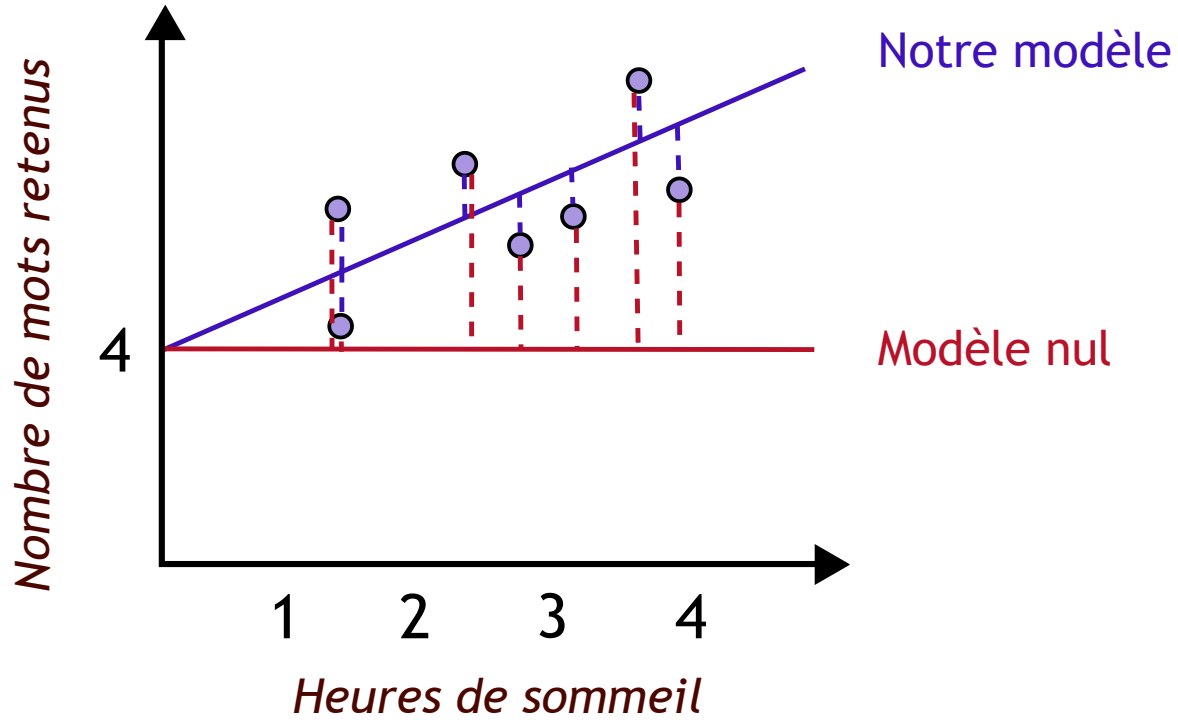
Rappel

L'analyse de variances (ANOVA)?

La durée de sommeil est-elle associée au nombre de mots mémorisés ?

Nos hypothèses:

- **H_0 : $\beta_1 = 0$**, le sommeil n'est pas associé au nombre de mots mémorisés
- **H_1 : $\beta_1 \neq 0$**, le sommeil est associé au nombre de mots mémorisés



L'ANOVA ici est le rapport

$$\frac{Variance\ \beta_1 \neq 0}{Variance\ restante}$$

Plus le F est grand mieux c'est

Si la *p-value < 0,05, nous pouvons rejeter* **H_0**

Régression linéaire et moindres carrés

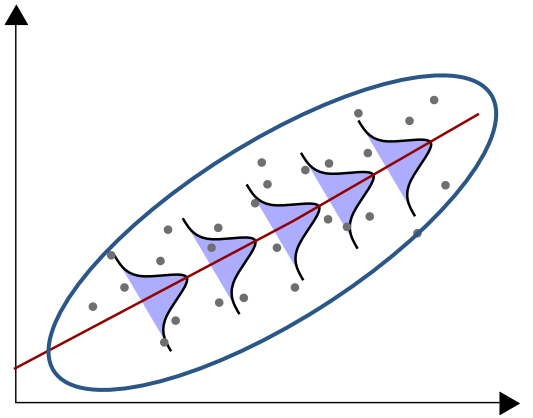
Valeur de y quand $x = 0$ autrement dit ⟶ $y = ax + b$

Prédicteurs

L'erreur ⟶ C'est-à-dire… ou résidu

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k} + \in_i$$

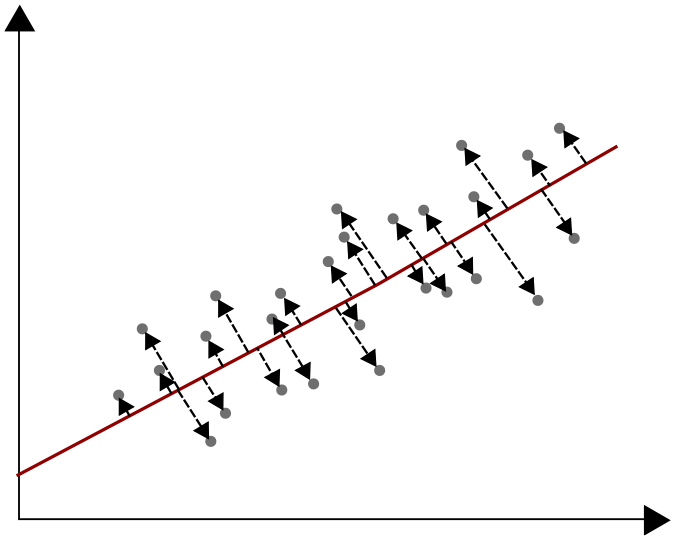
Effets (ou poids) de chaque prédicteur



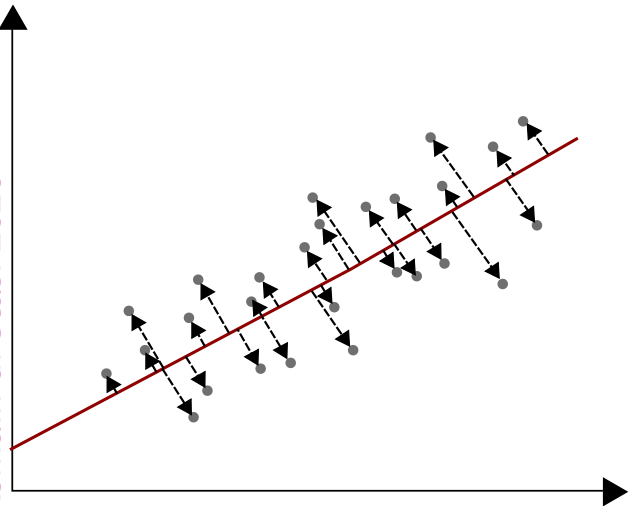
Comment se comporte tes points par rapport à ta droite de régression

L'objectif de toute régression est de bien positionner la droite, comme ici...

C'est là qu'il existe une différence entre statistiques bayésiennes et fréquentistes



Régression linéaire et moindres carrés



Pour placer la droite lorsqu'on fait des analyses fréquentistes on utilise la **méthode des moindre carrés**. L'idée est de **minimiser la somme des carrés des écarts résiduels (SCER)** qui se formalise mathématiquement comme ceci :

$\hat{\mu}_y$ la moyenne des prédiction y des observations y_i

Le paramètre β représente la pente, il peut y en avoir plusieurs dans les régressions multiples.

$$SCER = \sum_i (\hat{y}_i - y_i)^2 = \sum_i (\hat{\mu}_y - y_i + \beta(x_i - \hat{\mu}_x))^2$$

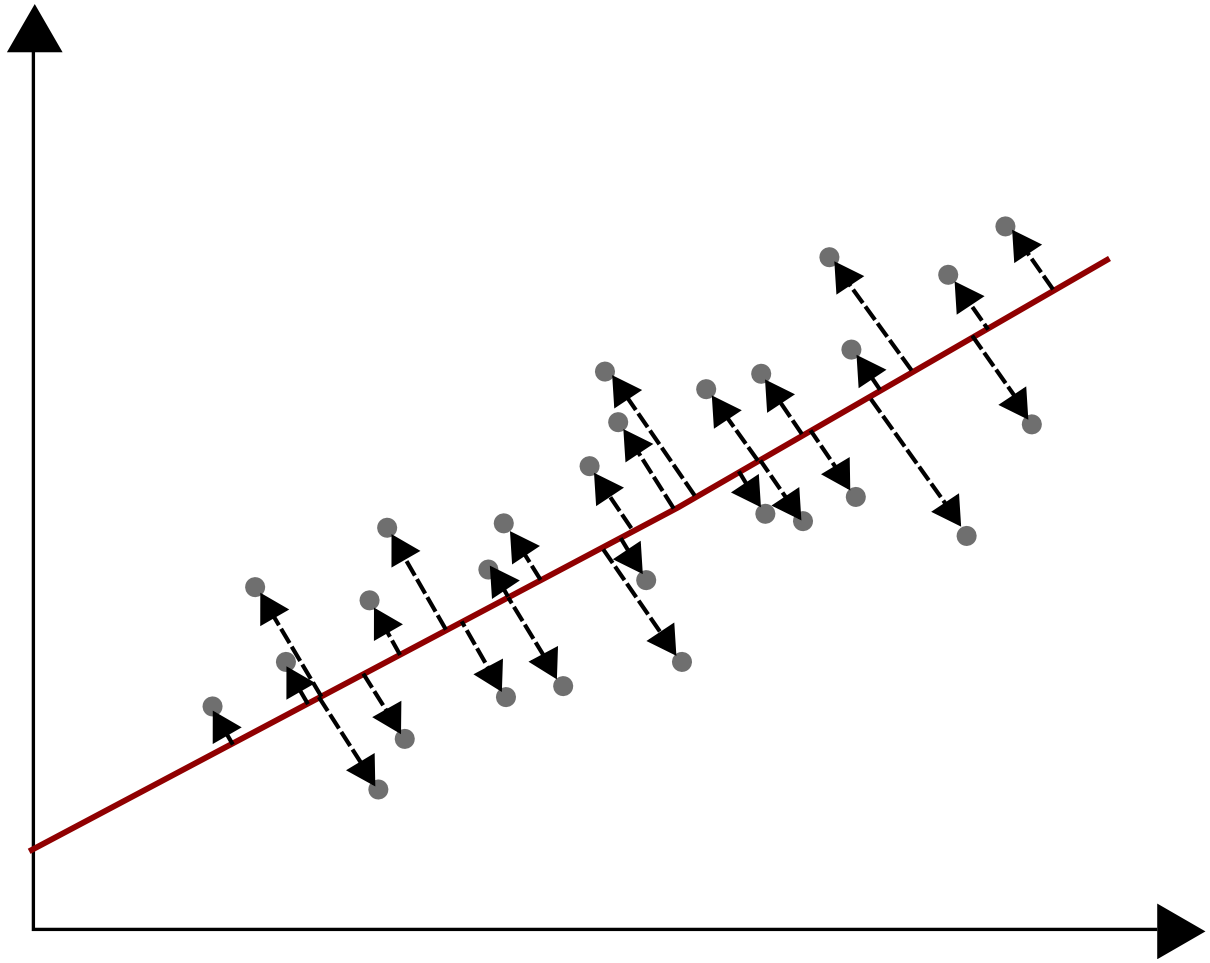
$\hat{y}_i$ représente la valeur prédite pour une observation i

Les observations y_i

Moyennes des x_i

Régression linéaire et moindres carrés

Pour faire simple l'idée est que la somme des carrés des distances des points autour **de sorte à ce que la somme de leurs distance aux carré soit la plus petite possible**.



Maximum de vraisemblance dans une régression linéaire

- On estime pour une observation y qu'elle suit une régression linéaire tel que : $y \sim N(\beta_0 + \beta_1 x, \sigma)$

- La probabilité d'une observation de y est donc :

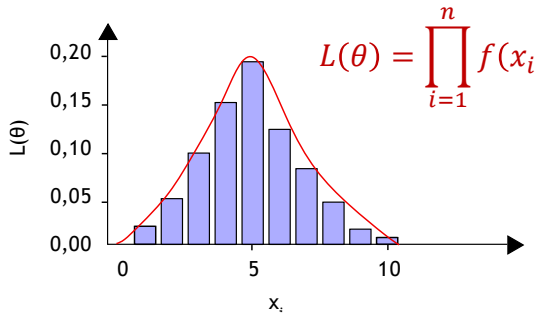
$$f(y|\beta_0, \beta_1, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma}\right)^2}$$

- y est la variable aléatoire dépendante
- x est la variable explicative (ou prédicteur)
- β_0 est l'ordonnée à l'origine (intercept)
- β_1 est le coefficient de régression (pente)
- σ est l'écart type de l'erreur supposé normale (pente)

- Pour n observations indépendantes il faut considérer l'ensemble des probabilités conjointes en utilisant la formule de la fonction de densité $L(\theta)$ que nous avons vue tout à l'heure.

observation ↘ paramètres ↘

$$\underbrace{L(\beta_0, \beta_1, \sigma)}_{L(\theta)} = f(y_1, \dots, y_n | \underbrace{\beta_0, \beta_1, \sigma}_{\theta}) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma}\right)^2}}_{f(x_i|\theta)}$$



Maximum de vraisemblance dans une régression linéaire

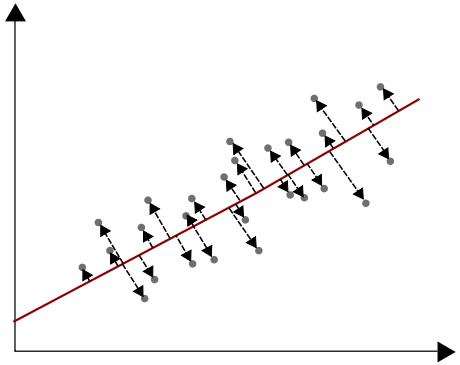
- Pour simplifier tous cela on fait souvent le logarithme de cette fonction

$$L(\beta_0, \beta_1, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\beta_0-\beta_1 x}{\sigma}\right)^2}$$

$$l(\beta_0, \beta_1, \sigma) = \sum_{i=1}^{n} \left(\log \left(\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{y_i - \beta_0 - \beta_1 x_i}{\sigma}\right)^2\right)\right)$$

$$l(\beta_0, \beta_1, \sigma) = n \log \left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_{1x_i})^2$$

Somme des écarts aux carré



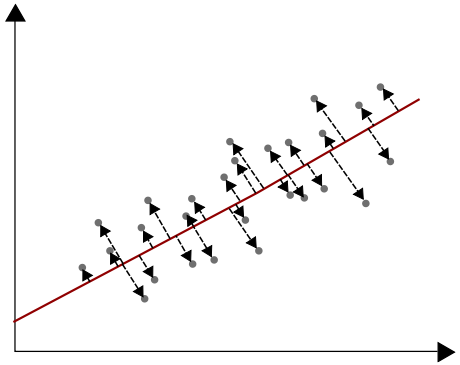
Comme elle soustrait, plus cette somme est petite plus la vraisemblance sera grande

Maximum de vraisemblance dans une régression linéaire

- Pour simplifier tous cela on fait souvent le logarithme de cette fonction

$$l(\beta_0, \beta_1, \sigma) = n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_{1x_i})^2$$

Somme des écarts aux carré



Comme elle soustrait, plus cette somme est petite plus la vraisemblance sera grande

En **régression linéaire classique**, lorsque les erreurs sont supposées normales, indépendantes et de variance constante, maximiser la vraisemblance revient exactement à minimiser la somme des carrés des résidus. C'est cette hypothèse de normalité qui fait apparaître la somme des carrés dans la log-vraisemblance. Ainsi, la méthode des moindres carrés est un cas particulier du maximum de vraisemblance. Le MLE apporte en plus une interprétation probabiliste et permet de généraliser l'approche à des modèles où l'erreur n'est pas normale (Poisson, binomiale, Gamma, etc.).

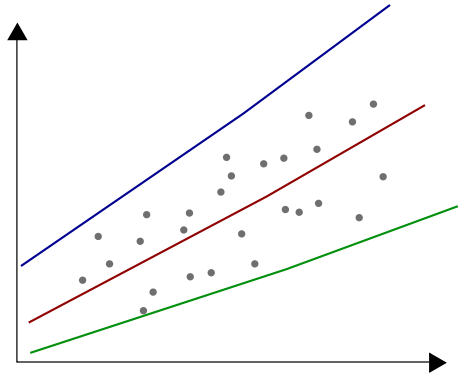
Maximum de vraisemblance dans une régression

- Concrètement que fait le maximum de vraisemblance par étapes :

Etape 1: Ce que l'on cherche

On a des points dispersés sur un graphique (valeurs de x et y) — par exemple, taille de l'enfant et nombre de mots qu'il connaît. On veut **trouver une droite** qui **colle le mieux** à ces points.

Etape 2: On essaye des droites



Certaines passent trop en dessous ou au-dessus des points, d'autres passent **pile au bon endroit** : c'est **la meilleure droite**.

Maximum de vraisemblance dans une régression

- Concrètement que fais le maximum de vraisemblance par étapes :

***Etape 3*: Une idée de "chance"**

Pour chaque droite, on peut se demander : **“À quel point cette droite rend mes données probables ?”**

Autrement dit : si cette droite était la vraie, **à quel point serait-ce normal d’observer ces points aussi proches (ou éloignés)** ?

***Au final* la méthode du maximum de vraisemblance**

C’est une méthode qui :

- **Teste plein de droites différentes.**
- **Mesure pour chacune** la "probabilité d’avoir obtenu ces données".
- **Choisit la droite qui rend les données les plus probables.**

Vraisemblance

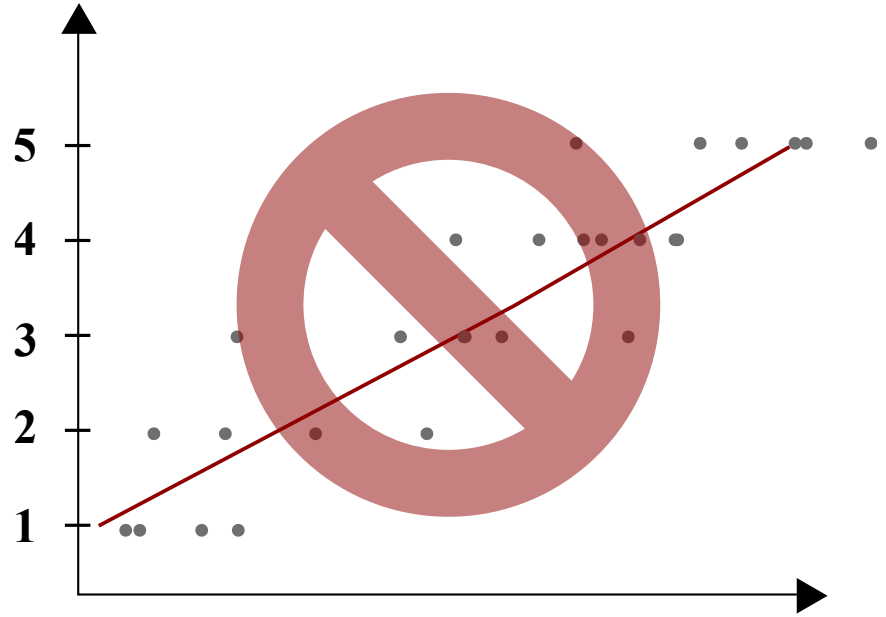
- Quelle application ont ces deux méthodes : **Moindre carrée** *vs.* **Maximum de vraisemblance.**

Moindre carrée: est utilisé lorsque vous êtes dans le cadre d'une régression linéaire classique

→ Ce qui veut dire que vos résidus suivent une loi normal centrée autour de la droite

→ Qu'ils sont homoscédastique (i.e., homogénéité de la variances)

→ Que votre Variable dépendante contiens au minimum 20 modalités possibles.



→ *Cumulative link models* ('clm') du package « ordinal » (Christensen, 2023)

<u>Puis</u> *anova.clm* du package « RVaideMêmoire » (Herve 2011).

Vraisemblance

- Quelle application ont ces deux méthodes : **Moindre carrée** *vs.* **Maximum de vraisemblance.**

Maximum de vraisemblance : méthode couramment utilisée pour estimer les paramètres d'une régression linéaire généralisée (mais pas exclusivement). Elle consiste à ajuster les données en choisissant la loi de probabilité appropriée pour les résidus (e.g., binomiale, Poisson, Gamma, *etc.*).

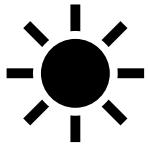
Si l'on adopte une approche plus proche de la philosophie bayésienne, puisque le maximum de vraisemblance restreint l'espace dans lequel notre modèle va chercher les paramètres à l'image du *prior*. Par exemple, un temps de réaction de 1000 secondes n'est pas réaliste ; de même, une loi Gamma n'ira pas explorer des valeurs aberrantes de ce type.

L'approche bayésienne pousse cette logique encore plus loin : au lieu de chercher une seule "meilleure valeur" des paramètres, elle considère toutes les valeurs possibles et indique leur degré de plausibilité (en combinant les données et nos croyances préalables, appelées *priors*). Comme il est souvent trop compliqué de calculer cette distribution directement, on utilise des **simulations par chaînes de Markov (MCMC)** voir ici pour un exemple, qui permettent d'explorer progressivement l'espace des paramètres. On obtient ainsi non pas un seul chiffre, mais une distribution complète de possibilités.

Notion de Chaîne de Markov

- L'idée clé : le futur ne dépend pas du présent, mais du chemin qui l'y a mené.

Un exemple très simple la météo : on imagine un cas très simple de météo ou il n'y a que deux possibilités :



On dit qu'aujourd'hui il fait soleil. Demain selon notre modèle il y a donc 70 % de chance qu'il fasse beau et 30 % de chance qu'il pleuve. Et si aujourd'hui il pleut alors : il y a 60 % de chance qu'il continue de pleuvoir et 40 % de chance que le soleil revienne.

Ce modèle est un **Chaîne de Markov**:

- Tu passes d'un état (temps) à un autre
- Avec certaines probabilités de transition
- Et à chaque étape, seul l'état actuel compte pour prédire le suivant

Notion de Chaîne de Markov

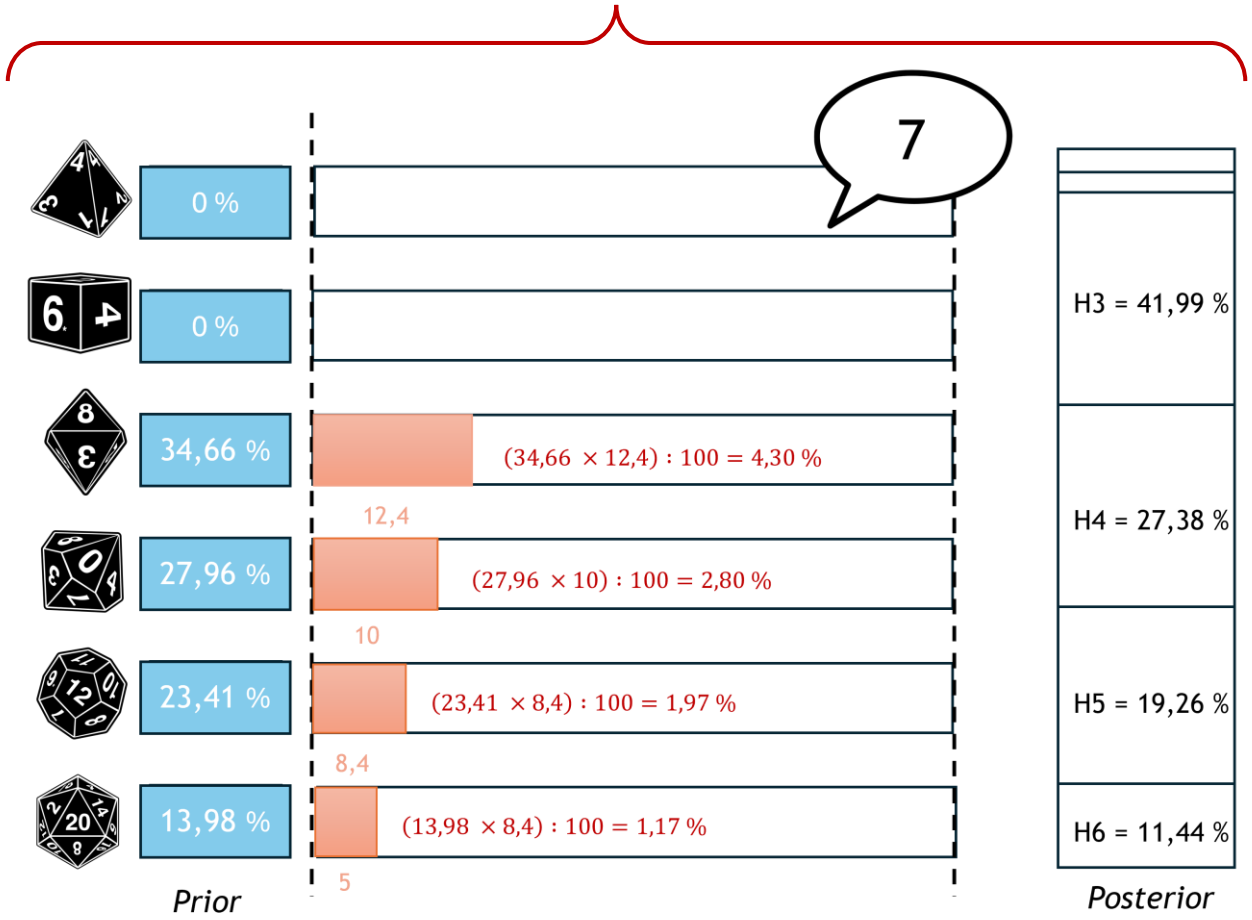
Une chaîne de Markov, c’est comme un jeu

Imagine que tu jettes un dé ou que tu avances sur un jeu de plateau : Ton **prochain mouvement dépend de là où tu es maintenant**. Tu n’as pas besoin de te rappeler comment tu es arrivé là.

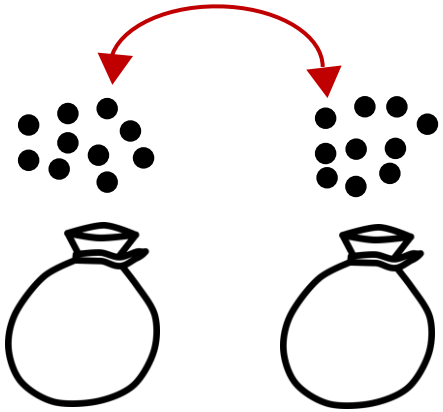
Une **chaîne de Markov**, c’est une suite d’étapes où chaque nouvelle étape **dépend uniquement de la précédente**, pas de toute l’histoire. En statistiques, on l’utilise pour **simuler** plein de scénarios possibles et **explorer** les valeurs probables d’un paramètre.

Notion de Chaîne de Markov

Va répéter ça x fois en resimulant les données et comparer nos données observées à nos données simulées pour voir quelle est l'hypothèse la plus vraisemblable.

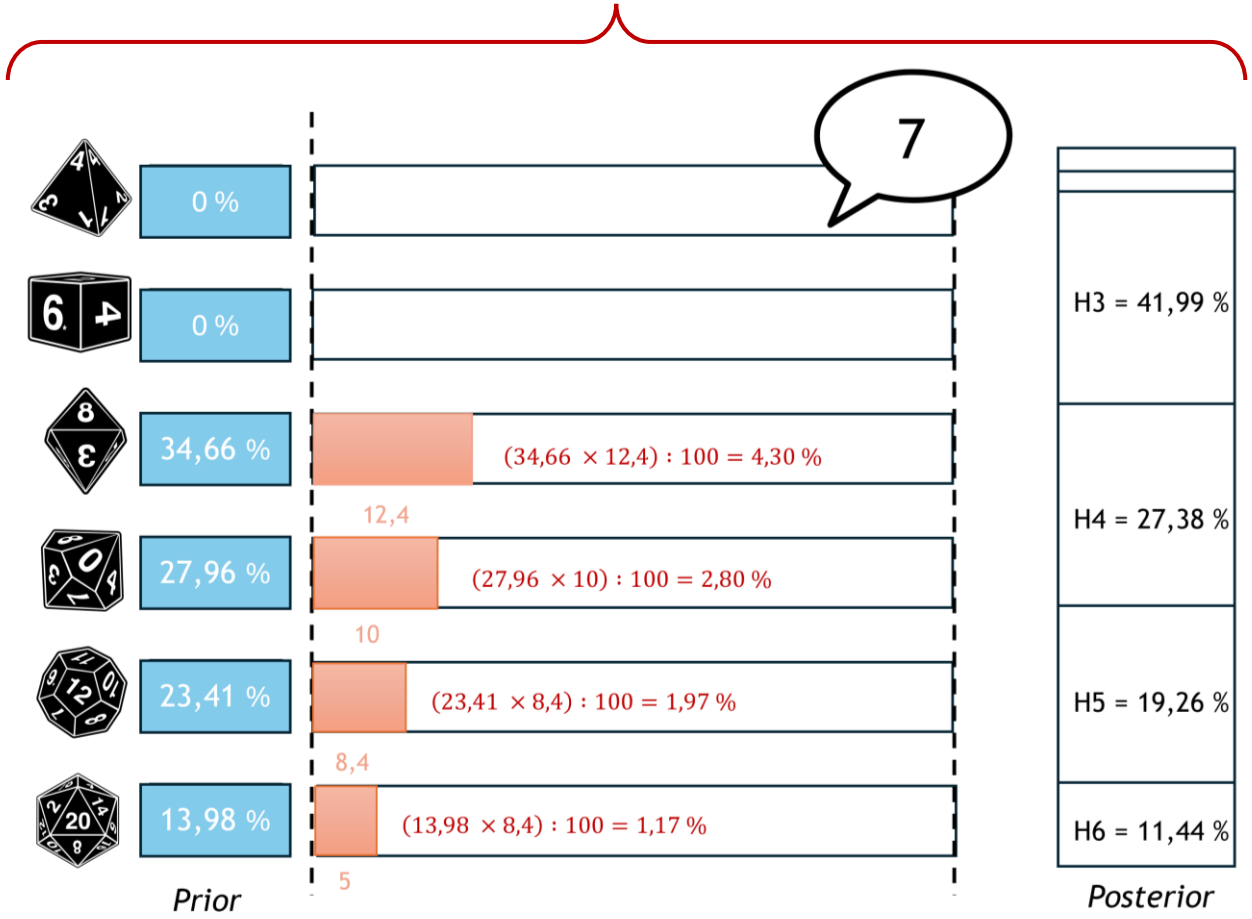


Contrairement aux tests de permutations qui vont ***redistribuer aléatoirement*** les données dans les différentes conditions comme illustré ici avec les deux sacs de billes…

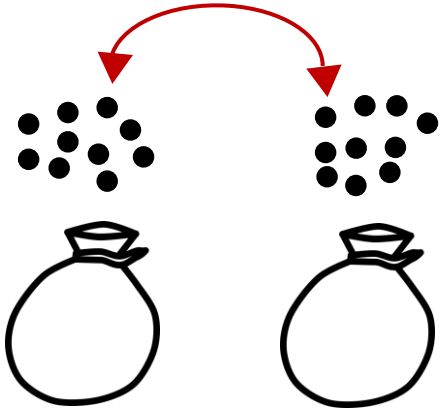


La chaîne de Markov ne modifie pas les données, elle applique différentes hypothèses et calcule leur vraisemblance.

Notion de Chaîne de Markov



Contrairement aux test de permutations qui vont ***redistribuer aléatoirement*** les données dans les différentes conditions comme illustré ici avec les deux sacs de billes…



La chaîne de Markov ne modifie pas les données, elle applique différentes hypothèses et calcule leur vraisemblance.

Une chaîne de Markov propose une hypothèse (ex : un dé à 8 faces), puis envisage d'en tester une autre (ex : un dé à 10 faces), et **accepte ou rejette** la nouvelle hypothèse selon la vraisemblance.

Critique de la *p-value* et Facteur Bayésien

- Je suis enquêteur et je veux savoir combien j'ai de personnes dopées chez mes 19 premiers coureurs.

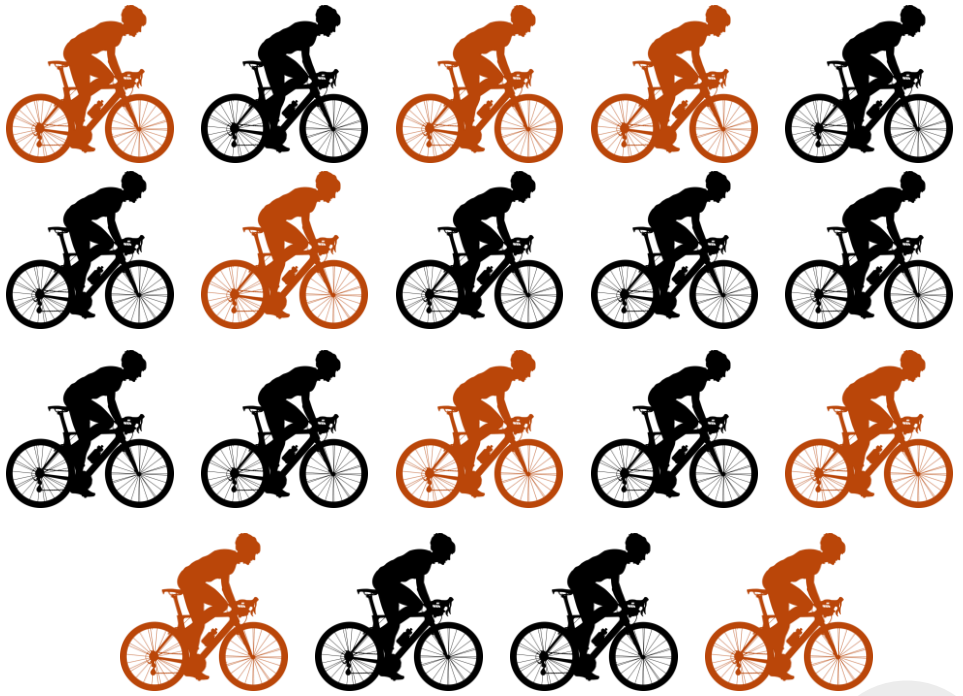
  → Je sais que l'un des principaux symptômes de dopages est un taux d'hématocrite (i.e., quantité de globules rouges dans le sang) particulièrement élevé – 50%

- 8 sur les 19 cyclistes sont testés avec un taux élevé.

  → Donc 42 % de mon échantillon

- On sait que dans la population générale seul 13 % des sujets ont un hématocrite

  → L'idées est de calculé la probabilité qu'on ait $Prop_{observé} = 42\ \%$ alors que $Prop_{attendue} = 13\ \%$



© exemple tiré de Thierry Ancelle

Test du chi deux *Ou* *p-value*

Loi binomiale

Critique de la *p-value* et Facteur Bayésien

- 8 sur les 19 cyclistes sont testés avec un taux élevé.

  → Donc 42 % de mon échantillon

- On sait que dans la population générale seul 13 % des sujets ont un hématocrite

  → L'idée est de calculer la probabilité qu'on ait $Prop_{observé} = 42\ \%$ alors que $Prop_{attendue} = 13\ \%$

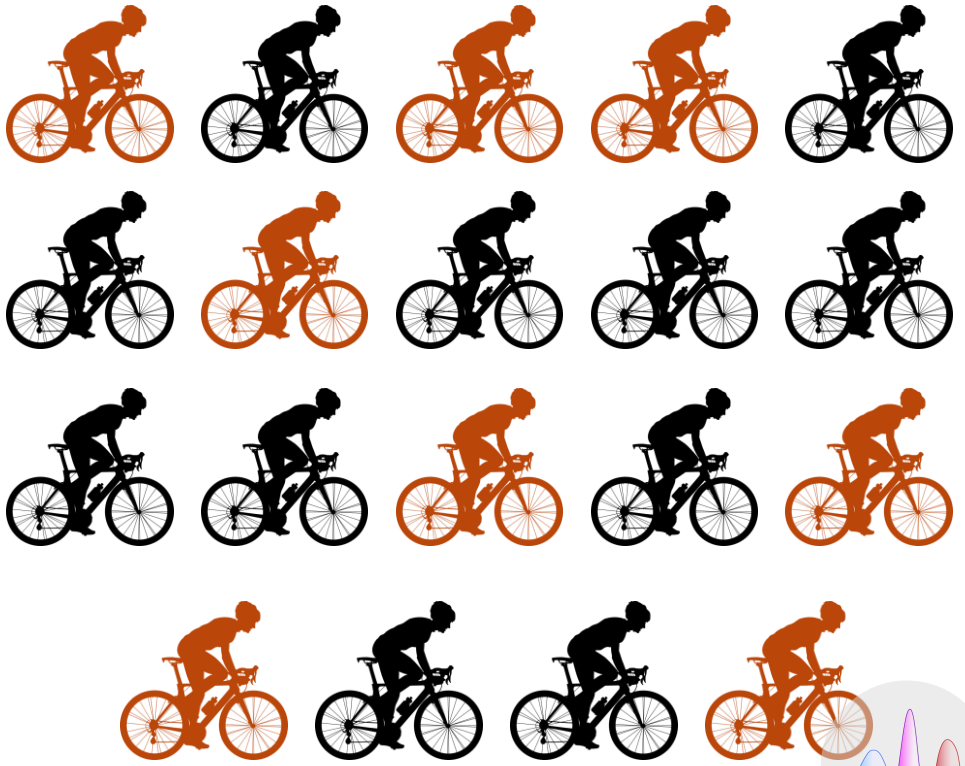
$$H_0 : Prop_{observe} = Prop_{attendu}$$

$$H_1 : Prop_{observe} > Prop_{attendu}$$

$$p = 0{,}002$$

- Mais ce n'est qu'un symptôme. Il y a d'autres raisons pouvant expliquer un taux élevé d'hématocrite.

  → Un séjour à la montagne



© exemple tiré de Thierry Ancelle

Critique de la *p-value* et Facteur Bayésien

- 8 sur les 19 cyclistes sont testés avec un taux élevé (> 50%).

  → On trouve une étude qui montre que lorsque les individus consomment un produit X, 50% des participants ayant consommé ce produit dopant montent à 50%

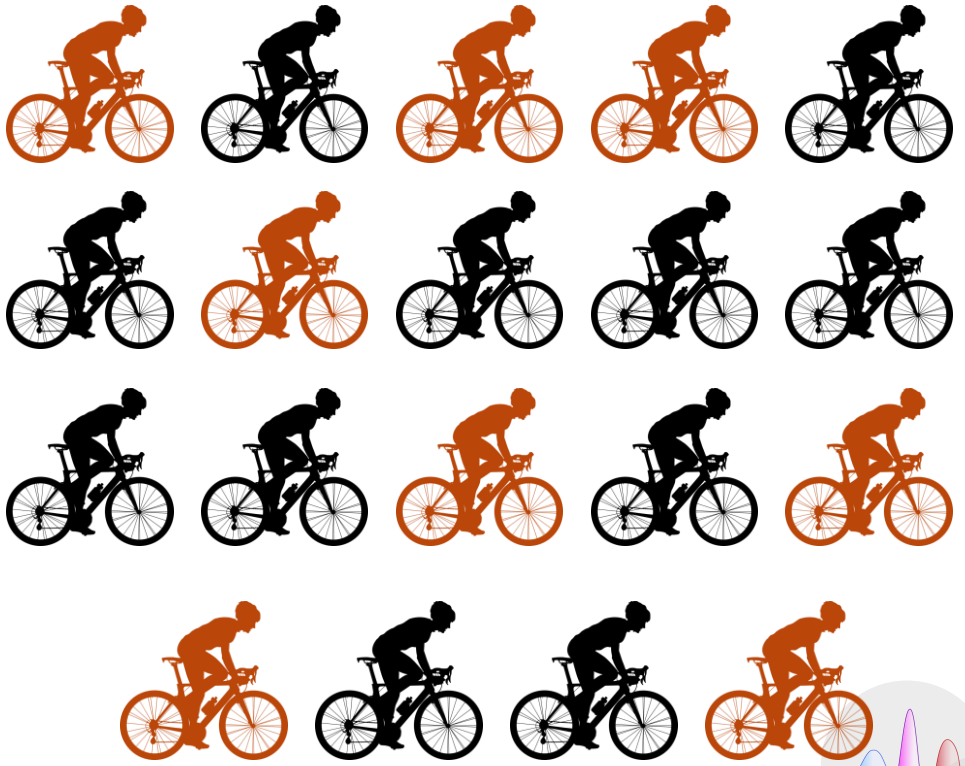
- On peut donc réaliser un autre test en évaluant le pourcentage de personnes ayant un taux d'hématocrite supérieur à 50 % par rapport à une probabilité attendue de 50 % si 100 % des 19 coureurs avait consommé un produit dopant.

  → Ici on aurait $Prop_{observé} = 42\ \%$ alors que $Prop_{attendue} = 50\ \%$

$$H_0 : Prop_{observe} = Prop_{attendu}$$

$$H_1 : Prop_{observe} > Prop_{attendu}$$

$$p = 0{,}32\ ns$$



© exemple tiré de Thierry Ancelle

⚠ Ici on est très ennuyé puisque on ne peux ni accepter H_0 ni la rejeter…

Critique de la *p-value* et Facteur Bayésien

- Pour résumer on a deux études
  - → La première montre un résultat significatif mais n'est pas claire dans H1, on ne sait pas combien sont dopés, on a aucune probabilité de trouver un individu dopé dans la population
  - → La seconde ne permet pas de conclure…

- Si on part de l'hypothèse que la population de cycliste étudiée est une population dopée en utilisant une approche bayésienne basée sur la vraisemblance on aura…

$$V(H_{dopage}) = \Pr(D \mid H_{dopage}) = \Pr(8/19)\ si\ P = 0{,}50$$

Puisque pour rappel la formule de la loi binomiale est :

$$\Pr(X = K) = \frac{n!}{k!\,(n-k)!} p^k (1-P)^{n-k}$$

$n = 19$ $k = 8$ $P = 0{,}5$

$$V(H_{dopage}) = 0{,}144$$

Attention la vraisemblance n'est pas un probabilité et ne s'interprète sûrement pas comme la probabilité d'observer cet échantillon si mon hypothèse est vraie. Elle n'est pas suffisante seule.

Critique de la *p-value* et Facteur Bayésien

- On peut faire l'hypothèse inverse en mesurant la vraisemblance des données selon l'hypothèse que mon échantillon appartienne à ma population normale.

$$V(H_{clean}) = \Pr(D \mid H_{clean}) = \Pr(8/19) \; si \; P = 0{,}13$$

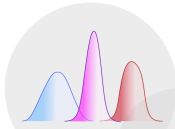
$$V(H_{clean}) = 0{,}00133$$

- On voit bien que nos deux vraisemblances

$$\begin{cases} V(H_{dopage}) = 0{,}144 \\ V(H_{clean}) = 0{,}00133 \end{cases} \rightarrow \frac{V(H_{dopage}) = 0{,}144}{V(H_{clean}) = 0{,}00133} = 108{,}27 \approx 108$$

- Cela signifie que maintenant qu'on a recueilli les données la vraisemblance de l'hypothèse dopage (H_{dopage}) est 108 fois plus forte que celle d'une absence de dopage (H_{clean}).

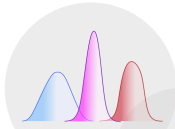
→ C'est ce qu'on appelle **le Facteur de Bayes :** plus il est **élevé,** plus l'hypothèse est **vraisemblable.**



Critique de la *p-value* et Facteur Bayésien

- Juste pour être sûr : si j'ai un facteur de Bayes de 2, que cela signifie-t-il ?

$$\frac{\mathrm{P}(\mathrm{D}\,|H_1)}{\mathrm{P}(\mathrm{D}\,|H_2)} = 2 \quad \longrightarrow \quad \text{Le résultat est deux fois plus vraisemblable sous } H_1$$

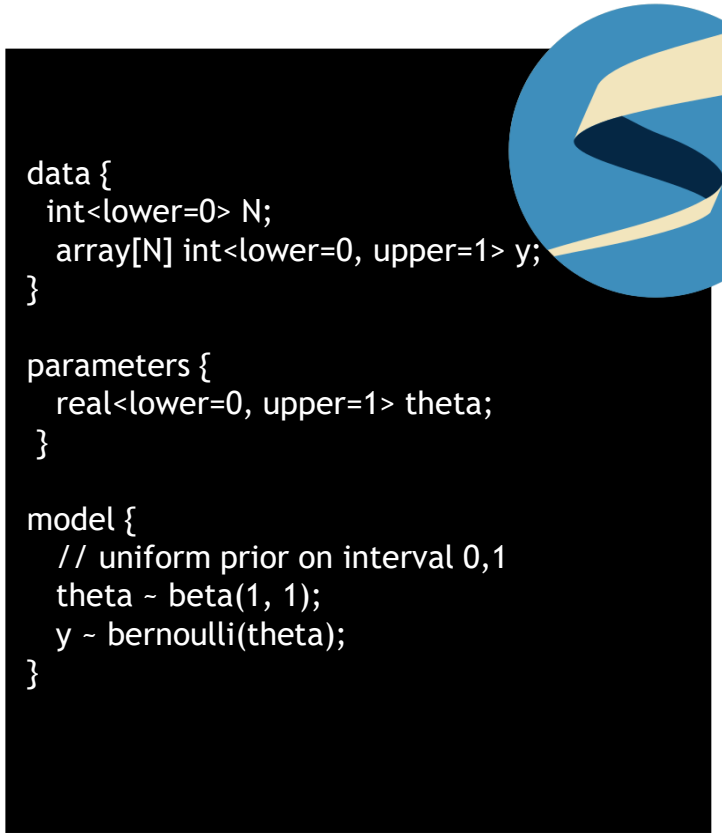


Partie Pratique utilisation du package brms

- Le langage privilégié pour les analyses bayésiennes est Stan. Heureusement, il est possible d'utiliser le package *brms*, qui fait l'interface entre R et Stan (installé localement sur l'ordinateur), ce qui facilite grandement l'écriture et l'estimation des modèles.

- *Si brms facilite l'écriture de modèles bayésiens grâce à une syntaxe proche de lme4, il n'est pas le seul outil disponible : d'autres packages comme Rstan permettent également d'interfacer R avec Stan.*

- *brms présente l'avantage de bénéficier d'une large communauté active, notamment sur GitHub, et de nombreux tutoriels sont disponibles en ligne. Je vous recommande tout particulièrement celui proposé par Our coding club !*



Example de Stan

Installation à faire sur son ordinateur !

R Version 4.4.2

IDE recommandées*

Vs Code

** Integrated Development Environement*

Installation des packages

```
Install.packages('brms', dependencies = TRUE) # pour réaliser des modèles bayésiens
Install.packages('emmeans') # pour les comparaisons post-hoc
Install.packages('ggplot2') # pour produire les graphes
```

Charger les librairies

```
library(brms)
library(emmeans)
library(ggplot2)
```

Importer vos données dans R

```
setwd("../Data_for_bayes")
Data Bayes = read.table(file = "yourdata.csv", header= TRUE, sep= "," , dec= ".  , na.string = c("", "NAN"),
stringsAsFactors = TRUE)
```

Modèle Bayésien – régression linéaire simple

Modèle brms

```
Model_brm <- brms::brm(y~ x,
data = data, family = poisson(), chains = 3,
iter = 3000, warmup = 1000)

# saveRDS(Model_brm, “Model_brm.RDS")
# you can save the model as an RDS (Rdata) that way you don't need to run the model again if you come back to this code
```

- L’argument « ***Chains = 3*** » veut dire « fais 3 chaînes de simulation différentes pour bien explorer toutes les possibilités »

- L’argument « **iter = 3000, warmup = 1000** » veut dire « Simule 3000 itérations, dont 1000 pour l’échauffement ». Le modèle explore plein d’hypothèses, et :

  - 3000 = nombre total de simulations,
  - 1000 = les premières sont un “échauffement” pour que le modèle trouve une zone stable (elles ne sont pas gardées pour l’analyse)

Modèle Bayésien – régression linéaire simple

Modèle brms

```
Model_brm <- brms::brm(y~ x,
data = data, family = poisson(), chains = 3,
iter = 3000, warmup = 1000)

# saveRDS(Model_brm, “Model_brm.RDS")
# you can save the model as an RDS (Rdata) that way you don't need to run the model again if you come back to this code
```

Ce code crée un **modèle bayésien** pour déterminer si la variable x prédit la variable y, qui représente un nombre d’événements. Nous utilisons ici une **distribution de Poisson**, qui ne prend en compte que des entiers naturels.

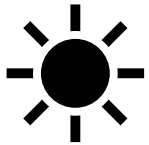
Modèle Bayésien – régression linéaire simple

Modèle brms – notion de Chaîne de Markov

Rappel

- L'idée clé : le futur ne dépend pas du présent, mais du chemin qui l'y a mené.

Un exemple très simple la météo : on imagine un cas très simple de météo ou il n'y a que deux possibilités :



On dit qu'aujourd'hui il fait soleil. Demain selon notre modèle il y a donc 70 % de chance qu'il fasse beau et 30 % de chance qu'il pleuve. Et si aujourd'hui il pleut alors : il y a 60 % de chance qu'il continue de pleuvoir et 40 % de chance que le soleil revienne.

Ce modèle est un **Chaîne de Markov**:

- Tu passes d'un état (temps) à un autre
- Avec certaines probabilités de transition
- Et à chaque étape, seul l'état actuel compte pour prédire le suivant

Modèle Bayésien – régression linéaire simple

Modèle brms – notion de Chaîne de Markov

Rappel

Une chaîne de Markov, c’est comme un jeu

Imagine que tu jettes un dé ou que tu avances sur un jeu de plateau : Ton **prochain mouvement dépend de là où tu es maintenant**. Tu n’as pas besoin de te rappeler comment tu es arrivé là.

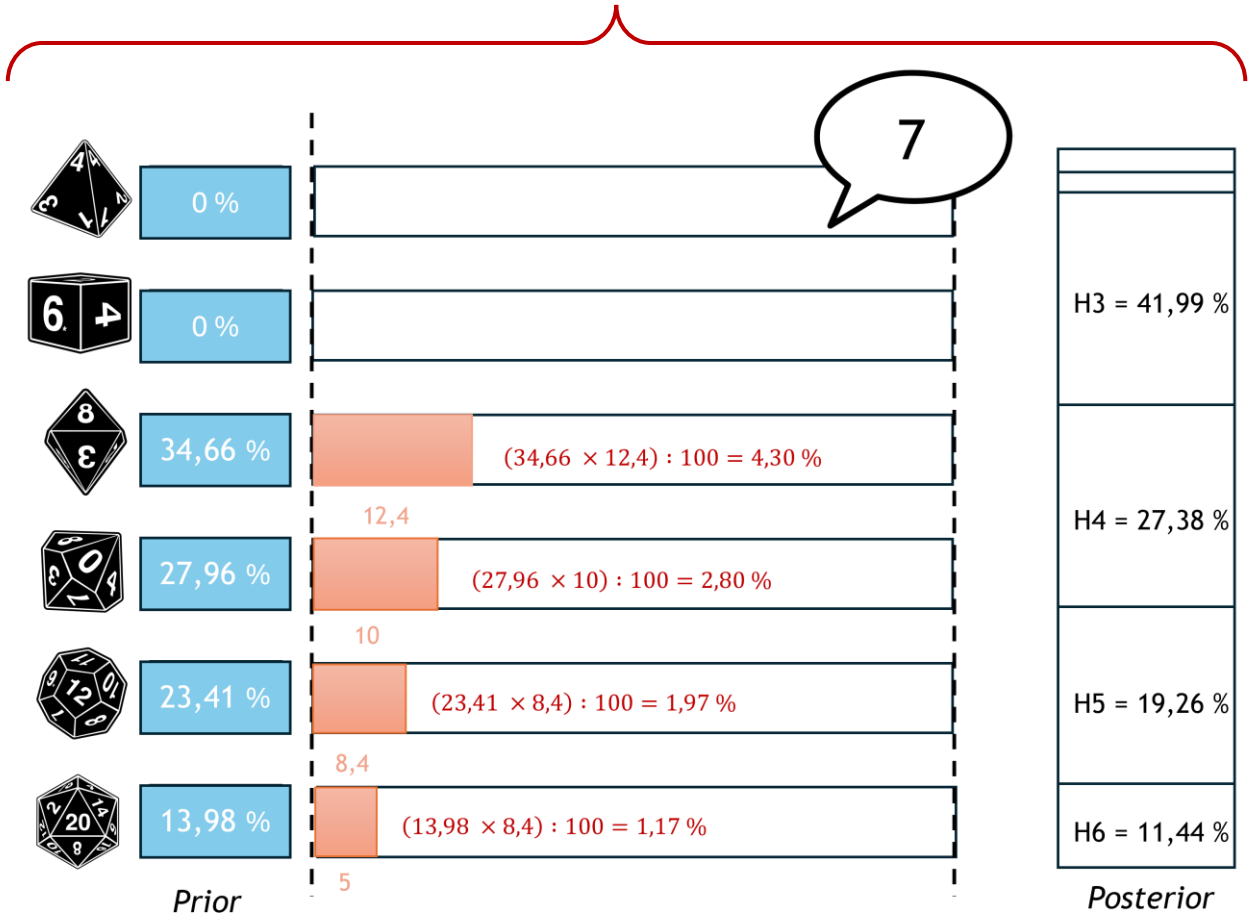
Une **chaîne de Markov**, c’est une suite d’étapes où chaque nouvelle étape **dépend uniquement de la précédente**, pas de toute l’histoire. En statistiques, on l’utilise pour **simuler** plein de scénarios possibles et **explorer** les valeurs probables d’un paramètre.

Modèle Bayésien – régression linéaire simple

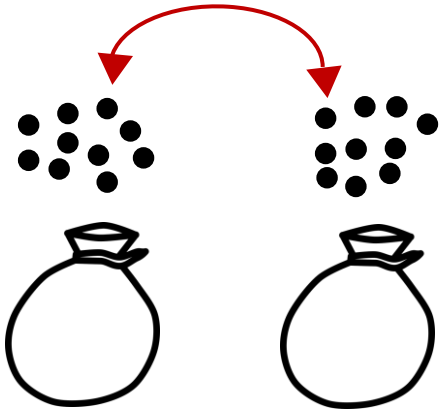
Modèle brms – notion de Chaîne de Markov

Rappel

Va répéter ça x fois en resimulant les données et comparer nos données observées à nos données simulées pour voir quelle est l'hypothèse la plus vraisemblable.



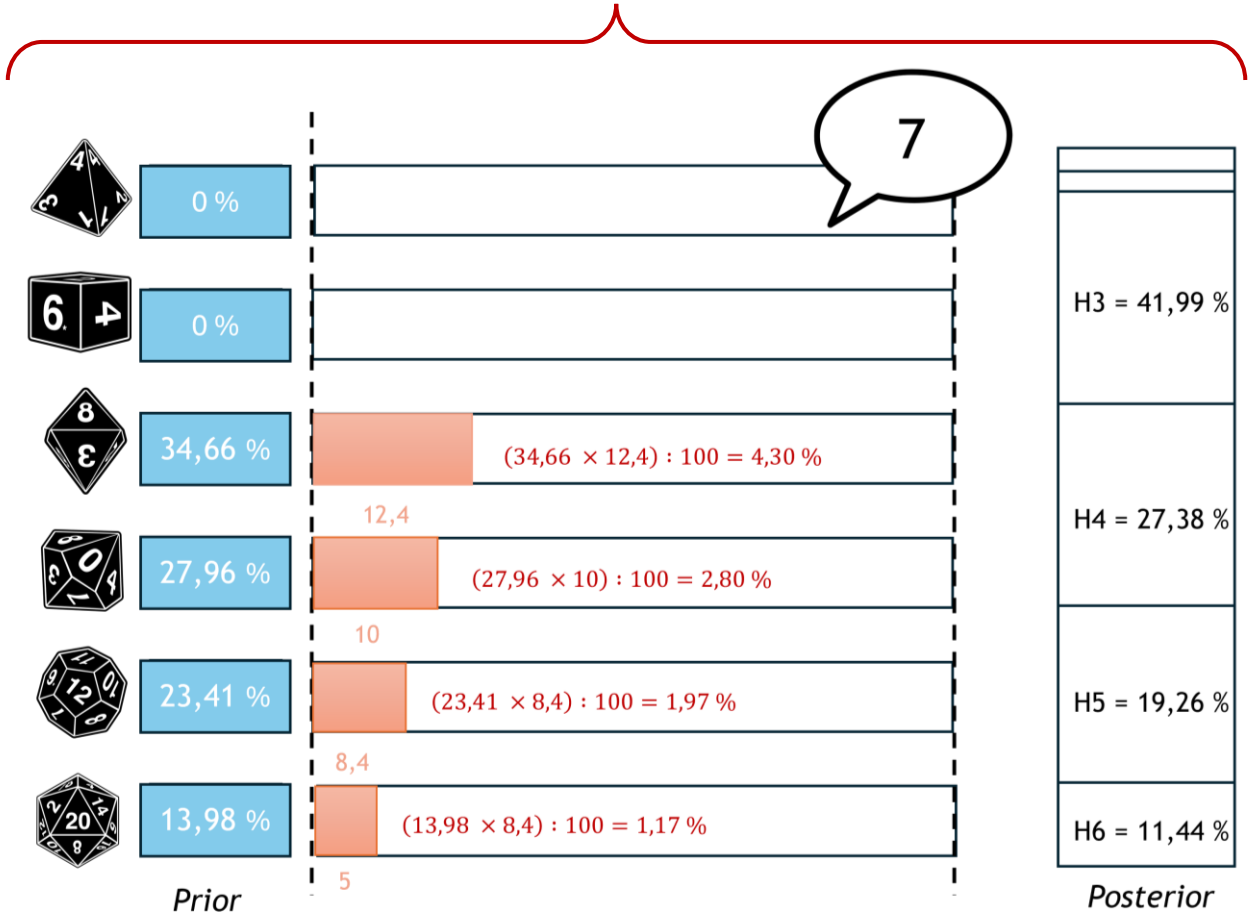
Contrairement aux tests de permutations qui vont ***redistribuer aléatoirement*** les données dans les différentes conditions comme illustré ici avec les deux sacs de billes…



La chaîne de Markov ne modifie pas les données, elle applique différentes hypothèses et calcule leur vraisemblance.

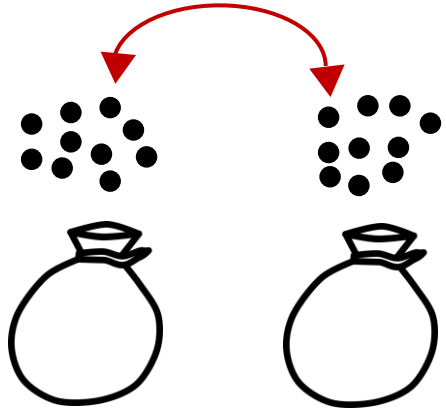
Modèle Bayésien – régression linéaire simple

Modèle brms – notion de Chaîne de Markov



Rappel

Contrairement aux test de permutations qui vont ***redistribuer aléatoirement*** les données dans les différentes conditions comme illustré ici avec les deux sacs de billes…



La Chaîne de Markov ne change rien aux données, colles nos différentes hypothèses dessus et calculs sa vraisemblance.

Une chaîne de Markov propose une hypothèse (ex : un dé à 8 faces), puis envisage d'en tester une autre (ex : un dé à 10 faces), et **accepte ou rejette** la nouvelle hypothèse selon la vraisemblance.

Modèle Bayésien – régression linéaire simple

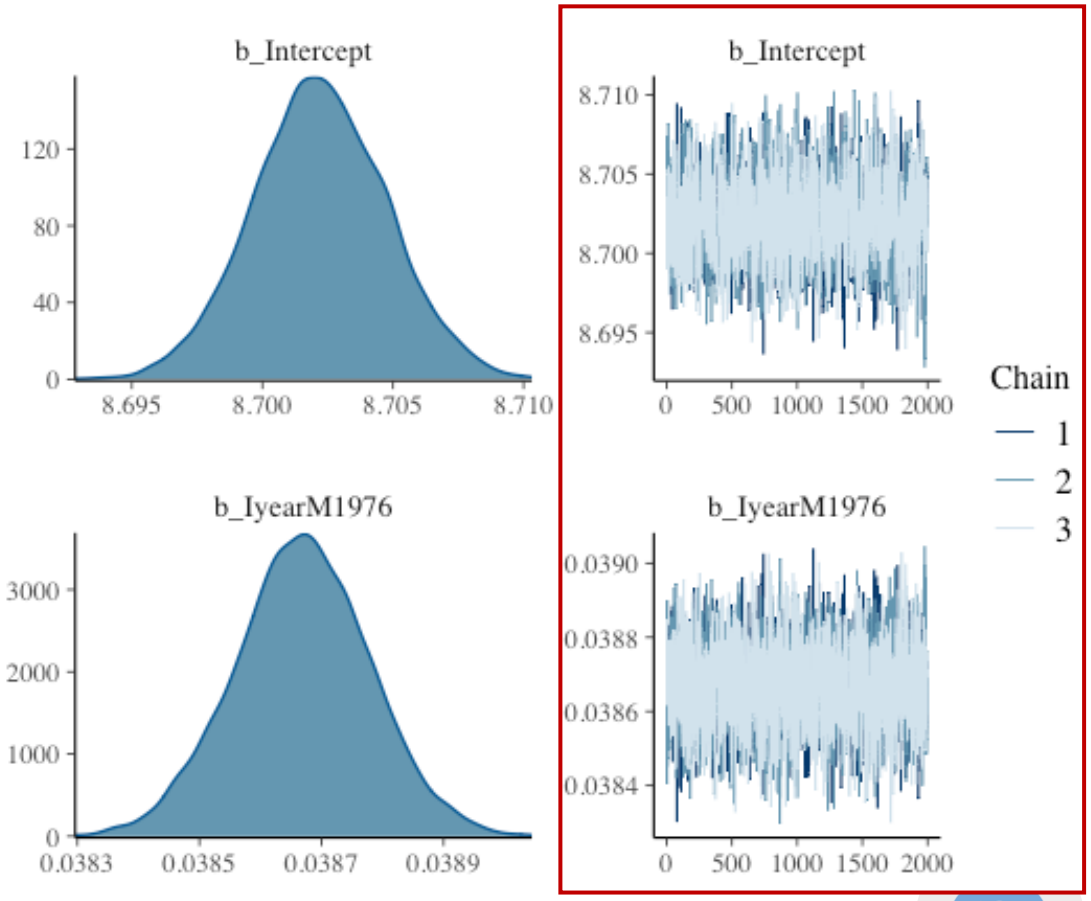
Modèle brms – notion de Chaîne de Markov

Attention une Chaîne de Markov doit absolument tester toutes les possibilités. Pour le tester rien de plus simple.

```
plot(Model_brm)
```

Les graphiques de type *Caterpillar* doivent être aussi détaillés que possible…

Plus ils le sont plus cela veut dire que votre chaîne de Markov explore l'ensemble des possibilités.



Modèle Bayésien – régression linéaire simple

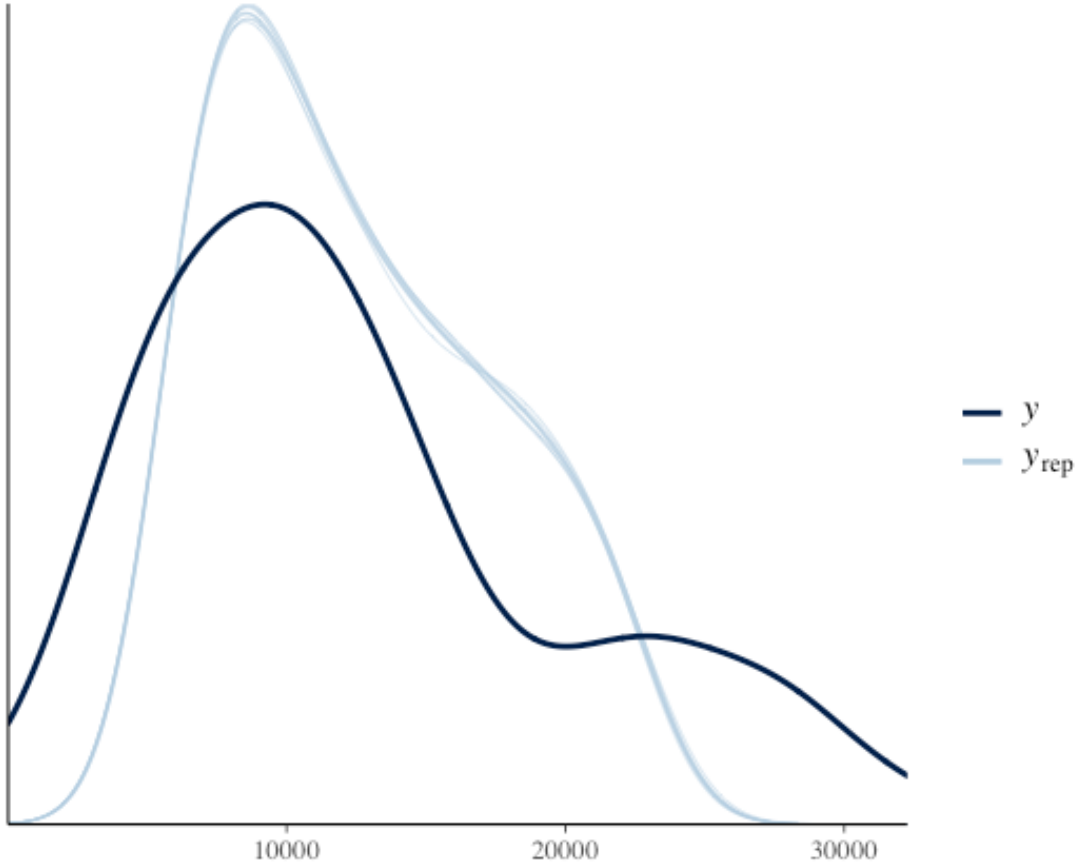
Modèle brms – notion de Chaîne de Markov

Attention une Chaîne de Markov doit absolument tester toutes les possibilités. Pour le tester rien de plus simple.

```
pp_check(Model_brm, ndraws = 10)
```

10 jeux de données simulées par les Chaînes de Markov dans le modèle (i.e., distribution théorique)

Distribution observée

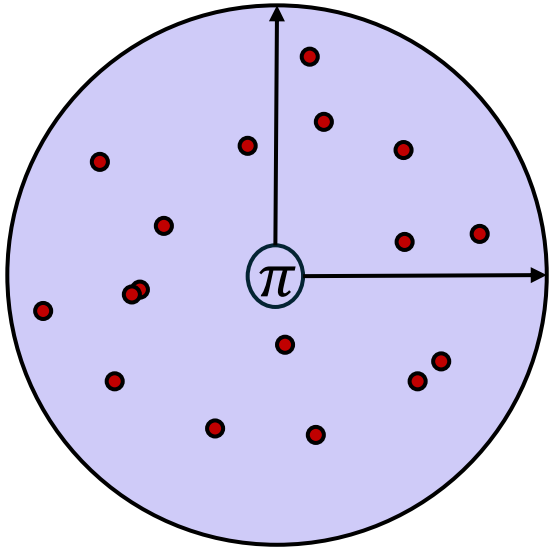


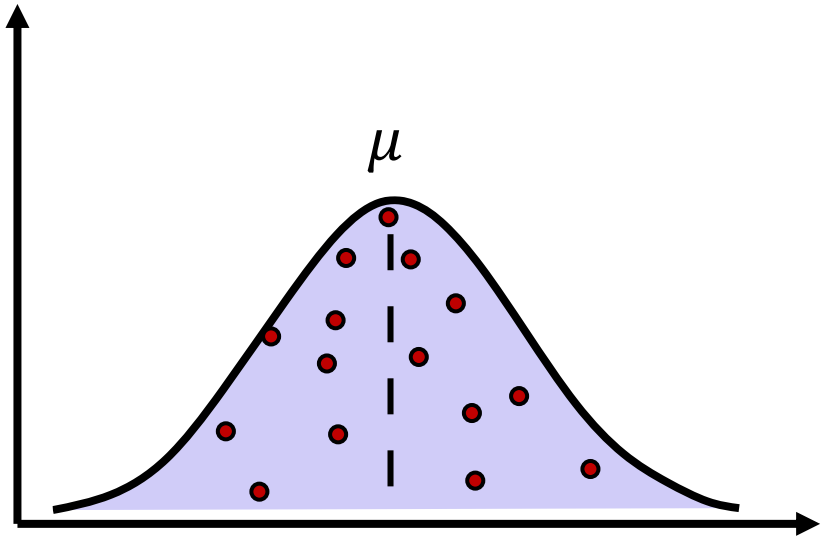
Dans Our Coding Club ils suggèrent d'accepter cette différence, mais d'expérience étant donné le quasi bimodal observé j'aurais ajusté légèrement en faisant ce qu'on appelle des mixtures model.

Modèle Bayésien – régression linéaire simple

Modèle brms – notion de Chaîne de Markov

La MCMC cherche à estimer le paramètre π dans une distribution circulaire





Voici deux liens **GitHub** particulièrement utiles pour visualiser comment fonctionne une MCMC :

© Chandan Relekar

© Chi Feng

Modèle Bayésien – régression linéaire simple

Modèle brms – les priors

```
Model_brm <- brms::brm(y~ x,
data = data, family = poisson(), chains = 3,
iter = 3000, warmup = 1000)
```

↳ Modèle sans *prior* → Ce modèle ne prend pas en compte nos hypothèses préalables.

Dans la littérature, tu as vu précédemment que l'effet d'un phénomène était d = 0,5 et que son écart-type était de 2.La différence moyenne attendue est donc $\approx 0{,}5 \times 2 = 1$

↳ On s'attend donc à une différence d'environ 1

Les coefficients des variables explicatives se sont les effets fixes

```
prior = prior(normal(1, 1), class = "b")
```

Un écart type de 1 si tu es sûr que tu vas observer cette différence

Modèle Bayésien – régression linéaire simple

Modèle brms – les priors

Les coefficient des variables explicatives se sont les effets fixes

```
prior(normal(1, 1), class = "b")
```

Dans $y_i = \beta_0 + \beta_1 x_{i,1} + \in_i$

Ce n'est pas le seul *prior* que l'on peut mettre :

Tu peux également aider le modèle en lui disant la moyenne et l'écart-type que le groupe de référence à…

```
prior(normal(3, 1), class = "Intercept")
```

L'intercept correspond à la **moyenne du groupe de référence** (groupe A) et à β_0 dans l'équation de la droite. Puisqu'une comparaison de moyenne correspond à un cas particulier de la régression linéaire.

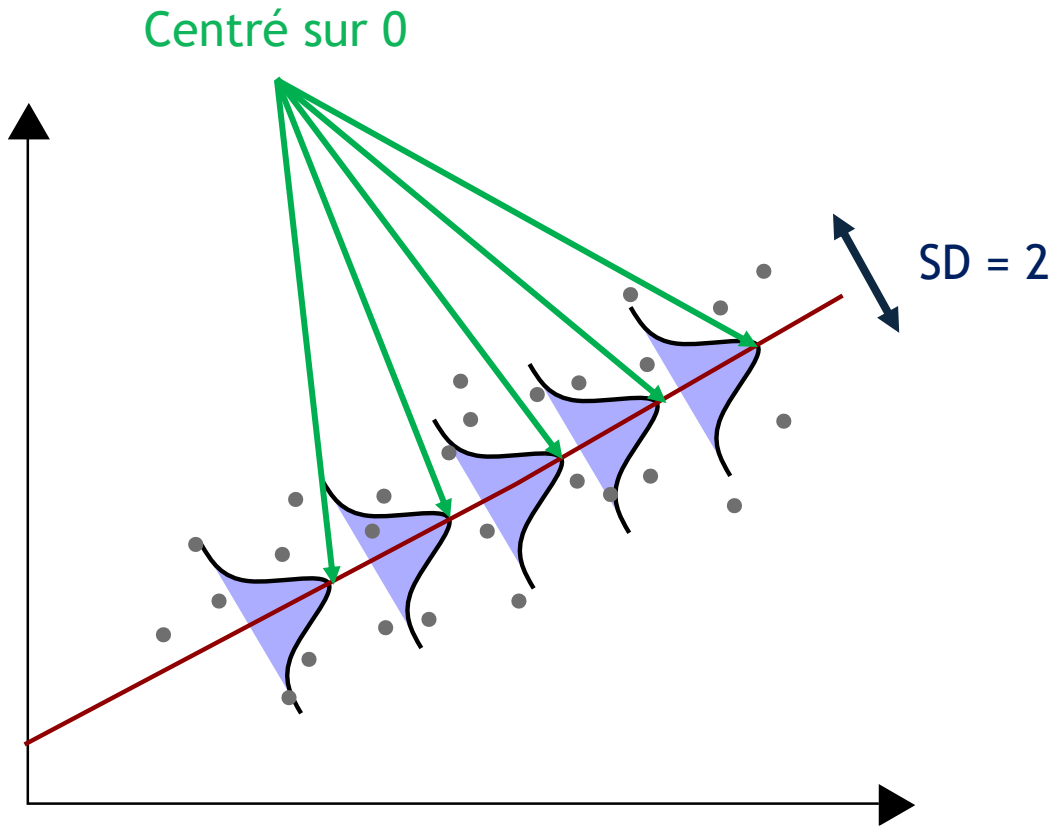
Modèle Bayésien – régression linéaire simple

Modèle brms – un dernier *prior*, celle de l'erreur résiduel (ou résidues)

```
prior(normal(0, 2), class = "sigma")
```

↳ Cas classique de la régression (par défaut).

Puisque…



Modèle Bayésien – régression linéaire simple

Mes priors

```
prior_model = prior(normal(0, 1), class = "b") + prior(normal(3, 1), class = "Intercept") + prior(normal(0, 2), class = "sigma")
```

Mon modèle

```
Model_brm <- brms::brm(y~ x,
data = data, family = poisson(), chains = 3,
prior = prior_model ,
iter = 3000, warmup = 1000)
```

L'extractions des statistiques

```
summary (Model_brm)
```

Mes post-hoc

```
emmeans(Model_brm, ~ pairwise~X)
```

Modèle Bayésien : régression linéaire mixte multiple

Mon modèle

```
Model_mixte_brm <- brm(y ~ x*z
+ (1 | ID)
,data = data
, family = zero_inflated_beta()
, prior = prior_model ,
, iter = 3000, warmup = 1000)
```

↳ Il est possible d'ajouter un facteur aléatoire en utilisant la même syntaxe que le package « lme4 »

L'extractions des statistiques

```
summary(Model_mixte_brm)
```

Les analyses post-hoc

```
emmeans(Model_mixte_brm, ~ pairwise~ z | x)
```

Modèle Bayésien : Intervalle de crédibilité et ESS

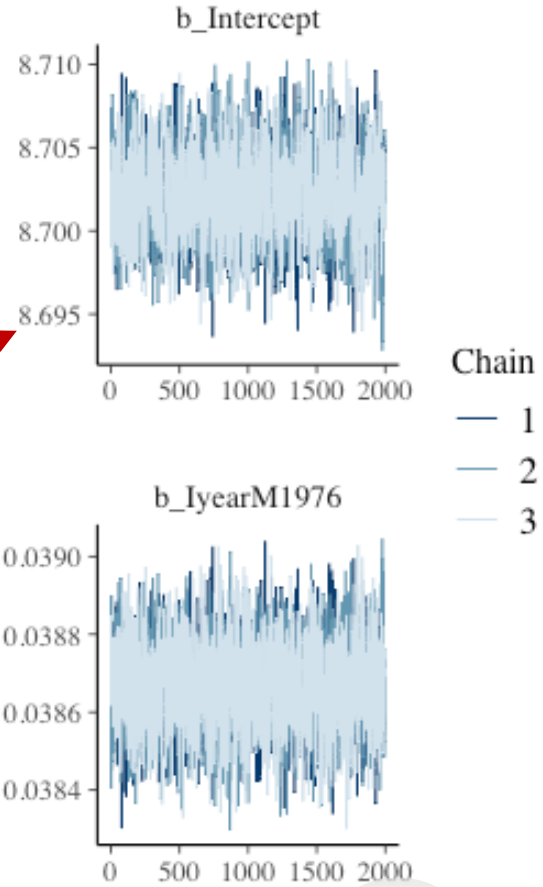
Summary

```
 Family: poisson
  Links: mu = log
Formula: pop ~ I(year - 1976)
   Data: France (Number of observations: 70)
  Draws: 3 chains, each with iter = 3000; warmup = 1000; thin = 1;
         total post-warmup draws = 6000

Population-Level Effects:
           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept      8.70      0.00     8.70     8.71 1.00     1603     2216
IyearM1976     0.04      0.00     0.04     0.04 1.00     3479     3798

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

<1.01

> à **1000**

> à **1000**



L'intervalle de crédibilité (CrI) ne passe pas par zéro, ce qui indique que la différence est « crédible » ce qui se traduit en termes fréquentistes par « significative ».

Modèle Bayésien : régression linéaire mixte multiple

emmeans

```
Condition = Control:
Demo   emmean lower.HPD upper.HPD
E1      -1.62233    -5.54     2.25
E2      -1.09361    -5.08     2.66
E3      -0.83152    -4.78     3.05
E4      -0.55457    -4.41     3.41
E5      -0.35519    -4.12     3.63

Condition = Test:
Demo   emmean lower.HPD upper.HPD
E1     -0.23794    -4.03     3.72
E2      0.00363    -3.81     3.84
E3     -0.21479    -4.09     3.64
E4     -0.78565    -4.68     3.04
E5     -1.56607    -5.43     2.22

Point estimate displayed: median
HPD interval probability: 0.95
```

Modèle Bayésien : régression linéaire mixte multiple

emmeans

```
$contrasts
Condition = Control:
contrast estimate lower.HPD upper.HPD
 E1 - E2      -0.533   -2.1561    0.990
 E1 - E3      -0.793   -2.3210    0.767
 E1 - E4      -1.079   -2.6245    0.456
 E1 - E5      -1.283   -2.7779    0.266
 E2 - E3      -0.264   -1.8297    1.278
 E2 - E4      -0.539   -2.1010    1.031
 E2 - E5      -0.746   -2.2344    0.852
 E3 - E4      -0.283   -1.7713    1.305
 E3 - E5      -0.491   -2.0657    1.036
 E4 - E5      -0.203   -1.6892    1.361
```

Modèle Bayésien : régression linéaire mixte multiple

emmeans

```
Condition = Test:
contrast estimate lower.HPD upper.HPD
  E1 - E2   -0.246  -1.7289    1.246
  E1 - E3   -0.027  -1.5327    1.417
  E1 - E4    0.543  -0.9075    2.020
  E1 - E5    1.331  -0.1638    2.782
  E2 - E3    0.219  -1.2040    1.709
  E2 - E4    0.793  -0.6733    2.255
  E2 - E5    1.568   0.0934    3.009
  E3 - E4    0.574  -0.9363    1.971
  E3 - E5    1.356  -0.0715    2.816
  E4 - E5    0.773  -0.6938    2.191
```

Des modèles Bayésiens aux modèles multiverses

- Vous trouvez que les **analyses Bayésiennes sont trop subjectives** ?

- Vous pensez qu'il est **trop difficile de caractériser des *priors*** surtout quand la recherche est novatrice.

  ↳ Rappelez-vous qu'une ***p-value* sans analyse de puissance** (par ex. avec G*Power) est difficile à interpréter. Si vous avez trop peu d'individus par rapport à ce qu'indique G*Power, vous augmentez vos chances de faux négatifs (ne pas détecter un effet qui existe). À l'inverse, si votre échantillon est trop grand, vous risquez de détecter comme significatifs de très petits effets, statistiquement réels mais peu pertinents d'un point de vue biologique.

  ↳ L'analyse de puissance reste par définition assez subjective, puisqu'il est rare que votre condition expérimentale soit exactement identique à celle d'une étude précédente. On fait bien sûr de notre mieux pour s'en approcher, mais il subsiste toujours des différences.

Donc

La **subjectivité des *priors*** en bayésien n'est pas plus problématique que celle des analyses de puissance en fréquentiste.

Des modèles Bayésiens aux modèles multiverses

- En outre, vous pouvez tester votre *prior*, via ce qu'on appelle des **tests de sensibilités.**

(Kallioinen et al., 2023; van de Schoot et al., 2021)

Le principe est de **tester plusieurs *priors*** dans un liste et voir lequel est **suffisamment informatif pour aider la convergence** de votre modèle sans pour autant trop fortement influencer vos « *Posterior* ».

Ma recommandation serait si votre hypothèse est assez forte de ne changer que votre paramètre d'incertitude, l'écart-type (*Standard Deviation, SD*). Comme dans l'exemple ci-dessous.

```
priors_test_sensitivity<- list(
strong = prior(normal(0.11, 0.15), class = "b", coef = " Independent_variable1.L"),
medium = prior(normal(0.11, 0.33), class = "b", coef = " Independent_variable1.L"),
weak = prior(normal(0.11, 0.5), class = "b", coef = " Independent_variable1.L"),
very_weak = prior(normal(0.11, 1), class = "b", coef = " Independent_variable1.L"),
minimal = prior(normal(0.11, 1.67), class = "b", coef = " Independent_variable1.L"))
```

Des modèles Bayésiens aux modèles multiverses

- En outre, vous pouvez tester votre *prior*, via ce qu'on appelle des tests de sensibilités.

(Kallioinen et al., 2023; van de Schoot et al., 2021)

Facteur aléatoire, l'individu.

```
model_strong_dependant_variable <- brm(dependant_variable ~ Independent_variable1*Independent_variable2
+ (1 | ID)
, data = data
, family = zero_inflated_poisson(link = "log", link_zi = "logit")
, prior = priors_test_sensitivity $strong
, iter = 2000, warmup = 1000
, chains = 2, cores = 2
, seed = 123
, save_pars = save_pars(all = TRUE))
```

Je diminue le nombre d'itérations et de warm-up puisque, comme je réalise cette analyse pour chaque *prior* du vecteur, il peut être long d'obtenir l'ensemble des résultats.

Ici je demande de chercher dans le vecteur que j'ai créer dans la diaporama suivante le *prior* « *strong* » ce qui correspond à :
« prior(normal(0.11, 0.15), class = "b", coef = "Condition.L") »

Des modèles Bayésiens aux modèles multiverses

- En outre, vous pouvez tester votre *prior*, via ce qu'on appelle des tests de sensibilités.

(Kallioinen et al., 2023; van de Schoot et al., 2021)

↳ Il faut faire un modèle pour chaque *prior* créée dans cette liste.

```
priors_test_sensitivity<- list(
strong = prior(normal(0.11, 0.15), class = "b", coef = " Independent_variable1.L"),
medium = prior(normal(0.11, 0.33), class = "b", coef = " Independent_variable1.L"),
weak = prior(normal(0.11, 0.5), class = "b", coef = " Independent_variable1.L"),
very_weak = prior(normal(0.11, 1), class = "b", coef = " Independent_variable1.L"),
minimal = prior(normal(0.11, 1.67), class = "b", coef = " Independent_variable1.L"))
```

Dans la diaporama précédente le modèle correspondait à ce *prior*.

Des modèles Bayésiens aux modèles multiverses

- En outre, vous pouvez tester votre *prior*, via ce qu'on appelle des tests de sensibilités.

(Kallioinen et al., 2023; van de Schoot et al., 2021)

```
# Test the sensitivity of each of them
powerscale_sensitivity(model_strong_ dependant_variable, variable = "b_ Independent_variable1.L")
powerscale_sensitivity(model_medium_ dependant_variable, variable = "b_ Independent_variable1.L")
powerscale_sensitivity(model_weak_ dependant_variable, variable = "b_ Independent_variable1.L")
powerscale_sensitivity(model_very_weak_ dependant_variable, variable = "b_ Independent_variable1.L")
powerscale_sensitivity(model_minimal_ dependant_variable, variable = "b_ Independent_variable1.L")
powerscale_sensitivity(model_flat_ dependant_variable, variable = "b_Independent_variable1.L”)
```

La fonction *powerscale_sensitivity* compare la valeur attendue d'après le *prior* seul à celle issue des données (*posterior*). En pratique, la différence entre les deux ne doit pas être trop importante : un écart de 20 à 30 % est généralement considéré comme acceptable.

Des modèles Bayésiens aux modèles multiverses

- En outre, vous pouvez tester votre *prior*, via ce qu'on appelle des tests de sensibilités.

(Kallioinen et al., 2023; van de Schoot et al., 2021)

```
powerscale_sensitivity(model_strong_ dependant_variable, variable = "b_ Independent_variable1.L")

Likelihood selection: all data

# variable                    prior    likelihood      diagnosis
# b_ Independent_variable1    0.14       0.171
```

Ici, le prior est acceptable puisque l'écart relatif entre le prior et le posterior (likelihood) est |0.171 − 0.14| / 0.171 ≈ 18 %, ce qui reste dans la tranche de 20–30 % mentionnée plus haut. Toutefois, il est recommandé de choisir un *prior* qui soit à la fois cohérent avec nos hypothèses (donc informatif) et qui n'influence pas excessivement le *posterior*. Il faut aussi impérativement que le prior soit en dessous du likelihood, voir Kallioinen et al. 2022 tableau)

Des modèles Bayésiens aux modèles multiverses

- Vous pouvez visuellement voir ce que vos prior ont comme impact sur vos données avec les commandes suivantes :

(Kallioinen et al., 2023; van de Schoot et al., 2021)

```
draws_strong_ dependant_variable <-  model_strong_GazeFre1s %>%  spread_draws(b_Independent_variable1.L) %>%
mutate(Prior = "Strong")

draws_medium_ dependant_variable <- model_medium_GazeFre1s %>%  spread_draws(b_Independent_variable1.L)
%>%  mutate(Prior = "Medium")

draws_weak_ dependant_variable <- model_weak_GazeFre1s %>%  spread_draws(b_Independent_variable1.L) %>%
mutate(Prior = "Weak")

draws_very_weak_ dependant_variable <- model_very_weak_GazeFre1s %>%
spread_draws(b_Independent_variable1.L) %>%  mutate(Prior = "Very Weak")

draws_minimal_ dependant_variable <- model_minimal_GazeFre1s %>% spread_draws(b_Independent_variable1.L)
%>%  mutate(Prior = "Minimal")

draws_flat_ dependant_variable <- model_flat_GazeFre1s %>%  spread_draws(b_Independent_variable1.L) %>%
mutate(Prior = "Flat")
```

Des modèles Bayésiens aux modèles multiverses

- Vous pouvez visuellement voir ce que vos prior ont comme impact sur vos données avec les commandes suivantes :

(Kallioinen et al., 2023; van de Schoot et al., 2021)

Puis

```
all_draws_prior <-
bind_rows (draws_strong_ dependant_variable,
draws_medium_ dependant_variable,
draws_weak_ dependant_variable,
draws_very_weak_ dependant_variable,
draws_minimal_ dependant_variable,
draws_flat_ dependant_variable )
```

Des modèles Bayésiens aux modèles multiverses

- Vous pouvez visuellement voir ce que vos prior ont comme impact sur vos données avec les commandes suivantes :

(Kallioinen et al., 2023; van de Schoot et al., 2021)

Enfin

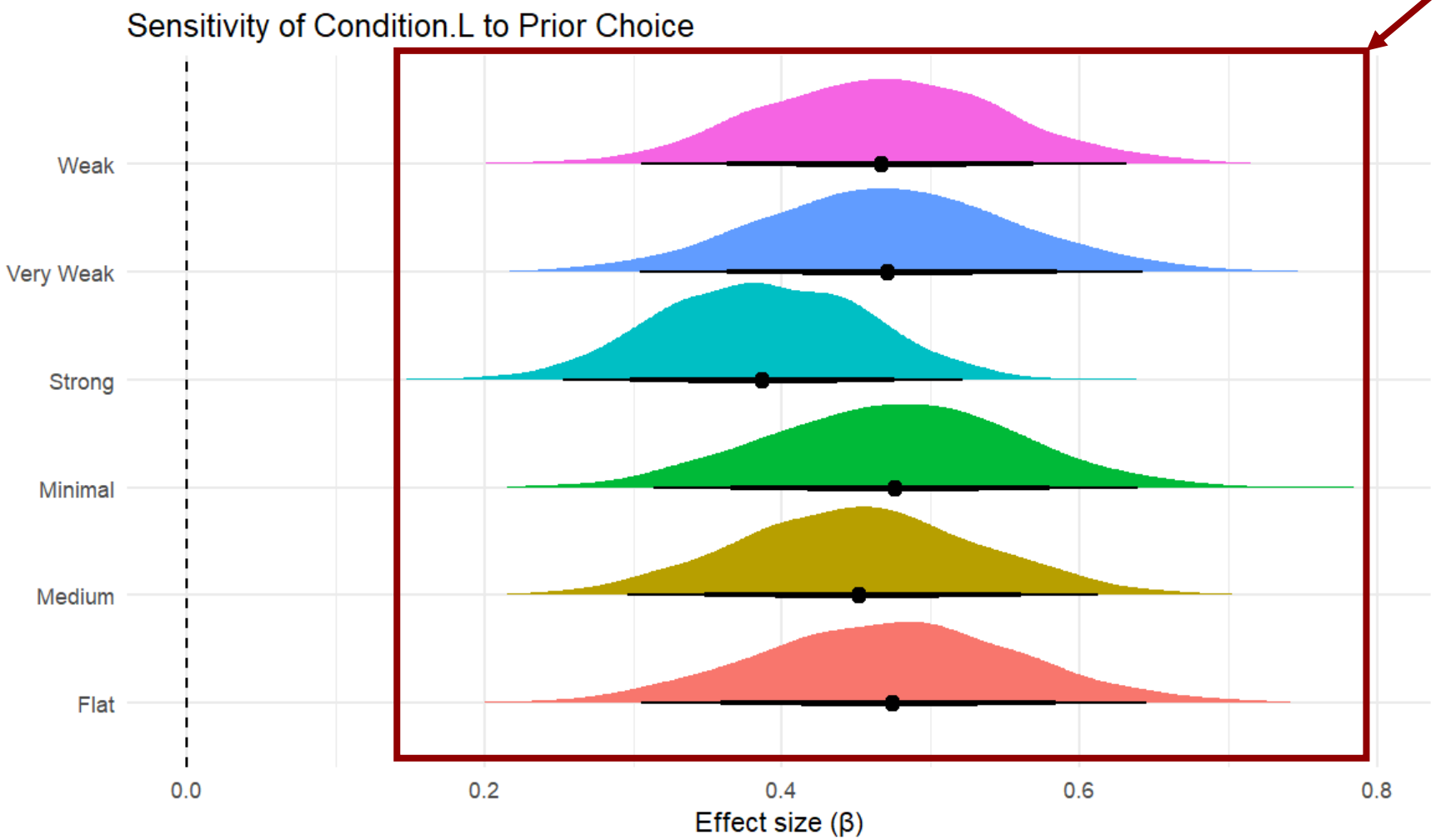
```
prior_sensitivity_ dependant_variable = ggplot(all_draws_ dependant_variable, aes(x = b_
Independant_variable1.L, y = Prior, fill = Prior))
+ stat_halfeye(.width = c(.5, .8, .95))
+ geom_vline(xintercept = 0, linetype = "dashed")
+ labs(x = "Effect size (β)", y = NULL, title = "Sensitivity of Condition.L to Prior Choice")

+theme_minimal()
+theme(legend.position = "none")

prior_sensitivity_GazeFre1s
```

Des modèles Bayésiens aux modèles multiverses

- Vous pouvez visuellement voir ce que vos prior ont comme impact sur vos données avec les commandes suivantes :

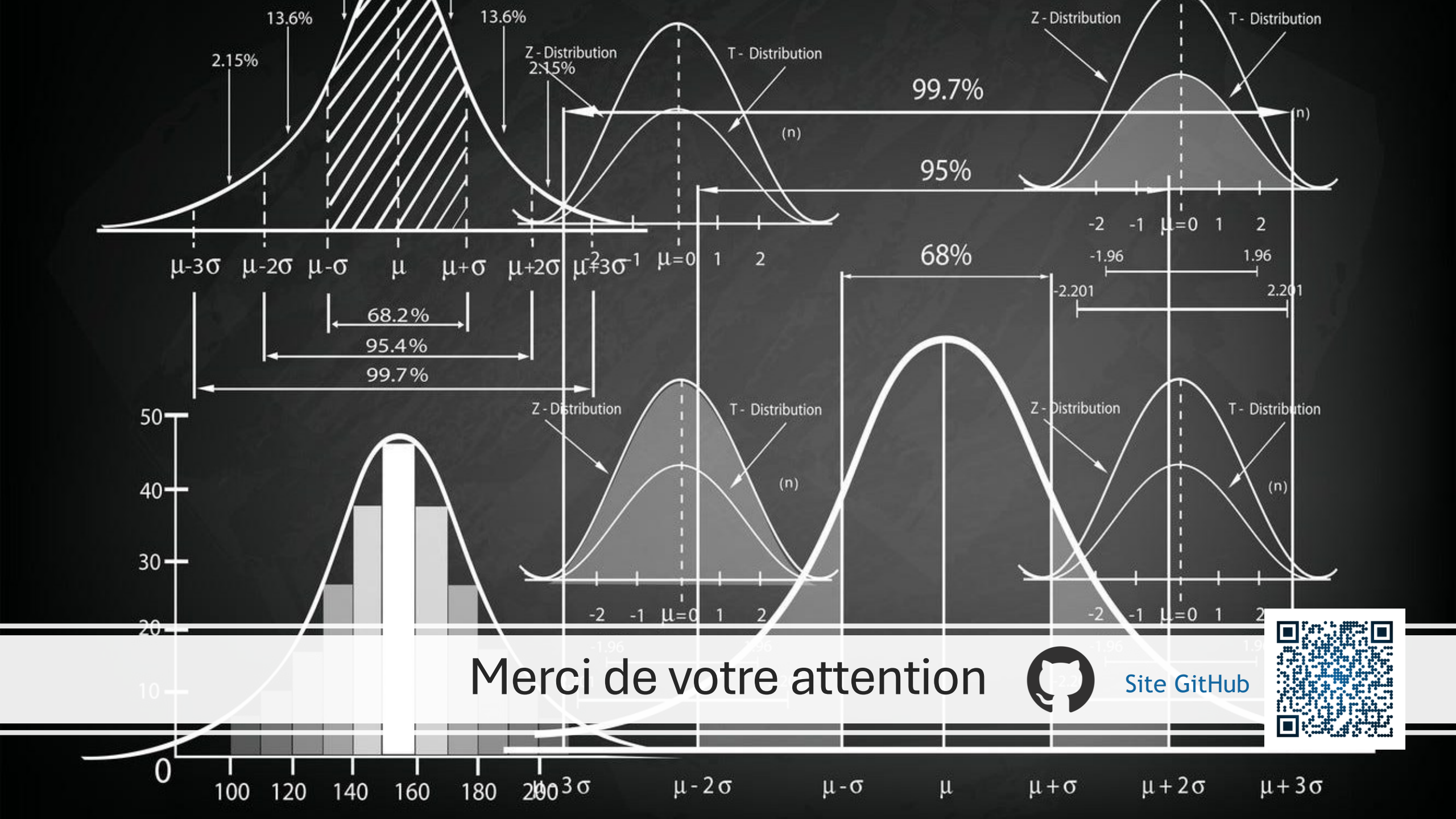
(Kallioinen et al., 2023; van de Schoot et al., 2021)

Pour obtenir



Ici, vous voyez concrètement comment vos données se comportent en fonction du *prior* utilisé.

Preprint

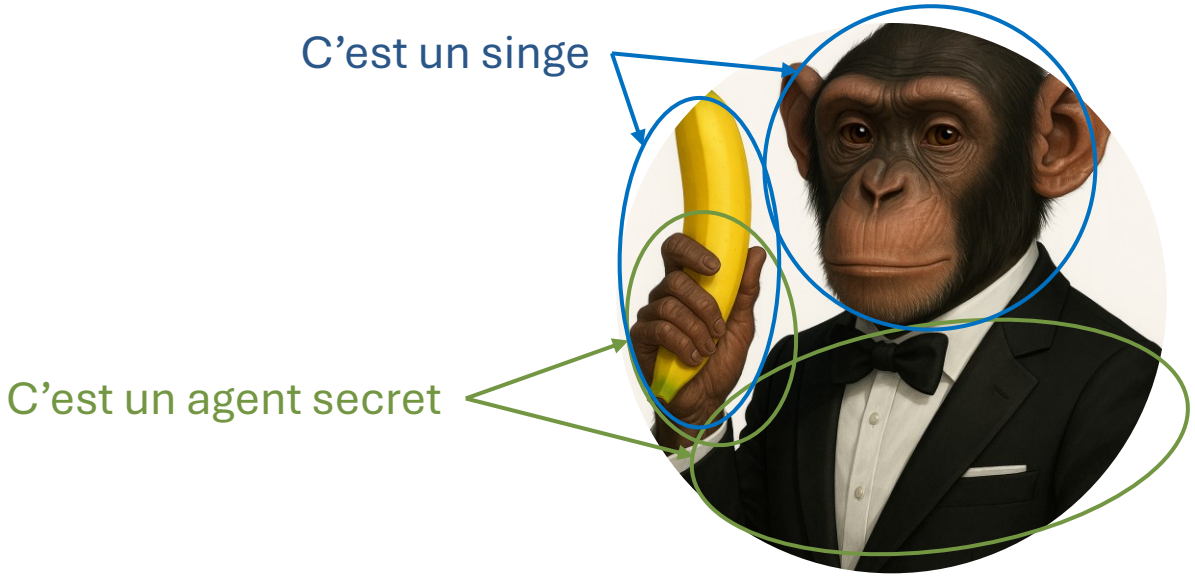


Merci de votre attention

Site GitHub

Le théorème de Bayes

- Un théorème que nous utilisons tous les jours sans le savoir.

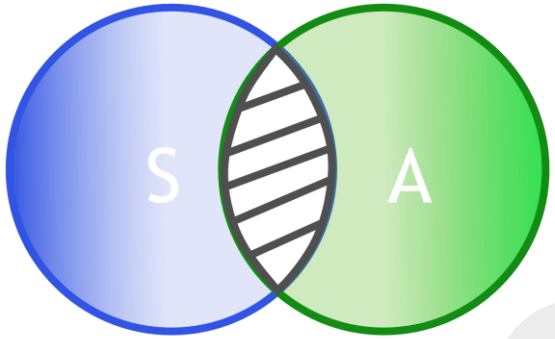


→ La **question** que poserait le théorème de Bayes…

Quelle est la probabilité qu'il s'agisse d'un agent secret sachant qu'il a une tête de singe et une banane ?

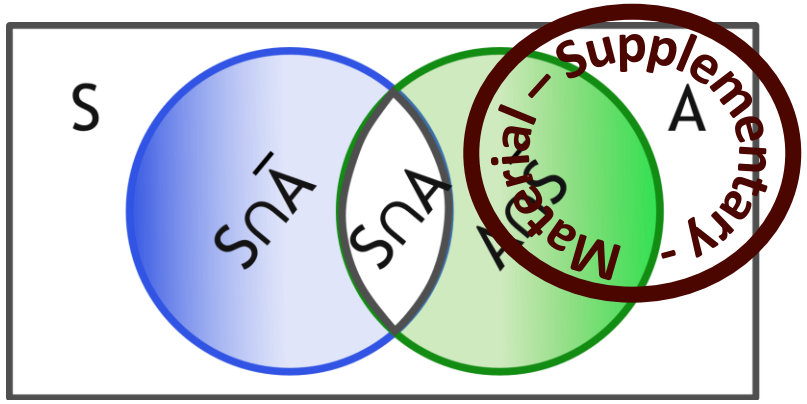
→ Si on regroupait les indices « banane » et « tête de singe » en une seul catégorie ça donnerait…

$$\mathrm{P}(S|A) = \frac{\mathrm{P}(S|A) \cdot P(A)}{P(S)}$$



(Voir aussi l'exemple de Kruschke 2015, p. 102, table 5.2)

Le théorème de Bayes



$$P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{P(S|A).P(S)}{P(A|S).P(S) + P(A|\overline{S}).P(\overline{S})}$$

Probabilité qu'il s'agisse d'un singe alors qu'il ressemble à James Bond

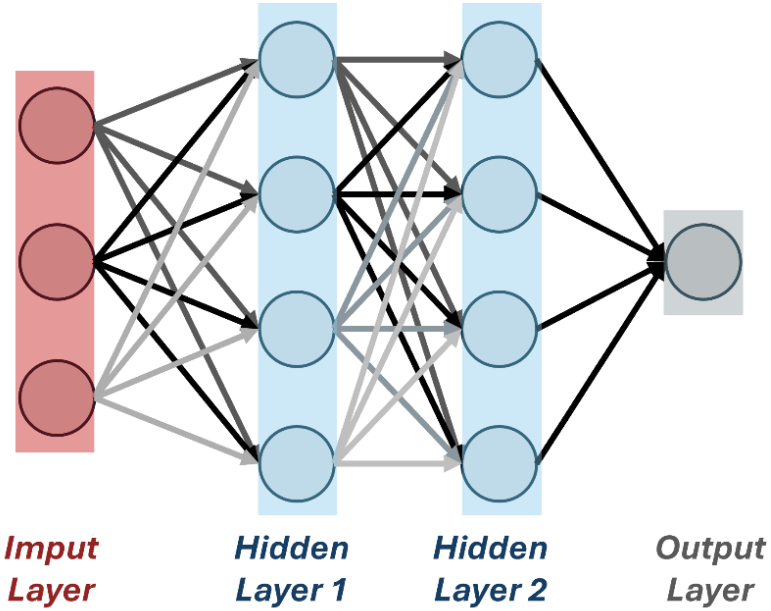
Mettons que nous ayons une chance sur dix de croiser un singe, puisque nous sommes à proximité d'un cirque ou d'un zoo. On a donc $P(S) = 0{,}1$, où S désigne l'événement "c'est un singe". Supposons également que, si c'est un singe, il y a 95 % de chances d'observer cette image : $P(S|A) = 0{,}95$. $P(\overline{S})$ est la probabilité inverse de $P(S)$ donc $P(\overline{S})$ = 1 - $P(S)$. Enfin il faut supposer également que la probabilité d'avoir cette photo ne s'agissant pas d'un singe est de 1 % donc $P(A|\overline{S}) = 0{,}01$. Cela donne…

$$P(S|A) = \frac{0{,}95.0{,}1}{0{,}95.0{,}1+0{,}01.0{,}9} = \frac{0{,}095}{0{,}095+0{,}009} = \frac{0{,}095}{0{,}104} \approx 0{,}913$$

(Voir aussi l'exemple de Kruschke 2015, p. 102, table 5.2)

Le théorème de Bayes est appliqué dans de nombreux domaines!

Un exemple d'intelligence artificielle et naturelle



Intelligence Artificiel
(e.g., réseau de neurones artificiels)

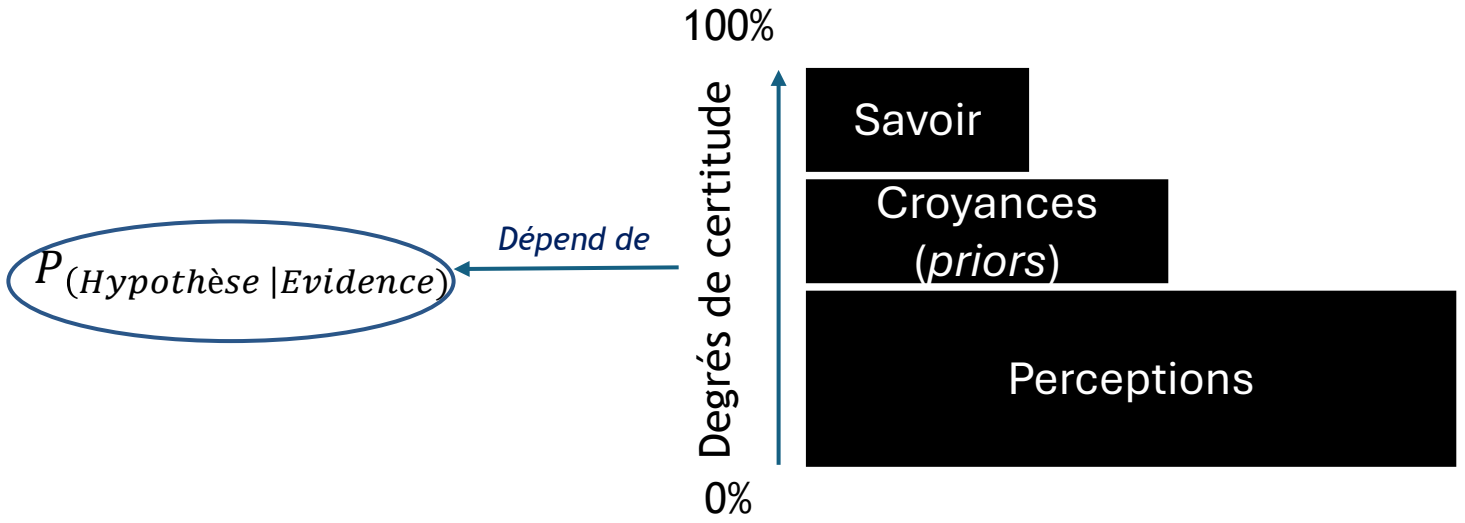
- Une machine apprend en réactualisant continuellement ses attentes qu'en statistiques bayésiennes ont appelle *a priori* (ou *prior)*.
- Un humain fonctionne de la même manière : il a des attentes, et des observations qui vont plus ou mois être bouleversées. L'écart entre ses attentes et ses observations, s'il est mémorisé, constitue un apprentissage (voir article de Poli et al. 2024, ou Cook et al. 2011 pour une discussion détaillée sur les modèles de *reinforcement learning*).

(Poli et al. 2024; Cook et al. 2011)

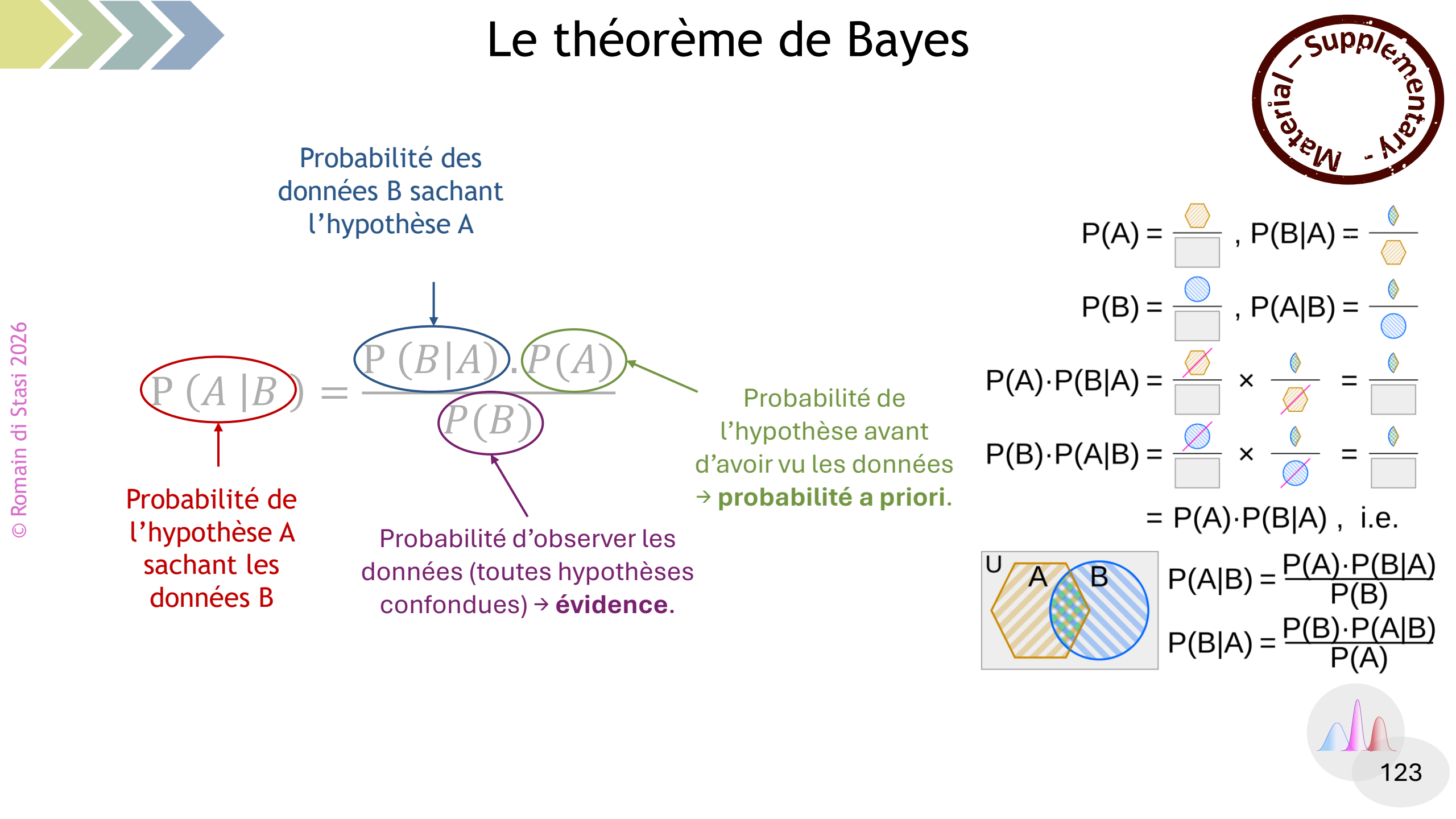
Le théorème de Bayes est appliqué dans de nombreux domaines!

Un exemple l'Intelligence artificiel et naturelle

Ce qu'une IA fait, tout comme nous, c'est privilégier les réponses les plus étayées par des évidences, en fonction des hypothèses existantes, tout en évitant le surajustement (*overfitting*).



Le théorème de Bayes



Probabilité des données B sachant l'hypothèse A

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probabilité de l'hypothèse A sachant les données B

Probabilité de l'hypothèse avant d'avoir vu les données → **probabilité a priori**.

Probabilité d'observer les données (toutes hypothèses confondues) → **évidence**.

$$P(A) = \frac{\text{⬡}}{\text{▭}}, \quad P(B|A) = \frac{\text{◆}}{\text{⬡}}$$

$$P(B) = \frac{\text{●}}{\text{▭}}, \quad P(A|B) = \frac{\text{◆}}{\text{●}}$$

$$P(A) \cdot P(B|A) = \frac{\text{⬡}}{\text{▭}} \times \frac{\text{◆}}{\text{⬡}} = \frac{\text{◆}}{\text{▭}}$$

$$P(B) \cdot P(A|B) = \frac{\text{●}}{\text{▭}} \times \frac{\text{◆}}{\text{●}} = \frac{\text{◆}}{\text{▭}}$$

$$= P(A) \cdot P(B|A), \text{ i.e.}$$

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(B)}$$

$$P(B|A) = \frac{P(B) \cdot P(A|B)}{P(A)}$$

Comparaison de moyenne et régression linéaire

- Une comparaison de moyennes est un cas particulier de la régression linéaire

$$y_i = \beta_0 + \beta_1 x_{i,1} + X_i + \in_i$$

- y_i Score de l'individu
- X_i 0 si l'individu est dans le groupe A (contrôle), $X_i = 1$ s'il est dans le groupe B (traitement)
- β_0 Moyenne du groupe A
- β_1 Différence de moyenne entre B et A
- $\in_i$ Bruit aléatoire

Un test t pour deux échantillons équivaut à une régression linéaire avec un prédicteur binaire.

Maximum de vraisemblance
Loi Binomiale

Autre Exemple

- On souhaite estimer **le taux de germination d'un lot de semences** en faisant germer $\boldsymbol{n} = 20$ semences dans des condition similaires.

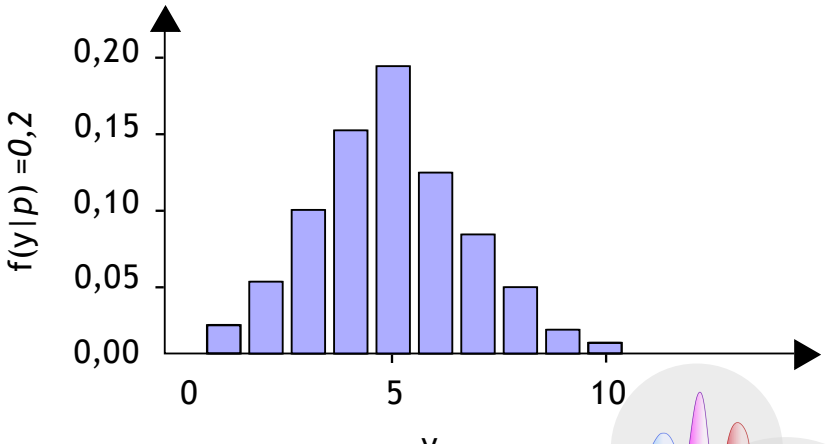
- Nous avons la **réponse** $\boldsymbol{y}$, qui correspond aux nombre de semences ayant germé avec succès.

- Nous savons que y suit un distribution binomiale ou p est la probabilité de germination de la population:

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

- $f(y|p)$ représente la distribution de y, conditionnelle à une valeur de p

Par exemple si on considère que j'ai une probabilité que mes plantes ont 20 % de chance de germer on va avoir $f(y|p) = 0{,}2$.



Il correspond à notre *prior* rappelez-vous l'exercice avec les dés

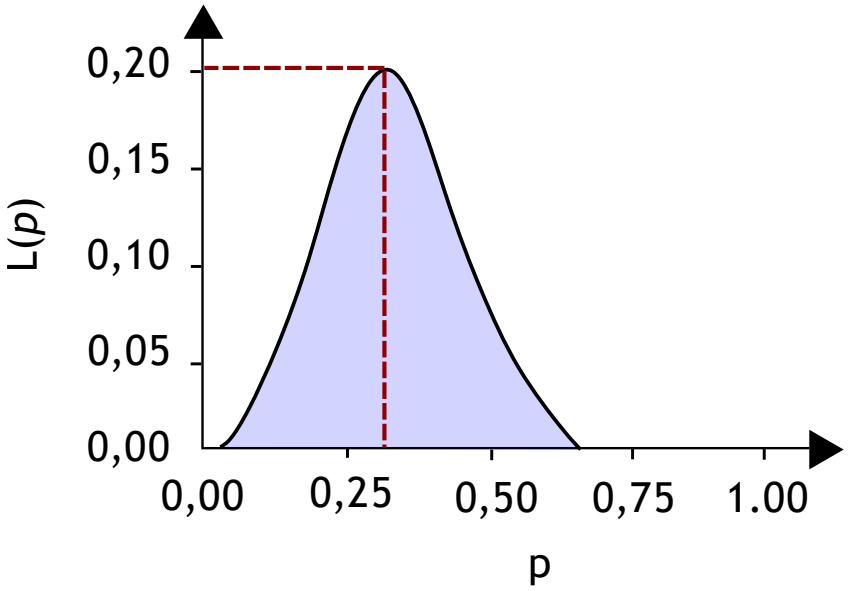
Maximum de vraisemblance
Loi Binomiale

- L'estimé du maximum de vraisemblance pour p correspond à :

$$\hat{p} = \frac{y}{n}$$

- Dans notre exemple, avec $y = 6$ et $n = 20$ nous aurons :

$$\hat{p} = \frac{6}{20} = 0{,}3$$



Maximum de vraisemblance
Loi Binomiale

- Subtilité importante – la fonction de densité des deux équations peut s'écrire de deux façons

Cas de figure 1 : On **ne sait pas** ce qu'il s'est **passé essai par essai**, mais on **a le total**. Tu sais que tu as 10 essais (N =10) et il y a y = 6 succès.

Ici pas besoin de faire le produit de probabilités au lieu de ça on peut simplement écrire :

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y}$$

Cas de figure 2 : Tu observes chaque essai **séparément.**

Tu n'as pas le résultat final, il faut donc le calculer comme un produit :

$$f(y|p) = \prod_{i=1}^{n} p^y (1-p)^{n-y}$$

Maximum de vraisemblance
Loi Binomiale

Cas de figure 2 : Tu observes chaque essai **séparément.**

Tu n'as pas le résultat final, il faut donc le calculer comme un produit.

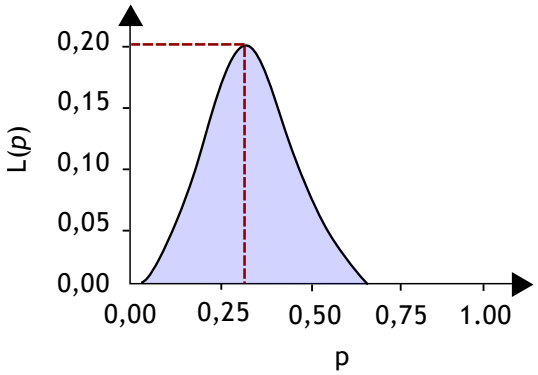
$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$\underbrace{L(p)}_{L(\theta)} = \prod_{i=1}^{n} \underbrace{p^{y}\,(1-p')^{n-y}}_{f(x_i \mid \theta)}$$

Fonction de vraisemblance qui représente la probabilité d'observer y succès sur n si la proba de succès et p

Pour simplifier tous cela on fait souvent le logarithme de cette fonction

$$\log L(p) = \sum_{i=1}^{n} [y_i \log(p) + (1-y_i)\log(1-p)]$$



Comme précédemment l'objectif est de maximiser cette fonction en trouvant la valeur p qui rend les données les plus probables.