Cours Probabilités L2
Université Nice Sophia-Antipolis

François Delarue

Table des matières

CHAPITRE 1

Probabilités finies

Nous établissons dans ce premier chapitre les principes essentiels de la théorie des probabilités lorsque les expériences considérées admettent un nombre fini d'issues possibles. Nous parlerons de « probabilités finies ».

1. Description ensembliste d'une expérience aléatoire

On appelle expérience aléatoire tout phénomène dont les issues sont a priori incertaines : lancer d'une pièce à pile ou face, roulement d'un dé, tirage d'une carte parmi un paquet mélangé, instant de désintégration d'un atome instable ...

Une première approche consiste à décrire sous la forme d'un ensemble toutes les issues possibles de l'expérience étudiée.

Par exemple, les issues possibles d'un lancer d'une pièce à pile ou face sont modélisées par $\{0, 1\}$, celles d'un lancer de dé à 6 faces par $\{1, 2, 3, 4, 5, 6\}$ et ainsi de suite.

1.1. Vocabulaire ensembliste. La tradition retient comme notation générique pour l'ensemble des issues possibles d'une expérience aléatoire la lettre grecque Ω.

Nous comprenons que les parties de Ω, i.e. les sous-ensembles de Ω, jouent un rôle essentiel en pratique. Ils permettent de décrire précisément la réalisation d'événements :

DÉFINITION 1.1. *Etant donné un ensemble fini Ω, on appelle **événement** tout sous-ensemble de Ω. En particulier, l'ensemble des événements coïncide avec l'ensemble des parties de Ω, à savoir $\mathcal{P}(\Omega)$.*

Insistons : d'un point de vue pratique, un événement, au sens ensembliste du terme, décrit un événement, au sens usuel du terme, observable à l'issue de l'expérience. Par exemple, si l'expérience consiste à lancer un dé à six faces, l'événement “obtenir un chiffre pair” est représenté par l'ensemble (ou événement) $\{2, 4, 6\}$.

Les éléments de Ω sont désignés, de façon générique, par la lettre ω. En particulier, pour chaque ω dans Ω, le singleton $\{\omega\}$ est un événement : en pratique, il signifie que l'expérience étudiée a exactement ω pour issue. De façon plus générale, nous comprenons que le vocabulaire ensembliste usuel permet de décrire finement les observations

Observations	Terme ensembliste
Décrire un résultat possible	$\{\omega\},\ \omega \in \Omega$
Décrire un événement	$A, \quad A \subset \Omega$
Décrire l'implication d'un événement par un autre	$A \subset B$
Décrire la réalisation d'un événement **ou** d'un autre	$A \cup B$
Décrire la réalisation d'un événement **et** d'un autre	$A \cap B$
Décrire l'absence de réalisation d'un événement	$A^{\complement}$
Evénement impossible	$\emptyset$
Evénement certain	Ω
Incompatibilité de deux événements	$A \cap B = \emptyset$

1.2. Observations de plusieurs expériences. L'observation de plusieurs expériences ou d'expériences multiples peut toujours être comprise comme l'observation d'une seule et même expérience, familièrement décrite comme "grosse".

Donnons un exemple simple : le lancer de deux dés peut être compris comme une seule et même expérience ou comme la succession de deux expériences. D'un point de vue ensembliste, le rassemblement de deux expériences en une seule et même expérience est obtenu par multiplication cartésienne :

Définition 1.2. *Etant donnés deux ensembles Ω_1 et Ω_2, l'ensemble $\Omega = \Omega_1 \times \Omega_2$ est l'ensemble des couples $\{(\omega_1, \omega_2), \quad \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}$.*

En particulier, nous pouvons modéliser le lancer de deux dés par l'ensemble $\{1, \dots, 6\} \times \{1, \dots, 6\}$. Insistons : l'ordre des lancers est préservé par la structure de couple. La première coordonnée de chaque couple représente le premier lancer et la seconde le second lancer.

1.3. Règles de dénombrement. Nous verrons dans la suite que la théorie des probabilités s'appuie, pour partie, sur quelques propriétés essentielles de dénombrement. Nous rappelons ici quelques uns de ces fondamentaux.

La première des notions est celle de permutation : il s'agit de déterminer le nombre de façons de classer successivements n objects distincts ou tout simplement de classer (dans le désordre) les entiers naturels de 1 à n.

Proposition 1.3. *Le nombre de permutations de $\{1, \dots, n\}$, i.e. le nombre de n-uplets de la forme $(i_1, \dots, i_n)$ à coordonnées entre 1 et n et deux à deux distinctes est égal à $n!$.*

Preuve. Il s'agit d'un rappel.

Il existe un lemme de dénombrement bien pratique permettant de déduire les cardinaux de d'autres ensembles usuels.

LEMME 1.4. *(Lemme des bergers.) Etant donnés un entier $k \geq 1$, deux ensembles finis E et F et f une application de E dans F telle que*

$$\forall y \in F, \quad |\{x \in E : f(x) = y\}| = k,$$

alors $|E| = k|F|$. (Ici, $|\cdot|$ désigne le cardinal.)

Preuve. Admis. (Résultat par ailleurs tout-à-fait intuitif.)

Nous donnons maintenant deux exemples

PROPOSITION 1.5. *Le nombre de façons de choisir, **en tenant compte de l'ordre**, k éléments parmi n distincts est égal à $A_n^k = n!/(n-k)!$, $1 \leq k \leq n$. Précisément,*

$$\big|\big\{(i_1, \dots, i_k), \quad i_j \in \{1, \dots, n\},\ i_j \neq i_\ell \text{ pour } j \neq \ell\big\}\big| = \frac{n!}{(n-k)!}.$$

(Nous poserons, par convention, $A_n^0 = 1$.)

Preuve. La preuve est un peu formelle. Mais, il s'agit d'une belle application du lemme des bergers. Considérons

$$E = \big\{(i_1, \dots, i_n), \quad i_j \in \{1, \dots, n\},\ i_j \neq i_\ell \text{ pour } j \neq \ell\big\},$$
$$F = \big\{(i_1, \dots, i_k), \quad i_j \in \{1, \dots, n\},\ i_j \neq i_\ell \text{ pour } j \neq \ell\big\},$$

de même que

$$f : (i_1, \dots, i_n) \in E \mapsto (i_1, \dots, i_k) \in F.$$

Clairement, tout k-uplet dans F est atteint. (Autrement dit, f est une surjection). En fait, nous remarquons que, pour $(i_1, \dots, i_n)$ et $(i'_1, \dots, i'_n)$ dans E, leurs images sont égales si et seulement si les k premières coordonnées sont égales. Dit autrement, les antécédents de $(i_1, \dots, i_k)$ par f sont construits en complétant les coordonnées restantes à l'aide d'une permutation de $\{1, \dots, n\} \setminus \{i_1, \dots, i_k\}$. Mais, pour tout $(i_1, \dots, i_k)$, il y a $(n-k)!$ permutations de ce type possibles.

$\square$

Exactement sur le modèle, on obtient

PROPOSITION 1.6. *Le nombre de façons de choisir, **sans tenir compte de l'ordre**, k éléments parmi n distincts est égal à $C_n^k = n!/[k!(n-k)!]$, $1 \leq k \leq n$. Précisément,*

$$\big|\big\{\{i_1, \dots, i_k\}, \quad i_j \in \{1, \dots, n\},\ i_j \neq i_\ell \text{ pour } j \neq \ell\big\}\big| = \frac{n!}{k!(n-k)!}.$$

(Nous poserons, par convention, $C_n^0 = 1$.)

Insistons : C_n^k désigne le cardinal du nombre de parties à k éléments parmi un ensemble à n éléments. En particulier, la convention $C_n^0 = 1$ fait tout-à-fait sens.

Les coefficients $(C_n^k)_{0\leq k\leq n}$ sont aussi appelés coefficients binomiaux en raison de la formule du binôme de Newton :

Proposition 1.7. *Etant donnés deux réels a et b et un entier $n \geq 1$:*

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}.$$

En particulier,

$$\sum_{k=0}^{n} C_n^k = 2^n.$$

Il s'agit du nombre de parties d'un ensemble à n éléments.

2. Espace de probabilité fini

2.1. Mesure de probabilité. La théorie des probabilités vise à « mesurer » les événements observables à l'issue d'une expérience, i.e. à décrire la chance de les observer concrètement en pratique.

Par exemple, lorsque une pièce est lancée à pile ou face, la probabilité d'obtenir face est, si la pièce est équilibrée, 1/2 et, le cas échéant, celle d'obtenir pile est aussi 1/2. Dans ce cas précis, la mesure des événements consiste en un simple comptage. Mais, en réalité, mesurer peut s'avérer plus subtil que compter. Donnons encore un exemple : imaginons, qu'à l'issue du lancer successif de deux pièces, nous fassions la somme des résultats de chacune des pièces (avec la règle simple : 0 pour pile et 1 pour face). Alors, l'ensemble des issues possibles est $\{0, 1, 2\}$. Intuitivement, nous comprenons très bien que la probabilité d'avoir 1 est égale à 1/2, celles d'avoir 0 ou 2 valant 1/4. Ici, la mesure ne se résume par un simple comptage, mais plutôt à un *comptage avec pondération.*

Définition 2.1. *Etant donné un ensemble fini Ω, on appelle mesure de probabilité sur Ω, toute fonction $\mathbb{P}$ de $\mathcal{P}(\Omega)$ à valeurs dans $[0, 1]$ vérifiant :*

(1) $\mathbb{P}(\Omega) = 1$,

(2) *Pour $A, B \in \mathcal{P}(\Omega)$ t.q. $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.*

Le couple $(\Omega, \mathbb{P})$ s'appelle espace de probabilité (fini en l'espèce).

La propriété numéro 2 porte le nom d'*additivité*. Elle traduit une idée simple : la mesure d'un événement se calcule par la somme des mesures des singletons qu'il contient. Remarquons, en choisissant $A = \Omega$ et $B = \emptyset$, que $\mathbb{P}(\emptyset) = 0$.

Proposition 2.2. *Etant donné un espace de probabilité fini* $(\Omega, \mathbb{P})$*, pour tout événement* A*,*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}).$$

En particulier, la mesure de probabilité $\mathbb{P}$ *est entièrement définie par la famille des poids*

$$\big(p_\omega = \mathbb{P}\{\omega\}\big)_{\omega \in \Omega}.$$

Ces poids sont positifs et, par additivité, sont de somme égale à 1.

Réciproquement, toute famille $(q_\omega)_{\omega \in \Omega}$ *de poids positifs de somme égale à 1 définit une probabilité sur* Ω *par la formule :*

$$\mathbb{Q}(A) = \sum_{\omega \in A} q_\omega.$$

Toute famille finie de poids positifs et de somme égale à 1 est appelée loi de probabilité finie.

Preuve. A détailler en cours.

Soulignons que les lois de probabilité finies existent indépendamment du choix de Ω : ainsi $(1/2, 1/2)$ ou $(1/3, 1/2, 1/6)$. Elles définissent des mesures de probabilité dès lors que l'ensemble Ω est précisé : un choix typique (ou encore canonique) consiste à poser $\Omega = \{1, \ldots, N\}$ où N est égal au nombre de poids sous-jacents. En pratique, il est de fait d'usage de confondre les notions de mesure de probabilité et de loi de probabilité. Nous essaierons autant que possible de réserver le mot "mesure" lorsque l'ensemble Ω sera précisé et de parler de loi lorsque la famille de poids sera donnée sans information sur Ω.

2.2. Exemples.

Définition 2.3. *(Loi uniforme.) Etant donné un ensemble fini* Ω *de cardinal* N*, on appelle mesure de probabilité uniforme sur* Ω*, la mesure définie par*

$$\forall A \in \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \frac{|A|}{N}.$$

Elle est associée à la famille de poids $(p_\omega = 1/N)_{\omega \in \Omega}$*. La loi* $(1/N, \ldots, 1/N)$ *est appelée loi uniforme sur un ensemble à* N *éléments (ou, de façon équivalente, sur* $\{1, \ldots, N\}$*).*

Définition 2.4. *(Loi binomiale.) Etant donnés un entier* $N \geq 1$ *et un réel* $0 < p < 1$*, on appelle loi de probabilité binomiale de paramètres* N *et* p *(sous-entendu sur* $\{0, \ldots, N\}$*), la famille de poids*

$$p_k = C_n^k p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

Elle définit de façon canonique une mesure de probabilité sur $\{0, \ldots, N\}$*.*

2.3. Formules de calcul. L'axiome d'additivité permet de montrer

Proposition 2.5. *Etant donnés deux événements* A, B *d'un espace de probabilité fini* $(\Omega, \mathbb{P})$,

$$\mathbb{P}\big(A^{\complement}\big) = 1 - \mathbb{P}(A),$$
$$\mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(A \setminus B), \quad B \subset A,$$
$$\mathbb{P}\big(A \cup B\big) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Preuve. La première formule est évidente. La deuxième découle de l'additivité. Pour la troisième, il suffit d'écrire $A \cup B$ comme la réunion disjointe de A et de $B \setminus (A \cap B)$ et d'appliquer la deuxième formule.

□

Le résultat se généralise au cas de n événements

Proposition 2.6. *Etant donnés* n *événements* $A_1, \ldots, A_n$ *d'un espace de probabilité fini* $(\Omega, \mathbb{P})$,

$$\mathbb{P}\bigg(\bigcup_{i=1}^{n} A_i\bigg) = \sum_{\ell=1}^{n} (-1)^{n-\ell} \sum_{1 \leq i_1 < \cdots < i_\ell \leq n} \mathbb{P}\bigg(\bigcap_{j=1}^{\ell} A_{i_j}\bigg).$$

Preuve. Voir en TD.

3. Variables aléatoires

3.1. Expériences multiples. En présence d'expériences multiples (telles le lancer successif d'un dé), il peut-être pertinent de se focaliser sur une expérience précisément. Dit autrement, lorsque l'expérience considérée s'attache à l'étude de plusieurs traits, il peut être pertinent de ne s'intéresser qu'à un seul d'entre eux.

Définition 3.1. *Etant donné un ensemble* Ω *fini, on appelle variable aléatoire toute application de* Ω *dans* $\mathbb{R}$.

Il peut paraître un peu superflu de rebaptiser une notion déjà existante en mathématique. En réalité, le nom cherche ici à traduire l'essence même de l'objet étudié : il s'agit de se focaliser sur un trait aléatoire particulier. Par exemple, si Ω désigne l'ensemble de la population vivant en France, une variable aléatoire peut être la taille des individus ou leur poids ou encore leur âge.

Un exemple plus concret est donné par le modèle de l'expérience multiple :

Exemple 3.2. *Etant donné l'ensemble* $\{1, \ldots, 6\}^N$ *décrivant les résultats de* N *lancers successifs d'un dé à six faces, l'application* X_i, *pour* $1 \leq i \leq N$,

$$X_i : (x_1, \ldots, x_N) \in \{1, \ldots, 6\}^N \mapsto x_i,$$

est une variable aléatoire décrivant les résultats du i*ème lancer.*

3.2. Image réciproque. En pratique, les valeurs prises par une variable aléatoire sont essentielles. Il s'agit des résultats possibles de l'expérience étudiée.

D'un point de vue de la modélisation, ces valeurs sont les images par la variable aléatoire des aléas de l'espace de probabilité initiale, i.e. ces valeurs correspondent à l'ensemble $\{X(\omega),\ \omega \in \Omega\}$. Il est aisément compréhensible que, pour une observation donnée, l'expérimentateur cherche à décrire, au moins formellement, les ω susceptibles d'expliquer l'observation.

Précisément, nous considérerons souvent, dans la suite, $\{\omega \in \Omega : X(\omega) = x\}$ pour un x réel donnée. Pour simplifier les notations, nous écrirons la plupart du temps $\{X = x\}$ ou encore $X^{-1}(\{k\})$ (au sens de l'image réciproque de la théorie des ensembles). Plus généralement, nous préférerons écrire $\{X \in A\}$ ou $X^{-1}(A)$ au lieu de $\{\omega \in \Omega : X(\omega) \in A\}$, pour $A \subset \mathbb{R}$.

3.3. Loi d'une variable aléatoire. Une variable aléatoire décrit, par nature, une expérience dont les résultats sont aléatoires. Il s'agit, comme dans le cas d'un espace de probabilité, de mesurer la façon dont le hasard se répartit à l'arrivée.

Proposition 3.3. *Etant donnés un espace de probabilité fini* $(\Omega, \mathbb{P})$ *de cardinal* N *et une variable aléatoire* X *à valeurs dans* $\{x_1, \dots, x_P\} \subset \mathbb{R}$, P *étant nécessairement plus petit que* N, *on appelle loi de* X *les poids*

$$\big(\mathbb{P}\{X = x_k\}\big)_{1 \leq k \leq P}.$$

De façon canonique, il s'agit d'une loi de probabilité sur $\{x_1, \dots, x_P\}$.

Preuve. Les poids sont clairement positifs. Il reste à en faire la somme. Nous écrivons

$$\sum_{k=1}^{P} \mathbb{P}\{X = x_k\} = \mathbb{P}\Big(\bigcup_{k=1}^{P} \{X = x_k\}\Big).$$

En effet, la variable X ne peut pas prendre deux valeurs différentes au même point. Autrement dit, les événements considérés sont deux à deux disjoints. Maintenant, nous remarquons que

$$\bigcup_{k=1}^{P} \{X = x_k\} = \bigcup_{k=1}^{P} X^{-1}\big(\{X = x_k\}\big) = X^{-1}\left(\bigcup_{k=1}^{P} \{X = x_k\}\right) = \Omega.$$

$\square$

Exemple 3.4. *(Loi binomiale) Les lancers successifs de* N *pièces équilibrées sont modélisés par l'espace* $\{0,1\}^N$ *muni de la probabilité uniforme, notée* $\mathbb{P}$. *Le nombre de succès est modélisé par la variable aléatoire*

$$S : (\omega_1, \dots, \omega_N) \in \{0,1\}^N \mapsto \sum_{i=1}^{N} \omega_i.$$

La loi de S est la loi binomiale de paramètre $1/2$ sur $\{0,\ldots,N\}$, i.e.

$$\forall k \in \{0,\ldots,n\}, \quad \mathbb{P}\{S=k\} = 2^{-N} C_N^k.$$

Preuve. Il suffit de montrer que le cardinal de $\{S = k\}$ est C_N^k. Mais, l'ensemble $\{S=k\}$ est en bijection avec l'ensemble des parties à k éléments de $\{1,\ldots,N\}$: à chaque suite de k succès correspond une et une seule partie, donnée par les numéros des succès.

$\square$

Remarquons que le résultat est généralisable dans le cas d'une pièce déséquilibrée. Le cas échéant, la loi de S est la loi binomiale de paramètres N et p.

Exemple 3.5. *(Loi hypergéométrique - loi des sondages.) Dans une population de N individus, n apprécient l'action publique d'une personnalité politique A et $N-n$ la désapprouvent. Une fraction de p individus de cette population est interrogée, au hasard.*

Ce tirage au sort est modélisé par l'espace de probabilité fini

$$\Omega = \mathcal{P}_p(\{1,\ldots,N\}),$$

muni de la probabilité uniforme. (Ici, $\mathcal{P}_p$ désigne l'ensemble des parties à p éléments.) Par convention de modélisation, les soutiens de A sont représentés par des entiers entre 1 et n et ses opposants par des entiers entre $n+1$ et N.

Le nombre de soutiens parmi l'échantillon tiré est modélisé par la variable aléatoire

$$H : \omega \in \mathcal{P}_p(\{1,\ldots,N\}) \mapsto |\omega \cap \{1,\ldots,n\}|.$$

Sa loi est donnée par

$$\mathbb{P}\{H=k\} = \frac{C_n^k C_{N-n}^{p-k}}{C_N^p}, \quad \max(n-N+p,0) \le k \le \min(n,p).$$

Preuve. Nous commençons par remarquer que H prend ses valeurs dans $\{\max(0,N-(n-p)),\min(p,n)\}$. Il y a en effet au plus $\min(n,p)$ personnes favorables à A dans les personnes interrogées et au plus $\min(N-n,p)$ défavorables. Autrement dit , H est toujours inférieure à $\min(n,p)$ et supérieure à $p-\min(N-n,p) = \max(n-N+p,0)$.

Maintenant, pour $\max(n-N+p,0) \le k \le \min(n,p)$, l'ensemble $\{H = k\}$ s'écrit comme les parties de $\{1,\ldots,N\}$ ayant exactement k éléments dans $\{1,\ldots,n\}$ et $p-k$ dans $\{n+1,\ldots,N\}$. Il est donc en bijection avec $\mathcal{P}_k(\{1,\ldots,n\}) \times \mathcal{P}_{p-k}(\{n+1,\ldots,N\})$. De fait, son cardinal est $C_n^k \times C_{N-n}^{p-k}$. Comme le cardinal de Ω est C_N^p, on en déduit le résultat.

$\square$

Feuille de TD Numéro 1

1. Dénombrement

Dans chacun des exercices qui suivent, on essaiera d'expliquer la démarche de dénombrement de façon aussi rigoureuse que possible.

Exercice 1.1. Etant donnés deux entiers n et k positifs, $k \leq n-1$, montrer la relation du triangle de Pascal (on essaiera de donner deux preuves différentes) :

$$C_n^k + C_n^{k+1} = C_{n+1}^{k+1}.$$

Que peut-on dire si $k = n$?

Pour $k \leq n$, montrer la relation d'absorption

$$kC_n^k = nC_{n-1}^{k-1}.$$

En déduire la valeur de

$$\sum_{k=0}^{n} kC_n^k.$$

Exercice 1.2. Etant donnés deux entiers $n \geq 1$ et $0 \leq k \leq n$, vérifier que $C_n^k = C_n^{n-k}$. En déduire, par un raisonnement de pur dénombrement, que

$$\sum_{k=0}^{n} (C_n^k)^2 = C_{2n}^n.$$

Exercice 1.3. Soient a et b deux réels et n un entier supérieur à 2. En écrivant (par exemple par récurrence)

$$(a+b)^n = \sum_{(i_1,\dots,i_n)\in\{0,1\}^n} \prod_{j=1}^{n} \varepsilon_{i_j}, \quad \varepsilon_0 = a,\ \varepsilon_1 = b,$$

retrouver le binôme de Newton.

Exercice 1.4. Les nouvelles plaques minéralogiques comptent exactement 4 lettres et 3 chiffres. En comptant les éventuelles combinaisons exclues, donner le nombre de plaques possibles.

Comparer ce résultat au système des anciennes plaques, comptant 2 lettres et 6 chiffres.

Exercice 1.5. Combien de successions de lettres peut-on former à partir du mot « REVER » ?

Exercice 1.6. L'équipe de football d'un grand club européen compte (par ordre alphbétique) 1 Brésilien, 2 Espagnols, 2 Français, 4 Italiens et 2 Ivoiriens. En supposant que la feuille de match, numérotée de 1 à 11, ne mentionne que les nationalités des joueurs, donner le nombre de feuilles de match possibles.

Exercice 1.7. De combien de façon peut-on ranger p boules (indiscernables) dans n cases numérotées ? (Avec $p \leq n$.)

2. Calcul de probabilités

Dans chacun des exercices qui suivent, on prendra un grand soin à modéliser l'expérience considérée à l'aide d'un espace de probabilité, dont on précisera la mesure de probabilité.

Exercice 2.1. Quelle est la probabilité qu'en jetant six dés équilibrés et discernables (par exemple par la couleur), toutes les faces exhibent un chiffre différent ?

Exercice 2.2. Quelles sont les probabilités que, parmi les familles de n enfants, $n \geq 2$, une famille soit constituée d'enfants des deux sexes (événement A), puis des garçons et d'au plus une fille (événement B). Calculer la probabilité de $A \cap B$.

Exercice 2.3. Une urne contient M jetons numérotés de 1 à M. On tire successivement n en jetons en remettant chaque fois le jeton tiré et en brassant bien. Donner la probabilité qu'aucun jeton ne soit tiré plus d'une fois.

Application : un groupe de n étudiants sont réunis dans une même salle. Quelle est la probabilité qu'au moins deux étudiants aient leur anniversaire le même jour. (On suppose qu'aucun n'est né le 29 février et que n est inférieur à 365.)

Exercice 2.4. Des tickets au nombre de M sont édités et numérotés de 1 à M. Pour simplifier, on suppose que les n premiers (avec $2n \leq M$) sont gagnants. (Naturellement, les acheteurs ne le savent pas.) Quelle est la probabilité qu'un acheteur de n billets achète au moins un billet gagnant ?

Exercice 2.5. Etant donné un espace de probabilité fini $(\Omega, \mathbb{P})$, montrer que $\mathbb{P}(A) \leq \mathbb{P}(B)$ si A et B sont deux événements tels que $A \subset B$.

Exercice 2.6. Etant donné un espace de probabilité fini $(\Omega, \mathbb{P})$, montrer pour toute famille $A_1, \ldots, A_n$ de n événements que

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Exercice 2.7. Démontrer la formule de Poincaré : étant donnés n événements $A_1, \ldots, A_n$ d'un espace de probabilité fini $(\Omega, \mathbb{P})$,

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{\ell=1}^{n} (-1)^{n-\ell} \sum_{1 \leq i_1 < \cdots < i_\ell \leq n} \mathbb{P}\Big(\bigcap_{j=1}^{\ell} A_{i_j}\Big).$$

Application : un facteur distrait distribue le courrier au hasard dans une rue comptant n numéros. En supposant qu'il dépose exactement une lettre par habitation et que, sur les n distribuées au total, une et une seule soit destinée à une habitation donnée, quelle est la probabilité que personne ne reçoive de lettre lui ayant été explicitement adressée ?

3. Loi de variables aléatoires

Exercice 3.1. On lance deux dés à 6 faces. Donner la loi de la somme des deux résultats obtenus.

Exercice 3.2. Etant donnée une variable aléatoire X de loi uniforme sur $\{1, ..., N\}$ (définie sur un espace de probabilité fini $(\Omega, \mathbb{P})$), donner les lois de $N + 1 - X$ et de X/N.

On suppose $N = 2n$. On construit alors la variable aléatoire

$$\forall \omega \in \Omega,\ Y(\omega) = \lfloor \frac{X(\omega)}{2} \rfloor,$$

où $\lfloor \cdot \rfloor$ désigne la partie entière. Donner la loi de Y.

Exercice 3.3. Etant donnés un espace de probabilité fini $(\Omega, \mathbb{P})$ et un événement A, donner la loi de la variable aléatoire indicatrice de A

$$\mathbf{1}_A : \omega \in \Omega \mapsto \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Exercice 3.4. On procède à N lancers successifs d'une pièce déséquilibrée de paramètre de succès $0 < p < 1$. Donner la loi du nombre de succès.

Exercice 3.5. Une urne contient n boules rouges et $N - n$ noires. On tire, sans remise, p boules parmi les N. Donner la loi du nombre de boules rouges tirées.

Exercice 3.6. On rappelle qu'une variable aléatoire H est dite de loi hypergéométrique de paramètres N, n et r si

$$\mathbb{P}\{H = k\} = \frac{C_n^k C_{N-n}^{r-k}}{C_N^r}, \quad \max(n - N + r, 0) \leq k \leq \min(n, r).$$

Que dire de la limite de $\mathbb{P}\{H = k\}$ lorsque N tend vers l'infini et n/N tend vers $p \in]0, 1[$? (Le paramètre r étant fixé.)

Exercice 3.7. En considérant l'espace produit $\{0, 1\}^N$, donner une modélisation du lancer de N pièces déséquilibrées de paramètre $p \in (0, 1)$. Donner alors la loi de la variable aléatoire :

$$T : \omega = (\omega_1, \ldots, \omega_N) \in \{0, 1\}^N \mapsto \begin{cases} k & \text{si } \omega_1 = \cdots = \omega_{k-1} = 0 \text{ et } \omega_k = 1, \\ N + 1 & \text{si } \omega_1 = \cdots = \omega_N = 0. \end{cases}$$

Que représente T ?

CHAPITRE 2

Probabilités discrètes

Nous cherchons ici à étendre les principes de la théorie des probabilités, énoncés jusqu'ici dans le cadre d'un espace de probabilité fini, au cas d'un espace de probabilité discret.

D'un point de vue abstrait, il s'agit de modéliser des expériences dont les issues sont a priori en quantité dénombrable. D'un point de vue concret, il peut s'agir de modéliser des expériences dont les issues sont en quantité finie mais très grande. En pratique, il peut en effet être pertinent de donner une description asymptotique, c'est-à-dire obtenue par passage à la limite, d'un phénomène physique gouverné par une variable d'état à très grandes valeurs.

Donnons un exemple :

Une façon de décrire simplement le trafic autoroutier à une caisse de péage consiste à tirer, toutes les 10 secondes (ou éventuellement toutes les secondes pour obtenir une description plus fine), une pièce déséquilibrée de petit paramètre p, l'obtention d'un 1 traduisant l'arrivée d'une voiture. (Naturellement, aucune voiture n'arrive si jamais un 0 est tiré.)

Supposons, qu'à des fins commerciales (tarification du péage), nous cherchons à étudier le nombre total d'arrivées sur une plage horaire de 10 heures. (Par exemple, la plage 08h00-18h00).

Si le pas de temps retenu est de 10 secondes, il ne peut pas arriver plus de 360 voitures par heure ou encore plus de 3600 voitures sur une plage horaire de 10 heures. Si le pas de temps retenu est de 1 seconde (ce qui suppose une modification du paramètre p), le modèle accèpte l'arrivée éventuelle de 36000 voitures par plage de 10 heures. Bien que la probabilité d'un tel événement soit certainement très proche de 0, la modélisation suppose de choisir Ω comme un ensemble de cardinal 36000.

De façon plus générale, nous comprenons que le cardinal de Ω est amené à exploser au fur et à mesure de la décroissance du pas de temps (dixième de seconde, centième de seconde ...) : le cas asymptotique est à modéliser par un ensemble de cardinal infini. Il correspond au cas idéal où le comptage des voitures s'effectue en temps continu.

1. Mesure d'événements de cardinal infini

1.1. Ensembles (au plus) dénombrables. Dans le Chapitre 1, nous avons choisi de représenter, lorsqu'en quantité finie, les issues possibles d'une expérience par les éléments d'un ensemble (ou d'un *univers*). Ce principe demeure vrai lorsque l'expérience considérée admet une quantité dénombrable d'issues. Nous commençons de fait par rappeler :

Définition 1.1. *On appelle ensemble dénombrable un ensemble en bijection avec $\mathbb{N}$, i.e. dont les éléments peuvent être numérotés par les entiers naturels. Plus généralement, on appelle ensemble au plus dénombrable un ensemble fini ou dénombrable.*

Les exemples les plus typiques d'ensemble dénombrable sont : $\mathbb{Z}$, $\mathbb{N}^2$, $\mathbb{Q}$. Plus généralement, tout produit cartésien (fini) d'ensembles dénombrables est dénombrable. En revanche, $\mathbb{R}$ n'est pas dénombrable. (Au passage, rappelons, qu'à travers le développement décimal, $\mathbb{R}$ peut se penser comme le produit infini $\{0, \ldots, 9\}^{\mathbb{N}}$: un produit cartésien (infini) d'ensembles finis n'est pas dénombrable.)

Donnons deux exemples :

Exemple 1.2. *Imaginons que nous cherchions à modéliser les lancers d'une infinité de pièces à pile ou face. La modélisation proposée, au Chapitre 1, dans le cas de deux lancers, suggère de considérer, comme ensemble de représentation, le produit cartésien infini $\{0, 1\}^{\mathbb{N}}$, i.e. l'ensemble des suites dont les éléments sont soit 0 soit 1. Au regard de la discussion précédente, il ressort que l'ensemble proposé n'est pas dénombrable.*

En revanche, si nous nous attardons plus spécifiquement sur l'instant (i.e. le numéro) d'apparition du premier 1 en oubliant l'ensemble des lancers lui donnant suite, nous comprenons que l'ensemble à même de modéliser cette expérience est $\mathbb{N}^*$.

1.2. Mesure sur un ensemble au plus dénombrable. Etant donné un ensemble au plus dénombrable Ω, nous conservons le vocabulaire adopté dans le premier chapitre : les sous-ensembles de Ω sont appelés *événements*.

Il s'agit, dans la suite, de mesurer ces événements. Malheureusement, la définition de la mesure donnée dans le Chapitre 1 ne saurait suffire dans le cas des espaces dénombrables. Ceci est aisément compréhensible : lorsque Ω est fini, les événements sont eux-mêmes en nombre fini, de sorte que la propriété d'additivité porte, tout au plus, sur des réunions de taille finie. Lorsque Ω est dénombrable, $\mathcal{P}(\Omega)$ est lui-même infini (il est même non dénombrable car en bijection avec $\{0, 1\}^{\mathbb{N}}$) : la question se pose de fait de l'éventuelle extension de la propriété d'additivité à des réunions non finies d'événements.

La résolution du problème de l'additivité se résume à la résolution d'un dilemme : à ne passer assez imposer sur la façon de mesurer les événements, le risque est de ne pouvoir guère mesurer plus d'événements que dans le cas d'un univers fini ; à trop demander, le risque est à l'inverse de ne pas trouver de mesure à même de satisfaire les conditions exigées.

Il s'agit, pour trouver le bon équilibre, de s'interroger sur la finalité de la modélisation : l'introduction d'un univers dénombrable permet, comme nous l'avons déjà dit, d'assimiler (ou d'approcher) des populations de très grande taille par des populations de taille dénombrable. De même, l'observation d'un nombre infini d'événements peut-elle être comprise, d'un point de vue pratique, comme l'observation d'un très grand nombre d'événements : l'infinité est obtenue par passage à la limite le long d'événements en nombre fini. Pour être mené à bien, un passage à la limite nécessite une forme de continuité :

Définition 1.3. *Etant donné un ensemble dénombrable Ω, on appelle mesure de probabilité sur Ω, toute fonction $\mathbb{P}$ de $\mathcal{P}(\Omega)$ à valeurs dans $[0,1]$ vérifiant :*

(1) $\mathbb{P}(\Omega) = 1$,

(2) *Pour* $A, B \in \mathcal{P}(\Omega)$ *t.q.* $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,

(3) *Pour toute suite* $(A_n)_{n\geq 1}$ *d'événements deux à deux disjoints*

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \lim_{n\to+\infty} \mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big).$$

Le couple $(\Omega, \mathbb{P})$ *s'appelle espace de probabilité (dénombrable en l'espèce).*

Naturellement, nous remarquons que

$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathbb{P}(A_k).$$

Le passage à la limite dans (3) suggère de fait de développer la probabilité d'une réunion dénombrables d'événements deux-à-deux disjoints sous la forme d'une série.

1.3. Quelques rappels sur les séries à termes positifs. Nous rappelons quelques propriétés essentielles des séries à termes positifs :

Proposition 1.4. *Soit une suite de réels positifs* $(a_n)_{n\geq 0}$. *Alors, deux cas et deux seulement sont possibles : soit la série des* $(a_n)_{n\geq 0}$ *est convergente et* $\sum_{n\geq 0} a_n = \lim_{n\to+\infty} \sum_{k=0}^{n} a_k$, *soit la série des* $(a_n)_{n\geq 0}$ *est divergente et* $\lim_{n\to+\infty} \sum_{k=0}^{n} a_k = \infty$. *Cette dichotomie est résumée de la façon suivante : soit* $\sum_{n\geq 0} a_n < +\infty$, *soit* $\sum_{n\geq 0} a_n = +\infty$.

Par exemple, s'il existe une constante C vérifiant

$$\forall n \geq 0, \quad \sum_{k=0}^{n} a_k \leq C,$$

alors la série des $(a_n)_{n\geq 0}$ est convergente et

$$\lim_{n\to+\infty} \sum_{k=0}^{n} a_k = \sum_{n\geq 0} a_n.$$

En particulier, étant donnée une autre suite de réels positifs $(b_n)_{n\geq 0}$ telle que

$$\forall n \geq 0, \quad a_n \leq b_n,$$

la convergence de la série des $(b_n)_{n\geq 0}$ implique la convergence de la série des $(a_n)_{n\geq 0}$.

Par contraposition, la divergence de la série des $(a_n)_{n\geq 0}$ implique la divergence de la série des $(b_n)_{n\geq 0}$.

1.4. Définitions équivalentes d'une mesure de probabilité. La théorie des séries nous permet maintenant d'affirmer que

PROPOSITION 1.5. *Une fois supposés vérifiés les axiomes* (1) *et* (2) *de la définition d'une mesure, l'axiome* (3) *est équivalent à*

(3′) *Pour toute suite $(A_n)_{n\geq 1}$ d'événements deux à deux disjoints*

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n\geq 1} \mathbb{P}(A_n).$$

(Remarquons que le membre de gauche est automatiquement majoré par 1 : ceci garantit la convergence du membre de droite)

(3″) *Pour toute suite* **croissante** *d'événements $(A_n)_{n\geq 1}$, i.e. $A_n \subset A_{n+1}$, $n \geq 1$,*

$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \lim_{n\to+\infty} \mathbb{P}(A_n).$$

La propriété (3′) *porte le nom de σ-additivité. La propriété* (3″) *porte le nom de continuité de la mesure par réunion croissante.*

Preuve. A détailler en cours. □

En particulier, en appliquant (3′) à la décomposition d'un événement en une réunion dénombrable d'événements, nous obtenons, sur le modèle du Chapitre 1 :

PROPOSITION 1.6. *Etant donné un espace de probabilité dénombrable $(\Omega, \mathbb{P})$, pour tout événement A,*

$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}).$$

En particulier, la mesure de probabilité $\mathbb{P}$ est entièrement définie par la famille dénombrable des poids

$$\big(p_\omega = \mathbb{P}\{\omega\}\big)_{\omega\in\Omega}.$$

Ces poids sont positifs et, par σ-additivité, sont de somme égale à 1.

Réciproquement, toute famille $(q_\omega)_{\omega\in\Omega}$ de poids positifs de somme égale à 1 définit une probabilité sur Ω par la formule :

$$\mathbb{Q}(A) = \sum_{\omega\in A} q_\omega.$$

Toute famille dénombrable de poids positifs et de somme égale à 1 est appelée loi de probabilité dénombrable.

Preuve. A détailler en cours.

Nous comprenons finalement le cas dénombrable comme une extension du cas fini. Dans les deux cas, il est possible de résumer une mesure de probabilité par une suite de poids de somme égale à un.

2. Exemples

2.1. Loi de Poisson. Reprenons l'exemple discuté en introduction.

Supposons que toutes les Δ secondes, il puisse passer une voiture avec probabilité $p \in (0,1)$. Alors, au bout de 10 heures, le nombre de voitures passées suit une loi binomiale de paramètre N et p, où $N = 36000/\Delta$. (Pensons Δ comme 10 ou 1, ou encore 0.1, de sorte que N soit un entier.)

Précisément, les poids donnant le nombre de voitures passées sur une place de 10 heures sont donnés par

$$q_k^\Delta = C_N^k p^k (1-p)^{N-k}, \quad 0 \le k \le N.$$

Supposons maintenant que p soit proportionnel à Δ, i.e. $p = \lambda\Delta$. Lorsque Δ tend vers 0, N tend vers l'infini : nous obtenons, à la limite, une famille de poids indexée par les entiers. Un simple calcul montre que les poids limites ont pour forme

$$q_k = \frac{\lambda^k}{k!}\exp(-\lambda), \quad k \ge 0.$$

Naturellement, la somme de ces poids est égale à 1. La famille ainsi obtenue définit une loi de probabilité sur $\mathbb{N}$: elle porte le nom de loi de Poisson de paramètre λ.

2.2. Loi géométrique. Considérons maintenant le lancer de N pièces déséquilibrées de paramètre $p \in (0,1)$. En utilisant les exercices du TD 1, nous savons que la variable aléatoire

$$T^N : \omega = (\omega_1, \dots, \omega_N) \in \{0,1\}^N \mapsto \begin{cases} k & \text{si } \omega_1 = \dots = \omega_{k-1} = 0 \text{ et } \omega_k = 1, \\ N+1 & \text{si } \omega_1 = \dots = \omega_N = 0, \end{cases}$$

admet pour loi

$$\mathbb{P}\{T^N = k\} = (1-p)^{k-1}p \quad 1 \le k \le N,$$
$$\mathbb{P}\{T^N = k\} = 1 - (1-p)^N \quad 1 \le k \le N.$$

Lorsque N tend vers l'infini, les poids convergent vers la famille de poids

$$q_k = (1-p)^{k-1}p \quad k \ge 1.$$

Ces poids sont de somme égale à 1. Ils forment une loi de probabilité sur $\mathbb{N}^*$, appelée loi géométrique de paramètre p. Ils doivent être compris comme la loi du premier instant de succès d'une suite infinie de lancers à pile ou face.

3. Variable aléatoire

Nous suivons les définitions données dans le cas d'un univers fini :

DÉFINITION 3.1. *Etant donné un espace de probabilité* $(\Omega, \mathbb{P})$ *(au plus) dénombrable, on appelle variable aléatoire toute application* X *de* Ω *dans* $\mathbb{R}$.

Le cas échéant, $X(\Omega)$ *est un ensemble au plus dénombrable : il peut s'écrire sous la forme* $(x_i)_{i\in I}$ *où* I *est un intervalle de* $\mathbb{N}^*$ *de borne inférieure égale à 1. On appelle alors loi de* X *la famille des poids*

$$\big(\mathbb{P}\{X = x_i\}\big)_{i\in I}.$$

De façon canonique, il s'agit d'une loi de probabilité sur $\{x_i,\ i \in I\}$.

Feuille de TD Numéro 2

1. Probabilités d'événements

Ici, $(\Omega, \mathbb{P})$ désigne un espace de probabilité dénombrable.

Exercice 1.1. Un signal périodique (de fréquence entière) est transmis depuis un récepteur jusqu'à un émetteur. La fréquence du signal de sortie peut malheureusement être différente de celle du signal d'entrée.

Le problème est modélisé par l'espace produit $\Omega = (\mathbb{N}^*)^2$: la première coordonnée désigne la fréquence du signal d'entrée et la seconde la fréquence du signal de sortie. La fréquence du signal d'entrée est supposée aléatoire et la transmission est elle-aussi supposée être soumise à des perturbations aléatoires. Les aléas sont répartis selon une probabilité $\mathbb{P}$ sur Ω. Exprimer la probabilité que les signaux d'entrée et de sortie aient la même fréquence en fonction des poids $(p_n = \mathbb{P}(\{n, n\}))_{n \geq 0}$.

Exercice 1.2. Etant donnée une suite d'événements $(A_n)_{n \geq 1}$, montrer que

$$\lim_{n \to +\infty} \mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \mathbb{P}\left(\bigcup_{n \geq 1} A_n\right).$$

(Commencer par vérifier que le membre de gauche est bien défini.)

En déduire que

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \mathbb{P}(A_n).$$

(Remarquer le membre de droite a toujours un sens, même si la série ne converge pas.)

Exercice 1.3. Soit une suite $(A_n)_{n \geq 1}$ d'événements.

(1) Montrer que

$$\left(\bigcap_{n \geq 1} A_n\right)^{\complement} = \bigcup_{n \geq 1} A_n^{\complement}.$$

(2) On suppose que la suite $(A_n)_{n\geq 1}$ est décroissante, i.e. $A_n \subset A_{n+1}$, $n \geq 1$. Que peut-on dire de la suite $(A_n^{\complement})_{n\geq 1}$? En déduire que

$$\lim_{n\to+\infty} \mathbb{P}\Big(\bigcap_{k=1}^{n} A_k\Big) = \mathbb{P}\Big(\bigcap_{n\geq 1} A_n\Big).$$

Exercice 1.4. Donner un exemple d'espace de probabilité (dénombrable) et de suite $(A_n)_{n\geq 1}$ d'événements tels que

$$\lim_{n\to+\infty} \mathbb{P}(A_n) \neq \mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big).$$

(Remarquer au passage que le membre de gauche n'existe pas nécessairement.)

Exercice 1.5. Montrer qu'il n'est pas possible de trouver une mesure de probabilité sur $\mathbb{N}$ sous laquelle tous les singletons aient la même masse.

2. Variable aléatoire

Exercice 2.1. La durée de vie d'un composant électronique est modélisé comme suit. Toutes les Δ secondes, une pièce déséquilibrée de (petit) paramètre p est tirée à pile ou face. Si la pièce tombe sur 1, le composant tombe en panne ; sinon, il demeure en fonctionnement. Donner la loi du premier instant de panne du composant.

Exercice 2.2. Une machine de haute précision fabrique des pièces détachées pour l'aéronautique. Expliquer pourquoi il est légitime d'approcher la loi du nombre de défauts d'une pièce fabriquée par la machine par une loi de Poisson.

Supposons que le paramètre de la loi de Poisson soit donné par 2. Calculer la probabilité que la pièce n'ait aucun défaut, qu'elle ait plus de 3 défauts et qu'elle ait 1 ou 2 défauts.

Exercice 2.3. Soit X une variable aléatoire définie sur $(\Omega, \mathbb{P})$.

(1) Montrer que

$$\bigcap_{n\geq 1} \{X \geq n\} = \emptyset.$$

(2) En déduire que

$$\lim_{n\to+\infty} \mathbb{P}\{X \geq n\} = 0.$$

(3) On appelle fonction de répartition de la variable aléatoire X la fonction

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}\{X \leq t\}.$$

Montrer que F_X tend vers 1 lorsque t tend vers l'infini.

EXERCICE 2.4. On appelle fonction de répartition d'une variable aléatoire X la fonction

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}\{X \leq t\}.$$

(1) Calculer la fonction de répartition d'une variable aléatoire de loi géométrique de paramètre p sur $\mathbb{N}^*$.

(2) On suppose que p s'écrit sous la forme $p = \lambda/N$ avec $\lambda > 0$ fixé et N entier naturel très grand. Montrer que, pour tout $t \geq 0$

$$F_X(tN) \to 1 - \exp(-\lambda t).$$

(3) On reprend maintenant le modèle de l'Exercice **??**. On suppose que $p = \lambda\Delta$ avec $\lambda > 0$ et Δ suffisamment petit pour que $p < 1$.

Donner, pour Δ tendant vers 0, une approximation de la probabilité que le composant soit encore en fonctionnement une heure après sa mise en marche.

EXERCICE 2.5. Soit X une loi géométrique de paramètre $1-\exp(-1)$ sur $\mathbb{N}^*$. Calculer la loi de $\exp(-X+1)$. Tracer sa fonction de répartition. Que peut-on observer ?

CHAPITRE 3

Probabilités conditionnelles et indépendance

Les deux chapitres précédents se sont attaché au calcul de probabilités d'un événement en particulier. Un des problèmes principaux de la théorie des probabilités consiste en réalité à calculer la probabilité d'un événement sachant qu'un autre s'est effectivement réalisé, i.e. à modéliser un lien de causalité entre les événements.

Ce lien de causalité existe par essence même dans la nature. A titre d'exemple, la probabilité qu'il pleuve sur les côtes méditerranéennes un après-midi de printemps est plus ou moins élevée selon la météorologie du matin.

Les probabilités conditionnelles offrent une réponse théorique, très générale, à la modélisation de la causalité. En l'absence de causalité, les probabilités conditionnelles conduisent à la définition d'un concept essentiel : l'indépendance.

1. Probabilités conditionnelles

1.1. Principe. Pour illustrer le principe des probabilités conditionnelles, focalisons-nous sur un exemple simple : un agent financier, au service d'une banque, doit prendre des positions, au cours de la journée, à l'achat ou la vente sur une action, sur une matière première ou sur une monnaie. Par essence, il ne sait pas, au début de la journée, quel sera le cours de l'actif financier en fin de journée. (A moins de disposer d'informations spécifiques, dites d'initiées, dont l'utilisation est punie par la loi.) En revanche, il peut être en mesure de prédire, avec plus ou moins d'incertitude, quel sera le cours en fin de journée. Naturellement, au fur et à mesure de la journée, l'incertitude est amenée à diminuer : l'agent bénéficie d'informations lui permettant d'affiner ses prévisions.

La théorie des probabilités doit permettre de tenir en compte du supplément d'informations dont bénéficie l'agent. De façon heuristique, tout se passe comme si, au fur et à mesure du temps, l'univers des possibles venait à se rétrécir.

Les probabilités conditionnelles obéissent exactement à ce principe : il s'agit de restreindre l'univers des possibles pour modéliser la connaissance d'informations supplémentaires. Si je lance deux dés à six faces et que je m'interroge sur la probabilité d'obtenir une paire, l'univers des possibles n'est pas le même avant ou après le lancer du premier dé. Après le premier lancer, il est clairement restreint.

La restriction de l'univers des possibles conduit à redéfinir les probabilités avec lesquelles surgissent les événements. Précisément, il s'agit de décider de la place (ou encore de la mesure) d'un événement à l'intérieur du nouvel univers. (Faire un dessin.)

Définition 1.1. *Etant donnés un espace de probabilité (au plus dénombrable)* $(\Omega, \mathbb{P})$ *et un événement* B *de mesure non nulle, la probabilité conditionnelle de* A *sachant* B *mesure la proportion occupée par* A *dans* B *ou encore par* $A \cap B$ *dans* B*. Elle est égale à* $\mathbb{P}(A \cap B)/\mathbb{P}(B)$ *et notée* $\mathbb{P}(A|B)$*.*

Précisément, il faut voir l'application

$$\mathbb{P}(\cdot|B) : A \in \mathcal{P}(\Omega) \mapsto \mathbb{P}(A|B),$$

comme une nouvelle mesure de probabilité. Sous cette nouvelle mesure, tous les événements disjoints de B sont de mesure nulle. On dit qu'ils ne sont pas chargés.

La contrainte de non-nullité de la mesure de B est tout-à-fait naturelle. Un événement de mesure nulle est en effet un événement que l'on n'observera jamais en pratique.

1.2. Formule des probabilités totales. En pratique, il est fréquent que la probabilité d'un événement ne soit connue que conditionnellement à un autre. Ceci est par exemple vrai en médecine : les tests de dépistage, d'une maladie virale ou d'une affection génétique, ne sont jamais totalement fiables. Les résultats sont toujours entâchés d'une marge d'erreur, heureusement faible. Le cas échéant, il y a dissymétrie de l'erreur dans la détection d'un faux positif et d'un faux négatif : la probabilité de déclarer sain un patient malade est différente de celle de déclarer malade un sujet sain.

D'une certaine façon, le réel ou encore l'univers des possibles admet alors une structure arborescente. (Faire un dessin.) Le réel est structuré en branches et les probabilités conditionnelles sont les probabilités de "sauter" d'un noeud de l'arbre à un autre. D'un point de vue calculatoire, cette idée porte le nom de *formule des probabilités totales* :

Proposition 1.2. *Etant donnés un espace (au plus dénombrable)* $(\Omega, \mathbb{P})$ *et une partition de* Ω *en* N *événements* $B_1, \dots, B_N$ *deux-à-deux disjoints et de mesure non nulle, la mesure d'un événement* A *s'écrit*

$$\mathbb{P}(A) = \sum_{i=1}^{N} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

1.3. Formule de Bayes. La formule des probabilités totales doit être comprise comme un parcours de l'arbre depuis la racine jusqu'aux feuilles. Par exemple, dans l'exemple du dépistage des maladies, la formule des probabilités totales permet de déterminer la probabilité d'observer un faux moyennant la connaissance des probabilités de saut d'un noeud de l'arbre à un autre. Il est néanmoins possible de renverser le point de vue : sachant que l'on a observé un faux, quelle est la probabilité qu'il s'agisse d'un patient sain ? Dit autrement, il est possible de parcourir l'arbre des feuilles jusqu'à la racine.

Proposition 1.3. *Etant donnés un espace (au plus dénombrable)* $(\Omega, \mathbb{P})$, *une partition de* Ω *en* N *événements* $B_1, \ldots, B_N$ *deux-à-deux disjoints et de mesure non nulle et un événement* A *de mesure non-nulle*

$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(A|B_1)\mathbb{P}(B_1)}{\sum_{i=1}^{N} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Cette formule donne les probabilités de sauter à rebours d'un noeud de l'arbre à un autre.

Exemple. (Sources : Wikipédia.) Le dépistage de la trisomie 21 tel qu'il est couramment pratiqué en France conduit à un taux de détection de 90%, pour un taux de faux-positif de 5%.

Le taux de détection de 90% signifie que si cent femmes enceintes d'un foetus trisomique passent ce test, en moyenne 90 grossesses anormales seront correctement identifiées, et 10 ne seront pas détectées.

Parallèlement, le taux de faux positif de 5% signifie que si cent femmes enceintes d'un foetus normal passent ce test, en moyenne 5 seront identifiées à tort comme des grossesses anormales.

Sachant que la probabilité qu'un foetus soit porteur d'une trisomie 21 est 1/650, écrire l'arbre des possibles et calculer la probabilité qu'un foetus soit sain sachant qu'il a été déclaré anormal.

A.N. La probabilité que le foetus soit effectivement malade sachant qu'il est déclaré porteur de la trisomie 21 est environ de 2.7 %. La probabilité que le foetus soit malade sachant qu'il est déclaré sain est environ .01 %.

2. Indépendance d'événements

2.1. Indépendance de deux événements. La notion d'indépendance de deux événements exprime l'absence de causalité entre eux :

Définition 2.1. *Etant donné un espace de probabilité (au plus dénombrable)* $(\Omega, \mathbb{P})$, *deux événements* A *et* B, *de mesure non nulle, sont dits indépendants si* $\mathbb{P}(A|B) = \mathbb{P}(A)$ *ou* $\mathbb{P}(B|A) = \mathbb{P}(B)$, *i.e.* $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Par extension, un événement de mesure nulle est indépendant de tout autre.

Exemple. Soit l'ensemble $\{1,\ldots,6\}^2$ muni de la probabilité uniforme. Alors, pour tout $A, B \subset \{1,\ldots,6\}$, $A \times \{1,\ldots,6\}$ et $\{1,\ldots,6\} \times B$ sont indépendants.

Ceci traduit l'idée que, dans le lancer de deux dés, les résultats sont indépendants.

2.2. Indépendance de n événements. L'indépendance de n événements doit être comprise dans un sens très strict : il s'agit de dire que l'occurence d'une partie de ces n événements ne favorise ni ne défavorise l'occurence d'une autre partie de ces n événements. En aucun cas, il ne peut suffire de comparer deux à deux les événements.

Définition 2.2. *Etant donné un espace de probabilité (au plus dénombrable) $(\Omega, \mathbb{P})$, n événements $A_1, \ldots, A_n$ sont dits indépendants si*

$$\forall 1 \leq \ell \leq n,\ \forall 1 \leq i_1 < i_2 < \cdots \leq i_\ell, \quad \mathbb{P}\big(\bigcap_{j=1}^{\ell} A_{i_j}\big) = \prod_{j=1}^{\ell} \mathbb{P}(A_{i_j}).$$

Contre-exemple 1. Considérons l'ensemble $\{1,2,3,4\}$ muni de la probabilité uniforme et les événements

$$A = \{1,2\},\ B = \{2,3\},\ C = \{3,1\}.$$

Montrer que A et B sont indépendants, B et C sont indépendants, A et C sont indépendants, mais que A, B et C ne sont pas indépendants.

Contre-exemple 2. Considérons le lancer de deux dés et les événements :

$$\begin{aligned} A &= \{1,\ldots,6\} \times \{1,2,5\}, \\ B &= \{1,\ldots,6\} \times \{4,5,6\}, \\ C &= \{(i,j) \in \{1,\ldots,6\}^2 : i + j = 9\}. \end{aligned}$$

Montrer que $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ mais que A et B ne sont pas indépendants, B et C ne sont pas indépendants, et que A et C ne sont pas indépendants.

3. Indépendance de variables aléatoires

3.1. Définition.

Définition 3.1. *Etant donné un espace de probabilité (au plus dénombrable) $(\Omega, \mathbb{P})$, n variables aléatoires $X_1, \ldots, X_n$, à valeurs dans n ensembles* *au plus dénombrables $E_1, \dots, E_n$, sont dites indépendantes si*

$$\mathbb{P}\big\{X_1 = i_1, \dots, X_n = i_n\big\} = \prod_{j=1}^{n} \mathbb{P}\{X_j = i_j\}, \quad 1 \leq j \leq n.$$

Exemple typique. Le lancer de deux dés à six faces est modélisé par l'ensemble $\Omega = \{1, \dots, 6\}^2$, muni de la probabilité uniforme. Chacun des deux lancers peut être modélisé par une application coordonnée :

$$X_i : (\omega_1, \omega_2) \in \Omega \mapsto \omega_i, \quad i = 1, 2.$$

Alors, les applications X_1 et X_2 sont indépendantes.

Afin de généraliser cet exemple, nous rappelons quelques propriétés des séries à termes positifs.

3.2. Théorème de Fubini.

PROPOSITION 3.2. *Etant donnés $(a_{i,j})_{i,j\in\mathbb{N}}$ une famille de réels positifs, la famille*

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j}$$

admet une limite finie lorsque n et m convergent vers l'infini si et seulement si la série de terme général

$$A_i = \sum_{j\geq 1} a_{i,j}$$

(toujours défini comme un élément de $\mathbb{R}_+ \cup \{+\infty\}$) est convergente. Le cas échéant, la limite

$$\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} a_{i,j}$$

est égale à

$$\sum_{i=1}^{+\infty} A_i.$$

3.3. Poids produits. Nous pouvons maintenant donner une généralisation de l'exemple typique ci-dessus :

PROPOSITION 3.3. *Etant donnés Ω_1 et Ω_2 deux ensembles (au plus) dénombrables et $(p^1_i)_{i\in\Omega_1}$ et $(p^2_j)_{j\in\Omega_2}$ deux familles de poids de probabilité sur $\Omega = \Omega_1 \times \Omega_n$ la famille de poids*

$$q_{i_1,i_2} = p_{i_1} p_{i_2}, \quad (i_1, i_2) \in \Omega,$$

définit une famille de poids de probabilité sur Ω, appelée famille de poids produits.

De plus, sous ces poids de probabilité, les applications coordonnées

$$X_i : (\omega_1, \omega_2) \in \Omega \mapsto \omega_i$$

sont indépendantes.

3.4. Caractérisation de l'indépendance. Le théorème de Fubini permet également de caractériser l'indépendance de deux variables aléatoires :

Proposition 3.4. *Etant donnés un espace de probabilité (au plus dénombrable)* $(\Omega, \mathbb{P})$ *et deux variables aléatoires* X *et* Y *à valeurs dans* E *et* F*, sous-ensemble au plus dénombrable de* $\mathbb{R}$*,* X *et* Y *sont indépendantes si et seulement si, pour toutes parties* A *et* B *dans* E *et* F*,*

$$\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}.$$

Preuve. Supposons pour simplifier que A et B sont dénombrables. Alors, les éléments de A et B peuvent être numérotés par des entiers naturels :

$$A = \{a_n,\ n \geq 1\},\ B = \{b_m,\ m \geq 1\}.$$

Ici, les $(a_n)_{n\geq 1}$ sont deux-à-deux distincts. De même pour les $(b_n)_{n\geq 1}$. Alors, nous pouvons écrire (sans hypothèse d'indépendance sur X et Y)

$$\begin{aligned}
\{X \in A, Y \in B\} &= \bigcup_{n\geq 1, m\geq 1} \{X = a_n, Y = b_m\} \\
&= \bigcup_{N\geq 1} \bigcup_{1\leq n\leq N, 1\leq m\leq N} \{X = a_n, Y = b_m\}. \\
&:= \bigcup_{N\geq 1} C_N.
\end{aligned}$$

Nous remarquons que les événements $(C_N)_{N\geq 1}$ sont croissants de sorte que

$$\begin{aligned}
\mathbb{P}\big(\{X \in A, Y \in B\}\big) &= \lim_{N\to+\infty} \mathbb{P}(C_N) \\
&= \lim_{N\to+\infty} \mathbb{P}\bigg(\bigcup_{1\leq n\leq N, 1\leq m\leq N} \{X = a_n, Y = b_m\}\bigg) \\
&= \lim_{N\to+\infty} \sum_{n=1}^{N} \sum_{m=1}^{N} \mathbb{P}\{X = a_n, Y = b_m\},
\end{aligned}$$

en raison du caractère deux-à-deux disjoints des événements.

Supposons maintenant X et Y indépendantes. Alors, nous obtenons sont croissants de sorte que

$$\mathbb{P}\big(\{X \in A, Y \in B\}\big) = \lim_{N\to+\infty} \sum_{n=1}^{N} \sum_{m=1}^{N} \mathbb{P}\{X = a_n\}\mathbb{P}\{Y = b_m\}.$$

Par application du théorème de Fubini, nous obtenons

$$\mathbb{P}\big(\{X \in A, Y \in B\}\big) = \left(\sum_{n\geq 1} \mathbb{P}\{X = a_n\}\right)\left(\sum_{m\geq 1} \mathbb{P}\{Y = b_m\}\right) = \mathbb{P}(A)\mathbb{P}(B).$$

Ceci montre l'implication dans l'énoncé de la proposition. La réciproque est triviale.

Autre preuve. On commence par traiter le cas fini. Puis on écrit A et B comme deux réunions croissantes.

Feuille de TD Numéro 3

1. Probabilités conditionnelles

Exercice 1.1. Dans une famille de deux enfants, quelle est la probabilité que le cadet soit une fille sachant que l'aîné est un garçon ? Sachant que l'un des deux est un garçon, quelle est la probabilité que l'autre soit une fille ?

Exercice 1.2. (On pourra utiliser l'exercice précédent.) Lorsque le téléphone sonne dans une famille de deux enfants composée exactement d'une fille et un garçon, la fille répond, en l'absence des parents, avec probabilité p.

Les « Castagnier » ont deux enfants. Ils les ont laissés seuls pour la soirée. Le téléphone sonne. Une fille décroche l'appareil. Quelle est la probabilité que l'autre enfant soit un garçon ?

Exercice 1.3. Lorsque la sonnette résonne à la porte d'entrée d'un logement occupé par une famille de deux enfants, l'aîné, en l'absence des parents, ouvre la porte avec la probabilité p.

Les « Fourchaume » ont deux enfants. Ils les ont laissés seuls pour la matinée. Le facteur sonne. Une fille lui ouvre la porte. Quelle est la probabilité que l'autre enfant soit un garçon ?

Exercice 1.4. Dans cet exercice, nous supposons pour simplifier que les yeux d'un être humain sont soit de couleur bleue soit de couleur marron.

Le gène de la couleur bleue est supposé récessif : il faut avoir être hérité des deux parents pour qu'il soit effectivement exprimé. Le « génotype » correspondant s'écrit « bb ». Les autres s'écrivent « Mb » et « MM ».

(1) La soeur de François Pignon a les yeux bleus, mais ses parents les yeux marrons. Quelle est la probabilité que François Pignon aient les yeux bleus ?
(2) La femme de François Pignon a également les yeux bleus. Quelle est la probabilité que leur deuxième enfant aient les yeux bleus sachant que leur premier a les yeux marrons ?

Exercice 1.5. Le restaurant parisien « Chez Septime » reçoit une grosse livraison de boîtes d'oeuf frais avant la réception du Président Novalès, chef d'un Etat Sud Américain. Son patron, Monsieur Septime, estime lors de la réception que deux pourcents des boîtes sont abimées. Il accepte la livraison moyennant réduction de son prix tout en sachant que

(1) 2/3 des boîtes abimées contiennent au moins un oeuf cassé,

(2) 98% des boîtes non-abimées ne contiennent pas d'oeufs cassés.

Un apprenti cuisinier range une boîte d'oeufs lorsqu'il est surpris par Monsieur Septime avec un oeuf cassé dans la main. De réputation colérique, le patron s'en prend à son employé et le menace de déduire l'oeuf de sa paie. En réalité innocent, l'employé s'apprête à désigner la boîte du bout des doigts pour plaider sa bonne foi : quelle est la probabilité que celle-ci soit effectivement abîmée ?

Exercice 1.6. On lance deux dés. Quelle est la probabilité qu'au moins l'un d'entre eux donne 6 sachant que les deux dés affichent des résultats différents ?

Exercice 1.7. Une usine fabrique des pièces, avec une proportion de 0,05 de pièces défectueuses. Le contrôle de fabrication est tel que : (a) si la pièce est bonne, elle est acceptée avec la probabilité 0,96 (b) si la pièce est mauvaise, elle est refusée avec la probabilité 0,98. Une pièce est contrôlée au hasard. Quelle est la probabilité qu'il y ait une erreur de contrôle ? Quelle est la probabilité qu'une pièce acceptée soit mauvaise ?

Exercice 1.8. Une urne contient b boules blanches et r boules rouges. Une boule est tirée au hasard et est remplacée dans l'urne par $d+1$ boules de la même couleur. Quelle est la probabilité que la seconde boule tirée soit rouge ? Quelle est la probabilité que la première boule tirée soit rouge sachant que la seconde tirée est rouge ?

Exercice 1.9. Une compagnie d'assurance répartit ses assurés en trois catégories : conducteur à faible risque, conducteur à risque moyen et conducteur à haut risque. Les statistiques de la compagnie indiquent que la probabilité d'accident sur une période de un an est 0,05, 0,15 et 0,30 selon la catégorie. Par ailleurs, la répartition des assurés est la suivante : 20% sont à bas risque, 50% à risque moyen et 30% à haut risque. Un assuré est choisi au hasard : quelle est la probabilité qu'il ait un accident au cours de l'année ? Sachant que l'assuré n'a pas eu d'accident lors de l'année écoulée, quelle est la probabilité qu'il soit à faible risque ?

Exercice 1.10. Un lot de 100 dés contient 25 dés pipés dont la probabilité de sortie du 6 est 1/2. Un dé est choisi au hasard et lancé : il donne 6. Quelle est la probabilité qu'il soit pipé ?

Exercice 1.11. Soit $(\Omega, \mathbb{P})$ un espace au plus dénombrable. Etant donnés $n+1$ événements $A_1, \dots, A_{n+1}$, $n \geq 1$, tels que $\mathbb{P}(\cap_{i=1}^n A_i) > 0$, montrer que

$$\mathbb{P}\Big(\bigcap_{i=1}^{n+1} A_i\Big) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_{n+1}|A_1 \cap A_2 \cap \dots \cap A_n).$$

Exercice 1.12. On se propose de démontrer par une méthode probabiliste la relation :

$$\sum_{p+q=n, p\geq 1, q\geq 1} pq = \frac{(n+1)n(n-1)}{6}.$$

Pour cela, on va calculer de deux manières différentes la probabilité de tirer, sans remise et successivement, dans une urne contenant $n+1$ boules numérotées de 1 à $n+1$, trois boules d'indices croissants. On note A cet événement.

(1) Calculer, pour tout (i, j, k) tel que $1 \leq i < j < k \leq n+1$, la probabilité conditionnelle de A sachant que l'ensemble des trois numéros obtenus est $\{i, j, jk\}$. En déduire que $\mathbb{P}(A) = 1/6$.

(2) Montrer, par dénombrement, que

$$\mathbb{P}(A) = \big[(n+1)n(n-1)\big]^{-1} \sum_{p+q=n, p\geq 1, q\geq 1} pq.$$

Conclure.

2. Evénements indépendants

Exercice 2.1. Considérons l'ensemble $\{1, 2, 3, 4\}$ muni de la probabilité uniforme et les événements

$$A = \{1, 2\},\ B = \{2, 3\},\ C = \{3, 4\}.$$

Montrer que A et B sont indépendants, B et C sont indépendants, A et C sont indépendants, mais que A, B et C ne sont pas indépendants.

Exercice 2.2. Considérons le lancer de deux dés et les événements :

$$A = \{1, \dots, 6\} \times \{1, 2, 5\},$$
$$B = \{1, \dots, 6\} \times \{4, 5, 6\},$$
$$C = \{(i, j) \in \{1, \dots, 6\}^2 : i + j = 9\}.$$

Montrer que $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ mais que A et B ne sont pas indépendants, B et C ne sont pas indépendants, et que A et C ne sont pas indépendants.

Exercice 2.3. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable. Montrer que deux événements A et B sont indépendants si et seulement si $A^\complement$ et $B^\complement$ sont indépendants. Que dire dans le cas de n événements, $n \geq 3$?

Exercice 2.4. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable muni de trois événements A, B et C indépendants. Montrer que A est indépendant de $B \cup C$.

Exercice 2.5. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable.

(1) Etant donnés n événements indépendants $A_1, \ldots, A_n$, $n \geq 2$, montrer que $\mathbb{P}(\cup_{k=1}^n A_k) = 1 - \prod_{k=1}^n (1 - \mathbb{P}(A_k))$.

(2) Peut-on trouver une partition de Ω en n événements indépendants, $n \geq 2$, de même probabilité ?

Exercice 2.6. Une classe de CP compte 4 garçons et 6 filles. Elle est mélangée avec une classe de CE1 composée de 6 garçons et de n filles, $n \geq 0$. Les deux classes sont réunies dans une même salle. Un élève (garçon ou fille) est alors interrogé au hasard. Comment choisir n de sorte que les événements « l'élève est un garçon » et l'élève est en CP » soient indépendants ?

Exercice 2.7. Un signal est transmis le long de n relais montés en série. Pour simplifier, le signal est réduit à un 0 ou un 1. Chaque relais transmet le signal avec probabilité $0 < p < 1$ et le déforme 1 en 0 ou 0 en 1 avec probabilité $1-p$. Les relais fonctionnent indépendamment les uns des autres. Quelle est la probabilité que le chiffre transmis à l'arrivée soit le bon ? Décrire le comportement asymptotique de cette probabilité lorsque n tend vers l'infini.

Exercice 2.8. Une urne contient 12 boules numérotées de 1 à 12. On en tire une au hasard et on consdière les événements $A =$ « la boule tirée porte un numéro pair » et $B =$ « la boule tirée porte un numéro multiple de 3 ». Les événements A et B sont-ils indépendants. Reprendre la question en remplaçant 12 par 13.

3. Variables aléatoires indépendantes

Exercice 3.1. Etant donnés un entier $n \geq 1$ et un réel $0 < p < 1$, nous considérons l'ensemble $\Omega = \{0,1\}^n$ muni de la mesure de probabilité $\mathbb{P}$ définie par

$$\mathbb{P}\{(\varepsilon_1, \ldots, \varepsilon_n)\} = p^{\sum_{i=1}^n \varepsilon_i}(1-p)^{n-\sum_{i=1}^n \varepsilon_i}, \quad \varepsilon_i \in \{0,1\}.$$

Nous considérons également les variables aléatoires

$$X_i : (\varepsilon_1, \ldots, \varepsilon_n) \mapsto \varepsilon_i, \quad 1 \leq i \leq n.$$

Vérifier que les variables $X_1, \ldots, X_n$ sont indépendantes.

Exercice 3.2. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable muni de X et Y variables aléatoires indépendantes de loi de Bernoulli de paramètres $p \in]0,1[$ et $q \in]0,1[$. Calculer la loi du produit XY.

Exercice 3.3. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable muni de $X_1, \ldots, X_n$ n variables aléatoires indépendantes de même loi de Bernoulli de paramètre $0 < p < 1$, donner la loi de $S = \sum_{i=1}^{n} X_i$.

Exercice 3.4. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable muni de X et Y variables aléatoires de lois binomiales de paramètres (m, p) et (n, p), pour deux entiers m et n plus grands que 1 et un réel $p \in]0, 1[$. Montrer que $X + Y$ suit une loi binomiale de paramètres $(m + n, p)$.

Exercice 3.5. Soit $(\Omega, \mathbb{P})$ un espace de probabilité au plus dénombrable muni de X et Y variables aléatoires de lois de Poisson de paramètres $\lambda > 0$ et $\mu > 0$, donner la loi de $X + Y$.

Exercice 3.6. Deux joueurs jouent au jeu suivant : un dé à six faces est lancé suivi d'une pièce à pile ou face. Le joueur A gagne, en euro, le résultat du dé si la pièce tombe sur pile et perd, en euro, le résultat du dé si la pièce tombe sur face. Modéliser l'expérience et donner la loi du gain du joueur A.

Exercice 3.7. On s'intéresse au nombre de boîtes vides lors de lancers de n balles dans N boîtes, les lancers étant uniformes et indépendants les uns des autres. L'univers retenu pour la modélisation est l'ensemble $\Omega = \{1, \ldots, N\}^n$ muni de la probabilité uniforme, notée $\mathbb{P}$. Dans la suite, on note X_i le numéro de la boîte dans laquelle la balle i tombe, Y_k le nombre de balles tombées dans la boîte k, Z_k la variable de Bernoulli valant 1 si la boîte k est vide et V le nombre de boîtes vides.

(1) Montrer que les $(X_i)_{1 \leq i \leq n}$ sont indépendantes.

(2) Exprimer l'événement $\{Y_k = s\}$, $1 \leq k \leq N$, à l'aide des variables $(X_i)_{1 \leq i \leq n}$. En déduire la loi de Y_k et en particulier la valeur de $\mathbb{P}\{Y_k = 0\}$.

(3) Exprimer Z_k à l'aide de Y_k. Quelle est la loi de Z_k ? Les $(Z_k)_{1 \leq k \leq N}$ sont-elles indépendantes ? Exprimer V à l'aide des $(Z_k)_{1 \leq k \leq N}$.

CHAPITRE 4

Espérance

L'espérance est une notion essentielle de la théorie des probabilités, à même de prédire le comportement moyen observable lors de la répétition d'une même expérience dans des conditions similaires et indépendantes.

A titre d'exemple, lorsque le lancer d'une pièce à pile ou face est répété un grand nombre de fois, il est intuitivement attendu de la fréquence d'apparition du pile (ou du face) d'être proche de $1/2$ si la pièce est bien équilibrée. Une telle fréquence, calculée sur les observations, est appelée empirique, car fondée sur l'observation de l'expérience. Par ailleurs, la quantité $1/2$ apparaît comme un gain moyen à espérer lors de la répitition d'un très grand nombre de lancers.

La notion d'espérance vise à généraliser cette idée et à définir, pour une expérience, une quantité théorique susceptible de représenter un gain imaginaire à espérer à l'issue de la répétition de l'expérience, un très grand nombre de fois.

1. Espérance d'un jeu à deux issues

1.1. Variable aléatoire de Bernoulli.

Définition 1.1. *Etant donnés un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ au plus dénombrable et un événement A, la variable aléatoire $\mathbf{1}_A$ égale à 1 sur A et à* 0 *sur son complémentaire est appelée indicatrice de l'événement A.*

Sa loi est donnée par $\mathbb{P}\{\mathbf{1}_A = 1\} = \mathbb{P}(A)$ et $\mathbb{P}\{\mathbf{1}_A = 0\} = 1 - \mathbb{P}(A)$. Il s'agit d'une loi de Bernoulli de paramètre $\mathbb{P}(A)$.

1.2. Espérance de la loi de Bernoulli. Comme expliqué ci-dessus, l'idée d'espérance est naturellement comprise lors du lancer d'un grand nombre de pièces à pile ou face. Très clairement, si la pièce est déséquilibrée et la probabilité qu'elle tombe sur face égale à un réel p entre 0 et 1, la fréquence d'apparition du face lors de la répétition des lancers est espérée proche de p.

L'expérience du pile ou face peut être comprise dans un sens très large : en fait, tout expérience à valeurs dans $\{0, 1\}$ peut être, avec un peu d'imagination, comprise comme un lancer de pile de face.

En particulier, si une variable aléatoire X donnée prend les valeurs 1 ou 0 avec probabilités p et $1-p$, son gain moyen ou encore son espérance apparaît naturellement comme égale à p.

Définition 1.2. *Etant donnés un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ au plus dénombrable et une variable aléatoire X de loi de Bernoulli de paramètre p, l'espérance de X (ou encore sa moyenne), notée $\mathbb{E}(X)$, est égale à p, i.e.*

$$\mathbb{E}\big[X\big] = p.$$

En particulier, si $X = \mathbf{1}_A$, pour un événement A, alors

$$\mathbb{E}\big[\mathbf{1}_A\big] = \mathbb{P}(A).$$

1.3. Jeu à deux issues autres que 0 et 1. La formulation précédente a fait apparaître l'espérance d'un jeu de pile ou face à deux issues : 0 et 1. Dans ce cas, l'espérance peut être comprise comme le barycentre de 0 et 1 affectée des poids de probabilité $(1-p)$ ="probabilité de prendre la valeur 0" et p ="probabilité de prendre la valeur 1", i.e.

$$p = 0 \times (1-p) + 1 \times p.$$

En réalité, cette vue de l'esprit se généralise au cas d'un jeu à deux issues éventuellement différentes de 0 et 1. Imaginons par exemple, qu'au cours d'un lancer à pile ou face, un joueur ramasse 2 euros à chaque 1 et perdre 1 euro à chaque 0. Son gain moyen ou gain empirique, car calculé à partir de l'osbservation, est espéré proche de 2 fois la probabilité d'obtenir 1 plus -1 fois la probabilité d'obtenir la valeur 0.

Etant donnés un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ au plus dénombrable et une variable aléatoire X de loi de Bernoulli de paramètre p, l'espérance de X (ou encore sa moyenne), notée $\mathbb{E}(X)$, est égale à p, i.e.

$$\mathbb{E}\big[X\big] = p.$$

En particulier, si $X = \mathbf{1}_A$, pour un événement A, alors

$$\mathbb{E}\big[\mathbf{1}_A\big] = \mathbb{P}(A).$$

Définition 1.3. *Etant donnés un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ au plus dénombrable et une variable aléatoire X prenant deux valeurs $\{a, b\}$ selon la loi*

$$\mathbb{P}\{X = a\} = p_a$$
$$\mathbb{P}\{X = b\} = p_b = (1 - p_a),$$

alors, l'espérance de X (ou encore sa moyenne théorique), notée $\mathbb{E}(X)$, est égale à

$$\mathbb{E}\big[X\big] = ap_a + bp_b.$$

Une autre façon de voir cette égalité consiste à voir X comme

$$X = a\mathbf{1}_A + b(1 - \mathbf{1}_A),$$

où A désigne l'événement $A = \{X = a\}$. Alors, nous sommes tentés d'écrire

$$\mathbb{E}[X] = a\mathbb{E}\big[\mathbf{1}_A\big] + b\mathbb{E}\big[1 - \mathbf{1}_A\big].$$

Un simple calcul montre que nous retrouvons la formule de la définition.

En particulier, nous avons la proposition suivante :

Proposition 1.4. *Etant donné un espace de probabilité* $(\Omega, \mathcal{A}, \mathbb{P})$ *au plus dénombrable muni d'une variable aléatoire* X *prenant deux valeurs. Alors*

$$\forall \lambda \in \mathbb{R}, \quad \mathbb{E}[\lambda X] = \lambda \mathbb{E}(X),$$

et plus généralement,

Proposition 1.5. *Etant donné un espace de probabilité* $(\Omega, \mathcal{A}, \mathbb{P})$ *au plus dénombrable muni d'une variable aléatoire* X *prenant deux valeurs* a *et* b. *Alors, pour toute fonction* f *de* $\mathbb{R}$

$$\forall \lambda \in \mathbb{R}, \quad \mathbb{E}[f(X)] = f(a)\mathbb{P}\{X = a\} + f(b)\mathbb{P}\{X = b\}.$$

2. Espérance d'une variable aléatoire finie

Dans le paragraphe précédent, nous avons compris l'espérance d'une variable aléatoire à deux issues comme le barycentre de ces deux issues, pondéres par les poids de probabilité de chacune des deux issues. Très naturellement, cette idée demeure dans le cas d'une variable aléatoire prenant un nombre fini de valeurs.

2.1. Espérance d'une combinaison linéaire d'indicatrices. Pour comprendre l'espérance d'une combinaison linéaire d'indicatrices, nous travaillons, comme dans le paragraphe 1, par analogie avec le gain à espérer dans un jeu.

Supposons que X désigne une variable aléatoire définie sur un espace de probabilités au plus dénombrable $(\Omega, \mathbb{P})$ et à valeurs dans $\{1, \dots, n\}$: à cette variable aléatoire, il est possible d'associer les événements $\{X = 1\}$, $\{X = 2\}$, ... et $\{X = n\}$. Pour une famille de n réels $\lambda_1, \dots, \lambda_n$, nous considérons alors le jeu suivant : si la réalisation expérimentale associée à X est égale à i, un joueur gagne la quantité algébrique λ_i. De fait, le gain qu'il réalise à la suite de l'expérience s'écrit exactement comme :

$$\sum_{i=1}^{n} \lambda_i \mathbf{1}_{\{X=i\}}.$$

Imaginons maintenant que l'expérience modélisée par X soit répétée un grand nombre de fois, dans des circonstances indépendantes ; lorsque l'expérience donne le chiffre i, le joueur empoche la quantité algébrique λ_i. Le gain moyen par expérience qu'il peut espérer au bout d'un grand nombre d'expérience s'écrit exactement comme

$$\sum_{i=1}^{n} \lambda_i \mathbb{P}\{X = i\}.$$

De sorte que, sur le modèle du premier paragraphe, nous pouvons écrire :

$$\mathbb{E}\Big[\sum_{i=1}^{n} \lambda_i \mathbf{1}_{\{X=i\}}\Big] = \sum_{i=1}^{n} \lambda_i \mathbb{P}\{X = i\}. \tag{2.1}$$

Nous observons alors, comme précédemment, un phénomène de linéarité. Chaque indicatrice $\mathbf{1}_{\{X=i\}}$ admet $\mathbb{P}\{X = i\}$ pour espérance et tout se passe comme si

$$\mathbb{E}\Big[\sum_{i=1}^{n} \lambda_i \mathbf{1}_{\{X=i\}}\Big] = \sum_{i=1}^{n} \lambda_i \mathbb{E}\big[\mathbf{1}_{\{X=i\}}\big].$$

Nous comprenons que l'espérance est linéaire. Nous reviendrons sur cette notion par la suite.

2.2. Décomposition d'une v.a. comme somme d'indicatrices. En réalité, toute variable aléatoire prenant un nombre fini de valeurs peut se décomposer comme une combinaison linéaire d'indicatrices. En effet, si X est une variable aléatoire, définie sur un espace au plus dénombrable $(\Omega, \mathbb{P})$ prenant les valeurs $\{x_1, \ldots, x_n\}$, nous pouvons toujours écrire :

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{\{X=x_i\}}.$$

De fait, en vertu de la relation (**??**), nous pouvons poser

Définition 2.1. *Etant donnés un espace de probabilité au plus dénombrable* $(\Omega, \mathbb{P})$ *et* X *une variable aléatoire prenant les valeurs* $\{x_1, \ldots, x_n\}$, *alors l'espérance de* X, *notée* $\mathbb{E}(X)$, *est le barycentre des* $x_1, \ldots, x_n$ *pondérés par les poids de probabilité* $\mathbb{P}\{X = x_1$, …, $\mathbb{P}\{X = x_{xn}$, *i.e.*

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \mathbb{P}\{X = x_i\}.$$

En particulier, l'espérance de X *ne dépend que de la loi de* X.

2.3. Linéarité. En vertu de la discussion menée en début de section, nous comprenons que l'espérance est linéaire :

DÉFINITION 2.2. *Etant donnés un espace de probabilité au plus dénombrable* $(\Omega, \mathbb{P})$ *et* X *et* Y *deux variables aléatoires prenant un nombre fini de valeurs, pour tout réel* λ *et* μ*,*

$$\mathbb{E}\big(\lambda X + \mu Y\big) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y).$$

2.4. Exemples. Examions le lancer d'un ou plusieurs dés :

Exemple 1. Etant donné le lancer d'un dé à 6 faces modélisé par une v.a. X de loi uniforme sur $\{1, \ldots, 6\}$, l'espérance de X vaut :

$$\mathbb{E}(X) = \frac{1}{6} \sum_{i=1}^{6} i = \frac{6 \times 7}{2 \times 6} = \frac{7}{2}.$$

Exemple 2. L'espérance de la somme des lancers de deux dés à 6 faces vaut $7/2 + 7/2 = 7$.

3. Espérance d'une v.a. prenant un nombre dénombrable de valeurs

La propriété barycentrique de l'espérance tient encore lorsque la variable considérée prend un nombre dénombrable de valeurs. Mais, malheureusement, il peut arriver que le barycentre ne soit pas défini.

3.1. Contre-exemple. Rappelons un résultat d'analyse :

$$\sum_{n \geq 1} n^{-2} = \frac{\pi^2}{6}.$$

En particulier, il est possible de définir les poids de probabilités $(p_n)_{n \geq 1}$:

$$p_n = \frac{6}{\pi^2 n}.$$

Ils sont effectivement positifs de somme égale à 1.

Examinons maintenant la forme d'un éventuel barycentre associé à ces poids de probabilités. Il s'agit de calculer la somme

$$\sum_{n \geq 1} n p_n = \sum_{n \geq 1} \frac{6}{\pi^2 n}.$$

Mais, nous savons (d'un autre résultat d'analyse) que cette somme est divergente et vaut $+\infty$, de sorte que le barycentre n'est ici pas défini.

3.2. Convergence absolue. Pour garantir la bonne définition du barycentre, il est de fait nécessaire de garantir la convergence d'une somme. En réalité, nous conviendrons de vérifier la convergence absolue de cette somme pour définir le barycentre (la semi-convergence est en effet trop faible : la valeur du barycentre pourrait dépendre de l'ordre de sommation des éléments) :

DÉFINITION 3.1. *Une suite de réels* $(a_n)_{n\geq 1}$ *est dite absolument sommable si la série de termes positifs* $(|a_n|)_{n\geq 1}$ *est convergente.*

3.3. Définition de l'espérance. Nous arrivons à la définition :

DÉFINITION 3.2. *Etant donnés un espace de probabilité dénombrable* $(\Omega, \mathbb{P})$ *et une v.a.* X *prenant un nombre dénombrable de valeurs* $(x_n)_{n\geq 1}$*, l'espérance de* X *est bien définie si*

$$\sum_{n\geq 1} |x_n| \mathbb{P}\{X = x_n\} < +\infty.$$

Le cas échéant, elle est notée $\mathbb{E}(X)$ *et vaut :*

$$\mathbb{E}(X) = \sum_{n\geq 1} x_n \mathbb{P}\{X = x_n\}.$$

La variable X *est alors dite intégrable.*

Si Y *désigne une autre variable aléatoire intégrable, alors pour tous réels* λ *et* μ*,*

$$\mathbb{E}\big(\lambda X + \mu Y\big) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y).$$

4. Espérance d'une composition

4.1. Composition. Il s'agit de déterminer l'espérance d'une composition de la forme $f(X)$ où f est une fonction réelle de la variable réelle et X est une variable aléatoire.

Commençons par examiner le cas où X est une variable aléatoire prenant un nombre fini de valeurs. Alors, nous pouvons décomposer X de façon canonique sous la forme :

$$X = \sum_{i=1}^{n} x_i \mathbf{1}_{\{X = x_i\}},$$

où $x_1, \dots, x_n$ désignent les différentes valeurs de X. Assez clairement, lorsque $X = x_i$, $f(X)$ vaut $f(x_i)$ de sorte que

$$f(X) = \sum_{i=1}^{n} f(x_i) \mathbf{1}_{\{X = x_i\}}.$$

En utilisant la linéarité de l'espérance, nous en déduisons :

Proposition 4.1. *Etant donnés un espace de probabilité au plus dénombrable $(\Omega, \mathbb{P})$ muni d'une variable aléatoire X prenant un nombre fini de valeurs et une fonction f de $\mathbb{R}$ dans $\mathbb{R}$, l'espérance $\mathbb{E}[f(X)]$ de $f(X)$ vaut*

$$\mathbb{E}[f(X)] = \sum_{i=1}^{n} f(x_i)\mathbb{P}\{X = x_i\}.$$

Cette propriété reste vraie dans le cas dénombrable pour peu que le critère d'absolue convergence soit vérifié :

Proposition 4.2. *Etant donnés un espace de probabilité dénombrable $(\Omega, \mathbb{P})$ muni d'une variable aléatoire X prenant un nombre dénombrable de valeurs et une fonction f de $\mathbb{R}$ dans $\mathbb{R}$, l'espérance $\mathbb{E}[f(X)]$ de $f(X)$ existe si*

$$\sum_{n\geq 1} |f|(x_n)\mathbb{P}\{X = x_n\} < +\infty.$$

Le cas échéant, elle vaut

$$\mathbb{E}[f(X)] = \sum_{n\geq 1} f(x_n)\mathbb{P}\{X = x_n\}.$$

4.2. Inégalité triangulaire. Nous remarquons, en choisissant $f = |\cdot|$, que l'intégrabilité d'une variable aléatoire équivaut à l'intégrabilité de sa valeur absolue. Le cas échéant, il existe une comparaison naturelle entre les deux :

Proposition 4.3. *Soit X une variable aléatoire telle que $\mathbb{E}(|X|)$ existe. Alors, $\mathbb{E}(X)$ existe aussi et $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.*

Preuve. Dans le cas fini, les espérances existent trivialement. Dans le cas dénombrable, il s'agit de la remarque précédente. Ensuite, nous remarquons (par exemple dans le cas dénombrable) que

$$\begin{aligned}|\mathbb{E}(X)| &= \Big|\sum_{n\geq 1} x_n\mathbb{P}\{X = x_n\}\Big| \\ &\leq \sum_{n\geq 1} |x_n|\mathbb{P}\{X = x_n\}\Big| \\ &= \mathbb{E}(|X|).\end{aligned}$$

Ceci finit la preuve. $\square$

En particulier, nous comprenons que X et Y intégrables implique $X + Y$ intégrable.

4.3. Exemple du carré. Un exemple fréquemment utilisé est celui où f est la fonction carré. Dans le cas où X prend un nombre fini de valeurs, X^2 admet toujours une espérance. En revanche, si X prend un nombre dénombrable de valeurs, il est nécessaire de vérifier que X^2 est intégrable.

Nous dirons alors que X est intégrable. Un résultat célèbre est l'inégalité de Cauchy-Schwarz :

PROPOSITION 4.4. *Etant donnés un espace de probabilité dénombrable muni d'une variable aléatoire X prenant un nombre dénombrable de valeurs $(x_n)_{n\geq 1}$, X est intégrable si elle est de carré intégrable.*

Le cas échéant,

$$\mathbb{E}(|X|) \leq \left[\mathbb{E}(X^2)\right]^{1/2}.$$

Preuve. Nous supposons que

$$\sum_{n\geq 1} x_n^2 p_n < +\infty,$$

avec $p_n = \mathbb{P}\{X = x_n\}$. Alors,

$$\sum_{n\geq 1} |x_n| p_n \leq \left(\sum_{n\geq 1} |x_n|^2 p_n\right)^{1/2} \left(\sum_{n\geq 1} p_n\right)^{1/2} = \left(\mathbb{E}(X^2)\right)^{1/2}.$$

En particulier,

$$\mathbb{E}(|X|) \leq \left[\mathbb{E}(X^2)\right]^{1/2}.$$

Ceci achève la preuve. □

Ajouter que l'espérance est croissante.

4.4. Variance. Nous avons compris l'espérance comme une valeur moyenne. Une question naturelle est de mesurer la distance moyenne à la moyenne :

DÉFINITION 4.5. *Soit X une variable aléatoire de carré intégrable. Alors, on appelle variance de X la quantité*

$$\mathbb{V}(X) = \mathbb{E}\left[\left(X - \mathbb{E}(X)\right)^2\right] = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

La variance est toujours positive. Sa racine carrée s'appelle l'écart-type : il s'agit de la distance typique entre les valeurs prises par X et l'espérance de X.

Exemple. La variance du lancer d'un dé à 6 faces est donnée par :

$$\frac{1}{6}\sum_{i=1}^{6} i^2 - \frac{49}{4} = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

Attention, la variance n'est pas linéaire : $\mathbb{V}(\lambda X) = \lambda^2 \mathbb{V}(X)$.

4.5. Espérance et indépendance. Il se trouve qu'espérance et indépendance sont profondément reliées. Rappelons en effet que l'indépendance de deux événements s'écrit :

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

Dorénavant, nous savons écrire la probabilité d'un événement sous la forme d'une espérance. Nous pouvons écrire l'égalité ci-dessus comme :

$$\mathbb{E}\big[\mathbf{1}_{A\cap B}\big] = \mathbb{E}\big[\mathbf{1}_A\big]\mathbb{E}\big[\mathbf{1}_B\big]$$

Mais, nous remarquons maintenant que

$$\mathbf{1}_{A\cap B} = \mathbf{1}_A \mathbf{1}_B,$$

de sorte que l'indépendance de deux événements s'écrit également

$$\mathbb{E}\big[\mathbf{1}_A\mathbf{1}_B\big] = \mathbb{E}\big[\mathbf{1}_A\big]\mathbb{E}\big[\mathbf{1}_B\big].$$

En résumé, l'indépendance sépare les éspérances d'indicatrices.

Une question naturelle est de savoir si elle sépare les espérances en général. Voici la réponse à la question :

Proposition 4.6. *Etant données deux variables aléatoires X et Y définies sur un espace de probabilité au plus dénombrable $(\Omega, \mathbb{P})$ et deux fonctions f et g telles que*

$$\mathbb{E}\big[|f(X)|\big] < +\infty, \quad \mathbb{E}\big[|g(Y)\big] < \infty.$$

Si les variables aléatoires X et Y sont indépendantes alors le produit $f(X)g(Y)$ est intégrable et

$$\mathbb{E}\big[f(X)g(Y)\big] = \mathbb{E}\big[f(X)g(Y)\big].$$

En particulier, si X et Y sont elles-mêmes intégrables, l'espérance du produit XY est égal au produit des espérances.

Attention, l'hypothèse d'indépendance est fondamentale. Prenons par exemple le cas où $X = Y$ suit une loi uniforme sur $\{-1, 1\}$, alors $\mathbb{E}[XY] = 1 \neq 0 = \mathbb{E}(X)\mathbb{E}(Y)$.

Nous allons maintenant faire la preuve. Pour simplifier, nous allons nous restreindre au cas de variables prenant un nombre fini de valeurs.

Preuve. Supposons de fait que X soit à valeurs dans $\{x_1, \ldots, x_n\}$ et Y à valeurs dans $\{y_1, \ldots, y_m\}$. Alors,

$$\mathbb{E}\big[f(X)g(Y)\big] = \sum_{i=1}^{n}\sum_{j=1}^{m} f(x_i)g(y_j)\mathbb{P}\{X = x_i, Y = y_j\}.$$

En utilisant l'indépendance, il vient

$$\mathbb{P}\{X = x_i, Y = y_j\} = \mathbb{P}\{X = x_i\}\mathbb{P}\{Y = y_j\}.$$

En séparant les sommes, nous sommes capables de séparer les espérances.

Un corollaire important est le suivant

Par ailleurs,

Proposition 4.7. *La variance de la somme de n variables aléatoires indépendantes est égale à la somme des variances.*

Preuve. A détailler en cours.

4.6. Loi faible des grands nombres. Nous allons démontrer le phénomène qui nous a servi intuitivement à définir l'espérance : la répétition de l'expérience conduit à la connaissance.

Proposition 4.8. *Soient $X1, \ldots, X_n$ n variables aléatoires indépendantes de même loi de carré intégrable. Alors, pour tout $\varepsilon > 0$,*

$$\mathbb{P}\{\big|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\big| > \varepsilon\} \leq \frac{\mathbb{V}(X_1)}{\varepsilon^2 n}.$$

En particulier, lorsque n, la probabilité que la moyenne dite empirique et la vraie moyenne soient éloignées de plus de ε est très petite.

Preuve. La preuve est menée en deux temps. Nous commençons par calculer la variance de $\mathbb{V}(\bar{X}_n)$. Ensuite, il suffit d'appliquer l'inégalité de Bienaymé-Chebychev, dont la preuve est

$$\mathbb{E}[X^2] \geq \mathbb{E}[X^2\mathbf{1}_{\{X>a\}}] \geq a^2\mathbb{P}\{X > a\},$$

en utilisant la croissance de l'espérance.

Feuille de TD Numéro 4

1. Espérance

Exercice 1.1. On considère une variable aléatoire X à valeurs dans $\{1, \ldots, 5\}$ de loi

$$\mathbb{P}\{X = 1\} = 1/4,\ \mathbb{P}\{X = 2\} = 1/6,\ \mathbb{P}\{X = 3\} = 1/12,$$
$$\mathbb{P}\{X = 4\} = 1/3,\ \mathbb{P}\{X = 5\} = 1/6.$$

Calculer son espérance.

Exercice 1.2. Calculer l'espérance d'une loi uniforme sur $\{1, \ldots, N\}$, $N \geq 1$.

Exercice 1.3. Calculer de deux façons l'espérance de la somme des lancers de deux dés à 6 faces :

(1) en utilisant la loi de la somme des deux lancers,
(2) en utilisant la linéarité de l'espérance.

Exercice 1.4. Calculer de deux façons l'espérance d'une loi binomiale de paramètres $n \geq 1$ et $p \in [0, 1]$:

(1) en utilisant la relation d'absorption $kC_n^k = nC_{n-1}^{k-1}$, $1 \leq k \leq n$,
(2) en écrivant une variable de loi binomiale comme la somme de variables de Bernoulli.

Exercice 1.5. En utilisant la relation d'absorption rappelée dans l'exercice précédent, calculer l'espérance d'une loi hypergéométrique (N, n, p) avec $0 \leq n \leq N$ et $1 \leq p \leq N$. La comparer à l'espérance d'une binomiale.

Exercice 1.6. Soit X une variable à valeurs dans $\{0, \ldots, N\}$, $N \geq 1$. Montrer que

$$\mathbb{E}(X) = \sum_{k=0}^{N-1} \mathbb{P}\{X > k\}.$$

Application. Dans une urne contenant N boules numérotées de 1 à N, on tire avec remise n boules successives. Le numéro de la plus grande boule tirée est modélisé par une variable aléatoire X :

(1) Calculer $\mathbb{P}\{X > k\}$, pour $k \in \{0, \dots, N-1\}$. En déduire l'espérance de X.

(2) Donner la loi de X.

Exercice 1.7. Un joueur a $2^N - 1$ euros en poche, $N \geq 1$, et joue sur rouge/noir au casino. (La probabilité de gain est supposée égale à $1/2$.) Au premier coup, il mise un euro sur le noir. Si le noir sort, il empoche, en plus de sa mise, un euro et arrête de jouer. S'il perd, il rejoue en misant deux euros sur le noir. S'il gagne lors cette deuxième mise, il empoche, en plus de ses trois euros de mise, un euro et arrête de jouer. S'il perd, il rejoue en doublant sa mise, i.e. en misant quatre euros sur le noir. Et ainsi de suite...

(1) Expliquer pourquoi le joueur est sûr de pouvoir jouer N fois.

(2) Les issues des N lancers rouge/noir sont modélisés par N variables aléatoires de Bernoulli $X_1, \dots, X_N$. Décrire à partir de $X_1, \dots, X_N$ l'événement : « le joueur perd toute sa fortune au bout des N lancers ». Calculer sa probabilité.

(3) On désigne par S le gain algébrique empoché par le joueur au bout des N lancers. Montrer qu'il vaut 1 ou $-2^N + 1$. Donner la loi de S.

(4) Calculer l'espérance de S. Commenter.

Exercice 1.8. Une chaîne de frabrication produit des objets qui peuvent être défectueux, selon les hypothèses de modélisation suivantes :

(1) la chaîne produit en une heure un nombre aléatoire Y d'objets, Y à valeurs dans $\{0, \dots, N\}$, $N \geq 1$,

(2) chaque objet produit a une probabilité $p = 1/10$ d'être défectueux, indépendamment de tous les autres.

Le nombre d'objets défectueux produits en une heure par la chaîne de fabrication est modélisé par une variable aléatoire X.

(1) Pour deux entiers $k, n \geq 0$, calculer la probabilité conditionnelle $\mathbb{P}\{X = k | Y = n\}$. (Bien distinguer les cas $k \geq n$ et $k > n$.)

(2) Quelle est l'espérance de X sous la probabilité conditionnelle $\mathbb{P}(\cdot | Y = n)$, i.e. l'espérance de la loi conditionnelle de X sachant $Y = n$?

(3) En déduire que $\mathbb{E}(X) = p\mathbb{E}(Y)$.

Exercice 1.9. Montrer que la loi géométrique de paramètre $p \in]0, 1]$ est intégrable. Calculer son espérance.

Exercice 1.10. Montrer que la loi de Poisson de paramètre $\lambda > 0$ est intégrable. Calculer son espérance.

2. Espérance et indépendance. Variance.

Exercice 2.1. Soit X une variable aléatoire de loi uniforme sur le doublon $\{-N, N\}$, $N \geq 1$. Calculer l'espérance et la variance de X. Commenter.

Exercice 2.2. Calculer, par deux méthodes différentes, la variance d'une loi binomiale de paramètres n et p.

Exercice 2.3. Calculer la variance d'une loi uniforme sur $\{1, \ldots, N\}$. On rappelle que

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Exercice 2.4. Etant donnée une variable aléatoire X, on appelle fonction génératrice de X la fonction

$$\forall s \in [0,1], \quad G_X(s) = \mathbb{E}[s^X].$$

(1) Montrer que $G_X(s)$ est bien définie pour tout $s \in [0,1]$.
(2) Calculer $G_X(s)$ pour X de loi binomiale de paramètres $n \geq 1$ et $p \in [0,1]$, puis pour X de loi de Poisson de paramètre $\lambda > 0$.
(3) Montrer que $G_{X+Y}(s) = G_X(s)G_Y(s)$, $s \in [0,1]$, si X et Y sont deux variables aléatoires indépendantes.
(4) On suppose que X prend un nombre fini de valeurs. Montrer que G_X est dérivable en 1 : que vaut $G'_X(1)$?

Exercice 2.5. Montrer que la loi géométrique de paramètre $p \in]0,1]$ est de carré intégrable. Calculer sa variance.

Exercice 2.6. Montrer que la loi de Poisson de paramètre $\lambda > 0$ est de carré intégrable. Calculer sa variance.

Exercice 2.7. Un joueur décide d'aller jouer sur rouge/noir à la roulette d'un casino (probabilité 18/37 de gagner en raison du zéro) en s'imposant la règle suivante : une fois arrivé devant la table de jeu, il tirera au hasard uniforme entre 1 et N le nombre de coups qu'il jouera.

Le nombres de coups à jouer est modélisé par une variable aléatoire T et les issues des N lancers rouge/noir auquel le joueur participera éventuellement sont modélisés par N variables $X_1, \ldots, X_N$ à valeurs dans $\{0, 1\}$.

(1) Montrer que le gain du joueur s'écrit sous la forme

$$G = \sum_{i=1}^{T} (2X_i - 1).$$

(2) En décomposant T comme une combinaison d'indicatrices, calculer $\mathbb{E}(G)$ et $\mathbb{V}(G)$.

Exercice 2.8. Démontrer que la variance de la somme de n variables aléatoires indépendantes de carré intégrable est égale à la somme des n variances.

Exercice 2.9. Soient $X_1, \ldots, X_n$ n variables indépendantes de même loi de Bernoulli de paramètre $p \in [0, 1]$. Montrer que

$$\lim_{n \to +\infty} \mathbb{E}\big[\big|\frac{1}{n}\sum_{i=1}^{n} X_i - p\big|^2\big] = 0.$$

Exercice 2.10. A la sortie d'un restaurant, le garçon remet, à n hommes venus déjeuner ensemble, leurs chapeaux au hasard.

Les clients sont numérotés de 1 à n : pour le client i, on désigne par X_i la variable aléatoire valant 1 si le chapeau distribué est le bon et 0 sinon. Le nombre de chapeaux correctement distribués est $S = X_1 + \cdots + X_n$.

(1) Donner la loi de chaque X_i. En déduire l'espérance de S.

(2) Donner la valeur de $\mathbb{E}(X_i X_j)$ pour $i \neq j$. En déduire la variance de S.

CHAPITRE 5

Variables aléatoies continues

Jusqu'ici, les phénomènes que nous avons modélisé étaient supposés prendre un nombre fini ou au plus dénombrable de valeurs. En pratique, ceci se révèle en réalité insuffisant pour modéliser certaines expériences.

Donnons deux exemples simples :

(1) Supposons que nous cherchions à décrire, à l'aide d'une loi de probabilité, la taille d'un individu. En première approximation, nous pourrions nous contenter d'arrondir la taille au centimètre près de façon à modéliser le problème à l'aide d'une variable prenant des valeurs entières. Néanmoins, nous comprenons que cette démarche est limitée dès lors que nous cherchons à décrire plus finement la taille d'un individu.

(2) Supposons que nous cherchions à tirer un nombre au hasard entre 0 et 1. Là encore, nous pourrions, en première approximation, tirer au hasard en tronquant volontairement le tirage après 10 ou 20 chiffres après la virgule, mais cette démarche, bien qu'intéressante, laisse un goût d'inachevé.

Dans ce chapitre nous allons donc introduire des variables aléatoires prenant un nombre infini de valeurs.

1. Lois à densité

1.1. Notion d'histogramme. Supposons que nous recensions la taille dans une population de N individus et que nous représentions les valeurs obtenues à l'aide d'un histogramme : nous comptons combien de personnes ont une taille entre n et $n+1$ cm, pour n entier naturel, et nous représentons la proportion d'individus correspondants à l'aide d'un bâton dont la hauteur est égale à la proportion.

FIGURE 1. Histogramme de la taille d'une population de 10000 individus

Nous observons une certaine régularité de l'histogramme : son contour ressemble à une fonction régulière. Cette fonction contour permet de lire la proportion de personnes dont la taille est comprise entre n cm et m cm : il suffit de calculer la somme des aires des bâtons entre n et n, i.e. l'aire sous la courbe.

1.2. Densités. Le principe précédent est celui sur lequel nous allons fonder l'utilisation de variables aléatoires prenant un nombre continu de valeurs : au lieu de calculer la probabilité de prendre une valeur donnée, nous calculerons la probabilité que l'expérience soit entre deux valeurs données à l'aide d'une aire sous une courbe.

Pour cela, nous devons spécifier quelles courbes seront susceptibles d'être utilisées : nous voulons lire une probabilité comme une aire sous une courbe. Nous comprenons de fait deux choses :

(1) La courbe doit être à valeurs positives.

(2) L'aire totale sous la courbe doît être égale à 1 : cette aire représente la probabilité que l'expérience soit entre $-\infty$ et $+\infty$.

Nous obtenons donc comme définition :

DÉFINITION 1.1. *On appelle fonction de densité toute fonction f continue (ou continue par morceaux) positive et d'intégrale égale à 1, i.e.*

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

Donnons deux exemples

(1) $f(x) = \mathbf{1}_{]0,1[}(x)$, $x \in \mathbb{R}$,

(2) $f(x) = \lambda \mathbf{1}_{]0,+\infty[}(x) \exp(-\lambda x)$, $x \in \mathbb{R}$ et $\lambda > 0$.

1.3. Espace de probabilité sous-jacent. Nous voulons modéliser une expérience prenant un nombre continu de valeurs à l'aide d'une variable aléatoire. Pour cela, il serait nécessaire de définir un espace de probabilité Ω, muni d'événements et d'une mesure, et une application X, appelée variable aléatoire, sur cet espace et à valeurs dans $\mathbb{R}$.

Comme X est à valeurs continues, il est claire que Ω contient lui-même un nombre continu de valeurs. Il n'est de fait pas vraisemblable de représenter Ω comme un ensemble dénombrable et encore moins comme un ensemble fini. La difficulté, lorsque Ω est très gros, est de préciser ce que sont les événements : vu la quantité de points dans Ω, qu'est-il possible d'observer ? Mathématiquement, cette question se résume de la façon suivante : peut-on choisir $\mathcal{P}(\Omega)$ comme ensemble des événements. Essentiellement, il se trouve que la réponse est non : avec un tel choix, la théorie dit qu'il n'est pas (toujours) possible de construire une mesure de probabilité modélisant la réalité et mesurant toutes les parties. Il est donc nécessaire de restreindre les événements à une sous-famille d'ensembles.

En pratique, cette sous-famille d'ensemble est notée $\mathcal{A}$: elle porte le nom de tribus. Elle doit vérifier les mêmes axiomes de stabilité que dans le cas dénombrable :

DÉFINITION 1.2. *Une tribu est une collection $\mathcal{A}$ de parties de Ω telles que*

(1) $\Omega \in \mathcal{A}$

(2) $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$

(3) $(A_n)_{n\geq 0} \in \mathcal{A} \Rightarrow \cup_{n\geq 0} A_n \in \mathcal{A}$.

Une fois $\mathcal{A}$ donnée, il est nécessaire de mesurer les événements :

DÉFINITION 1.3. *Une mesure de probabilité sur $(\Omega, \mathcal{A})$ est une application de $\mathcal{A}$ dans $[0,1]$ telle que*

(1) $\mathbb{P}(\Omega) = 1$

(2) $(A_n)_{n\geq 0} \in \mathcal{A}$*,* $A_n \cap A_m = \emptyset$ *si* $n \neq m$*,*

$$\mathbb{P}\Big(\bigcup_{n\geq 0} A_n\Big) = \sum_{n\geq 0} \mathbb{P}(A_n).$$

En pratique, nous n'éssaierons pas de construire $\mathcal{A}$ et $\mathbb{P}$: il s'agit du programme de L3. Nous nous contenterons simplement d'admettre leur existence.

1.4. Variables aléatoires sur un espace continu. Une variable aléatoire sur un espace de probabilité $(\Omega, \mathcal{A})$ comme précédemment est non seulement une application X de Ω dans $\mathbb{R}$, mais aussi une application telles qu'il soit possible de déterminer la probabilité que X tombe dans un intervalle I (ouvert, fermé, semi-ouvert ou semi-fermé). Autrement dit

DÉFINITION 1.4. *Une variable aléatoire X sur $(\Omega, \mathcal{A})$ est une application de Ω dans $\mathbb{R}$ telle que pour tout intervalle $I \subset \mathbb{R}$, $\{\omega : X(\omega) \in I\} \in \mathcal{A}$, i.e. telle que l'on puisse écrire et calculer $\mathbb{P}\{X \in I\}$, où $\{X \in I\}$ est un raccourci de notations pour $\{\omega : X(\omega) \in I\}$.*

En pratique, nous ne construirons pas les variables aléatoires continues : nous admettrons leur existence.

1.5. Loi d'une variable aléatoire continue.

DÉFINITION 1.5. *Etant donnée une fonction de densité f, nous dirons que X est une variable aléatoire de loi de densité f si pour tout intervalle $I = [a,b]$, $]a,b[$, $[a,b[$ ou $]a,b]$, (avec a et b éventuellement infinis)*

$$\mathbb{P}\big\{X \in I\big\} = \int_a^b f(x)dx.$$

Dit autrement,

$$\mathbb{P}\big\{a \leq X \leq b\big\} = \int_a^b f(x)dx.$$

(Ici, il est possible de remplacer les inégalités larges par des inégalités strictes.)

Remarquons en particulier que $\mathbb{P}\{X = a\} = 0$ si X est à densité !

Exemple. Nous dirons que X suit une loi uniforme sur l'intervalle $]0,1[$ si X est une loi à densité de densité $\mathbf{1}_{]0,1[}$. Le cas échéant,

$$\mathbb{P}\{a \leq X \leq b = b - a,$$

si $0 \leq a <\leq b \leq 1$. Si $a \leq 0 \leq b \leq 1$, nous obtenons b. Si $0 \leq a \leq 1 \leq b$, nous obtenons $1 - a$. Enfin, si $a \leq 0 \leq 1 \leq b$, nous obtenons 1.

Il faut comprendre : là où l'indicatrice est nulle, il n'y a pas d'intégration.

Remarque : la loi uniforme sur l'intervalle $]0,1[$ est la loi du tirage au hasard d'un nombre entre 0 et 1 à l'aide d'un ordinateur. Par exemple, en Scilab, il s'agit de la fonction `rand`.

FIGURE 2. Histogramme d'un tirage de 10^6 valeurs entre 0 et 1

Exemple. Nous dirons que X suit une loi exponentielle de paramètre λ si X admet pour densité la fonction $x \mapsto \lambda \mathbf{1}_{]0,+\infty[}(x)\exp(-\lambda x)$.

1.6. Fonction de répartition. Usuellement, nous résumons la loi d'une variable aléatoire à densité en utilisant sa fonction de répartition :

DÉFINITION 1.6. *Soit X une variable aléatoire. Nous appelons fonction de répartion de X la fonction*

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}\{X \leq t\}.$$

Lorsque X est à densité et admet f pour densité, la fonction de répartion s'écrit

$$F_X(t) = \int_{-\infty}^t f(x)dx.$$

Exemple. La fonction de répartition de la loi exponentielle de paramètre λ est la fonction $F(t) = 1 - \exp(-\lambda t)$ pour $t \geq 0$ et 0 sinon.

Nous remarquons que la fonction de répartition est croissante, vaut 0 en $-\infty$ et 1 en $+\infty$. Par ailleurs, nous remarquons également que, dans le cas à densité, F_X est la primitive de f qui vaut 0 en $-\infty$: si f est continue, la dérivée de la fonction de répartition est la densité elle-même.

Enfin, nous admettrons que deux variables aléatoires ayant la même fonction de répartion ont la même loi.

1.7. Comment caractériser la loi d'une v.a. ? Il existe de fait plusieurs façons, équivalentes pour caractériser la loi d'une v.a. continues : il peut s'agir de calculer la fonction de répartition. Il peut aussi s'agir d'écrire,

$$\mathbb{P}\{a \leq X \leq b\} = \int_a^b f(x)dx,$$

où f est une fonction de densité. Donnons un exemple. Soit U une variable aléatoire de loi uniforme sur $]0,1[$, donner la loi de $-\ln(U)$. Nous calculons

$$\begin{aligned}\mathbb{P}\{a \leq -\ln(U) \leq b\} &= \mathbb{P}\{\exp(-b) \leq U \leq \exp(-a)\} \\ &= \int_{\exp(-b)}^{\exp(-a)} \mathbf{1}_{]0,1[}(x)dx \\ &= \int_{\max(\exp(-b),0)}^{\min(\exp(-a),1)} dx\end{aligned}$$

Pour x dans $]0,1[$, nous pouvons effectuer le changement de variable $x = -\ln(u)$ de sorte que

$$\begin{aligned}\mathbb{P}\{a \leq -\ln(U) \leq b\} &= \int_{\max(\exp(-b),0)}^{\min(\exp(-a),1)} dx \\ &= \int_{\min(a,0)}^{b} \exp(-u)du \\ &= \int_a^b \mathbf{1}_{]0,+\infty[}(u)\exp(-u)du.\end{aligned}$$

2. Calcul d'espérances

2.1. Définition. Dans le cas fini (ou dénombabrable), l'espérance d'une variable aléatoire X a été définie comme un barycentre pondéré par les poids de la loi de X. Précisément, si X est à valeurs dans $\{x_n,\ n \geq 0\}$ et si

$$\sum_{n\geq 0} |x_n|\mathbb{P}\{X = x_n\} < +\infty,$$

alors l'espérance de X vaut

$$\mathbb{E}[X] = \sum_{n\geq 0} x_n\mathbb{P}\{X = x_n\}.$$

La question se pose de savoir comment étendre cette définition au cas d'une variable continue de loi de densité f.

Pour comprendre la définition, nous utilisons l'image suivante. Nous savons que

$$\mathbb{P}\{a \leq X \leq b\} = \int_a^b f(x)dx.$$

Si a et b sont proches, la probabilité que $\mathbb{P}\{a \leq X \leq b\}$ doit être comprise comme la probabilité que X se trouve au voisinage de a. A ce moment là, $\mathbb{P}\{a \leq X \leq b\}$ joue le rôle d'un $\mathbb{P}\{X = x_n\}$ dans le cas dénombrable. L'équivalent du produit

$$x_n \mathbb{P}\{X = x_n\}$$

est donc

$$a \int_a^b f(x)dx$$

ou encore

$$\int_a^b x f(x)dx.$$

Si maintenant, nous sommons sur tous les découpages de la droite réelle, nous obtenons

Définition 2.1. *Soit X une variable aléatoire de loi de densité f. Si*

$$\int_{-\infty}^{+\infty} |x| f(x)dx < +\infty,$$

alors l'espérance de X est définie comme

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)dx.$$

Nous admettons que l'espérance ainsi construite est linéaire.

Il faut bien comprendre cette formule comme un barycentre continu : les poids sont microscopiques, de la forme $f(x)dx$. (Faire un dessin : la droite réelle est subdivisée en petits morceaux et les poids sont les aires microscopiques sous la densité.)

Exemple. Calculer l'espérance de U de loi uniforme sur $]0, 1[$. Nous obtenons

$$\mathbb{E}(U) = \int_0^1 u du = \frac{1}{2}.$$

2.2. Composition. Exactement, comme dans le cas dénombrable, nous pouvons nous interroger sur l'espérance d'une composition, i.e. d'une quantité de la forme $g(X)$ où g est une fonction continue (ou éventuellement continue par morceaux).

En réalité, tout se passe comme dans le cas fini, l'espérance s'écrivant comme le barycentre des valeurs de g contre les poids infinitésimaux $f(x)dx$. Nous obtenons

Proposition 2.2. *Soient X une variable aléatoire de loi de densité f et g une fonction continue par morceaux telle que*

$$\int_{-\infty}^{+\infty} |g(x)| f(x)dx < +\infty,$$

alors l'espérance de $g(X)$ existe et vaut

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

En particulier, si g est la fonction valeur absolue, alors

$$\mathbb{E}[|X|] = \int_{-\infty}^{+\infty} |x|f(x)dx.$$

Exemple. Calculer l'espérance de $\exp(U)$ si U est une loi uniforme sur $]0,1[$. Nous obtenons

$$\mathbb{E}[\exp(U)] = \int_0^1 \exp(u)du = \exp(1) - 1.$$

2.3. Variance. Pour mesurer la façon dont les valeurs d'une variable aléatoire X sont étalées autour de l'espérance (si elle existe), on utilise, comme dans le cas discret, la variance :

Définition 2.3. *Etant donnée une variable aléatoire X de loi de densité f, X est dite de carré intégrable si*

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f(x)dx < +\infty.$$

Le cas échéant, X est intégrable et la variance de X est la quantité

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}\big[(X - \mathbb{E}(X)^2)^2\big] \\ &= \int_{-\infty}^{+\infty} \Big(x - \int_{-\infty}^{+\infty} xf(x)dx\Big)^2 f(x)dx.\end{aligned}$$

La variance s'écrit

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X).$$

Exemple. Calculer la variance d'une loi uniforme sur $]0,1[$.

3. Lois gaussiennes

3.1. Loi gaussienne centrée réduite. Rappelons que la loi gaussienne centrée réduite est la loi de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{x^2}{2}\big), \quad x \in \mathbb{R}.$$

Voici sa représentation graphique

Figure 3. Densité gaussienne

Le nom de la gaussienne centrée réduite s'explique facilement : l'espérance vaut 0 et la variance vaut 1.

3.2. Gaussiennes générales. Etant donnée une variable aléatoire X de loi gaussienne centrée réduite, la variable aléatoire $m + \sigma X$, m réel et $\sigma > 0$, a pour espérance m et pour variance σ^2. Elle porte le nom de loi gaussienne de moyenne m et de variance σ^2. Il sera vu en TD que sa densité est la fonction

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\big(-\frac{(x-m)}{2\sigma^2}\big).$$

Le paramètre de moyenne m ne fait que translater la densité. Examinons l'influence du paramètre σ.

FIGURE 4. Evolution de la densité gaussienne avec la variance de 0.2 à 2 par pas de 0.2

3.3. Théorème Central Limite. L'introduction des gaussiennes est motivée par le résultat suivant que nous ne démontrerons pas.

THÉORÈME 3.1. *Soient* $(X_n)_{n\geq 1}$ *une collection de variables aléatoires de Bernoulli indépendantes de même paramètre* $0 < p < 1$ *(ici indépendant veut dire que* $X_1, \ldots, X_n$ *sont indépendantes pour tout* n*), alors pour* $a < b$*,*

$$\lim_{n\to+\infty} \mathbb{P}\big\{a \leq \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leq n\big\} = \mathbb{P}\{a \leq Z \leq b\},$$

où Z *suit une loi gaussienne centrée réduite.*

Autrement dit, lorsque l'on répète un grand nombre de lancers à pile ou face déquilibré de paramètre p, la distance entre la moyenne empirique et le paramètre est de l'ordre de $\sqrt{p(1-p)}/\sqrt{n}$. Précisément, pour $b = -a = 1,960$

$$\mathbb{P}\{a \leq Z \leq b\} \approx 0.95,$$

de sorte que, à 95% de chance, la distance entre la moyenne empirique et le paramètre est de l'ordre de $1,960\sqrt{p(1-p)}/\sqrt{n}$ lorsque n est grand.