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

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

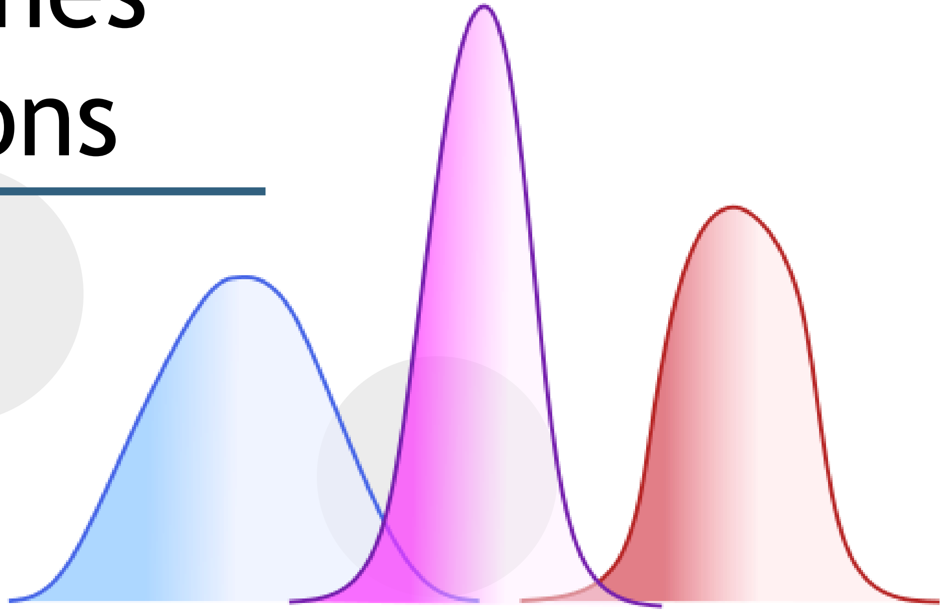
SCIENTIA VINCERE TENEBRAS
UNIVERSITAS BRUXELLENSIS

Statistiques bayésiennes – rappel et illustrations

Cours 5 – OMN

Romain di Stasi, PhD

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



Belgian Association for Psychological Sciences

ULB UNIVERSITÉ LIBRE DE BRUXELLES

CENTER FOR RESEARCH IN COGNITION & NEUROSCIENCES /UNESCOG

- Pour comprendre ce que sont les **statistiques bayésiennes**, il faut comprendre :

  → Le **théorème de Bayes**

  → La **loi binomiale**

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

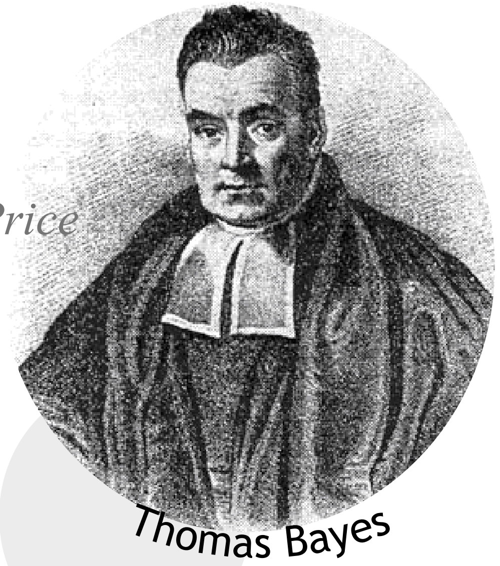
Mais pas de panique, nous allons revenir sur ces aspects ici.

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



Probabilité de l'hypothèse A sachant les données B

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

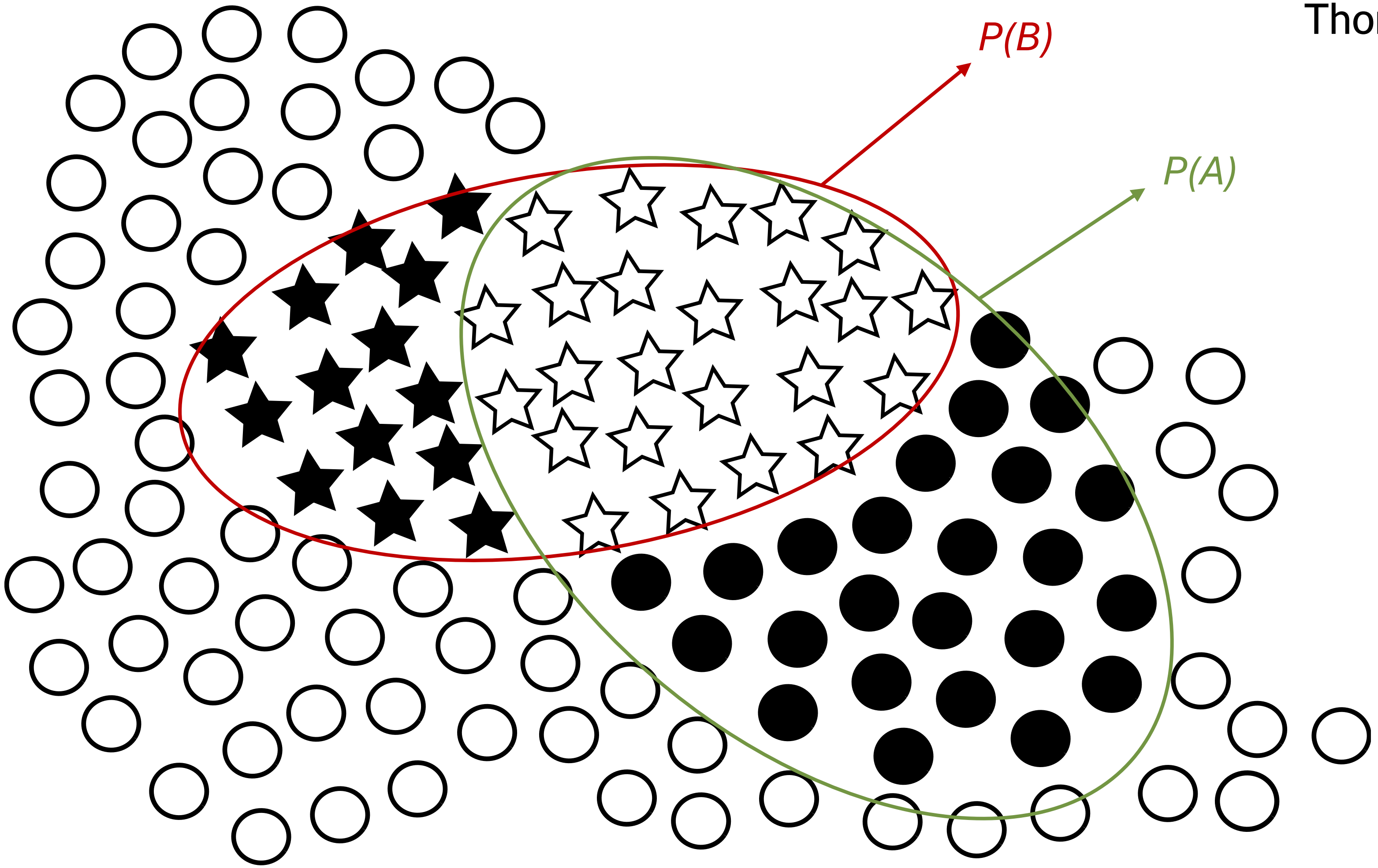
Probabilité de l'hypothèse A sachant les données B

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

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

(Kruschke 2015, p. 99-118)

Le théorème de Bayes – une brève histoire



Thomas Bayes montre ceci…

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

Cela revient au même que faire l'inverse :

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

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.

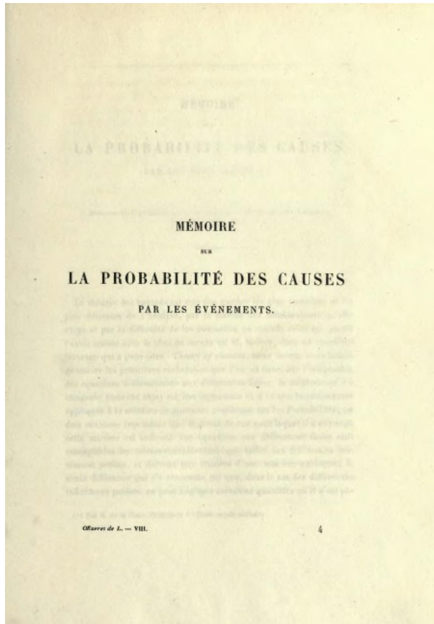
Puis Pierre-Simon Laplace a découvert l'intérêt de l'égalité mise en évidence par Thomas Bayes qu'il a reformulée ainsi:

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

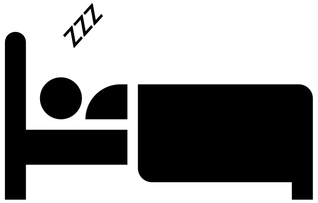
$$P(Ho|x)$$

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.

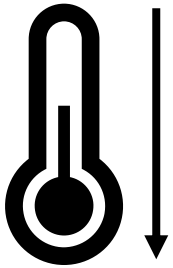
Le théorème de Bayes – une brève histoire

Rappelez vous l’exemple de l’OVNI

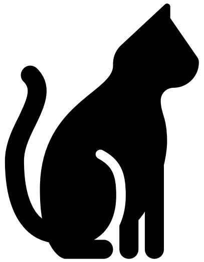


J’entends un bruit de grincement de parquet, c’est à cause de

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**

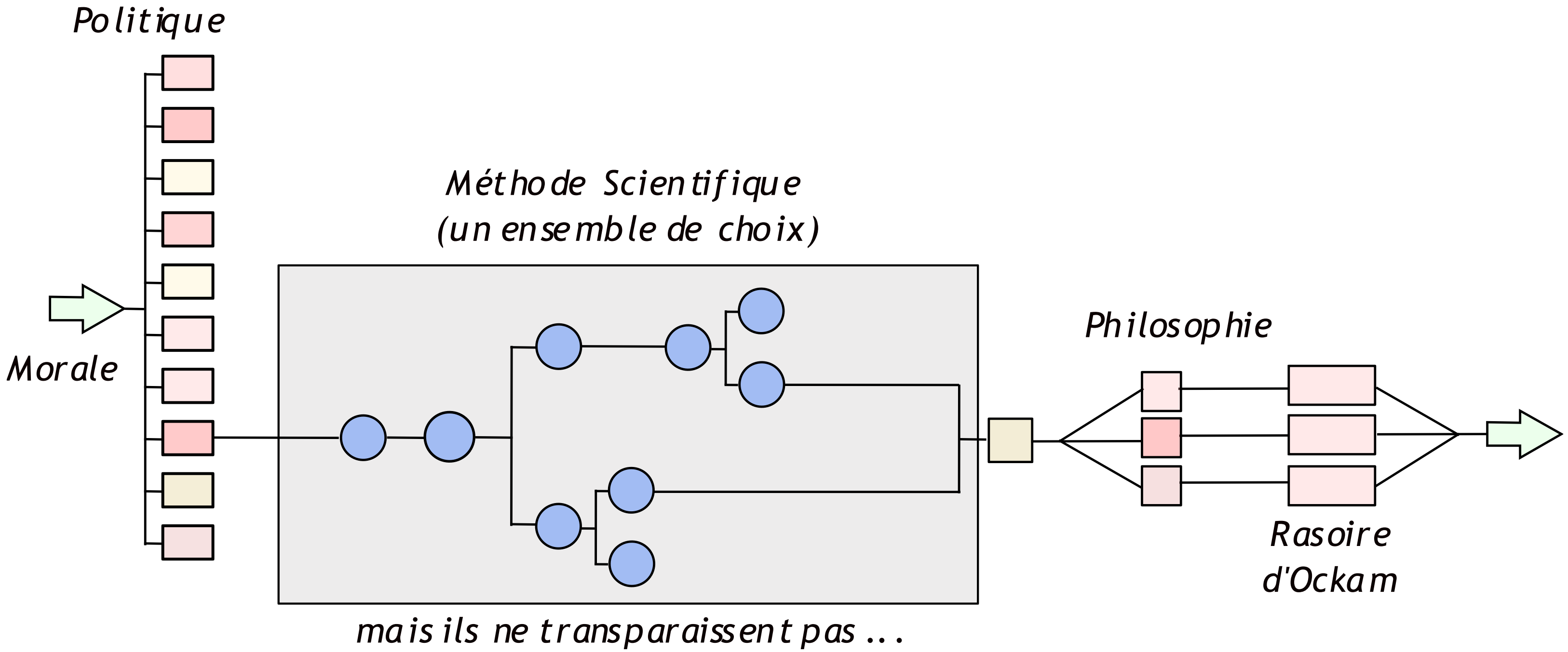
Les statistiques fréquentistes ne le sont pas moins puisqu’elles reposent sur des choix.

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.



© Romain di Stasi, inspiré par la Chaîne d'Hygiène mentale

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.

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

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.

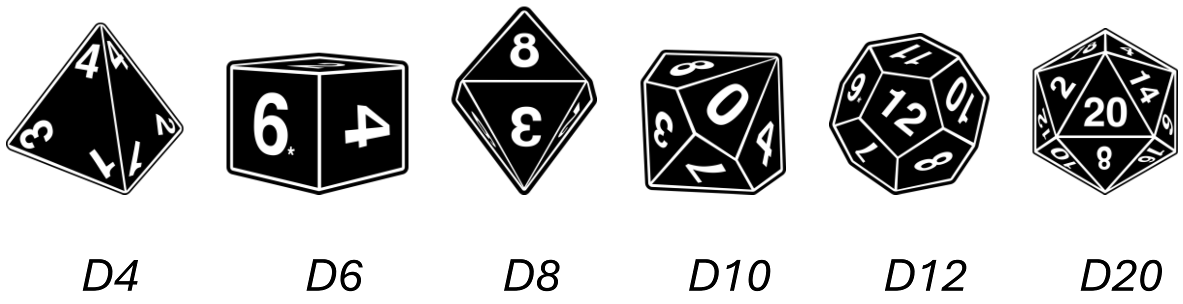
~~**Croyance A priori (Prior 1)** – tous les moutons sont blanc ou noir.~~

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.

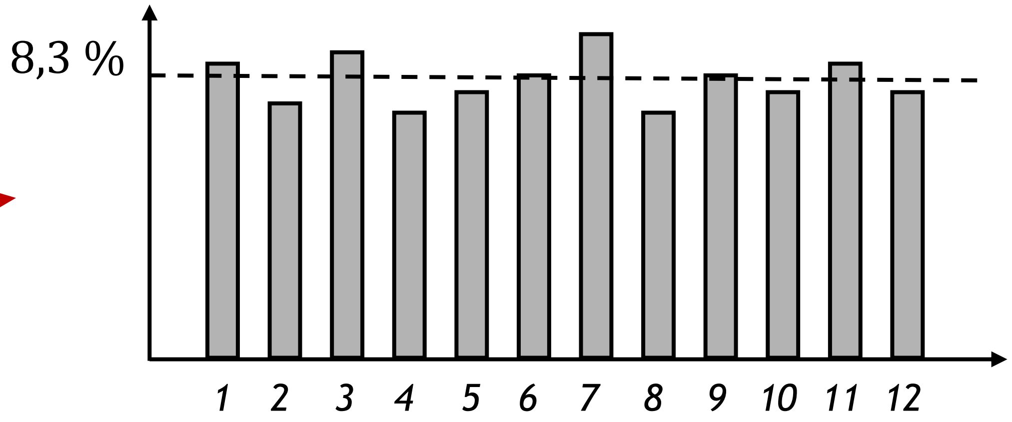


J'ai le D12, quelles sont les chances que j'aie chaque résultat possible ?

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.

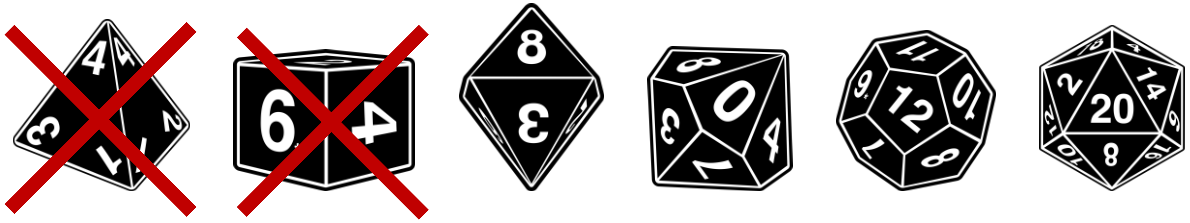
Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.

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…

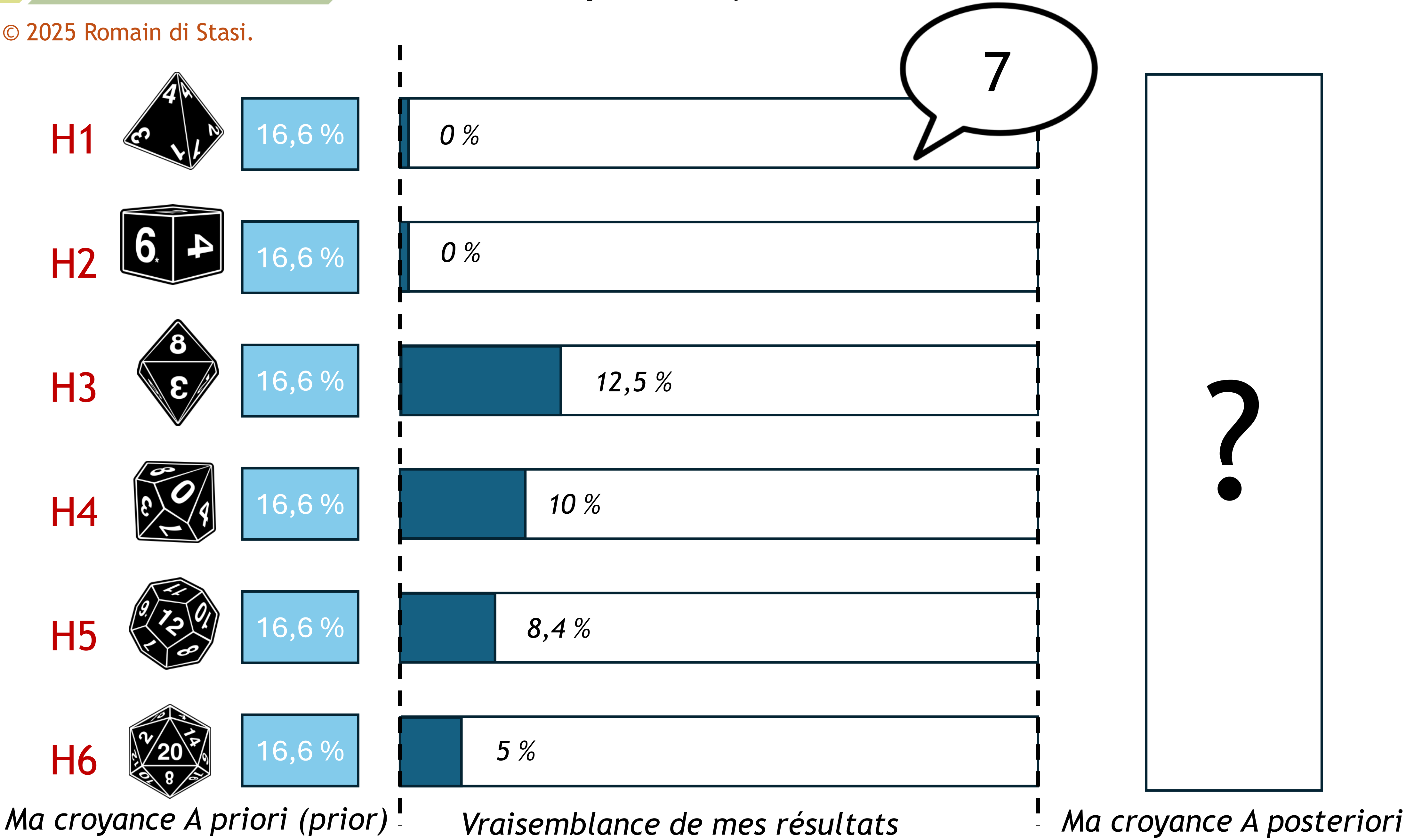
Statistiques Fréquentistes vs. bayésiennes, une illustration dans un jeu de rôle.

Différents dés sont les hypothèses pour modéliser le résultat obtenu.

H1 H2 H3 H4 H5 H6

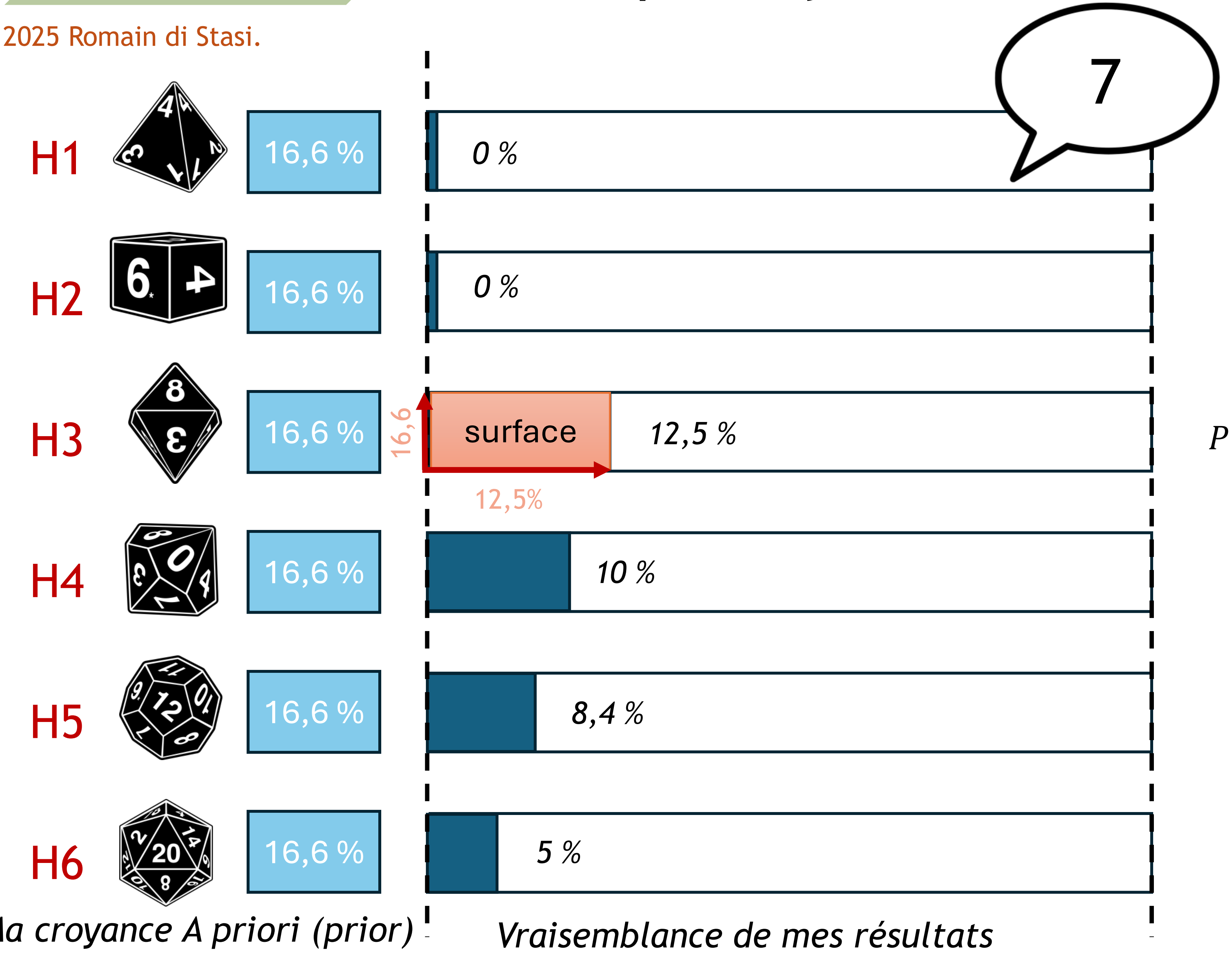
D4 D6 D8 D10 D12 D20

Statistique bayésienne – et lancé de dés



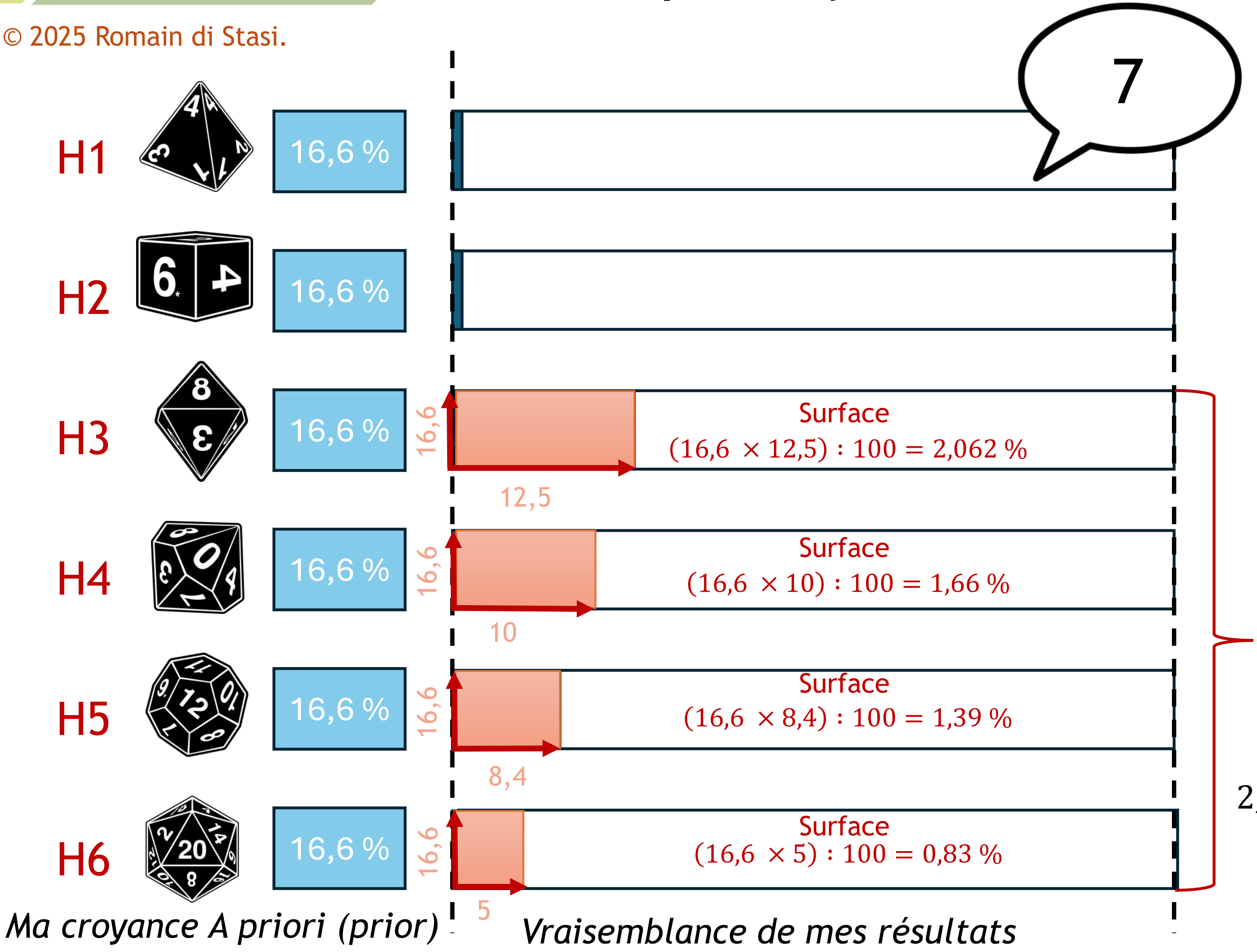
Le théorème

Statistique bayésienne – et lancé de dés



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

Statistique bayésienne - et lancé de dés



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

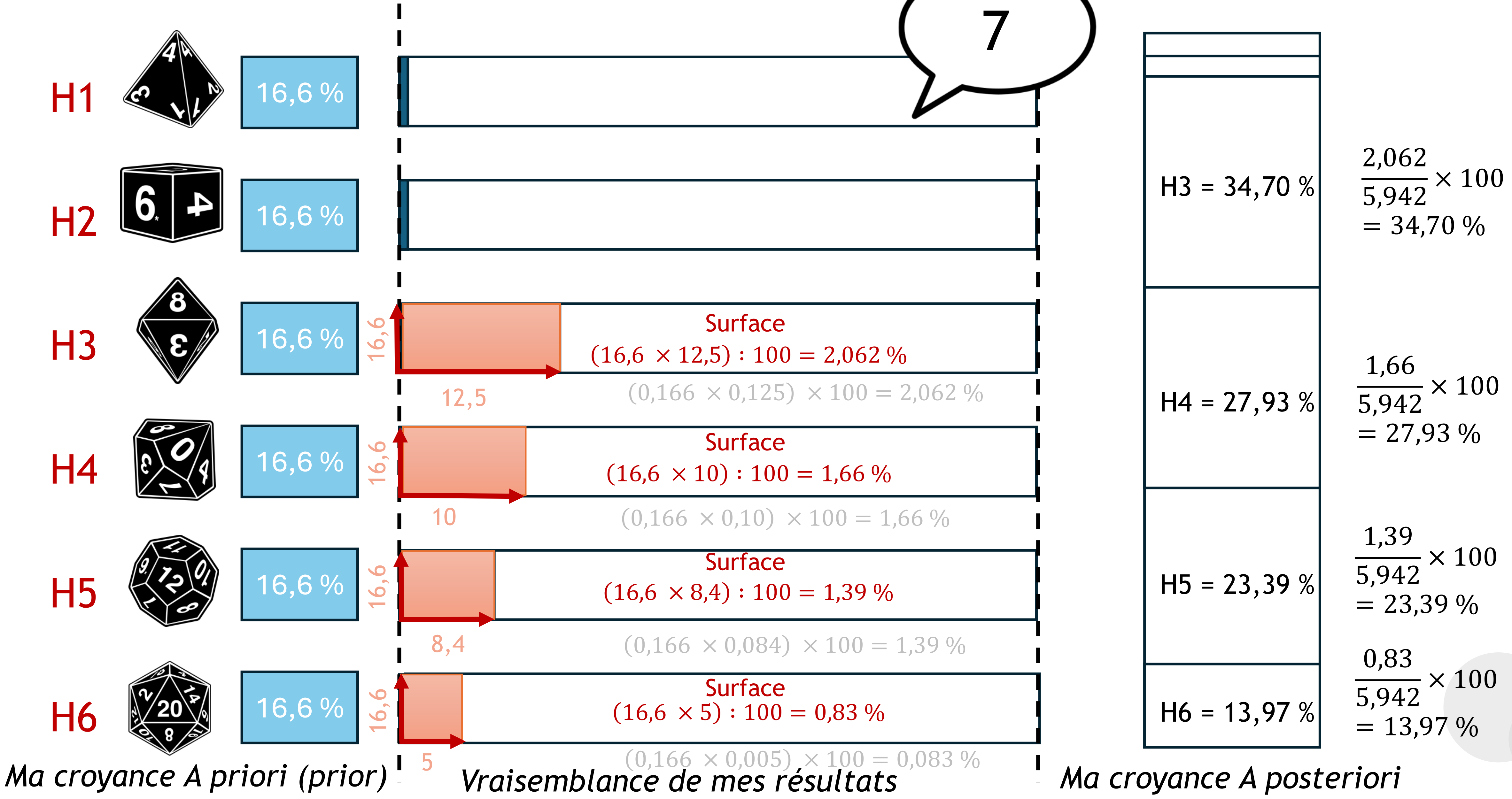
$$P(H3\,|x) = \frac{P(x\,|H3).P(H3)}{P(x)}$$

$$\sum Surfaces$$

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

H1  0 %

H2  0 %

H3  34,70 %

H4  27,93 %

H5  23,39 %

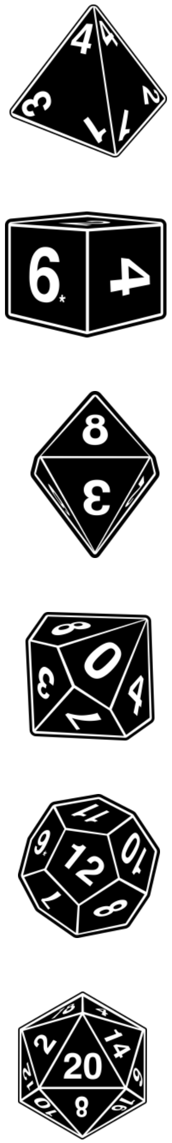
H6  13,97 %

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

7

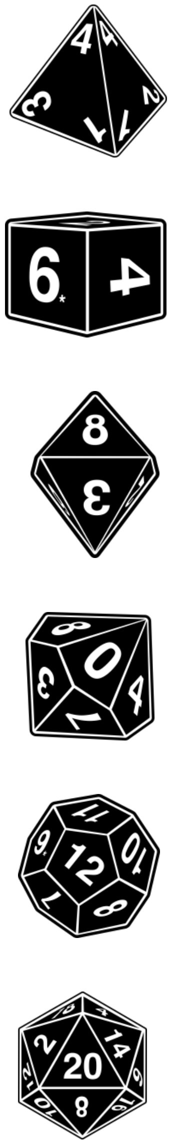


H1	0 %
H2	0 %
H3	34,70 %
H4	27,93 %
H5	23,39 %
H6	13,97 %

Nouvelle croyance nouveau A priori

Ma croyance A posteriori

Statistique bayésienne – et lancé de dés



	Nouvelle croyance nouveau a priori	Vraisemblance	Calcul
H1	0 %		
H2	0 %		
H3	34,70 %	12,5	$(34{,}70 \times 12{,}5) : 100 = 4{,}34\ \%$
H4	27,93 %	10	$(27{,}93 \times 10) : 100 = 2{,}79\ \%$
H5	23,39 %	8,4	$(23{,}39 \times 8{,}4) : 100 = 1{,}96\ \%$
H6	13,97 %	5	$(13{,}97 \times 8{,}4) : 100 = 1{,}17\ \%$

7

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

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

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

H3 = 42,30 %

H4 = 27,19 %

H5 = 19,10 %

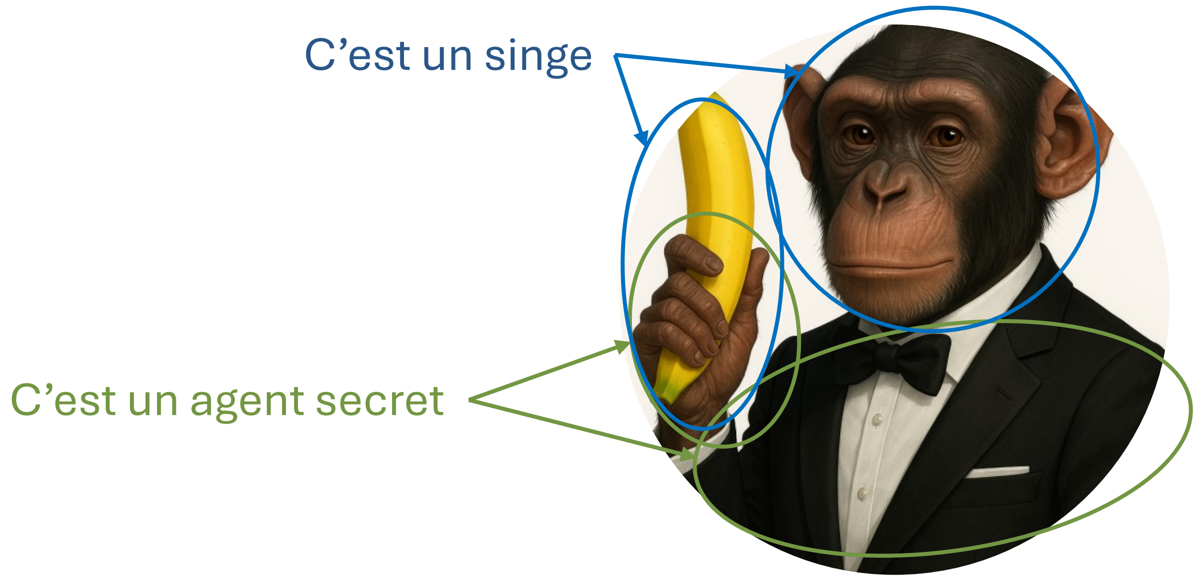
H6 = 11,40 %

Nouvelle croyance nouveau a priori

Ma croyance a posteriori

Le théorème de Bayes

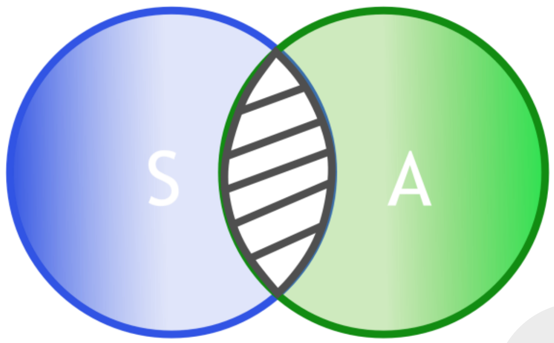
- Un théorème que nous utilisons tous les jours sans le savoir.



→ 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 ?

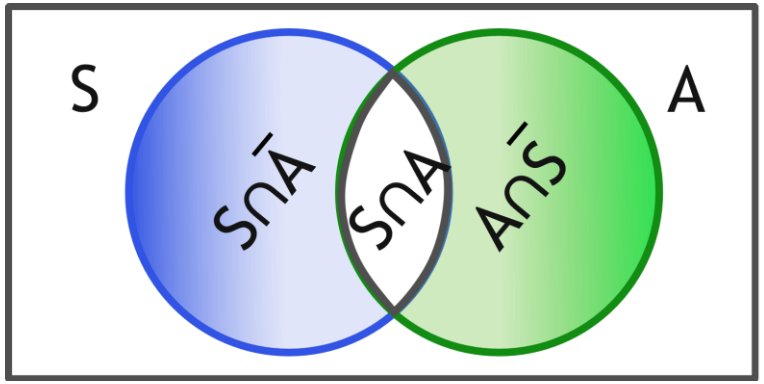
→ Si on regroupait les indices « banane » et « tête de singe » en une seul catégorie ça donnerait…



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

(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

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**.

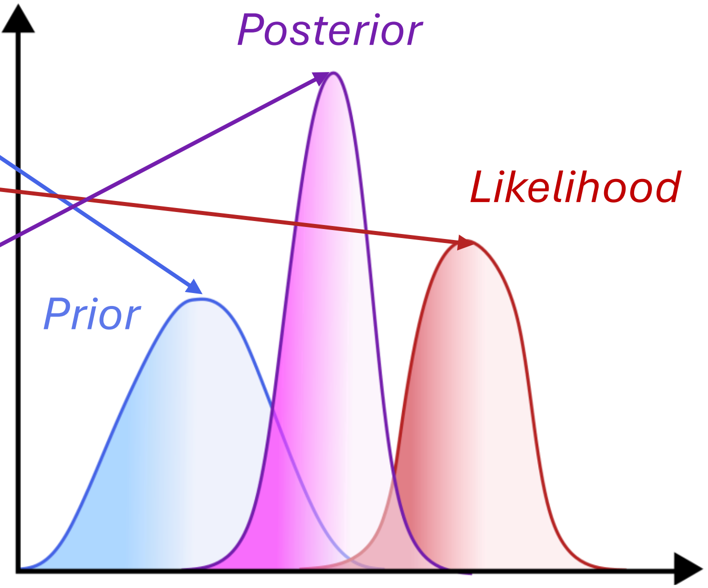
Ne pas confondre vraisemblance et croyance *A posteriori*

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

Vraisemblance vs. Probabilité *a posteriori*

Cas du dépistage

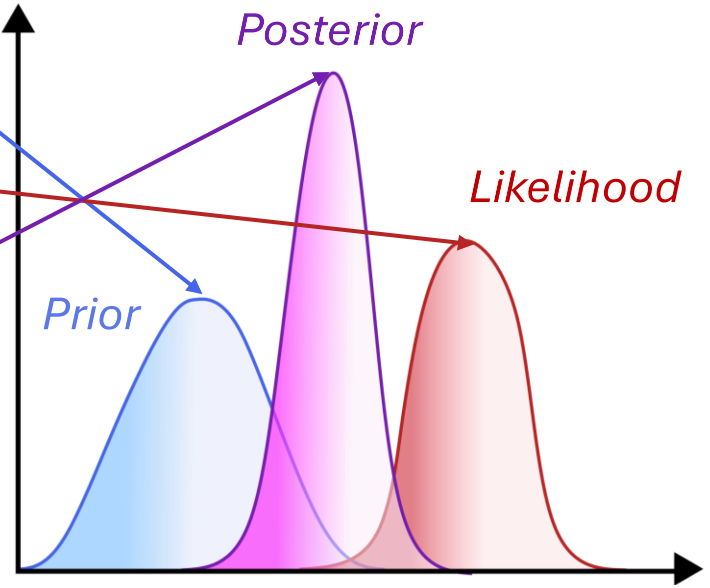
Terme	**Représente quoi**	**Example dans le test**
Croyance A priori (*prior*)	Ce que je pense avant de voir le test	$P(Mal) = 1\%$
Vraisemblance (*Likelihood*)	Probabilité d'observer le test positif *si la personne est malade*	$P(Pos)$
Croyance *A posteriori* (*posterior*)	Ce que je crois après avoir 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 posteriori* (*posterior*)	Ce que je crois après avoir vu le test	P (Dé = D10)

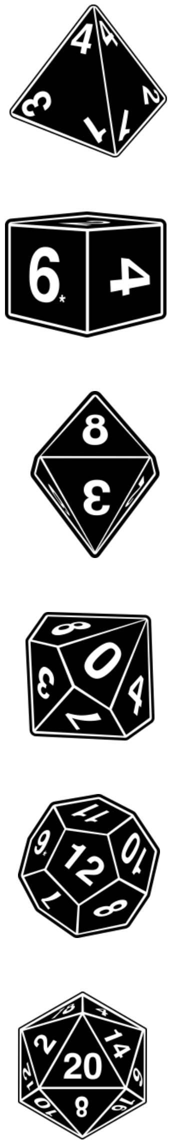


- La croyance *A posteriori* (ou *posterior*) représente la combinaison de l'information qui provient de l'observation des données et nos croyances *A priori* (*prior*)

Vraisemblance vs. Probabilité *a posteriori*

Rappel

7



Prior	Vraisemblance	Calcul
0 %		
0 %		
34,66 %	12,4	(34,66 × 12,4) : 100 = 4,30 %
27,96 %	10	(27,96 × 10) : 100 = 2,80 %
23,41 %	8,4	(23,41 × 8,4) : 100 = 1,97 %
13,98 %	5	(13,98 × 8,4) : 100 = 1,17 %

H3 = 41,99 %

H4 = 27,38 %

H5 = 19,26 %

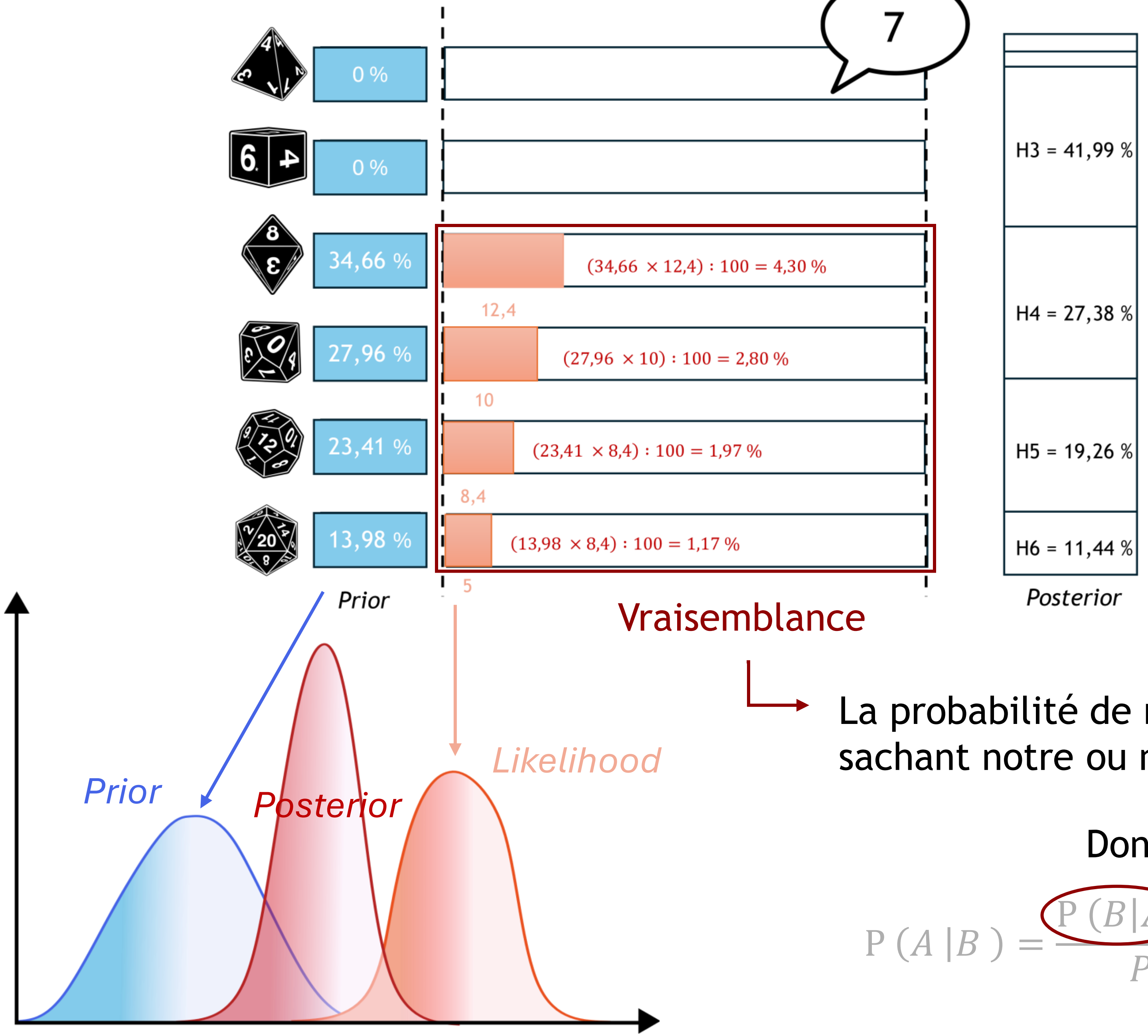
H6 = 11,44 %

Prior

Posterior

Vraisemblance vs. Probabilité *a posteriori*

Rappel



→ La probabilité de nos données sachant notre ou nos hypothèses

Donc

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

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*.

- 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 ?

© exemple tiré de Primer

(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 ?

© exemple tiré de Primer

(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 ?

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

© exemple tiré de Primer

(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 ?

© exemple tiré de Primer

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 ?

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

© exemple tiré de Primer

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 ?

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

→ Deux fois sur les deux essais ?

© exemple tiré de Primer

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 ?

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

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 $p = 0{,}17 \text{ ou } 17\%$.

→ 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.

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

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

© exemple tiré de Thierry Ancelle

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

Loi Binomiale

Caractéristique = trouble du sommeil	$\boldsymbol{k}$	0
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.

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

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

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

Loi Binomiale – cumul des probabilités

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

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

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

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

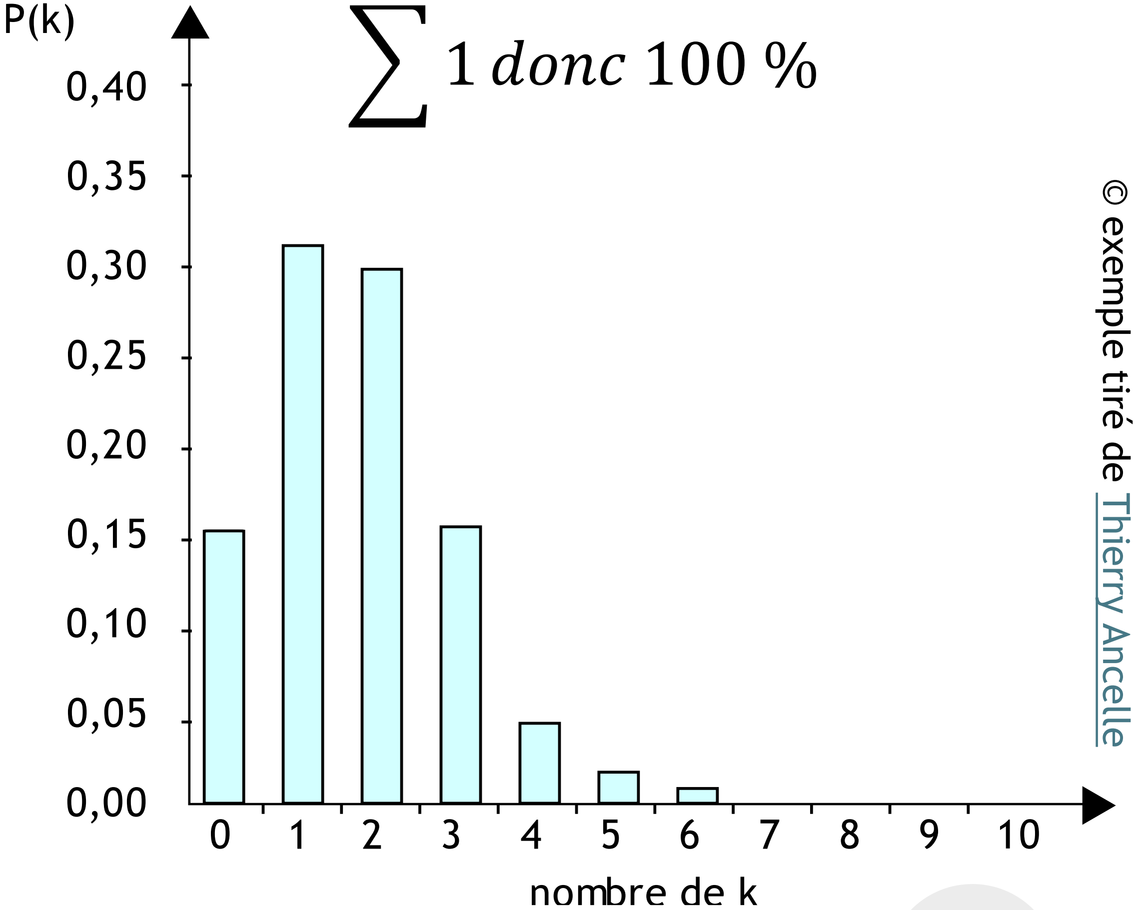
$$\Pr(4) = \frac{10!}{4!\,(10-4)!} 0{,}17^4 (1-0{,}17)^{10-4} = 0{,}057 = 5{,}7\ \%$$

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

$$\vdots$$

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

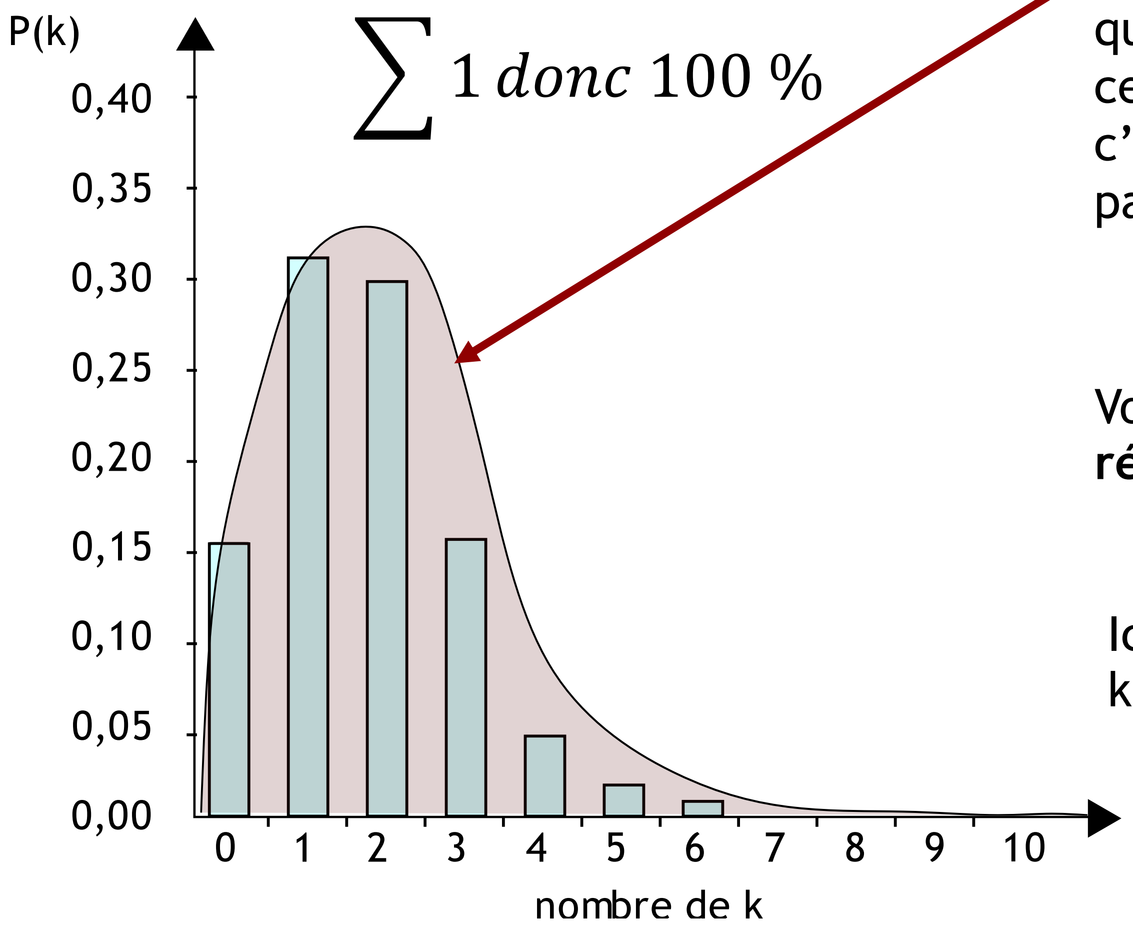
$$\sum 1\ donc\ 100\ \%$$



© exemple tiré de Thierry Ancelle

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

Loi Binomiale – cumul des probabilités

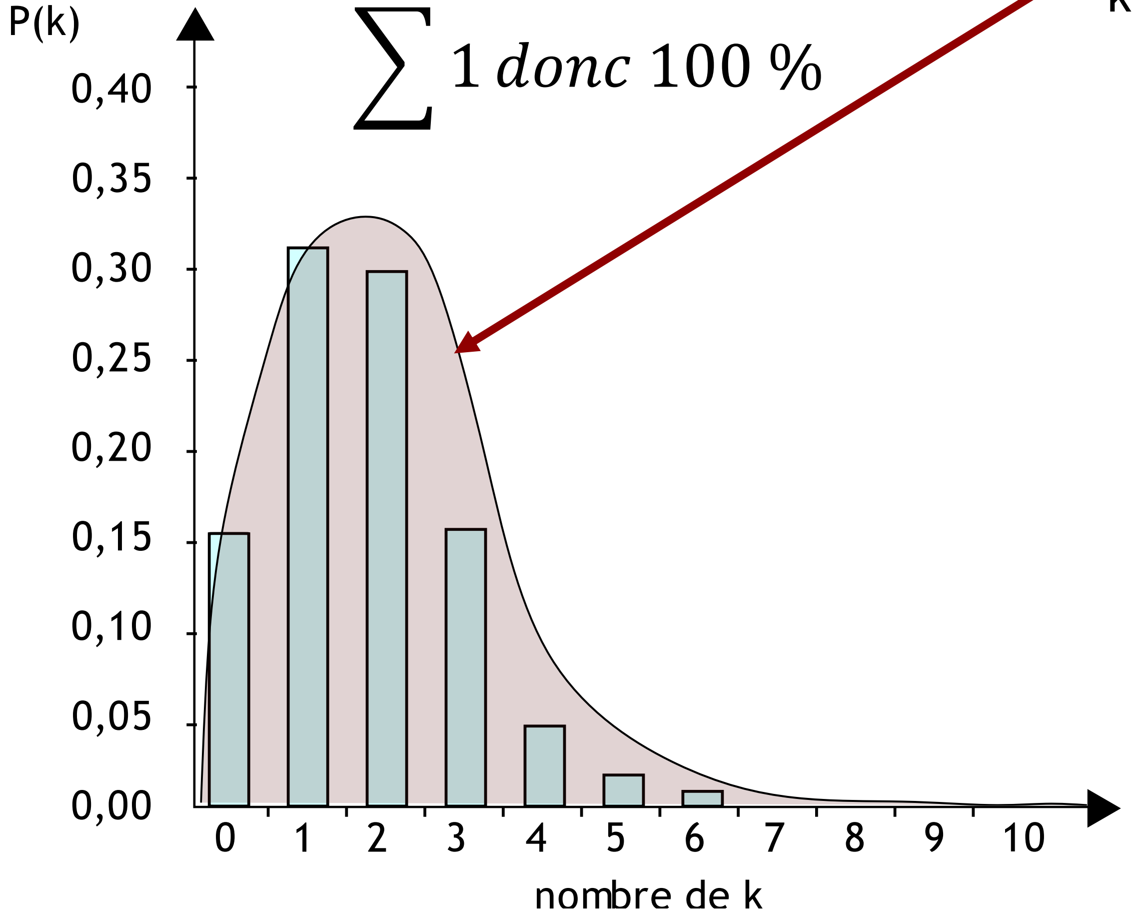


Ceci est une **fonction de densité**. La distribution théorique que vous êtes sensé obtenir si vous échantillonnez un certain nombre de participant. Elle est très importante car c'est elle qui va nous permettre de comprendre ce qui se passe statistiquement en traçant une droite de régression.

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

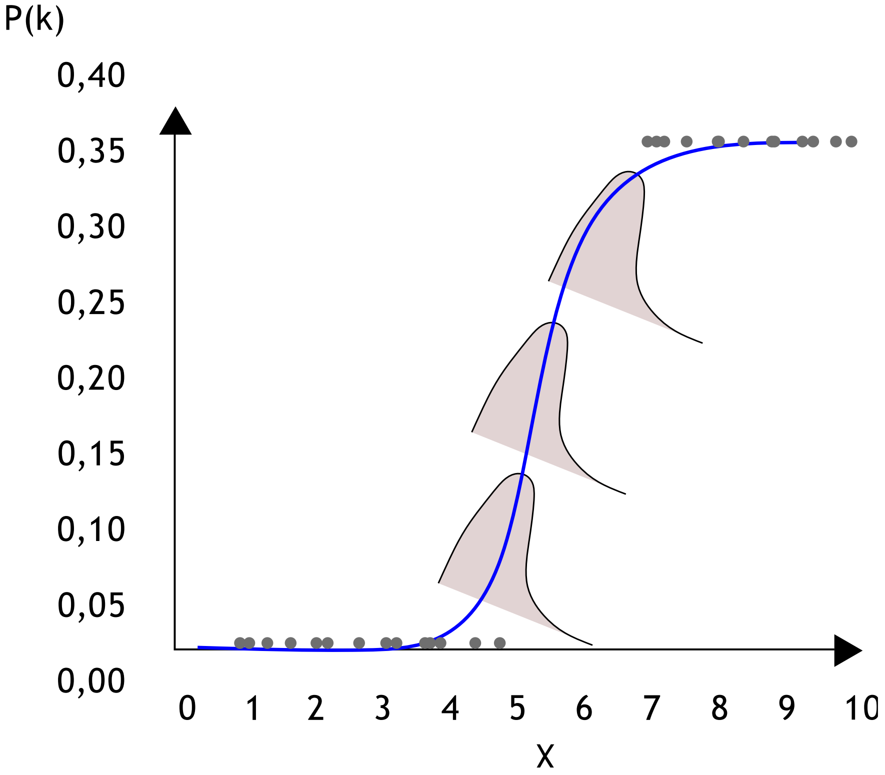
Ici ce serait une **régression logistique binomiale** puisque k ne peut prendre que deux modalités 0 ou 1.

Loi Binomiale – cumul des probabilités

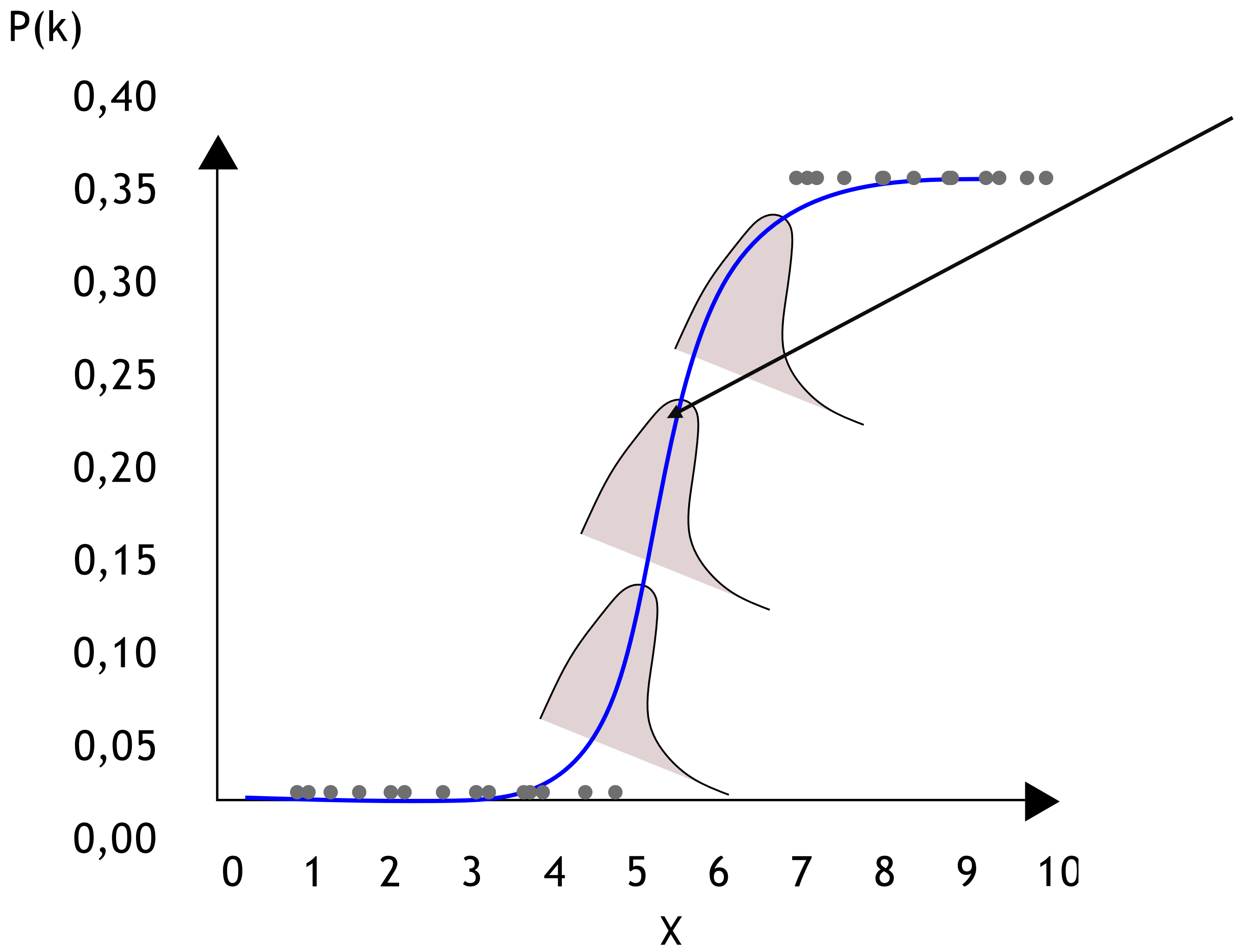


$\sum 1\, donc\ 100\ \%$

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



Loi Binomiale – cumul des probabilités



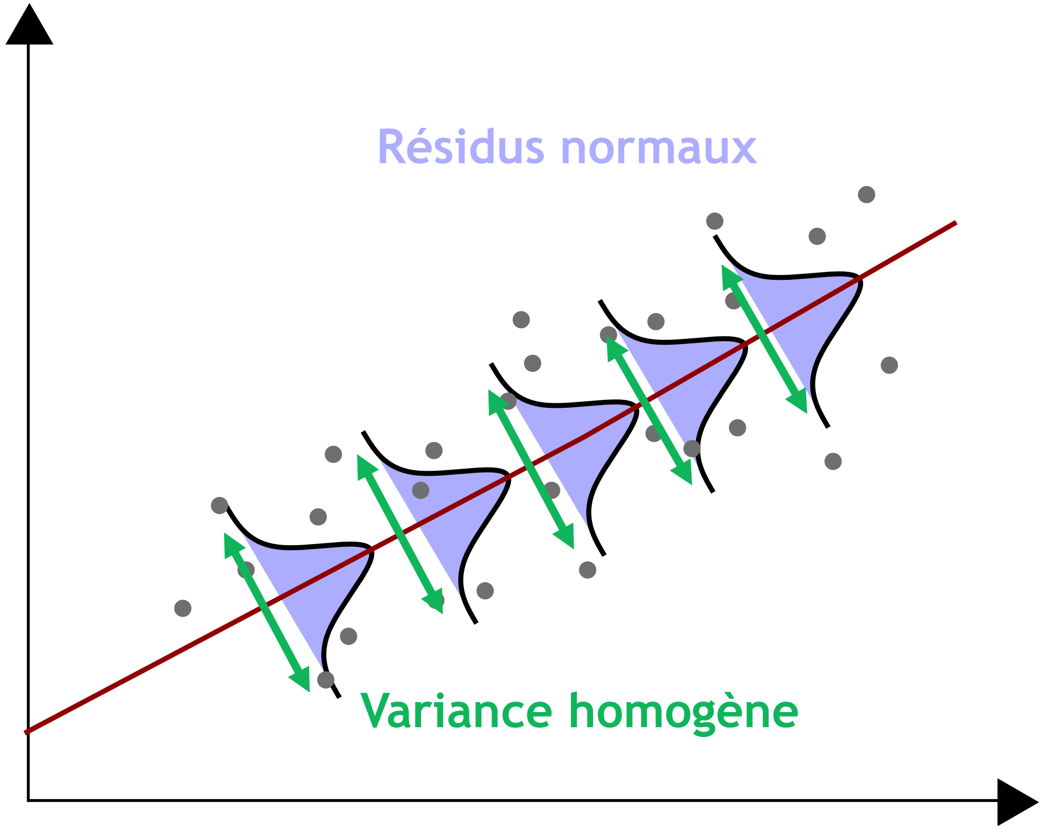
En traçant cette droite, nous pouvons comprendre la relation qui existe entre deux variables, qu'il s'agisse de deux variables quantitatives ou d'une variable quantitative et d'une variable qualitative (par exemple : groupe contrôle vs groupe test)

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.



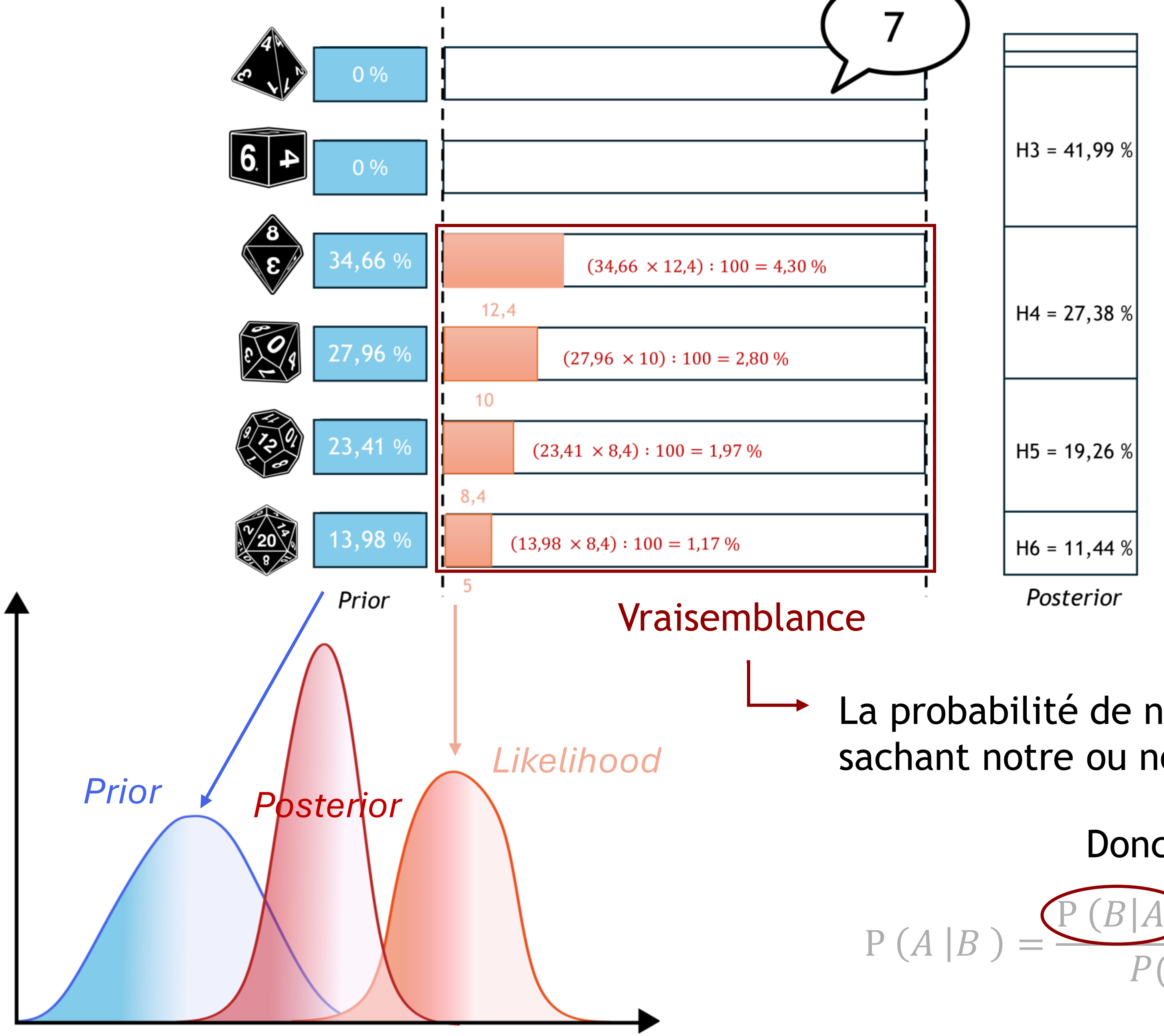
Pour tous les autres cas nous appliquons le **maximum de vraisemblance.**

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

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

Maximum de vraisemblance
Loi Binomiale

- La vraisemblance est une fonction notée :

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

$$L(\theta) = P(données\,|\theta)$$

- 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 $\{x_1, x_2, \ldots, x_n\}$, 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

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

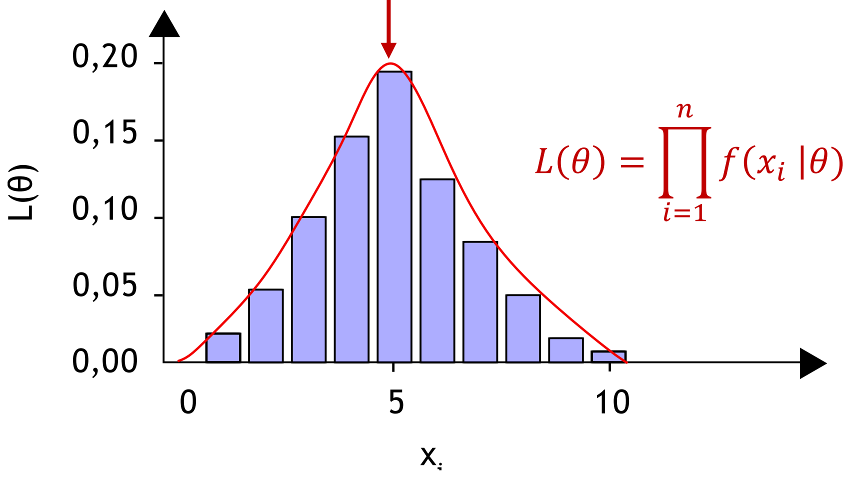
Le produit de toutes les probabilités en supposant qu'elles soient indépendantes

Maximum de vraisemblance Loi Binomiale

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

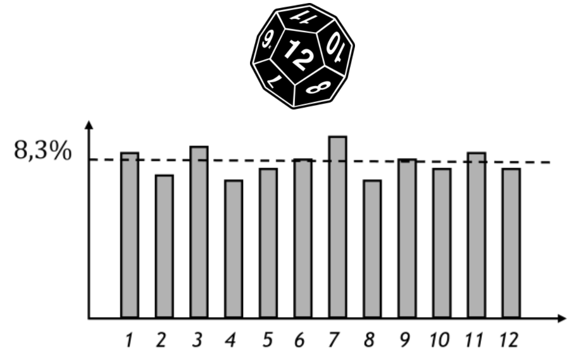
L'idée ici est de trouver la valeur de θ qui rend les données les plus probables. Formellement ça donne:

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



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

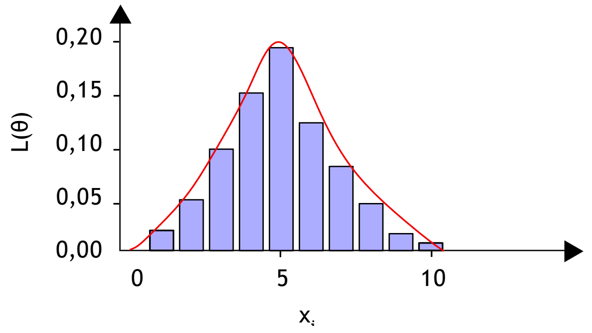
- Le maximum de vraisemblance correspond au p qui maximise cette expression

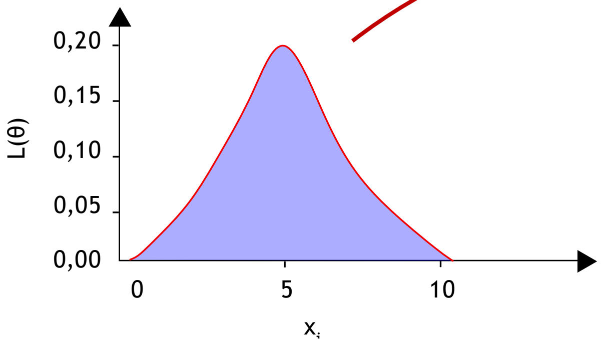
$$\hat{p}MLE = \frac{7}{10} = 0{,}7$$

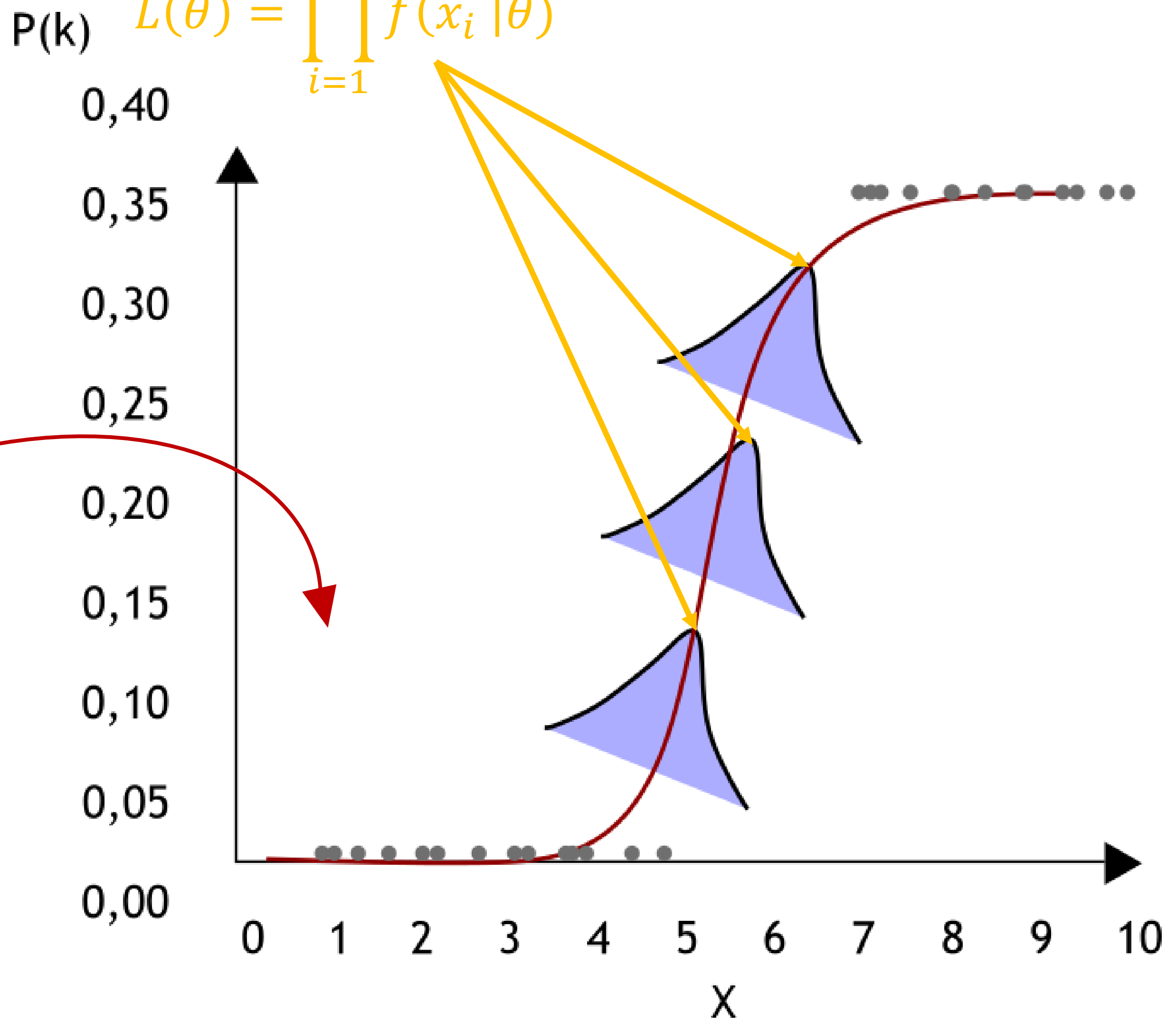
Maximum de vraisemblance

Loi Binomiale

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







Cette méthode peut s'appliquer à tous les type de distributions

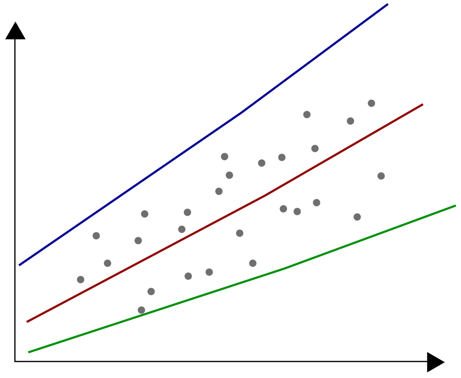
Maximum de vraisemblance dans une régression

- 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 fait 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** $\underbrace{\textbf{maximum de vraisemblance}}$

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.**

Le 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.*).

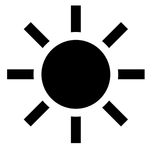
Si l'on adopte une approche plus proche de la philosophie bayésienne, puisque le maximum de vraisemblance restreint l'espace dans lequel notre modèle va chercher les paramètres à l'image du *prior*. Par exemple, un temps de réaction de 1000 secondes n'est pas réaliste ; de même, une loi Gamma n'ira pas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

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.

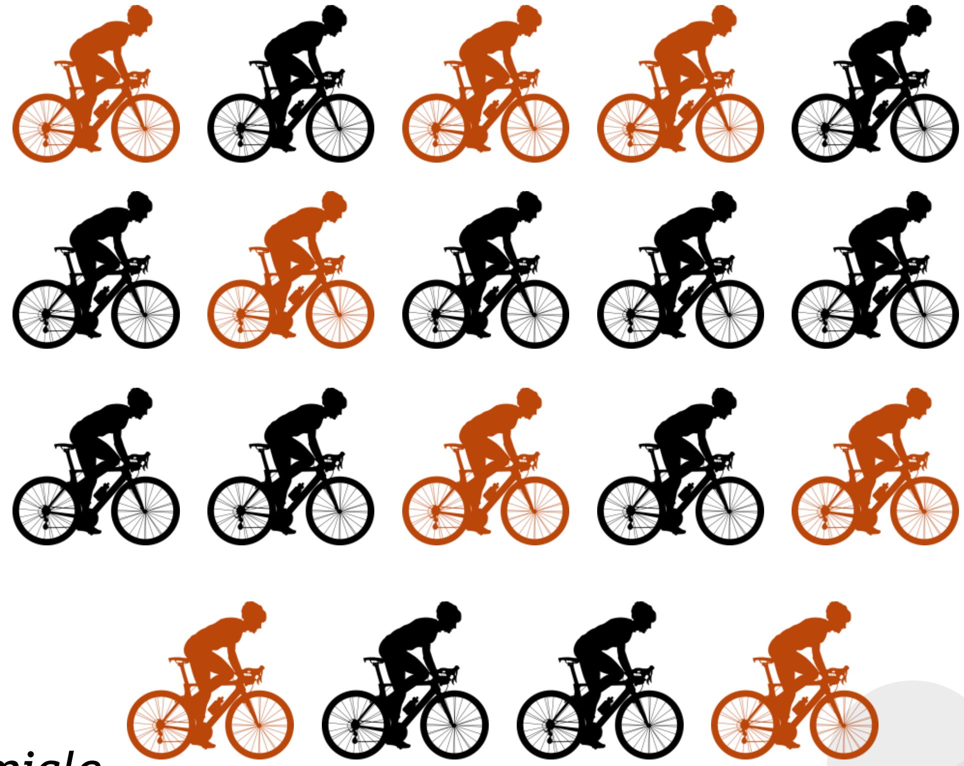
  → Je sais que l'un des principaux symptômes de dopages est un taux d'hématocrite (i.e., quantité de globules rouges dans le sang) particulièrement élevé – 50%

- 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

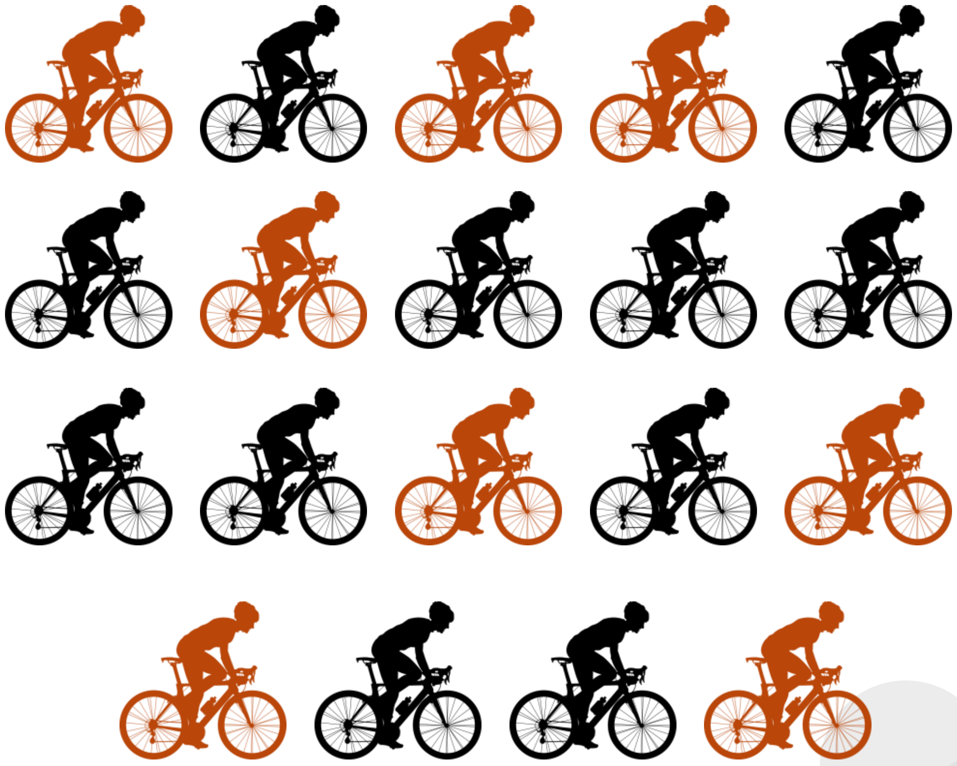
- On sait que dans la population générale seul 13 % des sujets ont un hématocrite

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

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

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

$$p = 0{,}002$$



© exemple tiré de Thierry Ancelle

- 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

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%

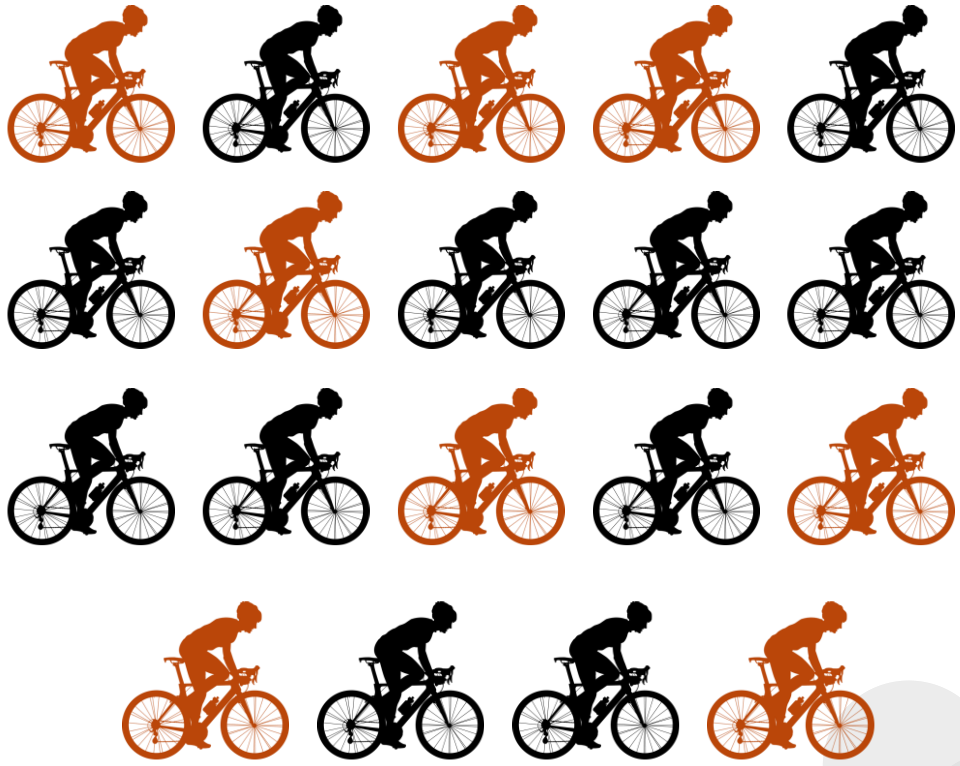
- On peut donc réaliser un autre test en évaluant le pourcentage de personnes ayant un taux d'hématocrite supérieur à 50 % par rapport à une probabilité attendue de 50 % si 100 % des 19 coureurs avait consommé un produit dopant.

  → 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 :

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

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

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

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.

© exemple tiré de Thierry Ancelle

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

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

- 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.**

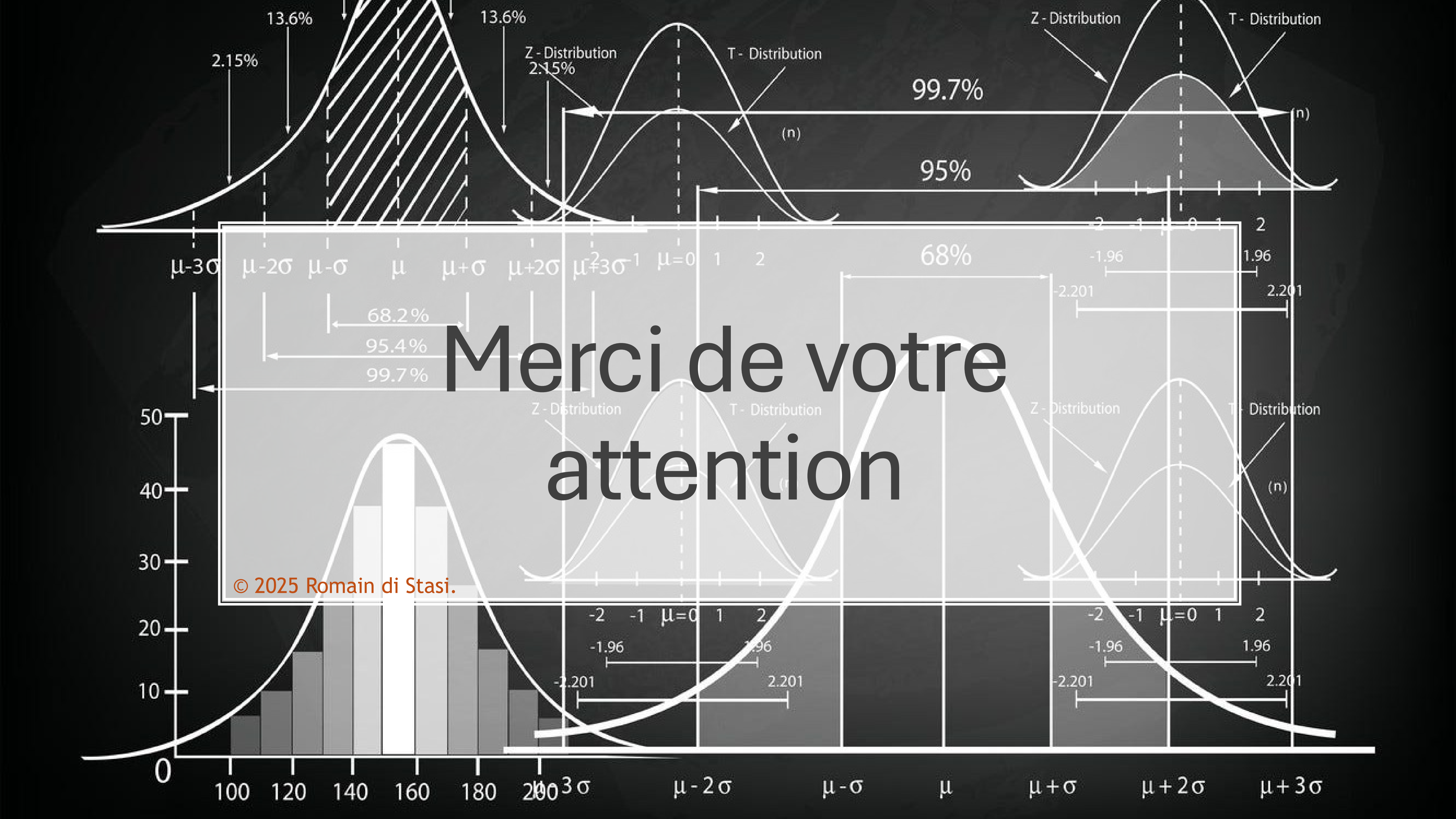
© exemple tiré de Thierry Ancelle

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 ?

$$\frac{\mathrm{P(D}\,|H_1)}{\mathrm{P(D}\,|H_2)} = 2 \longrightarrow$$

Le résultat est deux fois plus vraisemblable sous H_1



Merci de votre attention