

LE MANS UNIVERSITÉ

Licence SPI semestre 4

Introduction aux probabilités

Alexandre POPIER

Année 2017-2018

Table des matières

Introduction.	1
1 Espaces de probabilité	7
1.1 Espace de probabilité	7
1.1.1 L'espace des observables Ω	7
1.1.2 Les événements	8
1.1.3 La probabilité \mathbb{P}	8
1.2 Probabilités discrètes finies, équiprobabilité, cas dénombrable	12
1.2.1 Équiprobabilité sur les espaces finis	13
1.2.2 Extension au cas dénombrable	13
1.3 Indépendance et probabilités conditionnelles	14
1.3.1 Indépendance	14
1.3.2 Probabilités conditionnelles	15
2 Variables aléatoires discrètes	21
2.1 Définition	21
2.2 Loix discrètes classiques	23
2.2.1 Loi uniforme sur $\{1, \dots, N\}$	23
2.2.2 Loi de Bernoulli	23
2.2.3 Loi binômiale	24
2.2.4 Loi géométrique	24
2.2.5 Loi de Poisson	25
2.3 Moments	25
2.4 Médiane, écart moyen minimum	28
2.5 Caractéristiques des loix classiques	28
2.5.1 Loi uniforme sur $\{1, \dots, N\}$	28
2.5.2 Loi de Bernoulli	29
2.5.3 Loi binômiale	29
2.5.4 Loi géométrique	29
2.5.5 Loi de Poisson	30
3 Variables aléatoires à densité	31
3.1 Densité de probabilité	31
3.2 Fonction de répartition d'une v.a. réelle	34
3.3 Moments	36

3.4	Médiane	37
3.5	Fonctions caractéristiques	37
3.6	Lois continues classiques	37
3.6.1	Loi uniforme	37
3.6.2	Loi exponentielle	38
3.6.3	Loi normale (ou de Gauss)	39
3.7	Récapitulatif des lois continues	40
3.8	Correspondance v.a. discrète - v.a. continue	40
4	Vecteurs aléatoires, indépendance	41
4.1	Fonction de répartition, densités	41
4.2	Indépendance	44
4.3	Moments, corrélations	45
4.4	Extension en dimension d	46
4.4.1	Matrice de covariance	47
4.4.2	Indépendance	47
5	Convergence et théorèmes limites	49
5.1	Loi des grands nombres	49
5.1.1	Inégalités de Markov et de Bienaymé-Tchebychev	50
5.1.2	Probabilité, limite des fréquences	51
5.1.3	Loi (forte) des grands nombres	52
5.2	Théorème central limite	52
5.3	Méthode de Monte-Carlo	53
6	Annexe	55
6.1	Dénombrement : quelques rappels	55
6.1.1	Arrangements et permutations	55
6.1.2	Indiscernabilité des objets : combinaisons	56
6.2	Séries	57
6.2.1	Vocabulaire	57
6.2.2	Critères de convergence pour les séries à terme général positif	60
6.2.3	Méthodologie	63
6.3	Intégration sur un segment	63
6.3.1	Intégrale des fonctions en escalier	63
6.3.2	Fonctions intégrables (au sens de Riemann)	64
6.3.3	Primitives	66
6.3.4	Changement de variable et intégration par parties	68
6.3.5	Calcul de primitives	70
6.4	Intégrales généralisées	73
6.4.1	Définitions et premières propriétés	74
6.4.2	Intégrales de Riemann	77
6.4.3	Fonctions à valeurs positives	78
6.4.4	Fonctions à valeurs quelconques	82
6.4.5	Méthodologie	82

Introduction

Un cours de métrologie explique que le résultat d'une mesure correspond à une estimation de la valeur du mesurande plus une incertitude sur cette estimation :

$$G_{\text{résultat}} = G_{\text{mesure}} + \Delta G.$$

Une mesure est toujours une estimation parce qu'il existe différentes sources d'erreurs de mesure.

- Erreurs systématiques :
 - Erreurs dues aux imperfections de l'instrument ou du procédé de mesure.
 - Surestimation (ou sous-estimation) systématique de la valeur vraie du mesurande.
- Erreurs aléatoires :
 - Erreurs qui surviennent lors de la répétition d'une même mesure.
 - Surestimation (ou sous-estimation) aléatoire de la valeur vraie du mesurande.

Ainsi pour un appareillage, on parle de :

- *Justesse* : qualité de l'appareillage de mesure dont les erreurs systématiques sont faibles.
- *Fidélité* : qualité de l'appareillage de mesure dont les erreurs aléatoires sont faibles.
- *Précision* : qualité de l'appareillage de mesure à la fois juste et fidèle.
- *Répétabilité / reproductibilité* : étroitesse de l'accord entre les mesurages successifs du même mesurande, mesurages effectués dans la totalité avec les mêmes conditions de mesure (répétabilité) ou non (reproductibilité).

L'objet de ce cours est comprendre la signification du terme aléatoire, ainsi que de justifier les techniques utilisées l'an dernier comme :

- Pour réduire l'effet des erreurs aléatoires, on mesure N fois la grandeur et on moyenne.
- L'écart-type est représentatif d'un intervalle de confiance pour la valeur mesurée.

Mais l'utilisation des probabilités ou des statistiques ne s'arrête pas au problème de mesure. Voici d'autres domaines de la physique où l'aléatoire est une propriété des objets :

- physique statistique (Maxwell, Boltzmann, Van der Waals, etc.) ;
- mécanique quantique ;
- théorie du chaos.

De plus le hasard s'est invité aussi dans l'informatique :

- algorithmes stochastiques (recuit simulé, réseaux de neurones, etc.) ;
- modélisation des réseaux informatiques et de communication (Internet).

Une expérience - le vocabulaire

Il est simple de faire l'expérience suivante :

- découper des cartes de dimensions égales dans du carton de couleur (rouge, jaune, vert et bleu) sur lesquelles on inscrira des chiffres de 0 à 9 ;
- mettre les cartes dans une boîte et les tirer au hasard une à une en les remettant chaque fois dans la boîte.

On effectue une centaine de tirages en comptabilisant les résultats ; un résultat est le couple « couleur et chiffre ».

Il s'agit de **tirages avec remise** ou **tirages non exhaustifs** ; c'est la représentation d'une mesure physique ne détériorant pas le système étudié (mesure de la longueur d'une table) ou le perturbant peu (mesure d'une tension électrique). Les **tirages sans remise** (ou **tirages exhaustifs**), où le contenu de la boîte varie, peuvent représenter une expérience qui modifie, voire qui détruit le système (bombardement d'un noyau atomique). Les **effectifs** n_i obtenus pour $N = 100$ tirages sont données en gras dans le tableau ci-dessous.

Couleurs	Effectifs n_i (décompte des tirages)								Totaux horizontaux
Rouge	8								8
Jaune	11 9								20
Vert	17 11 6								34
Bleu	7 10 12 9								38
Nombre X	...	1	2	3	4	5	6	...	
Totaux verticaux		0	24	40	27	9	0		100

On y voit deux **entités aléatoires** : couleur et nombre. Le nombre peut symboliser le résultat d'une mesure et la couleur l'état du système physique. Si le système est en rouge, la mesure de X est 3 ; le résultat est moins bien déterminé s'il est en jaune, en vert ou en bleu. Si on photographie un pendule en mouvement, l'image est nette et la mesure précise en bout de course (angle maximal, vitesse faible) ; près de la verticale, par contre, l'image est floue.

L'entité aléatoire « couleur » apparaît sous 4 aspects : rouge, jaune, vert ou bleu. Dans la dernière colonne, on a décompté l'apparition de chaque couleur en sommant les valeurs de chaque horizontale. Ces valeurs donnent une idée du comportement aléatoire des couleurs *quels que soient les nombres* : il s'agit d'une des deux **répartitions marginales** (écrite dans la colonne de droite).

L'entité aléatoire X est une variable pouvant prendre plusieurs valeurs avec différentes fréquences d'apparition ; on l'appelle **variable aléatoire** (v.a. en abrégé). C'est un objet aléatoire important : le physicien caractérisant ses observations par des nombres, nous travaillerons surtout avec des v.a. Sur la dernière ligne du tableau, on a décompté l'apparition des valeurs de X : X est compris entre 2 et 5, la valeur 3 est la plus fréquente. Cette ligne donne la **répartition marginale** de X *quelles que soient les couleurs*. Avec la ligne précédente des valeurs de X , elle constitue le **tableau de distribution** de X . Évidemment la somme des valeurs de la dernière ligne ou de la dernière colonne est le nombre total de tirages, il est indiqué en bas à droite : $N = 100$.

En statistique on travaille plutôt avec les **fréquences** $f_i = n_i/N$. Elles sont indiquées dans le tableau appelé **tableau de contingence** (ici à deux entrées) des couleurs et des valeurs de X .

Couleurs	Fréquence f_i des tirages							Totaux horizontaux
Rouge	0,08							0,08
Jaune	0,11 0,09							0,20
Vert	0,17 0,11 0,06							0,34
Bleu	0,07 0,10 0,12 0,09							0,38
Nombre X	...	1	2	3	4	5	6	...
Totaux verticaux	0	0,24	0,40	0,27	0,09	0		1,00

On y retrouve les répartitions marginales des fréquences, la répartition en gras au cœur du tableau étant la **répartition conjointe** des fréquences. La somme des fréquences indiquée en bas à droite est 1 (100%) : on dit que les fréquences sont normalisées. La **normalisation** des fréquences (puis celle des probabilités) signifie que l'on a envisagé tous les résultats. C'est une propriété fondamentale : il ne faut jamais oublier de normaliser une répartition !

Examinons le tableau de répartition de X : il donne toutes les informations fournies par les tirages sur le comportement aléatoire de X .

Nombre X	...	1	2	3	4	5	6	...	
Répartition de la v.a. X		0	0,24	0,40	0,27	0,09	0		1,00

On remarque

1. qu'il y a beaucoup d'informations et
2. qu'une seconde série de tirages conduira à un tableau analogue difficile à comparer au premier.

Il est donc nécessaire de résumer ces données en quelques **caractéristiques numériques** permettant une comparaison rapide et efficace entre plusieurs tableaux. Les premières sont les **paramètres de centrage** donnant une valeur *typique* de X : par exemple la **moyenne**, définie par une méthode barycentrique qui associe à chaque valeur de X un poids statistique égal à sa fréquence (**moyenne arithmétique** ou **moyenne pondérée**), notée \bar{x} :

$$(1) \quad \bar{x} = \frac{1}{N} \sum_i n_i x_i = \sum_i f_i x_i = 3,21.$$

Il faut distinguer la moyenne arithmétique \bar{x} , déduite des fréquences (données expérimentales), de la **moyenne probabiliste** ou **espérance mathématique**, calculée de la même façon mais avec des probabilités, et notée $\mathbb{E}(X)$. Les fréquences étant des estimations des probabilités, \bar{x} est une estimation de $\mathbb{E}(X)$. On ne peut calculer $\mathbb{E}(X)$ que si l'on connaît les probabilités des valeurs de X , c'est-à-dire la **distribution de probabilité** de la v.a. X .

- En physique on associe une **incertitude** à une valeur estimée ; le tout donne un **intervalle d'incertitude** où se trouve « presque sûrement » la vraie valeur.
- En statistique, l'idée est identique mais les termes en sont différents et les notions plus précises. Il n'y a pas d'incertitude mais une **caractéristique de dispersion** de la distribution (ou **fluctuation**). On l'évalue à partir de la valeur typique des écarts $x_i - \bar{x}$ entre les valeurs observées et la valeur moyenne. Si on raisonne brutalement en moyennant ces écarts par pondération, on obtient zéro car les écarts négatifs compensent exactement les écarts positifs. On laisse au lecteur le soin de faire le calcul. Pour s'affranchir des signes, on peut raisonner sur les valeurs absolues $|x_i - \bar{x}|$ ou les carrés $(x_i - \bar{x})^2$.

La première méthode a des avantages, mais la seconde mène à des développements plus importants. On notera s^2 la moyenne des carrés des écarts, estimation de la **variance** de X , $\text{Var}(X)$ calculée, elle, avec des probabilités.

$$(2) \quad s^2 = \sum_i f_i (x_i - \bar{x})^2 = 0,83.$$

Pour des raisons de dimension, la dispersion de X est la racine carrée s ou **écart quadratique moyen** (EQM) estimation de l'**écart-type** de X , noté σ_X . Ici $s = 0,91$.

Il faut retenir tout de suite que l'écart-type σ_X ainsi que son estimation s sont une bonne évaluation de l'incertitude du physicien : $\sigma_X \approx s \approx \Delta X$. Nous verrons que la statistique est plus précise avec la notion de **probabilité de confiance**. On ne dira pas que la vraie valeur de X est « presque sûrement » dans $[\bar{x} - \Delta X, \bar{x} + \Delta X]$, mais on associera à l'intervalle $[\bar{x} - r\sigma_X, \bar{x} + r\sigma_X]$ (ou **domaine** ou **intervalle de confiance**) la probabilité $P(r)$ pour que la vraie valeur y soit. Le facteur r est de l'ordre de 1 (ses valeurs usuelles vont de 0,7 à 4 pour des probabilités de confiance de 50 à 99,99 %). Le point de vue statistique est donc plus prudent : il n'y a plus d'intervalle d'incertitude où l'on a toutes les chances de trouver la vraie valeur de la quantité mesurée (ce qui est faux en pratique) mais un intervalle de confiance où la probabilité de trouver la vraie valeur n'est jamais 100%.

Aspect probabiliste. Avant de passer à la partie formelle de ce chapitre, revenons sur la notion de **probabilité** dont nous avons parlé sans la définir. Il y a deux façons de l'aborder. Si on jette un dé non truqué, la probabilité de voir apparaître le 1 est $1/6$; le calcul est immédiat et « intuitif » : on compte le nombre de cas favorables et on le divise par le nombre de possibilités. On peut compliquer les exemples mais ce ne sera qu'une question technique de **dénombrement** basée sur des hypothèses (équilibrage d'un dé, discernabilité ou non-discernabilité d'objets, etc.), donc sur une **modélisation** du système permettant de calculer théoriquement des **fréquences idéales**. Cette *voie théorique* se base sur l'intuitive confusion des fréquences idéales et des probabilités mais n'apporte aucune explication immédiate au lien existant entre probabilités et fréquences observées.

L'autre façon d'aborder les probabilités est la *voie expérimentale* : quel est donc le lien entre fréquence (notion expérimentale) et probabilité (notion théorique) ? La réponse est que si l'on répète une infinité de fois le tirage, les fréquences observées tendront vers les probabilités. Si on lance le dé un fois, on obtiendra 2 ou non, et non le sixième

TABLE 1 – Tableau de contingence en probabilités

Couleurs	Probabilité p_i des tirages						Répartition marginale des couleurs
Rouge	0,10						0,10
Jaune	0,10 0,10						0,20
Vert	0,10 0,10 0,10						0,30
Bleu	0,10 0,10 0,10 0,10						0,40
Nombre X	... 1	2	3	4	5	6 ...	
Répartition de la v.a. X	0	0,20	0,40	0,30	0,10	0	1,00

de 2; mais si on recommence 60000 fois l'expérience, le nombre de sorties de 2 sera proche de 10000. Généralement cela est trop bien admis car cette stabilité, propriété très riche, n'est pas triviale. Vérifier qu'après un grand nombre d'épreuves et malgré le caractère aléatoire de l'expérience, la limite existe et reste implacablement stable est toujours surprenant. La probabilité apparaît donc comme une limite, et cette stabilité expérimentale lui confère son importance. Ceci est une conséquence de la **loi des grands nombres**. Mais paradoxalement il faut aborder les probabilités par la théorie, définir les outils pour finalement démontrer cette loi. L'introduction axiomatique des probabilités avant la démonstration de leur existence est une démarche déroutante constituant un édifice autocohérent essentiellement basé sur l'expérience.

Jusqu'ici on a fait des statistiques : on a estimé les probabilités, l'espérance et l'écart-type, car on ne connaissait pas le système : en expérimentateurs, on l'a mesuré pour mieux le connaître. La loi des grands nombres dit qu'avec une infinité de tirages on pourrait le connaître exactement et aborder le calcul des probabilités., mais dans la pratique on n'a pas assez de temps pour cela (ni d'argent car une expérience coûte cher). Par chance une seule indication supplémentaire suffit ici : il y a 10 cartes dans la boîte. D'où la composition du système : une carte « rouge et 3 » (R3), une J3 puis J4, V2, V3, V4, B2, B3, B4, B5. On connaît parfaitement le système, on peut faire des probabilités.

En effet on peut dire maintenant qu'on a « une chance sur dix » de tirer une des cartes de la boîte, et au cours de 100 tirages, chaque carte sera « raisonnablement » tirée 10 fois. Ce ne sera pas forcément le cas mais comme dans l'exemple du dé, plus le nombre tirages N sera grand, plus on se rapprochera en valeur relative de $N/10$ sorties de chaque carte. Le tableau ci-dessous est le **tableau de contingence** couleurs-nombres en termes de probabilités. La **répartition conjointe**, les **répartitions marginales** y sont des répartitions de probabilités. La moyenne de X calculée à partir d'elles est l'**espérance mathématique** $\mathbb{E}(X)$ (3,3 pour notre 3,21 expérimental), la moyenne des carrés écarts est la variance $\text{Var}(X)$ (0,81 pour 0,83) et sa racine carrée, l'**écart-type** σ_X (0,90 pour 0,91). La comparaison des valeurs montre que nos évaluations statistiques sont bonnes : les écarts relatifs avec les véritables valeurs de $\mathbb{E}(X)$, $\text{Var}(X)$

et σ_X sont de l'ordre de quelques centièmes. Ce n'est pas un hasard : on verra que les écarts relatifs des fréquences sont de l'ordre de $1/\sqrt{N}$ (ici 1/10). On peut vérifier la fluctuation en 1/10 en comparant fréquences et probabilités conjointes. De fait des sommations, les distributions marginales, et à plus forte raison les caractéristiques numériques, gommant partiellement les effets des fluctuations statistiques inhérents à toute étude d'**échantillon** limité, ce qui améliore d'autant leurs estimations (on parle de la **robustesse** des estimateurs des caractéristiques numériques).

Plan du cours.

Le chapitre 1 donne les définitions essentielles et les calculs des probabilités. C'est un chapitre essentiel pour ce cours. Les chapitres 2 et 3 sont consacrés aux variables aléatoires en dimension 1, soit discrètes (chapitre 2) et soit à densité (chapitre 3). Le chapitre 4 est l'extension de ces notions à des vecteurs aléatoires. Le chapitre 5 traite des deux principaux théorèmes de convergence : la loi des grands nombres et le théorème central limite, qui sont la base de très nombreuses applications : méthode de Monte Carlo, statistiques, etc.

Chaque chapitre comporte une partie d'exercices qui seront distribués séparément.

Enfin dans l'annexe (chapitre 6), on rappelle quelques résultats d'analyse sur le dénombrement, les séries et l'intégration. **Attention** : toutes ces notions seront utilisées dans les chapitres 1 à 5 et doivent être maîtrisées. Elles sont supposées connues et nous ne reviendrons pas dessus en cours ou en T.D.

Chapitre 1

Espaces de probabilité

« La théorie des probabilités en tant que discipline mathématique peut et doit être développée à partir d'axiomes de la même manière qu'en géométrie et en algèbre. »

A. N. Kolmogorov (1903-1987).

Le calcul des probabilités est la science qui modélise les expériences dont l'issue n'est pas prévisible *a priori*. Le jet d'un dé, le tirage du Loto sont des exemples classiques de ces expériences, dites aléatoires. Une modélisation implique une simplification des phénomènes observés dans ces expériences, mais cette simplification conduit à une quantification, donc à la possibilité de faire des calculs et de prédire. La modélisation du calcul des probabilités a été inventée par A. N. Kolmogorov dans un livre (intitulé « Les fondements de la théorie des probabilités ») paru en 1933.

Dans ce chapitre nous allons donner les définitions fondamentales des probabilités, ainsi que les bases du calcul des probabilités. L'accent sera mis sur le cas discret fini. Pour l'instant aucune connaissance particulière n'est exigée sur la notion de série.

1.1 Espace de probabilité

1.1.1 L'espace des observables Ω

Nous conviendrons qu'effectuer une expérience aléatoire, c'est sélectionner par un procédé quelconque un élément ω dans un ensemble Ω . Par exemple, jeter un dé revient à sélectionner un élément de $\Omega = \{1, 2, 3, 4, 5, 6\}$. Cet ensemble Ω est appelé l'**espace des observables** (ou **espace des issues** ou espace des événements élémentaires ou **univers**). Ses points ω sont appelés **possibles**, **aléas**, **éventualités** ou **observables** ou **événements élémentaires**.

Il est très important qu'il soit clairement défini afin de regrouper au moins tous les résultats possibles de l'expérience aléatoire.

Exemple 1.1 On considère un dé à 6 faces, numérotées de 1 à 6. L'ensemble Ω est l'ensemble des résultats possibles d'un lancer, soit $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Exemple 1.2 On considère le même dé, sauf que sur la sixième face le nombre 6 est remplacé par le 5. Il y a donc deux faces où 5 est inscrit. L'ensemble des résultats est différent : $\tilde{\Omega} = \{1, 2, 3, 4, 5\}$.

Exemple 1.3 Si on reprend l'exemple des cartes, on a :

$$\Omega = \{R3, J3, J4, V2, V3, V4, B2, B3, B4, B5\}.$$

1.1.2 Les événements

Le résultat d'une expérience aléatoire est par définition imprévisible. Toutefois, une fois fixé A un sous-ensemble de l'espace des observables Ω , on s'intéresse à la possibilité qu'a le résultat ω de l'expérience de tomber dans A . Les parties de Ω pour lesquelles on se pose ce genre de question sont appelées des **événements**. Un événement est donc un **sous-ensemble** ou une **partie** de Ω . On note $\mathcal{P}(\Omega)$ l'**ensemble des parties** de Ω et on supposera que toute partie de Ω est un événement¹.

Dans l'exemple 1.3, on peut s'intéresser au tirage d'une carte verte : cela est réalisé pour trois événements élémentaires : $V2$, $V3$ et $V4$. La probabilité d'obtenir $V = \{V2, V3, V4\}$ est donc $3/10$ dans notre exemple. On peut ne s'intéresser qu'à la sortie de la carte $B3$. Mais cet **événement élémentaire** sera représenté par le **sous-ensemble élémentaire** à un élément $\{B3\}$ et non son élément $B3$. En probabilité les événements/sous-ensembles jouent un rôle plus important que les possibles/éléments.

Le langage de la théorie des ensembles permet des calculs systématiques sur les événements. Toutefois, il faut savoir que le langage courant, que nous utilisons dans une première étape pour décrire des événements a sa traduction ensembliste.

Définition 1.1 (Opérations sur les ensembles)

- Ensemble Ω : **événement certain**,
- Ensemble vide \emptyset : **événement impossible**,
- **Réunion** de A et B , $A \cup B$: A ou B sont réalisés (« ou » non exclusif),
- **Intersections** de A et B , $A \cap B$: A et B sont réalisés,
- A et B sont **disjoints**, i.e. $A \cap B = \emptyset$: les événements A et B sont **incompatibles**,
- $\bar{A} = A^c = \Omega \setminus A$: événement **contraire** de A ou **complémentaire** de A .

Ainsi un possible ω est dans $A \cup B$ si ω est dans A ou dans B (ou les deux!). L'aléa ω est dans $A \cap B$ si ω est à la fois dans A et dans B .

1.1.3 La probabilité \mathbb{P}

Pour l'instant l'univers Ω , espace probabilisable, est une image incomplète de l'expérience, car il manque les probabilités.

Reprenons l'exemple 1.3. On probabilise l'univers Ω en ajoutant pour chaque événement élémentaire la valeur $1/10$. Ils ont donc tous la même probabilité : on dit qu'ils sont

1. Un des premiers points délicats de la théorie est qu'on ne doit pas considérer tous les sous-ensembles de Ω comme des événements, si l'univers n'est pas dénombrable. Il faut parler de **tribu**.

TABLE 1.1 – Tableau de contingence en probabilités

Couleurs	Probabilité p_i des tirages							
Rouge	2/20							
Jaune	1/20		1/10					
Vert	3/20	1/20	4/20					
Bleu	5/20	1/20	1/20	1/20				
Nombre X	...	1	2	3	4	5	6	...

équiprobables. Il y a **équiprobabilité** parce qu'on a admis qu'il n'y avait aucune raison de tirer une carte plutôt qu'une autre. De la même façon on admet l'équiprobabilité des six résultats du lancer d'un dé non truqué. Cette hypothèse fréquente en pratique est due à notre ignorance : « il n'y a pas de raison de privilégier l'un plutôt que l'autre ». On peut l'imposer a priori pour aborder un calcul direct, puis après informations supplémentaires, faire une correction a posteriori (voir section 1.3.2). L'équiprobabilité dans un univers *fini* permet d'obtenir l'expression bien connue de la probabilité d'un événement (cf. section 1.2.1).

L'équiprobabilité n'est pas la règle générale. Si on ne considère que la v.a. X de notre exemple, le nouvel univers se réduit à 4 éléments de probabilités différentes.

X	2	3	4	5
Probabilité	0,20	0,40	0,30	0,10

On peut reprendre l'exemple avec dans la boîte deux cartes $R3$ au lieu d'une, une $J3$, une $J4$, trois $V2$, une $V3$, quatre $V4$, cinq $B2$, une $B3$, une $B4$ et une $B5$, soit 20 cartes au total, conduisant à une nouvelle probabilisation de Ω . L'univers est identique mais les probabilités sont différentes. Plusieurs probabilisations sont donc possibles pour le même univers.

On quantifie désormais la possibilité qu'a le résultat ω de l'expérience de tomber dans un événement $A \in \mathcal{P}(\Omega)$. La puissance du calcul des probabilités réside dans le fait qu'il ne manipule pas directement le résultat ω de l'expérience aléatoire mais mesure les chances d'arrivée d'un événement A du au hasard. La définition qui suit est la définition mathématiquement rigoureuse d'une probabilité.

Définition 1.2 *Étant donné un espace d'observables Ω , une probabilité \mathbb{P} est une application de $\mathcal{P}(\Omega)$ dans $[0, 1]$, donc une fonction qui associe à tout événement A un nombre $\mathbb{P}(A)$ compris entre 0 et 1 appelé **probabilité** de A , et qui satisfait aux axiomes suivants :*

1. *l'événement certain est de probabilité 1 :* $\mathbb{P}(\Omega) = 1.$

2. **Axiome d'additivité dénombrable :**

pour toute suite $A_1, A_2, \dots, A_n, \dots$ d'événements qui sont deux à deux disjoints, c'est-à-dire tels que $A_k \cap A_j = \emptyset$ si $k \neq j$, alors la série $\sum_{k=1}^{\infty} \mathbb{P}(A_k)$ converge et a

pour somme $\mathbb{P}\left(\bigcup_{k \geq 1} A_k\right)$:

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \mathbb{P}\left(\bigcup_{k \geq 1} A_k\right).$$

Le couple (Ω, \mathbb{P}) est alors appelé un **espace de probabilité** ou **espace probabilisé**.

Avant de voir les propriétés conséquences de ces axiomes, faisons quelques remarques sur cette définition.

1. Une probabilité est une mesure ou un poids d'un événement. Elle se doit d'être **positive**.
2. La condition $\mathbb{P}(\Omega) = 1$ est une **normalisation**. La probabilité que n'importe quel aléa ω se réalise est 1.
3. L'axiome d'additivité dénombrable est plus compliqué à comprendre. C'est une **règle d'addition** pour les probabilités. Nous l'avons utilisé plusieurs fois en ajoutant des probabilités d'événements élémentaires pour obtenir celle d'un événement à plusieurs aléas. Ainsi dans l'exemple 1.3, la probabilité d'avoir $X = 2$ était 0,2 car cet événement correspond au sous-ensemble $\{V2, B2\}$. L'axiome est donc l'extension de cette règle aux sous-ensembles disjoints et en quantité dénombrable. Pour deux ensembles A et B , elle s'écrit :

$$A \cap B = \emptyset \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Voici quelques conséquences immédiates des axiomes.

Propriétés 1.1 Soit (Ω, \mathbb{P}) un espace de probabilité. Alors

1. $\mathbb{P}(\emptyset) = 0$.
2. Si A_1, A_2, \dots, A_n dans $\mathcal{P}(\Omega)$ sont deux à deux disjoints, alors

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n);$$

en particulier $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

3. Si A et B sont dans $\mathcal{P}(\Omega)$ et si $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
4. Si A et B sont dans $\mathcal{P}(\Omega)$, mais ne sont pas nécessairement disjoints, alors

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

Si les A_1, A_2, \dots, A_n dans $\mathcal{P}(\Omega)$ ne sont pas nécessairement deux à deux disjoints, alors $\mathbb{P}(A_1 \cup \dots \cup A_n) \leq \mathbb{P}(A_1) + \dots + \mathbb{P}(A_n)$.

5. Soit $(A_n)_{n \in \mathbb{N}}$ une partition finie (ou infinie dénombrable) de Ω (i.e. les $A_n \in \mathcal{P}(\Omega)$ sont deux à deux disjoints, et leur union est Ω). Alors pour tout $B \in \mathcal{P}(\Omega)$,

$$\mathbb{P}(B) = \sum_{n=0}^{+\infty} \mathbb{P}(B \cap A_n).$$

Pour le lecteur motivé, la preuve est donnée dans la suite.

Preuve.

1. L'axiome d'additivité dénombrable est applicable à la suite constante définie par $A_n = \emptyset$, qui est effectivement formée d'événements deux à deux disjoints. La série dont le terme général $\mathbb{P}(\emptyset)$ est constant, ne peut converger que si ce terme général est 0.
2. Sa première partie se démontre en appliquant l'axiome d'additivité dénombrable à A_1, A_2, \dots, A_n continuée par $\emptyset = A_{n+1} = A_{n+2} = \dots$, et en utilisant le point 1. Appliquer ça à $n = 2$, $A_1 = A$ et $A_2 = A^c$ fournit $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ en utilisant le premier axiome d'une probabilité.
3. On écrit $B = A \cup (B \setminus A)$ comme réunion de deux ensembles disjoints (notez que $B \setminus A = B \cap A^c$ est bien dans $\mathcal{P}(\Omega)$), et on applique le point précédent :

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

4. On écrit comme dans la démonstration précédente : $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$, $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B)$, puis on écrit $A \cup B = (A \cap B) \cup (B \setminus A) \cup (A \setminus B)$ comme réunion de trois ensembles deux à deux disjoints et on applique le point 2. :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \setminus B) \\ &= \mathbb{P}(A \cap B) + (\mathbb{P}(B) - \mathbb{P}(A \cap B)) + (\mathbb{P}(A) - \mathbb{P}(A \cap B)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \end{aligned}$$

Pour terminer le 4., on démontre le résultat par récurrence sur n . C'est trivial pour $n = 1$. Si c'est démontré pour n , appliquons la première partie de ce point 4. à $A = A_1 \cup \dots \cup A_n$ et à $B = A_{n+1}$. On obtient, à l'aide de l'hypothèse de récurrence

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B) \leq \left(\sum_{k=1}^n \mathbb{P}(A_k) \right) + \mathbb{P}(A_{n+1}).$$

5. Soient $B_n = A_n \cap B$. Les B_n sont deux à deux disjoints, avec $\bigcup_{n=0}^{+\infty} B_n = B$. Par σ -additivité,

$$\mathbb{P}(B) = \sum_n \mathbb{P}(B_n) = \sum_n \mathbb{P}(B \cap A_n).$$

□

Reprenons l'exemple 1.3 et illustrons ce qui précède sur quelques cas.

- **Union.** Intéressons nous au tirage d'une carte verte ou de 4. Cet événement est l'union des sous-ensembles $V = \{V2, V3, V4\}$ et de $\{X = 4\}$. La probabilité $\mathbb{P}(V \cup \{X = 4\})$ se calcule en comptant ici le nombre de cas favorables. La table donne la valeur 5/10. On vérifie ici que $\mathbb{P}(V \cup \{X = 4\})$ n'est pas égale à $\mathbb{P}(V) + \mathbb{P}(\{X = 4\})$. En additionnant les probabilités de V et de $\{X = 4\}$, on compte deux fois celle de l'intersection. Ainsi :

$$\begin{aligned} 5/10 &= \mathbb{P}(V \cup \{X = 4\}) = \mathbb{P}(V) + \mathbb{P}(\{X = 4\}) - \mathbb{P}(V \cap \{X = 4\}) \\ &= 3/10 + 3/10 - 1/10. \end{aligned}$$

- **Incompatibilité.** Les événements $J = \text{Jaune}$ et $X = 2$ sont incompatibles. Donc on vérifie bien que :

$$\mathbb{P}(J \cup \{X = 2\}) = \mathbb{P}(J) + \mathbb{P}(\{X = 2\}).$$

1.2 Probabilités discrètes finies, équiprobabilité, cas dénombrable

On considère d'abord un espace fini $\Omega = \{\omega_n, 1 \leq n \leq N\}$ à N éléments.

Soit $(p_n)_{1 \leq n \leq N}$ N nombres réels tels que

1. $p_n \geq 0$, pour tout n ,

2. $p_1 + \dots + p_N = \sum_{n=1}^N p_n = 1$.

Ces deux conditions impliquent en particulier que $0 \leq p_n \leq 1$ pour tout entier n .

L'application qui à tout événement A fait correspondre la valeur

$$\mathbb{P}(A) = \sum_{\substack{1 \leq n \leq N \\ \omega_n \in A}} p_n$$

est une probabilité sur Ω . En effet

$$\mathbb{P}(\Omega) = \sum_{\substack{1 \leq n \leq N \\ \omega_n \in \Omega}} p_n = \sum_{n=1}^N p_n = 1.$$

De plus si A et B sont disjoints, si $\omega_n \in A$, alors $\omega_n \notin B$ et vice-versa. Ainsi si ω_n est dans $A \cup B$, soit il est dans A , soit il est dans B , et :

$$\mathbb{P}(A \cup B) = \sum_{\substack{1 \leq n \leq N \\ \omega_n \in A \cup B}} p_n = \sum_{\substack{1 \leq n \leq N \\ \omega_n \in A}} p_n + \sum_{\substack{1 \leq n \leq N \\ \omega_n \in B}} p_n = \mathbb{P}(A) + \mathbb{P}(B).$$

Réciproquement si \mathbb{P} est une probabilité sur Ω , on calcule la probabilité d'un événement élémentaire, et on pose

$$p_n = \mathbb{P}(\{\omega_n\}) = \mathbb{P}(\omega_n).$$

Comme \mathbb{P} est une probabilité, pour tout entier n , $p_n \geq 0$ et

$$\sum_{n=1}^N p_n = \sum_{n=1}^N \mathbb{P}(\{\omega_n\}) = \mathbb{P}\left(\bigcup_{1 \leq n \leq N} \{\omega_n\}\right) = \mathbb{P}(\Omega) = 1.$$

Ce type de probabilités est appelé **probabilité discrète finie** (portée par les éléments ω_n pondérés par les poids p_n). Ainsi, si Ω est fini, toute probabilité est discrète. Elles sont représentées souvent sous forme de tableau, comme la table .

Exemple 1.4 Toujours pour le dé à six faces, $\Omega = \{1, 2, \dots, 6\}$. Toute probabilité sur Ω se représente par six nombres compris entre 0 et 1, p_1, p_2, \dots, p_6 , tels que $\sum_{i=1}^6 p_i = 1$. Si le dé est équilibré alors pour tout i , $p_i = 1/6$.

1.2.1 Équiprobabilité sur les espaces finis

Définition 1.3 (Équiprobabilité) Soient un entier $N \geq 1$ et $\Omega = \{\omega_1, \dots, \omega_N\}$ un ensemble fini. L'équiprobabilité sur Ω est défini comme la probabilité :

$$\mathbb{P}(A) = \frac{1}{N} \sum_{n=1}^N \mathbf{1}_A(\omega_n) = \frac{\text{card } A}{\text{card } \Omega}.$$

La fonction $\mathbf{1}_A$ vaut 1 si $\omega \in A$, 0 sinon. Cette probabilité ne privilégie aucun élément particulier de Ω ; tous les p_n sont égaux. C'est un **cas particulier** de probabilité discrète, et ce n'est pas un cas universel. Dans ce cas, le calcul des probabilités se ramène au calcul de cardinaux d'ensembles, donc au *dénombrement*. On renvoie à l'annexe pour un rappel sur les factoriels, les arrangements, les coefficients binomiaux et les calculs simples sur le dénombrement.

Exemple 1.5 On jette trois fois une pièce de monnaie parfaite. On peut représenter l'espace Ω comme l'ensemble des applications de $\{1, 2, 3\}$ (trois jets) dans $\{P, F\}$ ($P =$ pile, $F =$ face). Autrement dit, $\Omega = \{P, F\}^3$. Donc $N = \text{card } \Omega = 2^3 = 8$. Il est alors facile de voir par exemple que

$$\mathbb{P}(\text{on sort exactement une fois P}) = \frac{3}{8},$$

$$\mathbb{P}(\text{on sort au moins une fois P}) = 1 - \mathbb{P}(\text{on sort trois fois F}) = 1 - \frac{1}{8} = \frac{7}{8}.$$

Exemple 1.6 Le jeu du loto consiste à choisir 6 numéros distincts parmi $\{1, 2, \dots, 49\}$. On suppose que les boules qui portent les 49 numéros sont toutes parfaites. On ne s'intéresse qu'aux résultats des 6 boules. L'espace Ω est

$$\Omega = \{(a_1, a_2, \dots, a_6), 1 \leq a_i \leq 49, \text{ les } a_i \text{ sont deux à deux différents}\}.$$

Par exemple $(1, 2, \dots, 6) = (6, 5, \dots, 1)$. On a $N = \text{card } \Omega = C_{49}^6 = \binom{49}{6}$. Par conséquent,

$$\mathbb{P}(\text{on gagne le premier prix avec un bulletin}) = \frac{1}{C_{49}^6} \approx 7 \times 10^{-8}.$$

1.2.2 Extension au cas dénombrable

On suppose maintenant que Ω est un espace infini dénombrable, c'est-à-dire qu'il existe une bijection entre Ω et \mathbb{N} , l'ensemble des entiers naturels (positifs ou nuls). Autrement dit on peut numéroter les éléments de Ω :

$$\Omega = \{\omega_n, n \in \mathbb{N}\}.$$

Les ensembles dénombrables les plus connus sont \mathbb{N} , \mathbb{Z} (tous les entiers positifs ou négatifs), \mathbb{Q} (tous les nombres rationnels, quotient de deux entiers). Les ensembles \mathbb{R} et \mathbb{C} ne

sont pas dénombrables. Ces ensembles infinis vont être utilisés dans différents modèles et sont des cas limites de problèmes pratiques finis. Ainsi si on joue à pile ou face jusqu'à obtenir pile, le nombre de lancers est a priori un nombre entier positif. Mais aucune borne a priori ne peut être mise dessus. Donc Ω sera \mathbb{N}^* (il faut jouer au moins une fois).

Sur ces ensembles infinis dénombrables, $\mathcal{P}(\Omega)$ sera l'ensemble de tous les événements. Pour définir une probabilité \mathbb{P} sur $\Omega = \{\omega_n, n \in \mathbb{N}\}$, tout se passe comme si Ω était fini. À chaque élément ω_n de Ω , on associe un poids $p_n \in [0, 1]$ tel que

$$\sum_{n=0}^{+\infty} p_n = 1.$$

Cela signifie que la série de terme général $(p_n, n \in \mathbb{N})$ converge et la somme vaut 1. Alors l'application qui à tout événement A fait correspondre la valeur

$$\mathbb{P}(A) = \sum_{n \geq 0, \omega_n \in A} p_n$$

est une probabilité sur Ω . Et réciproquement toute probabilité sur un espace Ω infini dénombrable se "décomposera" comme précédemment. Tout se passe formellement comme si Ω était fini, sauf qu'on manipule des séries.

Terminons par un exemple. On considère $\Omega = \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ et pour $p \in]0, 1[$, on pose

$$\forall n \geq 1, \quad p_n = p(1-p)^{n-1} \in [0, 1].$$

Alors

$$\sum_{n \geq 1} p_n = \sum_{n \geq 1} p(1-p)^{n-1} = p \sum_{n \geq 1} (1-p)^{n-1} = p \sum_{\ell \geq 1} (1-p)^\ell = p \frac{1}{1-(1-p)} = 1.$$

On a utilisé un résultat à connaître sur les séries géométriques (voir annexe). Donc on a défini une probabilité sur $\Omega = \mathbb{N}^*$.

1.3 Indépendance et probabilités conditionnelles

1.3.1 Indépendance

Définition 1.4 (Indépendance de deux événements) Soient A et B dans $\mathcal{P}(\Omega)$. On dit qu'ils sont *indépendants* si

$$(1.1) \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Dans l'exemple de la table 1, les événements « carte jaune » et « $X = 3$ ou 5 » sont indépendants. En effet

$$\mathbb{P}(\text{Carte jaune}) = 2/10, \quad \mathbb{P}(X = 3 \text{ ou } 5) = 5/10, \quad \mathbb{P}(\text{Carte jaune et } X = 3 \text{ ou } 5) = 0, 1.$$

Il n'y a pas de raison à donner : il suffit de vérifier la relation (1.1).

Il faut souligner le rôle essentiel de \mathbb{P} . Si on change de probabilité, il n'y a pas de raison pour que des événements, qui étaient indépendantes, le restent sous la nouvelle probabilité. Ainsi si on prend la probabilisation donnée par la table 1.1.3, les deux événements « carte jaune » et « $X = 3$ ou 5 » ne sont pas indépendants :

$$\mathbb{P}(\text{Carte jaune}) = 3/20, \quad \mathbb{P}(X = 3 \text{ ou } 5) = 6/20, \quad \mathbb{P}(\text{Carte jaune et } X = 3 \text{ ou } 5) = 1/20.$$

Contrairement à l'incompatibilité, l'indépendance n'est pas une propriété intrinsèque des événements.

L'indépendance n'est pas transitive : si A est indépendant de B et B de C , A n'est pas forcément indépendant de C .

Propriétés 1.2 *Si A et B sont indépendants, alors A^c et B sont indépendants, de même que A et B^c ou que A^c et B^c .*

Plus généralement, si $\{A_n\}_{n \in \mathbb{N}}$ est une suite d'événements :

Définition 1.5 *$\{A_n\}_{n \in \mathbb{N}}$ sont indépendants si pour tout ensemble fini d'indices $I \subset \mathbb{N}$:*

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Par exemple pour trois événements $\{A, B, C\}$ il faut vérifier que

$$\begin{aligned} \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C), \quad \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C), \quad \text{et } \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

1.3.2 Probabilités conditionnelles

On s'intéresse au lancer successif de deux dés. On considère les deux événements suivants : A : « la somme des deux dés est égale à 11 » et B : « le premier dé donne 6 ». On vérifie que la probabilité de A est $1/18$.

Mais si on observe le premier dé et que l'événement B est réalisé, en quoi cela affecte la probabilité d'avoir une somme égale à 11? Autrement dit, sachant que B est réalisé, quelle est la probabilité de A ? Pour obtenir 11, le deuxième dé doit donner 5. Et ainsi la probabilité d'avoir A , sachant B , est de $1/6$.

Cet exemple motive la définition suivante.

Définition 1.6 *Soient deux événements A et B sur (Ω, \mathbb{P}) , avec $\mathbb{P}(B) > 0$. La **probabilité conditionnelle de A sachant B** est*

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

Si B ne se réalise jamais, i.e. $\mathbb{P}(B) = 0$, cette définition n'a aucun sens.

Théorème 1.1 *Si $\mathbb{P}(B) > 0$, alors la fonction $\mathbb{P}(\cdot|B) : \mathcal{P}(\Omega) \rightarrow [0, 1]$ est une probabilité sur Ω .*

Preuve. D'abord on va vérifier que cette fonction est correctement définie. Soit $A \in \mathcal{P}(\Omega)$. Comme $A \cap B \subset B$, on a $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$. Donc par définition,

$$0 \leq \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1.$$

Vérifions maintenant les deux axiomes d'une probabilité.

1. Si $A = \Omega$, $\Omega \cap B = B$. Donc $\mathbb{P}(\Omega|B) = 1$.
2. Soit $A_1, A_2, \dots, A_n \dots$ une suite d'événements de $\mathcal{P}(\Omega)$, deux à deux disjoints. On veut calculer

$$\mathbb{P}\left(\bigcup_{k \geq 1} A_k \middle| B\right). \text{ On a}$$

$$\begin{aligned} \mathbb{P}\left(\bigcup_{k \geq 1} A_k \middle| B\right) &= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left[\left(\bigcup_{k \geq 1} A_k\right) \cap B\right] = \frac{1}{\mathbb{P}(B)} \mathbb{P}\left[\bigcup_{k \geq 1} (A_k \cap B)\right] \\ &= \frac{1}{\mathbb{P}(B)} \sum_{k \geq 1} \mathbb{P}(A_k \cap B) = \sum_{k \geq 1} \frac{\mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} \\ &= \sum_{k \geq 1} \mathbb{P}(A_k|B). \end{aligned}$$

Le passage de la première ligne à la deuxième est justifié par le fait que les ensembles $A_k \cap B$ sont deux à deux disjoints.

□

On déduit de la définition et du théorème précédent quelques règles de calcul sur les probabilités conditionnelles.

Propriétés 1.3

1. Si $A \cap B = \emptyset$, $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C)$.
2. $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$.
3. $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

Proposition 1.1 Soit deux événements A et B sur (Ω, \mathbb{P}) , avec $\mathbb{P}(B) > 0$. A et B sont indépendants si et seulement si

$$\mathbb{P}(A|B) = \mathbb{P}(A),$$

ou encore avec $\mathbb{P}(A) > 0$, $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Dire que deux événements sont indépendants, c'est dire que la connaissance de l'un n'influence pas l'autre.

Exemple 1.7 On modélise le lancer de deux pièces semblables indépendantes par

$$\Omega = \{\omega = (\omega_1, \omega_2); \omega_i \in \{P, F\}\} \text{ et } \mathbb{P}(\omega) = \mathbb{P}_1(\omega_1)\mathbb{P}_1(\omega_2),$$

avec $\mathbb{P}_1(P) = 1 - \mathbb{P}_1(F) = p$. On s'imagine qu'on a lancé la première pièce et que $\omega_1 = P$. On associe à chaque issue du lancer de la deuxième pièce une probabilité (sachant que $\omega_1 = P$) :

$$\mathbb{P}(\omega_2 = P | \omega_1 = P) = \frac{\mathbb{P}(\{P, P\})}{\mathbb{P}(\{\omega; \omega_1 = P\})} = \frac{pp}{p(1-p) + pp} = p.$$

L'indépendance des deux pièces signifie que la connaissance du résultat du lancer de la première ne nous apprend rien sur le résultat du lancer de la seconde.

Exemple 1.8 On suppose qu'on a deux paires de pièces différentes : la 1ère paire consiste en deux pièces indépendantes et pile sort 99 fois sur 100 en moyenne ; la deuxième paire consiste en deux pièces indépendantes mais pile sort 1 fois sur 10 en moyenne. Un jeu consiste en : (i) choisir au hasard une d'entre les deux paires, (ii) lancer les deux pièces correspondantes. On modélise l'espace des issues d'un jeu par :

$$\Omega = \{\omega = (\omega_1, \omega_2, n); \omega_i \in \{P, F\}, n = 1, 2\},$$

et

$$\mathbb{P}(\omega_1, \omega_2, 1) = \frac{1}{2} \mathbb{P}_1(\omega_1) \mathbb{P}_1(\omega_2) \text{ et } \mathbb{P}_1(P) = \alpha = 0,99$$

et

$$\mathbb{P}(\omega_1, \omega_2, 2) = \frac{1}{2} \mathbb{P}_2(\omega_1) \mathbb{P}_2(\omega_2) \text{ et } \mathbb{P}_2(P) = \beta = 0,1.$$

On calcule ici

$$\begin{aligned} \mathbb{P}(\omega_2 = P | \omega_1 = P) &= \frac{\mathbb{P}(\{\omega; \omega_2 = P, \omega_1 = P\})}{\mathbb{P}(\{\omega; \omega_1 = P\})} \\ &= \frac{\mathbb{P}(P, P, 1) + \mathbb{P}(P, P, 2)}{\mathbb{P}(P, P, 1) + \mathbb{P}(P, F, 1) + \mathbb{P}(P, P, 2) + \mathbb{P}(P, F, 2)} \\ &= \frac{\alpha^2/2 + \beta^2/2}{\alpha/2 + \beta/2} = \frac{\alpha^2 + \beta^2}{\alpha + \beta} \end{aligned}$$

Par ailleurs si on ne savait rien sur la première pièce

$$\mathbb{P}(\omega_2 = P) = \mathbb{P}(\{P, P, 1\}, \{F, P, 1\}, \{P, P, 2\}, \{F, P, 2\}) = \frac{1}{2}(\alpha + \beta).$$

On note que $\mathbb{P}(\omega_2 = P | \omega_1 = P) > \mathbb{P}(\omega_2 = P)$ (resp. 0,91 et 0,54). Intuitivement, comme pile est sorti la 1ère fois, on a plus de chance que la paire choisie soit la paire 1.

Les probabilités conditionnelles sont à l'origine de deux formules.

Proposition 1.2 (Formule des probabilités totales) Soit $\{A_i\}_{i=1, \dots, n}$ une partition de Ω , telle que pour tout i , $\mathbb{P}(A_i) > 0$. Pour tout $B \in \mathcal{P}(\Omega)$,

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B | A_i) \mathbb{P}(A_i).$$

Cette première formule provient immédiatement de la formule :

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

De la formule des probabilités totales, on déduit la formule de Bayes.

Proposition 1.3 (Formule de Bayes) *De plus si $\mathbb{P}(B) > 0$, on a pour tout entier k :*

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

Elle consiste à « inverser » les probabilités conditionnelles et est à la base des statistiques dites bayésiennes. Voyons sur un exemple comment elle s'utilise.

Exemple 1.9 Soit une urne contenant des boules blanches ou noires. Pour en savoir plus sur son contenu, on effectue 6 tirages successifs avec remise qui donnent tous une boule blanche. Que peut-on dire de la composition de l'urne ?

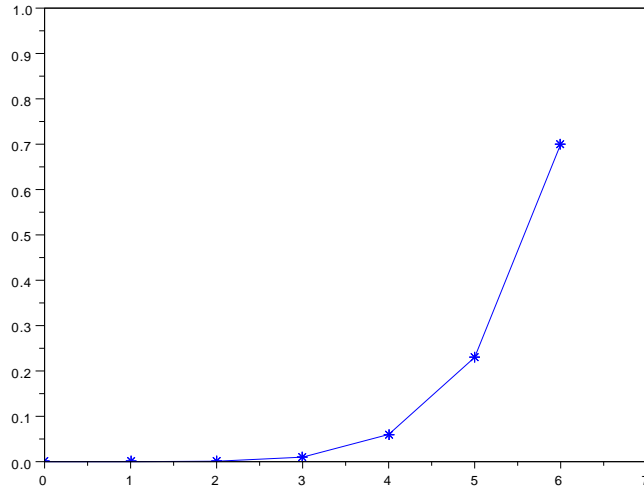
Si on sait que l'urne contient 6 boules au total, on a deux renseignements a priori (nombre et couleurs des boules) et une information a posteriori (résultat R : les 6 tirages ont tous donné une boule blanche). On peut imaginer les compositions possibles de l'urne (modélisation) : il y en a 7, depuis 0 boule blanche et 6 noires ($0B, 6N$) jusqu'à 6 boules blanches et 0 noire ($6B, 0N$), qui sont les états dits accessibles A_k . Il n'y a pas a priori de raison de privilégier l'un ou l'autre, d'où l'équiprobabilité $\mathbb{P}(A_k) = 1/7$. La probabilité de tirer une boule blanche dans une urne en contenant k est $k/6$; celle de tirer successivement 6 boules blanches est $(k/6)^6$: c'est $\mathbb{P}(R|A_k)$. Les probabilités a posteriori des compositions sont :

$$\mathbb{P}(A_k|R) = \frac{k^6}{\sum_{m=0}^6 m^6} = \frac{k^6}{67171}.$$

Ceci donne la table suivante :

Composition A_k	$0B, 6N$	$1B, 5N$	$2B, 4N$	$3B, 3N$	$4B, 2N$	$5B, 1N$	$6B, 0N$
k	0	1	2	3	4	5	6
$\mathbb{P}(A R)$	0	10^{-5}	10^{-3}	0,01	0,06	0,23	0,70

Et le graphe suivant :



Si maintenant le nombre total de boules M est inconnu, la composition A_k est

$$\{kB, (M - k)N\}.$$

Les probabilités $\mathbb{P}(A_k)$ sont inconnues mais l'équiprobabilité les rend inutile. La probabilité $\mathbb{P}(R|A_k)$ de tirer 6 boules blanches dans l'urne contenant k blanches sur M est $(k/M)^6$. Et la probabilité a posteriori cherchée est :

$$\mathbb{P}(A_k|R) = \frac{k^6}{f(M)}, \quad f(M) = \sum_{m=0}^M m^6.$$

$f(M)$ n'est pas connu, mais est indépendant de k . On a un renseignement relatif sur la composition de l'urne : les compositions contenant beaucoup de blanches sont favorisées (facteur k^6).

Chapitre 2

Variables aléatoires discrètes

2.1 Définition

Dans l'exemple du lancer de deux dés, avec $\Omega = \{1, \dots, 6\}^2$, l'éventualité $\omega = (i, j)$ représente l'éventualité « le premier dé donne i et le deuxième j ». Pour un ω donné, le résultat du premier dé est donc la première coordonnée de ω . Soit $X : \Omega \rightarrow \{1, \dots, 6\}$ l'application définie par

$$X((i, j)) = i.$$

Le résultat du premier dé pour l'éventualité ω est alors donné par $X(\omega)$. L'ensemble A de toutes les éventualités pour lesquelles le premier dé a donné 2, c'est-à-dire l'événement « le premier dé a donné 2 », est :

$$A = \{\omega \in \Omega, X(\omega) = 2\} = \{X = 2\} = X^{-1}(\{2\}) = \{(2, 1), \dots, (2, 6)\}.$$

Ainsi, si l'on ne s'intéresse qu'à un caractère de l'éventualité ω , cela revient à considérer une fonction $X(\omega)$, c'est-à-dire *une fonction du hasard*.

Un autre exemple de fonction du hasard est donné par le X de l'exemple 1.3.

Définition 2.1 *Soit E un ensemble quelconque. Si (Ω, \mathbb{P}) est un espace de probabilité, on appelle **variable aléatoire** définie sur (Ω, \mathbb{P}) et à valeurs dans E toute application X de Ω dans E ¹*

Pour $B \subset E$, l'événement $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ pourra être noté $\{X \in B\}$. Les ω sont souvent omis dans les notations probabilistes. Ainsi $\{\omega \in \Omega, X(\omega) \in B\}$ sera noté $\{X \in B\}$.

Cas particuliers :

- Si $E = \mathbb{R}$, on dit que X est une **variable aléatoire réelle** (en abrégé v.a.r.).
- Si $E = \mathbb{R}^d$, X est aussi appelée **vecteur aléatoire**, et dans ce cas, les composantes X_i de $X = (X_1, \dots, X_d)$ sont des v.a.r.

Exemple :

1. En théorie on doit ajouter application mesurable.

- Une fonction constante $f \equiv a$ est une variable aléatoire, même si dans ce cas il n'y a plus rien d'aléatoire dans la valeur de f .
- Si $A \in \mathcal{P}(\Omega)$ est un événement, alors la fonction indicatrice $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$ définie par

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon} \end{cases}$$

est une variable aléatoire réelle.

Définition 2.2 Soit $X : \Omega \rightarrow E$ une v.a. de l'espace de probabilité (Ω, \mathbb{P}) à valeurs dans un espace E . On définit l'application \mathbb{P}_X de $\mathcal{P}(E)$ dans $[0, 1]$ par :

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) \quad \forall B \in \mathcal{P}(E).$$

Alors \mathbb{P}_X est une probabilité sur E appelée loi de probabilité de X .

Ainsi, quand dans une expérience aléatoire on ne s'intéresse qu'à la variable aléatoire X , cela revient à se placer dans le nouvel espace fondamental (E, \mathbb{P}_X) . Ainsi, dans bien des cas, on ne précisera pas l'espace Ω choisi.

Exemple :

On lance deux dés. L'espace fondamental est $\Omega = \{\omega = (i, j) : 1 \leq i, j \leq 6\}$. Si on ne s'intéresse qu'à la somme $X(i, j) = X(\omega) = i + j$, on définit une variable aléatoire à valeurs dans $T = \{2, 3, \dots, 12\}$ et \mathbb{P}_X est définie par : $p_2 = 1/36, p_3 = 2/36, \dots, p_{11} = 2/36, p_{12} = 1/36$.

Exemple :

Si on reprend l'exemple 1.3, la loi de la variable aléatoire X est donnée par le tableau :

Nombre X	...	1	2	3	4	5	6	...
Répartition de la v.a. X		0	0,24	0,40	0,27	0,09	0	

Dans ce paragraphe, on va s'intéresser à un type particulier de variables aléatoires : celles dites **discrètes**, i.e. celles qui ne prennent qu'un nombre fini ou infini dénombrable de valeurs $\{x_i \in E, i \in \mathbb{N}\}$. Dans ce cas,

$$\forall A \in \mathcal{P}(E), \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{i=0}^{+\infty} \mathbb{P}(X = x_i) \mathbf{1}_A(x_i).$$

Soit $p_i = \mathbb{P}(X = x_i) \geq 0$. Comme $\mathbb{P}_X(E) = 1$, on a $\sum_{i=0}^{+\infty} \mathbb{P}(X = x_i) = \sum_{i=0}^{+\infty} p_i = 1$.

Réciproquement, soit une probabilité \mathbb{Q} sur E de la forme

$$\forall A \in \mathcal{P}(E), \mathbb{Q}(A) = \sum_{i=0}^{+\infty} p_i \mathbf{1}_A(x_i),$$

où les x_i sont des éléments de E et les p_i sont des poids positifs tels que $\sum_{i=0}^{+\infty} p_i = 1$.

Alors on peut montrer que \mathbb{Q} est la loi d'une v.a. X , portée par les x_i . En effet si A ne

contient aucun des x_i , $\mathbb{P}_X(A) = 0$, et $\mathbb{P}(X = x_i) = \mathbb{Q}(x_i) = p_i$: les valeurs prises par X sont les x_i . Autrement dit, **la loi de X est caractérisée par la donnée des (x_i, p_i)** . D'où la représentation

X	x_0	x_1	x_2	x_3	\dots	(valeurs prises par X)
\mathbb{P}_X	p_0	p_1	p_2	p_3	\dots	(probabilité)

Par ailleurs si un des p_i est nul, on « supprime » le x_i correspondant de la liste.

Exemple 2.1 On prend $E = \mathbb{R}$ et $x_0 = -1$, $x_1 = 0$, $x_2 = 3$, et $p_0 = 1/6$, $p_1 = 1/3$ et $p_2 = 1/2$. Dans ce cas

X	-1	0	3
\mathbb{P}_X	$1/6$	$1/3$	$1/2$

Exemple 2.2 Toujours avec $E = \mathbb{R}$, et pour tout $i \in \mathbb{N}$, $x_i = i$, $p_0 = 0$ et pour $i \geq 1$, $p_i = p(1 - p)^{i-1}$ avec $p \in [0, 1]$. Comme $\sum_{i=1}^{+\infty} p(1 - p)^{i-1} = 1$, c'est bien une loi de probabilité, dite loi géométrique de paramètre p . Elle est portée par les entiers strictement positifs.

2.2 Lois discrètes classiques

2.2.1 Loi uniforme sur $\{1, \dots, N\}$

X prend pour valeurs tous les entiers compris entre 1 et N .

Définition 2.3 X suit la loi uniforme si :

$$\forall 1 \leq k \leq N, \mathbb{P}(X = k) = 1/N.$$

Elle correspond à l'équiprobabilité du chapitre 1.

2.2.2 Loi de Bernoulli

On réalise une fois une expérience dont les résultats possibles sont un succès noté 1, ou un échec noté 0. Le cas typique est le jeu *pile* ou *face*. L'espace Ω_1 est

$$\Omega_1 = \{1; 0\}.$$

Si la pièce n'est pas truquée, pile et face ont la même chance de sortir. On associe la probabilité \mathbb{P}_1

$$\mathbb{P}_1(1) = \frac{1}{2}.$$

En général, on va supposer qu'une des faces est favorisée : il existe $p \in [0, 1]$ tel que

$$\mathbb{P}_1(1) = p \text{ et } \mathbb{P}_1(0) = 1 - p.$$

Les cas $p = 0$ et $p = 1$ n'ont aucun intérêt et on suppose donc que $0 < p < 1$. On définit la variable aléatoire $X : \Omega_1 \rightarrow \mathbb{R}$, telle que $X(1) = 1$ et $X(0) = 0$.

Définition 2.4 (loi de Bernoulli) X suit une loi de Bernoulli de paramètre p si :

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

Autrement dit, X ne prend que les valeurs 0 et 1, avec probabilité $1 - p$ et p .

2.2.3 Loi binômiale

L'exemple type d'un tel schéma consiste à lancer une pièce n fois et à spécifier l'état, *pile* noté 1 ou *face* noté 0, à chaque jet. X est la variable aléatoire qui compte le nombre de 1. Elle prend donc toutes les valeurs entières entre 0 et n , et suit comme loi

Définition 2.5 (loi binômiale) Une v.a. X à valeurs dans $\{0, \dots, n\}$ suit une loi binômiale de paramètres (n, p) , avec $n \geq 1$ et $p \in [0, 1]$, si :

$$\forall k = 0, \dots, n, \mathbb{P}_X(k) = \mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}.$$

On peut montrer que si X_1, \dots, X_n sont n variables de Bernoulli de paramètre p , définies sur un même espace de probabilité (Ω, \mathbb{P}) , et indépendantes, alors la somme $S_n = \sum_{i=1}^n X_i$ suit une loi binômiale de paramètres n et p . La réciproque est également vraie :

Théorème 2.1 B suit une loi binômiale de paramètres (n, p) si et seulement s'il existe n variables $(X_k)_{k=1, \dots, n}$ indépendantes et de Bernoulli de paramètre p telles que :

$$B = \sum_{k=1}^n X_k.$$

2.2.4 Loi géométrique

Un joueur procède à une suite de parties indépendantes de pile ou face et décide d'arrêter de jouer dès que, *pour la première fois* il aura amené pile. On s'intéresse au nombre X de parties qu'il lui faudra jouer pour réaliser son objectif. On est dans le cadre du schéma Succès-Échec infini du chapitre 2. Si $p \in]0, 1[$ est la probabilité d'obtenir pile en un lancer, la probabilité que $X = k$ est $p(1 - p)^{k-1}$. Si p est égal à 0, on n'obtiendra jamais pile, et si p est égal à 1, on obtient toujours pile au premier lancer. Donc ces deux cas n'ont plus rien d'aléatoire.

Définition 2.6 (loi géométrique) X suit une loi géométrique de paramètre $p \in]0, 1[$ dont la loi est donnée par :

$$\forall k \in \mathbb{N}, k \geq 1, \mathbb{P}(X = k) = p(1 - p)^{k-1}.$$

X prend donc toutes les valeurs entières supérieures ou égales à 1.

2.2.5 Loi de Poisson

La principale utilité de cette loi se trouve dans les modélisations de files d'attente, de réseaux, de fiabilité de machines. C'est aussi une approximation de la loi binômiale quand n est grand.

Définition 2.7 La probabilité de Poisson (*S. Poisson, 1781-1840*) sur \mathbb{N} est définie par :

$$\forall n \in \mathbb{N}, p_n = \frac{e^{-\lambda} \lambda^n}{n!}.$$

où λ est un paramètre réel strictement positif. Une v.a.r. X suit une loi de Poisson de paramètre $\lambda > 0$, si X ne prend que des valeurs entières et

$$\forall n \in \mathbb{N}, \mathbb{P}_X(n) = \frac{e^{-\lambda} \lambda^n}{n!}.$$

Théorème 2.2 (approximation de la loi binômiale) Si n tend vers $+\infty$ et p_n est tel que np_n tend vers $\lambda \neq 0$, si B suit une loi binômiale de paramètres n et p_n , alors :

$$\forall k \geq 0 \quad \lim_{n \rightarrow +\infty} \mathbb{P}(B = k) = \mathbb{P}(P = k),$$

où P suit une loi de Poisson de paramètre λ .

On utilise ce théorème pour faire le calcul approché suivant : si B suit une loi binômiale de paramètres n et p , avec les conditions suivantes : $n > 30$ et $np \leq 5$, alors pour $\lambda = np$:

$$\mathbb{P}(B = k) \simeq \frac{\lambda^k}{k!} \exp(-\lambda).$$

Remarque. Plus p est petit, meilleure est l'approximation. Plus généralement, pour B_i , pour $i = 1, \dots, r$, des v.a. indépendantes, de loi binômiale de paramètres respectifs (n_i, p_i) , si les p_i sont « assez petits », si on pose : $\lambda = n_1 p_1 + \dots + n_r p_r$, et si P est une v.a. de loi de Poisson de paramètre λ , on a :

$$\forall k \geq 0, \mathbb{P}(B_1 + \dots + B_r = k) \simeq \mathbb{P}(P = k).$$

2.3 Moments

Définition 2.8 (Espérance) L'espérance d'une v.a.r. discrète X de la loi \mathbb{P}_X définie par (x_i, p_i) est définie par

$$\boxed{\mathbb{E}(X) = \sum_{i=0}^{+\infty} x_i p_i, \text{ si } \mathbb{E}|X| = \sum_{i=0}^{+\infty} |x_i| p_i < +\infty.}$$

Remarque. Si X ne prend qu'un nombre fini de valeurs, la condition précédente est toujours vérifiée.

Propriétés 2.1 Soient X et Y deux variables aléatoires. On a les propriétés :

1. $\mathbb{E}(X)$ finie si et seulement si $\mathbb{E}(|X|)$ finie ;
2. $|X| \leq Y$ et $\mathbb{E}(Y)$ finie entraînent $\mathbb{E}(X)$ finie ;
3. $-\infty \leq a \leq X \leq b < +\infty \implies a \leq \mathbb{E}(X) \leq b$;
4. si $\mathbb{P}(X = a) = 1$, alors $\mathbb{E}(X) = a$;
5. $\mathbb{E}(X)$ finie implique $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

Le théorème suivant consigne les propriétés usuelles de l'espérance mathématique.

Théorème 2.3 Soient X et Y deux variables aléatoires discrètes. Si $\mathbb{E}|X| < +\infty$ et $\mathbb{E}|Y| < +\infty$, alors on a les propriétés :

A. Linéarité

- (A1.) $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
 (A2.) $\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$ pour tout $\lambda \in \mathbb{R}$.

B. Monotonie

- (B1.) $X \geq 0 \implies \mathbb{E}(X) \geq 0$;
 (B2.) $X \geq Y \implies \mathbb{E}(X) \geq \mathbb{E}(Y)$;
 (B3.) $\mathbb{P}(X = Y) = 1 \implies \mathbb{E}(X) = \mathbb{E}(Y)$.

Plus généralement pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$, on peut définir :

$$\mathbb{E}(g(X)) = \sum_{i \in I} g(x_i) p_i,$$

(si le second membre converge absolument). L'espérance mathématique d'une variable aléatoire X ne dépend que de la loi de X et indique la *valeur moyenne* autour de laquelle X prend ses valeurs.

On introduit d'autres caractéristiques de la loi de X qui rendent compte de la *dispersion* de cette loi, par exemple les *moments*. Débutons par un lemme qui permet de comparer les moments de différents ordres.

Lemme 2.1 Soient r et r' deux nombres réels tels que $0 < r < r'$ et X une variable aléatoire réelle. Si $\mathbb{E}(|X|^{r'})$ est fini, $\mathbb{E}(|X|^r)$ est aussi fini.

Définition 2.9 (Moments) Soit X une v.a.r. discrète, de loi \mathbb{P}_X donnée par (x_i, p_i) . Soit a et r deux nombres réels. Si $\mathbb{E}(|X - a|^r)$ est fini, alors le **moment d'ordre r** de X centré en a est défini par :

$$m_r^a = \mathbb{E}((X - a)^r) = \sum_{i \in I} p_i (x_i - a)^r.$$

Le moment d'ordre r (centré en 0) est défini par :

$$m_r = \mathbb{E}(X^r).$$

De même, si $\mathbb{E}|X|$ est fini, ainsi que $\mathbb{E}(|X - \mathbb{E}X|^r)$, le moment centré d'ordre r (à la moyenne) est défini par :

$$\mu_r = \mathbb{E}[(X - \mathbb{E}X)^r].$$

Si $r = 1$, alors $m_1 = \mathbb{E}X$ et $\mu_1 = 0$. Pour $r = 2$, le moment centré μ_2 est encore appelé variance de X et noté

$$\boxed{\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2].}$$

La variance est donc une quantité toujours positive. Sa racine carrée positive est appelée *écart-type* et notée $\sigma(X)$.

Propriétés 2.2 Soit X une v.a. ayant un moment d'ordre 2, et soient a et b deux réels. Alors $aX + b$ admet aussi un moment d'ordre 2, avec

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Lemme 2.2 Les variables aléatoires $X - \mathbb{E}X$ et $(X - \mathbb{E}X)/\sigma(X)$ sont respectivement centrée et centrée réduite, car

$$\mathbb{E}(X - \mathbb{E}(X)) = 0, \text{ et } \mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = 0, \text{ Var}\left(\frac{X - \mathbb{E}(X)}{\sigma(X)}\right) = 1.$$

Proposition 2.1 Une v.a.r. X a un moment d'ordre deux $\mathbb{E}(X^2)$ fini si et seulement si son espérance mathématique $\mathbb{E}X$ et sa variance $\text{Var}(X)$ existent et sont finies, et

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

Preuve. En effet,

$$\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - 2\mathbb{E}(X\mathbb{E}X) + (\mathbb{E}X)^2 = \mathbb{E}X^2 - 2(\mathbb{E}X)(\mathbb{E}X) + (\mathbb{E}X)^2.$$

□

Proposition 2.2 Soit X une variable aléatoire ayant un moment d'ordre 2. Alors

$$\mathbb{E}|X - \mathbb{E}(X)| \leq \sqrt{\text{Var}(X)} = \sigma.$$

Ainsi l'espérance est une caractéristique de position, alors que la variance (ou l'écart-type) en est une de dispersion.

Propriétés 2.3 Une variable aléatoire est constante avec probabilité 1 si et seulement si sa variance est nulle. La v.a. est alors égale à sa moyenne avec probabilité 1.

Ainsi à moins de travailler avec une constante, la variance est toujours strictement positive!

2.4 Médiane, écart moyen minimum

Nous introduisons maintenant une nouvelle caractéristique, appelée médiane, qui a l'avantage, sur l'espérance, d'exister pour toute variable aléatoire.

Définition 2.10 Soit X une variable aléatoire réelle. On appelle médiane de X tout nombre M vérifiant :

$$\mathbb{P}(X \leq M) \geq \frac{1}{2}, \quad \mathbb{P}(X \geq M) \geq \frac{1}{2}.$$

Il résulte immédiatement de la définition que si M est un médiane de X , on a :

$$\mathbb{P}(X \leq M) \geq \frac{1}{2} \geq \mathbb{P}(X > M), \quad \mathbb{P}(X \geq M) \geq \frac{1}{2} \geq \mathbb{P}(X < M).$$

Remarque. Toute v.a. X admet au moins une médiane. Elle peut en admettre plusieurs qui jouent toutes le même rôle. Si la fonction de répartition est continue et strictement croissante, alors X admet une médiane et une seule et l'on a $F(M) = 1/2$.

Théorème 2.4 Soit X une v.a. vérifiant $\mathbb{E}|X| < +\infty$. Si M est une médiane de X , alors pour tout nombre réel a , on a l'inégalité :

$$\mathbb{E}(|X - a|) \geq \mathbb{E}(|X - M|).$$

Donc si X est une v.a. vérifiant $\mathbb{E}|X| < +\infty$, l'expression $\mathbb{E}(|X - M|)$ prend la même valeur pour toute médiane M de X . Ceci justifie la définition suivante :

Définition 2.11 Soit X une v.a. vérifiant $\mathbb{E}|X| < +\infty$. Alors $\mathbb{E}(|X - M|)$ est appelé écart moyen minimum ou écart médian de X .

Si une médiane M est une caractéristique de position, on lui associe une caractéristique de dispersion, comme l'espérance et la variance.

Proposition 2.3 L'écart médian est majoré par l'écart-type.

Preuve. En effet on a $\mathbb{E}|X - \mathbb{E}X| \leq \sqrt{\text{Var}(X)}$ et $\mathbb{E}|X - M| \leq \mathbb{E}|X - \mathbb{E}X|$. □

2.5 Caractéristiques des lois classiques

2.5.1 Loi uniforme sur $\{1, \dots, N\}$

Si X suit une loi uniforme sur $\{1, \dots, N\}$, alors

Lemme 2.3 $\mathbb{E}(X) = (N + 1)/2$ et $\text{Var}(X) = \frac{N^2 - 1}{12}$.

De plus

Lemme 2.4 si N est impair, $(N + 1)/2$ est la médiane, tandis que si N est pair, tous les réels compris strictement entre $N/2$ et $(N/2) + 1$ sont des médianes.

2.5.2 Loi de Bernoulli

Lemme 2.5 *Si X suit une loi de Bernoulli de paramètre p , alors $\mathbb{E}(X) = p$ et $\text{Var}(X) = p(1 - p)$.*

Exercice 2.1 *Montrer que X a des moments de tout ordre et vérifier les résultats de la proposition précédente.*

Exemple important : ce qui suit est un exemple particulièrement important de loi de Bernoulli. Si (Ω, \mathbb{P}) est un espace de probabilité, et si $A \in \mathcal{P}(\Omega)$, on pose $X = \mathbf{1}_A$. C'est une variable aléatoire discrète de Bernoulli avec $p = \mathbb{P}(A)$. Ainsi on obtient

$$\mathbb{E}X = \mathbb{E}\mathbf{1}_A = \mathbb{P}(A).$$

Lemme 2.6 *Si X suit une loi de Bernoulli de paramètre p , alors*

- *si $p < 1/2$, 0 est la médiane de X ,*
- *si $p > 1/2$, 1 est la médiane de X ,*
- *si $p = 1/2$, tous les nombres compris dans l'intervalle $[0, 1]$ sont des médianes de X .*

2.5.3 Loi binômiale

Rappelons que si X suit une loi binômiale de paramètres (n, p) , alors il existe n v.a. de Bernoulli X_i indépendantes (voir paragraphe 4.2) de même paramètre p telles que $X = \sum_{i=1}^n X_i$. Ceci permet de calculer très facilement l'espérance et la variance d'une loi binômiale, en utilisant les propriétés de l'espérance et de la variance de v.a. indépendantes.

Lemme 2.7 *Si X suit une loi binômiale de paramètres n et p , alors $\mathbb{E}(X) = np$ et $\text{Var}(X) = np(1 - p)$.*

2.5.4 Loi géométrique

Pour calculer l'espérance et la variance d'une telle v.a., rappelons les formules suivantes. Soit q un réel tel que $0 < q < 1$. Alors

$$\begin{aligned} \sum_{i=0}^n q^i &= \frac{1 - q^{n+1}}{1 - q} \\ \sum_{i=0}^{+\infty} q^i &= \frac{1}{1 - q} \\ \sum_{i=1}^{+\infty} i q^{i-1} &= \frac{1}{(1 - q)^2}. \end{aligned}$$

Ici la condition de convergence dans la définition de l'espérance prend tout son sens.

Lemme 2.8 *Si X suit une loi géométrique de paramètre $p \in]0, 1[$, alors $\mathbb{E}(X) = 1/p$ et $\text{Var}(X) = \frac{1-p}{p^2}$.*

Exercice 2.2 *Redémontrer le lemme précédent et vérifier les résultats sur l'espérance et la variance de X .*

2.5.5 Loi de Poisson

Lemme 2.9 *Si X suit une loi de Poisson de paramètre λ , alors $\mathbb{E}(X) = \text{Var}(X) = \lambda$.*

Chapitre 3

Variables aléatoires à densité

Rappelons les définitions suivantes.

Définition 3.1 Soit E un ensemble quelconque. Si (Ω, \mathbb{P}) est un espace de probabilité, on appelle **variable aléatoire** définie sur (Ω, \mathbb{P}) et à valeurs dans E toute application X de Ω dans E ¹

Pour $B \subset E$, l'événement $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$ pourra être noté $\{X \in B\}$.

Définition 3.2 Soit $X : \Omega \rightarrow E$ une v.a. de l'espace de probabilité (Ω, \mathbb{P}) à valeurs dans un espace E . On définit l'application \mathbb{P}_X de $\mathcal{P}(E)$ dans $[0, 1]$ par :

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) \quad \forall B \in \mathcal{P}(E).$$

Alors \mathbb{P}_X est une probabilité sur E appelée loi de probabilité de X .

Définition 3.3 On dira que deux variables aléatoires X et Y ont même loi, ou sont identiquement distribués, si leurs lois \mathbb{P}_X et \mathbb{P}_Y sont des probabilités identiques. Cette relation sera notée

$$X \sim Y.$$

Dans le chapitre précédent, les variables aléatoires ne prenaient qu'un nombre au plus dénombrable de valeurs et ainsi leur loi était décrite par des poids associés à chaque valeur possible de la variable aléatoire. Ici la loi de la variable aléatoire va être décrite par une densité, c'est-à-dire une fonction de \mathbb{R} dans $[0, +\infty[$ telle que son intégrale sur \mathbb{R} fasse 1.

3.1 Densité de probabilité

Nous allons maintenant définir les probabilités absolument continues, qui constituent une classe importante de probabilités sur \mathbb{R} .

1. En théorie on doit ajouter application mesurable.

Définition 3.4 Une probabilité \mathbb{P} de (\mathbb{R}, \mathbb{P}) est dite absolument continue s'il existe une fonction $f : \mathbb{R} \rightarrow [0, +\infty[$ (donc positive), telle que

$$(3.1) \quad \mathbb{P}(A) = \int_A f(x) dx \quad \forall A \subset \mathbb{R}.$$

La fonction f est alors appelée densité de probabilité de \mathbb{P} , et on a :

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

Lemme 3.1 La relation (3.1) définit complètement la probabilité \mathbb{P} . On pourra ainsi parler de la densité de probabilité f , à partir du moment où f est une fonction positive dont l'intégrale sur \mathbb{R} vaut 1.

Dans le cas d'univers dénombrables (ou variables discrètes), si une éventualité a une probabilité nulle, on peut l'ôter de l'univers : ainsi, on peut toujours se ramener à un ensemble probabilisé pour lequel toutes les éventualités sont de probabilité non nulle. Par contre, dans le cas d'univers non dénombrables (en particulier sur \mathbb{R}), ce n'est pas toujours le cas. En effet si $A = \{a\}$ est réduit à un nombre réel a , alors

$$\mathbb{P}(A) = \mathbb{P}(a) = \int_a^a f(x) dx = 0.$$

Autrement dit la probabilité d'un nombre est nulle. D'où l'intérêt de la définition suivante :

Définition 3.5 Pour (Ω, \mathbb{P}) , un événement A est dit presque-impossible ou \mathbb{P} -négligeable si $\mathbb{P}(A) = 0$. Son complémentaire est dit presque-sûr.

Exemple :

La probabilité uniforme sur $[a, b]$ est la probabilité continue définie par une densité f constante sur $[a, b]$ (la constante de normalisation $1/(b - a)$ est là pour garantir que l'intégrale de f sur \mathbb{R} vaut un) :

$$(3.2) \quad f(x) = \frac{1}{b - a} \mathbf{1}_{[a, b]}(x).$$

On remarquera que si on « tire au hasard » un nombre réel dans l'intervalle $[0, 1]$ (avec la loi uniforme \mathbb{P}), il n'est presque-sûrement pas rationnel : $\mathbb{P}([0, 1] \cap \mathbb{Q}) = 0$. En effet, comme $[0, 1] \cap \mathbb{Q}$ est dénombrable,

$$\begin{aligned} \mathbb{P}([0, 1] \cap \mathbb{Q}) &= \mathbb{P} \left(\bigcup_{x \in [0, 1] \cap \mathbb{Q}} \{x\} \right) = \sum_{x \in [0, 1] \cap \mathbb{Q}} \mathbb{P}(\{x\}) \\ &= \sum_x \int_x^x f(u) du = \sum_x 0 = 0. \end{aligned}$$

Remarque. En modifiant la densité d'une variable absolument continue en un point (ou en un nombre dénombrable de points), on ne change pas sa loi (regarder dans (3.1)). C'est pourquoi on dit que f est *une* densité de la probabilité \mathbb{P} , et non pas *la* densité de \mathbb{P} . Malgré tout, on a intérêt à ne pas modifier artificiellement la densité f , et la garder la plus régulière possible.

Pourquoi f est-elle appelée la *densité de probabilité* de \mathbb{P} ? C'est ce que nous explique le lemme suivant, où on reconnaît la définition des densités utilisées en physique : densité de masse, etc.

Lemme 3.2 *Soit \mathbb{P} une probabilité absolument continue, de densité f . Alors, pour tout $x \in \mathbb{R}$ tel que f est continue en x ,*

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{\mathbb{P}([x - \varepsilon, x + \varepsilon])}{2\varepsilon}.$$

Preuve. On peut encadrer l'intégrale de f sur l'intervalle $[x - \varepsilon, x + \varepsilon]$ à l'aide des bornes inférieures et supérieures de f sur cet intervalle :

$$2t \inf_{x-\varepsilon \leq t \leq x+\varepsilon} f(t) \leq \mathbb{P}([x - \varepsilon, x + \varepsilon]) = \int_{x-\varepsilon}^{x+\varepsilon} f(t) dt \leq 2t \sup_{x-\varepsilon \leq t \leq x+\varepsilon} f(t).$$

Il ne reste plus qu'à utiliser la continuité de f en x , qui entraîne :

$$\lim_{\varepsilon \rightarrow 0} \inf_{x-\varepsilon \leq t \leq x+\varepsilon} f(t) = \lim_{\varepsilon \rightarrow 0} \sup_{x-\varepsilon \leq t \leq x+\varepsilon} f(t) = f(x).$$

□

Le calcul des probabilités absolument continues de (\mathbb{R}, \mathbb{P}) revient donc à la donnée

d'une densité $f : \mathbb{R} \rightarrow \mathbb{R}$. Dans ce contexte, l'analyse des fonctions réelles, que nous ne développerons pas ici, fournit de nombreux outils théoriques.

Le vocabulaire relatif aux probabilités absolument continues sera aussi appliqué aux variables aléatoires :

Définition 3.6 *Si X est une variable aléatoire réelle telle que \mathbb{P}_X est une probabilité absolument continue de densité f , on dira que X est absolument continue ou à densité, et a pour densité f .*

Alors pour tout $A \subset \mathbb{R}$,

$$\mathbb{P}(X \in A) = \int_A f(x) dx.$$

Ainsi si X est à densité f , alors

$$\mathbb{P}(X = a) = \int_a^a f(x) dx = 0$$

pour tout $a \in \mathbb{R}$. Ce qui constitue une différence majeure avec les variables aléatoires discrètes !

3.2 Fonction de répartition d'une v.a. réelle

Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle de l'espace de probabilité (Ω, \mathbb{P}) et \mathbb{P}_X sa loi de probabilité. Nous nous sommes ainsi rapporté à l'étude de l'espace de probabilité $(\mathbb{R}, \mathbb{P}_X)$.

Définition 3.7 On appelle fonction de répartition de X la fonction F_X définie sur \mathbb{R} , et à valeurs dans $[0, 1]$, par :

$$\forall x \in \mathbb{R}, F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x).$$

La valeur de \mathbb{P}_X sur tout intervalle $]a, b]$ s'obtient simplement à partir de la fonction de répartition :

$$\mathbb{P}_X(]a, b]) = \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

La valeur de \mathbb{P}_X sur n'importe quel sous-ensemble de \mathbb{R} peut en fait être déduite de la fonction de répartition de X :

Théorème 3.1 La fonction de répartition caractérise la loi : si deux variables aléatoires réelles X et Y ont même fonction de répartition, alors elles ont même loi.

La fonction de répartition d'une variable aléatoire a par ailleurs les propriétés suivantes :

Propriétés 3.1 Soit F_X la fonction de répartition d'une variable aléatoire réelle X . Alors :

1. F_X est croissante.
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
3. F_X est continue à droite.
4. F_X a une limite à gauche en tout point.
5. Si F_X admet une discontinuité en x_0 , $F_X(x_0) - \lim_{\substack{x \rightarrow x_0 \\ x < x_0}} F_X(x) = \mathbb{P}(X = x_0)$.

Preuve.

1. Si $x \leq y$, alors $X \leq x \Rightarrow X \leq y$ donc $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$.
2. En effet, comme F_X est croissante,

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X(x) &= \lim_{n \rightarrow +\infty} F_X(-n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(]-\infty, -n]) \\ &= \mathbb{P}_X\left(\bigcap_{n=1}^{+\infty}]-\infty, -n]\right) = \mathbb{P}_X(\emptyset) = 0. \end{aligned}$$

De même,

$$\begin{aligned} \lim_{x \rightarrow +\infty} F_X(x) &= \lim_{n \rightarrow +\infty} \mathbb{P}_X(]-\infty, n]) = \mathbb{P}_X\left(\bigcup_{n=1}^n]-\infty, n]\right) \\ &= \mathbb{P}_X(\mathbb{R}) = 1. \end{aligned}$$

3. Comme F_X est croissante,

$$\begin{aligned} \lim_{x^+} F_X &= \lim_{n \rightarrow +\infty} F_X\left(x + \frac{1}{n}\right) = \lim_{n \rightarrow +\infty} \mathbb{P}_X \left(\left] -\infty, x + \frac{1}{n} \right] \right) \\ &= \mathbb{P}_X \left(\bigcap_{n \geq 1} \left] -\infty, x + \frac{1}{n} \right] \right) = \mathbb{P}_X (] -\infty, x]) = F_X(x). \end{aligned}$$

4. Comme F_X est croissante et bornée, elle admet une limite à gauche en tout point.

5. On a :

$$\begin{aligned} F_X(x_0^-) &= \lim_{x \uparrow x_0} \mathbb{P}_X (] -\infty, x]) = \lim_{n \rightarrow +\infty} \mathbb{P}_X \left(\left] -\infty, x_0 - \frac{1}{n} \right] \right) \\ &= \mathbb{P}_X \left(\bigcup_{n=1}^{\infty} \left] -\infty, x_0 - \frac{1}{n} \right] \right) = \mathbb{P}_X (] -\infty, x_0[) \\ &= \mathbb{P}(X < x_0). \end{aligned}$$

Or, $\mathbb{P}(X = x_0) = \mathbb{P}(X \leq x_0) - \mathbb{P}(X < x_0)$.

□

Proposition 3.1 *Si X est une v.a. discrète, sa fonction de répartition est constante par morceaux.*

Exemple :

Soit X la variable aléatoire indiquant le résultat du lancé d'un dé équilibré. Sa fonction de répartition f saute de $1/6$ aux points $1, \dots, 6$ (voir figure 3.1).

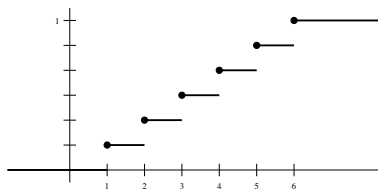


FIGURE 3.1 – fonction de répartition de X

Mais dans le cas à densité on a

Propriétés 3.2 *La fonction de répartition F_X d'une variable aléatoire réelle absolument continue de densité f_X s'écrit :*

$$(3.3) \quad F_X(x) = \int_{-\infty}^x f_X(u) du \quad \forall x \in \mathbb{R}.$$

En conséquence, F_X est continue. Elle est aussi dérivable aux points de continuité de f_X avec $F'_X(x) = f_X(x)$.

Remarque. Il existe des variables aléatoires réelles qui ne sont ni discrètes, ni absolument continues mais de loi de probabilité dite mixte, dont la fonction de répartition peut s'écrire :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \alpha F_d(x) + \beta F_C(x),$$

où $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$ et où F_d est la fonction de répartition d'une variable discrète et F_C est celle d'une variable absolument continue.

Il existe aussi des variables aléatoires ni continues, ni discrètes, ni mixtes... mais nous n'en rencontrerons pas dans ce cours.

Exemple 3.1 Soit X une v.a.r. dont la fonction de répartition est F_X . On s'intéresse à la fonction de répartition des v.a. suivantes :

$$\begin{aligned} Y &= X + a, \\ Z &= X^2, \\ U &= X^+ = \max(X, 0), \end{aligned}$$

où $a \in \mathbb{R}$ est un réel connu.

On a

$$\forall x \in \mathbb{R}, \quad F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(X \leq x - a) = F_X(x - a).$$

Pour calculer $F_Z(x)$, on remarque que $F_Z(x) = \mathbb{P}(Z \leq x) = 0$ lorsque $x < 0$. Si $x \geq 0$,

$$\begin{aligned} F_Z(x) &= \mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) \\ &= \mathbb{P}(X \leq \sqrt{x}) - \mathbb{P}(X < -\sqrt{x}) \\ &= F_X(\sqrt{x}) - F_X(-\sqrt{x} - 0) = F_X(\sqrt{x}) - \lim_{\varepsilon \rightarrow 0} F_X(-\sqrt{x} - \varepsilon). \end{aligned}$$

Donc

$$F_Z(x) = \begin{cases} F_X(\sqrt{x}) - F_X(-\sqrt{x} - 0) & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

On calcule maintenant $F_U(x)$. Il est clair que $F_U(x) = 0$ pour $x < 0$. Lorsque $x \geq 0$,

$$F_U(x) = \mathbb{P}(\max(X, 0) \leq x) = \mathbb{P}(X \leq x) = F_X(x).$$

En conclusion,

$$F_U(x) = \begin{cases} F_X(x) & \text{si } x \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

3.3 Moments

La loi d'une v.a. absolument continue X est donc caractérisée par la densité f_X .

Définition 3.8 (Espérance) L'espérance d'une v.a. absolument continue X de la loi \mathbb{P}_X définie par f_X est définie par

$$\mathbb{E}(X) = \int_{\mathbb{R}} x f_X(x) dx,$$

si le second membre converge absolument.

De même pour g une fonction de \mathbb{R} dans \mathbb{R} , on peut définir

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f_X(x)dx.$$

Tous les résultats suivants restent vrais : les propriétés 2.1, le théorème 2.3, la définition 2.9, la proposition 2.1, la propriété 2.2, la proposition 2.2 et la propriété 2.3.

3.4 Médiane

Toutes les définitions et tous les résultats valables pour une v.a. discrète sont encore vrais pour une v.a. à densité.

3.5 Fonctions caractéristiques

Définition 3.9 Soit X une v.a.r. On appelle fonction caractéristique de X (ou de la loi de X) la fonction de la variable réelle t définie par :

$$\phi_X(t) = \mathbb{E} [e^{itX}] = \int_{\mathbb{R}} e^{itx} f_X(x)dx.$$

Si X est une variable discrète, cette définition est encore valable avec $\phi_X(t) = \sum_{k \in I} e^{itx_k} p_k$.

Théorème 3.2 Désignons par ϕ la fonction caractéristique d'une v.a.r. X . Alors

1. ϕ est une fonction définie et continue pour tout $t \in \mathbb{R}$;
2. ϕ est bornée, et en fait, pour tout $t \in \mathbb{R}$, on a : $|\phi(t)| \leq \phi(0) = 1$;
3. pour tous a, b réels, on a, pour tout t réel, l'identité $\phi_{aX+b}(t) = e^{ibt} \phi_X(at)$.

Théorème 3.3 La fonction caractéristique d'une v.a. détermine la loi de cette variable. En d'autres termes, si deux v.a. admettent la même fonction caractéristique, elles ont même loi.

Proposition 3.2 Soit X une v.a. de fonction caractéristique ϕ . Si $\mathbb{E}|X| < \infty$, alors ϕ est continument dérivable et $\phi'(0) = i\mathbb{E}(X)$.

3.6 Lois continues classiques

3.6.1 Loi uniforme

Définition 3.10 Une v.a.r. X suit une loi uniforme sur l'intervalle $[a, b]$ avec $-\infty < a < b < +\infty$ si la densité de X est donnée par

$$\forall x \in \mathbb{R}, f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

Remarquer que

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(x)dx = 1, \quad \mathbb{P}(X = a) = \mathbb{P}(X = b) = 0.$$

Ainsi on ne peut pas distinguer une variable aléatoire de loi uniforme sur $[a, b]$, d'une autre de loi uniforme sur $]a, b]$ ou $[a, b[$ ou $]a, b[$.

Propriétés 3.3 Si X suit une loi uniforme sur $[a, b]$, alors X admet des moments de tout ordre et $\mathbb{E}(X) = (a + b)/2$ et $\text{Var}(X) = (b - a)^2/12$.

Preuve. En effet

$$\mathbb{E}(X) = \int_{\mathbb{R}} xf(x)dx = \frac{1}{b-a} \int_a^b xdx = \frac{1}{b-a} \frac{1}{2}(b^2 - a^2) = \frac{a+b}{2}.$$

De même

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f(x)dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \frac{1}{3}(b^3 - a^3) = \frac{a^2 + b^2 + ab}{3}$$

et ainsi

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{a^2 + b^2 + ab}{3} - \frac{(a+b)^2}{4} = \frac{1}{12}(b^2 + a^2 - 2ab) = \frac{(b-a)^2}{12}.$$

□

3.6.2 Loi exponentielle

Elle joue un rôle clé dans les processus de Markov et dans bon nombre de modèles.

Définition 3.11 Une v.a.r. X suit une loi exponentielle de paramètre $\lambda > 0$ si X est à valeurs positives et si la densité de X est donnée par

$$\forall x \in \mathbb{R}, f(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\mathbb{R}_+}(x).$$

Propriétés 3.4 Si X suit une loi exponentielle de paramètre $\lambda > 0$, alors X admet des moments de tout ordre et $\mathbb{E}(X) = 1/\lambda$ et $\text{Var}(X) = 1/\lambda^2$.

Preuve. En exercice. □

Exemple d'application Durée de bon fonctionnement d'un système

On admet que cette durée est généralement imprévisible. On la représente donc par une variable aléatoire positive X . Comme nous l'avons déjà remarqué, il n'est pas nécessaire de décrire l'espace fondamental sur lequel elle peut être définie.

Définition 3.12 On appelle *taux de défaillance du système* la fonction du temps définie par :

$$\forall t \geq 0, \quad \lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{P}(t < X \leq t + \Delta t | X > t).$$

Le *taux de défaillance au temps t* est la densité de probabilité pour le système de tomber en panne juste après le temps t , sachant qu'il est encore en marche au temps t .

Quelle est la loi de probabilité de X , si l'on suppose que cette fonction est constante ($\forall t \geq 0, \lambda(t) = \lambda > 0$; ce qui signifie que le système est **sans usure**) ?

En fait,

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < X \leq t + \Delta t \text{ et } X > t)}{\Delta t \cdot \mathbb{P}(X > t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < X \leq t + \Delta t)}{\Delta t \cdot \mathbb{P}(X > t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{F_X(t + \Delta t) - F_X(t)}{\Delta t(1 - F_X(t))} \end{aligned}$$

Si on suppose que X admet une densité, alors :

$$\lambda(t) = \frac{f_X(t)}{1 - F_X(t)}.$$

La question posée revient à résoudre l'équation différentielle :

$$F'_X(x) = \lambda(1 - F_X(x)),$$

avec la condition initiale $F_X(0) = 0$ (car X est à valeurs positives), dont la solution est : $F_X(x) = 1 - e^{-\lambda x}, \forall x \geq 0$. D'où la densité de X :

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

3.6.3 Loi normale (ou de Gauss)

On dit que X est une variable aléatoire gaussienne (K. Gauss, 1777-1855) de moyenne $m \in \mathbb{R}$ et d'écart-type $\sigma > 0$ si elle est absolument continue et a pour densité

$$f_X(x) = f_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

On note $\mathcal{N}(m, \sigma^2)$ cette loi.

Elle joue un rôle central en probabilité (voir chapitre suivant et le théorème central limite).

Propriétés 3.5 Si X suit une loi normale de paramètres $m = 0$ et $\sigma^2 = 1$ (loi normale centrée réduite), alors $\mathbb{E}(X) = 0, \text{Var}(X) = 1$;

Proposition 3.3 Si $X \sim \mathcal{N}(0, 1)$, alors $Y = \sigma X + m \sim \mathcal{N}(m, \sigma^2)$.

Comme F_Y n'est pas calculable explicitement, on utilisera la proposition précédente pour se ramener au cas $\mathcal{N}(0, 1)$ et on utilisera des tables.

3.7 Récapitulatif des lois continues

Loi	Densité	Espérance	Variance
Uniforme $\mathcal{U}(a, b)$ $a < b$	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponentielle $\exp(\lambda)$ $\lambda \in \mathbb{R}_+^*$	$\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale $\mathcal{N}(m, \sigma^2)$ $m \in \mathbb{R}, \sigma \in \mathbb{R}_+^*$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2

3.8 Correspondance v.a. discrète - v.a. continue

Dans la suite X désigne une v.a. réelle. De plus les x_i sont des nombres réels croissants.

Quantité calculée	V.a. discrète valeurs x_i probabilités p_i	V.a. continue densité $f(x)$
$\mathbb{P}(x_k \leq X < x_j)$	$\sum_{i=k}^{j-1} p_i$	$\int_{x_k}^{x_j} f(x) dx$
Normalisation	$\sum_i p_i = 1$	$\int_{-\infty}^{+\infty} f(x) dx = 1$
Moyenne ou espérance $\mathbb{E}(X)$	$\sum_i x_i p_i$	$\int_{-\infty}^{+\infty} x f(x) dx$
Variance $\text{Var}(X)$	$\sum_i (x_i - \mathbb{E}(X))^2 p_i$	$\int_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f(x) dx$
Moments d'ordre k	$\sum_i x_i^k p_i$	$\int_{-\infty}^{+\infty} x^k f(x) dx$

L'écart-type est toujours donné par $\sqrt{\text{Var}(X)}$.

Chapitre 4

Vecteurs aléatoires, indépendance

Le résultat d'une expérience aléatoire est souvent décrit par plusieurs v.a. L'exemple le plus classique est celui du tir sur une cible : le point d'impact est déterminé par des coordonnées aléatoires X et Y sur le plan de la cible. On dit alors qu'on a un **système de v.a.** ou un **vecteur aléatoire**.

Nous allons traiter surtout le cas d'un système à deux v.a. par souci de simplification. Une réalisation (x, y) du couple de v.a. (X, Y) donne un point aléatoire de coordonnées (x, y) . Comme pour une v.a. réelle, la distribution peut être discrète, à densité ou mixte. Pour un vecteur discret la représentation se fait par des points du plan ; la troisième dimension est utilisée pour les probabilités associées. Pour un système à densité, la représentation est constituée de régions du plan ; la troisième dimension sert à représenter la densité qui est alors une surface au-dessus du plan.

4.1 Fonction de répartition, densités

Soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle de l'espace de probabilité (Ω, \mathbb{P}) et \mathbb{P}_X sa loi de probabilité. Nous nous sommes ainsi rapporté à l'étude de l'espace de probabilité $(\mathbb{R}, \mathbb{P}_X)$. On rappelle que la fonction de répartition F_X de X est définie sur \mathbb{R} , et à valeurs dans $[0, 1]$, par :

$$\forall x \in \mathbb{R}, F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x).$$

La valeur de \mathbb{P}_X sur tout intervalle $]a, b]$ s'obtient simplement à partir de la fonction de répartition :

$$\mathbb{P}_X(]a, b]) = \mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

La valeur de \mathbb{P}_X sur n'importe quel sous-ensemble de \mathbb{R} peut en fait être déduite de la fonction de répartition de X et ainsi la fonction de répartition caractérise la loi : si deux variables aléatoires réelles X et Y ont même fonction de répartition, alors elles ont même loi.

Soit X une variable aléatoire réelle de fonction de répartition F . On note pour $u \in]0, 1[$,

$$q(u) = \inf\{x \in \mathbb{R}, F(x) > u\},$$

qui est la fonction quantile de X . Si la fonction F est inversible, alors $q = F^{-1}$, c'est-à-dire

$$\forall x \in \mathbb{R}, q(F(x)) = x, \quad \forall u \in]0, 1[, F(q(u)) = u.$$

Théorème 4.1 Soit X une variable aléatoire réelle, de fonction de répartition F . Alors la variable aléatoire $Y = F(X)$ suit une loi uniforme sur l'intervalle $[0, 1]$.

Réciproquement si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors $Y = q(U)$ a la même loi que X .

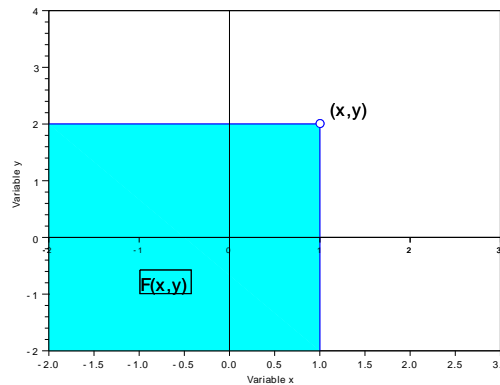
Ce théorème montre que toutes les v.a.r. peuvent être représentées à l'aide de la loi uniforme.

La définition qui suit est l'extension naturelle de la fonction de répartition d'une v.a. réelle.

Définition 4.1 Soit (X, Y) un vecteur aléatoire à valeurs dans \mathbb{R}^2 . La loi de (X, Y) est caractérisée par

$$F(x, y) = \mathbb{P}(X \leq x \text{ et } Y \leq y), \quad \forall (x, y) \in \mathbb{R}^2.$$

F est la **fonction de répartition** (bivariée) de (X, Y) .

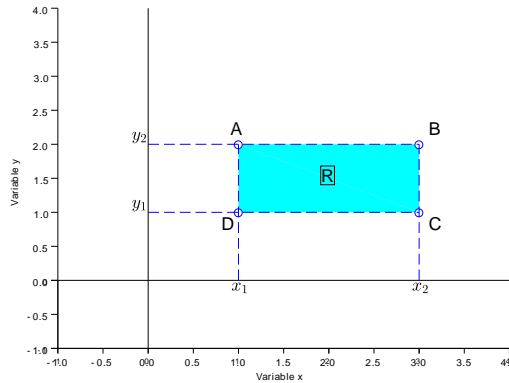


La quantité $F(x, +\infty) = \lim_{y \rightarrow +\infty} F(x, y)$ est égale à $\mathbb{P}(X \leq x \text{ et } Y \leq +\infty) = \mathbb{P}(X \leq x) = F_1(x)$. C'est la **fonction de répartition marginale** de X . De même on définit $F_2(y) = \mathbb{P}(Y \leq y) = F(+\infty, y) = \lim_{x \rightarrow +\infty} F(x, y)$. On appelle **lois marginales** ou simplement **marginales** de (X, Y) , les lois des v.a.r. X et Y prises séparément. Elles sont décrites par respectivement F_1 et F_2 . Enfin on a une normalisation :

$$F(+\infty, +\infty) = \mathbb{P}(X < +\infty, Y < +\infty) = 1.$$

Avec $F(x, y)$ on peut calculer la probabilité que le couple (X, Y) soit dans un rectangle :

$$\mathbb{P}((X, Y) \in R) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1).$$



On peut ensuite passer « à la limite ». Si $(x, y) = (x_1, y_1)$ et $(x + dx, y + dy) = (x_2, y_2)$, on a

$$\mathbb{P}((X, Y) \in R) = F(x + dx, y + dy) - F(x, y + dy) - F(x + dx, y) + F(x, y).$$

En développant les fonctions de répartition au voisinage de (x, y) ¹, on obtient

$$\mathbb{P}((X, Y) \in R) = \frac{\partial^2 F}{\partial x \partial y}(x, y) dx dy.$$

Ceci permet de définir la **densité conjointe** du couple (X, Y) au voisinage de (x, y) par

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y).$$

Alors si A est une partie quelconque de \mathbb{R}^2 , on a

$$\mathbb{P}((X, Y) \in A) = \int \int_A f(x, y) dx dy.$$

Et en fonction de f , on a

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x f(u, v) du dv, \quad \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(u, v) du dv = 1.$$

Les **densités marginales** f_1 et f_2 de X et Y s'obtiennent aussi à partir de f par :

$$f_1(x) = F'_1(x) = \frac{d}{dx} \left(\int_{-\infty}^x \int_{-\infty}^{+\infty} f(u, v) dv du \right) = \int_{-\infty}^{+\infty} f(x, v) dv,$$

$$f_2(y) = F'_2(y) = \int_{-\infty}^{+\infty} f(u, y) du.$$

1. Sous réserve que ce soit possible !

Nous disposons donc de trois types de probabilités infinitésimales : la probabilité marginale de X , $f_1(x)dx = \mathbb{P}(x < X \leq x + dx)$, celle de Y , $f_2(y)dy = \mathbb{P}(y < Y \leq y + dy)$ et la probabilité conjointe de (X, Y) , $f(x, y)dxdy$ qui est une probabilité d'intersection $\mathbb{P}(x < X \leq x + dx \text{ et } y < Y \leq y + dy)$. Nous pouvons donc évaluer les probabilités conditionnelles $\mathbb{P}(x < X \leq x + dx | y < Y \leq y + dy)$ et $\mathbb{P}(y < Y \leq y + dy | x < X \leq x + dx)$ en utilisant la définition 1.6 :

$$\begin{aligned} \mathbb{P}(x < X \leq x + dx | y < Y \leq y + dy) &= \frac{\mathbb{P}(x < X \leq x + dx \text{ et } y < Y \leq y + dy)}{\mathbb{P}(y < Y \leq y + dy)} \\ &= \frac{f(x, y)dxdy}{f_2(y)dy} = \frac{f(x, y)}{f_2(y)}dx. \end{aligned}$$

On peut alors définir les **densités de probabilité conditionnelles** :

$$f(x|y) = \frac{f(x, y)}{f_2(y)} = \frac{f(x, y)}{\int_{\mathbb{R}} f(u, y)du}, \quad f(y|x) = \frac{f(x, y)}{f_1(x)} = \frac{f(x, y)}{\int_{\mathbb{R}} f(x, v)dv},$$

et la formule de probabilités composées pour les densités :

$$f(x, y) = f_1(x)f(y|x), \quad f(x, y) = f_2(y)f(x|y).$$

4.2 Indépendance

Nous allons définir maintenant la notion d'indépendance pour deux v.a. X et Y .

Définition 4.2 (Indépendance) Soit (X, Y) un vecteur aléatoire de dimension 2. Les v.a. X et Y sont *indépendantes* si

$$f(x|y) = f_1(x) \iff f(y|x) = f_2(y) \iff f(x, y) = f_1(x)f_2(y).$$

Autrement dit la densité conjointe est le produit des marginales.

Proposition 4.1 Les composantes X et Y sont indépendantes si et seulement si

$$\boxed{f(x, y) = g_1(x)g_2(y)}$$

et alors X a pour densité $f_1(x) = \frac{g_1(x)}{\int_{\mathbb{R}} g_1(u)du}$; Y a pour densité $f_2(y) = \frac{g_2(y)}{\int_{\mathbb{R}} g_2(v)dv}$.

La définition qui suit est équivalente à la précédente si (X, Y) a une densité conjointe. Mais elle est plus générale car valable aussi si on n'a pas de densité.

Définition 4.3 Deux variables aléatoires X et Y sont *indépendantes* si pour tout $(x, y) \in \mathbb{R}^2$ on a

$$\mathbb{P}(X \leq x \text{ et } Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y) = F_1(x)F_2(y).$$

Dans le cas d'un vecteur aléatoire (X, Y) de loi discrète portée par les points (x_i, y_j) avec $i \in I$ et $j \in J$, la définition précédente devient :

Proposition 4.2 Si X et Y sont deux v.a.r. discrètes, X et Y sont indépendantes si, pour tout $i \in I$ et tout $j \in J$, on a :

$$\mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}) = \mathbb{P}(\{X = x_i\})\mathbb{P}(\{Y = y_j\}).$$

L'indépendance permet de montrer les résultats suivants :

Proposition 4.3 Si X et Y sont indépendantes, alors pour toutes fonctions f et g bornées,

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

En utilisant les fonctions caractéristiques, on obtient que

Théorème 4.2 Soit (X, Y) un couple de variables aléatoires. X et Y sont indépendantes si et seulement si pour tout u, v dans \mathbb{R}

$$\mathbb{E}(e^{iuX+ivY}) = \phi_X(u)\phi_Y(v).$$

Ceci implique que si X et Y ont un moment d'ordre 1 (resp. d'ordre 2) et sont indépendantes, alors

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y), \quad \text{et resp.} \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Théorème 4.3 La somme de deux variables aléatoires gaussiennes indépendantes de paramètres respectifs (μ_1, σ_1^2) et (μ_2, σ_2^2) est encore une gaussienne de paramètre $(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

4.3 Moments, corrélations

De la même façon que pour une v.a.r. à densité, on peut définir l'espérance d'un vecteur (X, Y) :

$$\mathbb{E}(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)f(x, y)dx dy$$

où f est la densité conjointe de (X, Y) et g est une fonction de \mathbb{R}^2 dans \mathbb{R} . Dans le cas discret, on aura

$$\mathbb{E}(g(X, Y)) = \sum_{i \in I} \sum_{j \in J} g(x_i, y_j)p_{i,j},$$

où $p_{i,j}$ est la probabilité d'avoir $X = x_i$ et $Y = y_j$.

Voyons quelques cas particuliers. Si $g(x, y) = x^k y^l$ avec k et l entiers, $\mathbb{E}(X^k Y^l)$ est appelé **moment mixte** d'ordres k et l du vecteur aléatoire (X, Y) . Pour $l = 0$, on a : $\mathbb{E}(X^k Y^0) = \mathbb{E}(X^k)$.

Définition 4.4 On appelle **covariance** de X et de Y la quantité :

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))].$$

La covariance de X et de X est simplement la variance de X .

Lemme 4.1 Si X et Y sont indépendantes, alors $Cov(X, Y) = 0$.

Des v.a. indépendantes sont donc *décorrélées*.

Définition 4.5 Le *coefficient de corrélation* entre X et Y est défini par

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y},$$

où σ_X et σ_Y sont les écart-types de X et de Y .

Le coefficient de corrélation vérifie : $-1 \leq Corr(X, Y) \leq 1$ et si X et Y sont indépendantes, alors $Corr(X, Y) = 0$.

Il est important de noter que corrélation et dépendance sont deux choses différentes. La corrélation ne traduit qu'un aspect de cette dépendance. Ainsi les seuls points dont nous sommes sûrs sont :

- Si X et Y sont indépendantes, alors $Corr(X, Y) = 0$. La réciproque n'est pas vraie en générale.
- Si $Corr(X, Y) = 1$ (resp. -1), alors X et Y sont liées par : $Y = aX + b$ avec a et b deux constantes non aléatoires avec $a > 0$ (resp. $a < 0$). La réciproque est vraie.

Ainsi certains vecteurs aléatoires peuvent être dépendants (mais pas linéairement) et avoir une corrélation nulle. De plus la capacité pour la corrélation de « mesurer » un lien même linéaire est sujette à caution (voir paragraphe ?? sur la régression).

4.4 Extension en dimension d

Un vecteur aléatoire de dimension d est une v.a. à valeurs dans \mathbb{R}^d . On note $X = (X_1, \dots, X_d)$. Les coordonnées de $X = (X_1, \dots, X_d)$ sont alors des v.a.r., dont les lois $\mu_{X_1}, \dots, \mu_{X_d}$ sont les **lois marginales de X** . La fonction de répartition de X est

$$\begin{aligned} \forall (x_1, \dots, x_d) \in \mathbb{R}^d, F_X(x_1, \dots, x_d) &= \mathbb{P}_X([\!-\infty, x_1] \times \dots \times [\!-\infty, x_d]) \\ &= \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d). \end{aligned}$$

La densité f de X est une fonction de \mathbb{R}^d dans \mathbb{R} positive telle que

$$\int_{\mathbb{R}^d} f(x_1, \dots, x_d) dx_1 \dots dx_d = 1$$

et telle que

$$F_X(x_1, \dots, x_d) = \int_{-\infty}^{x_d} \dots \int_{-\infty}^{x_1} f(u_1, \dots, u_d) du_1 \dots du_d.$$

Proposition 4.4 Si X est un vecteur aléatoire de densité $f : \mathbb{R}^d \rightarrow \mathbb{R}$, alors X_k a pour densité

$$f_k(y) = \int_{\mathbb{R}^{d-1}} f(x_1, \dots, x_{k-1}, y, x_{k+1}, \dots, x_d) dx_1 \dots dx_{k-1} dx_{k+1} \dots dx_d.$$

4.4.1 Matrice de covariance

On rappelle quelques notations.

- Si M est une matrice, M^T est sa transposée.
- Un vecteur $x \in \mathbb{R}^d$ est une matrice $d \times 1$.
- $\langle x, y \rangle = x^T y$.
- $\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_d))$.

Définition 4.6 La *matrice de covariance de X* est définie par

$$K(X) = \mathbb{E} [(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T] = \mathbb{E}(XX^T) - \mathbb{E}(X)\mathbb{E}(X)^T.$$

C'est donc une matrice de taille $d \times d$, de coefficients :

$$K(X)_{ij} = \text{Cov}(X_i, X_j).$$

Cette matrice vérifie les propriétés suivantes.

- $K(\alpha X) = \alpha^2 K(X)$, $\alpha \in \mathbb{R}$; $K(X + a) = K(X)$, $a \in \mathbb{R}^d$.
- **Symétrie** : $K^T(X) = K(X)$.
- **Positivité** : pour tout $\lambda \in \mathbb{R}^d$, $\lambda^T K(X) \lambda \geq 0$.
- Si M est une matrice déterministe $r \times d$, alors

$$K(MX) = MK(X)M^T.$$

- $\mathbb{P}(X - \mathbb{E}(X) \in \text{Im}(K(X))) = 1$.

4.4.2 Indépendance

Rappelons que si X est une variable aléatoire définie sur (Ω, \mathbb{P}) à valeurs dans \mathbb{R}^d , on note \mathbb{P}_X la loi de X , i.e.

$$\forall A \in \mathcal{P}(\mathbb{R}^d), \mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

Définition 4.7 (Indépendance) Les v.a.r. X_i , $1 \leq i \leq n$ sont indépendantes si et seulement si pour tout $(A_1, \dots, A_n) \in \mathcal{B}(\mathbb{R})^n$:

$$\mathbb{P}_{(X_1, \dots, X_n)}((A_1, \dots, A_n)) = \mathbb{P}_{X_1}(A_1) \times \dots \times \mathbb{P}_{X_n}(A_n).$$

Corollaire 4.1 Les v.a. réelles X_i , $1 \leq i \leq n$, sont indépendantes si et seulement si pour tout $x_i \in \mathbb{R}$,

$$\mathbb{P} \left(\bigcap_{i=1}^n \{X_i \leq x_i\} \right) = \prod_{i=1}^n F_{X_i}(x_i);$$

F_{X_i} étant la fonction de répartition de X_i .

Si X est un vecteur de composantes X_1, \dots, X_d , alors les composantes X_1, \dots, X_d sont indépendantes si et seulement si

$$f(x_1, \dots, x_d) = f_1(x_1) \dots f_d(x_d) \quad \text{p.p.}$$

avec f_k densité de X_k .

Corollaire 4.2 *Les composantes X_1, \dots, X_d sont indépendantes si et seulement si*

$$\boxed{f(x_1, \dots, x_d) = g_1(x_1) \dots g_d(x_d) \quad p.p.}$$

et alors X_k a pour densité $f_k(y) = g_k(y) / \int_{\mathbb{R}} g_k$.

Enfin si X et Y sont des vecteurs aléatoires indépendants, alors $K(X + Y) = K(X) + K(Y)$.

Chapitre 5

Convergence et théorèmes limites

Lorsqu'on lance une pièce en l'air on ne sait jamais si l'on obtiendra pile ou face. Mais on sait que malgré la stricte indépendance des lancers, un excédent de sorties « pile » sera tôt ou tard compensé par des sorties « face », de sorte qu'au bout d'un nombre suffisant de lancers les nombres de résultats pile et face seront aussi proches que l'on veut en pourcentage. Il y a là comme une sorte de magie qu'il a fallu deux siècles à analyser depuis J. Bernoulli au tout début du XVIII-ème siècle. C'est l'un des aspects de la loi des grands nombres. La démonstration (presque) générale date seulement de la fin du XIX-ème siècle et elle est due à Tchebychev.

Tout revient à dire que lors de l'accumulation d'un très grand nombre d'aléatoires, l'aléatoire disparaît et la limite est certaine, du moins en valeur relative. Mais que se passe-t-il « juste avant » cette limite ? S'il y a universellement limite, y a-t-il un comportement statistique limite universel ? Et dans ce cas y a-t-il une ou des lois vers lesquelles tendent les distributions ? Ce problème a été abordé dès les débuts des probabilités par de Moivre (1732) et le nom de Poisson reste attaché aux premiers résultats importants. On a longtemps cru qu'il n'y avait qu'une loi limite, la loi de Gauss ou normale, d'où son importance dans les théories physiques et les méthodes statistiques élaborées jusqu'à nos jours. Il a fallu attendre les années 1930 pour établir qu'en fait il existait une infinité de lois privilégiées : les lois stables¹.

5.1 Loi des grands nombres

D'abord rappelons

Définition 5.1 (Presque sûrement) *Pour (Ω, \mathbb{P}) , un événement A est dit presque-impossible ou \mathbb{P} -négligeable si $\mathbb{P}(A) = 0$. Son complémentaire est dit presque-sûr.*

Exemple 5.1 *Si on joue à pile ou face une infinité de fois, « obtenir uniquement des pile » est négligeable.*

1. qui ne feront pas partie de ce cours.

5.1.1 Inégalités de Markov et de Bienaymé-Tchebychev

Les résultats qui suivent sont valables aussi bien pour une v.a. discrète que pour une v.a. à densité.

Proposition 5.1 (Inégalité de Markov) *Soit une variable aléatoire X qui admet un moment d'ordre 1. Pour tout $a > 0$, on a*

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}(|X|)}{a}.$$

Preuve. En effet

$$\begin{aligned} \mathbb{E}(|X|) &= \mathbb{E}(|X|\mathbf{1}_{]a,+\infty[}(|X|)) + \mathbb{E}(|X|\mathbf{1}_{[0,a]}(|X|)) \geq \mathbb{E}(|X|\mathbf{1}_{]a,+\infty[}(|X|)) \\ &\geq a\mathbb{E}(\mathbf{1}_{]a,+\infty[}(|X|)) = a\mathbb{P}(|X| > a). \end{aligned}$$

□

Proposition 5.2 (Inégalité de Bienaymé-Tchebychev) *Soit X une v.a. qui admet un moment d'ordre 2. Pour tout réel $a > 0$,*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

Preuve. Si on pose $Y = X - \mathbb{E}(X)$, on a

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}(|Y| \geq a) = \mathbb{P}(Y^2 \geq a^2) \leq \frac{\mathbb{E}(Y^2)}{a^2},$$

d'après l'inégalité de Markov appliquée à Y^2 . Or $\mathbb{E}(Y^2) = \text{Var}(X)$.

□

L'inégalité de Bienaymé-Tchebychev permet de travailler avec des v.a. dont la loi des inconnue. En effet si on désire une confiance d'au moins 90 % sur un résultat, comme :

$$\mathbb{P}(|X - \mathbb{E}(X)| < \varepsilon) = 1 - \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2},$$

on en déduit que si on prend $\varepsilon > \sigma_X \sqrt{10} \approx 3\sigma_X$, alors

$$\mathbb{P}(|X - \mathbb{E}(X)| < \varepsilon) \geq 0,9.$$

Ainsi quelle que soit la distribution de X , on a certain d'avoir

$$X \in [\mathbb{E}(X) - 3\sigma_X, \mathbb{E}(X) + 3\sigma_X]$$

avec une confiance de 90%.

Si maintenant on se donne n v.a. X_1, \dots, X_n indépendantes et identiquement distribuées (ou de même loi), et si on pose

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n),$$

qui est la moyenne arithmétique (on dit aussi empirique) des X_i , alors

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}(X_1), \quad \text{Var}(\bar{X}_n) = \frac{1}{n} \text{Var}(X_1).$$

En utilisant l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2}.$$

Donc pour tout $\delta > 0$, il existe un entier n_0 tel que pour tout $n \geq n_0$,

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) \leq \delta.$$

Théorème 5.1 (de Tchebychev) *Pour tout $\varepsilon > 0$,*

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|\bar{X}_n - \mathbb{E}(X_1)| \geq \varepsilon) = 0.$$

La moyenne arithmétique converge en probabilité vers l'espérance de X_1 .

5.1.2 Probabilité, limite des fréquences

Considérons de nouveau le cas d'une expérience à deux issues, échec (noté 0) et succès (noté 1). On se fixe un nombre $0 < p < 1$, la probabilité d'obtenir un succès après une réalisation de l'expérience. On va modéliser le cas de cette expérience qui se répète infiniment dans des conditions identiques et indépendantes. Ici :

$$\Omega = \{\omega_i = (\eta_j, j \in \mathbb{N}) : \eta_j \in \{0, 1\}\}.$$

On veut vérifier que p correspond à notre intuition d'être la fréquence asymptotique de réalisations des *succès* :

$$(5.1) \quad \lim_{n \rightarrow \infty} \frac{|\{i \leq n : \eta_i = \text{succès}\}|}{n} = p.$$

On note X_1, \dots, X_n nos variables de Bernoulli modélisant chacune une réalisation de l'expérience (i.e. $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$ pour tout $0 \leq i \leq n$), et $S_n = X_1 + \dots + X_n$. Le terme de gauche de l'équation précédente (5.1) correspond alors à une réalisation de la variable aléatoire S_n/n .

Une autre façon de se poser cette question est de se demander s'il y a beaucoup de réalisations où $S_n/n \neq p$. On reformule alors la question

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}\left(\left\{\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \epsilon\right\}\right) = 0.$$

La réponse affirmative est la célèbre loi (faible) des grands nombres, qui relie les notions de fréquence et de probabilité :

Théorème 5.2 (Bernoulli, 1685)

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}\left(\left\{\omega : \left|\frac{S_n(\omega)}{n} - p\right| > \epsilon\right\}\right) = 0.$$

5.1.3 Loi (forte) des grands nombres

Ce premier résultat obtenu pour une suite de variables aléatoires de Bernoulli a été généralisé à toute suite de v.a. $(X_n)_{n \in \mathbb{N}^*}$ indépendantes et identiquement distribuées (en abrégé i.i.d.), c'est-à-dire toutes ont la même loi sur un même espace Ω et sont indépendantes.

Théorème 5.3 *Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, de loi commune celle d'une variable aléatoire générique X . On suppose que l'espérance de X existe, i.e. $\mathbb{E}(|X|) < +\infty$. Alors on a presque sûrement :*

$$\lim_{n \rightarrow +\infty} \frac{1}{n}(X_1 + \dots + X_n) = \mathbb{E}(X).$$

Autrement dit, en introduisant les notations :

$$S_n = \sum_{i=1}^n X_i, \quad \overline{X}_n = \frac{1}{n} S_n,$$

le théorème stipule que les moyennes \overline{X}_n des X_i converge presque sûrement, c'est-à-dire qu'il existe un ensemble $\Omega_0 \subset \Omega$ tel que $\mathbb{P}(\Omega_0) = 1$ et si $\omega \in \Omega_0$, alors

$$\lim_{n \rightarrow +\infty} \frac{1}{n}(X_1(\omega) + \dots + X_n(\omega)) = \lim_{n \rightarrow +\infty} \overline{X}_n(\omega) = \mathbb{E}(X).$$

Par rapport au théorème de Tchebychev on parle de loi forte des grands nombres car on a une limite presque sûre qui implique une convergence en probabilité, plus faible.

5.2 Théorème central limite

La loi des grands nombres « lie » la fréquence et la probabilité. Remarquons que quelque chose d'aléatoire converge vers un point complètement déterministe. On peut se demander ce qui se passe si au lieu de diviser par n , on divise par quelque chose qui tend aussi vers $+\infty$, mais moins vite que n . Est-ce qu'on peut obtenir à la limite un objet encore aléatoire? Dans ce contexte, on va voir que la loi normale joue un rôle particulier, qui lui confère une grande importance en probabilité et en statistique. Rappelons un résultat concernant les lois gaussiennes.

Proposition 5.3 *Si X et Y sont deux v.a. normales indépendantes de loi respective $\mathcal{N}(\mu, \sigma^2)$ et $\mathcal{N}(\nu, \theta^2)$, alors $X+Y$ est aussi une variable normale, de loi $\mathcal{N}(\mu+\nu, \sigma^2+\theta^2)$.*

Corollaire 5.1 *Soient $\{X_i, i = 1, \dots, n\}$ des variables $\mathcal{N}(\mu, \sigma)$ indépendantes, alors*

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

On peut donc représenter \bar{X}_n comme

$$\bar{X}_n = \frac{\sigma}{\sqrt{n}}Z + \mu, \text{ avec } Z \sim \mathcal{N}(0, 1),$$

d'où

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}.$$

Ce phénomène est en fait beaucoup plus général et fait l'objet du théorème suivant :

Théorème 5.4 (Théorème central limite) *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a. indépendantes et identiquement distribuées, de moyenne μ et de variance σ^2 . Alors la distribution de \bar{X}_n approche celle de la loi $\mathcal{N}(0, 1)$ au sens suivant. Pour tout $a < b$ on a*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(a < \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < b \right) = \mathbb{P}(a < Z < b),$$

où $Z \sim \mathcal{N}(0, 1)$. De façon équivalente, si on note $S_n = X_1 + \dots + X_n$:

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} < b \right) = \mathbb{P}(a < Z < b).$$

On note que $n\sigma^2 = \text{Var}(S_n)$ et $n\mu = \mathbb{E}(S_n)$. Ainsi le théorème dit qu'en gros :

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \simeq Z.$$

5.3 Méthode de Monte-Carlo

C'est un **algorithme probabiliste de calcul numérique d'intégrales**.

On souhaite calculer la valeur numérique d'une intégrale de la forme $\int_0^1 g(x)dx$, avec g continue. La méthode dite de Monte-Carlo consiste à « simuler » N v.a. $(U_n)_{1 \leq n \leq N}$ indépendantes et identiquement distribuées de loi uniforme sur $[0, 1]$. La loi des grands nombres donne :

$$\lim_{n \rightarrow +\infty} \frac{g(U_1) + \dots + g(U_n)}{n} = \mathbb{E}(g(U_1)) = \int_0^1 g(x)dx.$$

Donc si N est assez grand,

$$\frac{g(U_1) + \dots + g(U_N)}{N} \simeq \int_0^1 g(x)dx.$$

Plus généralement si on sait simuler des v.a. $(X_n)_{n \geq 1}$ ayant une densité f , on va avoir une valeur numérique pour des intégrales de la forme :

$$\frac{g(X_1) + \dots + g(X_N)}{N} \approx \mathbb{E}(g(X_1)) = \int g(x)f(x)dx = \mu.$$

Un algorithme n'est pertinent que si on peut en donner la vitesse. Grâce au théorème central limite, on a une idée de la vitesse de convergence. En effet pour N assez grand

$$\mathbb{P} \left(\frac{\sqrt{N}}{\sigma} \left| \frac{g(X_1) + \dots + g(X_N)}{N} - \mu \right| \geq a \right) \approx \mathbb{P}(|Z| \geq a).$$

σ^2 est la variance de $g(X_1)$. Si on choisit $a = 1,96$, la table de la loi normale (cf. ??) nous dit que $\mathbb{P}(|Z| \geq a) = 0,05$ et ainsi

$$\mathbb{P} \left(\frac{\sqrt{N}}{\sigma} \left| \frac{g(X_1) + \dots + g(X_N)}{N} - \mu \right| \geq 1,96 \right) \approx 0,05.$$

Autrement dit

$$\mu \in \left[\frac{g(X_1) + \dots + g(X_N)}{N} - 1,96 \frac{\sigma}{\sqrt{N}}, \frac{g(X_1) + \dots + g(X_N)}{N} + 1,96 \frac{\sigma}{\sqrt{N}} \right]$$

avec une confiance proche 95%. L'intervalle précédent est un **intervalle de confiance à 95%**. L'erreur commise en approchant μ par la moyenne empirique est donc de l'ordre de $2 \frac{1,96\sigma}{\sqrt{N}}$. La vitesse de convergence de la méthode de Monte Carlo est donc en $1/\sqrt{N}$. Et on ne pourra pas aller plus vite avec cette méthode.

Les méthodes numériques déterministes sont plus rapides, mais leur rapidité dépend beaucoup de la régularité de la fonction à intégrer (au moins dérivable) et de la dimension de l'espace dans lequel la fonction prend ces valeurs. La méthode de Monte-Carlo a l'avantage de ne pas trop dépendre de la régularité, ni de la dimension. Remarquons pour finir que la variance σ joue un rôle important dans le théorème et on peut voir que plus cette variance est petite, plus le résultat numérique est proche du résultat théorique. Il existe quelques principes généraux qui permettent de réduire la variance.

Chapitre 6

Annexe

6.1 Dénombrement : quelques rappels

Ce court paragraphe est là pour rappeler quelques principes de base du dénombrement. Il doit être compris mais n'est pas traité en cours.

6.1.1 Arrangements et permutations

Arrangements avec répétition

On effectue p tirages *non exhaustifs* dans une urne contenant n boules *distinctes*. Combien y a-t-il de résultats différents en tenant compte de l'ordre des tirages? Au premier tirage il y a n possibilités, au second n , etc. jusqu'au p -ième. Le nombre de résultats possibles est donc n^p . Ici les boules sont discernables ainsi que les tirages. À chaque tirage est associée une boule et une seule; il s'agit donc d'une application de l'ensemble des p tirages dans l'ensemble des n boules. Ces sont des **arrangements avec répétition**, leur nombre est $a_n^p = n^p$.

Arrangements sans répétition

On effectue p tirages *exhaustifs* dans une urne contenant n boules *distinctes*. Le nombre de résultats possibles en tenant compte de l'ordre se calcule de la même façon : au premier tirage on a n choix, au second $n - 1$, etc. jusqu'au p -ième où il reste $n - p + 1$ boules. On obtient donc : $n(n - 1) \dots (n - p + 1)$. Boules et tirages sont discernables, mais à chaque tirage est associée une boule et une seule. C'est une **injection** de l'ensemble des p tirages dans celui des n boules. Une boule peut ne pas être tirée, et p ne peut pas être supérieur à n . On parle de tirages *exhaustifs et ordonnés*. Ce sont des **arrangements sans répétition**, leur nombre A_n^p est :

$$A_n^p = n(n - 1) \dots (n - p + 1) = \frac{n!}{(n - p)!}.$$

On rappelle que $n! = 1 \times 2 \times 3 \times \dots \times n$ avec par convention $0! = 1$.

Permutations

C'est un arrangement sans répétition avec $p = n$. Ceci correspond aussi au nombre de numérotations des n éléments d'un ensemble, ou encore au nombre de distribution de n objets distincts dans n cases distinctes. Ce sont les **permutations** de n éléments ; il y en a :

$$P_n = n!.$$

6.1.2 Indiscernabilité des objets : combinaisons

Combinaisons sans répétition

Soit p tirages exhaustifs dans une urne contenant n objets *distincts* ; on ne s'intéresse maintenant qu'à la nature des objets tirés et non à l'ordre des tirages¹. : le nombre donné par A_n^p est donc trop important. Les tirages constitués des mêmes objets sont regroupés en classes, chacune contenant $p!$ permutations d'objets. Une classe est une **combinaison** de p éléments parmi n ; elle est ici sans répétition car les objets n'apparaissent jamais plus d'une fois. Leur nombre est :

$$C_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!}$$

qu'on appelle **coefficient binomial**. On peut attacher à ce cas :

- la distinction de p objets (atomes, particules) *indiscernables* dans n cases (sites, états) *distinctes, sans répétition* ;
- le nombre de façons de partitionner un ensemble à n éléments (distincts) en deux sous-ensembles à p et $n - p$ éléments.

Coefficient multinomial ou polynomial

On généralise au nombre de partitions d'un ensemble à n éléments distincts en r sous-ensembles à n_1, \dots, n_r éléments avec $n_1 + n_2 + \dots + n_r = n$. On obtient alors le **coefficient multinomial** ou **polynomial** :

$$C_n^{n_1, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}.$$

Ce coefficient représente aussi le nombre de façons de tirer n_1 objets de type 1, n_2 objets de type 2, ... et n_r objets de type r lors de n tirages exhaustifs dans une urne contenant des objets de ces types.

1. le tirage peut se faire par poignées

Relations entre combinaisons

On peut prouver les relations suivantes :

$$\begin{aligned} C_n^0 &= C_n^n = 1, \\ C_n^1 &= C_n^{n-1} = n \\ C_n^p &= C_n^{n-p} \\ pC_n^p &= nC_{n-1}^{p-1} \\ C_n^p &= C_{n-1}^{p-1} + C_{n-1}^p. \end{aligned}$$

La dernière relation est dite du triangle de Pascal. On peut ajouter la formule dite de Vandermonde

$$\sum_{k=0}^n C_n^k C_m^{p-k} = C_{n+m}^p,$$

et celle dite du binôme de Newton :

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

6.2 Séries

6.2.1 Vocabulaire

Définition 6.1 On appelle série à valeurs dans \mathbb{R} tout couple (u, \mathcal{S}) de deux suites d'éléments de \mathbb{R} telles que pour tout $n \in \mathbb{N}$, $S_n = u_0 + u_1 + \dots + u_n$, soit le terme général de la suite \mathcal{S} , avec u_k terme général de la suite u .

En somme une série, c'est une suite $u = (u_n)_{n \in \mathbb{N}}$ mais à partir de laquelle on s'intéresse à une nouvelle suite, de terme général $S_n = u_0 + \dots + u_n$, appelé somme partielle d'ordre n de la série.

Par abus de notation, on notera encore $u = (u_n)_{n \in \mathbb{N}}$ la série de terme général u_n , et par abus de langage, on parlera de la série des u_n sans autre précision, en sous-entendant la série de terme général u_n , donc le couple (u, \mathcal{S}) qui précède. À bien distinguer de la suite des u_n !

Définition 6.2 Soit u une série de terme général u_n . Elle est dite convergente, de somme S si et seulement si la suite des sommes partielles S_n converge, avec $S = \lim_{n \rightarrow +\infty} S_n$.

La série sera dite divergente si elle ne converge pas. En cas de convergence, l'élément $R_n = S - S_n$ est appelé reste d'ordre n de la série.

Propriétés 6.1 On peut faire les opérations suivantes :

1. **Addition** : si u et v sont deux séries convergentes de terme général u_n et v_n , alors la série de terme général $w_n = u_n + v_n$ converge. En cas de convergence,

$$\sum_{n=0}^{+\infty} \lambda w_n = \sum_{n=0}^{+\infty} u_n + \sum_{n=0}^{+\infty} v_n.$$

2. **Multiplication par un scalaire** : si $\lambda \in \mathbb{R}^*$, la série de terme général λu_n est de même nature que la série de terme général u_n . En cas de convergence,

$$\sum_{n=0}^{+\infty} \lambda u_n = \lambda \sum_{n=0}^{+\infty} u_n.$$

Un exemple : la série géométrique

On va considérer un exemple simple où on pourra calculer effectivement les sommes partielles.

Soit la série de terme général $u_n = \alpha^n$, avec α fixé dans \mathbb{R} . L'identité

$$1 - \alpha^{n+1} = (1 + \alpha + \alpha^2 + \dots + \alpha^n) (1 - \alpha) \stackrel{\alpha \neq 1}{\iff} S_n = \sum_{k=0}^n \alpha^k = \frac{1 - \alpha^{n+1}}{1 - \alpha}.$$

On sait que si $|\alpha| < 1$, alors $\lim_{n \in \mathbb{N}} \alpha^n = 0$, donc $\lim_{n \in \mathbb{N}} S_n = \frac{1}{1 - \alpha}$: la série converge et on connaît sa somme, et aussi son reste d'ordre n $R_n = \frac{\alpha^{n+1}}{1 - \alpha}$.

Par contre si $|\alpha| \geq 1$, la série des α^k diverge. À ce stade ce n'est pas tout à fait évident, mais on va établir un résultat qui va nous permettre de conclure.

Théorème 6.1 Si une série de terme général u_n est convergente, alors $\lim_{n \rightarrow +\infty} u_n = 0$.

Application pour la série géométrique des α^n : si $|\alpha| \geq 1$, pour tout n , $|\alpha^n| \geq 1$, donc le terme général ne tend pas vers 0 : la série diverge.

Remarque. Il ne suffit pas que le terme général d'une série tende vers 0 pour qu'elle converge.

Prenons par exemple la série de terme général $u_n = \frac{1}{n+1}$. Il tend vers 0 et pourtant la série diverge car, pour les sommes partielles, on a

$$S_{2n-1} - S_{n-1} = \frac{1}{n+1} + \dots + \frac{1}{2n} \geq n \frac{1}{2n} = \frac{1}{2}.$$

Et si la série convergeait, en notant S sa somme, on devrait avoir

$$\lim_{n \rightarrow +\infty} (S_{2n-1} - S_{n-1}) = S - S = 0.$$

Ceci est absurde !

Autre exemple : lien avec une suite

Soit la suite de terme général x_n . On pose $u_n = x_{n+1} - x_n$, pour tout entier naturel n , d'où

$$S_n = \sum_{k=0}^n u_k = x_{n+1} - x_0,$$

et la série des u_n converge si et seulement si la suite des x_n est convergente. En cas de convergence, on aura $S = \left(\lim_{n \rightarrow +\infty} x_n \right) - x_0$.

Ces deux exemples (séries géométriques et séries dont le terme général est la différence de deux termes consécutifs d'une suite) sont des cas particuliers, où l'on sait calculer les sommes partielles. Le plus souvent ceci n'est pas possible.

Convergence absolue

On introduit alors la notion de convergence absolue.

Définition 6.3 Soit u une série de terme général u_n . Elle est dite **absolument convergente** si la série des $|u_p|$ est convergente.

Proposition 6.1 Pour qu'une série numérique de terme général u_n converge absolument, il faut et il suffit que la suite $\left(\sum_{k=0}^n |u_k| \right)_{n \in \mathbb{N}}$ soit majorée.

Preuve. En effet, la suite réelle $\left(\sum_{k=0}^n |u_k| \right)_{n \in \mathbb{N}}$ est croissante. Si elle n'est pas majorée, elle tend vers $+\infty$ quand n tend vers $+\infty$. Sinon elle converge vers $\sup_{n \in \mathbb{N}} \left(\sum_{k=0}^n |u_k| \right)$. \square

Théorème 6.2 La convergence absolue implique la convergence.

Dans ce qui suit, on va pratiquement toujours chercher la convergence absolue, ceci non par masochisme, mais parce que la structure d'ordre sur \mathbb{R} va nous permettre, pour les séries à terme positif, d'établir des critères de comparaison dont on déduira des critères de convergence.

Un dernier résultat d'ordre général avant de passer aux séries à termes positifs :

Théorème 6.3 On ne change pas la nature (convergence ou divergence) d'une série si on en modifie un nombre fini de termes.

Preuve. Soit une série de terme général u_n . On se donne un nombre fini, p , d'indices : n_1, n_2, \dots, n_p , et on remplace u_{n_k} (pour $k = 1, \dots, p$) par une valeur donnée a_k . Notons v la série de terme général v_n défini par $v_{n_k} = a_k$ si $k \in \{1, 2, \dots, p\}$ et $v_n = u_n$ sinon.

Pour tout $n > \sup \{n_1, n_2, \dots, n_p\}$, on a de manière évidente

$$V_n = U_n + \sum_{k=1}^p (a_k - u_{n_k})$$

donc $V_n - U_n$ est constant par rapport à $n > \sup\{n_1, n_2, \dots, n_p\}$. On a donc convergence de la série des u_n si et seulement si celle des v_n converge (même équivalence pour la convergence absolue). \square

6.2.2 Critères de convergence pour les séries à terme général positif

Dans toute cette partie, nous nous intéressons uniquement au cas des séries de terme général positif. Nous connaissons la nature des séries géométriques. Après avoir établi des critères de comparaison, nous en déduisons des critères de convergence.

Critères de comparaison

Théorème 6.4 *Soient deux séries u et v de termes généraux u_n et v_n positifs, et vérifiant l'inégalité $u_n \leq v_n$ à partir d'un certain rang. Alors si u diverge, v diverge ; et si v converge, u converge.*

La phrase à partir d'un certain rang signifie : $\exists n_0 \in \mathbb{N}, \forall n \geq n_0 \dots$

Comme on conclut sur la nature des séries on fait la justification en supposant les u_n et v_n positifs pour tout n et l'hypothèse $u_n \leq v_n$ valable aussi pour tout n (utilisation du théorème 6.3).

En sommant les inégalités on obtient $U_n \leq V_n$ pour les sommes partielles. Donc si v converge, la suite des V_n est majorée, celle des U_n aussi, donc la série u converge (application de la proposition 6.1). Par contraposition si u diverge, v diverge.

Corollaire 6.1 *Soient deux séries u et v de termes généraux u_n et v_n positifs à partir d'un certain rang. S'il existe deux constantes a et b strictement positives telles que $au_n \leq v_n \leq bu_n$ à partir d'un certain rang, les deux séries sont de même nature.*

Preuve. En utilisant les propriétés 6.1, la série des u_n et celle des au_n sont de même nature. Donc si v converge, d'après le théorème précédent, celle des au_n converge, donc celle des u_n aussi ; si u converge, celle des bu_n converge, donc v converge. Ainsi les séries sont simultanément convergentes, donc aussi divergentes. \square

Corollaire 6.2 *Soient deux séries u et v de termes généraux u_n et v_n positifs à partir d'un certain rang. Si $\lim_{n \rightarrow +\infty} \frac{u_n}{v_n} = \lambda$ existe avec $\lambda > 0$, les deux séries sont de même nature.*

Preuve. L'existence de la limite des u_n/v_n suppose les v_n non nuls, donc strictement positifs à partir d'un certain rang n_0 .

Soit $\varepsilon \in]0, \lambda[$. Il existe $n_1 \geq n_0$ tel que pour tout $n \geq n_1$,

$$\lambda - \varepsilon \leq \frac{u_n}{v_n} \leq \lambda + \varepsilon \Leftrightarrow (\lambda - \varepsilon)v_n \leq u_n \leq (\lambda + \varepsilon)v_n.$$

D'après le corollaire précédent, les deux séries sont de même nature. □

Séries de Riemann.

Théorème 6.5 Soit $\alpha \in \mathbb{R}$. La série de terme général $1/n^\alpha$, $n \geq 1$, converge si et seulement si $\alpha > 1$.

Preuve. On a déjà vu que la série de terme général $1/(n+1)$ était divergente. Or pour tout entier $N \geq 1$,

$$\sum_{n=1}^N \frac{1}{n} = \sum_{m=0}^{N-1} \frac{1}{m+1}.$$

Donc la série de Riemann diverge si $\alpha = 1$.

Notons $u_n = 1/n^\alpha$ pour $n \geq 1$. Si $\alpha < 1$, alors $\lim_{n \rightarrow +\infty} nu_n = \lim_{n \rightarrow +\infty} n^{1-\alpha} = +\infty$. Donc il existe un entier $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geq n_0$,

$$nu_n \geq 1 \iff u_n \geq \frac{1}{n}.$$

Par comparaison, la série de terme général u_n diverge.

Si $\alpha > 1$, posons

$$v_n = \frac{1}{\alpha-1} \left(\frac{1}{n^{\alpha-1}} - \frac{1}{(n+1)^{\alpha-1}} \right).$$

$v_n = x_{n+1} - x_n$ est la différence de deux termes consécutifs de la suite $x_n = -\frac{1}{(\alpha-1)n^{\alpha-1}}$. Comme la suite des x_n converge (elle tend vers zéro), la série des v_n est convergente. De plus

$$\begin{aligned} v_n &= \frac{1}{\alpha-1} \frac{1}{(n+1)^{\alpha-1}} \left(\left(\frac{n+1}{n} \right)^{\alpha-1} - 1 \right) \\ &= \frac{1}{\alpha-1} \frac{1}{(n+1)^{\alpha-1}} \left(\exp \left((\alpha-1) \ln \frac{n+1}{n} \right) - 1 \right) \\ &= \frac{1}{\alpha-1} \frac{1}{(n+1)^{\alpha-1}} \left(\exp \left((\alpha-1) \ln \left(1 + \frac{1}{n} \right) \right) - 1 \right) \\ &\approx \frac{1}{\alpha-1} \frac{1}{(n+1)^{\alpha-1}} \left(\exp \left((\alpha-1) \frac{1}{n} \right) - 1 \right) \\ &\approx \frac{1}{\alpha-1} \frac{1}{(n+1)^{\alpha-1}} \left(1 + \frac{\alpha-1}{n} - 1 \right) = \frac{1}{n(n+1)^{\alpha-1}} \end{aligned}$$

et ainsi $\lim_{n \rightarrow +\infty} \frac{u_n}{v_n} = 1$. □

Règles de convergence

Venons en aux **critères de convergence pour les séries à termes positifs**.

Règle de Riemann :

Théorème 6.6 Soit une série u de terme général positif u_n . Supposons qu'il existe α réel tel que $\lim_{n \rightarrow +\infty} n^\alpha u_n = \lambda$ avec $\lambda > 0$. Si $\alpha > 1$, la série converge, et si $\alpha \leq 1$, la série diverge.

Remarque. S'il existe $\alpha > 1$ tel que $\lim_{n \rightarrow +\infty} n^\alpha u_n = 0$, la série converge ; et s'il existe $\alpha \leq 1$ tel que $\lim_{n \rightarrow +\infty} n^\alpha u_n = +\infty$, la série diverge.

Règle de Cauchy :

Théorème 6.7 (Règle de Cauchy) Soit une série u de terme général positif u_n . Supposons que $\lim_{n \rightarrow +\infty} \sqrt[n]{u_n} = \lambda$ existe. Si $\lambda < 1$, la série converge ; si $\lambda > 1$, la série diverge ; et si $\lambda = 1$, on ne peut pas conclure.

Remarque. En cas de convergence, si $k \in]0, 1[$, et n_0 entier tel que $\forall n \geq n_0, \sqrt[n]{u_n} \leq k$, on peut évaluer le reste d'ordre $n \geq n_0 - 1$. En effet pour tout $p \geq n_0$, on a $u_p \leq k^p$. Donc

$$\sum_{p=n+1}^q u_p \leq k^{n+1} + \dots + k^q = \frac{k^{n+1} - k^{q+1}}{1 - k}.$$

Si q tend vers $+\infty$, on a $R_n \leq \frac{k^{n+1}}{1 - k}$. Une telle majoration permet de trouver n pour avoir une incertitude connue.

Règle de d'Alembert :

Théorème 6.8 (Règle de d'Alembert) Soit une série de terme général u_n strictement positif à partir d'un certain rang, et telle que $\lim_{n \rightarrow +\infty} \frac{u_{n+1}}{u_n} = \lambda$. Si $\lambda < 1$, la série converge ; si $\lambda > 1$, elle diverge. Si $\lambda = 1$, on ne peut pas conclure.

Preuve. Si $\lambda < 1$, avec $\varepsilon > 0$ tel que $\lambda + \varepsilon = k < 1$, et n_0 tel que pour tout $n \geq n_0, \frac{u_{n+1}}{u_n} \leq \lambda + \varepsilon = k < 1$, le lemme qui précède permet de conclure.

Si $\lambda > 1$, avec $\varepsilon > 0$ tel que $\lambda - \varepsilon \geq 1$, et n_0 tel que pour tout $n \geq n_0, \frac{u_{n+1}}{u_n} \geq \lambda - \varepsilon \geq 1$, il y a divergence car u_n ne tend pas vers 0.

Si $\lambda = 1$, la limite étant atteinte par valeurs supérieures, il existe n_0 tel que pour tout $n \geq n_0, 1 \leq \frac{u_{n+1}}{u_n} \leq 1 + \varepsilon$. On conclut à la divergence, encore par le lemme précédent.

Si $\lambda = 1$, l'exemple de $u_n = 1/n^\alpha$, qui donne $\frac{u_{n+1}}{u_n} = \left(\frac{n}{n+1}\right)^\alpha$, montre que suivant les valeurs de α , il y aura convergence ou divergence. \square

Remarque. Évaluation du reste d'ordre n , en cas de convergence. Supposons trouvés

$k \in]0, 1[$ et n_0 tel que $\forall n \geq n_0, \frac{u_{n+1}}{u_n} \leq k$. On a donc

$$u_{n+1} \leq k u_n, u_{n+2} \leq k^2 u_n, \dots, u_{n+p} \leq k^p u_n,$$

donc

$$R_n = \sum_{p=1}^q u_{n+p} \leq u_n (k + k^2 + \dots + k^q) \leq k u_n \frac{1 - k^q}{1 - k} \leq \frac{k u_n}{1 - k}.$$

Si on veut, sans calculer les u_n , estimer le nombre de termes à calculer pour obtenir une incertitude donnée, avec les mêmes hypothèses on a

$$u_{n+1} \leq k^{n-n_0+1} u_{n_0}, \dots, u_{n+p} \leq k^{n-n_0+p} u_{n_0},$$

ce qui conduit à $R_n \leq \frac{k^{n-n_0+1}}{1-k} u_{n_0}$.

Lien entre les règles de d'Alembert et de Cauchy : si le critère de d'Alembert ne permet pas de conclure (limite égale à 1), celui de Cauchy non plus.

6.2.3 Méthodologie

Pour étudier la convergence ou non d'une série de terme général u_n , on suivra le canevas suivant.

1. D'abord on justifie que u_n est bien défini à partir d'un certain entier qu'on précise, et on étudie le signe de u_n .
2. Deux cas se présentent :
 - soit tous les u_n sont positifs ;
 - soit tous les u_n ne sont pas positifs, auquel cas on considère $v_n = |u_n|$.
 Dans le premier cas, $v_n = |u_n| = u_n$.
3. On essaie d'appliquer les critères de comparaison ou de convergence à la série des v_n , ce qui est possible car $v_n \geq 0$. Si cette dernière converge, alors on obtient la convergence absolue de la série des u_n .
4. Si la série des v_n converge via les règles de convergence de Cauchy ou de d'Alembert, on peut donner une estimation du reste d'ordre n .
5. Si la série des v_n diverge ou si on ne peut pas conclure, alors il y aura des questions supplémentaires pour étudier la série des u_n .

6.3 Intégration sur un segment

Dans la suite, $[a, b]$ est un intervalle fermé de \mathbb{R} avec $-\infty < a < b < +\infty$.

6.3.1 Intégrale des fonctions en escalier

Une subdivision croissante de points, σ , de $[a, b]$ est une suite finie strictement croissante de points de $[a, b]$ telle que :

$$a = a_0 < a_1 < \dots < a_n = b.$$

L'ensemble des subdivisions de $[a, b]$ est ordonné par la relation : σ' est plus finie que σ si et seulement si $\sigma \subset \sigma'$. Autrement dit, σ' contient tous les points de σ , i.e. si $\sigma = \{a_0, a_1, \dots, a_n\}$ et $\sigma' = \{b_0, b_1, \dots, b_p\}$:

$$\forall i \in \{1, 2, \dots, n\}, \exists j \in \{1, 2, \dots, p\}, b_j = a_i.$$

Définition 6.4 Soit $f : [a, b] \rightarrow \mathbb{R}$. On dit que f est une fonction en escalier s'il existe une subdivision σ de $[a, b]$, $\sigma = \{x_0 < x_1 < \dots < x_n\}$, telle que f soit constante sur chacun des intervalles $]x_i, x_{i+1}[$, $0 \leq i \leq n-1$.

Une telle subdivision est dite associée à f .

Définition 6.5 Soit f une fonction en escalier sur $[a, b]$ et soit $\sigma = \{x_0 < x_1 < \dots < x_n\}$ une subdivision associée à f . On pose

$$I(f, \sigma) = \sum_{i=0}^{n-1} (x_{i+1} - x_i) c_i,$$

où $c_i = f|_{]x_i, x_{i+1}[}$.

Interprétation géométrique : $I(f, \sigma)$ représente la somme algébrique des aires des rectangles.

Proposition 6.2 Le réel $I(f, \sigma)$ ne dépend pas de la subdivision σ associée à f .

$I(f, \sigma)$ ne dépendant pas de σ on notera simplement $I(f)$.

Définition 6.6 Soit f une fonction en escalier définie sur $[a, b] \subset \mathbb{R}$. L'intégrale de f sur $[a, b]$ est

$$I(f) = \int_a^b f(x) dx = I(f, \sigma), \quad \forall \sigma \text{ subdivision associée à } f.$$

Propriétés 6.2 On considère f et g des fonctions en escalier sur $[a, b]$.

1. **Relation de Chasles** : soit $c \in [a, b]$. Alors $\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$.
2. **Linéarité de l'intégrale** : pour tout $(\lambda, \mu) \in \mathbb{R}^2$

$$\int_a^b (\lambda f(x) + \mu g(x)) dx = \lambda \int_a^b f(x) dx + \mu \int_a^b g(x) dx.$$

3. **Croissance** : si pour tout $x \in [a, b]$, $f(x) \leq g(x)$, alors $\int_a^b f(x) dx \leq \int_a^b g(x) dx$.

Conséquence : si $f \geq 0$ sur $[a, b]$, alors $\int_a^b f(x) dx \geq 0$.

4. **Majoration** : $\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx$.

Conséquence : s'il existe k tel que $|f(x)| \leq k$ pour tout $x \in [a, b]$, alors

$$\left| \int_a^b f(x) dx \right| \leq k(b - a).$$

6.3.2 Fonctions intégrables (au sens de Riemann)

Maintenant que l'on sait intégrer des fonctions en escalier, nous allons pouvoir intégrer des fonctions plus générales en les encadrant par des fonctions en escalier.

Soit alors une subdivision de $[a, b]$, $\sigma = \{x_0 < x_1 < \dots < x_n\}$, et une fonction $f : [a, b] \rightarrow \mathbb{R}$ **bornée** :

$$\exists M \in \mathbb{R}_+, \quad \forall x \in [a, b], \quad |f(x)| \leq M \iff -M \leq f(x) \leq M.$$

On introduit :

$$m_i = \inf\{f(x), x \in [x_i, x_{i+1}]\} \text{ et } M_i = \sup\{f(x), x \in [x_i, x_{i+1}]\}$$

qui existent puisque f est bornée.

On pose $\Delta x_i = x_{i+1} - x_i$ et on introduit les sommes supérieure et inférieure de Riemann :

$$S(f, \sigma) = \sum_{i=0}^{n-1} M_i \Delta x_i, \text{ et } I(f, \sigma) = \sum_{i=0}^{n-1} m_i \Delta x_i.$$

Comme f est une fonction bornée, pour toute subdivision σ ,

$$-M(b-a) \leq I(f, \sigma) \leq S(f, \sigma) \leq M(b-a).$$

Lemme 6.1 *Si σ' est plus fine que σ , i.e. $\sigma \subset \sigma'$, alors $S(f, \sigma') \leq S(f, \sigma)$ et $I(f, \sigma) \leq I(f, \sigma')$.*

On définit alors

$$\inf_{\sigma} S(f, \sigma) = \overline{\int_a^b} f(x) dx, \text{ et } \sup_{\sigma} I(f, \sigma) = \underline{\int_a^b} f(x) dx.$$

Lemme 6.2 *Alors : $\underline{\int_a^b} f(x) dx \leq \overline{\int_a^b} f(x) dx$.*

Définition 6.7 *S'il y a égalité, on dit que f est intégrable (au sens de Riemann) et on note :*

$$\int_a^b f(x) dx = \underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx$$

l'intégrale de f sur $[a, b]$.

Si f est intégrable, alors on peut « encadrer » f entre deux fonctions en escalier g et h telles que

$$g \leq f \leq h \text{ et } \int_a^b (h - g)(x) dx \leq \varepsilon.$$

Remarque. [Important] On n'intègre que des fonctions bornées. Donc quand on dit d'une fonction qu'elle est intégrable, il est sous-entendu qu'elle est bornée.

Théorème 6.9 *On considère f et g deux fonctions définies sur $[a, b]$, intégrables sur $[a, b]$.*

1. **Linéarité** : pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $\lambda f + \mu g$ est intégrable et

$$\int_a^b (\lambda f(x) + \mu g(x)) dx = \lambda \int_a^b f(x) dx + \mu \int_a^b g(x) dx.$$

2. **Produit** : fg est intégrable.

3. **Relation de Chasles** : pour tout $c \in [a, b]$, f est intégrable sur $[a, c]$ et sur $[c, b]$, et

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

4. **Majoration** : $|f|$ est intégrable et $\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx$.

5. **Croissance** : si pour tout $x \in [a, b]$, $f(x) \leq g(x)$, alors $\int_a^b f(x)dx \leq \int_a^b g(x)dx$.

Remarque. [importante] Par convention $\int_b^a f(x)dx = - \int_a^b f(x)dx$ et $\int_a^a f(x)dx = 0$.

Classes de fonctions intégrables

Théorème 6.10 Si f est continue sur $[a, b]$, alors f est intégrable.

Proposition 6.3 Si f est une fonction continue et positive sur $[a, b]$ telle que $\int_a^b f(x)dx = 0$. Alors $f = 0$ sur $[a, b]$.

Terminons par une autre classe de fonctions intégrables.

Définition 6.8 Soit $f : [a, b] \rightarrow \mathbb{R}$. La fonction f est dite continue par morceaux sur un intervalle $[a, b]$ s'il existe une subdivision $\sigma = (a_0 < a_1 < \dots < a_n)$ du segment $[a, b]$ et des fonctions f_i , $i = 0, 1, \dots, n_1$, continues sur $[a_i, a_{i+1}]$, telles que $f = f_i$ sur $]a_i, a_{i+1}[$.

Théorème 6.11 Une fonction f continue par morceaux sur $[a, b]$ est intégrable. De plus

$$\int_a^b f(x)dx = \sum_{i=0}^{n-1} \int_{a_i}^{a_{i+1}} f_i(x)dx.$$

6.3.3 Primitives

Définition 6.9 Soit $f : [a, b] \rightarrow \mathbb{R}$. On appelle primitive de f toute application $F : [a, b] \rightarrow \mathbb{R}$ dérivable telle que pour tout $x \in [a, b]$, $F'(x) = f(x)$.

Remarquez que si F est une primitive de f , alors pour toute constante $C \in \mathbb{R}$, $t \mapsto F(t) + C$ est aussi une primitive de f . Donc il n'y a jamais unicité de la primitive.

Définition 6.10 Soit f une fonction continue sur $[a, b]$ (f est donc intégrable sur $[a, x]$ pour tout $x \in [a, b]$). On définit alors une fonction F sur $[a, b]$ par

$$(6.1) \quad F(x) = \int_a^x f(t)dt.$$

Propriétés 6.3 La fonction F a les propriétés suivantes :

— $F(a) = 0$;

— si $a \leq x \leq y \leq b$,

$$F(y) - F(x) = \int_x^y f(t)dt.$$

— Si f est positive sur $[a, b]$, alors F est croissante.

— $F(b) = \int_a^b f(t)dt$.

Théorème 6.12 Soit f une fonction continue sur $[a, b]$. Alors F définie par (6.1) est continue et dérivable sur $[a, b]$ et pour tout $x \in [a, b]$, $F'(x) = f(x)$.

Ainsi, d'après le théorème précédent, si f est continue sur $[a, b]$, f admet au moins une primitive. De plus si G est une autre primitive de f sur $[a, b]$, alors $(G - F)' = 0$, donc $G - F$ est une fonction constante. Comme $G(a) - F(a) = G(a)$, on en déduit

Lemme 6.3 si G est une primitive de f sur $[a, b]$, alors pour tout $x \in [a, b]$,

$$G(x) = \int_a^x f(t)dt + G(a) = F(x) + G(a).$$

Donc si on connaît une primitive G d'une fonction f sur un intervalle $[a, b]$, alors

$$\int_a^b f(x)dx = G(b) - G(a);$$

ceci quelle que soit la primitive choisie.

Donc pour calculer une intégrale, on pourra commencer par chercher une primitive de f , en sachant que si f est continue, $x \mapsto \int_a^x f(t)dt$ est la primitive de f qui s'annule en a .

Notation : une primitive sera notée $\int f(x)dx$. Ne pas oublier qu'elle est toujours déterminée à une constante près.

Exemple 6.1 (Catalogue des primitives) Ces exemples sont à retenir :

— Si $\alpha \in \mathbb{R} \setminus \{-1\}$, et si $c \in \mathbb{R}$, alors pour tout $x > c$,

$$\int (x - c)^\alpha dx = \frac{(x - c)^{\alpha+1}}{\alpha + 1} + Cte.$$

— Pour tout $x \neq c$,

$$\int \frac{1}{x - c} dx = \ln |x - c| + Cte.$$

— Si $\alpha \in \mathbb{R}^*$,

$$\int e^{\alpha x} dx = \frac{1}{\alpha} e^{\alpha x} + Cte.$$

Pour tout entier $n \in \mathbb{N}$, les fonctions $x \mapsto (x - c)^n$ sont définies sur \mathbb{R} et

$$\int (x - c)^n dx = \frac{(x - c)^{n+1}}{n + 1} + Cte.$$

Exemple 6.2 — Calculer l'intégrale de la fonction $x \mapsto x^2$ sur $[1, 2]$.

Cette fonction est continue, donc intégrable sur $[1, 2]$. On montre ensuite que $x^3/3$ est une primitive de cette fonction, puis

$$\int_1^2 x^2 dx = \frac{2^3}{3} - \frac{1^3}{3} = \frac{7}{3}.$$

— Calculer l'intégrale de la fonction $f : x \mapsto e^{-x}$ sur $[0, 4]$.

La fonction $x \mapsto -e^{-x} + 2$ est une primitive de la fonction f , donc

$$\int_0^4 f(x) dx = (-e^{-4} + 2) - (-e^{-0} + 2) = 1 - e^{-4}.$$

Notez que ces fonctions étant positives, leurs intégrales doivent être aussi positives !

6.3.4 Changement de variable et intégration par parties

Nous allons donner deux résultats très utiles pour calculer les intégrales ou les primitives.

Intégration par parties

Une fonction f est de classe C^1 sur $[a, b]$ si elle est continue, dérivable sur $[a, b]$ et si sa dérivée f' est continue sur $[a, b]$. En utilisant la formule de dérivation $(fg)' = f'g + fg'$, on obtient

Théorème 6.13 (Formule d'intégration par parties) Soient f et g deux fonctions de $[a, b]$ dans \mathbb{R} de classe C^1 . Alors :

$$\begin{aligned} \int_a^b f(x)g'(x) dx &= [f(x)g(x)]_a^b - \int_a^b f'(x)g(x) dx \\ &= f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x) dx. \end{aligned}$$

Exemple 6.3 Calculer une primitive de la fonction $x \mapsto xe^x$.

Ceci n'a rien d'évident. Une intégration par parties permet de le faire. Posons $u(x) = x$ et $v'(x) = e^x$. On a donc $u'(x) = 1$ et on convient que $v(x) = e^x$. Par intégration par parties :

$$\int_0^t xe^x dx = [u(x)v(x)]_0^t - \int_0^t u'(x)v(x) dx = te^t - \int_0^t e^x dx = te^t - e^t + 1.$$

Ainsi la primitive de $x \mapsto xe^x$ qui s'annule en 0 est $t \mapsto te^t - e^t + 1$.

Exemple 6.4 Calculer une primitive de la fonction $x \mapsto \ln x$.

On pose $u(x) = \ln x$ et $v(x) = 1$. Alors $u'(x) = 1/x$ et $v(x) = x$. Par intégration par parties pour $t \in \mathbb{R}_+^*$:

$$\int_1^t \ln x dx = [x \ln x]_1^t - \int_0^t 1 dx = t \ln t - t.$$

Donc une primitive de $x \mapsto \ln x$ est donnée par $t \mapsto t \ln t - t + \text{Cte}$.

Exemple 6.5 Pour $n \in \mathbb{N}$, $a \in \mathbb{R}^*$, $I_n(t) = \int_0^t x^n e^{ax} dx$. Trouver une relation de récurrence liant $I_{n+1}(t)$ et $I_n(t)$.

On a $I_{n+1}(t) = \int_0^t x^{n+1} e^{ax} dx$. Posons $u(x) = x^{n+1}$ et $v'(x) = e^{ax}$. Alors $u'(x) = (n+1)x^n$ et $v(x) = e^{ax}/a$. Ainsi

$$I_{n+1}(t) = \frac{1}{a} t^{n+1} e^{at} - \frac{n+1}{a} I_n(t).$$

Changement de variable

Théorème 6.14 Soit $\phi : [a, b] \rightarrow \mathbb{R}$ de classe C^1 , c'est-à-dire dérivable et à dérivée continue. Soit f définie et continue sur $\phi([a, b])$. Alors :

$$\int_{\phi(a)}^{\phi(b)} f(t) dt = \int_a^b f(\phi(x)) \phi'(x) dx.$$

C'est la formule de changement de variable.

En pratique quand on fait un changement de variable pour calculer une intégrale du type : $\int_a^b f(\phi(x)) \phi'(x) dx$, on commence par dire : « posons $t = \phi(x)$ ». Ensuite on utilise la formule mnémotechnique : $\phi'(x) dx = dt$. Enfin on ajuste les bornes d'intégration en remarquant que si x parcourt l'intervalle $[a, b]$, alors $t = \phi(x)$ parcourt l'intervalle $[\phi(a), \phi(b)]$.

Exemple 6.6 Calculer l'intégrale $\int_0^1 \frac{1}{x+1} dx$.

On pose $t = x + 1$. On a donc $dt = 1 \cdot dx = dx$. Donc

$$\int_0^1 \frac{1}{x+1} dx = \int_1^2 \frac{1}{t} dt = [\ln t]_1^2 = \ln 2 - \ln 1 = \ln 2.$$

6.3.5 Calcul de primitives

Reconnaître une dérivée

On sait que si u est une fonction dérivable, la dérivée de $\ln(u)$ (sous réserve que $\ln(u)$ ait un sens) est $\frac{u'}{u}$. On sait aussi que la dérivée de e^u est $u'e^u$. Et on sait que la dérivée de u^a (où a est une constante) est, sous réserve que u^a ait un sens, au^{a-1} . On peut présenter ces résultats « à l'envers » pour obtenir le tableau suivant :

Fonctions	Primitives	Remarques
$\frac{u'}{u}$	$\ln(u) + c$	u est à valeurs strictement positives.
$u'e^u$	$e^u + c$	
$u'u^a$	$\frac{1}{a+1}u^{1+a} + c$	u est à valeurs strictement positives et $a \neq -1$.

Exemple 6.7 Calculer $\int_0^1 2x(x^2 + 1)^{19} dx$.

On reconnaît la dérivée de la fonction $(1/20)(x^2 + 1)^{20}$. Donc

$$\int_0^1 2x(x^2 + 1)^{19} dx = \frac{1}{20} [(x^2 + 1)^{20}]_0^1 = \frac{1}{20} (2^{20} - 1).$$

Fractions rationnelles

Ce sont des fonctions du type

$$f(x) = \frac{P(x)}{Q(x)} = \frac{\sum_{i=0}^m a_i x^i}{\sum_{j=0}^n b_j x^j}.$$

Remarquez qu'elles ne sont pas définies en général sur \mathbb{R} tout entier. Par exemple, $f(x) = \frac{1}{x}$. D'abord on commence par factoriser le polynôme Q :

$$Q(x) = a \prod_{i=0}^r (x - \alpha_i)^{n_i} \prod_{j=0}^s (x^2 + b_j x + c_j)^{m_j},$$

avec n_i et m_j entiers supérieurs à 1, et $b_j^2 - 4c_j < 0$. Par exemple :

$$Q(x) = 3x^4 + 12x^3 + 18x^2 + 24x + 24 = 3(x + 2)^2(x^2 + 2).$$

Nous nous limiterons ici au cas où $m_j = 1$.

Pour ce type de fonctions, il existe toujours une décomposition dite en éléments simples de la forme

$$f(x) = R(x) + \sum_{i=0}^r \sum_{k=1}^{n_i} \frac{\lambda_{ik}}{(x - \alpha_i)^k} + \sum_{j=0}^s \sum_{k=1}^{m_j} \frac{r_{jk}x + q_{jk}}{(x^2 + b_jx + c_j)^k},$$

où R est un polynôme, λ_{ik} , r_{jk} et q_{jk} sont des réels. Par exemple :

$$f(x) = \frac{1}{(x-1)(x+2)} = \frac{1/3}{x-1} - \frac{1/3}{x+2}.$$

On sait trouver une primitive pour R . Si $n_i = 1$, on a à trouver une primitive de $\frac{1}{x - \alpha_i}$, qui est $\ln|x - \alpha_i|$. Si $n_i > 1$, alors il faut calculer $\int \frac{1}{(x - \alpha_i)^k} = \frac{1}{(1 - k)(x - \alpha_i)^{k-1}} + \text{Cte}$.

Il reste donc les fractions dont le dénominateur est un polynôme de degré 2. Cherchons une primitive de ce dernier terme pour $m_j = 1$.

$$\begin{aligned} \frac{r_{jk}x + q_{jk}}{x^2 + b_jx + c_j} &= \frac{r_{jk}}{2} \frac{2x + b_j}{x^2 + b_jx + c_j} + \frac{q_{jk} - \frac{r_{jk}b_j}{2}}{x^2 + b_jx + c_j} \\ &= \frac{r_{jk}}{2} \frac{2x + b_j}{x^2 + b_jx + c_j} + \left(q_{jk} - \frac{r_{jk}b_j}{2} \right) \frac{1}{\left(x + \frac{b_j}{2}\right)^2 + \left(c_j - \frac{b_j^2}{4}\right)} \end{aligned}$$

Or $b_j^2 - 4c_j < 0$. On pose $\nu_j^2 = c_j - \frac{b_j^2}{4} > 0$. Ainsi

$$\frac{r_{jk}x + q_{jk}}{x^2 + b_jx + c_j} = \frac{r_{jk}}{2} \frac{2x + b_j}{x^2 + b_jx + c_j} + \left(q_{jk} - \frac{r_{jk}b_j}{2} \right) \frac{1}{\left(x + \frac{b_j}{2}\right)^2 + \nu_j^2}.$$

On peut maintenant calculer une primitive pour cette fonction :

$$\int \frac{r_{jk}x + q_{jk}}{x^2 + b_jx + c_j} dx = \frac{r_{jk}}{2} \ln(x^2 + b_jx + c_j) + \left(q_{jk} - \frac{r_{jk}b_j}{2} \right) \int \frac{1}{\left(x + \frac{b_j}{2}\right)^2 + \nu_j^2} dx.$$

On fait le changement de variable $t = \left(x + \frac{b_j}{2}\right)$, qui donne

$$\begin{aligned} \int \frac{1}{\left(x + \frac{b_j}{2}\right)^2 + \nu_j^2} dx &= \int \frac{dt}{t^2 + \nu_j^2} = \frac{1}{\nu_j} \int \frac{d(t/\nu_j)}{1 + (t/\nu_j)^2} \\ &= \frac{1}{\nu_j} \int \frac{du}{1 + u^2} \end{aligned}$$

par changement de variable $t = \nu_j u$.

Proposition 6.4 Une primitive de la fonction $u \mapsto \frac{1}{1+u^2}$ est donnée par $t \mapsto \text{Arctan}(t) + Cte$.

La fonction Arctan est la fonction inverse de la fonction $\tan :]-\pi/2, \pi/2[\rightarrow \mathbb{R}$. C'est une fonction définie sur \mathbb{R} , continue, croissante et telle que

$$\lim_{x \rightarrow -\infty} \text{Arctan}(x) = -\frac{\pi}{2}, \quad \lim_{x \rightarrow +\infty} \text{Arctan}(x) = \frac{\pi}{2}.$$

C'est donc une fonction bornée.

Exemple 6.8 C'est un exercice type.

1. Trouver a, b et c réels tels que $\frac{1}{x(x^2+x+1)} = \frac{a}{x} + \frac{bx+c}{x^2+x+1}$.

2. Calculer l'intégrale $\int_1^2 \frac{1}{x(x^2+x+1)} dx$.

Tout d'abord, on remarque que

$$\frac{1}{x(x^2+x+1)} = \frac{1}{x} - \frac{x+1}{x^2+x+1}.$$

Donc $a = 1$ et $b = c = -1$. Ensuite on connaît une primitive de $1/x$. Reste à trouver une primitive du terme restant.

$$\begin{aligned} \frac{x+1}{x^2+x+1} &= \frac{1}{2} \frac{2x+1}{x^2+x+1} + \frac{1}{2} \frac{1}{x^2+x+1} \\ &= \frac{1}{2} \frac{2x+1}{x^2+x+1} + \frac{1}{2} \frac{1}{(x+1/2)^2 + (\sqrt{3}/2)^2} \\ &= \frac{1}{2} \frac{2x+1}{x^2+x+1} + \frac{1}{\sqrt{3}} \frac{2/\sqrt{3}}{1 + ((2/\sqrt{3})(x+1/2))^2}. \end{aligned}$$

On regroupe tous les termes :

$$\begin{aligned} \int_1^2 \frac{1}{x(x^2+x+1)} dx &= \int_1^2 \frac{dx}{x} - \frac{1}{2} \int_1^2 \frac{2x+1}{x^2+x+1} dx - \frac{1}{\sqrt{3}} \int_1^2 \frac{\frac{2}{\sqrt{3}}}{1 + (\frac{2}{\sqrt{3}}(x+1/2))^2} dx \\ &= [\ln(x)]_1^2 - \frac{1}{2} [\ln(x^2+x+1)]_1^2 - \frac{1}{\sqrt{3}} \int_{\sqrt{3}}^{5/\sqrt{3}} \frac{1}{1+t^2} dt \\ &= \ln(2) - \frac{1}{2} \ln(7) + \frac{1}{2} \ln(3) - \frac{1}{\sqrt{3}} [\text{Arctan}(t)]_{\sqrt{3}}^{5/\sqrt{3}} \\ &= \ln\left(2\sqrt{\frac{3}{7}}\right) - \frac{1}{\sqrt{3}} [\text{Arctan}(5/\sqrt{3}) - \text{Arctan}(\sqrt{3})]. \end{aligned}$$

Ceci vaut approximativement 0,16, ce qui est bien strictement positif (cohérent avec le signe de la fonction à intégrer).

Fonctions de la forme $P(x)e^{ax}$

P désigne un polynôme et a un nombre réel. Lorsque le degré de P est faible (inférieur ou égal à 2), on peut intégrer par parties. Cependant, le nombre d'intégration par parties à effectuer est égal au degré de P .

Exemple 6.9 $\int x^2 e^x dx = x^2 e^x - \int 2x e^x dx = x^2 e^x - 2x e^x + 2 \int e^x dx = x^2 e^x - 2x e^x + 2e^x + Cte.$

On remarque que la primitive est de la forme $Q(x)e^{ax} + Cte$ avec $degQ = degP$. Ce résultat est général. On trouve Q en dérivant, ce qui donne :

$$aQ + Q' = P.$$

L'existence de Q est assurée, car il n'est pas difficile de vérifier que l'équation ci-dessus donne un système triangulaire dont les inconnues sont les coefficients de Q .

Exemple 6.10 Calculer l'intégrale $\int_0^2 x^2 e^{-2x} dx.$

On recherche une primitive sous la forme $F(x) = (ax^2 + bx + c)e^{-2x}$. Soit en dérivant

$$F'(x) = (-2ax^2 - 2bx - 2c + 2ax + b)e^{-2x} = x^2 e^{-2x}.$$

Ceci donne le système suivant

$$\begin{cases} -2a & & = 1 \\ 2a & -2b & = 0 \\ & b & -2c = 0 \end{cases} \Leftrightarrow \begin{cases} a & = -1/2 \\ b & = a \\ c & = b/2 \end{cases} \Leftrightarrow a = b = -1/2, c = -1/4.$$

Donc

$$\int_0^2 x^2 e^{-2x} dx = \frac{1}{4} [(-2x^2 - 2x - 1)e^{-2x}]_0^2 = \frac{1}{4} [1 - 13e^{-4}]$$

Ceci vaut environ 0,19.

6.4 Intégrales généralisées

Dans le paragraphe précédent, on a défini l'intégrale d'une fonction réelle **bornée** sur un intervalle fermé et borné $[a, b]$ pour une certaine classe de fonctions, celles dites **Riemann intégrales**. Cette classe contient en particulier les fonctions en escalier, les fonctions continues, et les fonctions continues par morceaux.

Remarque.

1. Si f est intégrable sur $[a, b]$, f est bornée sur $[a, b]$ (la réciproque est fausse).
2. Si f est continue sur $[a, b[$ (semi-ouvert, borné ou non), alors f est intégrable sur $[a, x]$ pour tout $a \leq x < b$. La fonction $F : [a, b[\rightarrow \mathbb{R}$ définie par

$$\forall a \leq x < b, F(x) = \int_a^x f(t)dt$$

est la primitive de f satisfaisant $F(a) = 0$.

Soit I un intervalle de \mathbb{R} non vide (sans autre hypothèse a priori). On note $I = (a, b)$ avec $-\infty \leq a < b \leq +\infty$, tous les intervalles $[a, b]$, $[a, b[$, $]a, b]$ ou $]a, b[$. Dans la suite, **toutes les fonctions seront supposées continues par morceaux sur I** , c'est-à-dire :

Définition 6.11 f est continue par morceaux sur I s'il existe une subdivision finie $\sigma = (a_0 = a < a_1 < \dots < a_{n-1} < a_n = b)$ de l'intervalle $I = (a, b)$ et des fonctions f_i , $0 \leq i \leq n-1$, telles que

- pour tout $1 \leq i \leq n-2$, f_i est continue sur $[a_i, a_{i+1}]$,
- f_0 est continue sur $]a, a_1]$, f_{n-1} est continue sur $[a_{n-1}, b[$,
- et sur chaque intervalle $]a_i, a_{i+1}[$, f est égale à f_i :

$$\forall x \in]a_i, a_{i+1}[, f(x) = f_i(x).$$

Exemple 6.11 — La fonction f définie sur \mathbb{R} par

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ e^{-x} & \text{si } x > 0; \end{cases}$$

est continue par morceaux sur \mathbb{R} .

— La fonction définie sur \mathbb{R}_+^* par

$$f(x) = \begin{cases} \frac{1}{x} & \text{si } 0 < x \leq 2, \\ x^2 & \text{si } x > 2; \end{cases}$$

est aussi continue par morceaux sur \mathbb{R}_+^* .

Proposition 6.5 Si f est continue par morceaux sur $I = (a, b)$, alors pour tout $a < \alpha < \beta < b$, f est intégrable sur $[\alpha, \beta]$.

6.4.1 Définitions et premières propriétés

Définition 6.12 Soit f une fonction continue par morceaux sur $[a, b[$, où $a \in \mathbb{R}$, $a < b \leq +\infty$. On dit que l'intégrale de f sur $[a, b[$ converge si la fonction $F : [a, b[\rightarrow \mathbb{R}$ définie par $F(x) = \int_a^x f(t)dt$, admet une limite réelle en b . Dans ce cas on pose

$$\int_a^b f(t)dt = \lim_{x \rightarrow b} \int_a^x f(t)dt.$$

Dans le cas contraire, on dit que l'intégrale de f sur $[a, b[$ diverge.

En cas de divergence, le symbole $\int_a^b f(t)dt$ n'a pas de sens !

Exemple 6.12 1. Soit

$$f : [1, +\infty[\rightarrow \mathbb{R} \\ x \mapsto \frac{1}{x^2}.$$

Cette fonction est continue sur $I = [1, +\infty[$. L'intégrale de f sur I est convergente. En effet, la primitive de f valant 0 en 1 est : $F(x) = 1 - 1/x$ pour tout $x \geq 1$. Cette fonction admet 1 pour limite en $+\infty$. Donc

$$\int_1^{+\infty} \frac{1}{x^2} dx = 1.$$

2. La fonction $x \mapsto x$ est localement intégrable sur \mathbb{R}_+ , car continue. Sa primitive valant 0 en 0 est $x \mapsto x^2/2$, fonction qui n'a pas de limite en $+\infty$. Donc l'intégrale sur \mathbb{R}_+ de $x \mapsto x$ diverge.

Propriétés 6.4 Si f est continue par morceaux sur $[a, b[$ et si $c \in]a, b[$, alors l'intégrale de f sur $[a, b[$ converge si et seulement si l'intégrale de f sur $[c, b[$ converge.

Preuve. Pour tout $x \in [a, b[$, on a, par la relation de Chasles :

$$F(x) = \int_a^x f(t)dt = \int_a^c f(t)dt + \int_c^x f(t)dt = \int_a^c f(t)dt + G(x).$$

Les fonctions F et G diffèrent d'une constante, donc ont simultanément une limite réelle en b (ou une absence de limite). □

On dit que les intégrales $\int_a^b f(t)dt$ et $\int_c^b f(t)dt$ sont **de même nature**.

De même,

Définition 6.13 si f est une fonction continue par morceaux sur $]a, b]$ ($b \in \mathbb{R}$, $-\infty \leq a < b$), la fonction $x \mapsto \int_x^b f(t)dt$ est définie sur $]a, b]$. Si elle admet une limite en a , l'intégrale de f sur $]a, b]$ converge et on pose

$$\int_a^b f(t)dt = \lim_{x \rightarrow a} \int_x^b f(t)dt.$$

Dans le cas contraire, l'intégrale de f est dite *divergente*.

Exemple 6.13 Quelle est la nature de l'intégrale de la fonction logarithme sur $]0, 1]$?

La fonction $f : t \mapsto \ln(t)$ est continue sur $]0, 1]$. Pour $x \in]0, 1]$,

$$\int_x^1 \ln(t)dt = -1 - x \ln(x) + x;$$

et $\lim_{x \rightarrow 0} x \ln x = 0$. Donc l'intégrale de f sur $]0, 1]$ converge et vaut -1.

Définition 6.14 Soit f une fonction continue par morceaux sur $]a, b[$ ($-\infty \leq a < b \leq +\infty$). L'intégrale de f sur $]a, b[$ converge s'il existe un c dans $]a, b[$ tel que les intégrales de f sur $]a, c[$ et sur $]c, b[$ convergent. On définit alors l'intégrale de f sur $]a, b[$ en posant :

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt.$$

Dans le cas contraire, l'intégrale de f diverge.

On remarquera que le second membre est indépendant du choix de c .

Lemme 6.4 Si les intégrales de f sur $]a, c[$ et sur $]c, b[$ convergent, il en est de même pour tout $d \in]a, b[$.

Preuve. Si $d \in]a, b[$ et $x \in]a, b[$, alors avec la relation de Chasles :

$$\int_x^d f(t)dt = \int_x^c f(t)dt + \int_c^d f(t)dt,$$

d'où

$$(6.2) \quad \lim_{x \rightarrow a} \int_x^d f(t)dt = \int_a^c f(t)dt + \int_c^d f(t)dt = \int_a^d f(t)dt.$$

On montre de même que :

$$(6.3) \quad \lim_{x \rightarrow b} \int_d^x f(t)dt = \int_d^c f(t)dt + \int_c^b f(t)dt = \int_d^b f(t)dt.$$

Puisque $\int_d^c f(t)dt = - \int_c^d f(t)dt$, on obtient alors en sommant (6.2) et (6.3) :

$$\int_a^d f(t)dt + \int_d^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt.$$

□

Propriétés 6.5 (Linéarité, positivité) On suppose que f et g sont deux fonctions réelles continue par morceaux sur un même intervalle I d'extrémités a et b .

1. Si λ est un réel non nul, alors les intégrales $\int_a^b f(x)dx$ et $\int_a^b \lambda f(x)dx$ sont de même nature (i.e. toutes deux convergentes ou toutes deux divergentes), et si elles convergent, alors

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

2. Si les intégrales $\int_a^b f(x)dx$ et $\int_a^b g(x)dx$ convergent, alors l'intégrale $\int_a^b (f(x) + g(x))dx$ converge et

$$\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$$

3. Si pour tout $x \in I$, $f(x) \geq 0$, alors $\int_a^b f(x)dx \geq 0$.

Pour prouver ces résultats, on utilise les propriétés de linéarité et de positivité de l'intégrale sur un intervalle fermé et borné, et les propriétés des limites relatives à l'addition et la multiplication par un réel.

Rappelons que $\phi : [a, b[\rightarrow \mathbb{R}$ a une limite réelle en b si et seulement si pour tout $\varepsilon > 0$, il existe $c_\varepsilon \in]a, b[$ tel que

$$\forall (x, y) \in]c_\varepsilon, b[, |\phi(x) - \phi(y)| < \varepsilon.$$

Cette propriété est appelé *critère de Cauchy*. On en déduit

Proposition 6.6 (Critère de Cauchy) *Soit f une fonction localement intégrable sur $[a, b[$ ($a \in \mathbb{R}$, $a < b \leq +\infty$). Pour que l'intégrale de f sur $[a, b[$ converge, il faut et il suffit que pour tout $\varepsilon > 0$, il existe $c_\varepsilon \in]a, b[$ tel que*

$$\forall (x, y) \in]c_\varepsilon, b[, \left| \int_x^y f(t)dt \right| < \varepsilon.$$

6.4.2 Intégrales de Riemann

Théorème 6.15 (Comportement en $+\infty$) *Soit $x_0 \in \mathbb{R}_+^*$, et $\alpha \in \mathbb{R}$. Alors l'intégrale*

$$\int_{x_0}^{+\infty} \frac{1}{t^\alpha} dt \text{ converge si et seulement si } \alpha > 1.$$

Preuve. La fonction $t \mapsto 1/t^\alpha$ est continue sur \mathbb{R}_+^* . De plus pour $x \geq 0$,

$$\int_{x_0}^x \frac{1}{t^\alpha} dt = \begin{cases} \frac{1}{1-\alpha} (x^{1-\alpha} - x_0^{1-\alpha}) & \text{si } \alpha \in \mathbb{R} \setminus \{1\}; \\ \ln x - \ln x_0 & \text{si } \alpha = 1. \end{cases}$$

Comme $\lim_{x \rightarrow +\infty} x^{1-\alpha}$ vaut $+\infty$ si $1-\alpha > 0$, et 0 si $1-\alpha < 0$, et comme $\lim_{x \rightarrow +\infty} \ln x = +\infty$, on en déduit que

$$\int_{x_0}^{+\infty} \frac{1}{t^\alpha} dt = \lim_{x \rightarrow \infty} \int_{x_0}^x \frac{1}{t^\alpha} dt = \begin{cases} \frac{1}{\alpha-1} x_0^{1-\alpha} & \text{si } \alpha > 1; \\ +\infty & \text{si } \alpha \leq 1. \end{cases}$$

□

Proposition 6.7 (Comportement en un point fini) *De même l'intégrale $\int_0^{x_0} \frac{1}{t^\alpha} dt$ converge si et seulement si $\alpha < 1$. Et plus généralement, si $-\infty < a < b < +\infty$, les intégrales $\int_a^b \frac{dt}{(t-a)^\alpha}$ et $\int_a^b \frac{dt}{(b-t)^\alpha}$ convergent si et seulement si $\alpha < 1$.*

Remarque.

1. L'intégrale $\int_0^{+\infty} \frac{1}{t^\alpha} dt$ est toujours divergente (pour tout α).
2. Si $\alpha \leq 0$, la fonction $t \mapsto 1/t^\alpha$ est continue sur \mathbb{R}_+ , donc intégrable sur $[0, x_0]$ et l'intégrale $\int_0^{x_0} \frac{1}{t^\alpha} dt$ est une intégrale « ordinaire ».

6.4.3 Fonctions à valeurs positives

Dans toute cette partie les fonctions seront supposées ne prendre que des valeurs positives.

Théorème 6.16 *Soit f une fonction continue par morceaux sur un intervalle $I = [a, b[$ à valeurs dans \mathbb{R}_+ . Pour que l'intégrale de f sur I soit convergente, il faut et il suffit que la fonction $x \mapsto \int_a^x f(t)dt$ soit majorée sur I .*

Preuve. Comme f est à valeurs positives, en posant $F(x) = \int_a^x f(t)dt$, pour tout $x < x'$, on a

$$F(x') - F(x) = \int_x^{x'} f(t)dt \geq 0;$$

i.e. F est croissante. Cette fonction admet donc une limite en b si et seulement si l'ensemble $\{F(x), x \geq a\}$ est majoré et on sait que

$$\lim_{x \rightarrow b} F(x) = \sup \{F(x), x \geq a\}.$$

□

Critères de convergence

Dans ce qui suit, $I = [a, b[$ avec $a \in \mathbb{R}, a < b \leq +\infty$, ou $I =]a, b]$ avec $b \in \mathbb{R}$ et $-\infty \leq a < b$. De plus f et g sont deux fonctions continues par morceaux sur I , à valeurs positives.

Critère 1 *Si $f(t) \leq g(t)$ pour tout $t \in I$, alors la convergence de l'intégrale $\int_a^b g(t)dt$ implique la convergence de l'intégrale $\int_a^b f(t)dt$. Par contraposition, la divergence de $\int_a^b f(t)dt$ entraîne la divergence de $\int_a^b g(t)dt$.*

Preuve. On se limite au cas où $I = [a, b[$. Par positivité de l'intégrale, pour tout $x \in I$,

$$F(x) = \int_a^x |f(t)|dt \leq \int_a^x |g(t)|dt = G(x).$$

La convergence de $\int_a^b |g(t)|dt$ implique G bornée, d'où F bornée, d'où la convergence de $\int_a^b |f(t)|dt$. □

De ce critère, on en déduit immédiatement :

Critère 2 *S'il existe α et β tels que $\alpha f(t) \leq g(t) \leq \beta f(t)$ pour tout $t \in I$, alors les deux intégrales généralisées $\int_a^b f(t)dt$ et $\int_a^b g(t)dt$ sont de même nature.*

Critère 3 Si $I = [a, b[$ et et

$$\lim_{t \rightarrow b} \frac{f(t)}{g(t)} = \lambda > 0,$$

alors les intégrales $\int_a^b g(t)dt$ et $\int_a^b f(t)dt$ sont de même nature.

De plus, si les intégrales convergent, $\int_x^b |f(t)|dt$ et $\int_x^b |g(t)|dt$ sont des infiniment petits équivalents ; si les intégrales divergent, alors $\int_a^x |f(t)|dt$ et $\int_a^x |g(t)|dt$ sont des infiniment grands équivalents.

Preuve. Comme f et g sont équivalentes en b , on peut écrire $f(t) = g(t)(1 + \varepsilon(t))$ avec $\lim_{t \rightarrow b} \varepsilon(t) = 0$. Donc il existe $c \in [a, b[$ tel que pour tout $t \geq c$, $1/2 \leq 1 + \varepsilon(t) \leq 3/2$. Donc pour tout $t \in [c, b[$,

$$\frac{1}{2}|g(t)| \leq |f(t)| \leq \frac{3}{2}|g(t)|.$$

Le critère 2 implique donc que les intégrales $\int_c^b |f(t)|dt$ et $\int_c^b |g(t)|dt$ sont de même nature. Ainsi la première partie du critère s'en déduit immédiatement.

Soit $\eta > 0$. Il existe $B \in [a, b[$ tel que si $B \leq t < b$, alors $|\varepsilon(t)| \leq \eta$. Ainsi $(1 - \eta)|g(t)| \leq |f(t)| \leq (1 + \eta)|g(t)|$.

Supposons les intégrales convergentes : pour $B \leq x \leq x'$ on aura donc

$$(1 - \eta) \int_x^{x'} |g(t)|dt \leq \int_x^{x'} |f(t)|dt \leq (1 + \eta) \int_x^{x'} |g(t)|dt.$$

Ainsi en passant à la limite quand x' tend vers l'infini :

$$(6.4) \quad (1 - \eta) \int_x^b |g(t)|dt \leq \int_x^b |f(t)|dt \leq (1 + \eta) \int_x^b |g(t)|dt.$$

Finalement pour tout $\eta > 0$, il existe $B \in [a, b[$ tel que si $x \in [B, b[$ alors (6.4) : ceci traduit l'équivalence des infiniment petits $\int_x^b |f(t)|dt$ et $\int_x^b |g(t)|dt$.

Supposons maintenant les intégrales divergentes : pour $x \geq B$ on aura

$$(1 - \eta) \int_B^x |g(t)|dt \leq \int_B^x |f(t)|dt \leq (1 + \eta) \int_B^x |g(t)|dt,$$

soit encore

$$\begin{aligned} (1 - \eta) \left[\int_a^x |g(t)|dt - \int_a^B |g(t)|dt \right] &\leq \left[\int_a^x |f(t)|dt - \int_a^B |f(t)|dt \right] \\ &\leq (1 + \eta) \left[\int_a^x |g(t)|dt - \int_a^B |g(t)|dt \right], \end{aligned}$$

d'où

$$\begin{aligned} (1 - \eta) \left[\int_a^x |g(t)|dt - \int_a^B |g(t)|dt \right] + \int_a^B |f(t)|dt &\leq \int_a^x |f(t)|dt \\ &\leq (1 + \eta) \left[\int_a^x |g(t)|dt - \int_a^B |g(t)|dt \right] + \int_a^B |f(t)|dt. \end{aligned}$$

Mais les intégrales divergent et les fonctions $|f|$ et $|g|$ sont à valeurs positives. Donc $\lim_{x \rightarrow b} |g| = +\infty$ (fonction croissante de x sans limite). On suppose B assez grand pour avoir $\int_a^x |g| > 0$ pour tout $x \geq B$. On divise la double inégalité par $\int_a^x |g|$ et on constate que le minorant tend vers $1 - \eta$, le majorant vers $1 + \eta$ si $x \rightarrow b$. Donc il existe $C \geq B$ tel que $x \geq C$ implique

$$1 - 2\eta \leq \frac{\int_a^x |f|}{\int_a^x |g|} \leq 1 + 2\eta.$$

On a bien l'équivalence des infiniment grands. □

Dans le cas où $I =]a, b]$ et $f(t) \underset{t \rightarrow a}{\sim} g(t)$, les mêmes résultats sont vrais en considérant $\int_a^x |f(t)| dt$ et $\int_a^x |g(t)| dt$ si les intégrales convergent, et $\int_x^b |f(t)| dt$ et $\int_x^b |g(t)| dt$ si elles divergent.

En application des critères précédents, citons :

Critère 4 (Critère de Riemann) Si f est continue par morceaux dans $[a, +\infty[$ et si

$$\lim_{t \rightarrow +\infty} t^\alpha f(t) \lambda > 0,$$

pour un $\alpha \in \mathbb{R}$, alors l'intégrale $\int_a^{+\infty} f(t) dt$ converge si $\alpha > 1$ et diverge si $\alpha \leq 1$.

Preuve. On applique le critère 3 en remarquant que f est du signe de λ dans un intervalle $[c, +\infty[$ avec $c \geq a$ et $c > 0$. Donc les intégrales $\int_a^{+\infty} f(t) dt$, $\int_c^{+\infty} |f(t)| dt$ et $\int_c^{+\infty} \frac{dt}{t^\alpha}$ sont de même nature. □

Dans le cas où f n'est pas comparable avec une puissance de t , on peut utiliser également la règle de Riemann :

Critère 5 Soit f localement intégrable dans $[a, +\infty[$.

1. Si pour un réel α , $\alpha > 1$, $\lim_{t \rightarrow +\infty} t^\alpha f(t) = l \in \mathbb{R}$, alors l'intégrale $\int_a^{+\infty} |f(t)| dt$ converge.
2. Si pour un réel α , $\alpha' \leq 1$, $\lim_{t \rightarrow +\infty} t^{\alpha'} f(t) = L$ et $|L| \in]0, +\infty]$, alors l'intégrale $\int_a^{+\infty} f(t) dt$ diverge.

Remarquez que dans le premier cas, si $l \neq 0$, c'est le critère 4.

Exemple 6.14 Nature de $\int_0^\infty f(t) dt$ pour $f(t) = \frac{t^2 + 3}{t^4 + 1}$:

Ici f est rationnelle définie sur \mathbb{R} , donc continue sur \mathbb{R}_+ . Par ailleurs,

$$\lim_{t \rightarrow +\infty} t^2 f(t) = 1.$$

Donc l'intégrale $\int_0^{\infty} f(t) dt$ converge.

Exemple 6.15 Nature de $\int_1^{\infty} f(t) dt$ pour $f(t) = \frac{2 - \frac{1}{t}}{t}$:

La fonction f est continue et positive sur $[1, +\infty[$. De plus

$$\forall t \geq 1, f(t) \geq \frac{1}{t} > 0,$$

et l'intégrale $\int_1^{+\infty} \frac{1}{t} dt$ diverge, donc d'après le critère 1, $\int_1^{\infty} f(t) dt$ diverge.

Dans le cas où l'intervalle d'intégration I est semi-ouvert et **borné**, d'extrémités a et b , $-\infty < a < b < +\infty$, on a les règles suivantes.

Critère 6 — Si f est continue par morceaux sur $[a, b[$ et si

$$\lim_{t \rightarrow b} (b - t)^\alpha f(t) = \lambda > 0,$$

pour un $\alpha \in \mathbb{R}$, alors l'intégrale $\int_a^b f(t) dt$ converge si $\alpha < 1$ et $\int_a^b f(t) dt$ diverge si $\alpha \geq 1$.

— De même, si f est continue par morceaux sur $]a, b]$ et si

$$\lim_{t \rightarrow b} (t - a)^\alpha f(t) = \lambda > 0,$$

pour un $\alpha \in \mathbb{R}$, alors l'intégrale $\int_a^b f(t) dt$ converge si $\alpha < 1$ et $\int_a^b f(t) dt$ diverge si $\alpha \geq 1$.

Preuve. Appliquez le critère 3 avec les exemples cités dans la section 6.4.2. □

Critère 7 On suppose que f est localement intégrable sur $[a, b[$.

— Si $\lim_{t \rightarrow b} (b - t)^\alpha f(t) = l$ pour $l \in \mathbb{R}$ et $\alpha < 1$, alors l'intégrale $\int_a^b |f(t)| dt$ converge.

— Si $\lim_{t \rightarrow b} (b - t)^\alpha f(t) = L$ avec $|L| \in]0, +\infty]$ et $\alpha \geq 1$, alors l'intégrale $\int_a^b f(t) dt$ diverge.

On peut utiliser un critère équivalent si $I =]a, b]$.

6.4.4 Fonctions à valeurs quelconques

Définition 6.15 Soit f une fonction continue par morceaux sur un intervalle I . On dit que l'intégrale de f sur I est absolument convergente si l'intégrale de la fonction $|f|$ sur I converge.

Théorème 6.17 Si l'intégrale de f sur I est absolument convergente, alors l'intégrale de f sur I converge.

Remarque. Si f est de signe constant sur l'intervalle I , alors il n'y a pas de distinction entre l'absolue convergence et la convergence de l'intégrale.

Attention : le théorème 6.17 n'est pas une équivalence. L'intégrale de $|f|$ sur I peut diverger et l'intégrale de f sur I converger.

6.4.5 Méthodologie

On souhaite étudier la nature d'une intégrale de la forme $\int_I f(t)dt$ où I est un intervalle de \mathbb{R} et f une fonction à valeurs réelles. L'étude se fait en au moins quatre étapes :

1. D'abord on prend soin de justifier que f est continue par morceaux sur I et de repérer les points de I qui posent problème. En particulier si I est un intervalle borné, on pensera à identifier les points où f pourrait ne pas être définie.
2. On étudie le signe de f sur l'intervalle I (un tableau de variations peut être utile parfois). Si f n'est pas toujours de signe positif, on considère $|f|$.
3. On découpe l'intervalle I en autant de sous-intervalles que nécessaire.
4. On étudie la nature de l'intégrale de $|f|$ sur chaque sous-intervalle : il faut avoir à l'esprit les critères de convergence ou de divergence du cours.
5. On rassemble les résultats et on conclut.

Si vous n'arrivez pas à conclure, il y aura des questions supplémentaires pour vous guider vers la solution.

Exemple 6.16 Étudier l'intégrabilité sur $I = \mathbb{R}_+^*$ de la fonction

$$f(t) = \begin{cases} \frac{1 - e^t}{t^2} & \text{si } x \in]0, 1], \\ \frac{1}{t^2} & \text{si } x \in]1, +\infty[. \end{cases}$$

1. f est continue par morceaux sur \mathbb{R}_+^* . Il y a un problème éventuel en 0 et $+\infty$.
2. f est positive sur $]1, +\infty[$, négative sur $]0, 1[$.
3. Un découpage possible est $I =]0, 1/2] \cup]1/2, 2] \cup [2, +\infty[$.
4. Sur $]1/2, 2]$, la fonction f est continue par morceaux et bornée, donc intégrable (chapitre précédent). Sur $[2, +\infty[$, $|f(t)| = f(t) = \frac{1}{t^2}$: donc convergence absolue

(intégrale de Riemann). Sur $]0, 1/2]$, $|f(t)| = \frac{e^t - 1}{t^2}$. Comme $t \mapsto e^t$ est dérivable en zéro,

$$\lim_{t \rightarrow 0} \frac{e^t - e^0}{t - 0} = 1 \implies \lim_{t \rightarrow 0} t \frac{e^t - 1}{t^2} = 1 > 0.$$

Donc divergence de l'intégrale de $|f|$ sur $]0, 1/2]$. Comme sur cet intervalle $|f(t)| = -f(t)$, l'intégrale de f sur $]0, 1/2]$ diverge.

5. L'intégrale de f sur I diverge à cause de la divergence sur un des morceaux.