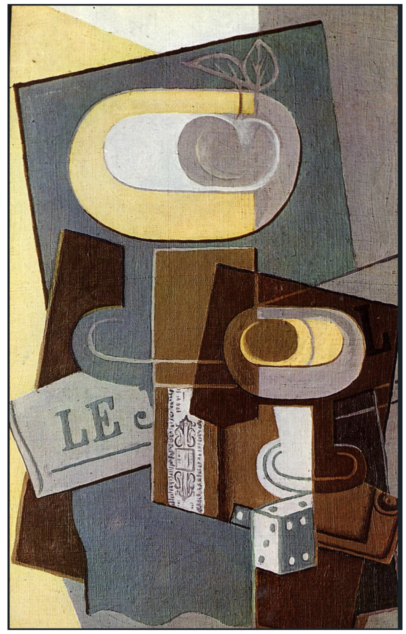
PB 1 | Espaces probabilisés

Après des considérations heuristiques, on aborde le cas des univers finis.



Le dé, Juan Gris

L'Histoire des probabilités a commencé avec celle des jeux de hasard. Bien que quelques calculs de probabilité soient apparus dans des applications précises au Moyen-Âge, ce n'est qu'au XVII-ième siècle que la théorie des probabilités se développe. Elle évolue sans aucun formalisme pendant deux siècles au moyen de problèmes d'urnes ou d'autres problèmes issus de jeux de hasard. La théorie moderne des probabilités, fondée sur la théorie de la mesure et la théorie de l'intégration, n'apparaît que vers 1930. Cette théorie s'est depuis diversifiée et ses apllications sont nombreuses (biologie, physique, finance, etc).

L'aube des probabilités

Durant l'Antiquité, le hasard n'a pas été un objet d'étude à proprement parler mais l'importance des jeux (de dés ou de cartes) a poussé les philosophes de la Grèce antique à aborder la notion d'incertitude.

Aristote

Le concept de probable chez *Aristote* (IV-ième siècle avant JC) est défini dans *les Topiques* :

Sont probables les opinions qui sont reçues par tous les hommes, ou par la plupart d'entre eux, ou par les sages, et parmi ces derniers, soit par tous, soit par la plupart, soit enfin par les plus notables et les plus illustres.

Cette idée d'Aristote a beaucoup été commentée et développée par la suite lors de traductions (Cicéron, Oresme, etc).

Certaines ébauches de calculs apparaissent dans des applications précises. Dans la Rome antique, le hasard est évalué pour décider de la valeur des rentes viagères, c'est-à-dire faire un pari sur la durée de vie. Ces calculs sont basés sur des tables de valeurs, certaines du jurisconsulte Ulpien (III-ième siècle) ont été retrouvées.

Bien que le concept de probabilité au sens statistique ne soit pas encore élaboré, des premiers *raisonnements probabilistes* apparaissent.

La notion de risque et les contrats d'assurance

L'apparition de la notion de *risque* est apparue au XII-ième siècle dans le cadre de l'évaluation de contrats commerciaux (cf. le *Traité des contrats* de Pierre de Jean Olivi).

Cette idée s'est développée aux XIII-ième et XIV-ième siècles, des bourses ont été créées pour assurer les transports maritimes (qui dépendent des fluctuations météorologiques) et au XVI-ième siècle avec la généralisation des contrats d'assurance maritime.

Une des premières références connues sur les probabilités est un calcul élémentaire dans *La Divine Comédie* de *Dante* au XV-ième siècle.

Dans un traité posthume de *Jérôme Cardan* publié en 1663 et intitulé *Liber de Ludo ALeae*, on trouve quelques éléments de probabilités. L'auteur étudie les différentes combinaisons possibles obtenues à partir de un, deux ou trois dés, truqués ou non et il expose le principe d'équiprobabilité.

Ces idées se propagent en Italie et en France. Galilée écrit entre 1620 en 1718 un bref mémoire sur le jeu de dés intitulé *Sopra le Scoperte de i Dadi* dans lequel il suppose l'équiprobabilité des lancers.

Aussi il y a une règle générale, que nous devons considérer le circuit entier[a], et le nombre de ces lancers qui représente en combien de façons les résultats favorables peuvent se produire, et comparer ce nombre au reste du circuit, et les paris mutuels devront être posés selon cette proportion, de sorte qu'on puisse disputer en termes égaux.

Cette approche est fondée sur l'intuition d'un comportement à long terme, elle est qualifiée de fréquentiste. La notion d'indépendance était connue de Cardan puisque qu'il utilise la multiplication des probabilités.

a. Note de LK : i.e. toutes les possibilités

Jérôme Cardan

Les débuts de la théorie : Pascal, Fermat et Huygens

L'acte de naissance de la théorie des probabilités est certainement la correspondance entre *Pierre de Fermat* et *Blaise Pascal* en 1654 au sujet d'une désormais célèbre question posée par Antoine Gombaud (dit chevalier de Méré) : le problème des partis ou problème des points.

Blaise Pascal

Il avait pour objet de déterminer la proportion suivant laquelle l'enjeu doit être partagé entre les joueurs lorsqu'ils conviennent de ne point achever la partie, et qu'il leur reste à prendre pour la gagner, des nombres de points inégaux. Pascal en donna le premier la solution, mais pour le cas de deux joueurs seulement ; il fut ensuite résolu pour Fermat, dans le cas général d'un nombre quelconque de joueurs.

Poisson

Suite à un séjour à Paris en 1655, *Christian Huygens* prend connaissance de cette discussion à l'Académie Parisienne et publie en 1657 le premier traité sur la théorie probabiliste, *De ratiociniis in ludo aleae*. C'est dans une lettre adressée à Frans Van Shooten, qui a traduit son traité en latin dans Mathematische Oefeningen, qu'il attribue la paternité de la théorie des probabilités à Pascal et Fermat :

Il faut savoir d'ailleurs qu'il y a un certain temps que quelques uns des plus Célèbres Mathématiciens de toute la France se sont occupés de ce genre de calcul, afin que personne ne m'attribue l'honneur de la première Invention qui ne m'appartient pas.

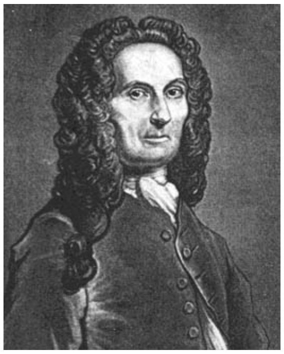
Le traité de Huygens est resté le seul ouvrage important de la théorie des probabilités jusqu'au début du XVIII-ième siècle, il fut traduit en anglais en 1692.

Christiaan Huygens

Les développements de la théorie au XVIII-ième siècle

Plusieurs ouvrages sur les probabilités suivent le traité de Huygens. *Pierre Rémond de Montmort* publie en 1708 l'un des premiers, l'*Essai sur les Jeux de Hasard*, qui contient la formule du binôme, des problèmes de jeux de cartes et le problème des partis.

Abraham de Moivre publie en 1718 *The Doctrine of Chances* contenant des problèmes combinatoires dont la formule de Stirling, des probabilités conditionnelles ainsi qu'une approximation de loi normale par une loi binomiale, c'est la première version du théorème central limite dit *théorème de De Moivre-Laplace*. Les énoncés mathématiques sont plus précis mais n'utilisent pas le formalisme actuel :



Abraham de De Moivre

Corollaire 8.
La rapport que, dans une puissance infinie d'une binomiale, dénotée par n, le plus grand terme porte à la somme de tout le reste, sera justement exprimé par la fraction $\frac{a+b}{\sqrt{abnc}}$*, où c désigne, comme auparavant, la circonférence d'un cercle dont le rayon égale l'unité.*

Par ce type d'énoncés, de Moivre en arrive à la courbe appelée normale, mais il ne la considère que comme approximation et non comme densité d'une loi normale.

Ce traité reste le traité majeur en théorie des probabilités jusqu'à la parution du grand traité de Laplace en 1812, *Théorie analytique des probabilités.*

L'œuvre posthume de Jacques Bernoulli, *Ars Conjectandi* est publiée en 1713. Elle reprend les calculs de Huygens de divers problèmes combinatoires notamment sur le binôme. Son traité contient également une description de l'estimation d'un phénomène aléatoire sous forme de fréquences : *Ce qu'il n'est pas donné d'obtenir a priori l'est du moins a posteriori, c'est-à-dire qu'il sera possible de l'extraire en observant l'issue de nombreux exemples semblables.*

C'est la relation entre la probabilité d'un évènement et l'observation de celui-ci, Bernoulli appelle ce résultat son *théorème d'or*. Le nom de loi des grands nombres pour ce type de résultats a été donné plus tard par Poisson. Ce résultat est cependant une version généralisée du théorème actuel appelé loi des grands nombres.

Jacques Bernoulli

Daniel Bernoulli

Nicolas Bernoulli

Nicolas Bernoulli publie en 1711 sa thèse de doctorat où apparait pour la première fois la loi uniforme discrète. *Daniel Bernoulli* étudie, autour de 1732, des applications du calcul des probabilités aux problèmes d'assurance, à l'astronomie, au calcul d'erreur ou au paradoxe de Saint-Pétersbourg. Dans les mêmes années, *Leonhard Euler* contribue également à des questions d'assurance. *Jean le Rond D'Alembert* étudie de même le paradoxe de Saint-Pétersbourg et énonce la règle actuelle du calcul de l'espérance d'une variable aléatoire discrète.

Thomas Bayes énonce dans un article posthume de 1764 intitulé *An essay towards solving a problem in the doctrine of chances*, une première version d'une formule appelée aujourd'hui formule de Bayes et utilisant des probabilités conditionnelles. Ce théorème est également appelé problème inverse ou probabilité inverse et a été étudié par de nombreux auteurs.

Cependant, le premier à donner une formulation précise de ce problème de probabilité inverse est *Pierre-Simon de Laplace* en 1774 dans son mémoire *Sur la probabilité des causes*.

Thomas Bayes

Le naturaliste *Georges-Louis Leclerc de Buffon* publie en 1777 son *Essai d'arithmétique morale* dans lequel apparaissent des liens entre les probabilités et la géométrie (cf. le problème de l'aiguille de Buffon).

Les travaux de Laplace

Joseph-Louis Lagrange publie vers 1770 un mémoire contenant des problèmes de durée d'un jeu de hasard et utilisant des lois continues. Débute alors la considération du caractère continu des probabilités.

Pierre-Simon de Laplace

Au début du XIX-ième siècle, plus précisément en 1812, *Laplace* publie son traité intitulé *Théorie analytique des probabilités* dans lequel il présente des résultats asymptotiques étant ainsi un des premiers ouvrages à distinguer des énoncés de principes probabilistes aux estimations des probabilités observées suite à une expérience. Cependant la distinction aujourd'hui usuelle entre le signe somme $\sum$ et le signe intégrale $\int$ n'était pas observée à l'époque de Laplace. Il énonce ces résultats à propos de problèmes d'urnes en utilisant des rationnels : *le rapport du nombre de billets blancs au nombre total des billets contenus dans l'urne peut être un quelconque des nombres fractionnaires compris depuis 0 jusqu'à 1.*

En augmentant le nombre de billets à l'infini, Laplace énonce alors une première version rigoureuse du théorème central limite : *on voit donc qu'en négligeant les quantités infiniment petites, nous pouvons regarder comme certain que le rapport du nmobre de billets blancs au nombre total de billets est compris est compris entre les limites* $p/(p+q)+w$ *et* $p/(p+q)-w$, w *étant égale à* $(p+q)^{-1/n}$, n *étant plus grand que 2 et moindre que 3; partant*w *peut être supposé moindre qu'aucune grandeur posée.*

Ici le théorème central limite est vu par l'intermédiaire d'intervalles de confiance, c'est-à-dire en donnant l'imprécision de la convergence d'une somme vers la moyenne obtenue pour la loi des grands nombres. Les travaux de Laplace resteront des travaux de référence jusqu'à la fin du XIX-ième siècle et au début du XX-ième siècle.

C'est à partir de Laplace que la théorie des probabilités est enseignée en France. À la fin du XVIIIe siècle, autour de 1792, le politicien et mathématicien *Nicolas de Condorcet* ambitionne le développement d'un enseignement des probabilités pour *donner la vérité aux hommes* et leur donner accès au bonheur. En 1816, *Sylvestre-François Lacroix* édite le premier ouvrage d'enseignement des probabilités intitulé *traité élémentaire de calcul des probabilités*. Dans ces mêmes années, *Condorcet* publie *les éléments du calcul des probabilités et son application aux jeux de hasard, à la loterie, et aux jugements des hommes* dans lequel il aborde la notion d'équiprobabilité dans des exemples combinatoires basés sur les jeux de hasard. Lorsque l'école polytechnique est créée en 1794, l'enseignement débute avec le cours du calcul des probabilités de *Joseph Fourier*. Citons également *François Arago* qui enseigna des calculs de probabilité dans son cours d'arithmétique sociale.

Le temps des critiques

La théorie des probabilités a toujours pris une place à part dans le domaine des mathématiques. La Faculté des sciences de Paris, créée en 1808, ne contenait pas de chaire de calcul des probabilités. La première chaire n'est créée qu'en 1834.

La théorie des probabilités a d'abord été considérée comme une partie des mathématiques appliquées. Le philosophe Auguste Comte la désigne comme *prétendue théorie des probabilités*, le courant philosophique du positivisme (1830-1845) en fait une critique radicale : *prétendre évaluer une probabilité est une escroquerie scientifique et une malhonnêteté morale*.

Associer une science à l'incertitude du hasard, a soulevé plusieurs questionnements chez les mathématiciens. *Joseph Bertrand* en 1900 écrit : *comment oser parler des lois du hasard ? Le hasard n'est-il pas l'antithèse de toute loi ?*.

Vers la théorie moderne des probabilités

La théorie des probabilités va prendre un nouvel essor vers le début du XX-ième siècle grÃ¢ce à l'introduction de nouveaux concepts comme les mesures et l'intégration.

Axel Harnack introduit en 1881 la notion d'ensemble discret, puis en 1882 le concept d'*égalité en général* qui est précurseur de l'*égalité presque sûre*. Les mesures apparaissent sous différentes définitions suivant les travaux de *Georg Cantor* (1884), *Giuseppe Peano* (1887) ou *Camille Jordan* (1892). Mais c'est à *Émile Borel* que l'on attribue la paternité de la théorie de la mesure avec l'introduction d'ensembles de mesure nulle en 1897, et de la classe des sous-ensembles ouverts de $\mathbb{R}$ qu'il mesure à partir de la longueur des intervalles. Cette classe portera plus tard le nom de *borélienne*. La théorie des probabilités est simplement une branche de la théorie de la mesure avec ses propres applications, comme le souligne plus tard *Joseph Leo Doob* en 1953.

En 1901, *Henri-Léon Lebesgue* utilise cette théorie de la mesure pour développer la théorie de l'intégration qui généralise l'intégrale de Riemann. Sous l'impulsion de *Johann Radon* en 1913, cette théorie se généralise à $\mathbb{R}^n$ puis sur un ensemble plus abstrait Ω muni d'une tribu. La fin de cette généralisation est estimée à 1930 avec la décomposition de Radon-Nikodym-Lebesgue et l'existence des densités.

C'est à partir de 1930 que *Andreï Kolmogorov* fonde mathématiquement la théorie des probabilités. Il publie en 1933 son travail fondamental, *Grundbegriffe des Warscheinlichkeitrechnung* en posant les trois axiomes des probabilités qui définissent de manière rigoureuse et consistante les probabilités[1].

Émile Borel

Henri Lebesgue

Andreï Kolmogorov

1. Le mathématicien *Jean Dieudonné* souligne plus tard en 1977 : *le calcul des probabilités, en tant que discipline, n'existe guère que depuis 1933, comme partie de la théorie moderne de l'intégration ; elle a hérité de ses propres problèmes et même de son langage, des trois siècles antérieurs au cours desquels le calcul des probabilités était un mélange de raisonnements d'allure mathématique et de considérations plus ou moins intuitives sur le rôle et l'évaluation du hasard dans les comportements humains ou les phénomènes naturels*

L'axiomatique de Kolmogorov n'a cependant pas fait directement le consensus auprès de la communauté mathématique et n'a été largement reconnue qu'en 1959. Par exemple, Gustave Choquet souligne en 1962 que le groupe Bourbaki ferment à ses disciples les portes du calcul des probabilités.

L'exploration des processus stochastiques

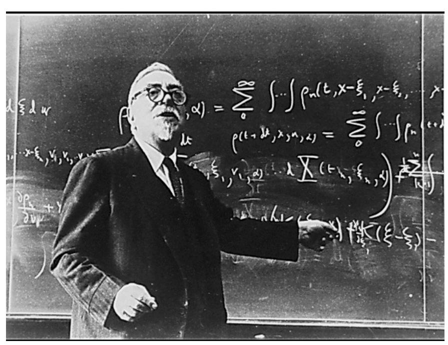
La théorie moderne des probabilités a permis une grande avancée en modélisation avec la création des processus stochastiques, qui ont occupé et occupent encore une grande part de l'activité de recherche en mathématique.

L'origine des processus stochastiques date du début du XX-ème siècle, ils ont été introduits pour modéliser des phénomènes temporels où intervient le hasard, en mécanique statistique par exemple. les premiers physiciens à les utiliser sont *Willard Gibbs*, *Ludwig Boltzmann*, *Henri Poincaré*, *Marian Smoluchowski* et *Paul Langevin*. Les bases théoriques sont apparus plus tard dans les années 1930-1940 grâce aux travaux des mathématiciens *Joseph Leo Doob* et *Andreï Kolmogorov*. Apparaît alors le terme *stochastique* issu du grec *stokastikos* signifiant *conjectural*.

Les processus de Markov

En 1902, *Andreï Markov* posa les bases de la théorie des processus à temps fini possédant la propriété de Markov : le futur du processus ne dépend que du présent et non du passé. Ces processus seront appelés processus de Markov. *Louis Bachelier* puis *Andreï Kolmogorov* généraliseront cette propriété en 1936 pour les processus à temps continu. Les marches aléatoires sont un exemple classique de processus de Markov. L'origine de ce type de processus remonte aux travaux de *John William Strutt Rayleigh* et et plus particulièrement son article de 1880 intitulé *Composition of n isoperiodic vibrations of unit amplitude and of phases distributed at random*. Le premier ouvrage sur les marches aléatoires isotropes est celui de *Karl Pearson* en 1905 sur les random flights (vols aléatoires), vient alors le traité de *George Polya* de 1921 dans lequel il traite les questions essentielles : récurrence, transience, point multiple, etc.

Andreï Markov



Norbert Wiener

Joseph Leo Doob

Le mouvement brownien

L'histoire du mouvement brownien commence en 1828 par *Robert Brown* avec une observation d'un mouvement de particules en suspension dans un liquide. Suivent plusieurs expérimentations de physiciens au XIX-ème siècle, pour arriver à une formalisation plus mathématique de ce phénomène en 1905 et 1906 par le physicien *Albert Einstein* dans trois articles fondateurs dont un des buts était une évaluation du nombre d'Avogadro. À la même époque, *Marian Smoluchowski* publie une théorie du mouvement brownien tout à fait analogue. Le physicien *Jean Perrin* utilise les travaux d'Einstein pour estimer le nombre d'Avogadro ce qui lui valut le prix Nobel de physique en 1926. D'autre part le chimiste *Theodor Svedberg* reçut le prix Nobel de chimie en 1926 pour son exploitation de la formule de

Smoluchowski. Le premier mathématicien à avoir construit le mouvement brownien (ou plutôt le processus de Wiener) est *Norbert Wiener* en 1923 avec l'article *Differential space* dans lequel il construit une mesure sur l'espace des fonctions continues telle que ses accroissements sur des intervalles de temps disjoints suivent une loi normale prévue par la théorie d'Einstein. D'autre part, Paul Langevin relit les approches de Smoluchowski et d'Einstein en 1908 ; ses travaux seront formalisés mathématiquement en 1942 par Joseph Leo Doob. C'est le début des équations différentielles stochastiques (EDS). Une autre approche des EDS vient de *Kiyoshi Itô* dans cette même année 1942, avec la formule d'Itô qui reprend les travaux de Kolmogorov sur les processus de Markov à temps continu. Les travaux sur le mouvement brownien continuent encore aujourd'hui et il serait impossible de citer tous les auteurs comme en atteste la biographie de cinq cents auteurs dans le livre de Daniel Revuz et Marc Yor publié en 1994.

Les martingales

Les premières idées de martingale, c'est-à-dire d'un processus vérifiant une propriété d'espérance conditionnelle, viennent des jeux de hasard avec les travaux de Pascal sur le problème des partis. Le terme martingale apparait pour la première fois dans la thèse de *Jean Ville* publiée en 1939 dans laquelle apparaît l'expression *système de jeu ou martingale*. Cette dénomination a été reprise par un de ses rapporteurs de thèse *Joseph Leo Doob* qui l'utilisa dans son ouvrage *Stochastic processes* publié en 1953. C'est ainsi au début du XX-ième siècle que les martingales sont formalisées mathématiquement avec les travaux de *Bernstein*, *Ville*, *Borel* et *Doob*.

La percolation

Le premier exemple de percolation a été publié par *Broadbent* en 1954, puis par *Broadbent* et *Hammersley* en 1957 dans *Percolation processes* comme un modèle de diffusion d'un fluide (liquide ou gaz) dans un milieu aléatoire. Le milieu est composé aléatoirement de canaux ouverts ou fermés, ce qui a donné l'idée à Broadbent et Hammersley de donner le nom de percolation. En 1960, Ted Harris donne une estimation de la valeur critique, dans le réseau plan $\mathbb{Z}^2$, d'obtenir une percolation (composante connexe de taille infinie). En 1980, *Harry Kesten* calcule cette valeur critique. Citons également *Stanislav Smirnov* qui obtint en 2010 la prestigieuse Médaille Fields pour son travail dans ce domaine.

1. Modéliser l'aléatoire

Nous commencerons ce cours par une approche informelle et intuitive de la théorie des probabilités. Le but de cette introduction est de dégager *un modèle mathématique minimal*.

Les « *définitions* » données dans cette partie seront reprises (et parfois modifiées, voire très largement rectifiées) à partir de la section **2.**, véritable début du cours *classique*.

1.1. Le hasard

L'objet de la théorie des probabilités est l'analyse mathématique de phénomènes dans lesquels le hasard[2] intervient. Ces phénomènes sont appelés des phénomènes aléatoires[3].

2. Ce mot vient de l'arable *al-zahr* qui signifie *dés*.
3. Ce mot vient du latin *alea* qui signifie *dés*.

Phénomène aléatoire

Un phénomène est dit aléatoire si, reproduit maintes fois dans des conditions identiques, il se déroule chaque fois différemment de telle sorte que le résultat change d'une fois sur l'autre de manière imprévisible.

Cette imprédictibilité (et donc le hasard) peut avoir deux origines :

- La connaissance imprécise de certains paramètres.
  - Le hasard est le reflet de la méconnaissance des conditions initiales.
  - *Cas d'un lancer d'un dé ou d'une pièce* : la pièce ou le dé ont des trajectoires parfaitement définies par la mécanique classique mais la sensibilité aux conditions initiales est telle qu'on ne peut en fait rien prédire.
- Trop de paramètres influent sur l'issue de l'expérience aléatoire étudiée.
  - Le hasard est le reflet de la multitude des facteurs.
  - *Durée de vie d'une ampoule, temps d'attente entre deux bus, mutations dans un génôme, prix d'une action, mouvement d'une particule fine en suspension dans un liquide* : dans ces exemples, de nombreux facteurs extérieurs influent sur le résultat de l'expérience et on ne sait pas les contrôler. La plupart des phénomènes déterministes sont accompagnés d'écarts aléatoires. Dans certains cas, il est possible de les négliger en remplaçant le phénomène réel par un schéma simplifié (comme par exemple l'étude de la chute libre par la mécanique classique). Mais cette approximation n'est pas toujours possible et il est souvent fondamental de pouvoir quantifier les écarts aléatoires (comme par exemple l'étude du mouvement de petites particules en suspension dans un liquide).

1.2. De l'expérience au modèle mathématique

Considérons l'expérience suivante : on lance consécutivement trois fois une même pièce de monnaie et l'on note le triplet obtenue sous la forme PPP, PFF, etc., où P et F désignent *Pile* et *Face*.

On emploie le vocabulaire suivant :

- *Cette expérience est aléatoire* : reproduite dans des conditions identiques, elle peut conduire à plusieurs résultats possibles, et si on ne peut prévoir le résultat par avance.
- *Univers associé à l'expérience aléatoire* : l'espace de tous les résultats possibles, appelé *espace d'états* ou encore *univers associé à l'expérience*, est noté Ω. On a donc

$$\Omega = \{\mathrm{PPP}, \mathrm{PPF}, \mathrm{PFP}, \mathrm{PFF}, \mathrm{FPP}, \mathrm{FPF}, \mathrm{FFP}, \mathrm{FFF}\}$$

  Un résultat possible de l'expérience est noté classiquement ω. Par exemple, $\omega := \mathrm{FFP} \in \Omega$.
- *Épreuves d'une expérience aléatoire* : on appelle *épreuve* chacune des réalisations de l'expérience.
- *Événements* : on appelle *événement aléatoire* tout sous-ensemble de Ω.
- *Ensemble des événements possibles* : l'ensemble des événements est donc $\mathscr{T} = \mathscr{P}(\Omega)$.

Ainsi, les événements aléatoires sont des ensembles. Nous allons utiliser le formalisme de la théorie des ensembles, en particulier les opérations élémentaires sur les ensembles, pour décrire diverses possibilités de réalisations d'événements. Pour cette expérience aléatoire :

- On peut décrire un événement en extension ou en compréhension. Par exemple :

$$\mathrm{A} = \underbrace{\text{« \textit{Pile est tiré au premier coup} »}}_{\text{en compréhension}} \quad \text{et} \quad \mathrm{A} = \underbrace{\{\mathrm{PPP}, \mathrm{PPF}, \mathrm{PFP}, \mathrm{PFF}\}}_{\text{en extension}}$$

  Le plus souvent en probabilité, on utilise la définition des événements en compréhension.

- Soit B lévénement « *Face est tiré la deuxième fois* », C l'événement « *Pile est tiré au premier coup puis Face au deuxième* » et D l'événement « *Pile n'est pas tiré au premier lancer* ». On a clairement

$$\mathrm{C} = \mathrm{A} \cap \mathrm{B}\,, \quad \mathrm{D} = \Omega \setminus \mathrm{A} = \overline{\mathrm{A}}$$

Nous cherchons à définir, pour un ensemble possible de réalisations de l'expérience $\mathrm{A} \in \mathscr{T}$, la vraisemblance accordée a priori à A (avant le résultat de l'expérience). Nous voulons donc associer à chaque événement A un nombre $\mathbf{P}(\mathrm{A})$ compris entre 0 et 1, qui représente la chance que cet événement soit réalisé à la suite de l'expérience. Voici ce l'observation nous apprend :

- En réalisant 5 épreuves de l'expérience, nous obtiendrions une suite de 5 résultats mais n'en tirerions presqu'aucune information.
- En revanche, si nous réalisions 10000 épreuves de l'expérience et si nous obtenions 5003 fois Pile au premier lancer, alors il serait tentant de dire que la probabilité de l'événement A vaut à peu près $1/2$.

Le fondement expérimental de la notion de probabilité est ici :

Principe de l'heuristique fréquentiste

Pour un grand nombre n d'épreuves de l'expérience, la fréquence

$$f_n(\mathrm{A}) := \frac{n_{\mathrm{A}}}{n} = \frac{\textit{nombre de fois où } \mathrm{A} \textit{ est réalisé}}{\textit{nombre total de répétitions de l'expérience}}$$

semble se stabiliser autour d'une valeur limite. L'approche intuitive et naturelle consiste à définir $\mathbf{P}(\mathrm{A})$ *comme étant la limite quand n tend vers l'infini des fréquences de réalisation $f_n(\mathrm{A})$. Intuitivement,*

$$\mathbf{P}(\mathrm{A}) = \text{limite de } f_n(\mathrm{A}) \text{ quand } n \to +\infty$$

Nous donnerons ultérieurement une justification et un sens précis à cette limite, grâce à la loi des grands nombres, qui est un des théorèmes fondamentaux de la théorie, justifiant toute la construction mathématique.

Les fréquences vérifient clairement les propriétés suivantes :

$$\begin{cases} f_n(\mathrm{A}) \in [0,1]\,; \\ f_n(\Omega) = 1\,; \\ \text{Si A et B sont disjoints, } f_n(\mathrm{A} \cup \mathrm{B}) = f_n(\mathrm{A}) + f_n(\mathrm{B}) \end{cases}$$

De ces propriétés intuitives, on conçoit la définition suivante :

Le modèle probabiliste

On postule l'existence d'une fonction $\mathbf{P} : \mathscr{P}(\Omega) \to [0,1]$ telle que

$$\begin{cases} \mathbf{P}(\Omega) = 1 \\ \forall (A,B) \in \mathscr{P}(\Omega)^2,\ \ A \cap B = \varnothing \implies \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) \end{cases}$$

1.3. Du choix d'un modèle

Dans des cas concrets d'expérience aléatoire, il faudra choisir *le modèle probabiliste adéquat*, ie l'espace probabilisé $(\Omega, \mathbf{P})$ qui modélise correctement l'expérience.

Prenons un exemple historique, l'erreur de D'Alembert.

> *On lance une seule pièce, deux fois de suite. Quelle est la probabilité de tirer Face au moins une fois ? Ce problème fut posé par d'Alembert en* 1754*. Il raisonna comme suit. Il n'y a que trois configurations possibles : on tire deux piles, deux faces ou un pile et un face. Il y a deux cas favorables sur les trois cas, la probabilité recherchée vaut donc* $2/3$*.*

Il est important de comprendre la *nature* de cette erreur : ce n'est pas *une erreur de calcul* mais *une erreur de modélisation*. On peut envisager deux modélisations de cette expérience aléatoire :

Modélisation I

- $\Omega := \{(P,P), (P,F), (F,F)\}$;
- $\mathscr{T} := \mathscr{P}(\Omega)$;
- $\forall \omega \in \Omega$, $\mathbf{P}(\{\omega\}) = \dfrac{1}{3}$.
- $A := \{(F,F), (P,F)\}$ donc $\mathbf{P}(A) = \dfrac{2}{3}$.

Modélisation II

- $\Omega := \{(P,P), (P,F), (F,P), (F,F)\}$;
- $\mathscr{T} := \mathscr{P}(\Omega)$;
- $\forall \omega \in \Omega$, $\mathbf{P}(\{\omega\}) = \dfrac{1}{4}$.
- $A := \{(F,F), (P,F), (F,P)\}$ donc $\mathbf{P}(A) = \dfrac{3}{4}$.

D'Alembert a choisi la modélisation I qui se place sur l'univers des configurations : Pile-Pile, Pile-Face, Face-Face avec équiprobabilité. Si on réalise cette expérience, on trouverait une fréquence s'approchant plutôt du résultat de $3/4$ donné par la modélisation II. C'est bien là que se trouve la difficulté du calcul des probabilités, trouver le bon modèle. L'erreur de D'Alembert peut sembler bien naïve de nos jours.

Signalons que l'on peut trouver *des modélisations différentes mais équivalentes*. Par exemple, on peut reprendre l'univers de la modélisation I et considérer la probabilité définie par

$$\mathbf{P}(\{(P,P)\}) = \mathbf{P}(\{(F,F)\}) = \frac{1}{4}\ ,\ \ \mathbf{P}(\{(F,P)\}) = \frac{1}{2}$$

Ce nouveau modèle est *équivalent* au modèle II.

Choisir le bon modèle probabiliste

Dans la pratique, les expériences seront décrites au moyen de mots-clés (comme *on tire simultanément*, *successivement*, *avec ordre*, *sans ordre* etc.) qui ne laisseront planer aucun doute sur le modèle probabiliste à choisir.

2. Probabilités dans le cadre des univers finis

Dans cette section, tous les univers Ω considérés seront supposés finis.

2.1. Espaces probabilisés finis

Bien qu'utilisant le formalisme de la théorie des ensembles, la théorie des probabilités continue à employer un vocabulaire antérieur beaucoup plus imagé. On retiendra les correspondances suivantes :

Vocabulaire probabiliste	**Vocabulaire ensembliste**
résultat possible	ω, élément de Ω
événement	A, sous-ensemble de Ω
événement élémentaire	$\{\omega\}$, où $\omega \in \Omega$
A est réalisé	$\omega \in \mathrm{A}$
A implique B	$\mathrm{A} \subset \mathrm{B}$
A ou B	$\mathrm{A} \cup \mathrm{B}$
A et B	$\mathrm{A} \cap \mathrm{B}$
contraire de A	$\complement_\Omega \mathrm{A}$ ou A^c ou $\overline{\mathrm{A}}$
événement impossible	$\varnothing$
événement certain	Ω
A et B sont incompatibles	$\mathrm{A} \cap \mathrm{B} = \varnothing$

Nous aurons besoin d'une notion proche de celle de partition, le caractère non vide en moins.

Définition 1.0. Système complet d'événements

Soit I un ensemble fini et $(\mathrm{A}_i)_{i\in \mathrm{I}}$ une famille de parties de Ω. On dit que $\{\mathrm{A}_i\,;\, i \in \mathrm{I}\}$ est un système complet d'événements de Ω si les A_i sont deux à deux disjoints et de réunion égale à Ω.

- On considère l'expérience suivante : lancer de deux pièces de monnaie.
  - On a clairement $\Omega := \{\mathrm{P}, \mathrm{F}\}^2$;
  - Le couple $\omega = (\mathrm{F}, \mathrm{P})$ est un résultat possible de l'expérience ;
  - L'événement A = « *on tire Pile au premier lancer* » s'écrit en extension $\mathrm{A} = \{(\mathrm{P},\mathrm{P})\,(\mathrm{P},\mathrm{F})\}$.
  - Les familles suivantes sont des systèmes complets d'événements :

$$\mathscr{S}_1 = (\varnothing, \{(\mathrm{F},\mathrm{P}), (\mathrm{F},\mathrm{F})\}, \{(\mathrm{P},\mathrm{P}), (\mathrm{P},\mathrm{F})\}) \quad \text{et} \quad \mathscr{S}_2 = (\{(\mathrm{F},\mathrm{F})\}, \{(\mathrm{F},\mathrm{P})\}, \{(\mathrm{P},\mathrm{P})\}, \{(\mathrm{P},\mathrm{F})\})$$

Définition 1.1. Espaces probabilisés finis

Soit Ω un univers fini.

- Une mesure de probabilité (ou probabilité tout court) sur Ω est une application $\mathbf{P} : \mathscr{P}(\Omega) \to [0,1]$ vérifiant les deux propriétés suivantes :
  - Normalisation : $\mathbf{P}(\Omega) = 1$;
  - Additivité finie : pour toutes parties disjointes A et B de Ω, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$.
- Un espace probabilisé fini est un couple $(\Omega, \mathbf{P})$ où Ω est un univers de cardinal fini et $\mathbf{P}$ une probabilité sur Ω.

Les axiomes d'additivité et de normalisation permettent d'étendre les propriétés d'additivité d'une probabilité à des réunions disjointes finies (*additivité finie généralisée*) et même à des réunions quelconques finies (*formule de Poincaré*).

Proposition 1.2. Propriétés d'une probabilité (↯ 1.1)

Soit $(\Omega, \mathbf{P})$ un espace probabilisé.

a. $\mathbf{P}(\varnothing) = 0$.

b. Pour tout événements A et B, $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

c. Pour tout événement A, $\mathbf{P}\left(\overline{A}\right) = 1 - \mathbf{P}(A)$.

d. Croissance d'une probabilité : si A et B sont deux événements avec $A \subset B$, alors $\mathbf{P}(A) \leqslant \mathbf{P}(B)$.

e. Additivité finie : si $A_1, \ldots, A_n$ sont deux à deux incompatibles, alors $\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}(A_i)$.

f. Si A et B sont deux événements, alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

g. Sous-additivité : pour des événements $A_1, \ldots, A_n$, $\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathbf{P}(A_i)$.

h. Formule des probabilités totales, version classique : si $(A_1, \ldots, A_n)$ est un système complet d'événements et si B est un événement, alors

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B \cap A_i)$$

En particulier, si $A \subset B$, alors $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$.

Comme tout événement A est réunion disjointe finie des singletons $\{\omega\}$ pour $\omega \in A$, on déduit de la propriété d'attitivité que l'on peut se contenter de définir une mesure de probabilité sur les singletons.

Proposition 1.3. Définition sur les singletons

Une probabilité est entièrement déterminée par les images des singletons : pour Ω ensemble fini et $(p_\omega)_{\omega\in\Omega} \in \mathbb{R}^\Omega$ vérifiant $\sum_{\omega\in\Omega} p_\omega = 1$ et $\forall \omega \in \Omega,\ \ p_\omega \geqslant 0$.

- Il existe une unique probabilité $\mathbf{P}$ sur Ω telle que $\mathbf{P}(\{\omega\}) = p_\omega$.
- On a $\forall A \subset \Omega,\ \ \mathbf{P}(A) = \sum_{\omega\in\Omega} \mathbf{P}(\{\omega\}) \mathbb{1}_A(\omega)$.

Cette proposition motive le vocabulaire suivant.

Définition 1.4. Loi ou distribution de probabilité

On appelle loi de probabilité ou encore distribution de probabilité sur Ω la donnée de $\{(\omega, p_\omega)\,;\, \omega \in \Omega\}$ telle que
$$\sum_{\omega\in\Omega} p_\omega = 1 \quad \text{et} \quad \forall \omega \in \Omega,\ p_\omega \geqslant 0$$

Ainsi, dans le cas d'un univers fini, se donner une mesure de probabilité $\mathbf{P}$ et se donner une distribution de probabilité sont équivalents.

Proposition 1.5. Formule de Poincaré

Pour des événements $A_1, \ldots, A_n$, on a $\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\substack{\Lambda\subset[\![1,n]\!] \\ \#\Lambda=k}} \mathbf{P}\left(\bigcap_{\ell\in\Lambda} A_\ell\right)$.

Nous démontrerons cette formule dans le chapitre suivant en utilisant des outils adéquats. Le lecteur abordera avec profit le test (⚡ **1.2**).

2.2. Probabilité uniforme sur un ensemble fini

L'équiprobabilité correspond à la situation où chaque résultat élémentaire de l'expérience, ie chaque singleton, a la même probabilité.

Définition 1.6. Équiprobabilité (également appelée probabilité uniforme)

Soit Ω un ensemble fini non vide. *La probabilité uniforme* sur Ω est définie par $\forall \omega \in \Omega,\ \mathbf{P}(\{\omega\}) = \dfrac{1}{\#\Omega}$.

Cette définition est justifiée par la proposition 2.1 à la page 14 On a clairement :
$$\forall A \subset \Omega,\ \mathbf{P}(A) := \frac{\#A}{\#\Omega} = \frac{\text{Nombre de cas favorables}}{\text{Nombre total de cas}} \quad \textit{(formule dite de Laplace)}$$

- On tire successivement et sans remise quatre cartes d'un jeu de cinquante-deux cartes. Quelle est la probabilité pour que, parmi ces quatre cartes, il y ait exactement deux rois ? On modélise par l'ensemble Ω des 4-arrangements de 52 cartes muni de l'équiprobabilité. On construit un tirage comportant exactement deux rois en choisissant d'abord les positions de ces deux rois sans ordre, puis deux rois avec ordre puis 2 cartes avec ordre parmi les 48 cartes qui ne sont pas des rois :
$$\frac{\binom{4}{2}\,2!\binom{4}{2}\,2!\binom{48}{2}}{4!\binom{52}{4}} = \frac{4\times 3\times 6\times 48\times 47}{52\times 51\times 50\times 49}$$
- Le problème du chevalier de Méré. Ce personnage marquant de la cour de Louis XIV qui « *avait très bon esprit, mais n'était pas très bon géomètre* » (cf. lettre de Pascal à Fermat) était un joueur impénitent, toujours à la recherche de règles cachées lui permettant d'avoir un avantage sur ses adversaires. Voici deux de ses règles dont nous nous allons étudier le bien fondé.

- « *Il est avantageux de parier sur l'apparition d'au moins un* 6 *en lançant un dé* 4 *fois de suite* ». Dans ce cas, on modélise par $\Omega = [\![1,6]\!]^4$ muni de l'équiprobabilité. On calcule plutôt la probabilité de l'événement contraire : n'obtenir aucun 6. Ce dernier correspond à un nombre de cas de figure valant 5^4 d'où

$$p_1 = 1 - \left(\frac{5}{6}\right)^4 > \frac{1}{2} \quad \text{car} \quad 5^4 = 625 < 648 = 3\times 6^3$$

  Cette première règle est donc vraie.

- « *Il est avantageux de parier sur l'apparition d'au moins un double-six en lançant deux dés* 24 *fois de suite.* » Dans ce cas, on modélise par $\Omega = \left([\![1,6]\!]^2\right)^2 4$ muni de l'équiprobabilité. On calcule plutôt la probabilité de l'événement contraire : n'obtenir aucun double 6. Ce dernier correspond à un nombre de cas de figure valant 35^{24} d'où

$$p_2 = 1 - \left(\frac{35}{36}\right)^{24} < \frac{1}{2} \quad \text{car} \quad 2\times 35^{24} > 36^{24}$$

  Le second pari n'est donc pas avantageux contrairement à la seconde règle.

3. Conditionnement

La notion de conditionnement est l'une des plus fructueuses de la théorie des probabilités.

3.1. Heuristique de l'idée de conditionnement

L'idée de base permettant la compréhension de cette notion est la suivante : une information supplémentaire concernant l'expérience modifie la vraisemblance que l'on accorde à l'événement étudié.

Cherchons, pour un lancer de deux dés, la probabilité de l'événement

A = « *la somme est supérieure ou égale à 10* »

$D_1\|D_2$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$D_1\|D_2$	1	2	3	4	5	6
1						7
2						8
3						9
4						10
5						11
6	7	8	9	10	11	12

Cette probabilité vaut :

- ⇛ 1/6 sans information supplémentaire ;
- ⇛ 5/11 si l'on sait que le résultat d'un des 2 dés est 6 ;
- ⇛ 0 si l'on sait a priori que le résultat d'un des dés est 2.

Pour obtenir ces résultats, nous avons dans chaque cas calculé le rapport du nombre de résultats favorables sur le nombre de cas possibles. Nous remarquons qu'il est indispensable de bien définir l'espace de probabilité lié à l'expérience munie de l'information a priori. Remarquons également que l'information a priori a changé la valeur de la probabilité de l'événement aléatoire.

L'approche intuitive pour formaliser cette notion consiste à revenir à la notion de fréquence empirique. La fréquence de réalisation de l'événement A sachant que l'evénement B est réalisé, sur n

épreuves de l'expérience, est égale au nombre de réalisations de A parmi celles pour lesquelles B est réalisé. Elle vaut donc

$$\frac{n_{\mathrm{A}\cap\mathrm{B}}}{n_{\mathrm{A}}} = \frac{f_n(\mathrm{A}\cap\mathrm{B})}{f_n(\mathrm{B})}$$

et en faisant tendre n vers l'infini, nous aboutissons à la définition suivante.

3.2. Probabilités conditionnelles

Définition 1.7. Probabilités conditionnelles

Pour deux événements A et B tels que $\mathbf{P}(\mathrm{B}) > 0$, on appelle probabilité conditionnelle de A sachant B le nombre $\mathbf{P}(\mathrm{A}\cap\mathrm{B})/\mathbf{P}(\mathrm{B})$. On la note $\mathbf{P}_{\mathrm{B}}(\mathrm{A})$ ou encore $\mathbf{P}(\mathrm{A}\,|\,\mathrm{B})$.

L'application $\mathbf{P}_{\mathrm{B}} : \mathrm{A} \longmapsto \mathbf{P}(\mathrm{A}\,|\,\mathrm{B})$ reflète la nouvelle information dont on dispose sur l'univers Ω : $(\Omega, \mathbf{P}_{\mathrm{B}})$ est un espace probabilisé.

Proposition 1.7. Probabilité conditionnelle (↯ 1.3)

Avec les notations précédentes,

a. L'application $\mathbf{P}_{\mathrm{B}}$ définit une probabilité sur Ω.

b. Si $\mathbf{P}(\mathrm{A}) > 0$ et $\mathbf{P}(\mathrm{B}) > 0$, alors $\mathbf{P}(\mathrm{A}\cap\mathrm{B}) = \mathbf{P}(\mathrm{B}\,|\,\mathrm{A})\,\mathbf{P}(\mathrm{A}) = \mathbf{P}(\mathrm{A}\,|\,\mathrm{B})\,\mathbf{P}(\mathrm{B})$.

Nous n'avons vu jusqu'ici aucune formule permettant de calculer la probabilité d'une intersection d'événements à l'aide des probabilités de ces événements. Une telle formule n'existe pas dans le cas général. Les probabilités conditionnelles fournissent une méthode générale tout à fait naturelle pour calculer une probabilité d'intersection.

Proposition 1.8. Formule des probabilités composées

Pour $n \in \mathbb{N}\setminus\{0,1\}$ et $\mathrm{A}_1,\dots,\mathrm{A}_n$ des événements tels que $\mathbf{P}(\mathrm{A}_1\cap\cdots\cap\mathrm{A}_{n-1}) > 0$, alors :

$$\mathbf{P}(\mathrm{A}_1\cap\cdots\cap\mathrm{A}_n) = \mathbf{P}(\mathrm{A}_1)\,\mathbf{P}(\mathrm{A}_2\,|\,\mathrm{A}_1)\,\mathbf{P}(\mathrm{A}_3\,|\,\mathrm{A}_1\cap\mathrm{A}_2)\cdots\mathbf{P}(\mathrm{A}_n\,|\,\mathrm{A}_1\cap\cdots\cap\mathrm{A}_{n-1})$$

✗ Un cas d'école. Une urne contient initialement 7 boules noires et 3 boules blanches. On tire successivement 3 boules : si on tire une noire, on l'enlève, si on tire une blanche, on la retire, et on ajoute une noire à la place. Quelle est la probabilité de tirer 3 blanches à la suite ? Notons B_i (resp. N_i) les événements *une blanche est tirée au tirage n° i* (resp. une noire). On a

$$\mathbf{P}(\mathrm{B}_1\cap\mathrm{B}_2\cap\mathrm{B}_3) = \mathbf{P}(\mathrm{B}_1)\,\mathbf{P}(\mathrm{B}_2\,|\,\mathrm{B}_1)\,\mathbf{P}(\mathrm{B}_3\,|\,\mathrm{B}_1\cap\mathrm{B}_2) = \frac{3}{10}\times\frac{2}{10}\times\frac{1}{10}$$

car, puisque $\mathbf{P}(\mathrm{B}_1) > 0$, on a $\mathbf{P}(\mathrm{B}_1\cap\mathrm{B}_2) = \mathbf{P}(\mathrm{B}_1)\,\mathbf{P}(\mathrm{B}_2\,|\,\mathrm{B}_1) = \frac{3\times 2}{10^2} \neq 0$.

Les probabilités conditionnelles permettent aussi de calculer la probabilité d'un événement en conditionnant par tous les cas possibles (de probabilité non nulle). La formule qui suit est une version conditionnelle de la formule des probabilités totales étudiée plus haut (cf. la proposition 2.1).

Proposition 1.9. Formule des probabilités totales (⚡ 1.4)

Si $(A_1,\dots,A_n)$ est un système complet d'événements ayant chacun une probabilité non nulle et si B est un événement, alors

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B \mid A_i)\,\mathbf{P}(A_i)$$

Dans la pratique, on étend cette formule au cas où certains A_i sont de probabilité nulle en convenant que $\mathbf{P}(B \mid A_i)\,\mathbf{P}(A_i) = 0$ (la formule reste valable car ce terme vaut $\mathbf{P}(B \cap A_i)$ qui est effectivement nul si $\mathbf{P}(A_i) = 0$).

Cette formule rend des services colossaux quand les expériences aléatoires se compliquent, par exemple lorsqu'au beau milieu d'une expérience aléatoire, la suite de l'expérience dépend d'un résultat aléatoire. Dans ce cas de figure, l'espace Ω devient un peu plus mystérieux et la combinatoire a ses limites. Le conditionnement permet de sortir de l'impasse…

✗ Des boules et des urnes. On considère deux urnes U_1 et U_2. L'urne U_1 contient r_1 boules rouges et v_1 boules vertes. L'urne U_2 contient r_2 boules rouges et v_2 boules vertes. On lance un dé. S'il indique le chiffre 1, on choisit l'urne U_1, sinon on choisit U_2. Dans chaque cas on effectue deux tirages avec remise dans l'urne choisie. Quelle est la probabilité d'obtenir une rouge au premier tirage ? deux rouges en tout ? On note U_i les événements *l'urne* U_i *a été choisie* puis R_i (resp. V_i) les événements *une boule rouge a été obtenue lors du tirage n° i* (resp verte). On a

$$\mathbf{P}(R_1) = \mathbf{P}(R_1 \mid U_1)\,\mathbf{P}(U_1) + \mathbf{P}(R_1 \mid U_2)\,\mathbf{P}(U_2) = \frac{r_1}{r_1+v_1} \times \frac{1}{6} + \frac{r_2}{r_2+v_2} \times \frac{5}{6}$$

et $$\mathbf{P}(R_1 \cap R_2) = \mathbf{P}(R_1 \cap R_2 \mid U_1)\,\mathbf{P}(U_1) + \mathbf{P}(R_1 \cap R_2 \mid U_2)\,\mathbf{P}(U_2) = \left(\frac{r_1}{r_1+v_1}\right)^2 \times \frac{1}{6} + \left(\frac{r_2}{r_2+v_2}\right)^2 \times \frac{5}{6}.$$

Les formules suivantes, dites de Bayes, permettent « *d'inverser le conditionnement* ».

Proposition 1.10. Formules de Bayes

a. Si A et B sont deux événements tels que $P(A) > 0$ et $P(B) > 0$, alors $\mathbf{P}(A \mid B) = \dfrac{\mathbf{P}(B \mid A)\,\mathbf{P}(A)}{\mathbf{P}(B)}$.

b. Si $(A_i)_{1 \leqslant i \leqslant n}$ est un système complet d'événements de probabilités non nulles et si B est un événement de probabilité non nulle, alors

$$\mathbf{P}\big(A_j \mid B\big) = \frac{\mathbf{P}\big(B \mid A_j\big)\,\mathbf{P}\big(A_j\big)}{\displaystyle\sum_{i=1}^{n} \mathbf{P}(B \mid A_i)\,\mathbf{P}(A_i)}$$

Voici une application typique de ces formules.

✗ Un classique de biostatistique. Un individu est tiré au hasard dans une population où l'on trouve une proportion 10^{-1} de malades du Covid-19. On lui fait passer un test de détection du Covid-19. Par ailleurs, des expérimentations antérieures ont permis de savoir que les probabilités d'avoir un résultat positif lors de l'application du test si l'individu est malade, ou s'il ne l'est pas, sont respectivement égales à $0,99$ (c'est la sensibilité du test) et à $0,17$ ($0,83 = 1-0,17$ est la spécificité du test). Sachant que le test donne un résultat positif, quelle est la probabilité pour que l'individu soit effectivement malade du Covid-19 ? On note T_+ l'événement *le test est positif* et C l'événement *l'individu est atteint du Covid-19*. Les hypothèses nous donnent les valeurs suivantes : $\mathbf{P}(C) = 10^{-1}$, $\mathbf{P}(T_+ \mid C) = 0,99$ et $\mathbf{P}\left(T_+ \mid \overline{C}\right) = 0,17$. On nous demande de calculer $\mathbf{P}(C \mid T_+)$. C'est typiquement une inversion du conditionnement donc une application des formules de Bayes :

$$\mathbf{P}(C \mid T_+) = \frac{\mathbf{P}(T_+ \mid C)\,\mathbf{P}(C)}{\mathbf{P}(T_+ \mid C)\,\mathbf{P}(C) + \mathbf{P}\left(T_+ \mid \overline{C}\right)\mathbf{P}\left(\overline{C}\right)} = \frac{99 \times 10^{-2} \times 10^{-1}}{99 \times 10^{-2} \times 10^{-1} + 17 \times 10^{-2} \times 9 \times 10^{-1}} \approx 0,39$$

- Nous recommandons au lecteur le test (⚡ **1.5**).

4. Indépendance

Intuitivement, deux événements A et B sont indépendants si le fait de savoir que A est réalisé ne donne aucune information sur la réalisation de B et réciproquement. Supposons que la réalisation de l'événement B n'ait aucune influence sur la réalisation de A. Alors, après n épreuve de l'expérience, la fréquence empirique de réalisation de A sera approximativement la même, que l'on sache ou non que B est réalisé. Ainsi donc,

$$f_n(A|B) := \frac{f_n(A \cap B)}{f_n(B)} \simeq f_n(A)$$

Le conditionnement ne change pas l'information que l'on a sur l'expérience. Par passage à la limite pour $n \to +\infty$, nous en déduisons les définitions suivantes. Si B est un événement de probabilité strictement positive, A sera dit indépendant de B si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$. On remarque que cette formule se symétrise et la notion d'indépendance se définit finalement plus efficacement comme suit.

4.1. Indépendance deux à deux et indépendance mutuelle

Définition 1.11. Événements indépendants

Les événements A et B sont dits indépendants si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B)$. On note cett propriété $A \perp\!\!\!\perp B$.

Cette notation n'est pas au programme mais nous l'utiliserons tout de même par comoditié (et surtout par paresse).

- Si A est un événement tel que $\mathbf{P}(A) = 0$ ou $\mathbf{P}(A) = 1$, alors il est indépendant de tout événement, y compris de lui-même (c'est le cas en particulier pour Ω et $\varnothing$).
  - ✔ Cas 1 : $\mathbf{P}(A) = 0$. Soit B un événement. Comme $A \cap B \subset A$, on a $\mathbf{P}(A \cap B) \leqslant \mathbf{P}(A)$ par croissance de la probabilité. Ainsi $\mathbf{P}(A \cap B) = 0 = \mathbf{P}(A)\,\mathbf{P}(B)$.
  - ✔ Cas 2 : $\mathbf{P}(A) = 1$. Soit B un événement. Comme $A \subset A \cup B$, on a $\mathbf{P}(A) \leqslant \mathbf{P}(A \cup B)$ par croissance de la probabilité. Ainsi $\mathbf{P}(A \cup B) = 1$ d'où $\mathbf{P}(A \cap B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cup B) = \mathbf{P}(A)\,\mathbf{P}(B)$.
- Deux événements incompatibles A et B de probabilités non nulles ne sont jamais indépendants.
- On lance 2 fois un dé. Si A_i est un événement qui ne dépend que du i-ème lancé, alors $A_1 \perp\!\!\!\perp A_2$.
- On tire une carte au hasard dans un jeu de 52 cartes. Les événements suivants sont indépendants :

  A = « *la carte est une dame* » et B = « *la carte est un coeur* »

  En effet, $\mathbf{P}(A \cap B) = \frac{1}{52} = \mathbf{P}(A)\,\mathbf{P}(B)$ car $\mathbf{P}(A) = \frac{4}{52}$ et $\mathbf{P}(B) = \frac{13}{52}$.

- Dans le premier cas, on a des épreuves différentes « *indépendantes* » (au sens du français) d'une même expérience aléatoire; dans cette situation, on aura toujours l'indépendance de p-uplets d'événements associés à des épreuves différentes : cette indépendance est supposée par le modèle. Dans le second cas, l'indépendance vient de la symétrie entre les couleurs. C'est le lemme des Bergers qui unifie ces différents calculs de probabilité.
- Le lecteur pourra aborder le test (⚡ **1.6**).

La propositon qui suit est une conséquence directe de la définition de l'indépendance :

Proposition 1.11. Indépendance et conditionnement

On suppose que $\mathrm{P}(\mathrm{B}) > 0$. On a $\mathrm{A} \perp\!\!\!\perp \mathrm{B} \iff \mathbf{P}(\mathrm{A} \mid \mathrm{B}) = \mathbf{P}(\mathrm{A})$.

Proposition 1.12. Indépendance et passage au complémentaire

Si $\mathrm{A} \perp\!\!\!\perp \mathrm{B}$, alors $\mathrm{A} \perp\!\!\!\perp \overline{\mathrm{B}}$, $\overline{\mathrm{A}} \perp\!\!\!\perp \overline{\mathrm{B}}$ et $\overline{\mathrm{A}} \perp\!\!\!\perp \overline{\mathrm{B}}$.

On étend la définiton de l'indépendance à un nombre fini d'événements.

Définition 1.13. Famille finie d'événements mutuellement indépendants

Les événements $\mathrm{A}_1, \ldots, \mathrm{A}_n$ sont dits mutuellement indépendants si, pour tout $\Lambda \subset [\![1, n]\!]$,

$$\mathbf{P}\left(\bigcap_{\ell \in \Lambda} \mathrm{A}_\ell\right) = \prod_{\ell \in \Lambda} \mathbf{P}(\mathrm{A}_\ell)$$

On note de même $\mathrm{A}_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \mathrm{A}_n$.

- Cas typique d'épreuves indépendantes d'une même expérience aléatoire. On lance n fois un dé. Si A_i est un événement qui ne dépend que du lancé $n^{\mathrm{o}}\ i$, alors $\mathrm{A}_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp \mathrm{A}_n$.
- L'indépendance des A_i deux à deux n'entraîne pas leur indépendance mutuelle si $n \geqslant 3$. Considérons par exemple la loi uniforme sur $\Omega = [\![1, 4]\!]$ et les événements $\mathrm{A} = \{1, 2\}$, $\mathrm{B} = \{1, 3\}$ et $\mathrm{C} = \{1, 4\}$. On a
$$\mathbf{P}(\mathrm{A} \cap \mathrm{B}) = \mathbf{P}(\{1\}) = \frac{1}{4} = \mathbf{P}(\mathrm{A})\,\mathbf{P}(\mathrm{B})$$
et de même pour $\mathrm{A} \cap \mathrm{C}$ et $\mathrm{B} \cap \mathrm{C}$. Cependant, $\mathbf{P}(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C}) = \dfrac{1}{4} \neq \dfrac{1}{8} = \mathbf{P}(\mathrm{A})\,\mathbf{P}(\mathrm{B})\,\mathbf{P}(\mathrm{C})$.
- Le lecteur abordera le test (⚡ **1.7**)avec profit.

On ne précise pas toujours *mutuellement*, en revanche il faudra toujours préciser *deux à deux* si c'est le cas.

La proposition 4.1 se généralise à la mutuelle indépendance.

Proposition 1.14.

Soit A_1, …, A_n des événements indépendants. En changeant certains A_i en leur contraire, on obtient à nouveau des événements indépendants.

4.2. Modélisation une série finie d'expériences indépendantes

Un des objectifs principaux de la théorie des probabilités est d'étudier les répétitions d'expériences aléatoires afin d'en dégager des propriétés asymptotiques (cf. les deux grands théorèmes de la théorie : la loi des grands nombres et le théorème Central-Limite). Le résultat abstrait suivant donne accès à la modélisation d'une répétition d'expériences indépendantes.

Proposition 1.15. Produit d'espaces probabilisés

Soit $n \in \mathbb{N}^*$ et $(\Omega_1, \mathbf{P}_1)$, …, $(\Omega_n, \mathbf{P}_n)$ des espaces probabilisés finis. Il existe une unique loi de probabilité $\mathbf{P}$ sur $\Omega := \Omega_1 \times \cdots \times \Omega_n$ telle que

$$\forall (A_1, \ldots, A_n) \in \prod_{i=1}^{n} \mathscr{P}(\Omega_i)\,,\ \mathbf{P}(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \mathbf{P}_i(A_i)$$

On note $\mathbf{P} = \mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_n$ cette probabilité dite produit des $\mathbf{P}_i$.

Dans ce modèle, deux événements portant sur des expériences différentes sont indépendants. C'est-à-dire que pour tout $(i, j) \in [\![1, n]\!]$ tel que $i < j$ et pour tout $(A_i, A_j) \in \mathscr{P}(\Omega_i) \times \mathscr{P}(\Omega_j)$, on a

$$\overbrace{\prod_{k=1}^{n} B_k}^{B:=} \perp\!\!\!\perp \overbrace{\prod_{k=1}^{n} C_k}^{C:=} \quad \text{où } \forall k \in [\![1, n]\!]\,,\ B_k = \begin{cases} \Omega_k & \text{si } k \neq i \\ A_i & \text{sinon} \end{cases} \text{ et } C_k = \begin{cases} \Omega_k & \text{si } k \neq j \\ A_j & \text{sinon} \end{cases}$$

En effet :

$$\mathbf{P}(B \cap C) = \mathbf{P}\left(\prod_{k=1}^{n} (B_k \cap C_k)\right) = \prod_{i=1}^{n} \mathbf{P}_i\,(B_k \cap C_k) = \mathbf{P}_i(A_i)\,\mathbf{P}_j(A_j) = \mathbf{P}(B)\,\mathbf{P}(C)$$

car $B_k \cap C_k = \begin{cases} A_k & \text{si } k \in \{i, j\} \\ \Omega_k & \text{sinon} \end{cases}$ pour tout k dans $[\![1, n]\!]$.

- Jouons à pile ou face. Soit $p \in [0, 1]$, $n \in \mathbb{N}^*$ et $\Omega := \{0, 1\}^n$. On peut modéliser n parties indépendantes de pile ou face de plusieurs manières :
  - On muni Ω de la probabilité définie sur les singletons par :

    $$\mathbf{P}(\{\omega_1, \ldots, \omega_n\}) = p^k (1-p)^{n-k} \quad \text{où } k := \#\{\, i \in [\![1, n]\!]\,;\, \omega_i = 1\}$$

  - On muni $\{0, 1\}$ de la loi de probabilité définie par $\mathbf{P}_1(\{0\}) = 1 - p$ et $\mathbf{P}_1(\{1\}) = p$ puis Ω de la probabilité produit $\mathbf{P}_1 \otimes \cdots \otimes \mathbf{P}_1$.

  En fait les deux modèles sont rigoureusement identiques, seule la construction diffère. On peut trouver le second point de vue plus « économique » et naturel.

5. Énoncés des tests

1.1.

Soit $(\Omega, \mathscr{T}, \mathbf{P})$ un espace probabilisé et A, B, C des événements. Les propriétés suivantes sont-elles vraies ou fausses ?

a. Il est possible d'avoir $\mathbf{P}(A) = 0.9$ et $\mathbf{P}(B) = -0.5$.

b. Si $B \cup C = \Omega$, alors $\mathbf{P}(A) = \mathbf{P}(A \cap B) + \mathbf{P}(A \cap C)$.

c. Il est possible d'avoir $\mathbf{P}(A) = 0.9$ et $\mathbf{P}(B) = 0.3$ et $A \cap B = \varnothing$.

d. Il est possible d'avoir $\mathbf{P}(A) = 0.8$, $\mathbf{P}(B) = 0.1$ et $\mathbf{P}(A \cap B) = 0.2$.

e. Il est possible d'avoir $\mathbf{P}(A) = 0.8$, $\mathbf{P}(B) = 0.4$ et $\mathbf{P}(A \cap B) = 0.3$.

f. Si $\Omega = [\![1,6]\!]$, $\mathbf{P}([\![1,2]\!]) = \mathbf{P}([\![2,3]\!]) = \mathbf{P}([\![4,5]\!]) = \dfrac{1}{3}$ et $\mathbf{P}([\![4,6]\!]) = \dfrac{1}{2}$, alors $\mathbf{P}(\{6\}) = \dfrac{1}{6}$.

g. Sous les mêmes hypothèses, $\mathbf{P}(\{4\}) = \dfrac{1}{6}$.

1.2.

Soit A et B deux événements d'un espace probabilisé fini $(\Omega, \mathbf{P})$. Montrer que

$$A \cup B = \Omega \implies \mathbf{P}(A \cap B) = \mathbf{P}(A)\,\mathbf{P}(B) - \mathbf{P}\big(\overline{A}\big)\mathbf{P}\big(\overline{B}\big)$$

1.3.

Soit A et B deux événements d'un espace probabilisé fini $(\Omega, \mathbf{P})$ tels que

$$\mathbf{P}(A) = \frac{1}{3}\,, \quad \mathbf{P}(B) = \frac{1}{4} \quad \text{et} \quad \mathbf{P}(A \cup B) = \frac{4}{9}$$

Calculer $\mathbf{P}(A \mid B)$, $\mathbf{P}\big(\overline{A} \mid B\big)$ et $\mathbf{P}(A \cap B \mid B)$.

1.4.

On dispose de six urnes numérotées de 1 à 6. L'urne numéro k comporte k boules blanches et une boule rouge. Un joueur lance un dé équilibré puis choisit une boule dans l'urne correspondant au résultat du dé. Déterminer la probabilité que la boule tirée soit blanche.

1.5.

Dans une usine, on fabrique des composants électroniques sur trois machines M_1, M_2 et M_3 produisant respectivement $50\%, 30\%$ et 20% du total. Un technicien estime que 2% des composants fabriqués par M_1, 3% des composants fabriqués par M_2 et 5% des composants fabriqués par M_3 sont défectueux.

a. Quelle est la probabilité qu'un composant pris au hasard à la sortie de l'usine soit défectueux ?

b. Quelle est la probabilité d'obtenir une pièce défectueuse provenant de M_1 ?

c. Les événements « *la pièce est défectueuse* » et « *la pièce provient de* M_1 » sont-ils indépendants ?

d. Un composant est défectueux. Quelle est la probabilité qu'il provienne de M_1 ?

1.6.

On jette deux fois le même dé. Les événements

$$\mathrm{A} = \text{« \textit{obtention d'un chiffre pair au premier lancer} »} \quad \text{et} \quad \mathrm{B} = \text{« \textit{obtention du} 1 \textit{au deuxième lancer} »}$$

sont indépendants.

1.7.

Soit $(\Omega, \mathbf{P})$ un espace probabilisé fini, A, B et C trois évènements mutuellement indépendants, de probabilités différentes de 0 et 1.

a. Montrer que les événements A et $\mathrm{B} \cup \mathrm{C}$ sont indépendants.

b. Montrez que $\mathbf{P}(\mathrm{B} \cup \mathrm{C})$ est strictement inférieure à 1.

6. Solutions des tests

1.1.

a. Faux. Une probabilité ne peut pas être négative.

b. Faux. Contre-exemple : on lance un dé idéal. Alors $\Omega = [\![1,6]\!]$. On considère $A = [\![3,4]\!]$, $B = [\![1,3]\!]$ et $C = [\![3,6]\!]$. Alors $B \cup C = \Omega$ mais

$$P(A) = \frac{1}{3} \neq \frac{1}{6} + \frac{1}{2} = P(A \cap B) + P(A \cap C).$$

c. Faux. Une probabilité ne peut pas être plus grande que 1, mais ici

$$P(A \cup B) = P(A) + P(B) = 0.9 + 0.3 = 1.2.$$

d. Faux. Puisque $A \cap B \subset B$ on a $P(A \cap B) \leqslant P(B)$, mais ici $P(A \cap B) = 0.2 > 0.1 = P(B)$.

e. C'est possible. Un exemple est $A = \{1,2\}$ et $B = \{2,3\}$ avec la loi de probabilité suivante :

x	1	2	3	4
$P(\{x\})$	0.5	0.3	0.1	0.1

f. Oui, on peut déduire que $P(\{6\}) = 1/6$ car

$$\begin{aligned} P(\{6\}) &= P(\{4,5,6\} \setminus \{4,5\}) \\ &= P(\{4,5,6\}) - P(\{4,5\}) \\ &= \frac{1}{2} - \frac{1}{3} = \frac{1}{6}. \end{aligned}$$

En revanche on ne peut pas déduire que $P(\{4\}) = \frac{1}{6}$. Voici un contre-exemple qui vérifie les hypothèses et où $P(\{4\}) \neq \frac{1}{6}$.

x	1	2	3	4	5	6
$P(\{x\})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	0	$\frac{1}{6}$

1.2.

Notons

$$\delta := \mathbf{P}(A)\,\mathbf{P}(B) - \mathbf{P}\big(\overline{A}\big)\mathbf{P}\big(\overline{B}\big)$$

On a

$$\begin{aligned} \delta &= \mathbf{P}(A)\,\mathbf{P}(B) - (1 - \mathbf{P}(A))(1 - \mathbf{P}(B)) \\ &= \mathbf{P}(A) + \mathbf{P}(B) - 1 \end{aligned}$$

Comme

$$\begin{aligned} \mathbf{P}(\Omega) = 1 &= \mathbf{P}(A \cup B) \\ &= \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \end{aligned}$$

on a

$$\delta = \mathbf{P}(A \cap B)$$

1.3.

On trouve $\mathbf{P}(A \mid B) = 5/9$, $\mathbf{P}\big(\overline{A} \mid B\big) = 4/9$ et $\mathbf{P}(A \cap B \mid B) = 5/9$.

1.4.

On considère le système complet d'événements $(A_1, \ldots, A_6)$ avec $A_k =$ « *le dé donne la valeur* k » et on étudie l'événement $B =$ « *la boule tirée est blanche* ». On a $\mathbf{P}(A_k) = 1/6$ et $\mathbf{P}(B \mid A_k)) = k/(k+1)$. Par formule des probabilités totales

$$\mathbf{P}(B) = \frac{1}{6} \sum_{k=1}^{6} \frac{k}{k+1} = \frac{617}{840}$$

1.5.

a. Notons D cette probabilité. On a

$$\begin{aligned} P(D) &= P(D|M_1)P(M_1) + P(D|M_2)P(M_2) \\ &\quad + P(D|M_3)P(M_3) \\ &= \frac{2}{100} \times \frac{50}{100} + \frac{3}{100} \times \frac{30}{100} + \frac{5}{100} \times \frac{20}{100} \\ &= \frac{29}{1000} \end{aligned}$$

b. On a

$$P(M_1 \cap D) = P(D|M_1)P(M_1) = \frac{1}{100}$$

c. Comme

$$P(D|M_1) = \frac{2}{100} \neq \frac{29}{1000} = P(D),$$

les événements « *la pièce est défectueuse* » et « *la pièce provient de* M_1 » ne sont indépendants.

d. On a

$$P(M_1|D) = \frac{P(D|M_1)P(M_1)}{P(D)} = \frac{10}{29}$$

1.6.

On choisit $\Omega = [\![1,6]\!]^2$ et l'équiprobabilité, on vérifie que

$$\mathbf{P}(A) = \frac{3\times 6}{36} = \frac{1}{2},\ \mathbf{P}(B) = \frac{6\times 1}{36} = \frac{1}{6}$$

et

$$\mathbf{P}(A\cap B) = \frac{3\times}{36} = \frac{1}{12} = \mathbf{P}(A)\,\mathbf{P}(B)$$

1.7.

a. $P\big(A\cap(B\cup C)\big) = P(A\cap B) + P(A\cap C) - P(A\cap B\cap C) = P(A)\big(P(B)+P(C)-P(B\cap C)\big) = P(A)P(B\cup C)$.

b. $1 - P(B\cup C) = P(\overline{B})P(\overline{C}) > 0$.