Cours de maths à l'usage du philosophe X

Les probabilités

Le hasard est par définition ce qu'on ne peut ni prévoir, ni calculer. Est-ce un reflet de notre ignorance ou une propriété intrinsèque de l'univers ? Les mathématiques, à vrai dire, ne se posent pas cette question. Leur objectif, au contraire, est de prévoir et de calculer. Mais quand un phénomène comprend une part de hasard, cela signifie-t-il que l'on ne peut rien en dire de précis ? Le développement de la théorie des probabilité montre le contraire : le hasard est une hypothèse que l'on peut modéliser et travailler pour en tirer des prévisions et les confronter au réel.

Ce chapitre comportera davantage d'exemples concrets que les autres : maintes situations peuvent être facilement modélisées en termes de probabilités. Cela ne signifie pas pour autant que les probabilités aient plus d'applications que d'autres domaines, mais simplement qu'elles sont exposables à un niveau plus élémentaire.

L'approche fréquentiste historique

Aussi étrange que cela paraisse, le développement rapide de la **théorie des probabilités** au XXème siècle doit beaucoup à la théorie de la mesure. L'étude des lois du hasard, qui commence avec **Blaise Pascal** et les frères **Bernoulli** au XVIIème siècle, trouve assez rapidement ses limites en raison de l'approche choisie, qui est dite **fréquentiste**. Il faut attendre l'approche **ensembliste** de Kolmogorov (*Fondements de la théorie des probabilités* 1933) pour que les probabilités s'intègrent complètement aux mathématiques et en deviennent un domaine important.

L'approche fréquentiste

Une **expérience aléatoire** est une expérience concrète dont le résultat ne peut être prédit. L'archétype en est le lancer d'un dé équilibré. Comment peut-on définir la probabilité du résultat 6 ? On va lancer notre dé un certain nombre N de fois et compter le nombre $n$de 6 obtenus.

Si le nombre N de lancers est petit, de l'ordre d'une dizaine, alors on ne peut pas dire grand chose sur le nombre n , il peut prendre toutes les valeurs entre 0 et 10, avec une préférence pour les petits nombres comme 1 ou 2. Si le nombre Nde lancers est moyen, de l'ordre d'une centaine, alors on commence à pouvoir préciser que le nombre n a tendance à se trouver entre 12 et 22 . Si le nombre N de lancers est grand, de l'ordre du millier, alors on est quasiment certain que le nombre de 6 obtenus se trouvera entre 155 et 175. Si on réalise un nombre immense Nde lancers, on devine que le nombre $n$de 6 sera très proche de $n = \frac{N}{6}$. Autrement dit, on obtient 6 en moyenne une fois sur 6. On définit alors la **probabilité** de 6 comme la limite de $p = \frac{n}{N}$ quand N tend vers l'infini : c'est la compréhension intuitive fréquentiste des probabilités.

Le problème des partis

Ce problème proposé à Blaise Pascal par le chevalier de Méré permet d'exposer la notion d'**espérance mathématique**, centrale dans la théorie des variables aléatoires.

Deux joueurs A et B décident de lancer répétitivement une pièce de monnaie équilibrée.

A parie sur l'apparition de trois fois Pile (pas forcément d'affilée).

B parie sur l'apparition de trois fois Face (pas forcément d'affilée).

Ils misent la même somme : 32 Euros chacun, dans un pot commun. Celui qui gagne emporte l'ensemble du pot.

Au premier lancer, Pile sort. Aussitôt après, le jeu doit s'arrêter pour une raison inconnue.

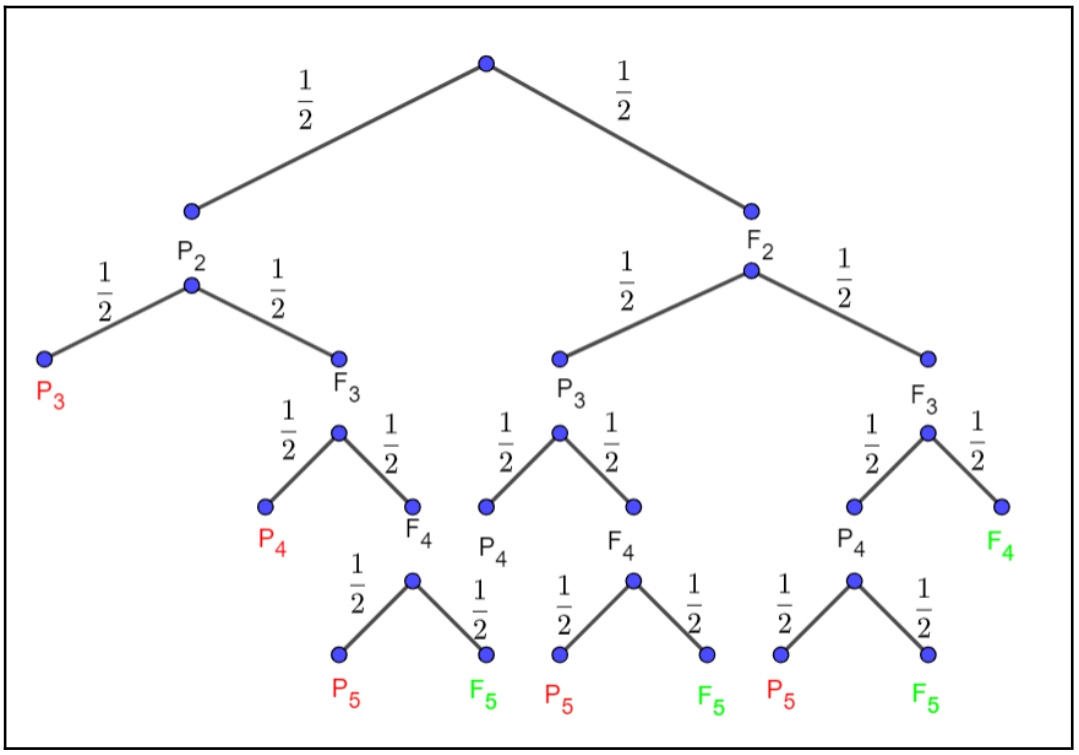
B veut que la partie soit annulée et les mises rendues à chaque joueur.

A n'est pas d'accord, il veut que l'on répartisse équitablement le pot commun en sachant que Pile était déjà sorti une fois.

Combien faudrait-il donner à chacun selon A ?

Des protagonistes célèbres réfléchissent alors à ce problème dans leur correspondance : Fermat, Huygens, Jacques Bernoulli s'en mêlent. La solution est qu'il faut partager le pot commun selon l'espérance du gain de chaque joueur.

Sur l'arbre de probabilités ci-dessous, on a entouré en rouge les cas où *A* gagne. Chaque branche porte une probabilité de $\frac{1}{2}$puisque la pièce est équilibrée.



Pour calculer la probabilité que A gagne, il faut multiplier les probabilités des branches qui y mènent. On obtient donc :

$$p(A\ gagne) = (\frac{1}{2})^2 + (\frac{1}{2})^3 + (\frac{1}{2})^4 + (\frac{1}{2})^3 + (\frac{1}{2})^4 + (\frac{1}{2})^4$$

$$p(A\ gagne) = \frac{11}{16}$$

A peut donc espérer gagner 64 euros avec une probabilité de $\frac{11}{16}$. Elle est supérieure à la probabilité $1 - \frac{11}{16} = \frac{5}{16}$ que B gagne car Pile est déjà sorti une fois…

Il faut donc donner $\frac{11}{16} \times 64 = 44$ *euros à A, et seulement 20 euros à B.*

On le voit, même s'il est impossible de prévoir quelle aurait été l'issue de la partie, le fait de pouvoir quantifier les probabilités des différentes issues donne une possibilité d'action, une solution à un problème.

L'approche ensembliste contemporaine

Dans l'approche fréquentiste, la probabilité est définie comme une donnée de l'expérience, et non une quantité qui peut être déterminée en dehors de tout recours à l'expérience. Elle ne se prête donc pas du tout à l'approche hypothético-déductive et structurale des mathématiques.

L'idée du soviétique Andreï Kolmogorov est donc d'inverser le point de vue. On appellera **espace probabilisé** un ensemble muni d'une mesure appelée **probabilité**. Le seul axiome

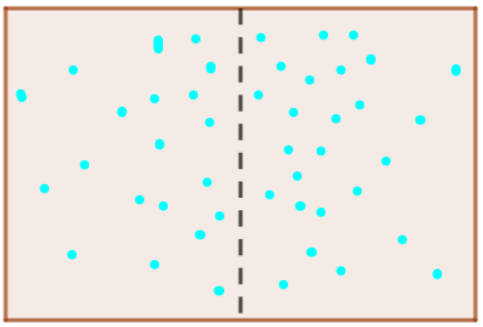
supplémentaire par rapport à une mesure au sens général est que la probabilité de l'ensemble total , appelé **univers**, vaut 1. Le miracle structural a alors lieu, et tous les théorèmes sur les mesures s'appliquent aux probabilités, qui deviennent un chapitre particulier de la théorie de la mesure. Si Blaise Pascal disait que les probabilités étaient la **géométrie du hasard**, les probabilistes aujourd'hui sont en fait plutôt des *analystes*, à ceci prêt que le formalisme des probabilités est fait pour stimuler la vision fréquentiste, qui reste la source d'intuition.

La gaussienne

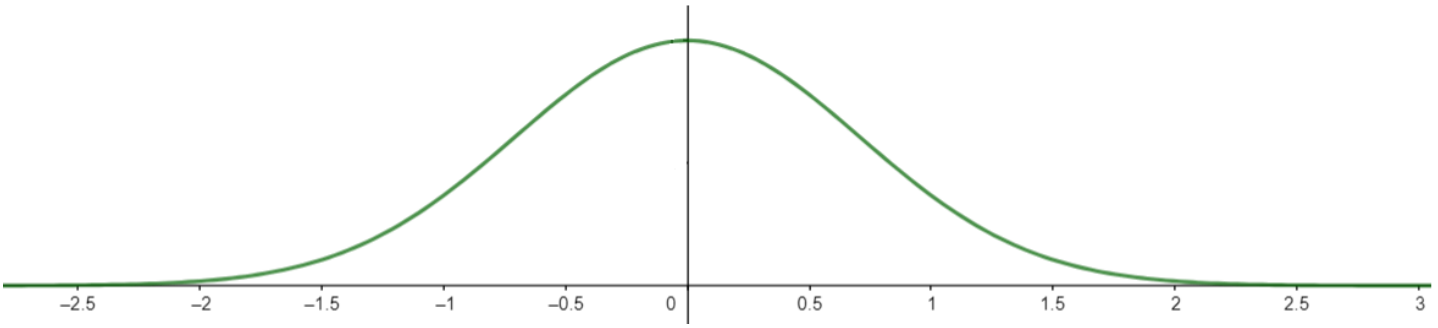
La théorie de la mesure permet de passer des **probabilités discrètes** aux **probabilités continues**. En probabilités discrètes, les résultats d'une expérience aléatoire sont dénombrables : on peut les numéroter. Or, si une expérience aléatoire peut avoir pour résultat un nombre réel quelconque, on a vu qu'il est impossible de numéroter ces résultats. La probabilité d'un événement, qui est l'union d'un certain nombre de résultats, n'est pas calculable.

Prenons un exemple : Supposons que l'on tire une fléchette sur une cible rectangulaire en visant le segment pointillé médian,

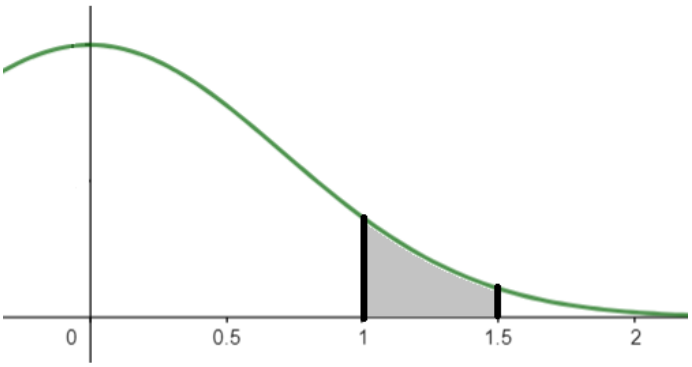
On peut admettre que la fléchette se plantera environ aussi souvent à gauche qu'à droite du segment, avec une tendance à être proche de ce dernier, on obtiendra une distribution semblable à :



Si, après un nombre très élevé de lancers, on place sur l'axe horizontal la distance au segment pointillé, et sur l'axe vertical les effectifs obtenus, on obtiendra une courbe appelée **Gaussienne**, ou courbe en cloche :



La probabilité que la fléchette soit éloignée du segment médian est faible, alors que la probabilité que la fléchette en soit proche est élevée. C'est ce qu'on appelle une **distribution normale** des résultats. Pour calculer la probabilité que la fléchette se trouve à une distance comprise entre 1 et 1,5 à droite du segment médian, on doit additionner les probabilités de l'infinité de valeurs comprises entre 1 et 1,5.
Au lieu d'une somme discrète, on utilisera une somme intégrale qui correspond à l'aire située sous la courbe verte, appelée **densité de probabilité**, entre les valeurs 1 et 1,5 :



$$p = \int_{1}^{1,5} f(x)dx$$

Le **théorème de la limite centrale**, affirme que, sous certaines conditions, à chaque fois que les résultats d'une expérience aléatoire sont influencés par une multitude de facteurs, ils se répartiront selon ce type de courbe en cloche. L'explication de ce phénomène tient au fait que la conjonction de k facteurs se répartit selon le **coefficient binomial** évoqué au chapitre précédent. De multiples études statistiques en sciences économiques et sociales ou en biologie aboutissent à cette répartition.

Applications à la prise de décision

Notre dernière remarque est à la base des **statistiques inférentielles**, utiles pour la prise de décision en situation d'incertitude . Pour tester une hypothèse, par exemple l'efficacité d'un médicament, on sait que les résultats de l'étude statistique doivent se répartir sur une courbe en cloche.

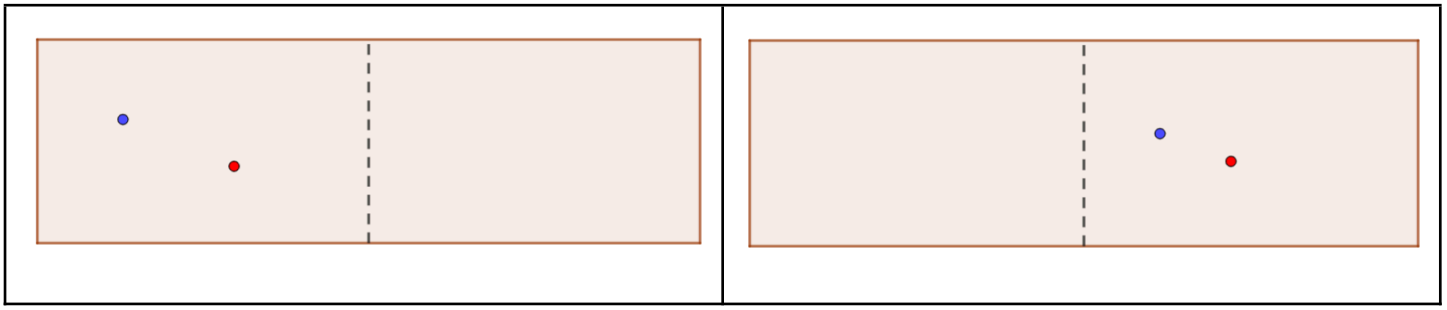
Ainsi, même si le médicament est efficace, il y aura des cas où il ne fonctionne pas. Mais si on l'a testé sur un nombre suffisant de cas, on peut déterminer un seuil de succès au-delà duquel on acceptera l'hypothèse que le médicament est efficace. La détermination de ce seuil utilise les propriétés de distributions continues de probabilités du même type que celle de Gauss évoquée au paragraphe précédent.

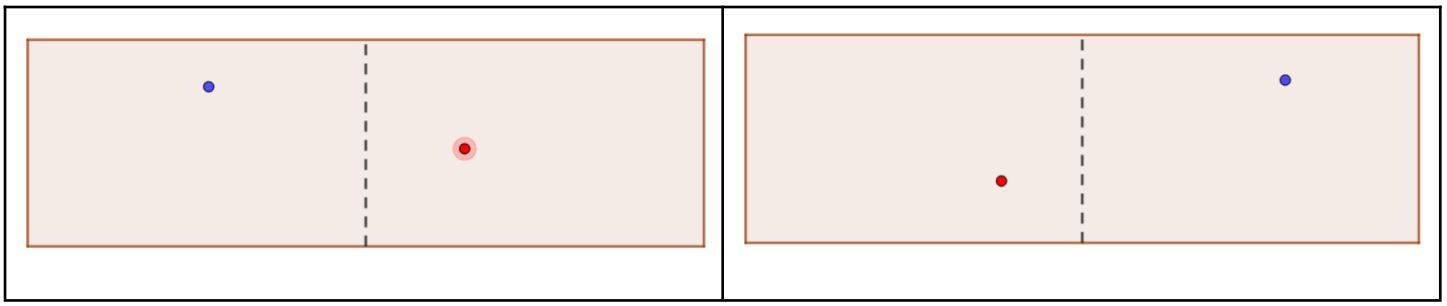
Ainsi, l'hypothèse d'efficacité peut être confirmée ou réfutée, en dépit des cas particuliers, inévitables en raison du caractère fondamentalement aléatoire des phénomènes biologiques, qui dépendent d'un grand nombre de facteurs qui ne peuvent pas être contrôlés.

Le second principe de la thermodynamique et l'entropie

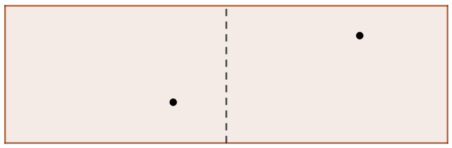
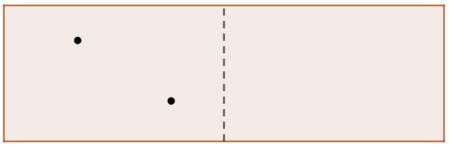
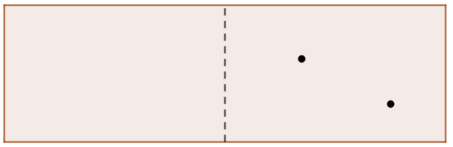
Le **deuxième principe de la thermodynamique** établit l'**irréversibilité** des phénomènes physiques, en particulier lors des échanges thermiques, et l'existence d'une fonction d'état d'un système, appelée **entropie**, qui ne peut qu'augmenter en l'absence d'un apport énergétique ordonné, appelé **travail** . C'est un principe d'évolution qui fut énoncé pour la première fois par Sadi Carnot en 1824. Nous n'exposerons pas ici les calculs nécessaires à la formulation mathématique du second principe, qui sont compliqués.

Imaginons une situation très simple, où deux atomes se déplacent librement dans une boîte divisée en deux par une paroi poreuse. Nous avons quatre situations possibles (des **micro-états**) si l'on est capable de distinguer les deux atomes :





Mais si l'on n'est pas capable de distinguer les atomes entre eux, il n'y a que trois situations observables (appelées **macro-états**) :

macro-état 1	macro-état 2	macro-état 3
		

L'**entropie** d'un macro-état est le logarithme du nombre de microétats qui lui correspondent :

$$E(1) = ln(2) \simeq 0,69 \qquad E(2) = ln(1) = 0 \quad E(3) = ln(1) = 0$$

La différence va s'accentuer si le nombre d'atomes, au lieu d'être égal à 2, est très grand, ce qui est bien sûr le cas pour un système physique macroscopique. Les états où il y a approximativement autant d'atomes d'un côté et de l'autre seront beaucoup plus nombreux que ceux où les atomes sont plus ou moins rassemblés d'un côté.

Il s'ensuit que les états d'entropie élevée sont plus probables que les autres, à un tel point que les autres ne se produisent statistiquement jamais, même s'ils ne sont pas théoriquement impossibles. Ainsi, entre deux états d'un système qui, d'un point de vue macroscopique, paraissent également possibles, la considération des micro-états montre que l'un se produira "toujours", et l'autre "jamais".

En thermodynamique, on est capable de calculer l'entropie des différents états observables d'un système en fonction d'autres variables d'état, comme la pression, la température ou le volume. De cette manière, on peut donc prévoir l'évolution du système physique.

Test de dépistage

La problématique des tests de dépistage aura connu une popularité inédite lors de la pandémie de Covid19. Il nous apparaît utile de la développer un peu dans la mesure où certaines informations diffusées dans les médias laissaient peu de place à la pensée critique.

Supposons que nous disposions d'un test destiné à détecter l'usage d'une drogue.

La **sensibilité** du test est la probabilité $p_D(+)$d'obtenir un test positif chez une personne droguée. On la supposera très élevée, égale à 99%.

La **spécificité** du test est la probabilité $p_{ND}(-)$d'obtenir un test négatif chez une personne non droguée. On la supposera également très élevée, égale à 98%.

Ces données statistiques sont obtenues en "testant" le test sur des personnes dont on sait à l'avance si elles sont droguées ou non.

Bien sûr, le test sera ensuite utilisé sur des personnes dont on ignore l'état. On supposera ici que la **prévalence** de l'usage de drogue est de 1% dans la population, c'est-à-dire que 1% des gens que l'on teste sont drogués.

Dans ce cas, on s'interrogera sur la probabilité $p_+(D)$qu'une personne soit droguée sachant que son test est positif.

Le pourcentage de drogués positifs est de 99% de 1%, c'est-à-dire 0,99%. Le pourcentage de non drogués positifs est 2% de 99%, c'est-à-dire 1,98%. Le pourcentage total de positifs est la somme O,99%+1,98%=2,97%. La probabilité d'être drogué sachant qu'on est positif est le quotient $p_+(D) = \frac{0{,}99\%}{2{,}97\%} = \frac{1}{3}$.

La conclusion est que le test, en dépit de ses excellentes performances sur le papier, génère un nombre énorme de faux-positifs : deux tiers des personnes testées positives ne sont pas droguées ! En revanche, si une personne est négative, le même genre de calcul montre qu'il est quasiment certain qu'elle n'est pas droguée : il y a très peu de faux-négatifs. La leçon à retenir de cela est que la prévalence joue un rôle important dans la problématique des tests de dépistage, qu'on ne peut pas résumer aux indicateurs de spécificité et de sensibilité.

Place croissante des probabilités

En raison de leur flexibilité et de l'implémentation du hasard comme hypothèse dans les calculs, les probabilités se prêtent extrêmement bien à la modélisation de phénomènes où l'incertitude liée aux mesures et aux données est naturellement présente. C'est évidemment le cas de presque tout ce que les sciences humaines étudient, mais c'est également souvent vrai en sciences physiques et naturelles.

Mais la place des probabilités a longtemps été limitée par la quantité de données que l'on pouvait traiter. Aujourd'hui, l'abondance des données enregistrées informatiquement permet aux probabilités d'être l'un des outils mathématiques les plus importants de la science.