Interpréter les probabilités

Mikaël Cozic & Isabelle Drouet (IHPST)

A paraître dans *Pour La Science* (édition française du *Scientific American*, Belin), Automne 2008

Introduction

Le langage des probabilités est aujourd'hui omniprésent : dans nos conversations quotidiennes, dans les sciences, aussi bien dans leurs parties les plus théoriques que dans leurs parties les plus appliquées. Pourtant, si la théorie mathématique des probabilités fait quasiment consensus, il n'en va pas de même de la théorie philosophique des probabilités, et en particulier de leur interprétation. Formellement, une distribution de probabilité est une mesure sur un espace d'événements qui attribue à chaque événement de cet espace un poids compris entre 0 et 1 et qui est additive : si *E* et *E'* sont deux événements incompatibles, la probabilité que l'un ou l'autre advienne est la somme de leurs probabilités respectives. Et l'on définit la probabilité conditionnelle, par exemple la probabilité de *E* étant donné que *E'* est le cas, comme le rapport entre la probabilité de *E* et *E*' et la probabilité de *E*. Les divergences naissent quand on se demande ce que signifient, dans différents contextes, les énoncés qui emploient des probabilités. Que signifient par exemple : « La probabilité qu'une pièce tombe sur pile est ½.», « La probabilité qu'Uranus gagne aujourd'hui étant donné qu'il a gagné hier est élevée. », ou encore « La probabilité que tel noyau radioactif se désintègre entre *t* et *t+h* est *p*. » ? Dans son livre célèbre *The Emergence of Probability*, I. Hacking fait valoir que le calcul des probabilités est marqué dès sa naissance par une dualité fondamentale d'interprétation avec, d'un côté, une conception des probabilités qui les tire vers les statistiques et, de l'autre, une conception des probabilités qui les met au cœur de la formation rationnelle des croyances en présence d'incertitude. Non seulement cette dualité n'a pas été résorbée, mais on dispose aujourd'hui d'un vaste spectre d'interprétations de la théorie des probabilités.

Les interprétations physiques

Selon une première famille d'interprétations, les probabilités sont des propriétés du monde physique. La proposition la plus simple consiste alors à considérer que les probabilités sont des fréquences relatives : la probabilité d'un caractère A est la proportion des A dans une certaine suite. Ainsi, la probabilité qu'une certaine pièce tombe sur *pile* est la proportion de *pile* dans une suite de lancers de cette pièce. Cette proposition s'accorde avec nos usages du concept de probabilité dans les contextes statisticiens. En outre, elle n'a aucune difficulté à rendre compte des propriétés des probabilités : les fréquences relatives satisfont sans doute possible les axiomes pour le calcul des probabilités. Toutefois, telle que nous l'avons présentée, la position fréquentiste requiert que soit précisé relativement à quelles suites les probabilités sont définies. Dans le cas de la pièce, la question est : quelle est la suite de lancers telle que la fréquence des *pile* dans cette suite est la probabilité que la pièce tombe sur *pile* ?

Les fréquentistes ont avancé deux réponses à cette question. Selon la première, qui a été défendue en particulier par le mathématicien britannique John Venn à la fin du dix-neuvième siècle, la probabilité que la pièce tombe sur *pile* est la fréquence de *pile* dans une suite réelle, et donc finie, de lancers de la pièce. La seconde réponse a été élaborée dans la première moitié du vingtième siècle, par Hans Reichenbach et Ludwig von Mises en particulier. Elle consiste à interpréter les probabilités comme des limites de fréquences relatives dans des suites infinies – et donc hypothétiques. Chacune de ces propositions se heurte à des difficultés qui lui sont propres. Concentrons-nous, toutefois, sur les difficultés qu'elles doivent affronter ensemble, et qui concernent génériquement les interprétations fréquentistes des probabilités.

La première de ces difficultés est la suivante : définir comme finies ou comme infinies les suites auxquelles les probabilités sont relatives ne suffit pas à isoler les contextes dans lesquels nous recourons à la notion de probabilité. Ainsi, ce ne sont pas les fréquences relatives dans *n'importe quelle* suite que nous appelons communément « probabilités ». A titre d'illustration, considérons la suite composée des éléments suivants : un atome d'oxygène, un éléphant, la Tour Eiffel et une bicyclette. Dans cette suite finie, la fréquence des éléments pesant moins d'une tonne est ½. Pourtant on voit mal quel sens il y a à faire de cette proportion la probabilité de peser moins d'une tonne. De façon analogue, dans une suite infinie de lancers de plusieurs dés différemment biaisés, la proportion de lancers qui donnent *6* semble n'être rien que nous appellerions « probabilité d'obtenir *6* ».

Admettons toutefois que nous ayons trouvé une caractérisation permettant d'exclure de telles suites de notre analyse des probabilités. Il reste pour les fréquentistes une seconde difficulté, qui apparaît quand on s'intéresse aux probabilités des événements uniques, singuliers. A titre d'illustration, considérons la probabilité qu'Hillary Clinton soit élue présidente des États-Unis en novembre 2008 et demandons-nous quelle valeur le fréquentisme conduit à attribuer à cette probabilité. S'agit-il de la fréquence des vainqueurs parmi les candidats à l'investiture du parti démocrate américain ? parmi les femmes candidates à la fonction de chef d'un grand État occidental ? ou même parmi les candidats à une élection dont le conjoint a déjà remporté une élection similaire ? Il apparaît que le fréquentisme laisse ouverte la question de savoir à quelle classe les événements singuliers doivent être référés. Plus fondamentalement, on peut considérer que le fréquentisme ne permet pas de rendre compte de ce que ces événements ont de singulier, puisqu'il ne conduit jamais à les envisager que comme appartenant à des classes. Selon Popper, dans le cadre fréquentiste, « un *énoncé de probabilité singulier* […] n'est singulier que sur le plan grammatical ».

L'interprétation propensionniste qu'il introduit dans le milieu des années 1950 permet de résoudre l'une et l'autre des deux difficultés que nous venons de discuter. En effet, elle consiste à la fois à attacher les probabilités à des ensembles de conditions physiques – excluant par là du domaine d'analyse les suites relativement auxquelles nous ne parlerions pas de probabilité – et à considérer que les probabilités sont d'abord des probabilités d'événements singuliers. Plus précisément, un propensionniste soutient qu'à un ensemble de conditions physiques donné sont attachées des tendances à réaliser les événements possibles, et que ces tendances sont mesurées par les probabilités. Ainsi, l'ensemble de conditions physiques constitué d'un dé, d'un dispositif de lancer de dé et d'un support destiné à recevoir le dé induit une tendance – ce que Popper appelle « propension » – à réaliser l'apparition du *6* et la probabilité de l'événement singulier « le lancer donne *6* » mesure cette propension.

Le propensionnisme est particulièrement adapté pour interpréter les probabilités de certains phénomènes quantiques, par exemple la production d'un photon par tel atome d'uranium radioactif dans les vingt minutes à venir. Elle se heurte pourtant à d'importantes difficultés. D'abord elle laisse inexpliquée la connexion entre probabilités et fréquences, et avec elle la façon dont les fréquences observées peuvent être utilisées pour évaluer les probabilités. En outre, il n'y a pas de raison de penser que les mesures de propensions ont bien les propriétés formelles des probabilités. Ensuite, la thèse selon laquelle les probabilités font référence à des propensions est ontologiquement peu économique. Enfin, l'interprétation proposée par Popper vise les probabilités absolues (telle que la probabilité que le lancer donne *6*) et il n'est pas clair si et comment elle peut être étendue aux probabilités conditionnelles (telle que la probabilité que le lancer donne *6* étant donné qu'il donne un nombre pair).

L'interprétation bayésienne

On oppose souvent aux interprétations physiques des probabilités leur interprétation bayésienne. L'interprétation bayésienne est épistémique : les probabilités sont alors des degrés de croyances. Les croyances, en effet, sont graduées : on peut croire plus ou moins que *P* est vrai, on peut croire plus en la vérité de *P* qu'en celle de *Q*, etc. La thèse fondamentale de cette interprétation est que les degrés de croyances d'un agent rationnel obéissent aux lois du calcul des probabilités. Par exemple, les degrés de croyances sont additifs : si *P* et *Q* sont deux propositions incompatibles, alors le degré de croyance en *P* ou *Q* est égal à la somme du degré de croyance en *P* et du degré de croyance en *Q*.

Pourquoi les degrés de croyances d'un agent rationnel devraient-ils obéir au calcul des probabilités ? Les tenants de l'interprétation bayésienne ont développé plusieurs arguments célèbres pour répondre à cette question. A peu près à la même époque, mais indépendamment, De Finetti (1937) et Ramsey (1926) ont construit un argument nommé le Pari Hollandais (*Dutch Book*) dont l'objectif est de montrer qu'un agent qui parie sur la base de ses degrés de croyance et dont les degrés de croyances violent le calcul des probabilités peut se voir proposer une série de paris qu'il accepterait alors même qu'elle le mènerait avec certitude à une perte monétaire (voir encadré). Plus généralement, on peut démontrer que les degrés de croyances d'un agent obéissent au calcul des probabilités si et seulement s'il est invulnérable à un Pari Hollandais. L'argument du Pari Hollandais appartient à une famille plus large d'arguments *pragmatiques* en faveur de l'interprétation bayésienne – des arguments qui prétendent montrer que la violation du calcul des probabilités engendre de l'irrationalité dans l'action ou dans la disposition à l'action.

Le Pari Hollandais

Supposons que Paul croit au degré 0.4 que *P* est vrai et au degré 0.7 que *P* n'est pas vrai (ce qui viole le calcul des probabilités); et que ses croyances sont reflétées dans ses coefficients de pari : Paul est prêt à payer 0.4 S euros pour un pari qui rapporte S euros si *P* est le cas et 0 euro sinon. Marie peut alors proposer deux paris à Paul qui lui vaudront une perte certaine : posons par exemple S = 10 euros et supposons que Marie propose le pari 1 sur *P* (pour 0.4×10 euros) et le pari 2 sur *non P* (pour 0.7×10 euros). Si *P* est le cas, alors Pierre obtiendra 10 – (0.4×10 + 0.7×10) = – 1 euro. Si *P* n'est pas le cas, alors Pierre obtiendra la même somme négative. Autrement dit, Pierre est perdant dans tous les cas de figure.

Les arguments pragmatiques en faveur du bayésianisme ont exercé une influence importante. Pour ce qui est de l'interprétation des probabilités, ils contribuent en particulier à distinguer l'interprétation bayésienne de l'interprétation propensionniste : nous l'avons vu, les propensionnistes ne donnent pas de raison de penser que les mesures de propensions devraient obéir au calcul des probabilités. En outre, contrairement à l'interprétation fréquentiste, l'interprétation bayésienne ne rencontre pas de difficulté particulière à l'endroit des probabilités d'événements singuliers.

L'interprétation bayésienne n'en soulève pas moins des questions et des difficultés assez nombreuses :

(1) pour ce qui est des fondements, la validité des arguments pragmatiques que nous venons d'évoquer est largement remise en cause. Certains, en particulier, considèrent qu'une telle approche réduit les croyances à leur rôle dans l'action et néglige leur dimension épistémologique. C'est ce qui motive la tentative récente de Joyce (1998) de fournir un argument purement épistémique (non pragmatique) à l'interprétation bayésienne. Joyce caractérise axiomatiquement un ensemble de conditions sur des mesures possibles de *précision* des croyances partielles et montre que pour toutes les mesures qui satisfont ces conditions, si les degrés de croyances d'un agent ne satisfont pas le calcul des probabilités, alors il existe des degrés de croyances qui sont strictement plus précis.

(2) pour ce qui est de l'interprétation bayésienne elle-même, elle soulève au moins deux questions :

(a) celle de savoir comment elle s'articule aux autres interprétations, et en particulier aux interprétations physiques des probabilités. La conception la plus extrême, à laquelle on rattache en général De Finetti, veut seule l'interprétation bayésienne soit valable et qu'il n'existe rien de tel qu'une probabilité objective. La difficulté est alors d'expliquer l'apparence d'objectivité que les probabilités ont dans de nombreux contextes.

(b) celle de savoir quel est le contenu exact des normes de rationalité qui sont retenues. La version radicale, dite « subjective », de l'interprétation bayésienne soutient que la seule norme qui pèse sur les croyances d'un agent rationnel est le respect du calcul des probabilités. Nous verrons dans la section suivante que d'autres versions en appellent à des contraintes plus fortes.

(3) pour ce qui est des applications philosophiques, l'une des principales réalisations de l'interprétation bayésienne est la *théorie bayésienne de la confirmation* qui se propose de caractériser en termes probabilistes les relations de confirmation entre données et hypothèses. Selon cette théorie, une théorie *T* est confirmée par la donnée *E* si la probabilité de *T* étant donné *E* est supérieure à la probabilité initiale de T. La théorie bayésienne de la confirmation se heurte à des difficultés importantes et largement débattues dans la littérature actuelle – ainsi du *problème des données connues* (voir encadré).

Le problème des données connues (*old evidence*)

Supposons que, en 1905, Albert Zweistein croit, comme la plupart de la communauté scientifique, que l'avance du périhélie de Mercure est significativement distincte de ce que prédit la théorie newtonienne et vaut environ 0.41 secondes d'arc par siècle (*E*). Il formule une nouvelle théorie *T* qui permet de prédire correctement cette avance – disons pour simplifier que *T* implique *E*. Puisque la probabilité initiale de *E* vaut 1 (la donnée est connue), la probabilité de *T* étant donné *E* est égale à la probabilité de *T* : alors même qu'A.Zweistein voit dans *E* le meilleur argument en faveur de sa théorie *T*, *E* ne confirme pas *T* du point de vue de la théorie bayésienne de la confirmation. De manière générale, une donnée connue ne peut confirmer une théorie dans le cadre bayésien.

L'interprétation bayésienne objective

Selon l'interprétation bayésienne subjective, la seule norme de rationalité qui pèse sur les degrés de croyance est le respect du calcul des probabilités. Or, il s'agit d'une exigence faible : elle autorise deux individus également rationnels et disposant des mêmes informations à entretenir des degrés de croyance différents. En particulier, il n'y a pour ces individus aucune obligation rationnelle d'endosser des degrés de croyance égaux aux probabilités physiques, si tant est que de telles probabilités existent. Ainsi, les subjectivistes considèrent qu'il n'est pas contraire à la raison de croire que la fréquence des A dans une certaine suite d'événements vaut p et de ne pas croire au degré p que le prochain événement relevant de cette suite sera un A.

La conformité non seulement avec les fréquences observées, mais plus généralement avec les informations disponibles, constitue une norme de rationalité supplémentaire que certains bayésiens font peser, en plus du respect du calcul des probabilités, sur les degrés de croyance. Cette norme s'énonce plus précisément ainsi : le degré de croyance en une proposition *P* doit appartenir au plus petit sous-intervalle de [0 ; 1] qui contient l'ensemble des valeurs compatibles avec les informations disponibles concernant *P*. A titre d'illustration, imaginons qu'un individu dispose des informations suivantes à propos d'une certaine pièce de monnaie : elle est équilibrée, l'usine qui fabrique les pièces de monnaie estime à 0,51 la fréquence de *pile* à l'occasion de leur lancer, une suite de lancers de la pièce considérée donne *pile* avec une fréquence de 0,45. La norme de rationalité qui nous intéresse veut que le degré *p* auquel l'individu croit que le prochain lancer de la pièce va donner *pile* appartienne à l'intervalle [0,45 ; 0,51]. Dans les limites de cet intervalle, les degrés de croyance d'individus qui disposeraient des informations que nous avons présentées peuvent différer.

Il existe toutefois une version de l'interprétation bayésienne qui impose, outre le respect du calcul des probabilités et la conformité aux informations disponibles, une norme de rationalité supplémentaire. Selon cette norme, les degrés de croyance d'un individu doivent être aussi neutres qu'il est possible une fois qu'ont été prises en compte les informations disponibles. Autrement dit, les degrés de coryance rationnelle d'un individu doivent être, parmi ceux compatibles avec les informations disponibles, ceux qui s'éloignent le plus de valeurs extrêmes. Dans l'exemple de la pièce de monnaie, cette contrainte supplémentaire impose de croire au degré 0,51 que le prochain lancer de la pièce va donner *pile*. Si l'on s'intéresse maintenant non plus à un degré de croyance pris isolément, mais à un ensemble de degrés de croyance, on considère généralement que cet ensemble satisfait la norme de neutralité s'il maximise une certaine grandeur appelée « entropie ». Il est possible de montrer que, étant donné certaines informations, il existe un seul ensemble de degrés de croyances 1) respectant le calcul des probabilités, 2) conforme à ces informations et 3) qui maximise l'entropie. Sous les trois normes de respect du calcul des probabilités, conformité aux informations disponibles et maximisation de l'entropie, deux individus qui disposent des mêmes informations ne peuvent donc pas à la fois être tous deux rationnels et entretenir des croyances de degrés différents. En ce sens, la version du bayésianisme qui impose ces trois normes de rationalité est objective.

La norme de maximisation de l'entropie a été introduite par Jaynes dès 1957 dans un article consacré à la mécanique statistique. Plus récemment, depuis le milieu des années 1990, le bayésianisme objectif est devenu une position particulièrement active en philosophie des probabilités. Le principal attrait de cette position réside précisément dans son objectivité : contrairement aux interprétations bayésiennes subjectives, le bayésianisme objectif n'a aucune difficulté à expliquer l'apparence d'objectivité des probabilités. En outre, il fait droit à l'intuition selon laquelle, certaines informations étant disponibles, certains degrés de croyance sont plus rationnels que d'autres. Les attraits de l'objectivité ne sont pas seulement théoriques, mais aussi pratiques : la détermination univoque des degrés de croyance par les informations disponibles facilite grandement la pratique de la science. Elle permet en particulier de développer des systèmes experts probabilistes sans avoir à révéler les croyances des experts. De façon similaire, elle autorise des gains computationnels significatifs dans le domaine de la statistique bayésienne.

Dans ces conditions, une famille de justifications du bayésianisme objectif consiste à partir de l'idée selon laquelle les degrés de croyance rationnelle doivent être objectifs et à dériver la norme de maximisation de l'entropie d'un certain nombre de contraintes qui semblent devoir peser sur les degrés de croyance rationnelle objectifs. Il existe également une seconde justification, de nature pragmatique. Elle consiste à faire valoir que les décisions prises sur la base de degrés de croyance bayésiens objectifs sont, de manière générale, des décisions prudentes. En d'autres termes, en s'en tenant aux informations dont on dispose, on prend généralement moins de risques (voir encadré).

Prudence et entropie

Paul envisage de prendre un train de la compagnie C. Il est possible que ce train, pour une raison ou pour une autre, ne circule pas. En conséquence, Paul prend un risque en allant à la gare. Imaginons maintenant que Paul apprenne que la compagnie C a dépassé la barre de 80% des trains annoncés circulant effectivement. Les informations disponibles imposent donc d'attribuer à la circulation du train qu'il compte prendre une probabilité comprise entre 0,8 et 1. La probabilité 0,8 qui vient par maximisation de l'entropie est celle qui conduit le plus souvent à ne pas prendre le risque d'aller à la gare.

Nous avons vu d'une part que l'objectivité des degrés de croyance rationnelle présente des attraits théoriques et pratiques, et d'autre part que le bayésianisme objectif se justifie indépendamment de ces attraits. Il n'en reste pas moins que la position bayésienne objective se heurte à des difficultés spécifiques. Outre des difficultés techniques sur lesquelles nous ne nous arrêtons pas ici, mentionnons

(1) une difficulté relative à la mise en œuvre du bayésianisme objectif : la tâche de maximisation de l'entropie est une tâche extrêmement complexe. Face à cette situation, les bayesiens objectifs ont proposé des méthodes permettant de maximiser l'entropie efficacement dans certains contextes familiers. Ces méthodes, toutefois, n'enlèvent rien au fait que le problème de maximisation de l'entropie n'admet pas de solution générale réalisable.

(2) une difficulté conceptuelle qui porte sur la justification du bayésianisme objectif. Quand bien même celui-ci conduirait à des actions prudentes, on peut se demander pourquoi il faudrait, précisément, être prudent. D'ou vient cette norme de rationalité ? Surtout, est-ce bien le rôle d'une *interprétation* des probabilités que d'affirmer une telle norme ? Plus généralement, le bayésianisme objectif se présente finalement au moins autant comme une théorie normative de l'évaluation des probabilités, que comme une interprétation de celles-ci.