GRÉGOIRE DUPONT

PROBABILITÉS ET STATISTIQUES POUR L'ENSEIGNEMENT

CAPES/CAPLP
AGRÉGATION INTERNE
MATHÉMATIQUES

NOUVEAU
CONCOURS

DUNOD

Création graphique de la couverture : Hokus Pokus Créations

5 rue Laromiguière, 75005 Paris
www.dunod.com

ISBN 978-2-10-072192-4

Table des matières

Introduction

Le contexte épistémologique et didactique

Des probabilités classiques aux statistiques inférentielles

Apparu au XVIIe siècle avec des mathématiciens comme Pierre de Fermat (1601–1655), Blaise Pascal (1623–1662) ou Jakob Bernoulli (1654–1705), le calcul des probabilités s'est initialement intéressé à la modélisation de jeux de hasard ou, plus généralement, de phénomènes aléatoires par nature. Cette étude consistait alors principalement en des problèmes de dénombrement, un champ que l'on nomme aujourd'hui les **probabilités classiques**.

En 1713, la publication par Jakob Bernoulli de l'*Ars Conjectandi* introduisit l'idée que la notion de probabilité était liée à celle de fréquence de réalisation d'un événement. Ce **principe fréquentiste**, appelé plus tard **loi des grands nombres**, créa une rupture épistémologique en montrant que le hasard tend à suivre des lois, si tant est que l'on observe des échantillons de taille suffisante. Ceci fut renforcé par les travaux d'Abraham De Moivre (1667–1754) et de Pierre-Simon de Laplace (1749–1827) sur ce qui allait devenir le **théorème *central limit***. Il s'agissait des prémisses de la statistique.

Ces idées furent reprises et développées au XIXe siècle par Karl Friedrich Gauss (1777–1855) qui montra que l'accumulation d'erreurs individuellement incontrôlables pouvait s'appréhender de manière globale. Ceci, ainsi que les travaux de James Clark Maxwell (1831–1879) sur la modélisation probabiliste de la cinétique des gaz ou ceux de Ludwig Boltzmann (1844–1906) sur la physique statistique furent à l'origine d'un nouveau glissement épistémologique : il était possible d'utiliser le calcul des probabilités dans des situations *a priori* déterministes, le hasard n'intervenant que comme une façon de modéliser l'ignorance d'informations trop nombreuses ou trop complexes à mesurer.

À la fin du XIXe siècle, le positivisme scientifique était encore de mise et il pouvait sembler que l'utilisation de méthodes probabilistes dans des situations déterministes n'était qu'une limitation de nos connaissances appelée à être dépassée. Mais la publication par Henri Poincaré (1854–1912) en 1892 du premier volume des *Méthodes nouvelles en mécanique céleste* allait changer la donne. En effet, ces travaux mirent en évidence l'existence de ce que l'on appelle aujourd'hui le **chaos déterministe**, c'est-à-dire de systèmes déterministes si sensibles aux conditions initiales que seule une connaissance exacte de l'ensemble de ces conditions permettrait de prédire avec certitude l'évolution du système. Puisqu'il n'est pas imaginable d'avoir accès à l'ensemble des informations contenues dans l'univers à un instant donné, Poincaré nous condamnait de ce fait à utiliser des méthodes non déterministes pour la modélisation de situations relevant de tels systèmes.

Ce nouveau saut épistémologique permit au calcul des probabilités de se développer dans de nombreux domaines où l'on ne l'attendait initialement pas. Mais encore fallait-il être en mesure d'exploiter les résultats obtenus pour savoir quelles conclusions tirer d'un modèle probabiliste, et quel modèle choisir face à un phénomène aléatoire. Le principal pas en avant dans cette direction fut effectué en 1900 par Karl Pearson (1857–1936) qui introduisit le concept de **test statistique**. Ces tests, dont l'étude a été systématisée par son fils Egon Pearson (1895–1980) et son collègue Jerzy Neyman (1894–1981), ont donné naissance à ce que l'on nomme aujourd'hui les **statistiques inférentielles**, lesquelles forment l'un des outils essentiels de nombreux domaines de la vie courante : sciences médicales, humaines et sociales, économie, finance, physique théorique, etc...

Mais bien qu'omniprésentes dans le monde qui nous entoure, ces méthodes statistiques et leur interprétation sont longtemps restées à l'écart des programmes d'enseignement, demeurant souvent mystérieuses pour nombre de nos concitoyens, ouvrant ainsi la porte à des raisonnements incorrects et à des décisions erronées face aux innombrables données statistiques qui sont communiquées quotidiennement. L'éducation à ces outils mathématiques contemporains est donc devenue une nécessité, comme cela a été acté par les nouveaux programmes du secondaire.

Des nouveaux programmes

Jusqu'en 2009, l'étude des probabilités et des statistiques dans les programmes de l'enseignement secondaire général et professionnel concernait seulement :

- les **probabilités classiques**, principalement à travers l'étude de la théorie des jeux et de problèmes de dénombrement,
- les **statistiques descriptives**, essentiellement à travers le calcul d'indicateurs de position et de dispersion pour des séries statistiques numériques.

Cependant, comme indiqué dans les documents ressources accompagnant les programmes de 2009 en lycée professionnel « *Pour décrypter le monde moderne, participer au débat démocratique, exercer son esprit critique, optimiser ses activités professionnelles, « l'honnête homme » du XXIe siècle doit être éduqué aux méthodes statistiques et aux probabilités.* » et « *Les précédents programmes de baccalauréat professionnel ne laissaient qu'une très faible place aux probabilités* [...] *avec une approche fondée sur le dénombrement des cas possibles. Cette approche a montré ses limites face aux enjeux décrits précédemment.* ».

Les nouveaux programmes, que ce soit pour la voie générale et technique ou pour la voie professionnelle, ont donc entrepris de mettre en adéquation la formation en probabilités et en statistiques avec les méthodes actuellement mises en œuvre dans notre société. C'est ainsi que les problèmes de dénombrement ont cédé la place à des problèmes d'échantillonnage et d'estimation s'attachant particulièrement à l'observation de la stabilisation des fréquences et au contrôle de ses fluctuations.

Les statistiques descriptives restent au programme mais l'accent est maintenant mis sur leur interprétation et leur exploitation pour la **prise de décision** face à un phénomène non déterministe, notamment par le biais d'intervalles de confiance et de fluctuations.

Un autre changement significatif s'opère dans l'approche de ces nouveaux programmes. Alors que jusqu'ici l'enseignement des mathématiques se faisait de manière essentiellement **logico-déductive**, c'est-à-dire formelle et démonstrative, les nouveaux programmes introduisent une approche **hypothético-déductive** fondée sur une démarche expérimentale. Ceci n'est pas sans soulever des difficultés pour des élèves, des parents d'élèves et des enseignants qui n'ont pas été formés à cette école. Et pourtant l'enseignant doit avoir d'autant plus de recul sur ce sujet pour en faire passer le sens qu'il n'est pas en mesure de donner toutes les démonstrations à ses élèves (convergence de l'intégrale de Gauss, théorème de Moivre-Laplace, etc...). Il lui faudra donc adopter dans ses cours une posture plus proche de celle de l'expérimentateur, présentant un théorème comme on présente une loi physique et montrant sa véracité à l'aide d'exemples et d'expérimentations bien choisis plutôt qu'en la démontrant à l'aide d'arguments formels.

Des nouvelles technologies

Ce revirement didactique opéré par les programmes repose en grande partie sur les possibilités offertes par les nouvelles technologies en matière de constructions et de manipulations d'exemples nombreux et variés. L'intégration du numérique est même l'une des conditions *sine qua non* de la réussite de la mise en œuvre de ces nouveaux programmes.

Les technologies sont en effet les outils idéaux pour multiplier les exemples et simuler des phénomènes aléatoires. Leur utilisation est d'autant plus nécessaire que les formules mises en œuvre pour les intervalles de confiances ou de fluctuations ne sont pas suffisamment naturelles pour être admises dès le premier abord par l'apprenant. Il est donc important que le futur enseignant prenne le recul nécessaire sur les technologies et sur les notions enseignées afin de voir comment celles-ci peuvent servir l'apprentissage de celles-là.

De notre point de vue, les outils les plus adéquats à ce jour pour enseigner le calcul des probabilités au secondaire semblent être le tableur, un logiciel d'algorithmique ainsi qu'un logiciel de géométrie dynamique. Favorisant les logiciels libres, nous préconisons donc LibreOffice, AlgoBox et Geogebra mais il ne s'agit là que de conseils, chacun adaptera son enseignement à son propre contexte (équipements de l'établissement, des élèves, etc...).

Des nouveaux concours

Les concours de l'enseignement, réformés pour partie en 2014, tiennent particulièrement compte des évolutions récentes des programmes et des technologies.

Les probabilités et les statistiques sont devenues depuis plusieurs années des thèmes incontournables lors des épreuves d'admissibilité et d'admission de l'agrégation interne, du CAPES et du CAPLP. Le fait que les nouveaux programmes de lycée aient décidé de mettre l'accent sur ces thèmes risque probablement de conserver cette dynamique.

Les jurys sont d'ailleurs d'autant plus vigilants quant à ces notions qu'ils déplorent souvent qu'elles soient mal maîtrisées des candidats, bien que la situation semble s'être améliorée lors des dernières sessions.

La capacité à intégrer le numérique dans l'enseignement fait partie des compétences évaluées aux concours. Cette intégration étant particulièrement souhaitable (et souhaitée) dans le cadre des probabilités et des statistiques, il est certain que le jury y sera particulièrement attentif.

Ceci est vrai pour l'ensemble des épreuves orales d'admission des différents concours. Mais il ne serait pas judicieux de croire que ces compétences ne pourraient pas être évaluées dès les épreuves écrites d'admissibilité ; la simulation de phénomènes probabilistes a par exemple fait l'objet d'exercices complets du CAPLP aux sessions 2013 et 2014 anticipée.

Les objectifs de l'ouvrage

Des contenus disciplinaires clairs et complets

Si la nouvelle approche de l'enseignement des probabilités et des statistiques demande de développer un enseignement moins axé sur le formalisme que sur l'intuition, il n'en est pas moins nécessaire pour l'enseignant d'avoir une excellente connaissance et compréhension du domaine. Ceci soulève plusieurs difficultés.

D'abord, même si cela tend à changer ces dernières années, la formation aux probabilités et aux statistiques s'est longtemps limitée à une approche classique fondée sur le dénombrement, les statistiques inférentielles ne se retrouvant que dans certains parcours, le plus souvent optionnels.

Une autre difficulté est causée par le fait que ce sujet a souvent été l'apanage de deux catégories de mathématiciens : les statisticiens théoriciens d'une part, produisant une bibliographie sophistiquée dépassant largement le cadre des programmes d'un premier cycle universitaire et des concours de l'enseignement ; les statisticiens appliqués d'autre part, produisant le plus souvent des ouvrages d'exercices d'application, indispensable pour développer d'excellentes compétences techniques mais insuffisants pour acquérir le recul théorique nécessaire à l'enseignement de ces notions.

L'objectif du présent ouvrage est donc de donner des éléments théoriques permettant de maîtriser ces nouveaux programmes tout en ne nécessitant qu'un bagage analytique classique de premier cycle universitaire. Et si nous nous permettons parfois certains écarts par rapport aux contenus des programmes du secondaire, c'est uniquement dans le but d'apporter plus de recul sur les notions enseignées.

Nous avons aussi fait en sorte de garder un point de vue très mathématique, privilégiant les énoncés analytiques ou algébriques formels afin de ne laisser aucune place au doute. Ces énoncés formels sont complétés par de nombreuses remarques heuristiques informelles visant à aider le candidat à donner du sens à ces notions.

Chaque chapitre de l'ouvrage contient aussi un certain nombre d'exercices-clés dont les corrigés détaillés permettent de vérifier la bonne acquisition des notions abordées.

Une contextualisation didactique

Dans la perspective des concours d'enseignement, il nous a semblé utile de régulièrement faire un point sur l'intégration des notions rencontrées dans les programmes de l'enseignement secondaire général ou professionnel ainsi que dans le programme des concours. L'ouvrage est donc régulièrement parsemé de **Points programmes** qui se réfèrent aux Bulletins Officiels stipulant les programmes en vigueur au jour où est écrit ce manuscrit.

Pour les classes de collège, il s'agit du BO spécial n°6 du 28 août 2008.

Concernant le baccalauréat professionnel, il s'agit du BO spécial n°2 du 19 février 2009.

Concernant l'enseignement général, pour les classes de Seconde, il s'agit du BO n° 30 du 23 juillet 2009. Pour les classes de Premières, il s'agit du BO n°9 du 30 septembre 2010. Enfin, pour les Terminales, il s'agit du BO n°8 du 13 octobre 2011.

Les programmes de concours se référant aux programmes de CPGE, ceux qui ont été retenus sont ceux des filières MPSI et MP parus respectivement au BO spécial n°3 du 30 mai 2013 et au BO spécial n°1 du 23 janvier 2014.

L'Annexe C présente aussi une synthèse de l'organisation des programmes de probabilités et de statistiques de l'enseignement général.

Une aide à l'aide à l'intégration du numérique

Nous avons déjà mentionné la grande utilité des technologies pour l'enseignement des nouveaux programmes de probabilités et de statistiques. Afin d'aider le candidat et le futur enseignant à intégrer au mieux le numérique dans son enseignement de ces disciplines, l'ouvrage est régulièrement parcouru de **Points numériques** qui donnent des conseils techniques ou pédagogiques afin de mobiliser le numérique de manière pertinente lors d'une séance d'enseignement. Toutes les fonctions de tableur proposées respectent la syntaxe du logiciel libre LibreOffice 4.0.

Une contextualisation historique

Afin de replacer ce nouvel enseignement dans son contexte historique, nous avons ajouté une note biographique à chaque apparition d'un nom de mathématicien. À cette fin, nous nous sommes souvent inspirés de l'excellent ouvrage *Des mathématiciens de A à Z* de Hauchecorne et Suratteau [HS96].

L'organisation de l'ouvrage

L'ouvrage s'articule autour de sept chapitres et de trois annexes.

Le Chapitre 1 présente les éléments théoriques nécessaires au calcul des probabilités. Il reprend de manière détaillée toutes les notions théoriques essentielles relatives aux lois de probabilités et aux variables aléatoires. Conformément à l'esprit de ces programmes, le but de ce chapitre n'est pas d'approfondir le calcul classique des probabilités mais de se familiariser avec les outils du calcul probabiliste en vue du développement de méthodes liées aux statistiques inférentielles dans les chapitres ultérieurs.

Le Chapitre 2 présente de manière détaillée les différentes lois qui jouent un rôle clé en modélisation et en statistique. Leurs propriétés mathématiques essentielles sont rappelées et démontrées et un point est systématiquement fait sur leur rôle pratique en modélisation.

Le Chapitre 3 complète ces rappels en présentant les grands résultats de convergences qui sous-tendent l'ensemble des programmes du secondaire. On y trouve en particulier la **loi des grands nombres** qui justifie le principe fréquentiste et le **théorème de Moivre-Laplace** qui permet de contrôler les fluctuations lors d'un échantillonnage.

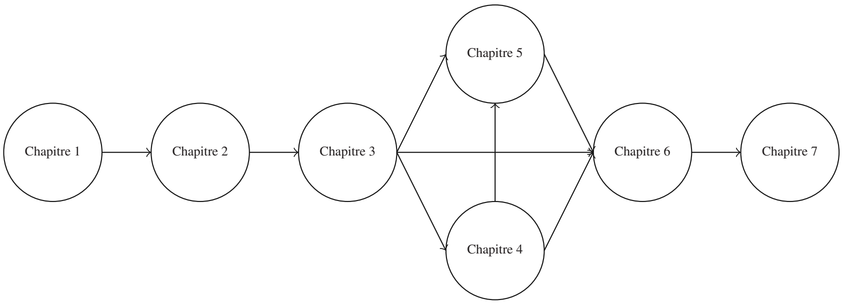
Le Chapitre 4 est dédié à l'étude de **méthodes de Monte-Carlo** permettant de simuler numériquement des variables aléatoires ou, plus généralement, de répondre à des problèmes déterministes à l'aide d'outils probabilistes.

Le Chapitre 5 introduit des éléments de théorie de l'**estimation ponctuelle** permettant d'obtenir des valeurs plausibles pour des paramètres inconnus de variables aléatoires à partir d'échantillons d'observations.

Le Chapitre 6 nous amène au cœur de la statistique inférentielle en introduisant les notions cruciales d'**intervalles de confiance et de fluctuations** pour l'estimation et l'échantillonnage. Les résultats présentés à ce chapitre sont mis en œuvre au Chapitre 7 pour répondre à des problèmes de **prise de décision** face à des phénomènes aléatoires.

Afin d'aider le lecteur, l'Annexe A propose un tableau récapitulatif des propriétés essentielles des lois vues au Chapitre 2. L'Annexe B propose les incontournables **tables de lois classiques** mais aussi – et peut-être surtout – des alternatives modernes pour les remplacer à l'ère du numérique. Enfin, l'Annexe C propose un tableau synthétisant l'organisation des programmes de probabilités et de statistiques dans l'enseignement secondaire général.

Le diagramme ci-dessous indiquera au lecteur des cheminements possibles tenant compte de l'interdépendance des différents chapitres.



Mode d'emploi à l'usage du futur enseignant

Pour la formation initiale

Cet ouvrage peut être utilisé comme un ouvrage d'introduction aux probabilités et aux statistiques par un étudiant dès la licence, le bagage mathématique nécessaire pour aborder l'essentiel des résultats de l'ouvrage est celui des deux premières années de licence de mathématiques.

Il nous semble cependant recommandable que l'étudiant ait déjà croisé les notions élémentaires du calcul classique des probabilités sans quoi le Chapitre 1 se trouvera être un peu rapide. Aucune aisance technique particulière n'est cependant requise.

L'étudiant pourra aussi gagner à se munir en parallèle d'un ouvrage d'exercices destiné à des études plus appliquées, comme par exemple le *Mini-Manuel de Probabilités et de Statistiques* [CDF07]. Ceci lui permettra de multiplier les exercices pratiques pour s'assurer de sa capacité à mettre en œuvre dans des situations pratiques les résultats vus dans cet ouvrage.

Pour l'agrégation interne

Pour les épreuves d'admissibilité de l'agrégation interne, il est absolument essentiel que le candidat maîtrise les grands théorèmes du calcul des probabilités présentés au Chapitre 1, les lois classiques vues au Chapitre 2 ainsi que les inégalités et les théorèmes de convergence du Chapitre 3.

Il est aussi nécessaire de bien maîtriser le contenu du Chapitre 6 relatif à l'estimation par intervalles. Le Chapitre 5 sur l'estimation ponctuelle aidera probablement le lecteur à mieux appréhender l'estimation par intervalles sans pour autant être explicitement au programme du concours.

Pour les épreuves d'admission, nous ne saurions que recommander la lecture attentive du Chapitre 4 et notamment les algorithmes qu'il contient. Aussi, le Chapitre 7 pourra être avantageusement mis à profit afin de mettre en œuvre des tests statistiques à l'aide des résultats vus sur l'estimation par intervalles au Chapitre 6.

Le candidat pourra également regarder attentivement les « points numériques » et les « points programmes » pour bien savoir situer et illustrer ces notions dans ses leçons et exercices. Les références historiques pourront aussi l'aider à agrémenter ses exposés lors des épreuves orales.

Enfin, nous conseillons au candidat à l'agrégation interne de se munir d'ouvrages du secondaire afin de s'assurer d'une parfaite maîtrise des applications pratiques des notions vues dans cet ouvrage. Dans ce domaine des mathématiques appliquées en particulier, la connaissance d'éléments théoriques sans la capacité de leur mise en œuvre simple est souvent lourdement sanctionnée par le jury du concours.

CAPES

Pour les épreuves d'admissibilité du CAPES, il nous semble indispensable que le candidat maîtrise le contenu du Chapitre 1, en particulier en ce qui concerne les variables aléatoires discrètes, ainsi que les lois classiques présentées au Chapitre 2, à l'exception des lois du χ^2 et de Student qui peuvent être omises en première lecture.

Il est fortement recommandé d'avoir une connaissance des inégalités classiques et des théorèmes de convergence du Chapitre 3 mais il nous semble encore plus important que le candidat se soit familiarisé avec le contenu du Chapitre 6 relatif aux intervalles de confiance et de fluctuations. Ici encore, la lecture du Chapitre 5 nous semble profitable pour mieux appréhender le Chapitre 6.

L'algorithmique faisant occasionnellement son apparition dans les sujets des épreuves d'admissibilité, il peut être bon de lire attentivement la Section 3 du Chapitre 4. Ces algorithmes pourront être avantageusement mis à profit lors des épreuves orales d'admission.

Comme précédemment, le Chapitre 7 donnera de la perspective quant à la mise en œuvre de tests statistiques lors des épreuves d'admission. Il faudra néanmoins prendre garde au fait que le vocabulaire des tests est explicitement hors des programmes du secondaire et que certains d'entre eux font appel à des lois dépassant le cadre de ces programmes.

Comme pour l'agrégation interne, nous renvoyons le candidat aux Points numériques et aux Points programmes pour transposer au mieux ces savoirs disciplinaires lors des épreuves orales. L'exploitation par le candidat des notes historiques pour contextualiser son enseignement lui sera évidemment bénéfique.

Là encore, nous conseillons au candidat de travailler en parallèle sur des ouvrages de niveau secondaire pour s'assurer de sa capacité à mettre en œuvre les différentes notions dans des cas concrets simples.

CAPLP

Pour les épreuves d'admissibilité du CAPLP, le candidat devra connaître les méthodes générales du calcul des probabilités présentées au Chapitre 1, en particulier celles qui touchent aux probabilités conditionnelles. Concernant les variables aléatoires, il nous semble surtout nécessaire de bien maîtriser les résultats du Chapitre 2, sauf les lois du χ^2 et de Student.

Le Chapitre 3 ne doit pas être approfondi mais il pourra néanmoins aider à l'indispensable lecture du Chapitre 6 sur l'estimation par intervalles, sujet incontournable au CAPLP.

Le Chapitre 4 – plus précisément la Section 3 sur la simulation – pourra être avantageusement mis à profit lors des épreuves écrites et orales.

Pour les épreuves d'admission, il pourra être intéressant de se référer aux Chapitres 1 et 6 pour se constituer une bibliothèque de démonstrations à présenter. Les Points numériques et Points programmes aideront aussi le candidat à transposer ces notions dans son enseignement.

Il est recommandé au candidat de se munir d'ouvrages de lycée professionnel pour s'assurer de sa maîtrise technique dans des cas simples et pour préparer ses séquences d'enseignement dans l'optique de ces épreuves.

Là encore, des références historiques lors des épreuves orales seront nécessairement appréciées par le jury du CAPLP, particulièrement sensible à la pédagogie.

Remerciements

Cet ouvrage doit beaucoup aux enseignants et futurs enseignants que j'ai pu avoir en formation, je souhaite donc leur rendre hommage ici. Je tiens aussi à remercier toutes les personnes qui ont participé à la relecture du manuscrit à divers stades de son évolution : Sylvain Porret-Blanc, Frédéric Doyon, Pascal Selbonne, Louis Beaudet, Marie-Paule Thomas, Geneviève et Alain Dupont. Je souhaite aussi remercier ma conjointe qui, probablement sans le vouloir, a pris une part active dans l'élaboration de ce projet. Je tiens enfin à remercier les éditions Dunod pour leur confiance et leur professionnalisme.

Chapitre 1

Le calcul des probabilités

Ce chapitre présente les notions et résultats essentiels du calcul des probabilités. Conformément à la logique des nouveaux programmes, nous n'avons pas pour objectif de revenir en détail sur tous les éléments du calcul classique des probabilités mais plutôt de mettre en place les notions et méthodes utiles pour pouvoir développer ensuite des outils performants vis-à-vis de problèmes de simulation, d'estimation et de prise de décision en situation probabiliste.

Le présent chapitre se divise en cinq grandes parties. La Section 1 est dédiée à l'étude théorique des espaces probabilisés et des lois de probabilité. La Section 2 présente les résultats les plus généraux concernant les variables aléatoires, résultats qui seront approfondis dans la Section 3 dans le cas des variables aléatoires discrètes et dans la Section 4 dans celui des variables aléatoires à densité. Enfin, la Section 5 est dédiée à l'étude des couples de variables aléatoires discrètes ou à densité.

1 Calcul des probabilités

1.1 Espaces probabilisés

Dans son incarnation la plus élémentaire, le calcul des probabilités consiste à associer à un ensemble A d'issues d'une expérience aléatoire un nombre réel $P(A)$ compris entre 0 et 1, appelé **probabilité de l'événement** A. Pour calculer $P(A)$, les techniques historiques ont montré l'utilité de pouvoir l'écrire comme réunion ou intersection d'autres événements et de considérer leurs complémentaires. Ce sont précisément ces propriétés qui ont été axiomatisées sous le nom d'**espace probabilisé** par Kolmogorov à l'aide de la théorie des ensembles dans son traité *Grundbegriffe der Wahrscheinlichkeitsrechnung* publié en 1933. Ce travail de formalisation a été la clé de l'essor du calcul des probabilités qui reposaient jusqu'alors sur l'intuition du **principe fréquentiste**, trop flou pour développer une théorie mathématique consistante.

Dans l'idée générale, un espace probabilisable est un couple $(\Omega, \mathscr{A})$ où Ω est l'ensemble de toutes les issues possibles de l'expérience aléatoire considérée, appelé **univers**, et $\mathscr{A}$ est un ensemble d'événements auxquels on souhaite pouvoir associer une probabilité. Une loi de probabilité sera alors une application P qui à chaque événement $A \in \mathscr{A}$ associe un nombre $P(A) \in [0,1]$ satisfaisant à certains axiomes assurant sa compatibilité avec les opérations ensemblistes usuelles.

Andrei Kolmogorov (1903 – 1987)

Andreï Nikolaïevitch Kolmogorov, mathématicien russe né à Tambov en 1903 et décédé à Moscou en 1987. Il est le créateur de l'axiomatisation contemporaine de la théorie des probabilités à l'aide de la théorie des ensembles.

a) Espaces probabilisables, tribus

Point programme

Les notions générales de tribus et d'espaces probabilisables/probabilisés débordent du cadre des programmes du secondaire et leur analyse théorique sort aussi de l'esprit des épreuves du CAPES et de l'agrégation interne. En CPGE, la notion de tribu est introduite en classe de MP et il est bien précisé qu'elle « *n'appelle aucun développement théorique* ». Il nous apparaît donc important que le candidat aux concours de l'enseignement s'attache d'avantage à la manipulation transparente de ces notions qu'à un approfondissement de leurs propriétés théoriques générales.

Dans toute cette section Ω est un ensemble non-vide quelconque.

Définition 1.1.1 (Tribu, σ-algèbre)**.** Une **tribu**, ou **σ-algèbre**, sur Ω est une famille $\mathscr{A} \subset \mathscr{P}(\Omega)$ telle que :

(T1) $\Omega \in \mathscr{A}$;

(T2) Pour tout $A \in \mathscr{P}(\Omega)$, si $A \in \mathscr{A}$, alors $\overline{A} \in \mathscr{A}$ (où $\overline{A}$ désigne le complémentaire de A dans Ω) ;

(T3) Pour toute famille dénombrable [1] $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$, on a

$$\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}.$$

Exemple 1.1.2.

1. $\mathscr{P}(\Omega)$ est une tribu sur Ω, appelée **tribu discrète** sur Ω ;
2. Si $A \subset \Omega$, alors $\left\{\emptyset, A, \overline{A}, \Omega\right\}$ est une tribu sur Ω, appelée **tribu engendrée** par A ;
3. $\{\Omega, \emptyset\}$ est une tribu sur Ω, appelée **tribu triviale** sur Ω.

(1). On prendra garde au fait qu'il existe plusieurs conventions pour la dénombrabilité. Dans cet ouvrage, nous dirons qu'un ensemble E est **dénombrable** s'il existe une bijection de E vers $\mathbb{N}$ et qu'il est **au plus dénombrable** s'il est fini ou dénombrable.

Remarque

Il est facile de montrer que si $\mathscr{A}$ est une tribu sur Ω, alors $\mathscr{A}$ est stable par réunions et intersections au plus dénombrables.

Définition 1.1.3 (Espace probabilisable)**.** Un **espace probabilisable** est un couple $(\Omega, \mathscr{A})$ où :

- Ω est un ensemble,
- $\mathscr{A}$ est une tribu sur Ω.

Ω est alors appelé l'**univers** et les éléments de $\mathscr{A}$ sont appelées les **événements** de l'espace probabilisable.

Remarques

1. On remarquera que si $A \subset \Omega$ est un événement, alors il suit de l'axiome (T2) que $\overline{A}$ est aussi un événement, appelé **événement contraire** de A.
2. Il est utile de noter que ce n'est pas Ω qui est un espace probabilisable mais Ω *muni d'une certaine famille* $\mathscr{A}$ *de sous-ensembles* stable sous les opérations de passage au complémentaire et de réunion dénombrable. L'analogie peut être faite avec la topologie où un espace topologique est un couple $(E, \mathscr{T})$ où E est un ensemble et où la topologie $\mathscr{T}$ est un certaine famille de parties de E stable sous réunions arbitraires et intersections finies. De ce fait, l'espace probabilisable ainsi que les lois de probabilité et les variables aléatoires qui vont lui être associées peuvent fortement dépendre du choix de la tribu $\mathscr{A}$ sur Ω. De la même manière, un espace topologique $(E, \mathscr{T})$ a une structure topologique et des applications continues qui dépendent fortement du choix de la topologie $\mathscr{T}$ sur l'ensemble E.

Si Ω est un ensemble fini ou dénombrable, nous le munirons en général de la tribu $\mathscr{P}(\Omega)$. Cependant, si Ω est infini non dénombrable cette tribu est trop fine pour définir un calcul des probabilités satisfaisant. En particulier, si nous considérons l'exemple fondamental où $\Omega = \mathbb{R}^n$, avec $n \geq 1$, on préférera considérer une plus petite tribu $\mathscr{B}_{\mathbb{R}^n}$, appelée **tribu borélienne** sur $\mathbb{R}^n$, en référence à Émile Borel.

Émile Borel (1871 – 1956)

Émile Borel, mathématicien français, né à Saint-Affrique en 1871 et décédé à Paris en 1956. Il est l'un des initiateurs du calcul moderne des probabilités.

Théorème 1.1.4 (Tribu borélienne)**.** *Soit $n \geq 1$. Alors il existe une plus petite tribu (au sens de l'inclusion) sur $\mathbb{R}^n$ contenant les ouverts de $\mathbb{R}^n$.*

*Cette tribu est appelée **tribu borélienne** de $\mathbb{R}^n$ et ses éléments sont appelés les **boréliens** de $\mathbb{R}^n$.*

Démonstration. Considérons l'ensemble

$$\mathscr{O} = \{U \in \mathscr{P}(\mathbb{R}^n) \mid U \text{ est ouvert dans } \mathbb{R}^n\}$$

de tous les ouverts de $\mathbb{R}^n$ et posons

$$\mathscr{T}_{\mathscr{O}} = \{\mathscr{T} \in \mathscr{P}(\mathscr{P}(\mathbb{R}^n)) \mid \mathscr{T} \text{ est une tribu, } \mathscr{O} \subset T\}$$

l'ensemble de toutes les tribus contenant $\mathscr{O}$. C'est un ensemble non-vide car la tribu discrète contient $\mathscr{O}$, c'est-à-dire $\mathscr{P}(\mathbb{R}^n) \in \mathscr{T}_{\mathscr{O}}$.
Posons alors

$$\mathscr{B}_{\mathbb{R}^n} = \bigcap_{\mathscr{T} \in \mathscr{T}_{\mathscr{O}}} \mathscr{T}$$

l'intersection de toutes les tribus contenant $\mathscr{O}$. C'est une tribu comme intersection de tribus, elle contient $\mathscr{O}$ et c'est la plus petite par construction. □

Puisqu'une tribu est stable par passage au complémentaire, il est équivalent de dire que la tribu borélienne est la plus petite tribu contenant tous les fermés de $\mathbb{R}^n$. Dans le cas particulier où $n = 1$, la tribu borélienne $\mathscr{B}_{\mathbb{R}}$ contient tous les intervalles de $\mathbb{R}$. De manière générale, la tribu borélienne $\mathscr{B}_{\mathbb{R}^n}$ contient tous les pavés de $\mathbb{R}^n$. On peut de plus montrer que la tribu borélienne sur $\mathbb{R}^n$ (avec $n \geq 1$) est engendrée par les produits d'intervalles semi-ouverts de la forme $]a,b]$ avec $a < b$ réels. Bien que ce soit là une propriété essentielle pour démontrer un certain nombre de résultats théoriques fondamentaux en calcul des probabilités, cela entre plutôt dans le cadre de la théorie de la mesure et sort de celui de nos programmes. Nous renvoyons le lecteur intéressé par ces questions théoriques à [Sch93, Chapitre V.2].

Conventions

Dans tout l'ouvrage, sauf mention contraire explicite, on adoptera les conventions suivantes :

- Si $n \geq 1$, alors $\mathbb{R}^n$ est muni de sa tribu borélienne,
- Si Ω est un ensemble fini ou dénombrable, alors Ω est muni de la tribu $\mathscr{P}(\Omega)$.

Ainsi, nous voyons implicitement $\mathbb{N}^n$, $\mathbb{Z}^n$ ou $\mathbb{R}^n$ ($n \geq 1$) comme des espaces probabilisables.

1.2 Lois de probabilité

Point programme

L'étude générale des lois de probabilité n'est pas au programme du secondaire où l'on se limite à des exemples particuliers dès la classe de Première. Si le candidat au CAPLP peut éventuellement se contenter de la seule connaissance des lois classiques (voir Chapitre 2), il nous semble nécessaire que les candidats au CAPES et à l'agrégation interne soient en mesure de travailler dans le cadre axiomatique général.

Définition 1.1.5 (Loi de probabilité, espace probabilisé)**.** Une **loi de probabilité** sur un espace probabilisable $(\Omega, \mathscr{A})$ est une application

$$P : \left\{ \begin{array}{ccc} \mathscr{A} & \longrightarrow & \mathbb{R}_+ \\ A & \mapsto & P(A) \end{array} \right.$$

qui vérifie les deux axiomes suivants :

(P1) $P(\Omega) = 1$;

(P2) P est **σ-additive**, c'est à dire que pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}} \subset \mathscr{A}$ telle que $A_i \cap A_j = \emptyset$ si $i \neq j$, on a

$$P\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n).$$

Le triplet $(\Omega, \mathscr{A}, P)$ est appelé un **espace probabilisé**.

Exemple 1.1.6. L'exemple historique de loi de probabilité est celui de la loi définie, si Ω est fini, par

$$P : \left\{ \begin{array}{ccc} \mathscr{P}(\Omega) & \longrightarrow & \mathbb{R}_+ \\ A & \mapsto & \dfrac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}. \end{array} \right.$$

On l'appelle la **loi uniforme** sur l'ensemble fini Ω. Ce sont précisément les propriétés élémentaires de cette loi qui ont motivé la définition axiomatique proposée ci-dessus.

> **Remarque**
>
> Il n'aura sans doute pas échappé au lecteur que nous utilisons la notation $\sum_{n\in\mathbb{N}} P(A_n)$ sans préciser la façon dont une telle somme infinie est définie. De manière générale, nous considérerons régulièrement dans cet ouvrage des sommes indexées par des exemples qui ne sont pas nécessairement finis, ni même ordonnés. Le bon cadre théorique pour la définition et la manipulation de ces sommes infinies est celui des **familles sommables** de réels. Il s'agit là d'une notion d'analyse que nous ne développerons pas et nous renvoyons le lecteur intéressé à [Cho69, Chapitre VII.3] ou [Ouv07, Chapitre 2]. En première approche, le lecteur pourra ignorer ces questions de convergence en travaillant sur des univers finis et en se concentrant sur la signification des calculs effectués. On pourra aussi penser la sommabilité en termes de convergence absolue (et donc commutative) de séries.

Il suit facilement des axiomes que si P est une loi de probabilité alors, pour tous événements A et B, on a les propriétés bien connues suivantes, que nous utiliserons librement au cours de l'ouvrage :

1. $P(A) \in [0,1]$;
2. $P(\overline{A}) = 1 - P(A)$ et en particulier $P(\emptyset) = 0$;

3. Si $A \subset B$, alors $P(A) \leq P(B)$ et $P(B \setminus A) = P(B) - P(A)$;
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Le dernier point se généralise à toute famille finie d'événements à l'aide de la **formule de Poincaré**, aussi appelée **formule du crible** :

Proposition 1.1.7 (Formule de Poincaré)**.** *Soient* $(\Omega, \mathscr{A}, P)$ *un espace probabilisé,* $n \geq 1$ *un entier et* $A_1, \ldots, A_n$ *des événements. Alors*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}\left((-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right)\right).$$

Démonstration. La preuve fait l'objet de l'Exercice 1.1. □

Henri Poincaré (1854 – 1912)

Henri Poincaré, né à Nancy en 1854 et décédé à Paris en 1912. Il est l'un des mathématiciens français les plus influents de son époque et est considéré comme l'un des derniers génies universels qu'a connu la science : il était ingénieur des mines, physicien, astronome et mathématicien. Ses travaux portent sur des domaines très variés allant de l'analyse à la géométrie en passant par la topologie, la théorie des nombres et les probabilités. Ses travaux sont aussi importants vis à vis de la théorie de la relativité d'Einstein et du problème des trois corps en mécanique céleste.

Remarque

La formule de Poincaré, malgré son apparente complexité, traduit une idée intuitive très simple. Pour calculer la probabilité d'une réunion de n événements, il faut commencer par sommer les probabilités de tous les événements. Mais on aura alors compté en trop toutes les intersections deux à deux donc il faut les retrancher. Mais en retranchant les intersections deux à deux, on enlève peut-être des intersections trois à trois, donc il faut les remettre. Et ainsi de suite jusqu'à traiter les intersections n à n.

Nous verrons de nombreux exemples de lois de probabilité au Chapitre 2. Pour la suite de ce chapitre, le lecteur peu familier avec le calcul des probabilités pourra simplement garder en tête l'exemple de la loi uniforme sur un ensemble fini mentionnée à l'Exemple 1.1.6.

1.3 Indépendance

L'indépendance est une notion essentielle dans le calcul des probabilités. Sa définition axiomatique est faite pour refléter la notion intuitive d'indépendance causale d'événements aléatoires. Ainsi, si $(\Omega, \mathscr{A}, P)$ est un espace probabilisé et $A, B \in \mathscr{A}$, on dira que A et B sont

indépendants si et seulement si

$$P(A \cap B) = P(A)P(B).$$

Point programme
Dans les programmes du secondaire, la notion d'indépendance n'est formellement introduite qu'en classe de Terminale S. Elle n'apparaît pas dans les autres programmes de l'enseignement général et est explicitement hors-programme en classe de Terminale professionnelle.

Remarque

On notera que l'indépendance des événements A et B ne leur est pas intrinsèque mais qu'elle dépend de la loi de probabilité P. Il ne s'agit donc pas d'une notion ensembliste mais bien d'une notion probabiliste.

En particulier, on prendra soin de ne pas confondre le fait pour deux événements A et B d'être **incompatibles**, c'est-à-dire disjoints, et le fait pour ces événements d'être indépendants. On pourra d'ailleurs remarquer que si A et B sont incompatibles et $P(A) > 0$ et $P(B) > 0$, alors $0 = P(A \cap B) \neq P(A)P(B)$.

Si la notion d'indépendance pour deux événements est relativement familière de tous, la définition de l'indépendance pour une famille de plus de deux événements est déjà plus délicate et souvent source d'erreurs.

Définition 1.1.8 (Indépendance)**.** Soit $(\Omega, \mathscr{A}, P)$ un espace probabilisé et $(A_i)_{i \in I}$ une famille d'événements. La famille $(A_i)_{i \in I}$ est dite **indépendante** si, pour tout sous-ensemble fini $J \subset I$, on a

$$\boxed{P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j).}$$

Remarque

On prendra garde au fait que pour vérifier l'indépendance d'une famille $(A_1, \ldots, A_n)$, avec $n \geq 1$, il ne suffit pas de vérifier que $P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$ mais il faut bien vérifier $2^n - n - 1$ égalités correspondant à toutes les sous-familles finies de cardinal ≥ 2.

Il ne suffit pas non plus de vérifier que l'on a indépendance deux à deux pour vérifier l'indépendance globale. Par exemple, l'indépendance des trois événements A_1, A_2 et A_3

signifie que l'on a les trois égalités

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad 1 \leq i < j \leq 3$$

ainsi que

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

Exemple 1.1.9. Considérons une expérience aléatoire consistant à lancer deux fois une pièce équilibrée. Considérons les événements A_1 : « la pièce fait pile au premier lancer », A_2 : « la pièce fait face au second lancer » et A_3 : « on obtient le même résultat aux deux lancers ». On peut alors modéliser cette expérience avec l'espace probabilisé $(\Omega, \mathscr{P}(\Omega), P)$ où $\Omega = \{0,1\}^2$ (0 pour « pile » et 1 pour « face ») et P est la loi uniforme, de sorte que $A_1 = \{0\} \times \{0,1\}$, $A_2 = \{0,1\} \times \{1\}$ et $A_3 = \{(0,0),(1,1)\}$. Ainsi,

$$P(A_1 \cap A_2) = P(\{(0,1)\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2),$$

$$P(A_1 \cap A_3) = P(\{(0,0)\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_3),$$

$$P(A_2 \cap A_3) = P(\{(1,1)\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2)P(A_3)$$

mais

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq \frac{1}{8} = P(A_1)P(A_2)P(A_3).$$

Autrement dit, les événements A_1, A_2 et A_3 sont indépendants deux à deux mais ne sont pas pour autant indépendants (2).

Remarque

Un événement A tel que $P(A) = 1$ est dit **presque-sûr** et un événement de probabilité nulle est dit **improbable** ou **négligeable**. Dans la littérature, un événement de probabilité 1 est parfois dit **certain** et un événement de probabilité 0 est dit **impossible** mais cette terminologie prête à confusion si l'univers est infini.

En effet, considérons par exemple une pièce équilibrée que nous lançons indéfiniment et notons A l'événement « la pièce fait face à chaque lancer ». Si, pour tout entier $i \geq 1$, on note A_i : « la pièce fait face au i-ème lancer ». On a alors,

$$A = \bigcap_{i \in \mathbb{N}^*} A_i$$

(2). Certains auteurs précisent que les événements sont indépendants **dans leur ensemble** pour bien démarquer cette notion de l'indépendance deux à deux. Pour notre part, nous ne le préciserons jamais car la seule notion d'indépendance vraiment pertinente d'un point de vue théorique est celle introduite à la Définition 1.1.8.