Tout est-il aléatoire[1] ?

Glenn Shafer[2] [3]

Prisme N° 20

Décembre 2010

[1] Ce texte est la retranscription de la présentation de Glenn Shafer lors du séminaire « Probabilismes ». La vidéo est disponible sur le site du Centre Cournot, www.centre-cournot.org.

[2] Glenn Shafer est professeur à l'Université Rutgers de Newark et au Centre d'informatique du Royal Holloway College de l'Université de Londres. Il est l'auteur de cinq ouvrages et de nombreux articles scientifiques publiés dans les revues de statistique, de philosophie, d'histoire, de psychologie, d'informatique, d'économie, d'ingénierie, de comptabilité et de droit.

[3] Je remercie le Centre Cournot de m'avoir donné l'occasion de présenter ces idées dans un cadre informel qui a permis de lever certaines barrières à la compréhension, d'ordinaire inévitables dans les publications à caractère plus classique. En retravaillant la transcription de cette communication, j'ai tenu compte des questions et des commentaires qu'ont bien voulu m'adresser plusieurs collègues, en particulier Michel Armatte, Darrell Rowbottom et Vladimir Vovk.

Résumé

Tout est-il aléatoire ? Kolmogorov a répondu non, Popper a dit oui. Ma propre sensibilité me rapproche de cet empiriste de la vieille garde qu'est Kolmogorov. Lorsque nous connaissons les résultats des coups précédents avant de faire de nouveaux paris, nous pouvons formuler des probabilités qui ont une valeur objective car elles passent avec succès les tests statistiques. Cela explique le succès de nombreuses méthodes adaptatives et présente un lien avec la notion de mesure *a priori* universelle de Levin. Celle-ci nous révèle qu'en effet, tout est aléatoire, mais qu'en un sens ce fait n'a pas de valeur empirique. Dès lors que l'on comprend que le succès des méthodes adaptatives n'est pas lié à un monde aléatoire au sens réfutable, on se montre sans doute plus mesuré à l'égard de la modélisation de la causalité et plus ouvert aux méthodes atypiques de jugement de probabilité.

Sommaire

1 Introduction

« Tout est-il aléatoire ? », cette question peut connaître diverses interprétations. Je commence par en étudier trois : celle d'Andrei Kolmogorov, celle de Karl Popper ainsi qu'une nouvelle interprétation, un peu audacieuse, faisant appel à un jeu dans lequel les probabilités font office de prévisions et de paris.

Ce jeu consiste à faire des prévisions successives : chaque nouvelle prévision doit suivre l'observation du résultat des *prévisions antérieures*. Dans ce cadre, le prévisionniste peut attribuer des probabilités qui sont objectives dans la mesure où elles réussissent certains tests statistiques. J'étudie les implications de ce fait en ce qui concerne l'existence, le sens et l'usage des probabilités.

2 Les trois approches du problème

Tout est-il aléatoire ? Tout événement a-t-il une probabilité objective ? Plus précisément, dans des circonstances données et une fois que l'on a fixé les conditions du dispositif expérimental, existe-t-il une probabilité objective pour chaque issue possible[4] ? Dans les années 1950, Kolmogorov et Popper se sont divisés sur la question. Le premier a répondu « non ». Le second « oui ». Je fais de même. Cependant, je suis d'accord avec la manière dont Kolmogorov conçoit le problème, tandis que je n'apprécie guère la façon dont Popper le conçoit et y répond. Ma propre interprétation du problème sera différente à la fois de celle de Kolmogorov et de celle de Popper, si bien que j'aboutirai à la même réponse que Popper tout en n'étant pas d'accord avec lui.

Kolmogorov

Voici ce que dit Kolmogorov en 1951 dans son article sur la probabilité dans la *Grande encyclopédie soviétique* :

---

[4] Certains lecteurs défendent une autre acception du terme « stochastique ». Je les prie d'excuser son usage ici destiné à capter leur attention et, je les invite à poursuivre la lecture pour autant qu'ils s'intéressent à la question posée : y a-t-il une probabilité objective pour chaque issue une fois que l'on a fixé le dispositif expérimental ?

Dans bien des contextes, on oppose « stochastique » à « déterministe ». Ici, je confère au terme « déterministe » une valeur stochastique : lorsque l'on est sûr qu'un événement va se produire, celui-ci a une probabilité objective, *une*, il est donc stochastique selon le sens que j'accorde à ce terme.

On ne saurait faire correspondre une probabilité déterminée à chaque événement. L'hypothèse qu'il existe en effet une probabilité précise correspondant à un événement donné, dans un contexte donné, doit être vérifiée, ou justifiée, en chaque cas[5].

Il se disait la chose suivante : soit une expérience que l'on peut faire et refaire indéfiniment, parfois la pièce tombe parfois sur face, parfois sur pile. La séquence des piles ou faces se comporte-t-elle comme le prédit la théorie mathématique des probabilités ? Comme le savait Kolmogorov, une séquence donnée peut très bien ne pas se comporter conformément aux probabilités. Il se peut que la fréquence des faces ne tende pas vers une limite. Le cas échéant, elle pourra y tendre, mais pas comme on s'y attendait : elle pourra par exemple converger par le haut. C'est empiriquement que l'on peut tester et vérifier que la convergence se fait bien et qu'elle se fait normalement. Tel était son point de vue et il paraît tout à fait raisonnable.

À l'époque où Kolmogorov développe ses intuitions sur la probabilité, dans les années 1920 et 1930, l'idée que les probabilités objectives ne se rencontrent qu'occasionnellement n'a rien d'original ni d'étrange. C'est sans doute une idée que partagent la grande majorité des mathématiciens, des économistes et des philosophes qui s'intéressent aux probabilités. À la suite de la récente débâcle économique déclenchée par des banquiers qui croyaient aux probabilités objectives et des mathématiciens laissant qu'ils en estimaient les valeurs, certains commentateurs se sont souvenus que Frank Knight et John Maynard Keynes avaient déclaré, à l'instar de Kolmogorov, qu'on ne pouvait attribuer de probabilité à toutes choses. Ces commentateurs firent remarquer en même temps qu'il s'agissait là d'une opinion bien peu orthodoxe et fort étrange, mais qui méritait qu'on s'y arrête un instant. Je crois qu'il faudrait en effet remonter un peu plus loin dans l'histoire de la pensée du vingtième siècle sur les probabilités afin de nous rendre compte que Knight, Keynes et Kolmogorov disent tout haut ce que quasiment *tout le monde* pense dans les années 20 et 30. Irving Fisher le dit aussi. *Personne* ne pense qu'il y a une probabilité objective pour tout. Bruno Finetti pense de son côté qu'à toute chose correspond une probabilité, mais qu'il s'agit d'une probabilité subjective. D'après lui, il n'y a pas de probabilité objective. Emile Borel cherche un compromis : on peut

[5] Je résume ainsi un passage de la traduction en anglais de cette encyclopédie.

toujours attribuer une probabilité, qui sera plus objective dans certains cas que dans d'autres. Entre 1900 et 1930, tout le monde est d'accord avec cela. « Certes, la probabilité objective est une chose bien particulière qu'on ne saurait attribuer à tous les événements, c'est une hypothèse très particulière ».

Kolmogorov est passé à la postérité en tant que mathématicien, et non en tant que philosophe. Parmi ses apports aux mathématiques figure une petite monographie publiée en 1933, les *Fondements de la théorie des probabilités*[6]. En rassemblant en système des idées que d'autres poursuivaient depuis des décennies, cette étude avance que les probabilités mathématiques sont une branche de la théorie de la mesure. Après la seconde guerre mondiale, les axiomes de la théorie de la mesure de Kolmogorov furent considérés par les mathématiciens comme la base de référence des probabilités. Philosophes et statisticiens n'eurent d'autre choix que d'accepter leur jugement sur ce point, tout en demeurant perplexes sur certaines choses. Si la théorie de la mesure constitue un cadre mathématique adéquat pour la probabilité, que faut-il en conclure sur la signification de la probabilité, sur son interprétation, sur son utilisation pour décrire le monde, sur ce que sont en réalité les probabilités ?

Les travaux de Popper

Selon Karl Popper, la science progresse par réfutation des hypothèses, et non par leur vérification. Cette thèse fut connue dans le monde entier grâce à sa *Logique de la découverte scientifique*[7], publiée à Vienne à la fin de 1934. Lorsqu'il rédigea cet ouvrage, Popper ne connaissait pas la théorie de la mesure et n'avait jamais entendu parler de Kolmogorov. En 1938, il propose sa propre axiomatique des probabilités, sans y faire référence. Ce n'est qu'à son retour en Angleterre à la fin des années 1940, après un séjour en Nouvelle-Zélande où il s'était consacré à la philosophie politique, qu'il découvrit que tout le monde pensait que c'est à Kolmogorov que l'on devait l'axiomatique des probabilités. Ce dernier occupait à

---

[6] Kolmogorov, Andrei N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Vienne, Autriche : Springer.

[7] Popper, Karl R. (1935), *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*, Vienne, Autriche : Springer.

présent le devant de la scène. Popper fut donc contraint de s'interroger sur l'interprétation à donner au nouveau cadre de la théorie de la mesure.

Il décide d'interpréter les probabilités propres à la théorie de la mesure comme des « propensions ». Voici la proposition qu'il fait dans *Le Réalisme et la science* :

> *Je propose une* nouvelle hypothèse physique *: tout dispositif expérimental engendre certaines tendances que l'on peut parfois tester au moyen d'hypothèses de fréquences*[8]*.*

Il semble que la question de savoir si l'on peut ou non tester les propensions ne compte pas pour Popper. Elles existent de toute façon.

Le Réalisme et la science est le premier ouvrage d'une trilogie qui constitue son *Post-scriptum à la logique de la découverte scientifique*. Dans le deuxième, *L'Univers irrésolu*, il fait place aux propensions en rejetant le déterminisme sous toutes ses formes – scientifiques, religieuses et métaphysiques, ainsi qu'il les désigne. Il n'existe aucune espèce de règle, affirme-t-il, en admettant que nous connaissions à la fois cette règle et la totalité de l'information concernant l'état actuel du monde sur laquelle elle s'appuie, qui nous dise à chaque fois ce qui va se passer par la suite. Toutefois, il semble penser qu'il y a toujours une règle qui donne des probabilités sur ce qui va se passer. Popper se disait adepte du « réalisme métaphysique », on pourrait dire qu'il l'était aussi de la « stochastique métaphysique ».

Le contraste entre sa pensée et celle de Kolmogorov est frappant. Ce dernier envisage de répéter effectivement une expérience pour voir ce qui se passe. Si l'on ne constate aucune fréquence stable, qui converge comme il faut, c'est qu'aucune probabilité objective ne correspond à l'événement. Popper pour sa part ne tient pas autant à la répétition effective. Même s'il n'est pas possible de répéter l'expérience, on peut du moins *imaginer* ces répétitions, de sorte qu'une fréquence *virtuelle* apparaîtra (virtuel voulant dire imaginaire, je suppose), ce qui suffit amplement, même si la répétition imaginée est impossible. Il y a donc toujours une probabilité objective.

[8] L'ouvrage, rédigé essentiellement dans les années 1950, ne fut publié qu'en 1983. Il s'agit ici de la reformulation d'un passage de la p. 360 (édition anglaise).

La manière de penser la probabilité a été adoptée par les statisticiens, ainsi que par tous ceux qui avaient recours aux mathématiques de la probabilité après la seconde guerre mondiale. Je ne pense pas qu'il ait fait autorité dans ce domaine[9], mais il est représentatif.

Avant-guerre, la plupart des mathématiciens travaillant avec ou sur les probabilités (comme la plupart des autres scientifiques) est empiriste plutôt que réaliste[10]. Leur proximité avec l'invention des nombres réels est trop grande pour qu'ils pensent que les nombres réels sont bien réels. Ils se méfient de l'infini. Lorsque Kolmogorov défend l'axiome selon lequel la probabilité de la disjonction d'un nombre infini d'événements disjoints est la somme de leurs probabilités, il ne soutient pas que cela reflète une quelconque vérité sur la réalité. Au contraire, il déclare que son axiome est une convention mathématique, utile du point de ce point de vue mais sans conséquence pratique dans la mesure où l'on ne saurait observer une série infinie d'événements. Pour Kolmogorov et Borel, les mathématiques ne sont utiles à la science que lorsque l'on peut établir un lien avec le monde de l'expérience, qui est nécessairement fini. Le concept d'infini est un merveilleux instrument mathématique car il simplifie nos représentations mathématiques, et par conséquent les clarifie. Ces détails qui font désordre et persistent en se complexifiant davantage lorsque l'on observe un nombre croissant d'objets, disparaissent souvent à la limite. Afin d'utiliser une représentation mathématique de type infini comme modèle pratique, il faut trouver le moyen de revenir en arrière pour se détacher de l'infini et affronter la confusion du fini.

Je ne sais si Popper a lu ce que Kolmogorov dit de la nécessité de rapporter les probabilités relatives à la théorie de la mesure au monde de l'expérience. Cela n'aurait pas changé grand-chose. Il s'intéresse, non pas au monde de l'expérience, mais à la réalité, et il n'hésite pas à supposer que la réalité contient un nombre

---

[9] Popper a sans doute été le philosophe des sciences le plus lu par les scientifiques. Il a peut-être contribué à légitimer l'idée que les probabilités sont des propensions. Je ne crois pas que ses travaux sur les probabilités aient réellement compté pour les mathématiciens. Alors que j'apprécie ses analyses sur le déterminisme dans *L'Univers irrésolu*, celles qui portent sur la probabilité dans *Le Réalisme et la science* me paraissent confuses et mal fondées.

[10] Cela ne veut pas dire qu'ils se seraient présentés comme des empiristes et auraient rejeté le titre de « réalistes » ; ils avaient d'autres manières de décrire leurs idées. Pour un point de vue empiriste sur l'opposition entre empiristes et réalistes, voir van Fraassen (1980). Pour un point de vue réaliste, voir Psillos (1999).

quelconque d'infinis et de nombres réels. Il prend sciemment le parti du réalisme en opposition à l'empirisme du cercle de Vienne que partagent la plupart des scientifiques dans les années 1930. Il n'est plus aussi isolé après-guerre ; en effet, l'échec du positivisme logique, version viennoise de l'empirisme, engendre une cohorte de réalistes parmi les philosophes. L'essor considérable de la formation en mathématiques pour les ingénieurs, les statisticiens et les économistes – formation qui se caractérise davantage par sa rigueur que par ses nuances – produit une foule de spécialistes des mathématiques appliquées pour lesquels les nombres réels et l'infini ont une réalité qu'ils n'auraient jamais eue pour Kolmogorov ou Borel.

Processus stochastiques

Je voudrais attirer l'attention sur un aspect qui, dans les citations tirées ci-dessus de Kolmogorov et Popper, semble aujourd'hui démodé. Ils se demandent tous deux comment une probabilité donnée peut se manifester sous forme d'une fréquence dans une séquence d'épreuves indépendantes et identiquement distribuées. Ils n'accordent toutefois pas la même importance aux processus stochastiques dans lesquels les épreuves ne sont pas indépendantes et identiquement distribuées. Dans les années 30, il est relativement fondamental que les épreuves soient indépendantes et identiquement distribuées dans les probabilités appliquées aux statistiques, à l'économie ou à d'autres sciences appliquées. Ainsi que le souligne Jerzy Neyman en 1960, on a fait des progrès dans les probabilités. Nos modèles sont désormais la plupart du temps des processus stochastiques plus complexes.

Le concept d'épreuves indépendantes et identiquement distribuées correspond à la représentation probabiliste de l'idée d'une répétition de l'expérience. Soit p la probabilité de E à chaque fois que se répète l'expérience. Posons que

$$y_i = \begin{cases} 1 & \text{si } E \text{ survient à l'épreuve } i \\ 0 & \text{s'il ne survient pas à l'épreuve } i \end{cases}$$

La fréquence de E au cours des n premières épreuves sera alors

$$\frac{\sum_{i=1}^{n} y_i}{n}$$

Les théorèmes classiques s'intéressent à la différence

$$\frac{\sum_{i=1}^{n} y_i}{n} - p. \qquad (1)$$

Selon le théorème de Bernoulli (loi faible des grands nombres), (1) est faible avec une probabilité élevée lorsque *n* est un grand nombre. Selon le théorème de Borel (loi forte des grands nombres) (1) tend vers zéro avec une probabilité égale à un à mesure que *n* tend vers l'infini. La loi du logarithme itéré nous indique la vitesse à laquelle (1) approche de zéro à mesure que *n* est plus grand, toujours avec une probabilité égale à un.

Kolmogorov et Popper semblent d'accord sur le fait que les conditions de ces théorèmes doivent être remplies pour que *p* corresponde à la probabilité objective de *E*[11]. Ils s'opposent sur l'idée qu'on puisse avoir une certitude *a priori* à propos de l'existence d'une *p* pour laquelle ces conditions sont remplies. Selon Kolmogorov, il faut le vérifier empiriquement ; selon Popper, le dispositif expérimental définit une telle *p* même si l'on ne peut répéter l'expérience suffisamment de fois pour obtenir le comportement attendu, ou même s'il est impossible de la répéter.

La raison pour laquelle cette discussion paraît obsolète, c'est que l'on travaille désormais spontanément avec des théorèmes plus généraux. Dans un processus stochastique, il se peut qu'aucun événement n'ait la même probabilité à chaque épreuve, mais on peut s'intéresser à des événements différents au fil de l'expérience. Soit E_i un événement attendu à l'épreuve *i*, et p_i la probabilité accordée à E_i, par le modèle à l'issue des épreuves antérieures. Dans ce cas, on remplace (1) par

$$\frac{\sum_{i=1}^{n} y_i}{n} - \frac{\sum_{i=1}^{n} p_i}{n} = \frac{\sum_{i=1}^{n} (y_i - p_i)}{n} \qquad (2)$$

[11] Dans le cas de Popper, il est peut-être plus juste de dire simplement qu'il considérait que les lois de la probabilité s'appliquent. Je ne suis pas sûr que Popper ait examiné la loi du logarithme itéré, et certains passages de *Le Réalisme et la science* suggèrent qu'il ne comprenait pas la différence entre le théorème de Bernoulli et celui de Borel.

C'est ce nombre qui devrait désormais tendre à devenir très petit et même à tendre vers zéro avec une probabilité égale à 1 à mesure n devient plus grand. Les choses se passent-elles ainsi ? La loi des grands nombres et celle du logarithme itéré restent-elles valables ? Oui, il est possible de démontrer les martingales de ces théorèmes classiques. Ce sont désormais des théorèmes portant sur une suite de probabilités p_1, p_2, ... dont chacune peut dépendre de l'issue des événements antérieurs, et non de théorèmes sur une probabilité unique p spécifiée à l'avance. Il ne s'agit plus désormais d'« interprétation de la fréquence » ni d'« interprétation de la propension » d'une probabilité unique p.

On peut facilement adapter la position de Kolmogorov à cette situation plus générale : il suffit de conserver l'idée qu'il faut faire subir au modèle une vérification empirique. Les tests possibles sont nombreux : on peut vérifier la martingale de la loi des grands nombres pour diverses suites d'événements, on peut aussi vérifier de nombreuses autres prévisions pour lesquelles le modèle donne une forte probabilité. Pour la plupart, ces tests ne permettent pas de découvrir des fréquences correspondant à chacune des probabilités au sein du modèle, et pour la plupart, ces dernières ne font pas l'objet d'interprétation de leur fréquence.

Il n'est pas si facile d'adapter la position de Popper. Il imagine ce qui se passerait s'il y avait un nombre infini d'épreuves. Serait-il satisfait si cette séquence imaginaire et infinie d'épreuves se déroulait continument au fil du processus stochastique infiniment long dans lequel (2) tend vers zéro tandis que les autres événements, censés avoir une probabilité égale à un, surviennent aussi ? Ou bien chercherait-il à pousser son imagination jusqu'à interpréter chacune des probabilités pour chaque moment de la série temporelle du modèle[12] ?

Je me sens donc proche de l'empirisme de Kolmogorov tandis que le réalisme de Popper me semble irréaliste. Je n'ai cependant pas l'intention de me gausser de Popper. Je décris plutôt le contexte dans lequel je me pose moi-même la question de savoir si tout est aléatoire. À l'instar de Kolmogorov, il me semble que le problème concerne les tests statistiques. Peut-on donner des probabilités qui réussissent ces tests ?

---

[12] Isabelle Drouet (2011) a mis en évidence les nombreux paradoxes issus de la tentative d'appliquer l'idée de Popper à toutes les probabilités, y compris conditionnelles, au sein d'un modèle.

Prévisions sous forme de probabilités objectives

Y a-t-il une probabilité objective pour tout événement ? Pour rester dans un cadre général, posons la question dans l'esprit des processus stochastiques. Ne supposons pas que nous répétons la même expérience indéfiniment, quel que soit le sens donné à « même ». Contentons-nous de supposer que l'on a une suite d'événements, dont notre événement fait partie, et qu'à chaque étape nous attribuons une probabilité à l'événement suivant, en cherchant à ce que ces probabilités soient validées par les tests statistiques. À mes yeux, la validation par les tests statistiques suffit à rendre les probabilités objectives. Que peut-on demander de plus[13] ? Tel est mon point de vue. C'est ce qui me pousse à affirmer qu'en effet *tout* événement pris dans une suite d'événements, sur laquelle on s'est mis d'accord, a une probabilité objective. On peut attribuer à ces événements des probabilités qui satisfont les tests statistiques.

Le processus stochastique dont je parle ne commence pas avec des probabilités. Il n'y a pas de théorie qui vous donne les probabilités à l'avance. Puisque nous n'avons pas encore déterminé si notre processus est stochastique, parlons donc simplement d'un « processus ». Puis-je faire des probabilités au sujet de ce processus, qui satisfont les tests statistiques ? Si vous me demandez de faire des probabilités avant que j'aie pu observer aucun des résultats obtenus, si vous me demandez d'énoncer dès à présent des probabilités sur ce qui se passera demain, le lendemain, le surlendemain et ainsi de suite, je ne suis pas sûr d'être en mesure de le faire. Il faut me donner une tâche plus facile : demandez-moi de donner une probabilité pour le lendemain et ensuite, le lendemain – après que j'aurai pu observer ce qui est arrivé, ce qui s'est passé – demandez-moi de donner une probabilité pour le jour suivant. Le jour suivant, une fois que je vois ce qu'il en est, je donnerai une probabilité concernant le jour suivant, et ainsi de suite. En d'autres termes, il faut que je donne mes probabilités pour chaque épreuve une fois que j'ai

[13] Surtout avant la seconde guerre mondiale, de nombreux mathématiciens du continent européen souscrivaient au « principe de Cournot » selon lequel la théorie de la probabilité mathématique n'avait d'autre contact avec les phénomènes que celui de ses prévisions égales à 1 ou proches de 1 ; voir Shafer (2007). Il découle de ce principe que l'on ne peut tester la validité d'une théorie probabiliste qu'en vérifiant les prévisions qu'elle donne avec une probabilité égale à 1 ou proche de 1. Pour une analyse plus récente de ces idées qui semblent avoir perdu de vue la tradition plus ancienne, voir Henning (2007).

pu observer le résultat des épreuves antérieures. Ceci est dans l'esprit de ce que les spécialistes de l'apprentissage automatique appellent désormais la « prévision en ligne ». Et de fait, il est possible de faire de bonnes prévisions de probabilité en ligne[14].

Tel est le contenu technique de mon intervention : je voudrais vous persuader que je suis en mesure de donner de manière séquentielle, des probabilités qui passent avec succès les tests statistiques. En ce sens donc, tout est aléatoire. Bien sûr, si tout est aléatoire, c'est que la stochastique n'est pas aussi extraordinaire qu'on l'a laissé entendre. Ou bien, peut-être qu'en ayant la possibilité d'observer tous les résultats antérieurs, j'ai la part trop belle. Nous reviendrons plus tard sur ces questions.

3 Un jeu de prévisions

On sait faire des tests statistiques lorsque l'on nous donne des probabilités. Je vous propose de faire la même chose sans probabilités, du moins pas dès le départ. Je veux construire les probabilités au fur et à mesure. La seule façon de procéder mathématiquement, c'est sous forme de jeu. L'un des joueurs détermine les probabilités, un autre décide de l'issue à donner, un troisième fait passer les tests.

Ma discussion ne porte que sur le cas le plus simple, dans lequel l'issue est binaire : oui ou non, pile ou face, un ou zéro. On peut travailler sur des cas beaucoup plus généraux, le domaine des résultats peut même changer d'une épreuve à l'autre, mais je veux m'en tenir à des choses simples.

Prévisionniste et Réalité

Commençons avec deux joueurs que nous appelons *Prévisionniste* et *Réalité*. Prévisionniste donne une probabilité, soit zéro soit un. Réalité annonce ensuite le résultat : disons, 0 si c'est pile, 1 si c'est face. Ils continuent ainsi indéfiniment. On pourrait ne retenir qu'un nombre fini d'épreuves, mais cela compliquerait les choses.

Je suppose qu'il s'agit d'un jeu à information parfaite. Prévisionniste commence en annonçant sa probabilité p_1. Après avoir observé le choix de

[14] Voir Cesa-Bianchi et Lugosi (2006).

Prévisionniste, Réalité annonce le résultat y_1. Ils recommencent, ce qui donne p_2 et y_2. Ainsi de suite. Chaque joueur voit ce que fait l'autre au moment où il le fait.

Dans ce cadre, on peut démontrer des théorèmes concernant ce que l'un ou l'autre des deux joueurs peut faire[15]. L'un de ces théorèmes nous apprend que l'un des joueurs a une stratégie qui le fera parvenir à un certain but indépendamment de ce que fait l'autre joueur. Beaucoup de ces théorèmes sont des généralisations de théorèmes classiques de la probabilité et les prennent pour corollaires. Un autre théorème, dont je ne suis pas sûr encore de pouvoir donner une formulation précise, nous apprend que Prévisionniste peut donner des probabilités qui ne seront pas réfutées par les tests statistiques.

Souvent, on peut abandonner l'hypothèse selon laquelle chaque joueur voit ce que fait l'autre. Si la stratégie de l'un des joueurs est gagnante, elle le restera de toute évidence si son adversaire dispose de moins d'informations. Il n'est donc pas toujours essentiel, eu égard aux résultats, que tout le monde dispose de toute l'information. Cela dit, il n'est pas question de tester Prévisionniste à partir d'information dont il ne dispose pas. Les joueurs peuvent par ailleurs acquérir d'autres informations au fil du jeu. Je n'ai pas l'opportunité d'écrire dire grand-chose là-dessus, mais en pratique, il se trouve toujours des x pour vous aider à prédire les y.

Pour définir le jeu, il ne faut pas seulement en expliquer le déroulement, il faut aussi déterminer qui est le gagnant. J'y viendrai. Pour l'instant, nous avons déjà assez d'éléments sur le jeu pour commencer à parler de stratégies. Qu'est-ce qu'une stratégie pour Prévisionniste ? C'est une règle qui lui permet de savoir quelle probabilité il va annoncer en fonction des coups antérieurs de l'autre joueur, Réalité. Cela revient donc plus ou moins à une distribution de probabilité des coups y_1, y_2, ...joué par Réalité. Si Prévisionniste commence avec une distribution de probabilité P pour y_1, y_2, ..., il peut proposer sous forme p_1 la probabilité que P donne pour $y_1 = 1$, après quoi, une fois qu'il a observé y_1, $y_1 = 0$ mettons, il peut proposer sous forme p_2 la probabilité que P donne pour $y_2 = 1$ à condition que $y_1 = 0$. Ainsi de suite. En d'autres termes, P dépend toujours de ce que le joueur a pu constater jusqu'à présent. Dans ces conditions,

[15] Voir, par exemple, les théorèmes dans Shafer et Vovk (2001).

la distribution de probabilité de Réalité = la stratégie de Prévisionniste

L'idée de faire dépendre une distribution de probabilité P sur ce que l'on a pu constater jusqu'à présent afin d'obtenir une prévision de probabilité pour ce qui va se passer par la suite est souvent tenue pour « bayésienne ». Cependant, ici, la distribution « a priori » P n'est pas construite selon un modèle bayésien traditionnel. Elle n'est pas construite à partir de notre idée sur les croyances de l'autre, ni à partir de la moyenne d'un modèle statistique (une classe de distributions de probabilité) concernant les croyances subjectives à propos desquelles le modèle statistique se révèle juste. Comme nous le verrons, on la choisit dans le but de mettre en échec un test (plus ou moins) universel.

Le Sceptique

Comment effectuer un test statistique sans partir d'une distribution de probabilité ? Les tests statistiques sont effectués par un autre joueur, Sceptique, qui parie en fonction des probabilités données par Prévisionniste. On peut se représenter la probabilité de Prévisionniste sous la forme du prix d'un ticket qui rétribue en dollars le coup joué par Réalité : 1$ ou 0$. Sceptique choisit le nombre de tickets qu'il est prêt à acheter. Il débute le jeu avec un capital de 1$ et tente de le réemployer de manière à se retrouver avec un capital important, voire infini. Il agit en sorte de ne pas risquer de se trouver avec un capital courant négatif. Si vous prenez le risque que votre capital soit négatif, cela signifie que vous exposez au risque le capital d'autrui dans la mesure où vous faites un pari à crédit. En fait, Sceptique essaie de multiplier le capital qu'il risque par un facteur important, voire infini.

Selon la théorie classique des probabilités, la probabilité de multiplier à l'infini le capital risqué – c'est-à-dire la probabilité qu'une variable aléatoire non-négative prenne une valeur infiniment plus grande que sa valeur attendue – est nulle. La probabilité qu'elle prenne de nombreuses fois une valeur conforme à sa valeur attendue est faible. Ainsi, Sceptique essaie de faire une chose dont la probabilité est faible, voire nulle. Normalement, dans les tests statistiques, on rejette l'hypothèse selon laquelle une distribution de probabilité est juste lorsque survient un événement auquel cette dernière accorde une probabilité faible ou nulle. Dans notre cas, où nous n'avons pas de distribution de probabilité au départ, nous pouvons quand même rejeter l'hypothèse que Prévisionniste est compétent si Sceptique parvient à multiplier son capital par un facteur important, voire infini. C'est à Jean-André Ville, qui soutint sa thèse de doctorat devant un jury présidé par Émile Borel en 1939, que l'on doit cette manière d'effectuer les tests statistiques[16].

Je me focaliserai sur le cas où Sceptique tente de multiplier son capital à l'infini, ce qui revient à une probabilité nulle, et où il dispose d'un nombre infini d'épreuves pour tenter sa chance. Il serait trop compliqué de présenter ici le cas où Sceptique cherche à multiplier son capital par un grand nombre fini au cours d'un nombre fini d'épreuves. J'évoquerai, par exemple, la loi forte des grands nombres plutôt que la loi faible car je n'ai pas le temps d'expliquer en détails dans quelle proportion on peut multiplier le capital au cours d'un nombre donné d'épreuves si la fréquence se situe à une distance donnée de la probabilité, et ainsi de suite. Ne croyez surtout pas qu'il n'existe pas de version finitaire ; comme Kolmogorov, j'utilise des infinis parce que c'est plus simple.

Que peut faire Sceptique ?

Afin d'illustrer ce que peut faire Sceptique, je vais faire appel à la démonstration par la théorie des jeux de la loi forte des grands nombres, telle que la propose Ville.

Pour faire simple, supposons que Prévisionniste donne toujours une probabilité de un sur deux. Cela ne préjuge pas de la façon dont Réalité va choisir y_n. Je ne formule aucune hypothèse stochastique et je ne suppose pas non plus que la probabilité objective pour que $y_n = 1$ soit de 1/2 au sens où on l'entend habituellement. Cette probabilité est de un sur deux au seul sens où Prévisionniste permet à Sceptique de faire des paris à deux contre un.

Comme je l'ai expliqué, le coup de Prévisionniste, un sur deux en l'occurrence, correspond au prix du ticket qui rétribue en dollars le coup y_n de Réalité, 1\$ ou 0\$. Le gain net provenant d'un ticket est de $\$(y_n - 1/2)$. Les coups de Sceptique, que nous désignerons par S_n, consistent à choisir le nombre de tickets qu'il

[16] Pour une comparaison plus détaillée avec d'autres méthodes de test statistique, voir Shafer *et al.* (2011) et Dawid *et al.* (2011).

va acheter. S_n correspond à n'importe quel chiffre réel, positif, nul ou négatif. Si S_n est positif, Sceptique parie sur $y_n = 1$. Si S_n est négatif, c'est que Sceptique vend des tickets plutôt qu'il n'en achète, il parie sur $y_n = 0$. La somme de ses gains correspondra au nombre de tickets multiplié par le gain net pour chaque ticket :
$S_n(y_n - 1/2)$.

Selon la loi des grands nombres, la part des coups de Réalité correspondant à 1 devrait tendre vers 1/2. Ville démontre que la stratégie du Sceptique (1) empêche son capital de prendre des valeurs négatives, quoi que fasse Réalité, et (2) lui permet de devenir infiniment riche si Réalité ne réalise pas la convergence. De toute évidence, Réalité peut empêcher cette convergence et il peut aussi empêcher Sceptique de gagner de l'argent, mais selon le théorème de Ville, il ne peut faire les deux à la fois. Sceptique joue en sorte que l'une des deux se produise.

Exprimons cela à l'aide de formules. Soit K_n le capital de Sceptique à l'issue de la énième épreuve. Mettons que K_0, son capital de départ, est égal à 1. On peut dès lors décrire le jeu comme suit :

$K_0 \coloneqq 1$
Pour $n = 1, 2, \ldots$:
  Sceptique annonce $s_n \in R$;
  Réalité annonce $y_n \in \{0,1\}$;
  $K_n \coloneqq K_{n-1} + s_n\left(y_n - \frac{1}{2}\right)$.

Sceptique gagne si
  K_n n'est jamais négatif, et si
  soit $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{1}{2}$ ou $\lim_{n\to\infty} K_n = \infty$.

Remarquez qu'il n'y a désormais plus que deux joueurs, Sceptique et Réalité. (Dans la mesure où nous avons fixé la stratégie de Prévisionniste, ce n'est plus la peine de le compter comme joueur.) L'un des deux joueurs sera le gagnant, l'autre le perdant[17].

[17] Dans les jeux à deux joueurs dont l'un est gagnant et l'autre perdant, on dit par convention que le gagnant obtient 1 et le perdant – 1, on dit alors que le jeu est un jeu à somme nulle. Ce résultat de 1 ou – 1, et non le capital K_i correspond à l'« utilité » que Sceptique cherche à maximiser. Notre usage de la théorie des jeux n'est

Selon le théorème de Ville, la stratégie de Sceptique est gagnante. Ville a explicitement élaboré une telle stratégie, avec une formule où le coup de Sceptique, S_n se joue en fonction des coups précédents de Réalité. La formule exprime les choses ainsi :

$$s_n(y_1, \dots, y_{n-1}) = \frac{4\left(r_{n-1} - \frac{n-1}{2}\right)}{n+1} K_{n-1}, \text{where } r_{n-1} = \sum_{i=1}^{n-1} y_i.$$

Cette formule incite Sceptique à parier une partie de son capital actuel sur face, en sorte que $y_n = 1$. Cette partie dépend de l'écart entre le nombre de fois où « face » est tombé et la moitié du nombre total d'épreuves ayant eu lieu. S'il y a trop de « faces », Sceptique parie que Réalité sortira encore des « faces », et s'il veut vaincre Sceptique, il faut que Réalité revienne à la moitié, et vice versa.

On peut aisément vérifier par induction que cette stratégie engendre le capital

$$K_n = 2^n \frac{r_n!\,(n - r_n)!}{(n+1)!}.$$

En prolongeant les factorielles par la formule de Stirling et en appliquant la divergence de Kullback-Leibler, Ville a immédiatement abouti à la loi des grands nombres de la théorie des jeux : pour que le capital ne soit pas borné, il faut que r_n/n tende vers 1/2. Ainsi, Sceptique s'enrichit infiniment si, une fois sur deux, Réalité ne choisit pas face ($y_i = 1$).

On peut facilement convertir cette démonstration remarquablement simple de la loi forte des grands nombres de la théorie des jeux en une démonstration de la loi forte de la théorie de la mesure, selon laquelle la convergence se produira sauf pour un ensemble de mesure nulle. Il suffit d'utiliser le fait que la probabilité qu'une stratégie multiplie à l'infini le capital qu'elle risque est nulle.

---

pas le même que celui que l'on fait d'ordinaire dans les sciences sociales, où chaque joueur cherche à maximiser une fonction d'utilité qui rassemble des besoins ainsi que des préférences complexes et revêt par conséquent de nombreuses valeurs possibles.

Pour tester des probabilités simples

Faisons revenir Prévisionniste et laissons-le annoncer une probabilité à sa guise. On a à présent trois joueurs. Prévisionniste essaie de donner des probabilités qui passent avec succès les tests statistiques. Sceptique met en œuvre les tests en essayant de s'enrichir grâce à ses paris sur les probabilités de Prévisionniste. Réalité décide. Voici la description complète de ce jeu plus général :

$K_0 := 1$
Pour $n = 1,2, \dots$:
  Prévisionniste annonce $p_n \in [0,1]$;
  Sceptique annonce $s_n \in R$;
  Réalité annonce $y_n \in \{0,1\}$;
  $K_n := K_{n-1} + s_n(y_n - p_n)$.

Sceptique gagne si
  K_n n'est jamais négatif, et si
  soit $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i) = 0$ ou $\lim_{n\to\infty} K_n = \infty$.

Ma démonstration du cas où *p* est toujours de 1/2 peut se généraliser et constituer la preuve que la stratégie de Sceptique est gagnante dans ce jeu. C'est la preuve villéenne par la théorie des jeux de la martingale de la loi forte des grands nombres, dont j'ai parlé dans la deuxième partie.

Ville publie ce résultat dès 1939, mais il apporte aux tests statistiques une dimension que l'on ignore encore souvent aujourd'hui. Dans la deuxième partie, je rappelle que l'on peut tester statistiquement la distribution de probabilité d'un processus aléatoire en utilisant les résultats d'une seule trajectoire effective de ce processus, même si les étapes successives ne sont pas indépendantes et identiquement distribuées selon cette distribution. Les résultats de Ville impliquent qu'il n'est pas même nécessaire d'avoir une distribution de probabilité entière pour le processus. Tout ce dont on a besoin pour effectuer des tests statistiques, c'est une suite de prévisions de probabilité à tester. C'est ce qu'a très clairement expliqué Phil Dawid dès 1984.

Dans *L'Univers irrésolu*, Popper souligne que la capacité de la science à prévoir est disséminée et fragmentaire. Les tests de Ville dans le cadre de la théorie des jeux nous permettent de pousser l'idée de Popper plus loin qu'il ne l'avait lui-même fait, car ils montrent que l'on peut légitimement tester des prévisions de probabilité même lorsqu'elles sont plus fragmentaires que les prévisions tirées de mesures de probabilité. Une mesure de probabilité est un système fermé qui spécifie à l'avance les possibilités d'information ainsi que les prévisions de probabilité (c'est-à-dire les probabilités conditionnelles) auxquelles donneront lieu ces diverses possibilités. Or le jeu présenté ci-dessus ne dit rien du moment où Prévisionniste va élaborer sa prévision suivante ni de quelle information il va se servir.

4 Comment effectuer des prévisions de probabilité

Après avoir étudié les possibilités d'action de Sceptique, voyons à présent ce que Prévisionniste peut faire. Pour commencer, voyons d'abord ce que Prévisionniste peut faire dans le pire des cas, lorsque Sceptique et Réalité s'allient contre lui. Il est évident qu'il ne saurait faire de bonnes prévisions dans ce cas, mais ce qui est évident se révèle être faux, ce qui nous invite à considérer la méthode dite de « prévision défensive » de Vovk.

La stratégie d'Hilary Putnam face au Prévisionniste

Comme je l'ai dit, il est évident que Prévisionniste n'a aucune chance dès lors que Sceptique et Réalité s'allient contre lui. On voit immédiatement ce qu'ils vont faire. Réalité fera le contraire de ce que Prévisionniste annonce comme probable, tandis que Sceptique, sachant ce que fera Réalité, pariera contre ce que Prévisionniste annonce comme probable. Sceptique fera ainsi des gains réguliers, et même infinis si le jeu se poursuit indéfiniment.

Le philosophe Hilary Putnam a pensé qu'il valait la peine de faire une description détaillée des stratégies respectives de Sceptique et de Réalité. Ce qui donne la chose suivante : supposons que Réalité fasse advenir l'événement à chaque fois que la probabilité annoncée par Prévisionniste est inférieure ou égale à 1/2 et l'empêche de se produire lorsque sa probabilité est supérieure à 1/2[18]. Supposons

[18] Ce que fait Réalité lorsque la probabilité de Prévisionniste est exactement de 1/2 n'a pas d'importance pourvu que Sceptique sache ce qu'il fera.

encore que sceptique le sache et parie en conséquence, mais qu'il limite l'enjeu à 1$ à chaque coup, son but étant de battre Prévisionniste, et non de l'humilier.

Pour $n = 1,2, \ldots$: Prévisionniste annonce $p_n \in [0,1]$; Sceptique annonce $s_n \in R$; Réalité annonce $y_n \in \{0,1\}$; Gains de Sceptique = $s_n(y_n - p_n)$.

Réalité met Prévisionniste dans la pire situation possible :

$$y_n = \begin{cases} 1 & \text{si } p_n < 0.5 \\ 0 & \text{si } p_n \geq 0.5. \end{cases}$$

Les gains de Sceptique sont alors réguliers :

$$s_n = \begin{cases} 1 & \text{si } p_n < 0.5 \\ -1 & \text{si } p_n \geq 0.5. \end{cases}$$

Les gains de Sceptique $s_n (y_n - p_n)$ s'élèvent à $\pm (\pm 0{,}5) = 0{,}5$. Il gagne 50 cents à chaque tour.

Cette façon de vaincre Prévisionniste paraît a priori tout à fait convaincante, mais elle a quelque chose d'excessivement artificiel. La stratégie de Sceptique en matière de tests, qui ne dépend que du dernier coup de Prévisionniste, p_n, est discontinue en tant que fonction de p_n. Si bien que :

$$s_n(p) = \begin{cases} 1 & \text{si } p < 0{,}5 \\ -1 & \text{si } p \geq 0{,}5. \end{cases}$$

Pourquoi cela n'est-il pas plausible ? Parce que l'idée que l'on puisse dire de quel côté d'une valeur exacte se trouve un nombre réel constitue une fiction intellectuelle qui ne rend pas compte de certaines distinctions que l'on peut faire en réalité. Une fonction à valeur réelle qui représente une opération que l'on peut réellement effectuer devrait être continue, de sorte qu'une déviation à l'entrée trop légère pour qu'on la remarque ne donnera pas lieu à une différence assez importante à la sortie pour qu'on la remarque[19]. Il faut peut-être accepter que la Réalité soit discontinue, la Réalité sera toujours ce qu'elle est, il ne nous appartient pas d'en décider. Mais Sceptique représente un test qui peut être effectué par une personne, un ordinateur ou du moins quelque mécanisme qui laisse une trace visible. Il n'est pas logique d'attribuer un comportement discontinu à Sceptique.

C'est le point de vue qui est soutenu dans un ouvrage écrit en collaboration avec Vovk, *Probability and Finance, It's only a game* (2001). Nous montrons que les tests statistiques conventionnels, y compris ceux qui sont fondés sur la loi des grands nombres, peuvent tous être mis en œuvre par l'intermédiaire de stratégies de paris continues. Cela n'est guère surprenant ; qui voudrait se fier à un statisticien qui ne se décide qu'après avoir coupé en quatre des cheveux invisibles ? Il est donc raisonnable de supposer que le Sceptique adoptera des stratégies de paris continues et de nous contenter de demander à Prévisionniste de mettre ces stratégies en échec.

Si vous êtes trop habitués aux fonctions discontinues en mathématiques appliquées pour accepter cet argument, il en est un autre qui vous plaira peut-être davantage. Il consiste à dire la chose suivante : si l'on demande à Prévisionniste de mettre en échec ces stratégies discontinues imaginaires, il faut lui laisser un peu de temps, et lui permettre de rendre les choses un peu aléatoires. Au lieu de lui demander d'annoncer sa probabilité avec une précision infinie, demandons-lui seulement de proposer un intervalle infime ; ensuite, une fois que ses adversaires ont joué, *vous* choisirez au hasard une valeur exacte à l'intérieur de cet infime intervalle. Par exemple, s'il projette d'annoncer 1/2, il faut qu'il annonce $1/2 \pm 1/10^{100}$ et ensuite, une fois que les autres ont joué, vous choisissez au hasard (si cela vous semble possible) un nombre infiniment précis à l'intérieur de cet intervalle. Il faut donner au Prévisionniste la possibilité d'échapper à la fiction d'une précision infinie. Soit, on le laisse introduire un peu d'aléatoire, soit on lui demande de mettre en échec des stratégies continues. Ces deux idées sont assez proches. L'introduction d'un peu d'aléatoire conduit à une moyenne du résultat final, ce qu'il faut, c'est que le processus soit continu.

[19] L.E.J. Brouwer le formule ainsi : la fonction constructive d'une variable réelle doit être continue.

La prévision défensive

Je vais à présent montrer que Prévisionniste peut mettre en échec n'importe quelle stratégie continue de la part de Sceptique, indépendamment de ce que fait Réalité.

Pour faire bonne mesure, introduisons un peu plus de réalisme en permettant à Prévisionniste d'avoir accès à davantage d'information. Soit x_n cette information supplémentaire, à laquelle les autres joueurs ont également accès. On suppose qu'elle est rendue publique avant le prochain tour. (Je rappelle qu'il s'agit d'un jeu à information parfaite, tous les joueurs peuvent voir ce qui est annoncé.) On peut prendre Réalité pour l'annoncer. Qui mieux que lui ferait l'affaire ?

Pour $n = 1,2, \ldots$:
Réalité annonce $x_n \in X$;
Prévisionniste annonce $p_n \in [0,1]$;
Sceptique annonce $s_n \in R$;
Réalité annonce $y_n \in \{0,1\}$;
Gains de Sceptique = $s_n(y_n - p_n)$.

Je vais démontrer que si Sceptique suit une stratégie connue toujours continue par rapport au dernier coup de Prévisionniste, ce dernier peut faire en sorte que Sceptique ne gagne jamais d'argent.

On peut formaliser l'idée que Prévisionniste connaît la stratégie de Sceptique en demandant à ce dernier d'annoncer sa stratégie au début du jeu. On peut ensuite faire sortir Sceptique du jeu et calculer ses gains à chaque épreuve à partir de la stratégie connue. On peut encore démontrer le théorème même si l'on en demande moins à Sceptique. Au lieu de lui demander de révéler sa stratégie au départ, on peut lui demander de révéler la stratégie qui sera adoptée pour le coup suivant immédiatement après l'annonce x_n de la part de Réalité. À cette étape du jeu, la p_n du Prévisionniste est la seule information non encore connue que la stratégie de Sceptique peut utiliser pour choisir s_n. De sorte que pour dire comment il s'y prendra pour choisir s_n, Sceptique peut se contenter de donner une fonction S_n sur [0, 1] que l'on appliquera à p_n pour déterminer s_n. C'est cette fonction qui doit être continue.

Le protocole est le suivant :

Pour $n = 1,2, \ldots$:
Réalité annonce $x_n \in X$;
Sceptique annonce S_n continue : $[0,1] \rightarrow R$;
Prévisionniste annonce $p_n \in [0,1]$;
Réalité annonce $y_n \in \{0,1\}$;
Gains de Sceptique = $S_n(p_n)(y_n - p_n)$.

Le théorème que je démontrerai à présent est le suivant : la stratégie de Prévisionniste empêche Sceptique de gagner de l'argent dans ce protocole :

- Si $S_n(p) > 0$ pour tout p, choisir $p_n = 1$;
- Si $S_n(p) < 0$ pour tout p, choisir $p_n = 0$;
- Sinon, choisir p_n de sorte que $S_n(p_n) = 0$.

Rappelons-nous le théorème des valeurs intermédiaires : une fonction continue est toujours positive, négative ou nulle. Si S_n est toujours positive, alors $p_n = 1$ produit des gains négatifs ou nuls pour Sceptique, quoi que fasse Réalité. Si S_n est toujours négative, alors $(p_n) = 0$ produit des gains négatifs ou nuls pour Sceptique, quoi que fasse Réalité. Si S_n a une valeur nulle, toute valeur de p_n telle que $S_n(p_n) = 0$ produits un gain nul pour Sceptique, quoi que fasse Réalité.

Tests universels et *a priori* universels

L'une des idées fondamentales de la probabilité en théorie des jeux, c'est que l'on peut mettre en œuvre un test statistique par l'intermédiaire d'une stratégie de pari visant à multiplier le capital engagé. Le but n'est pas, bien entendu, de ne passer avec succès qu'un seul test, mais d'en franchir plusieurs. Or, en général, il est possible de fusionner les tests que l'on veut passer en un test unique.

Si j'ai une stratégie qui, à partir d'un seul dollar engagé, dégage une somme infinie à chaque fois que les probabilités de Prévisionniste et les résultats de Réalité ne coïncident pas d'une certaine façon, et si j'ai une autre stratégie qui aboutit au même résultat lorsqu'ils ne coïncident pas selon une autre manière, que faire ? Il me faut jouer 50 cents sur l'une des stratégies et 50 sur l'autre, car multiplier 50 cents à l'infini vaut autant que de multiplier un dollar à l'infini. J'obtiens l'infinité dans les deux cas. Je m'enrichis à l'infini si l'une ou l'autre des deux infractions est commise.

Répartir son argent entre des stratégies revient à faire une moyenne de celles-ci. On peut utiliser une moyenne pondérée ou faire une moyenne à partir d'un nombre dénombrable de stratégies. Dès lors que l'on investit un peu d'argent dans chaque stratégie, on obtient l'infini lorsque l'une d'elles multiplie son capital initial à l'infini. Comme l'a expliqué Abraham Wald dans les années 1930, alors qu'il cherchait à éclairer le concept de « collectif » de Richard von Mises, aucun langage ne nous permet de concevoir au-delà d'un certain nombre, dénombrable, de tests. En pratique, on ne peut en concevoir qu'un nombre fini. Faisons donc une moyenne afin d'obtenir un test fourre-tout (porte-manteau), ou *test universel* si l'on veut[20]. Le problème de Prévisionniste revient donc à venir à bout d'un unique test et nous venons de voir comment cela se faisait.

Comme je l'ai expliqué plus haut, du point de vue formel, la stratégie de Prévisionniste revient à une distribution de probabilité P pour la suite entière y_1, y_2, des coups joués par Réalité ; pour obtenir la prévision de probabilité du Prévisionniste à propos de y_n au énième tour de jeu, il faut faire dépendre P de $y_1,\ldots,$ y_{n-1}[21]. Ainsi, appliquée à un test plus ou moins universel, la stratégie de prévision défensive donne une distribution a priori plus ou moins universelle relativement aux coups joués par Réalité. Comme cela l'indique, mon explication est une version finitaire de la notion de distribution a priori universelle développée par Leonid Levin dans les années 1970[22].

[20] Même en théorie, on ne peut pas réellement élaborer un test universel pour un langage donné, car l'ensemble dénombrable des tests concevables dans ce langage ne peut être dénombré de manière récursive.

[21] Pour que cette formulation soit strictement correcte lorsque la probabilité de $y_1,\ldots, y_{n-1}$ est nulle, en sorte que la probabilité conditionnelle ne soit pas définie, il faut revenir à la pratique du XIX^e siècle, avant que la théorie de la mesure ne devienne le fondement officiel des probabilités, et supposer que P soit donné en tant que système de probabilités conditionnelles, et non en tant que mesure de probabilité unique. Voir le dernier chapitre de Ville (1939).

[22] Levin fut le premier à démontrer l'existence de distributions qui résistent aux tests universels. Cependant, leur existence n'est que théorique dans la mesure où leurs valeurs ne sont pas calculables. La meilleure approche du

Calibrage

Mon analyse a pris un tour trop général, trop abstrait, trop infinitaire. Revenons à quelque chose de plus précis, concret et fini. Supposons qu'il n'y ait pas d'information supplémentaire x_n, de sorte que la tâche de Prévisionniste consiste simplement à prévoir y_n à partir de $y_1,..., y_{n-1}$ pour $n = 1, 2,...$. Supposons encore que nous nous contentons de faire subir un test simple aux prévisions de probabilité de Prévisionniste : nous voulons qu'elles soient calibrées.

Rappelons-nous qu'un prévisionniste est calibré à 30% si, lorsque l'on repère toutes les fois où il a déclaré $p = 0,3$, on constate que l'événement a eu lieu 30% du nombre total de fois. Bien entendu, pour être plus proche de la pratique, il ne faut pas être trop précis lorsque l'on observe le nombre de fois où p est proche de 30%, il faut constater que l'événement a eu lieu environ 30% du temps. Il faut aussi que cela puisse se passer 35% du temps. Trouvons un calibrage pour 21 valeurs différentes de p, disons, $p = 0,05$, $p = 0,10$, $p = 0,15$, et ainsi de suite, ainsi que des valeurs plus proches de 0 et de 1.

Dans la deuxième partie, nous avons vu la loi des grands nombres de Ville pour $p = 0,50$, qui nous indique la manière dont Sceptique peut tester Prévisionniste pour un calibrage à 50%. La stratégie $S_n(y_1, ..., y_{n-1})$ proposée, utilisée uniquement lors des tours n où Prévisionniste fixe p_n proche de 0,50, augmente le capital de Sceptique s'il y a de nombreux tours semblables à celui-ci et si Réalité ne fixe pas $y_n = 1$ pour environ 50% d'entre eux. Comme je l'ai également mentionné dans la deuxième partie, Sceptique adopte la même stratégie pour toutes les autres valeurs de p. Fusionnons les stratégies de Sceptique pour la vingtaine de valeurs de p retenues en en faisant une moyenne. Cela nous donne une stratégie qui enrichit Sceptique s'il est non-calibré pour toutes les valeurs de p (ou s'il est non-calibré pour l'une d'elles, car nous estimons que p est proche de l'une de ces valeurs).

Vovk a poussé plus loin cette idée en donnant à p bien plus de valeurs et en observant ce qu'il advient de la stratégie moyenne lorsque le nombre tend vers l'infini. Il dégagea une expression très simple correspondant à la stratégie de

---

sujet, du point de vue mathématique, se trouve chez Gács (2005), et on en trouvera l'historique chez Bienvenu *et al.* (2009).

Sceptique à chaque tour, *n*, du jeu :

$$S_n(p) = \sum_{i=1}^{n-1} e^{-C(p-p_i)^2} \, (y_i - p_i).$$

La constante *C* est un taux d'apprentissage, que l'on peut choisir plus ou moins à son gré.

Cette stratégie utilise toute l'information disponible depuis le début du jeu. Rappelons que yi correspond aux résultats de Réalité, 0 et 1, tandis que p_i correspond aux prévisions de Prévisionniste. Si l'on fait la moyenne des écarts entre prévision et résultat, $(y_i - p_i)$, il faut que la moyenne soit proche de zéro ; c'est notre martingale de la loi des grands nombres. Mais ici, nous avons une moyenne pondérée, utilisant un noyau gaussien :

$$e^{-C(p-p_i)^2}.$$

Ce noyau attribue un poids quasiment nul à tout tour *i* où p_i n'est pas proche de *p*, de sorte que l'on ne fait des moyennes que sur les tours où p_i est proche de *p*. N'importe quel autre noyau comportant cette caractéristique ferait aussi bien l'affaire.

Rappelons que Prévisionniste recourt à la fonction $S_n(p)$, qui définit la stratégie de Sceptique, afin de déterminer sa propre stratégie. Il se prémunit contre S_n en choisissant *p* de sorte que $S_n(p) = 0$. En d'autres termes, il recourt à une valeur de *p* qui lui a réussi par le passé. C'est ce que nous faisons, pour la plupart d'entre nous, dans la vie. À quel jeu jouons-nous ? À celui où l'on est bon !

Afin d'effectuer des tests statistiques dans l'environnement « en ligne », il faut empêcher l'émergence de toute tendance, et pour ce faire, il faut réitérer ce qui a bien fonctionné pour nous auparavant. En faisant cela, le pire qui puisse arriver, c'est que vous réussissiez un peu moins bien qu'avant, mais sans aller jusqu'à échouer au test. Mais si vous vous aventurez toujours sur des territoires où vous vous en êtes mal sorti auparavant, il y a des chances pour que Réalité transforme cette piètre performance en tendance franchement mauvaise qui vous conduise à échouer au test.

Comme je l'ai déjà indiqué, en plus du calibrage, il est d'autres vertus auxquelles le Prévisionniste peut aspirer. Il cherchera peut-être aussi à ce que la convergence se fasse comme il se doit, c'est-à-dire comme le prescrit la loi des logarithmes itérés. Il existe également une stratégie pour la loi des logarithmes itérés : on peut faire une moyenne si l'on veut, bien que cela puisse nécessiter un grand nombre d'épreuves pour que l'on voie la différence.

Mais surtout, dans les applications, il faut convoquer de nouveau x_n pour s'assurer que Prévisionniste a trouvé une bonne *résolution* ainsi qu'un bon calibrage. Cela implique qu'il ne faut pas seulement observer les fois où Prévisionniste avait déclaré que la probabilité qu'il pleuve était de 30%, mais aussi, et plus précisément, les fois où il avait annoncé 30% alors qu'il avait plu la veille. Il faut alors qu'il pleuve 30% du temps. De même pour d'autres informations annexes importantes $x_1,\ldots, x_n$. On peut obtenir la résolution, dans les limites du raisonnable, de la même manière que l'on a obtenu le calibrage : il faut un noyau qui combine la comparaison de p_i avec p et celle de x_i avec x_n.

On sait que l'on peut estimer une probabilité à partir d'un échantillon aléatoire. Ce que je vous ai dit est moins connu. En ligne, il n'est pas nécessaire que l'échantillon soit aléatoire, pour autant que des échantillons puissent être non-aléatoires. On peut effectuer des tests statistiques indépendamment du comportement de Réalité[23].

En ligne, tout est aléatoire

Ayant reçu une formation de statisticien en mathématiques dans les années 1970, le concept de processus aléatoire avec probabilités inconnues qu'il faut estimer au fur et à mesure m'est familier. Mais le succès de la prévision défensive m'a enseigné une chose que je n'avais pas comprise avant à propos de ce concept. Je croyais qu'il fallait savoir quelque chose de la réalité, ou du moins avoir des intuitions justes à son sujet, pour que cette *estimation au fur et à mesure* marche. Je croyais qu'il était indispensable de disposer d'un modèle adéquat de la réalité. La prévision défensive m'a appris que cela n'était pas aussi vrai que je le croyais[24]. Il

[23] C'est pourquoi j'appelle ce joueur « Réalité », et non Nature, car la Nature suit des lois tandis que la Réalité agit à sa guise.

[24] J'admets que ce concept conserve une certaine vérité : il faut avoir le bon noyau, c'est-à-dire qu'il faut identifier correctement les caractéristiques de notre information qui sont importantes pour faire des prévisions.

n'est pas nécessaire d'avoir un modèle adéquat de la réalité, on peut y arriver quel que soit le comportement de la réalité.

Je croyais qu'un processus pouvait parfois ne pas être stochastique ; que parfois il n'était pas possible de donner des probabilités qui passent avec succès les tests statistiques. Maintenant que je sais que je m'étais trompé, j'ai tendance à dire que tout est aléatoire, mais aussi que cette affirmation n'a aucun contenu empirique. Elle n'est pas réfutable dans le contexte des situations en ligne.

5 Prolongements

Cette intervention concerne les situations en ligne. Ma thèse est qu'il est possible d'y faire de bonnes prévisions de probabilité indépendamment de ce qui se passe dans la réalité. Mais nous ne sommes pas toujours en ligne. Aussi, je voudrais terminer en disant un mot à propos d'autres champs de l'expérience où l'on peut utiliser les probabilités mathématiques. J'en vois deux principaux : la causalité et le jugement.

Causalité probabiliste

Parfois, on sait quelque chose du comportement de la réalité. Nous disposons parfois de règles qui nous permettent de faire des prévisions de probabilité qui franchiront des tests statistiques, mais sans dépendre de la suite d'événements que nous avons décidé d'observer. Pour l'empiriste, notre connaissance de ces règles de prévision relève de la connaissance par causalité.

Cette connaissance par causalité est un vaste sujet que je ne peux aborder ici, mais je l'ai fait ailleurs (cf. Shafer, 1996). Dans cette étude, j'ai eu recours à la même notion de probabilité propre à la théorie des jeux que dans la présente communication : les prévisions de probabilité sont des propositions de pari que l'on teste par l'intermédiaire de stratégies destinées à multiplier le capital engagé. Ces propositions ne sont peut-être pas à proprement parler des prévisions de probabilité complètes, de sorte que l'on n'obtient que des probabilités supérieures et inférieures, comme dans l'ouvrage que j'ai écrit en collaboration avec Vovk (2001) et dans le cadre que Walley (1991) caractérise, à tort à mon avis, comme étant « imprécis ».

Du point de vue philosophique, le défi est de réconcilier les réalistes avec la thèse selon laquelle lorsque l'on parle de causalité, on parle avant tout de régularités dans des prévisions qui s'avèrent exactes. Dans ce genre de discussions philosophiques, personne n'a jamais le dernier mot, mais une fois que l'on a accepté le point de vue de la théorie des jeux sur les probabilités mathématiques, il apparaît clairement que la probabilité ne concerne que la prévision de phénomènes, et non leur « engendrement ».

Jugement de probabilité

L'environnement en ligne ne débute pas avec des probabilités, mais avec une structure presque aussi puissante : une suite d'expériences. J'ai indiqué comment trouver la probabilité d'un événement donné, mais seulement lorsque celui-ci était pris dans une suite. Je suppose que tout le monde était d'accord là dessus.

Dans la vie réelle, les gens sont souvent en désaccord sur la suite dans laquelle il s'agit de placer un événement donné. Un différend survient entre mon voisin et moi, nous allons au tribunal. Il replace ce qui s'est passé dans sa propre suite pour relater ce qu'il lui est arrivé tandis que je replace ce qui s'est passé la mienne pour relater ce qui m'est arrivé. Devant le juge, chacun cherche à défendre le bien-fondé de sa propre suite. Le rôle du juge est de discerner ce qui est pertinent. Les paris justifiés par l'une des suites se justifient-ils encore lorsque l'on voit l'information relative à l'autre suite ? C'est une affaire de jugement qui peut parfois prendre la forme d'un jugement pour savoir si les propositions de pari qui ne peuvent être mises en échec dans l'une des suites ne peuvent l'être non plus lorsque l'on recourt à l'information supplémentaire contenue dans l'autre série.

Ce type de raisonnement nous mène au conditionnement bayésien et à la theorie de Dempster-Shafer (cf. Shafer, à paraître) selon laquelle, encore une fois, on peut décider que seuls certaines propositions de pari demeurent fiables, et donc s'en tenir à des probabilités inférieures et supérieures.

Références

Cesa-Bianchi, Nicolò et Gábor Lugosi (2006), *Prediction, Learning, and Games*, Cambridge : Cambridge University Press.

Bienvenu, Laurent, Glenn Shafer et Alexander Shen (2009), « On the history of martingales in the study of randomness », *Electronic Journal for History of Probability and Statistics*, 5(1) (www.jehps.net).

Dawid, A. Philip (1984), « Statistical theory: the prequential approach », *Journal of the Royal Statistical Society A*, 147, pp. 278–92.

Dawid, A. Philip, Steven de Rooij, Glenn Shafer, Alexander Shen, Nikolai Vereshchagin et Vladimir Vovk (2011), « Insuring against loss of evidence in game-theoretic probability », *Statistics and Probability Letters*, 81, pp. 157–62.

Drouet, Isabelle (2011), « Propensions poppériennes et puissances aristotéliciennes. De l'interprétation des probabilités à la métaphysique », *Philosophie*, à paraître.

Gács, Péter (2005), « Uniform test of algorithmic randomness over a general space », *Theoretical Computer Science*, 341, pp. 91–137.

Hacohen, Malachi Haim (2000), *Karl Popper: The Formative Years, 1902–45*, Cambridge : Cambridge University Press.

Hennig, Christian (2007), « Falsification of propensity models by statistical tests and the goodness-of-fit paradox », *Philosophia Mathematica*, 15(2), pp. 166–92.

Kolmogorov, Andrei N. (1933), *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Vienne, Autriche : Springer.

Kolmogorov, Andrei N. (1951), « Probability (in Russian) », *Great Soviet Encyclopedia*, 7, 2e édition, pp. 508–10.

Neyman, Jerzy (1960), « Indeterminism in science and new demands on statisticians », *Journal of the American Statistical Association*, 55, pp. 625–39.

Popper, Karl R. (1935), *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft*, Vienne, Autriche : Springer.

Popper, Karl R. (1938), « A set of independent axioms for probability », *Mind*, 47, pp. 275–77.

Popper, Karl R. (1982), *The Open Universe: An Argument for Indeterminism,* Tome 2 de *Postscript to the Logic of Scientific Discovery*, Totowa, NJ : Rowman and Littlefield.

Popper, Karl R. (1983), *Realism and the Aim of Science*, Tome 1 de *Postscript to the Logic of Scientific Discovery*, Totowa, NJ : Rowman and Littlefield.

Psillos, Stathis (1999), *Scientific Realism: How Science Tracks Truth*, Londres: Routledge.

Shafer, Glenn (1996), *The Art of Causal Conjecture*, Cambridge, MA : The MIT Press.

Shafer, Glenn (2007), « Du principe de Cournot au marché efficient », in Jean-Philippe Touffut (éd.), *La société du probable. Les mathématiques sociales après Augustin Cournot*, Paris : Albin Michel, pp. 83–132 .

Shafer, Glenn (2011), « A betting interpretation for probabilities and Dempster–Shafer degrees of belief », *International Journal for Approximate Reasoning*, 52 (2), pp. 127-36.

Shafer, Glenn et Vladimir Vovk (2001), *Probability and Finance, It's Only a Game*, New York, NY : Wiley.

Shafer, Glenn, Alexander Shen, Nikolai Vereshchagin et Vladimir Vovk (2011), « Test martingales, Bayes factors, and p-values », *Statistical Science*, 26 (1), pp. 84-101.

van Fraassen, Bas C. (1980), *The Scientific Image*, Oxford : Oxford University Press.

Ville, Jean (1939), *Etude critique de la notion de collectif*, Paris : Gauthiers-Villars.

Walley, Peter (1991), *Statistical Reasoning with Imprecise Probabilities*, Londres : Chapman and Hall.