Introduction à l'inférence bayésienne

Cours 2 : le théorème de Bayes

Valentin ROUSSEL

Laboratoire S2HEP, Université de Lyon, France

Résumé du cours - Le bayésianisme est une forme d'épistémologie qui connaît un succès croissant dans de nombreux domaines du savoir. En confrontation direct avec le fréquentisme, le bayésianisme suggère de faire usage de l'inférence bayésienne dans le raisonnement scientifique comme d'un critère de démarcation entre la rationalité et l'irrationalité. Tantôt modèle mathématique, tantôt modèle de pensée, le bayésianisme tente de modéliser les formes de croyances en leur attribuant des degrés de crédibilité pouvant prendre des valeurs de 0 (« Je ne crois absolument pas en ... ») à 1 (« J'ai la certitude absolue que ... »). La force du modèle de pensée bayésien est en ceci paradoxale qu'elle est peut-être également sa plus grande faiblesse : son originalité, son anticonformisme à l'inférence fréquentiste – pourtant majoritaire dans les études scientifiques –, et la place prépondérante qu'elle accorde à la subjectivité de l'agent qui s'en réfère. Ce cours en plusieurs parties n'a pas la prétention d'être parfaitement exhaustif ; il propose en revanche une exploration mathématique et épistémologique du concept et des principaux mécanismes à l'œuvre dans la pensée bayésienne.

Mots-clefs - bayésianisme, inférence bayésienne, inférence fréquentiste, probabilités, épistémologie

Introduction

Nous commencerons ce cours par quelques rappels de mathématiques concernant le conditionnement de deux événements. Ceci nous conduira à l'expression du théorème de Bayes et de la formule généralisée de Bayes. Nous conclurons cette première grande partie par un exemple applicatif portant l'étude du dépistage d'une maladie. Par la suite, nous reviendrons sur la dénomination de chaque terme de la formule de Bayes, ce qui nous permettra d'aborder les notions de *vraisemblance*, *probabilité a priori* - ou *crédence* -, *probabilité a posteriori* - ou *plausibilité* -, *hypothèse*, *observation* et *fonction de partition*. Enfin, nous reviendrons sur la notion de *credence* et les différentes réponses apportées aux objections faites à l'encontre du subjectivisme qu'embarque ce terme.

Pourquoi introduire le conditionnement ?

C'est l'heure de ressortir le bon vieux dé ! On joue au lancer de dé avec un dé dont les faces paires sont de couleur blanche et les faces impaires de couleur noire. On souhaite utiliser ce dé pour formuler des paris. Toutefois, de loin, on ne peut distinguer que la couleur de la face obtenue. Instinctivement, il est assez naturel de modifier les probabilités des événements. En effet, si les faces du dé étaient toutes de la même couleur, pour toute face X du dé on aurait une probabilité $\mathbb{P}(X) = 1/6$ d'obtenir cette face à la suite d'un lancé. Toutefois, dans cette situation, si l'on distingue que la face du dé est noire, il vient alors que pour toute face p paire $\mathbb{P}(X_p) = 0$ et que pour toute face i impaire $\mathbb{P}(X_i) = 1/3$ au lieu de l'équirépartition initiale de probabilité $\mathbb{P}(X) = 1/6$ pour chaque résultat X élémentaire. La connaissance de la parité du résultat modifie les probabilités que l'on donne à chaque événement, le raisonnement est surbonné à la connaissance d'une condition.

Probabilité conditionnelle à un événement

Soit $(\Omega, \mathbb{A}, \mathbb{P})$ un espace probabilisé (voir cours 1) et B un événement de $\mathbb{A}$ de probabilité non nulle. On dit que l'événement A s'est réalisé si le résultat de l'expérience ω est tel que $\omega \in A$. On sait que l'événement B s'est réalisé, donc $\omega \in B$, et pour tout événement A de $\mathbb{A}$ on a :

$$\omega \in A \Leftrightarrow A \cap B$$

Considérons alors l'application :

$$\mu : \begin{cases} \mathbb{A} \to \mathbb{R}^+ \\ \mathbb{A} \mapsto (A \cap B) \end{cases}$$

L'application μ est une mesure sur $\mathbb{A}$ mais n'est généralement pas une probabilité car :

$$\mu(\Omega) = \mathbb{P}(\Omega \cap B) = (B)$$

et n'est ainsi pas forcément égale à 1. On considére de fait l'application :

$$\mathbb{P}^B = \frac{\mu}{\mathbb{P}(B)}$$

cette dernière est bien une probabilité sur $(\Omega, \mathbb{A})$

Définition - Pour tout événement B de probabilité non nulle, on appelle probabilité conditionnelle à B sur $(\Omega, \mathbb{A})$

$$\mathbb{P}^B : \begin{cases} \mathbb{A} \to [0, 1] \\ \mathbb{A} \mapsto \mathbb{P}^B(A) = \frac{\mathbb{P}(A \cap B}{\mathbb{P}(B)} \end{cases}$$

$\mathbb{P}^B(A)$ s'appelle probabilité conditionnelle à B de A (probabilité de A sachant B) également notée $\mathbb{P}(A|B)$.

Par suite on a donc, $\forall A \in \mathbb{A}$:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) = \mathbb{P}(A)\mathbb{P}(B|A)$$

Formule de Bayes

(Démonstration : voir cours 1)

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)} = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)}$$

***

Anatomie de la formule

$$\mathbb{P}(H_i|\Theta) = \frac{\mathbb{P}(H_i \cap \Theta)}{\mathbb{P}(\Theta)} = \frac{\mathbb{P}(\Theta|H_i)\mathbb{P}(H_i)}{\sum_{j \in J} \mathbb{P}(\Theta|H_j)\mathbb{P}(H_j)} \tag{1}$$

Ou encore :

$$\text{Plausibilite} = \frac{\text{Vraisemblance.Credence}}{\text{Fonction de partition}} \tag{2}$$

Ou encore :

$$\mathbb{P}(\textit{a posteriori}) = \frac{\text{Vraisemblance.}\mathbb{P}(\textit{a priori})}{\text{Fonction de partition}} \tag{3}$$

où :

➢ $\mathbb{P}(H_i|\Theta)$: est la plausibilité, ou probabilité *a posteriori* .
➢ $\mathbb{P}(\Theta|H_i)$: est la vraisemblance.
➢ $\mathbb{P}(H_i)$: est la « Credence », ou probabilité *a priori*.
➢ $\sum_{j \in J} \mathbb{P}(\Theta|H_j)\mathbb{P}(H_j)$: est la fonction de partition ou « marginale ».
➢ H_i : l'hypothèse à l'épreuve.
➢ Θ : l'observation permettant une réévaluation de l'hypothèse à l'épreuve.

Nous reviendrons sur la définition et l'usage de ces termes par la suite.

Exemple : dépistage de la schtroumpfite

Le village des schtroumpfs est touché par une épidémie de schtroumpfite, un test médical est appliqué sur les schtroumpfs du village pour déceler la maladie. On sait d'ores et déjà que :

- Si le schtroumpf est effectivement atteint, le test est positif dans 99 % des cas.
- Pour 2 % des cas, le résultat du test est positif alors que le schtroumpf est pourtant en bonne santé.
- Un schtroumpf sur 1000 est atteint de la schtroumpfite.

On souhaite désormais connaitre la probabilité qu'un schtroumpf soit atteint sachant que son test était positif. Commençons par nommer les événements :

- Notons H_1 l'événement « le schtroumpf est atteint » et H_2 l'événement « le schtroumpf n'est pas atteint ». H correspond ainsi ici à notre hypothèse théorique. Cette hypothèse ne peut prendre que deux états : H_1 et H_2. Il n'existe en effet pas d'autre état ni d'état intermédiaire, soit le schtroumpf est malade, soit il ne l'est pas.
- Notons Θ_+ l'événement « le résultat du test est positif » et Θ_- l'événement « le résultat du test est négatif ». Θ correspond ici à l'observation faite, qui doit nous permettre la réévaluation de l'hypothèse H. Cette observation ne peut prendre que deux états Θ_+ et Θ_-, soit le test est positif, soit il est négatif.

On a alors :

- $\mathbb{P}(\Theta_+|H_1) = 0.99$
- $\mathbb{P}(\Theta_+|H_2) = 0.02$
- $\mathbb{P}(H_1) = 0.001$

Rappelons l'expression de la formule de Bayes :

$$\mathbb{P}(H_i|\Theta) = \frac{\mathbb{P}(H_i \cap \Theta)}{\mathbb{P}(\Theta)} = \frac{\mathbb{P}(\Theta|H_i)\mathbb{P}(H_i)}{\sum_{j\in J}\mathbb{P}(\Theta|H_j)\mathbb{P}(H_j)}$$

De fait on a :

$$\mathbb{P}(H_1|\Theta_+) = \frac{\mathbb{P}(\Theta_+|H_1)\mathbb{P}(H_1)}{\sum_{j\in J}\mathbb{P}(\Theta_+|H_j)\mathbb{P}(H_j)} \tag{1}$$

Or, H peut prendre 2 états H_1 ou H_2, d'où :

$$\sum_{j\in J}\mathbb{P}(\Theta|H_j)\mathbb{P}(H_j) = \mathbb{P}(\Theta_+|H_1)\mathbb{P}(H_1) + \mathbb{P}(\Theta_+|H_2)\mathbb{P}(H_2) \tag{2}$$

Finalement :

$$\mathbb{P}(H_1|\Theta_+) = \frac{\mathbb{P}(\Theta_+|H_1)\mathbb{P}(H_1)}{\mathbb{P}(\Theta_+|H_1)\mathbb{P}(H_1) + \mathbb{P}(\Theta_+|H_2)\mathbb{P}(H_2)} \tag{3}$$

$$\mathbb{P}(H_1|\Theta_+) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.02 \times 0.999} = \frac{1}{21} = 4,76\% \tag{4}$$

En définitif, la probabilité qu'un schtroumpf soit réellement malade sachant que son test est positif est inférieure à 5%. **Le test proposé ne semble *a posteriori* pas très efficace**.

Définitions

Plausibilité - Dans l'exemple précédent, $\mathbb{P}(H_1|\Theta_+)$ est la *plausibilité* de l'hypothèse H_1 au vu du résultat de l'observation Θ_+. Elle correspond au degré de confiance *a posteriori* que l'on peut accorder au test au regard de toutes les informations dont on a connaissance et qui conditionnent ce résultat. La vraisemblance permet d'évaluer le niveau de crédibilité d'une hypothèse, ce qui est en quelque sorte la finalité du bayésianisme.

Vraisemblance - Dans l'exemple précédent, $\mathbb{P}(\Theta_+|H_1)$ est la *vraisemblance* du résultat de l'observation Θ_+ dans le cadre de l'hypothèse H_1. La vraisemblance décrit la plausibilité d'une valeur des paramètres d'un **modèle**, étant donné l'observation d'un certain nombre de réalisations d'une variable aléatoire. En inférence bayésienne, la vraisemblance est la densité de probabilité des données conditionnellement à une valeur des paramètres et comme une mesure de l'information apportée par les données sur la valeur des paramètres.

« Credence » - Nous impruntons ici le terme de « credence » à Lê Nguyên Hoang **(Hoang, 2018)**. La « credence », ou probabilité *a priori* mesure la force de conviction, exprimée en pourcentage, d'un agent rationnel, à propos d'une hypothèse. Par exemples :

(1) Un agent rationnel est confronté à la situation suivante : un sac contient 4 billes rouges et 1 bille bleue et il est demandé à l'agent rationnel de retirer l'une des billes du sac au hasard. L'agent rationnel doit alors croire, *a priori*, avoir 80 % de chance de tirer une boule rouge.
(2) Un agent rationnel est confronté à la situation suivante : un lancer de pièce. L'agent rationnel doit alors croire à 50 % que la pièce tombera sur face et 50 % que la pièce tombera sur pile.
(3) Un agent rationnel est confronté à la situation suivante : il arrive au village des schtroumpfs. 100.000 schtroumpfs habitent à schtroumpfville, et 100 schtroumpfs ont signalé être atteints de schtroumpfite. L'agent rationnel doit alors croire que chaque schtroumpf qu'il rencontre a $\frac{100}{100.000} = \frac{1}{1000} = 0.001\%$ de chance d'être atteint de schtroumpfite.

Oui mais :

➢ Dans la situation (1) : le sac est peut-être truqué ? Peut-être qu'il y a un double-fond ?
➢ Dans la situation (2) : la pièce est peut-être pipée ?
➢ Dans la situation (3) : « 100 schtroumpfs ont signalé être atteints de schtroumpfite », mais peut-être qu'il y a davantage de schtroumpfs atteints de schtroumpfite ? Peut-être que les schtroumpfs malades ne se sont pas tous signalés ?

Pour ces raisons, et à défaut d'avoir plus d'informations, l'agent rationnel doit considérer les probabilités *a priori* de chaque événement. Ces probabilités sont toutefois *subjectives* : les probabilités annoncées sont celles auxquelles l'agent rationnel est en mesure de s'attendre.

Fonction de partition - Une famille finie ou dénombrable d'ensembles non vides $(A_i)_{i\in I}$ est dite partition de Ω si on a : $A_i \cap A_j = \emptyset$ pour $i \neq j$ et $\Omega = \cup_{i\in I} A_i$. Supposons que la famille $(A_i)_{i\in I}$ soit une telle partition Ω. On suppose que, $\forall i \in I$, on ait $\mathbb{P}(A_i) \neq 0$. Pour tout événement B de probabilité non nulle on peut alors écrire :

$$\mathbb{P}(B) = \mathbb{P}(\Omega \cap B) = \mathbb{P}\left(\left(\bigcup_{i\in I} A_i\right) \cap B\right) = \sum_{i\in I} \mathbb{P}(A_i \cap B) = \sum_{i\in I} \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

Hypothèse - L'hypothèse est ici une sortee de pari épistémique, une conjecture que l'on cherche à inférer à l'aide de révisions successives du degré de crédibilité accordé à cette hypothèse. Ces révisions successives de l'opinion sont appuyées par des observations, permettant ainsi de rehausser ou abaisser le degré de crédibilité que l'on accorde à cette hypothèse. Par exemple, un agent rationnel tire d'une boite opaque une bille noire, puis une seconde, puis une troisième : il fait alors l'hypothèse que la boite ne contient que des billes noires. Au quatrième tirage, il tire une bille rouge. Cette nouvelle observation lui permet de réviser son hypothèse initiale qui devient : la boite contient une majorité de billes noires.

Observation - Un événement qui nous permet, suivant la règle des probabilités conditionnelles, la révision d'une hypothèse.

Inférence bayésienne : un exemple d'application

Pour cet exemple, je m'inspire très largement de l'excellente vidéo de Christophe Michel, « La pensée Bayésienne » [1] **(Michel, 2018)**. Je reprends son expérience des dés et détaille ici les calculs et processus à l'œuvre.

Un agent rationnel est confronté à la situation suivante :

Une boite contient 5 dés :
— Un dé à 20 faces
— Un dé à 12 faces
— Un dé à 8 faces
— Un dé à 6 faces
— Un dé à 4 faces

Une machine a pour consignes de choisir un dé, aléatoirement, le lancer et communiquer à l'agent uniquement le numéro de la face apparente. La machine conserve alors le même dé jusqu'à la fin de l'expérience et l'agent peut demander autant de fois qu'il le souhaite à la machine de relancer ce dé. L'agent cherche alors à répondre à cette question : **quel dé a été choisi par la machine ?** Par la suite nous noterons chaque étape intermédiaire par l'indice T.

Notons $\mathbb{P}(H_0)$ l'hypothèse : «Le dé tiré est le dé à 20 faces»
Notons $\mathbb{P}(H_1)$ l'hypothèse : «Le dé tiré est le dé à 12 faces»
Notons $\mathbb{P}(H_2)$ l'hypothèse : «Le dé tiré est le dé à 8 faces»
Notons $\mathbb{P}(H_3)$ l'hypothèse : «Le dé tiré est le dé à 6 faces»
Notons $\mathbb{P}(H_4)$ l'hypothèse : «Le dé tiré est le dé à 4 faces»

Notons que les probabilités « *a priori* » de tirer n'importe quel dé sont de $\frac{1}{5}$, soit 20% de chance. Ainsi, initialement et *a priori* :

$$\mathbb{P}(H_0) = \mathbb{P}(H_1) = \mathbb{P}(H_2) = \mathbb{P}(H_3) = \mathbb{P}(H_4)$$

T.1 : « Le numéro est 7 »

Avec un dé à 20 faces, il y a $\frac{1}{20}$ chance d'obtenir un 7. D'où : $\mathbb{P}(\Theta_7|H_0) = 5\%$

Avec un dé à 12 faces, il y a $\frac{1}{12}$ chance d'obtenir un 7. D'où : $\mathbb{P}(\Theta_7|H_1) = 8.3333...\%$

Avec un dé à 8 faces, il y a $\frac{1}{8}$ chance d'obtenir un 7. D'où : $\mathbb{P}(\Theta_7|H_2) = 12.5\%$

Avec un dé à 6 faces et avec un dé à 4 faces il n'y a aucune chance d'obtenir un 7. D'où : $\mathbb{P}(\Theta_7|H_3) = \mathbb{P}(\Theta_7|H_4) = 0$

A ce stade, la vraisemblance $\mathbb{P}(\Theta_7|H_2)$ indique qu'il est plus probable que le dé choisi par la machine soit le dé à 8 faces. Peut-on affiner le degré de confiance en cette hypothèse ? L'inférence bayésienne demande de ne pas tenir compte des *vraisemblances*, mais de la *plausibilité* de chaque hypothèse au regard de ces vraisemblances. Ce que l'on cherche à connaitre ici, ça ne sont pas les $\mathbb{P}(\Theta_7|H_i)$, mais les $\mathbb{P}(H_i|\Theta_7)$.

Calculons déjà la *marginale*, notée μ :

$\mu = \mathbb{P}(\Theta_7|H_0) \times \mathbb{P}(H_0) + \mathbb{P}(\Theta_7|H_1) \times \mathbb{P}(H_1) + \mathbb{P}(\Theta_7|H_2) \times \mathbb{P}(H_2)$
$\mu = 0.05 \times 0.2 + 0.083333 \times 0.2 + 0.125 \times 0.2 = 0.051666$

On connait alors tous les termes nous permettant de calculer les plausibilités associées :

$$\mathbb{P}(H_0|\Theta_7) = \frac{\mathbb{P}(\Theta_7|H_0)\mathbb{P}(H_0)}{\mu} = \frac{0.05 \times 0.2}{0.051666} = 0.19 \quad (1)$$

$$\mathbb{P}(H_1|\Theta_7) = \frac{\mathbb{P}(\Theta_7|H_1)\mathbb{P}(H_1)}{\mu} = \frac{0.083333 \times 0.2}{0.051666} = 0.32 \quad (2)$$

1. https ://www.youtube.com/watch ?v=x-2uVNze56s

$$\mathbb{P}(H_2|\Theta_7) = \frac{\mathbb{P}(\Theta_7|H_2)\mathbb{P}(H_2)}{\mu} = \frac{0.125 \times 0.2}{0.051666} = 0.48 \tag{3}$$

A ce stade, $\mathbb{P}(H_2|\Theta_7)$ indique qu'il est plus plausible que le dé choisi par la machine soit le dé à 8 faces. On souhaite désormais affiner ce degré de confiance par réévaluations successives. On demande à la machine de relancer le dé :

T.2 : « Le numéro est 3 »

A ce stade, l'agent rationnel révise alors une première fois ses croyances *a piori* . Les hypothèses H_3 et H_4 n'ont plus lieu d'être car il est impossible que le dé choisi par la machine soit le dé à 6 faces ou à 4 faces. Les anciennes croyances *a posteriori* $\mathbb{P}(H_0|\Theta_7)$, $\mathbb{P}(H_1|\Theta_7)$ et $\mathbb{P}(H_2|\Theta_7)$ deviennent les nouvelles croyances *a priori*, de manière à prendre en compte les observations précédentes, toujours au regard des probabiliés conditionnelles. Cette transpotions des anciennes croyances *a posteriori* en nouvelles croyances *a priori* est un processus fondamental suivant la démarche de l'inférence bayésienne.

On a donc :

$\mathbb{P}(H_0|\Theta_7) = \mathbb{P}(H_0)_{nouveau} = 0.19$
$\mathbb{P}(H_1|\Theta_7) = \mathbb{P}(H_1)_{nouveau} = 0.32$
$\mathbb{P}(H_2|\Theta_7) = \mathbb{P}(H_2)_{nouveau} = 0.48$

Et :

$\mathbb{P}(\Theta_3|H_0) = \frac{1}{20} = 5\%$
$\mathbb{P}(\Theta_3|H_1) = \frac{1}{12} = 8.3333\%$
$\mathbb{P}(\Theta_3|H_2) = \frac{1}{8} = 12.5\%$
Ces probabilités là ne changent pas.

En revanche, il faut calculer à nouveau la marginale :

$\mu = \mathbb{P}(\Theta_3|H_0) \times \mathbb{P}(H_0)_{nouveau} + \mathbb{P}(\Theta_3|H_1) \times \mathbb{P}(H_1)_{nouveau} + \mathbb{P}(\Theta_3|H_2) \times \mathbb{P}(H_2)_{nouveau}$
$\mu = 0.05 \times 0.19 + 0.08333 \times 0.32 + 0.125 \times 0.48 = 0.09616$

On connait alors tous les termes nous permettant de calculer les plausibilités associées :

$$\mathbb{P}(H_0|\Theta_3) = \frac{\mathbb{P}(\Theta_3|H_0)\mathbb{P}(H_0)_{nouveau}}{\mu} = \frac{0.05 \times 0.19}{0.09616} = 0.098 \tag{1}$$

$$\mathbb{P}(H_1|\Theta_3) = \frac{\mathbb{P}(\Theta_3|H_1)\mathbb{P}(H_1)_{nouveau}}{\mu} = \frac{0.083333 \times 0.32}{0.09616} = 0.271 \tag{2}$$

$$\mathbb{P}(H_2|\Theta_3) = \frac{\mathbb{P}(\Theta_3|H_2)\mathbb{P}(H_2)_{nouveau}}{\mu} = \frac{0.125 \times 0.48}{0.09616} = 0.621 \tag{3}$$

Le degré de crédibilité de l'hypothèse H_2 est donc réévalué de 0.48 à 0.62. Il est davantage plausible que le dé choisi par la machine soit le dé à 8 faces.

T.3 : « Le numéro est 2 »
Nous ne détaillerons pas ici les calculs qui reposent sur les mêmes principes que pour l'étape **T.2** et nous nous concentrerons uniquement sur les résultats :

$$\mathbb{P}(H_0|\Theta_2) = 0.048 \tag{1}$$

$$\mathbb{P}(H_1|\Theta_2) = 0.225 \tag{2}$$

$$\mathbb{P}(H_2|\Theta_2) = 0.7265 \tag{3}$$

T.4 : « Le numéro est 8 »

$$\mathbb{P}(H_0|\Theta_8) = 0.022 \tag{1}$$

$$\mathbb{P}(H_1|\Theta_8) = 0.1667 \quad (2)$$

$$\mathbb{P}(H_2|\Theta_8) = 0.8113 \quad (3)$$

(...) Accélérons jusqu'à T.200 (...)

T.200 : « Le numéro est 2 »

$$\mathbb{P}(H_0|\Theta_2) = 0.0016 \quad (1)$$

$$\mathbb{P}(H_1|\Theta_2) = 0.0081 \quad (2)$$

$$\mathbb{P}(H_2|\Theta_2) = 0.9903 \quad (3)$$

Tant que la machine continue de donner des numéros entre 1 et 8, l'hypothèse 2 devient de plus en plus crédible. Si bien qu'à partir du tirage 200, l'agent rationnel est en mesure d'attribuer un indice de confiance à 99% en faveur de l'hypothèse H_2. Même s'il est impossible d'écarter à 100% les hypothèses H_0 et H_1, l'hypothèse H_2 semble nettement plus crédible, et il serait possible de réitérer le processus un nombre de fois infini, jusqu'à que le taux de crédibilité, c'est-à-dire la *plausibilité* de l'hypothèse soit suffisamment significative. Ce taux peut-être plus ou moins grand selon l'importance de l'hypothèse, par exemple : entre 99% (plutôt faible) et 99.999% (très élevé).

Credence (ou probabilité *a priori*)

Cette dernière partie est l'occasion de revenir sur la notion de « credence », ou probabilité *a priori*, qui est la cible des critiques les plus souvent adressées à l'inférence bayésienne. Les « *priors* » **(Drouet, 2016)**, qui constituent les probabilités servant de point de départ à la succession des réévaluations de crédibilité d'une hypothèse, sont critiqués du fait de la subjectivité qu'ils sous-tendent. Or, comme nous l'avons vu dans la partie précédente, les *priors* sont absolument nécessaires à l'initialisation de tout processus bayésien. Quelles réponses les bayésiens apportent-ils à cette critique ?

(1) Les priors sont inconsistants - Le poids des *priors* dans la détermination des probabilités finales décroît à mesure que les informations disponibles s'accumulent. Ceci suggère qu'à l'infini, l'effet des *priors* tend à s'annuler. Les degrés de croyance tendent, comme De Finetti **(Gillies, 2000)** le montre, vers les fréquences relatives empiriquement observées. Considérons le cas de la répétition d'une expérience aléatoire, un lancer de pièce. Sous l'hypothèse d'« échangeabilité »[2] **(Drouet, 2016)** qu'il établit, De Finetti montre que le degré d'un agent rationnel, pratiquant la révision bayésienne, qui croit que le (n+1)-ème lancer donnera « pile » tend, quand le nombre de lancers tend vers l'infini, vers la fréquence relative des « pile » parmi les n premiers lancers. De plus, ce résultat vaut quel que soit le degré auquel l'agent rationnel croyait initialement, avec les n observations faites. En conclusion : quels que soient les degrés respectifs auxquels deux agents croient initialement une hypothèse, si les deux agents font les mêmes observations, alors les degrés actualisés de crédibilité qu'ils accordent à cette proposition tendent mutuellement l'un vers l'autre quand le nombre d'observations tend vers l'infini. De plus, ils tendent tous les deux vers les fréquences empiriquement observées.

(2) Le principe d'indifférence - En l'absence de raison de croire une hypothèse plus intensément qu'une autre, la rationalité requiert de croire les deux avec la même intensité. Ce principe peut être converti en une règle de choix portant directement sur les distributions de probabilité : la maximisation de l'entropie, où l'entropie d'une distribution de probabilité $\mathbb{P}$ portant sur un nombre fini de possibilités exclusives indexées sur les éléments θ d'une ensemble Θ vaut $\sum \mathbb{P}(h_\theta).log\mathbb{P}(h_\theta)$. Cette règle permet d'isoler une unique distribution de probabilité et est ainsi constitutive du « bayésianisme objectif ». Nous y reviendrons dans une prochaine partie du cours.

2. Selon cette hypothèse, deux suites de lancers d'une pièce qui ont la même longueur ont la même probabilité si elles comptent le même nombre de « pile ». Autrement dit, l'ordre ne compte pas.

Références

Drouet, I. (2016). *Le bayésianisme aujourd'hui. Fondements et pratiques*, volume 1. Editions Matériologiques, Paris, 1 edition. An optional note.

Gillies, D. (2000). *Philosophical Theories of Probability*. Routledge, UK, 1 edition.

Hoang, L. N. (2018). *La formule du savoir*. Edp sciences, France.

Michel, C. (2018). Ep26 la pensée bayesienne.