GRÉGOIRE DUPONT

PROBABILITÉS ET STATISTIQUES POUR L'ENSEIGNEMENT

CAPES/CAPLP
AGRÉGATION INTERNE
MATHÉMATIQUES

NOUVEAUX PROGRAMMES
ALGORITHMES EN PYTHON

2e ÉDITION

DUNOD

Direction artistique : Élisabeth Hébert

Conception graphique : Hokus Pokus Création

11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-080690-4

Table des matières

Introduction

1. Le contexte épistémologique et didactique

1.1. Des probabilités classiques aux statistiques inférentielles

Apparu au XVIIe siècle avec des mathématiciens comme Pierre de Fermat (1601-1655), Blaise Pascal (1623-1662) ou Jacques 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 à titre posthume de l'*Ars Conjectandi* de Jacques Bernoulli introduisit l'idée que la notion de probabilité était liée à celle de fréquence de réalisation d'un événement. Ce **principe fréquentiste**, appelé **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 de la limite centrée**, aussi connu sous le nom de TCL. 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 : intelligence artificielle, *Big data*, 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 et 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 en 2009 par les programmes du secondaire.

1.2. Les programmes de 2009

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.* ».

La réforme de 2009, que ce soit pour la voie générale et technologique ou pour la voie professionnelle, a 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.

L'accent a alors été mis sur l'interprétation et l'exploitation de ces phénomènes d'échantillonnage pour la **prise de décision** en situation d'incertitude, dans la logique hypothéticodéductive des tests statistiques. L'approche démonstrative a dans ces programmes laissé une place importante à une approche plus expérimentale s'appuyant notamment sur l'usage d'outils numériques et une pédagogie par l'exemple.

Ces nombreux changements de paradigmes ont suscité d'importants questionnements auxquels la première édition de cet ouvrage a proposé un certain nombre de réponses et d'outils facilitant la compréhension et la transmission des notions mathématiques à l'œuvre dans ces programmes.

1.3. Les programmes de 2019

La réforme de 2019 a profondément restructuré l'organisation du lycée et les programmes d'enseignement. En mathématiques, la démonstration a été remise au cœur des apprentissages de la Seconde à la Terminale, à l'exception peut-être de l'option Mathématiques complémentaires qui poursuit des objectifs légèrement différents en s'inscrivant dans une logique de mathématiques appliquées à d'autres domaines (sociaux, économiques, médicaux, etc.).

En outre, le constat avait été fait d'une dénaturation des exercices de prise de décision qui se sont au fil des années transformés en exercices standardisés, routiniers et strictement calculatoires, aux antipodes des compétences visées par les programmes de 2009.

Forts de ces constats, les programmes de 2019 proposent de continuer à étudier les phénomènes de fluctuations en échantillonnage tout en inscrivant cette étude dans un cadre plus démonstratif et moins directement appliqué.

Là où les programmes de 2009 s'appuyaient sur le théorème de Moivre-Laplace et les fluctuations de la loi normale, deux notions très difficiles à enseigner formellement au niveau secondaire, ceux de 2019 privilégient une approche directe à travers l'inégalité de Tchebychev et les seules variables aléatoires discrètes, souvent moins efficace mais beaucoup plus facile à expliquer, voire à démontrer.

La notion d'intervalles de fluctuations a aussi disparu des programmes au profit de la seule notion d'intervalles de confiance et, dans les problèmes d'estimation, l'accent est mis sur l'*écart* entre la valeur à estimer et son estimation, plutôt que sur les intervalles eux-mêmes. Les exercices-types de prise de décision ne font plus partie des attendus des programmes. On observe ainsi que les programmes se recentrent vers l'objectif essentiel : *comprendre* les phénomènes de stabilisation (loi des grands nombres) et de fluctuations (inégalité de Tchebychev).

Une autre évolution importante opérée entre 2009 et 2019 est celle de la place de l'algorithmique et du numérique. Les nouveaux programmes viennent ainsi préciser un certain nombre de situations algorithmiques dont la mise en œuvre viendra éclairer les enseignements. Les outils privilégiés pour cela étant Python ou un tableur.

Enfin, ces programmes proposent d'intégrer des éléments d'histoire des mathématiques au fil des enseignements afin d'inscrire les notions étudiées dans leur contexte historique et culturel et de faciliter ainsi leur appropriation.

2. Les objectifs de l'ouvrage

2.1. Des contenus disciplinaires clairs et adaptés aux concours

Le premier objectif de cet ouvrage est de proposer un développement clair et aussi efficace que possible des notions de probabilité à l'œuvre dans les programmes des concours de l'enseignement secondaire. Car s'il existe déjà une littérature pléthorique sur le calcul classique des probabilités et sur le dénombrement, les choses se compliquent quand on s'intéresse aux questions de fluctuations et d'estimation. En effet, 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 les notions clés sous-tendant les programmes de l'enseignement 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 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. La plupart des énoncés sont prouvés de manière détaillée, parfois en se restreignant aux cas pertinents pour les concours visés. Et si les preuves trop techniques ou dépassant le cadre des programmes de concours ont été omises, des références sont systématiquement proposées au lecteur souhaitant approfondir le sujet.

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.

2.2. 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 ainsi que dans le programme des concours. L'ouvrage est donc régulièrement parsemé de **Points programmes**, indiqués par l'icône 🕮, 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°30 du 26 juillet 2008. Concernant les classes de Seconde et de spécialité de Première, il s'agit du BO spécial n°1 du 22 janvier 2019. Pour les programmes de Terminale, il s'agit du BO spécial n°8 du 25 juillet 2019.

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

2.3. Une 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**, indiqués par l'icône @, qui donnent des conseils techniques ou pédagogiques afin de mobiliser le numérique de manière pertinente lors d'une séance d'enseignement.

Cette nouvelle édition contient des implémentations en `Python` de tous les algorithmes nécessaires à l'enseignement des probabilités.

Pour ce qui concerne les autres logiciels, les commandes proposées ont été testées avec LibreOffice 5.1 et Geogebra Classique 6.

2.4. Des éléments historiques

De nombreux **Points historiques** parcourent l'ouvrage afin de faciliter la mise en perspective des notions enseignées. Ils sont indiqués par une icône . Ces points concernent la plupart du temps des biographies de mathématiciens, auquel cas une source d'inspiration a souvent été l'excellent *Des mathématiciens de A à Z* de Hauchecorne et Suratteau [HS96] ainsi que la très instructive collection *Génies des Mathématiques* éditée par RBA.

D'autres compléments historiques instructifs, comme les paradoxes de Bertrand, l'origine de la loi normale ou l'expérience de l'aiguille de Buffon font l'objet de sections spécifiques.

3. 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 **lois classiques** 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 les inégalités de Markov, Tchebychev et de concentration, la loi des grands nombres qui justifie le principe fréquentiste et le TCL qui permet de contrôler les fluctuations lors d'un échantillonnage.

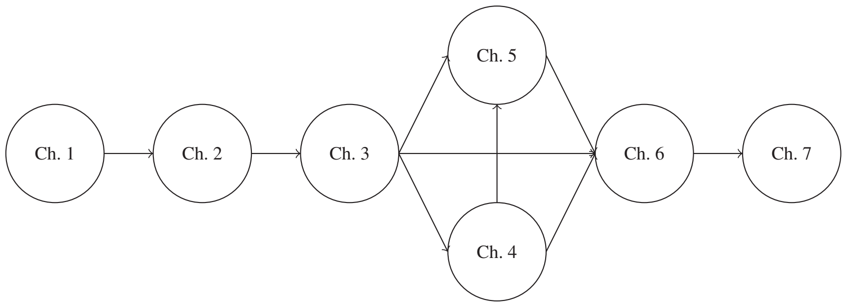
Le Chapitre 4 est dédié à l'étude de **méthodes de Monte-Carlo** permettant de simuler numériquement des variables aléatoires ou, plus généralement, de répondre à des problèmes déterministes à l'aide d'outils probabilistes. On y trouve aussi des algorithmes de **simulation** des lois classiques.

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 numériques pour les remplacer. 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.



4. Mode d'emploi à l'usage du futur enseignant

4.1. 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 concrètes les résultats vus dans cet ouvrage.

4.2. 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. Depuis 2019, les notions d'estimation traitées aux Chapitres 5 et 6 font aussi explicitement partie du programme.

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 aux Chapitre 5 et 6.

Le candidat pourra également regarder attentivement les Points numériques et les Points programmes pour mieux 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 le 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.

4.3. Pour le 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 régulièrement son apparition dans les sujets des épreuves d'admissibilité, il peut être bon de lire attentivement la Section 4.3 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.

4.4. Pour le 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.

Le Chapitre 4 – plus précisément la Section 4.3 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.

5. Principales modifications de la deuxième édition

Bien que les programmes aient évolué entre 2009 et 2019, les notions qui les sous-tendent restent fondamentalement les mêmes. Ce sont donc surtout les axes d'approfondissement des notions qui ont changé plutôt que les notions elles-mêmes. De ce fait, le fond et le plan général de l'ouvrage restent inchangés.

Néanmoins, le traitement de plusieurs notions, en particulier celles concernant les intervalles de fluctuations et de confiance au Chapitre 6, a été assez largement remanié pour le remettre en adéquation avec les nouveaux programmes. Le Chapitre 3 a aussi été complété de manière à être plus en phase avec la nouvelle approche des programmes fondée sur l'inégalité de concentration plutôt que le théorème de Moivre-Laplace.

Une autre modification importante concerne le Chapitre 4 et les points numériques qui sont maintenant adaptés à Python alors qu'ils l'étaient pour AlgoBox dans la précédente édition.

Nous avons aussi enrichi cette nouvelle édition avec de nouveaux résultats, exercices, exemples ou algorithmes qui rendent l'ouvrage plus complet.

6. Remerciements

Cet ouvrage doit beaucoup aux enseignants et futurs enseignants que j'ai pu avoir en formation. Je tiens aussi à remercier les personnes qui ont participé à la relecture du manuscrit : Sylvain Porret-Blanc, Frédéric Doyon, Pascal Selbonne, Louis Beaudet, Marie-Paule Thomas, Geneviève et Alain Dupont ainsi que Stéphane Bosquain et Alban Delmouly pour la deuxième édition. 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 rappelle les notions et résultats essentiels du calcul des probabilités. Notre objectif n'est pas de développer de manière trop détaillée cette théorie classique mais plutôt de mettre en place les notions et méthodes utiles pour répondre à des problèmes de modélisation, de simulation, d'estimation et de prise de décision en situation d'incertitude.

Le présent chapitre se divise en cinq grandes parties. La Section 1.1 est dédiée à l'étude théorique des espaces probabilisés et des lois de probabilité. La Section 1.2 présente les résultats les plus généraux concernant les variables aléatoires, lesquels seront approfondis à la Section 1.3 dans le cas des variables aléatoires discrètes et à la Section 1.4 dans celui des variables aléatoires à densité. Enfin, la Section 1.5 est dédiée à l'étude des couples de variables aléatoires réelles.

1.1 Calcul des probabilités

Le calcul des probabilités est apparu avec l'étude des jeux de hasard. On trouve trace de tels jeux dès l'Antiquité mais les premiers écrits mathématiques dont nous disposons à ce sujet datent de la fin du XVe avec les travaux de Luca Pacioli. Cette étude a ensuite été approfondie par des savants comme Cardan, Pascal, Fermat, Huygens et Bernoulli pour finir par constituer un corpus efficace de techniques de calcul.

Néanmoins, comme ce fut le cas dans de nombreux domaines des mathématiques au cours du XIXe, ce corpus fut ébranlé par l'apparition de paradoxes, les plus célèbres d'entre eux étant listés au début de l'ouvrage de Joseph Bertrand *Calcul des probabilités* publié en 1889. Ces paradoxes montrèrent la nécessité de fonder le calcul des probabilités sur des bases formelles solides. Ce travail intellectuel fut impulsé par David Hilbert dans sa célèbre conférence de 1900 où il énonça une liste de vingt-trois problèmes majeurs en mathématiques, parmi lesquels le sixième consistait à axiomatiser la physique, domaine auquel s'intégrait alors le calcul des probabilités.

La réponse à ce problème fut apportée en 1933 par Andrei Kolmogorov dans son ouvrage *Grundbegriffe der Wahrscheinlichkeitsrechnung* où il fonda le calcul des probabilités sur la théorie des ensembles à l'aide du concept d'espaces probabilisé que nous utilisons encore aujourd'hui.

Les travaux de Kolmogorov ont ainsi montré que les paradoxes de Bertrand n'étaient pas paradoxaux au sens mathématique du terme mais simplement, comme Bertrand en avait déjà eu l'intuition, des problèmes mal posés. Dit autrement, étant donné un choix de modèle pour une situation donnée, la théorie du calcul des probabilités est parfaitement définie et n'amène aucune contradiction. Mais le choix du modèle, lui, ne relève pas de la théorie mathématique et plusieurs choix peuvent être faits pour modéliser une même situation, amenant alors à des résultats contradictoires. Nous reviendrons plus en détail sur cette question à la Section 1.1.a.3.

Point programme

L'importance du choix du modèle est bien soulignée dans les programmes de Seconde : « *On insiste sur le fait qu'une loi de probabilité (par exemple une équiprobabilité) est une hypothèse du modèle choisi et ne se démontre pas. Le choix du modèle peut résulter d'hypothèses implicites d'équiprobabilité [...] qu'il est recommandable d'expliciter ; [...] Dans tous les cas, on distingue nettement le modèle probabiliste abstrait et la situation réelle.* »

Point historique: Andrei Kolmogorov (1903-1987)

Mathématicien soviétique né à Tambov en 1903 et décédé à Moscou en 1987. Ses travaux furent nombreux dans différents domaines de l'analyse et des probabilités. Il est en particulier reconnu pour avoir axiomatisé le calcul des probabilités à partir de la théorie des ensembles. Il a aussi donné naissance à la théorie Kolmogorov-Arnold-Moser (KAM) pour l'étude des systèmes dynamiques.

1.1.a. 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 $\mathbf{P}(A)$ compris entre 0 et 1, appelé **probabilité de l'événement** A. Pour calculer $\mathbf{P}(A)$, 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.

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 de parties de Ω, appelées **événements**, auxquelles on souhaite pouvoir associer une probabilité. Une loi de probabilité sera alors une application $\mathbf{P}$ qui à chaque événement $A \in \mathscr{A}$ associe un nombre $\mathbf{P}(A) \in [0,1]$ satisfaisant à certains axiomes assurant sa compatibilité avec les opérations ensemblistes usuelles.

1.1.a.1. 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 mais 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. On rappelle qu'un ensemble est dit **dénombrable** s'il est en bijection avec une partie de $\mathbf{N}$.

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 $(A_i)_{i\in I} \subset \mathscr{A}$, on a

$$\bigcup_{i\in I} A_i \in \mathscr{A}.$$

Exemple 1.1.2.

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

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

Point historique: Émile Borel (1871 - 1956)

Mathématicien français, né à Saint-Affrique en 1871 et décédé à Paris en 1956. Il est l'un des initiateurs de la théorie de la mesure et 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 $\mathbf{R}^n$ contenant les ouverts de $\mathbf{R}^n$.
Cette tribu est appelée **tribu borélienne** de $\mathbf{R}^n$ et ses éléments sont appelés les **boréliens** de $\mathbf{R}^n$.

Démonstration. Considérons l'ensemble

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

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

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

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 $\mathscr{P}(\mathbf{R}^n) \in \mathscr{T}_{\mathscr{O}}$.
Posons alors

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

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. $\square$

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 $\mathbf{R}^n$. Dans le cas particulier où $n = 1$, la tribu borélienne $\mathscr{B}_{\mathbf{R}}$ contient tous les intervalles de $\mathbf{R}$. De manière générale, la tribu borélienne $\mathscr{B}_{\mathbf{R}^n}$ contient tous les pavés de $\mathbf{R}^n$. On peut de plus montrer que la tribu borélienne sur $\mathbf{R}^n$ (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 $\mathbf{R}^n$ est muni de sa tribu borélienne,
- Si Ω est un ensemble dénombrable, alors Ω est muni de la tribu $\mathscr{P}(\Omega)$.

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

1.1.a.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

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

qui vérifie les deux axiomes suivants :

(P1) $\mathbf{P}(\Omega) = 1$;

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

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

Le triplet $(\Omega, \mathscr{A}, \mathbf{P})$ 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

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

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 $\sum_{n\in\mathbf{N}} \mathbf{P}(A_n)$ 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 ensembles 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 mais à laquelle nous faisons parfois référence pour gagner en concision, 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. Quand l'ensemble de sommation est dénombrable, on pourra aussi penser la sommabilité en termes de convergence absolue (et donc commutative) des séries correspondantes.

Il suit facilement des axiomes que si $\mathbf{P}$ 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. $\mathbf{P}(A) \in [0,1]$;
2. $\mathbf{P}\left(\overline{A}\right) = 1 - \mathbf{P}(A)$ et en particulier $\mathbf{P}(\emptyset) = 0$;
3. Si $A \subset B$, alors $\mathbf{P}(A) \leq \mathbf{P}(B)$ et $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$;
4. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ (**formule du crible**).

La formule du crible se généralise à toute famille finie d'événements à l'aide de la **formule de Poincaré** :

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

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

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

Point historique: Henri Poincaré (1854 - 1912)

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 les 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.

1.1.a.3. Complément historique : Les paradoxes de Bertrand

Nous revenons maintenant sur la remarque faite en introduction de ce chapitre sur l'importance du choix du modèle pour répondre à un problème concret. Dès les premières pages de son ouvrage *Calcul des probabilités* (1889), Joseph Bertrand met en garde contre les difficultés qui apparaissent avec la notion de hasard quand l'univers des possibles est infini. Le plus simple des paradoxes qu'il présente est le suivant :

« *On demande, par exemple, la probabilité pour qu'un nombre, entier ou fractionnaire, commensurable ou incommensurable, choisi au hasard entre 0 et 100, soit plus grand que 50. La réponse semble évidente : le nombre des cas favorables est la moitié de celui des cas possibles. La probabilité est* $\frac{1}{2}$.

Au lieu du nombre, cependant, on peut choisir son carré. Si le nombre est compris entre 50 et 100, le carré sera entre 2 500 et 10000.

La probabilité pour qu'un nombre choisi au hasard entre 0 et 10000 surpasse 2500 semble évidente : le nombre des cas favorables est les trois quarts du nombre des cas possibles. La probabilité est $\frac{3}{4}$.

Les deux problèmes sont identiques. D'où vient la différence des réponses ? Les énoncés manquent de précision. »

Comme l'a justement observé Bertrand, le problème initial est mal défini car le terme *au hasard* dans le cas d'un nombre infini d'issues ne renvoie à aucune notion mathématique précise. La différence des réponses vient ainsi de deux façons différentes de modéliser un même problème.

Dans sa première réponse, Bertrand propose de modéliser le problème à l'aide d'une variable aléatoire X, représentant le nombre choisi et suivant une loi uniforme sur $[0,100]$ (voir Section 2.6). Dans ce cas, la probabilité recherchée est

$$p_1 = \mathbf{P}(X \geq 50) = \int_{50}^{100} \frac{dt}{100} = \frac{1}{2}.$$

Dans sa seconde réponse, Bertrand propose de modéliser le problème à l'aide d'une variable aléatoire Y, représentant le carré du nombre choisi et suivant une loi uniforme sur $[0,10000]$. Dans ce cas, la probabilité recherchée est

$$p_2 = \mathbf{P}(Y \geq 2500) = \int_{2500}^{10000} \frac{dt}{10000} = \frac{3}{4}.$$

Les deux calculs de probabilité sont corrects mais correspondent à deux modèles différents. Ce que nous enseigne ce paradoxe est que le calcul des probabilités doit se limiter à des questions qui entrent dans son cadre formel, à savoir celui du calcul de probabilités d'événements précis dans un espace probabilisé donné, le reste relève de la *modélisation* et non de la théorie.

Remarque

Dans le paradoxe exposé ci-dessus, on pourrait néanmoins objecter que la première modélisation semble bien plus naturelle que l'autre et qu'il existe sûrement un moyen de montrer que $p = \frac{1}{2}$ est *la* bonne réponse au problème. Il est cependant possible de proposer d'autres paradoxes, notamment géométriques, pour lesquels aucun choix de modèle ne semble plus naturel qu'un autre. Nous invitons pour cela le lecteur à consulter l'ouvrage suscité de Bertrand, librement accessible sur le site de la Bibliothèque Nationale de France.

Point historique: Joseph Bertrand (1822-1900)

Mathématicien français, né en 1822 à Paris et décédé en 1900 dans cette même ville. Major d'agrégation, il a enseigné au lycée Saint-Louis, à l'École Polytechnique, à l'École Normale Supérieure et au Collège de France. Membre de l'Académie des Sciences à partir de 1856, il est élu en 1884 à l'Académie Française. Ses travaux portèrent principalement sur l'analyse mais il s'est aussi intéressé à l'histoire des sciences et à l'économie.

1.1.b. 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

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

Point programme

La notion d'indépendance est introduite dans le programme de spécialité de la classe de première.

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 $\mathbf{P}$. 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 $\mathbf{P}(A) > 0$, $\mathbf{P}(B) > 0$, alors $0 = \mathbf{P}(A \cap B) \neq \mathbf{P}(A)\mathbf{P}(B)$.

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 plus délicate et souvent source d'erreurs.

Définition 1.1.8: Indépendance

Soit $(\Omega, \mathscr{A}, \mathbf{P})$ 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

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

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 $\mathbf{P}(A_1 \cap \cdots \cap A_n) = \mathbf{P}(A_1) \cdots \mathbf{P}(A_n)$ 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