

Éléments de cours de Probabilités

Licence de mathématiques
Université de Versailles Saint-Quentin

Jean-François Marckert

Table des matières

I. Introduction	1
1. Un peu d'histoire	1
2. Modélisation et problèmes métaphysiques	1
II. Espace de Probabilité fini et dénombrable	3
1. Espace et mesure de probabilité	3
2. Équiprobabilité	4
<i>a.</i> Combinaison et arangements	4
3. Fonctions génératrices de dénombrement	5
<i>a.</i> Méthode générale	5
<i>b.</i> Un exemple	7
4. Indépendance d'événements et probabilité conditionnelle	8
5. Espaces discrets	10
<i>a.</i> Variables aléatoires discrètes	10
<i>b.</i> Lois images	11
<i>c.</i> Espérance, variance et moments	12
<i>d.</i> Espérance d'une fonction d'une variable aléatoire	13
<i>e.</i> Inégalité de Markov et de Bienaymé-Tchebichev	14
6. Exemples de lois discrètes	15
7. Famille de variables aléatoires	16
8. Fonctions génératrices de probabilité	18
9. Variance et covariance	19
III. Espaces de probabilités généraux	20
1. axiomatique de Kolmogorov :	20
2. Variables aléatoires, espérance	20
3. Lois de probabilité	21
4. Variables aléatoires admettant une densité	22

5.	Exemples de lois admettant une densité	23
6.	Fonction de répartition	24
	<i>a.</i> Définition et premières propriétés	24
	<i>b.</i> Fonction de répartition et densité	25
7.	Calcul de la loi d'une variable aléatoire	25
	<i>a.</i> Grâce à la fonction de répartition	25
	<i>b.</i> Grâce au calcul de "l'espérance"	26
8.	Lois jointes	28
9.	Indépendances	30
10.	Changement de variables	33
11.	Simulations de variables aléatoires	34
12.	La fonction caractéristique	35
	<i>a.</i> Convolée	39
	<i>b.</i> Transformée de Laplace	40
IV. Théorèmes limites en probabilité		41
1.	Lemmes de Borel-Cantelli	41
2.	Convergence en probabilité	42
	<i>a.</i> Loi faible des grands nombres	43
3.	Convergence presque sûre	43
4.	Convergence dans L^p	45
5.	Convergence en loi. Caractérisations	45
6.	Critères généraux de convergence en loi	47
7.	Théorème de la limite centrale	50
8.	Hierarchie des convergences	52
V. ANNEXES : Rappels d'intégration et de théorie des ensembles		54
1.	Tribus et mesures	54
2.	Intégration	54
3.	Quelques éléments sur les mesures sur \mathbb{R}	55
	<i>a.</i> Décompositions des mesures réelles	55
	<i>b.</i> Application aux mesures de probabilité	55

I. Introduction

Dans cette introduction nous allons commencer par nous poser les questions suivantes, naïves au premier abord :

- Que sont les probabilités ?

- Peut-on concrètement utiliser les résultats probabilistes dans la vie de tous les jours ?

Si oui, quelle confiance peut-on accorder aux résultats obtenus ?

Bien entendu, nous ne répondons pas vraiment à ces questions.

1. Un peu d'histoire

On date souvent le début du calcul des probabilités aux premiers travaux de Pascal et Fermat (qui correspondaient à ce sujet durant l'année 1654). Il va sans dire que les questions ayant trait au hasard (ou à certaines "tentatives de le mesurer") sont nettement antérieures à cette époque ; les philosophes grecs s'occupaient de toutes les questions imaginables et donc bien sûr des problèmes de hasard et de déterminisme.

Au début du 18^{ème} siècle, la famille Bernoulli d'une part et de Moivre d'autre part réalisent de nombreuses avancées. En particulier, on doit à la première, la loi des grands nombres et au second le théorème de la limite centrale (pour une somme de variables de Bernoulli) retrouvée par Laplace 80 ans plus tard. Puis, de nombreux scientifiques apportent des pierres à l'édifice : Bayes, Leibniz...

Il fallut attendre le début du 20^{ème} siècle pour lancer des fondements mathématiques solides à la théorie des probabilités avec Poincaré, Borel et surtout Kolmogorov. Aujourd'hui, la recherche en probabilité est très active et de nombreux résultats sont publiés chaque année. L'utilisation des outils probabilistes et statistiques dans toutes les sciences témoignent de l'efficacité et de l'importance de ces disciplines.

2. Modélisation et problèmes métaphysiques

Les problèmes posés à la construction de la théorie des probabilités sont grossièrement de deux ordres :

- on veut se servir des calculs que l'on fait en probabilité *dans la vraie vie*.
- on ne peut pas faire de mathématiques sans une axiomatique claire et "rigoureuse". En particulier, on ne peut pas (décentement) construire des mathématiques en utilisant une notion de *hasard* alors que l'existence et une définition de celui-ci sont sujets à discussion (et donnent du travail aux philosophes depuis toujours).

Ainsi, par exemple, on pourrait avoir envie de définir la probabilité d'un événement (penser à la probabilité que la face supérieure d'un dé donné tombe sur 6), comme la proportion asymptotique de résultats favorables par rapport au nombre d'expériences effectuées (la proportion asymptotique de 6 obtenue). Il est clair que ceci pose de nombreux problèmes : par exemple, comment prouver que cette proportion converge ? Et même si elle converge, comment vérifier que la proportion limite ne change pas si on refait une suite d'expériences ? Ou encore, comment lancer un dé un nombre infini de fois ?!

Le tour de force de Kolmogorov a été d'utiliser les travaux tout jeunes de la théorie de la mesure afin de construire une axiomatique cohérente et puissante. La probabilité des événements est perçue *a priori* : les résultats asymptotiques (dits *fréquentistes* comme la loi des grands nombres par exemple) deviennent *des théorèmes* et ne sont donc pas *des définitions* comme on pourrait s'y attendre.

Ainsi, les probabilités font parties des mathématiques. Ses théorèmes (résultats) sont inférés logiquement sur l'axiomatique de Kolmogorov et ainsi, les théorèmes établis en probabilité possèdent le même statut que tous les autres résultats mathématiques. On insiste au passage sur le fait que la théorie des probabilités ne suppose en aucune façon l'existence du hasard, ni encore moins quelque propriété qu'il pourrait avoir. Ceci n'est pas du tout un détail. Dans la théorie des probabilités on définit la probabilité d'un événement élémentaire sans parler du tout d'une suite d'expériences.

Maintenant, parlons un peu des applications des probabilités. Les probabilités et les statistiques sont utilisées dans toutes les sciences. Comment alors déterminer la probabilité d'un événement réel? Comme on l'a déjà plus ou moins laissé entendre, la question est mal posée (c'est-à-dire qu'il n'existe pas de bonnes réponses à cette question).

Ainsi, en pratique, on cherche un modèle probabiliste en accord (le plus possible) avec le phénomène observé. Par exemple, si une pièce ne nous semble pas suspecte, on dira que la probabilité qu'elle tombe sur pile est $1/2$. Si 100 lancers donnent 75 piles, on sera poussé à proposer un autre modèle.

De nombreux problèmes interviennent lors de la modélisation et sans trop rentrer dans les détails, il est important de se rendre compte que ce point est le seul qui est vraiment toujours discutable : une fois la modélisation faite, il s'agit de mathématiques et donc plus rien n'est discutable.

Considérons ce dernier exemple : on veut modéliser le jeu du loto ; on dit qu'un tirage est un sous ensemble à 7 éléments de $\Omega = \{1, \dots, 49\}$. Pour des raisons de symétries on donne à toute partie A à 7 éléments dans Ω , $\mathbb{P}(A) = 1/C_{49}^7$. Souvent, en probabilité, on suppose que les tirages sont indépendants de semaines en semaines. Il faut se rendre compte que l'on ne pourra jamais prouver ceci (à cent pour cent de certitude).

Pour en finir avec ce petit paragraphe qui peut donner une idée pessimiste de ce que l'on peut faire avec les probabilités il est bon de réfléchir au "échanges" entre la partie mathématique et la partie application des probabilités. Bien entendu, le but des probabilités est aussi de modéliser des problèmes ayant traits au hasard. Le "hasard" nourrit d'une part les notions probabilistes (c'est-à-dire, notre conception du hasard nous pousse à définir certaines notions, par exemple les notions d'indépendance ou de probabilité conditionnelle sont exactement les notions intuitives) d'autre part, en retour, le calcul des probabilités permet de "deviner" le comportement de certains phénomènes aléatoires.

L'efficacité des probabilités dans les applications valide que les modélisations choisies sont tout à fait convenables et constituent, si ce n'est une image parfaite de la *réalité*, une bonne approximation de celle-ci.

II. Espace de Probabilité fini et dénombrable

1. Espace et mesure de probabilité

Définition : On appelle espace de probabilité fini un ensemble quelconque de cardinal fini (on l'appelle également espace d'état, ou univers).

On le note habituellement Ω . Dans la pratique, on imagine que les éléments de Ω sont les résultats possibles d'une expérience aléatoire. On appelle **événement** toute partie de Ω ; les atomes de Ω sont appelés événements élémentaires.

Exemples :

- a) Pour le jeu de pile ou face on prendra $\Omega = \{p, f\}$ ou $\Omega = \{0, 1\}$.
- b) Pour une suite de 10 lancers d'une pièce, on prendra $\Omega = \{p, f\}^{10}$, l'ensemble des 10 uplets composés de p et de f .
- c) Nombre de lancers d'une pièce avant qu'elle ne tombe sur pile $\Omega = \mathbb{N}^*$ (on peut ajouter $+\infty$ à cet ensemble si on le souhaite).
- d) Durée de la prochaine communication téléphonique à la cabine du coin de la rue $\Omega = \mathbb{R}^+$.

Les ensembles (c) et (d) ne sont pas de cardinaux finis. On traitera ces ensembles ultérieurement.

On pourra remarquer que Ω peut très bien ne pas être un ensemble de nombres. Une question dérangeante se pose : dans l'exemple (a) on pourrait avoir envie d'ajouter $\{t\}$ à Ω pour la tranche de la pièce ou ce genre de chose. En fait, on peut mettre à peu près n'importe quoi dans Ω , ce n'est pas bien important. Ce qui est important c'est la mesure de probabilité que l'on met sur lui (si on met une probabilité telle que $\mathbb{P}(\{t\}) = 0$, tout se passe comme si $\{t\}$ n'appartenait pas à Ω).

Définition : Une *mesure de probabilité* sur Ω est une mesure (positive) sur l'espace mesurable $(\Omega, \mathcal{P}(\Omega))$ de masse totale 1, où $\mathcal{P}(\Omega)$ désigne l'ensemble des parties de Ω ($\mathcal{P}(\Omega)$ est une tribu).

Notons \mathbb{P} une telle mesure. \mathbb{P} est définie de $\mathcal{P}(\Omega)$ dans $[0, 1]$ et on doit avoir, pour tout A et B éléments de $\mathcal{P}(\Omega)$,

$$\mathbb{P}(\Omega) = 1, \quad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \text{ si } A \text{ et } B \text{ sont disjoints.} \quad (1)$$

Remarquons que l'additivité ici suffit car le cardinal de Ω est supposé fini (si Ω n'est pas fini, \mathbb{P} doit être σ -additif).

Exercice 1 : Montrer que si \mathbb{P} est une probabilité sur Ω (de cardinal fini), alors pour tout A et B dans $\mathcal{P}(\Omega)$, on a

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \\ \mathbb{P}(\emptyset) &= 0, \\ \mathbb{P}(A^c) &= 1 - \mathbb{P}(A), \\ \mathbb{P}(A) &\leq \mathbb{P}(A \cap B) + \mathbb{P}(B^c) \end{aligned}$$

Exemples de mesures de probabilité (correspondants aux exemples vus plus haut) :

- a) $\mathbb{P}(\{p\}) = \mathbb{P}(\{f\}) = 1/2$. Il s'agit de la mesure $\mathbb{P} = \frac{1}{2}(\delta_p + \delta_f)$ (où δ est la mesure de Dirac).
- b) On peut munir cet ensemble de la probabilité uniforme, c'est-à-dire que chacun des 2^{10} événements élémentaires de Ω ont même probabilité 2^{-10} (il s'agit de la mesure $\mathbb{P} = 2^{-10} \sum_{\omega \in \Omega} \delta_\omega$).

Remarque 1 (*A propos des mesures de probabilité discrète et des masses de Dirac*).

On rappelle tout d'abord que la mesure de Dirac δ_a est définie par : pour tout ensemble A on a :

$$\delta_a(A) = \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon} \end{cases}$$

Considérons la mesure \mathbb{P} qui donne la masse $1/6$ à chacun des éléments de $\Omega = \{1, 2, 3, 4, 5, 6\}$. Une petite vérification montre que $\mathbb{P} = \sum_{i=1}^6 \frac{1}{6} \delta_i$ (par exemple $(\sum_{i=1}^6 \frac{1}{6} \delta_i)(\{1, 3\}) = \sum_{i=1}^6 \frac{1}{6} \delta_i(\{1, 3\}) = 2/6$). Cette expression des probabilités à l'aide de la théorie de la mesure (passée sous silence dans les cours antérieurs) est fondamentale. Elle permet de traiter dans un même cadre les probabilités discrètes et continues (comme on le verra par la suite). Dans un deuxième temps, on verra que la plupart des résultats de la théorie des probabilités sont en quelques sortes des résultats de la théorie de l'intégration.

2. Équiprobabilité

On est toujours dans le cas où $\text{card}(\Omega) < +\infty$.

Proposition 1 *L'application*

$$\begin{aligned} \mathbb{P} : \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} \end{aligned}$$

est une probabilité sur Ω (preuve laissée en exercice) appelée équiprobabilité sur Ω .

Connaître la probabilité d'un événement quelconque se réduit donc à calculer le cardinal de A .

Preuve : Il s'agit de la mesure $\mathbb{P} = \frac{1}{\text{card}(\Omega)} \sum_{\omega \in \Omega} \delta_\omega$. \square

a. Combinaison et arrangements

Un k -uplet d'éléments de A est une liste ordonnée de k éléments de A (c'est-à-dire, un élément de A^k)

Exemple : couple = 2-uplets. Le point de coordonnée (1,2) est différent de celui de coordonnée (2,1).

Arrangements

Soit $B = \{b_1, \dots, b_n\}$. On note $U_k = \{k\text{-uplets composés d'éléments différents de } B\}$ (plus formellement,

$$U_k = \{x = (x_1, \dots, x_k) \in A^k \text{ t.q. } i \neq j \implies x_i \neq x_j\}.$$

U_k est appelé ensemble des arrangements de k éléments de B . Puisque $\#B = n$, le nombre de tels arrangements est donné par :

$$\#U_k = A_n^k = \frac{n!}{(n-k)!} \quad \text{pour } n \geq 0, \quad 0 \leq k \leq n. \quad (2)$$

$$A_n^n = n!, A_n^0 = 1$$

Preuve de (2) : Un k -uplet est la donnée de k éléments ordonnés. On peut choisir le premier de n façons, le deuxième de $n - 1$ façons, ..., le k ème de $(n - k + 1)$ façons. \square

Exemple : nombre de tiercés dans l'ordre avec 20 chevaux. L'ordre d'arrivée (4,7,12) est différent de l'ordre (7,12,4). Il y a donc $A_{20}^3 = 20 \times 19 \times 18 = 6840$ ordres possibles.

Combinaisons

On note $SE_k = \{\text{sous-ensembles de } B \text{ à } k \text{ éléments}\}$

$$SE_k = \{\{x_1, \dots, x_k\}, (\forall l, x_l \in B), i \neq j \implies x_i \neq x_j\}$$

Alors

$$\#SE_k = C_n^k = \frac{A_n^k}{k!} = \frac{n!}{(n-k)!k!} \quad \text{pour } n \geq 0, \quad 0 \leq k \leq n \quad (3)$$

Preuve de (3) : Le nombre de k -uplets est A_n^k . Pour chaque sous-ensemble de k éléments de B on peut fabriquer $k!$ k -uplets différents. Ainsi $\#SE_k = A_n^k/k!$. \square

Exemple : Nombre de grilles différentes au loto : on choisit 6 numéros parmi 49. Le choix est un ensemble de 6 éléments parmi $\{1, \dots, 49\}$; c'est-à-dire choisir $\{1, 5, 10, 15, 16, 42\}$ ou $\{1, 5, 10, 15, 42, 16\}$ c'est cocher les mêmes cases et c'est donc la même chose. Ainsi le nombre de choix est $C_{49}^6 = 49!/(43!6!) = 13983816$.

Propriété 1 Pour tout $n \geq 0$

$$\begin{aligned} C_n^p &= C_n^{n-p} && \text{pour tout } 0 \leq p \leq n \\ C_n^p &= C_{n-1}^{p-1} + C_{n-1}^p && \text{pour tout } 1 \leq p \leq n-1 \\ A_n^p &= nA_{n-1}^{p-1} && \text{pour tout } 1 \leq p \leq n-1 \end{aligned}$$

Pour tout $n \geq 0$

$$2^n = \sum_{k=0}^n C_n^k \quad \text{et} \quad \sum_{k=0}^n C_n^k (-1)^k = 0$$

Preuve laissée en exercice.

3. Fonctions génératrices de dénombrement

On introduit une méthode très utilisée (car très puissante) par les combinatoristes pour dénombrer des objets complexes. Certains lecteurs trouveront peut-être plus profitable de regarder d'abord la partie (b) qui suit, afin d'avoir un exemple.

a. Méthode générale

Ce qui suit sort du livre de P. Flajolet et R. Sedgewick que l'on peut télécharger depuis le site ouèbe de Philippe Flajolet.

On appelle classe de structures combinatoires ou plus simplement *classe* un ensemble fini ou dénombrable sur lequel une fonction *taille* est définie ; la taille d'un élément est un nombre entier positif (ou nul). Soit \mathcal{A} une classe et a un élément de \mathcal{A} . On note $|a|$ la taille de a . On note \mathcal{A}_n la partie de \mathcal{A} composée des objets de taille n . On note $a_n = \text{card}(\mathcal{A}_n)$. On suppose que tous les a_n

sont finis. La suite $(a_n)_n$ est appelée suite de dénombrement (ou de comptage) de \mathcal{A} . La fonction génératrice ordinaire (FGO) ou de dénombrement de la classe \mathcal{A} est la série entière

$$A(z) = \sum_{n \geq 0} a_n z^n.$$

(Il ne faudra pas confondre cette FGO avec la fonction génératrice de probabilité introduite un peu plus loin). Notez bien que le fait que cette série converge pour certains z est secondaire. Si elle converge, on pourra s'en servir comme d'une fonction, sinon, on pourra voir $A(z)$ comme une série formelle, c'est-à-dire un vecteur de longueur infinie sur lequel on pourra quand même faire certaines opérations (additions, multiplications, dérivations...).

$A(z)$ possède une seconde écriture qu'il est bon d'avoir à l'esprit pour des raisons qui apparaîtront plus tard :

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|};$$

on voit bien que le coefficient de z^n dans cette dernière somme est justement a_n .

Définition : Si Φ est une construction binaire qui associe à deux classes \mathcal{B} et \mathcal{C} une nouvelle classe

$$\mathcal{A} = \Phi(\mathcal{B}, \mathcal{C})$$

de sorte que chaque a_n ne dépende que d'un nombre fini de b_k et de c_j , on dit que Φ est admissible.

Dans ce cas, il existe un opérateur bien défini Ψ tel que

$$A(z) = \Psi(B(z), C(z)).$$

Le produit cartésien : Supposons par exemple que \mathcal{A} soit le produit cartésien $\mathcal{B} \times \mathcal{C}$, c'est à dire que \mathcal{A} est l'ensemble des couples de type (β, γ) avec $\beta \in \mathcal{B}$ et $\gamma \in \mathcal{C}$; supposons encore que la taille de l'élément $\alpha = (\beta, \gamma)$ soit $|\alpha|_{\mathcal{A}} = |\beta|_{\mathcal{B}} + |\gamma|_{\mathcal{C}}$, alors on a

$$a_n = \sum_{k=0}^n b_k c_{n-k}$$

ce qui se traduit d'une part par le fait que le produit cartésien est admissible et d'autre par sur les FGO par

$$A(z) = B(z)C(z).$$

Union de deux ensembles : Soient \mathcal{B} et \mathcal{C} deux classes disjointes munies de la même fonction taille; $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$ est admissible et on a

$$A(z) = B(z) + C(z).$$

Ensemble des listes finies : Soit \mathcal{C} une classe; on note $\mathfrak{S}(\mathcal{C})$ l'ensemble des listes finies (ordonnées) d'éléments (non nécessairement différents) de \mathcal{C} :

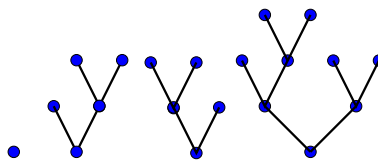
$$\mathfrak{S}(\mathcal{C}) = \{\varepsilon\} \cup \mathcal{C} \cup (\mathcal{C} \times \mathcal{C}) \cup (\mathcal{C} \times \mathcal{C} \times \mathcal{C}) \cup \dots$$

Le symbole ε désigne un élément de taille 0 dont on se sert parfois pour décrire "un élément nul" ou "la liste vide". Ici on autorise la liste vide. La FGO de $\mathcal{A} = \mathfrak{S}(\mathcal{C})$ s'exprime en fonction de celle de \mathcal{C} :

$$A(z) = 1 + C(z) + C^2(z) + C^3(z) + \dots = \frac{1}{1 - C(z)}.$$

b. Un exemple

On appelle arbre binaire un arbre ordonné dans lequel tous les noeuds ont 0 ou 2 fils. Sur le dessin ci-contre, les quatre arbres binaires sont différents. On définit la taille d'un arbre binaire t comme étant simplement son nombre de noeuds (on la note $|t|$). Sur le dessin ci-contre, les quatre arbres sont différents, le premier est de taille 1, les deux suivants de taille 5, et le dernier de taille 9. On se pose la question suivante : combien y a-t-il d'arbres binaires de taille n ?



On note donc \mathcal{C} la classe des arbres binaires, \mathcal{C}_n la classe des arbres binaires ayant n noeuds, et $c_n = \text{card } \mathcal{C}_n$. Nous cherchons à calculer c_n (pour tout n), et pour cela, nous allons décomposer les arbres binaires. Prenons un arbre t ; nous voyons que deux cas se présentent : soit t est réduit à un noeud, soit t est un noeud qui possède deux sous arbres qui sont eux même des arbres binaires. Il vient

$$c_0 = 0, c_1 = 1 \text{ et pour } j \geq 2, \quad c_j = \sum_{i=1}^{j-1} c_i c_{j-1-i}.$$

La raison de la dernière égalité est la suivante : un arbre de taille j peut être représenté par le couple des deux sous arbres de t . La taille du sous arbre de gauche étant i , il y a c_i arbres gauches possibles et alors, le sous arbre droit peut être de c_{j-1-i} manières. La classe \mathcal{C} possède une décomposition récursive :

$$\mathcal{C} = u + u \times \mathcal{C} \times \mathcal{C} \tag{4}$$

où u représente un arbre de taille 1 (un noeud isolé). On peut voir maintenant que

$$C(z) = z + zC^2(z); \tag{5}$$

pour voir cela, il faut considérer ceci : $c_j = \sum_{i=1}^{j-1} c_i c_{j-1-i} = \sum_{i=0}^{j-1} c_i c_{j-1-i}$. Cette dernière somme est le coefficient de z^{j-1} dans la série $C(z)^2$ (le produit de Cauchy des coefficients donne le coefficient du produit), donc de z^j dans $zC(z)^2$. Il reste à extraire les coefficients de $C(z)$ solution de (5). On a

$$zC^2(z) - C(z) + z = 0.$$

C est solution d'une équation de degré 2; on trouve

$$C(z) = \frac{1 \pm \sqrt{1 - 4z^2}}{2z}.$$

Le développement en série entière de la solution que l'on cherche ne possède que des coefficients positifs. On a donc C croissante partout et donc en 0^+ et donc

$$C(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z}.$$

Il reste à extraire les coefficients. Pour cela on utilise la formule de Taylor (on dérive...). Tout calcul fait, on obtient $c_n = 0$ pour n pair et $c_{2n+1} = C_{2n}^n / (n+1)$ pour n impair. $C_{2n}^n / (n+1)$ est communément appelé, le n ème nombre de Catalan.

Remarque 2 Les deux formules (4) et (5) doivent paraître étonnantes (au lecteur non habitué... des beautés de la combinatoire). L'équation (5) qui traduit sous forme de série génératrice la décomposition combinatoire (4) de la structure des arbres binaires lui ressemble à s'y méprendre. Ce n'est pas dû à l'exemple traité et cette ressemblance est tout à fait générique (et les combinatoristes passent de la première à la deuxième en quelques secondes).

Exercice 2 : (difficile) :

1) On appelle composition de n , une liste finie ordonnée $(x_i)_{1 \leq i \leq k}$ où k est un entier, les x_i sont des entiers positifs non nuls de somme n . Deux compositions $(x_i)_i$ et $(x'_i)_i$ sont dites différentes si il existe un indice j tel que $x_j \neq x'_j$. Combien y a-t-il de compositions de n ? (on utilisera un raisonnement par série génératrice).

2) Soit n un entier non nul. On appelle partition de n , une suite finie (x_1, \dots, x_k) telle que $i < j$ entraîne $x_i \leq x_j$ et telle que tous les x_i sont strictement positifs et telle que $\sum x_i = n$. On note c_n le nombre de partitions de n . Prouver que la série génératrice des c_n est

$$C(x) = \prod_{k \geq 1} \frac{1}{1 - x^k}.$$

4. Indépendance d'événements et probabilité conditionnelle

Les probabilités conditionnelles ont pour but d'évaluer "le changement de probabilité" dû à l'acquisition d'informations. Par exemple, si l'on dispose d'un dé juste, la probabilité d'obtenir un 1 est $1/6$. Si quelqu'un lance le dé pour nous et nous donne l'information suivante : "le résultat est impair". On peut écarter les événements $\{2, 4, 6\}$ et en déduire que maintenant, le résultat est 1 avec probabilité $1/3$. Formalisons tout cela...

Définition : Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé et $B \in \mathcal{P}(\Omega)$ tel que $\mathbb{P}(B) > 0$; soit A un élément de $\mathcal{P}(\Omega)$. La probabilité conditionnelle de A sachant B est définie par :

$$\mathbb{P}(A | B) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (6)$$

On a donc :

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

Le théorème suivant est important : il explique comment et pourquoi la "probabilité conditionnelle à l'événement B " est bien une probabilité qui vérifiera donc toutes les propriétés propres aux probabilités établies plus haut.

Proposition 2 Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé et $B \in \mathcal{P}(\Omega)$ tel que $\mathbb{P}(B) > 0$. L'application

$$\begin{aligned} \mathbb{P}_B : \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{P}_B(A) \end{aligned}$$

est une probabilité sur Ω .

Preuve : On a immédiatement $\mathbb{P}_B(\emptyset) = 0, \mathbb{P}_B(\Omega) = 1$. Soit C et D de $\mathcal{P}(\Omega)$ tels que $C \cap D = \emptyset$.

$$\begin{aligned} \mathbb{P}_B(C \cup D) &= \frac{\mathbb{P}((C \cup D) \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}((C \cap B) \cup (D \cap B))}{\mathbb{P}(B)} \end{aligned}$$

ces 2 ensembles (au numérateur) sont disjoints, donc

$$\begin{aligned}\mathbb{P}_B(C \cup D) &= \frac{\mathbb{P}(C \cap B)}{\mathbb{P}(B)} + \frac{\mathbb{P}(D \cap B)}{\mathbb{P}(B)} \\ &= \mathbb{P}_B(C) + \mathbb{P}_B(D) \quad \square\end{aligned}$$

La σ -additivité (que l'on doit montrer sur les espaces de probabilité non finis) se montre de la même manière.

La formule des probabilités conditionnelles correspond tout à fait au changement de probabilité intuitif. Le référent n'est plus Ω mais B . Ainsi (6) traduit le fait que les cas possibles sont dans B ; la probabilité de B sachant B vaut donc 1.

Formule des probabilités totales. Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé et A_1, \dots, A_k une partition de Ω , c'est-à-dire :

$$\bigcup_{i=1}^k A_i = \Omega \text{ et } \forall i, j, \quad i \neq j \implies A_i \cap A_j = \emptyset.$$

On a, pour tout $A \in \mathcal{P}(\Omega)$

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap (\bigcup_{i=1}^k A_i)) = \sum_{i=1}^k \mathbb{P}(A \cap A_i)$$

Si pour tout i , $\mathbb{P}(A_i) > 0$, on a de plus :

$$\mathbb{P}(A) = \sum_{i=1}^k \mathbb{P}(A | A_i) \mathbb{P}(A_i)$$

Formule de Bayes Sous les mêmes hypothèses. si on a de plus $\mathbb{P}(A) > 0$, alors :

$$\mathbb{P}(A_j | A) = \frac{\mathbb{P}(A_j \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A | A_j) \mathbb{P}(A_j)}{\sum_{i=1}^k \mathbb{P}(A | A_i) \mathbb{P}(A_i)}$$

La formule de Bayes est utilisée pour "retourner" des probabilités conditionnelles.

Définition : Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé. On dit que deux événements A et B sont indépendants si

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

On note $A \perp\!\!\!\perp B$.

L'indépendance est une notion primordiale en probabilité comme on va le voir par la suite. Si A et B sont indépendants, par (6) on a :

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

(Ceci bien sûr si $\mathbb{P}(B) \neq 0$.) L'interprétation est la suivante : si A et B sont indépendants, savoir que B est réalisé ne change pas la probabilité de la réalisation de A .

“Dans la vraie vie”, l’indépendance d’événements est souvent une question cruciale ; en médecine par exemple, à cause de l’effet placebo, on se pose souvent la question de savoir si la guérison est indépendante ou non de la prise d’un médicament. Plus ou moins honnêtement, on peut affirmer qu’une personne superstitieuse se distingue d’une non-superstitieuse en ce qu’elle perçoit de la dépendance entre deux événements, là où la seconde n’en voit pas.

Exercice 3 : a) Montrer que Ω est indépendant de tout événement. Montrer la même chose pour \emptyset .
 b) Montrer que si $A \perp\!\!\!\perp B$ alors $A^c \perp\!\!\!\perp B$, $A \perp\!\!\!\perp B^c$ et $A^c \perp\!\!\!\perp B^c$.

5. Espaces discrets

On suppose ici que le cardinal de Ω est au plus dénombrable. Deux cas sont possibles :

- $\text{card}(\Omega) < +\infty$ ce qui nous ramène au cas fini.
- $\text{card}(\Omega) = +\infty$; dans ce cas, il existe une bijection de Ω dans \mathbb{N} . Ainsi, les éléments de Ω sont numérotés par les entiers.

Une probabilité sur Ω est une mesure de masse totale 1 (sur la tribu $\mathcal{P}(\Omega)$) ; il est important maintenant de rappeler que \mathbb{P} (en plus des conditions (1) page 3) est, comme toute mesure, σ -additive, c’est-à-dire, pour toute suite $(A_n)_n$ d’éléments de $\mathcal{P}(\Omega)$ disjoints deux à deux,

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n).$$

La σ -additivité assure que \mathbb{P} est bien connue si on connaît $\mathbb{P}(\{\omega\})$ pour tout $\omega \in \Omega$ (on appelle les éléments ω de probabilité non nulle, les atomes de \mathbb{P}). De plus, on peut voir que,

$$\forall A \in \mathcal{P}(\Omega), \quad \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega).$$

En fait, la mesure \mathbb{P} est encore une somme de masses de Dirac pondérées que l’on peut exprimer comme suit :

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

a. Variables aléatoires discrètes

Les variables aléatoires sont des objets centraux en théorie des probabilités. Elles jouent le même rôle que les fonctions en analyse.

Définition : Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace de probabilité. On appelle variable aléatoire réelle toute application mesurable X de $(\Omega, \mathcal{P}(\Omega))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

En fait, il s’agit ici de la définition générale de variables aléatoires ; il est clair que si l’on prend $\mathcal{P}(\Omega)$ comme tribu sur Ω toute application de Ω dans \mathbb{R} est mesurable (quelle que soit la tribu considérée sur \mathbb{R} ; en effet demander $X^{-1}(B) \in \mathcal{P}(\Omega)$ est équivalent à ne rien demander). Ce ne sera plus le cas dès que l’espace Ω sera plus grand ou non muni de la tribu $\mathcal{P}(\Omega)$ (voir chapitre sur les espaces de probabilités généraux).

Remarque : Une variable aléatoire n’est pas une fonction aléatoire !! Ce qui est aléatoire (si on a envie de faire ce genre d’interprétation), c’est son argument.

b. **Lois images**

Définition : Soit X une variable aléatoire réelle sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$. X induit sur \mathbb{R} une nouvelle mesure, notée \mathbb{P}_X , dite loi image de \mathbb{P} par X . Elle est définie pour tout élément A de $\mathcal{B}(\mathbb{R})$ par :

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega, X(\omega) \in A\}).$$

Notons Ω_X l'ensemble image de Ω par X (c'est-à-dire $\Omega_X = \{X(\omega), \omega \in \Omega\}$). Puisque Ω est au plus dénombrable, Ω_X aussi. La mesure \mathbb{P}_X en fait, ne charge que Ω_X et on a :

$$\mathbb{P}_X(\Omega_X) = \mathbb{P}(X^{-1}(\Omega_X)) = \mathbb{P}(\Omega) = 1.$$

Ainsi, si $A \cap \Omega_X = \emptyset$ alors $\mathbb{P}_X(A) = 0$.

Définition : La mesure \mathbb{P}_X est appelée *loi* de X . On utilise, entre autres, les notations suivantes :

$$\begin{aligned} \mathbb{P}_X(\{a\}) &= \mathbb{P}(X = a) = \mathbb{P}(\{\omega, X(\omega) = a\}), \\ \mathbb{P}_X([a, +\infty[) &= \mathbb{P}(X \geq a) = \mathbb{P}(\{\omega, X(\omega) \geq a\}) \end{aligned}$$

Deux exemples de v.a. et de loi image :

a) $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{i\}) = 1/6 \forall i \in \Omega$. Soit $X : \Omega \rightarrow \mathbb{R}$ t.q. $X(\omega) = \omega(3 - \omega)$ pour tout $\omega \in \Omega$

b) $\Omega = \mathbb{N}^*$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{i\}) = 1/2^i$, Soit $Y : \Omega \rightarrow \mathbb{R}$ t.q. $Y(\omega) = \omega^2$.

Les deux applications X et Y sont des variables aléatoires. Dans le cas (a), on peut imaginer un jeu de dé (à un joueur). Si le dé tombe sur la face i le joueur gagne la somme $i(3 - i)$. Dans le cas (b), $\mathbb{P}(\{i\})$ est la probabilité qu'il soit nécessaire de lancer i fois une pièce avant de tomber sur *face*. $X(\omega)$ est le carré de ce nombre de coups.

Lois images des deux exemples

Dans le cas (a), $X(1) = 2, X(2) = 2, X(3) = 0, X(4) = -4, X(5) = -10, X(6) = -18$. Ainsi $\Omega_X = \{2, 0, -4, -10, -18\}$.

On a alors

$$\mathbb{P}_X(\{2\}) = \mathbb{P}(X = 2) = \mathbb{P}(\{\omega, X(\omega) = 2\}) = \mathbb{P}(\{1, 2\}) = 2/6.$$

En utilisant le même genre de "décomposition" on trouve $\mathbb{P}(X = 0) = \mathbb{P}(X = -4) = \mathbb{P}(X = -10) = \mathbb{P}(X = -18) = 1/6$.

Dans le cas (b), \mathbb{P}_Y ne charge que les entiers qui sont des carrés d'entier. On a, pour i entier,

$$\mathbb{P}_Y(\{i^2\}) = \mathbb{P}(Y = i^2) = \mathbb{P}(\{\omega, Y(\omega) = i^2\}) = \mathbb{P}(\{i\}) = 1/2^i$$

et $\mathbb{P}(Y = i) = 0$ si i n'est pas un carré d'entier.

Remarque 3 (À propos de la terminologie "loi image")

Dans l'exemple (a), la mesure sur Ω , \mathbb{P} s'écrit :

$$\mathbb{P} = \sum_{i=1}^6 \frac{1}{6} \delta_i \quad \text{et} \quad \mathbb{P}_X = \sum_{i=1}^6 \frac{1}{6} \delta_{X(i)}$$

Dans le cas (b), $\mathbb{P} = \sum_{i \geq 1} 2^{-i} \delta_i$ et $\mathbb{P}_Y = \sum_{i \geq 1} 2^{-i} \delta_{Y(i)} = \sum_{i \geq 1} 2^{-i} \delta_{i^2}$

c. **Espérance, variance et moments**

Définition : Soient X une variable aléatoire réelle et k un entier positif; on dit que X admet un moment d'ordre k si

$$\sum_{\omega \in \Omega} |X(\omega)|^k \mathbb{P}(\{\omega\}) < +\infty.$$

Dans ce cas, le moment d'ordre k , noté m_k est défini par

$$m_k = \sum_{\omega \in \Omega} X(\omega)^k \mathbb{P}(\{\omega\}) = \mathbb{E}(X^k);$$

On peut remarquer que la convergence ou non de cette somme ne dépend pas de l'ordre de sommation (c'est une conséquence de la théorie de l'intégration de Lebesgue). Lorsque X possède un moment d'ordre 1, on appelle moyenne de X (ou espérance mathématique de X), la quantité (qui est alors bien définie)

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) = m_1.$$

On somme alors par paquets, en utilisant la relation suivante : $\forall \omega \in \Omega, \sum_{x_i \in \Omega_X} \mathbf{1}_{\{x_i\}}(X(\omega)) = 1$ (car $X(\omega)$ vaut un et un seul x_i). On a

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} \left(\sum_{x_i \in \Omega_X} \mathbf{1}_{\{x_i\}}(X(\omega)) \right) X(\omega) \mathbb{P}(\{\omega\}) = \sum_{x_i \in \Omega_X} \left(\sum_{\omega \in \Omega} \mathbf{1}_{\{x_i\}}(X(\omega)) X(\omega) \mathbb{P}(\{\omega\}) \right)$$

on trouve :

$$\mathbb{E}(X) = \sum_{x_i \in \Omega_X} x_i \mathbb{P}(X = x_i). \tag{7}$$

et pour tout a et b réels, si X et Y admettent des moments d'ordre 1,

$$\mathbb{E}(aX + bY) = a \mathbb{E}(X) + b \mathbb{E}(Y). \tag{8}$$

On a juste besoin de la linéarité de la somme.

Il est intéressant de remarquer que la formule (8) est vraie même si les variables X et Y sont non indépendantes (cette notion est définie plus loin).

Exercice 4 : a) Soient p et q deux entiers tels que $0 < p < q$. Montrer que si X admet un moment d'ordre q alors il admet aussi un moment d'ordre p

b) Donner une v.a. X telle que X possède un moment d'ordre p mais pas d'ordre q .

Définition : Lorsque X possède un moment d'ordre 2, on appelle variance de X la quantité

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

L'écart type de X , noté $\sigma(X)$ est la quantité

$$\sigma(X) = \sqrt{\text{Var}(X)}.$$

Une identité souvent utilisée (et dont la preuve est laissée en exercice) est la suivante

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

Cette formule montre bien que la variance est la moyenne des écarts quadratiques de la variable X à sa moyenne. Plus la variance est grande, plus X est dispersée pour cette mesure.

Propriété 2 (i) Pour tout a et b réels, et X v.a. admettant un moment d'ordre 2

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

(ii) Soit X une variable admettant un moment d'ordre 2. La fonction $a \mapsto f(a) = \mathbb{E}((X - a)^2)$ admet un minimum unique pour $a = \mathbb{E}(X)$; ce minimum vaut $\text{Var}(X)$.

Preuve : La preuve de (i) est laissée en exercice. Voyons (ii).

$$\begin{aligned} \mathbb{E}((X - a)^2) &= \mathbb{E}((X - \mathbb{E}(X) + \mathbb{E}(X) - a)^2) \\ &= \mathbb{E}((X - \mathbb{E}(X))^2 + \mathbb{E}((\mathbb{E}(X) - a)^2) + 2\mathbb{E}((X - \mathbb{E}(X)))(\mathbb{E}(X) - a)) \\ &= \text{Var}(X) + \mathbb{E}((\mathbb{E}(X) - a)^2) \end{aligned}$$

car $\mathbb{E}((X - \mathbb{E}(X))) = 0$. Puisque $\mathbb{E}((\mathbb{E}(X) - a)^2) = (\mathbb{E}(X) - a)^2 \geq 0$ et ne vaut 0 que si $a = \mathbb{E}(X)$ on a le résultat annoncé. \square

Exercice 5 : On dit qu'une v.a. X est constante s'il existe un réel c tel que $\mathbb{P}(X = c) = 1$.
Montrer $(\text{Var}(X) = 0) \iff (X \text{ est constante.})$

d. Espérance d'une fonction d'une variable aléatoire

Soit X une variable aléatoire réelle et Φ une fonction de \mathbb{R} dans \mathbb{R} , alors il est aisé de vérifier, que dans le cas discret (où Ω est muni de la tribu $\mathcal{P}(\Omega)$), $\Phi \circ X$ est une variable aléatoire; de plus si

$$\sum_{\omega \in \Omega} |\Phi(X(\omega))| \mathbb{P}(\{\omega\}) < +\infty$$

l'espérance mathématique de la variable aléatoire $\Phi(X)$ est donnée par

$$\begin{aligned} \mathbb{E}(\Phi(X)) &= \sum_{\omega \in \Omega} \Phi(X(\omega)) \mathbb{P}(\{\omega\}) \\ &= \sum_{x \in \Omega_X} \Phi(x) \mathbb{P}(X = x) \\ &= \sum_{y \in \Omega_{\Phi(X)}} y \mathbb{P}(\Phi(X) = y). \end{aligned}$$

Ces trois formules permettent de calculer $\mathbb{E}(\Phi(X))$ à trois niveaux : sur Ω (avec \mathbb{P}), sur Ω_X (avec \mathbb{P}_X) sur $\Omega_{\Phi(X)}$ avec $\mathbb{P}_{\Phi(X)}$.

e. Inégalité de Markov et de Bienaymé-Tchebichev

Propriété 3 (Inégalité de Markov) Soit X une variable aléatoire positive sur Ω . Pour tout $x \in \mathbb{R}^{+\ast}$, on a :

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}$$

Preuve :

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) \geq \sum_{\{\omega \mid X(\omega) \geq x\}} X(\omega)\mathbb{P}(\omega) \geq \sum_{\{\omega \mid X(\omega) \geq x\}} x \mathbb{P}(\omega) = x \mathbb{P}(X \geq x). \quad \square$$

Ainsi, on voit que lorsque X a une moyenne, la queue de distribution de la variable X (la fonction $x \mapsto \mathbb{P}(X \geq x)$) est au plus de l'ordre de $1/x$. Pour prouver que la queue de distribution est plus petite encore, on peut utiliser les moments de X d'ordre supérieur ; par exemple, si X est positive et possède un moment d'ordre 6 (si $\mathbb{E}(X^6) < +\infty$) alors

$$\mathbb{P}(X \geq x) = \mathbb{P}(X^6 \geq x^6) \underset{\text{Markov}}{\leq} \frac{\mathbb{E}(X^6)}{x^6};$$

on voit maintenant que la queue de distribution est au plus de l'ordre de x^{-6} . Si $\mathbb{E}(e^X) \leq +\infty$ on peut montrer (cette fois ceci est valable même si X n'est pas positive) que

$$\mathbb{P}(X \geq x) = \mathbb{P}(e^X \geq e^x) \leq \frac{\mathbb{E}(e^X)}{e^x}.$$

Cette fois la queue de distribution est exponentiellement petite.

Ces deux exemples d'utilisation de l'inégalité de Markov montrent que bien souvent, l'inégalité de Markov ne donne pas une bonne majoration. Elle montre aussi qu'il y a un lien profond entre l'existence de moments d'ordre k et le comportement de la queue de distribution.

Inégalité de Bienaymé-Tchebichev

Propriété 4 (Inégalité de Bienaymé-Tchebichev) Soit X une variable aléatoire réelle possédant un moment d'ordre 2 et $x > 0$:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\text{Var}(X)}{x^2}. \tag{9}$$

Cette inégalité montre l'intérêt de la variance pour mesurer la dispersion d'une v.a..

Preuve :

$$\begin{aligned} \mathbb{P}(|X - \mathbb{E}(X)| \geq x) &= \mathbb{P}(|X - \mathbb{E}(X)|^2 \geq x^2) \\ &\underset{\text{Markov}}{\leq} \frac{\mathbb{E}(|X - \mathbb{E}(X)|^2)}{x^2} = \frac{\text{Var}(X)}{x^2}. \quad \square \end{aligned}$$

L'inégalité de BT peut se reformuler comme suit :

$$\mathbb{P}(X \notin [\mathbb{E}(X) - x, \mathbb{E}(X) + x]) \leq \frac{\text{Var}(X)}{x^2}.$$

6. Exemples de lois discrètes

- X est une variable de Bernoulli de paramètre p , pour $p \in [0, 1]$, on note $X \sim B(p)$, si $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = 1 - p$. On a immédiatement, $\mathbb{E}(X) = p$, $\text{Var}(X) = p(1 - p)$.

- X est une variable binomiale de paramètre n et p , pour $n \in \mathbb{N}$ et $p \in [0, 1]$, on note $X \sim B(n, p)$, si X est la somme de n variables de Bernoulli $B(p)$ indépendantes. On obtient par un petit dénombrement,

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k} \text{ pour tout } k \in \{0, \dots, n\}$$

et $\mathbb{E}(X) = np$, $\text{Var}(X) = np(1 - p)$.

- X suit la loi uniforme sur $\{1, \dots, n\}$, pour $n \in \mathbb{N}^*$, on note $X \sim \mathcal{U}\{1, \dots, n\}$, si $\mathbb{P}(X = k) = 1/n$ pour tout $k \in \{1, \dots, n\}$. On a $\mathbb{E}(X) = (n + 1)/2$, $\text{Var}(X) = (n^2 - 1)/12$.

- On dit que X suit une loi de Poisson de paramètre λ , pour $\lambda > 0$, on note $X \sim \mathcal{P}(\lambda)$, si la loi de X est donnée par

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ pour } k \in \mathbb{N}.$$

On a $\mathbb{E}(X) = \text{Var}(X) = \lambda$.

- On dit que X suit une loi géométrique (ou de Pascal) de paramètre p , pour $p \in]0, 1]$, on note $X \sim \mathcal{G}(p)$, si

$$\mathbb{P}(X = k) = p(1 - p)^{k-1} \text{ pour tout } k \in \mathbb{N}^*.$$

On a $\mathbb{E}(X) = 1/p$, $\text{Var}(X) = (1 - p)/p^2$.

Remarque 4 Sur la construction d’une variable géométrique par une suite infinie d’expériences (on pourra passer cette partie dans une première lecture) Souvent, on voit apparaître la loi géométrique comme le temps d’apparition d’un événement dans une suite d’expériences aléatoires. Par exemple, si on lance une pièce dont la probabilité de tomber sur pile est p , sur face $1 - p$, et si les lancers sont indépendants, le nombre de lancers nécessaires pour l’apparition de “pile” est une variable aléatoire dont la loi est $\mathcal{G}(p)$. Si on peut rapidement l’admettre, on peut néanmoins se demander dans quel espace de probabilité on a travaillé pour calculer cela. Il s’agit bien sûr de l’espace $\Omega = \{pile, face\}^{\mathbb{N}^*}$ ou $\Omega = \{0, 1\}^{\mathbb{N}^*}$ l’ensemble des suites infinies de 0 et de 1 (correspondant aux suites de pile-face). ω , un élément de Ω est une suite infinie $\omega = (\omega_1, \omega_2, \dots)$. Ce qui nous intéresse, c’est $A_k = \{\omega \mid \inf\{j \mid \omega_j = 1\} = k\}$, l’ensemble des suites dont le premier rang d’apparition d’un 1 est k . Il reste donc à définir une tribu \mathcal{A} sur Ω et une mesure \mathbb{P} sur cette tribu cohérente avec ce que l’on veut faire. On voudrait par exemple, que la probabilité d’avoir une suite commençant par 1, 0, 1, 0, 0 soit $p^2(1 - p^3)$. Eh bien, en fait ceci est relativement difficile à faire.

Si p vaut $1/2$ (ou rationnel), pour construire \mathbb{P} , on peut transporter la mesure de Lebesgue de $[0, 1]$ ainsi que la tribu des Lebesgue-mesurables; on procède comme suit : on identifie l’élément $(\omega_1, \omega_2, \omega_3, \dots)$ de Ω avec le nombre de $[0, 1]$ dont l’écriture en base 2 est $0, \omega_1 \omega_2 \omega_3 \dots$ (on a comme d’habitude un problème à cause du fait que certains nombres possèdent deux écritures en base 2, mais ils sont dénombrables et non gênants). Notons $T : \Omega \rightarrow [0, 1]$ cette application. On munit Ω de la tribu rendant mesurable T (c’est à dire on prend comme tribu sur Ω les ensembles $T^{-1}(A)$ où A parcourt l’ensemble des Lebesgue-mesurables de $[0, 1]$). On transporte également la mesure de Lebesgue λ comme d’habitude : $\mathbb{P} = \lambda \circ T$. De cette manière

$$\mathbb{P}(A_k) = \lambda\{x = 0, x_1 \dots x_{k-1} 1 x_{k+1} \dots; \text{ avec } x_i = 0 \text{ pour } i < k, x_i \in \{0, 1\} \text{ pour } i > k\} = 2^{-k}.$$

(Si p est rationnel $p = n/m$ on écrit les nombres en base m , et on regarde cette fois les décimales inférieures à n).

Si p n’est pas rationnel, ou si on observe une première réalisation dans une suite d’expériences dont la probabilité change à chaque fois, tout se complique. On utilise un théorème de Kolmogorov qui dit, en gros, que l’on peut construire un espace de probabilité (et une mesure donc) sur un espace produit infini si les mesures des cylindres sont “consistantes”, pour une certaine notion de consistance (ceci est expliqué dans le livre de Loève).

7. Famille de variables aléatoires

Définition : Soient (X_1, \dots, X_n) une famille de variables aléatoires définies sur un même espace $(\Omega, \mathbb{P}(\Omega))$ et à valeurs dans $E_1 \times \dots \times E_n$.

Les variables X_1, \dots, X_n sont dites indépendantes si, pour tout $(x_1, \dots, x_n) \in E_1 \times \dots \times E_n$,

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i).$$

Si X et Y sont indépendantes, on note $X \perp\!\!\!\perp Y$.

Remarque 5 • les virgules dans le membre de gauche doivent être lues “et” ou “intersection”.

- L'indépendance des variables aléatoires deux à deux n'implique pas l'indépendance .
- On trouvera des propriétés importantes des v.a. indépendantes dans la proposition 10 page 32 ; la définition de v.a. indépendantes dans le cas général est donnée page 30.

Proposition 3 Les variables aléatoires (X_1, \dots, X_n) sont indépendantes si et seulement si pour toute fonction bornée $f_i : E_i \rightarrow \mathbb{R}$

$$\mathbb{E}(f_1(X_1) \dots f_n(X_n)) = \prod_{i=1}^n \mathbb{E}(f_i(X_i)) \quad (10)$$

Preuve : Tout d'abord, on voit qu'en prenant $f_i = \mathbb{1}_{\{x_i\}}$ on obtient que (10) implique l'indépendance. Maintenant, supposons les X_i indépendants.

$$\begin{aligned} \mathbb{E}(f_1(X_1) \dots f_n(X_n)) &= \sum_{\omega \in \Omega} \left(\prod_{i=1}^n f_i(X_i(\omega)) \right) \mathbb{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} \left(\sum_{E_1 \times \dots \times E_n} \mathbb{1}_{\{(x_1, \dots, x_n)\}}(X_1(\omega), \dots, X_n(\omega)) \right) \left(\prod_{i=1}^n f_i(X_i(\omega)) \right) \mathbb{P}(\{\omega\}) \\ &= \sum_{E_1 \times \dots \times E_n} \left(\prod_{i=1}^n f_i(x_i) \right) \mathbb{P}(X_1(\omega) = x_1, \dots, X_n(\omega) = x_n) \\ &= \sum_{E_1 \times \dots \times E_n} \prod_{i=1}^n \left(f_i(x_i) \mathbb{P}(X_i(\omega) = x_i) \right) \\ &= \prod_{i=1}^n \mathbb{E}(f_i(X_i)) \end{aligned}$$

où il faut comprendre la somme sur $E_1 \times \dots \times E_n$ par somme pour tout $(x_1, \dots, x_n) \in E_1 \times \dots \times E_n$.

Explication : d'après le théorème de Fubini, puisque les fonctions étant bornées (sur des espaces finis), on a $\sum_{\omega \in \Omega} \left(\prod_{i=1}^n f_i(X_i(\omega)) \right) \mathbb{P}(\{\omega\}) \leq \prod_{i=1}^n \|f_i\|_\infty < +\infty$. On peut donc intervertir les signes sans craintes.

La première ligne est la définition de l'espérance ; dans la deuxième ligne, la première parenthèse dans le membre de droite vaut 1 ; dans la troisième, on a classé les ω en fonction des valeurs $X_i(\omega)$; dans la quatrième, on a fait jouer l'indépendance ; dans la cinquième, puisque les variables sont séparées, on a interverti le signe \prod et \sum . \square

Corollaire 1 Si X et Y sont deux variables aléatoires indépendantes admettant des moments d'ordre 2 alors

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Preuve :

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

Il reste à montrer que $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. On ne peut pas conclure tout de suite car $x \mapsto x$ n'est pas bornée. Mais on sait que X et Y ont des moments d'ordre 2 et donc d'ordre 1. Donc

$$\begin{aligned}\mathbb{E}(XY) &= \sum \sum xy \mathbb{P}(X = x, Y = y) = \sum \sum xy \mathbb{P}(X = x) \mathbb{P}(Y = y) \\ &= \sum x \mathbb{P}(X = x) \sum y \mathbb{P}(Y = y) = \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

L'application de Fubini ici est loisible grâce, comme on l'a dit, au fait que $\sum |x| \mathbb{P}(X = x) < +\infty$ et $\sum |y| \mathbb{P}(Y = y) < +\infty$; \square

8. Fonctions génératrices de probabilité

Définition : Soit X une variable aléatoire à valeurs dans \mathbb{N} . On appelle fonction génératrice de probabilité (ou série génératrice), la série entière

$$\mathbf{G}_X(s) = \mathbb{E}(s^X) = \sum_{n=0}^{+\infty} s^n \mathbb{P}(X = n) = \sum_{n=0}^{+\infty} s^n \mathbb{P}_X(n).$$

Cette série entière a un rayon de convergence supérieur à 1 car $\mathbf{G}_X(1) = 1$.

Proposition 4 *La série génératrice caractérise entièrement la loi de X*

Preuve : Il suffit d'extraire ses coefficients pour retrouver la loi de X . \square

De même, on voit facilement que

$$\mathbb{E}(X) = \mathbf{G}'_X(1),$$

et

$$\mathbb{E}(X^2) = \mathbf{G}''_X(1) + \mathbf{G}'_X(1);$$

on prendra garde au fait que ces quantités peuvent "valoir" $+\infty$ (habituellement, on dit plutôt que $\mathbb{E}(X) = \lim_{1^-} \mathbf{G}'_X(s)$).

La grande utilisation des fonctions génératrices provient surtout de la propriété suivante (utilisée en conjonction avec la Proposition 4) :

Proposition 5 *Si X et Y sont deux variables aléatoires indépendantes (à valeurs dans \mathbb{N}) alors*

$$\mathbf{G}_{X+Y} = \mathbf{G}_X \mathbf{G}_Y \quad \text{sur } [-1, 1]$$

Preuve : Soit s dans $[-1, 1]$.

$$\mathbf{G}_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y) = \mathbb{E}(s^X) \mathbb{E}(s^Y) = \mathbf{G}_X(s) \mathbf{G}_Y(s).$$

Seule la troisième égalité pose problème. On utilise la Proposition 3 avec $f_1(x) = f_2(x) = s^x$ ($s \in [-1, 1]$ est fixé ici, et x est dans \mathbb{N}). Les f_i sont bien bornées (sur \mathbb{N}), et comme X et Y sont indépendantes, on peut conclure. \square

Exemples de fonctions génératrices

- Si B est une variable de Bernoulli $B(p)$, on a

$$\mathbf{G}_B(s) = 1 - p + ps \quad \text{pour } s \in \mathbb{R}.$$

- Une variable X de loi binomiale $B(n, p)$ est égale en loi à la somme de n copies indépendantes de Bernoulli $B(p)$; ainsi

$$\mathbf{G}_X(s) = (1 - p + ps)^n \quad \text{pour } s \in \mathbb{R}.$$

- Pour Y suivant une loi de Poisson $\mathcal{P}(\lambda)$, on trouve

$$\mathbf{G}_Y(s) = \mathbb{E}(s^Y) = \sum_{k=0}^{+\infty} s^k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda + \lambda s} \quad \text{pour } s \in \mathbb{R}.$$

Il est vraiment aisé grâce à ce dernier point et à la proposition 5 de montrer que la somme de deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs λ et μ suit une loi de Poisson $\mathcal{P}(\lambda + \mu)$.

- Si X suit une loi géométrique de paramètre p , on a

$$\mathbf{G}_X(s) = \sum_{k=1}^{+\infty} p(1-p)^{k-1} s^k = \frac{ps}{1 - (1-p)s} \quad \text{pour } |s| \leq (1-p)^{-1}.$$

9. Variance et covariance

Définition : Soient X et Y deux variables possédant des moments d'ordre 2, la covariance de X et de Y est la quantité

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

Il est clair que $\text{cov}(X, Y) = \text{cov}(Y, X)$; on peut voir également que $\text{cov}(X, X) = \text{Var}(X)$. On a aussi

$$\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))),$$

ce que l'on vérifie aisément en développant le second membre de cette formule.

Cette deuxième formule montre que $\text{cov}(X, Y)$ est positive lorsque X et Y "ont tendance" à être en même temps du même côté de leur moyenne.

Propriété 5 Si X et Y sont indépendantes et possèdent un moment d'ordre 2, alors $\text{cov}(X, Y) = 0$ (la réciproque est fautive).

Si X et Y ont des moments d'ordre 2,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y).$$

La première propriété provient du fait que si X et Y sont indépendantes alors $\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(X - \mathbb{E}(X))\mathbb{E}(Y - \mathbb{E}(Y))$. Pour la deuxième, on écrit,

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X + Y))^2 \\ &= \mathbb{E}(X^2) + \mathbb{E}(Y^2) + 2\mathbb{E}(XY) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y). \quad \square \end{aligned}$$

III. Espaces de probabilités généraux

On introduit ici l'axiomatique de Kolmogorov qui est relativement simple (après avoir fait le cours d'intégration) et tout ce que l'on a déjà dit. Mais il ne faut pas se tromper : avant Kolmogorov, on ne faisait pas des probabilités comme maintenant. D'autre part, d'un point de vue métaphysique (ou en tout cas, pour comprendre les rapports entre "probabilité mathématique" et la conception intuitive des probabilités comme mesure de hasard), la construction de Kolmogorov est très importante.

1. axiomatique de Kolmogorov :

Ω est un ensemble muni d'une tribu \mathcal{A} . On appelle mesure de probabilité sur (Ω, \mathcal{A}) une mesure \mathbb{P} de masse totale 1.

Quelques remarques

- Il n'est fait – comme annoncé plus haut – aucune allusion au hasard ou à une suite quelconque d'expériences pour définir la probabilité d'un événement (un événement A est un élément de la tribu, sa probabilité est la mesure de A).
- La probabilité d'un événement (que l'on peut interpréter comme sa masse) est définie dans un espace mesurable Ω . Lorsque l'on modélise une expérience aléatoire, on définit un espace Ω que l'on munit ensuite d'une probabilité \mathbb{P} . Cette modélisation est une "idéalisée" de la réalité.

2. Variables aléatoires, espérance

On généralise ici, les définitions et théorèmes du chapitre sur les espaces discrets.

Définition : Soit (E, \mathcal{E}) un espace mesurable. Une application mesurable de (Ω, \mathcal{A}) dans (E, \mathcal{E}) est appelée *variable aléatoire* (c'est-à-dire, X est une v.a. si $\forall B \in \mathcal{E}, X^{-1}(B) \in \mathcal{A}$).

Dans ce cours, E sera presque toujours \mathbb{R}^d (avec $d = 1$ le plus souvent); la tribu \mathcal{E} sera le plus souvent la tribu borélienne sur \mathbb{R}^d (ou celle des Lebesgue mesurables).

Loi image

Définition : Soit X une variable aléatoire réelle sur un espace $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeur dans (E, \mathcal{E}) . X induit sur l'espace mesurable (E, \mathcal{E}) une nouvelle mesure, notée \mathbb{P}_X , dite loi image de \mathbb{P} par X . Elle est définie sur tout élément B de \mathcal{E} par :

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(\{\omega | X(\omega) \in B\}). \quad (11)$$

Ici, la nécessité pour X d'être mesurable est criante !! Sans cela, $X^{-1}(B)$ ne serait pas nécessairement dans \mathcal{A} , domaine où la mesure \mathbb{P} est définie.

On peut montrer aisément que \mathbb{P}_X est bien définie et est bien une probabilité sur (E, \mathcal{E}) , c'est-à-dire une mesure de masse totale 1.

Grâce à la mesure \mathbb{P}_X on définit une notion d'intégrale que l'on appelle espérance :

$$\mathbb{E}(f(X)) = \int_E f(x) d\mathbb{P}_X(x) = \int_\Omega f(X(\omega)) d\mathbb{P}(\omega)$$

(voir cours d'intégration).

Si la quantité

$$\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega), \quad (12)$$

est finie, on appelle *espérance mathématique* de X (ou moment d'ordre 1, ou moyenne), la quantité

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

qui est alors aussi définie et finie. (Il arrive que l'on dise que l'espérance de X soit $+\infty$ ou $-\infty$; mais sans (12), il est possible que l'espérance ne converge pas dans $\overline{\mathbb{R}}$).

On a, pour B dans \mathcal{E} ,

$$\mathbb{P}(\{\omega | X(\omega) \in B\}) = \int_{\Omega} \mathbf{1}_B(X(\omega)) d\mathbb{P}(\omega) = \mathbb{E}(\mathbf{1}_B(X)) = \mathbb{P}_X(B) = \int_E \mathbf{1}_B(x) d\mathbb{P}_X(x) = \int_B d\mathbb{P}_X(x).$$

Remarque 6 (sur l'apparition des tribus) Comme rappelé plus haut, une probabilité est une mesure. Comme telle, elle est une application définie sur une tribu de Ω et non pas sur Ω directement.

Sur des ensembles dénombrables, une mesure est entièrement déterminée par sa valeur sur les atomes. En général, lorsque Ω est discret, on prend comme tribu $\mathcal{P}(\Omega)$. Ainsi, dans ce cas, tous les événements sont mesurables et \mathbb{P} leur attribue donc une mesure. Regardons maintenant ce qui se passerait si la tribu sur Ω dénombrable ne contenait pas tous les atomes, (c'est-à-dire si elle était différente de $\mathcal{P}(\Omega)$). La mesurabilité d'une variable aléatoire X nécessite que $X^{-1}(\{a\})$ soit dans la tribu (pour tout a de \mathbb{R}). Une simple analyse montre que X doit être constante sur les éléments minimaux de la tribu. Si on considère un élément minimal A de la tribu non réduit à un point (A existe par hypothèse) \mathbb{P} attribue une probabilité à A mais pas à ses éléments. Tout cela nous montre que distinguer les atomes constituant les éléments minimaux n'est pas raisonnable. Quitte à construire des classes d'équivalence ("appartenir à un même élément minimal") on peut donc considérer, dans le cas des espaces dénombrables que la tribu est $\mathcal{P}(\Omega)$ sans perdre vraiment en généralité.

Maintenant, considérons la mesure de Lebesgue λ sur $[0, 1]$. C'est visiblement une mesure de probabilité. Mais, comme vu dans le cours d'intégration, λ agit sur la tribu des Lebesgue-mesurables qui est différente de $\mathcal{P}([0, 1])$ (car il existe des ensembles non Lebesgue-mesurables). D'autre part, la mesure de Lebesgue attribue un poids 0 à tous les éléments de Ω . On voit donc clairement, que tout ce qui se passait dans le cas discret tombe à l'eau : connaître la mesure sur tous les éléments de Ω ne permet pas de reconstituer la mesure et d'autre part, il n'y a aucun moyen d'étendre la mesure de Lebesgue sur toutes les parties de $[0, 1]$. Il s'ensuit qu'il va exister des événements dont on ne pourra pas définir la probabilité. On se contentera donc ici des ensembles que l'on peut mesurer (et qui sont déjà bien nombreux) : les boréliens de \mathbb{R}^+ (ou les Lebesgue-mesurables, si on veut).

Ce qui se passe dans le cas où la probabilité est la mesure de Lebesgue sur $[0, 1]$ va se passer de manière identique (la mesure n'agira pas sur $\mathcal{P}(\Omega)$ mais seulement sur une sous tribu) dès que la mesure ne chargera pas qu'un nombre dénombrable de points; c'est le cas par exemple dès que la mesure admet une densité.

3. Lois de probabilité

Le théorème de décomposition des mesures (voir annexe) prévoit que chaque mesure peut s'exprimer sous la forme d'une somme de trois mesures étrangères :

- une mesure (discrète) chargeant un nombre dénombrable de points.
- une mesure admettant une densité par rapport à la mesure de Lebesgue.
- une mesure chargeant un Lebesgue-négligeable mais ne chargeant pas les points.

Dans ce cours on ne considère que les mesures du premier et deuxième type ; un exemple de mesure du troisième type est donné en annexe.

Les mesures du premier type sont des sommes de masses de Dirac pondérées (par des poids positifs de somme 1). Celle du second type sont du type $\mu = g\lambda$ où g est une fonction mesurable, positive, d'intégrale 1 par rapport à la mesure de Lebesgue λ . La fonction g est appelée *densité de probabilité*.

Définition : Soit \mathbb{P} une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On dit que \mathbb{P} admet une densité si il existe une fonction g positive, mesurable, d'intégrale 1 (contre la mesure de Lebesgue) telle que

$$\mathbb{P} = g\lambda$$

c'est-à-dire, si \mathbb{P} admet la représentation suivante :

$$\text{pour tout borélien } B, \quad \mathbb{P}(B) = \int_B g(x) d\lambda(x) = \int_{\mathbb{R}} g(x) \mathbf{1}_B(x) d\lambda(x).$$

On voit que si g est d'intégrale 1 et positive, \mathbb{P} est bien une mesure de probabilité. D'après Radon-Nicodym et le théorème de décomposition de Lebesgue, la condition pour que \mathbb{P} admette une densité est qu'elle soit absolument continue par rapport à la mesure de Lebesgue.

Lemme 1 Si g et \tilde{g} sont deux fonctions mesurables telles que

$$\mathbb{P} = g\lambda = \tilde{g}\lambda$$

alors $g = \tilde{g}$, λ presque partout.

Preuve : Supposons que g et \tilde{g} diffèrent sur A , borélien non négligeable. Soient $A^+ = \{x, g(x) > \tilde{g}(x)\}$ et $A^- = \{x, g(x) < \tilde{g}(x)\}$. Alors, l'un des deux ensembles A^+ ou A^- n'est pas négligeable. Quitte à renommer g et \tilde{g} on peut supposer que $\lambda(A^+) > 0$. Mais $A^+ = \cup_n A_n^+$ avec $A_n^+ = \{x, g(x) \geq \tilde{g}(x) + 1/n\}$. Par suite, il existe n , tel que $\lambda\{A_n^+\} > 0$. Mais $\int_{A_n^+} g(x) d\lambda(x) > \int_{A_n^+} \tilde{g}(x) d\lambda(x)$ ce qui prouve que les deux mesures $g\lambda$ et $\tilde{g}\lambda$ diffèrent (puisqu'elles diffèrent sur A_n^+). \square

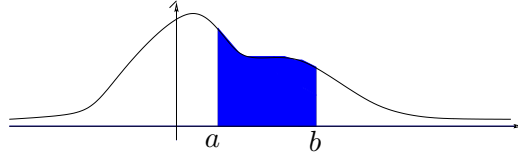
4. Variables aléatoires admettant une densité

Définition : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et X une variable aléatoire réelle (à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). On dit que X admet pour densité f (on devrait dire, pour être consistant que \mathbb{P}_X a pour densité f par rapport à la mesure de Lebesgue) si

$$\mathbb{P}_X = f\lambda.$$

On calcule alors la probabilité d'un événement par

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



Par exemple, si $A = [a, b]$, $\mathbb{P}(X \in [a, b]) = \int_a^b f(x) dx$.

Si X admet f pour densité, on a

$$\mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x) = \int_{\mathbb{R}} x f(x) dx$$

et pour h mesurable

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}} h(x) f(x) dx.$$

On rappelle que ces quantités sont définies si elles convergent absolument.

On a, par exemple

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f(x) dx$$

5. Exemples de lois admettant une densité

- **Loi uniforme :** On dit que X suit la loi uniforme sur $[0, 1]$ si la densité de X est $f(x) = \mathbb{1}_{[0,1]}(x)$; on note $X \sim \mathcal{U}[0, 1]$.

On a $\mathbb{E}(X) = 1/2$, $\text{Var}(X) = 1/12$.

On dit que Y suit la loi uniforme sur $[a, b]$ si sa densité est $f(x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a}$. On a $\mathbb{E}(Y) = (a+b)/2$ et $\text{Var}(Y) = (b-a)^2/12$

- **Loi normale :** On dit que X suit la loi normale de paramètres (m, σ^2) , pour $(m, \sigma^2) \in \mathbb{R} \times \mathbb{R}^{+\ast}$, on note $\mathcal{N}(m, \sigma^2)$ si la densité de X est

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

On a $\mathbb{E}(X) = m$, $\text{Var}(X) = \sigma^2$.

- **Loi exponentielle :** On dit que X suit la loi exponentielle de paramètre a (avec $a > 0$) si la densité de X est

$$f_a(x) = a \exp(-ax) \mathbb{1}_{\mathbb{R}^+}(x);$$

on note $X \sim \mathcal{E}(a)$. On a $\mathbb{E}(X) = 1/a$, $\text{Var}(X) = 1/a^2$.

- **Loi de Cauchy :** On dit que X suit la loi de Cauchy de paramètre $c > 0$, on note $X \sim C(c)$ si la densité de X est

$$f_c(x) = \frac{1}{\pi} \frac{c}{c^2 + x^2}.$$

Une variable de Cauchy ne possède pas de moyenne (et donc pas de variance).

6. Fonction de répartition

a. Définition et premières propriétés

Définition : Soit X une variable aléatoire dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La fonction

$$\begin{aligned} \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X(]-\infty, x]) \end{aligned}$$

est appelée fonction de répartition de X .

La fonction de répartition de X ne dépend que de la loi de X puisque $F_X(x) = \mathbb{P}_X(]-\infty, x])$. Un exercice classique d'intégration montre que la réciproque est également vraie : si on connaît la fonction de répartition de X , on peut calculer $\mathbb{P}_X(]a, b])$ pour tout a et b , puis, puisque les intervalles du type $]a, b]$ engendrent les boréliens, on peut reconstituer \mathbb{P}_X grâce à F_X (en fait, les ensembles du type $] - \infty, x]$ forment un π -système de Dynkin).

Propriété 6 Soit X une variable réelle.

a) F_X est croissante et on a $\lim_{x \rightarrow +\infty} F_X(x) = 1$, $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

b) F_X est continue à droite. La limite à gauche de F_X en x est $F_X(x-) = \mathbb{P}_X(]-\infty, x[)$. On a

$$\mathbb{P}_X(\{x\}) = F_X(x) - F_X(x^-).$$

Preuve : La croissance de F_X est évidente. Soit $A_n =]-\infty, n]$, A_n est une suite croissante d'ensembles. Comme $\bigcup_n A_n = \mathbb{R}$, on a $\lim \mathbb{P}_X(A_n) = \mathbb{P}_X(\mathbb{R}) = 1$. Donc, puisque F_X est croissante et que $F_X(n) \xrightarrow{n \rightarrow +\infty} 1$ on a $\lim_{x \rightarrow +\infty} F_X(x) = 1$. Faisons tendre maintenant n vers $-\infty$. A_n est décroissante vers \emptyset . Donc, $F_X(n) = \mathbb{P}_X(]-\infty, n]) \xrightarrow{n \rightarrow -\infty} 0$. On conclut en utilisant encore la croissance de F_X .

Soit x un point de \mathbb{R} . Montrons la continuité à droite en x . Soit x_n une suite de réels convergeant vers x et tels que $\forall n, x_n > x$. Supposons que $(x_n)_n$ décroît vers x . La suite d'ensembles $] - \infty, x_n]$ est décroissante vers $] - \infty, x]$ et on a donc (d'après les compléments), $\mathbb{P}(]-\infty, x_n]) \longrightarrow \mathbb{P}(]-\infty, x])$, autrement dit, $F_X(x_n) \longrightarrow F_X(x)$. Si cette fois x_n tend vers x , mais $x_n < x$, x_n croissante, la suite d'ensembles $] - \infty, x_n]$ est croissante vers $] - \infty, x[$ et donc $\mathbb{P}(]-\infty, x_n]) \longrightarrow \mathbb{P}(]-\infty, x[)$, d'où la conclusion.

Pour finir, on voit que $\mathbb{P}_X(\{x\}) = \mathbb{P}(]-\infty, x]) - \mathbb{P}(]-\infty, x[)$. \square

Remarque 7 Dans la preuve, pour montrer la continuité à gauche et l'existence d'une limite à droite, on a supposé que la suite $(x_n)_n$ était croissante dans le premier cas et décroissante dans le deuxième. On avait bien le droit : en effet, si y_n tend vers x avec $y_n > x$ (sans supposer que la suite y_n décroît vers x), la suite $(\tilde{y}_n)_n$ définie pour tout n par

$$\tilde{y}_n = \sup\{y_j, j \geq n\}$$

est une suite décroissante à qui on peut appliquer ce qui vient d'être dit. Donc $F_X(\tilde{y}_n) \longrightarrow F_X(x)$. Par ailleurs, $F_X(\tilde{y}_n) \geq F_X(y_n) \geq F_X(x)$ donc $F_X(y_n)$ converge aussi vers $F_X(x)$. Ainsi, pour démontrer la continuité à droite (resp. à gauche) on peut toujours supposer que la suite x_n que l'on considère est décroissante (resp. croissante). Par ailleurs, cette propriété ne tient pas à la croissance de F_X comme pourrait le laisser croire cette preuve.

b. Fonction de répartition et densité

Soit X une variable aléatoire réelle et F sa fonction de répartition. Il est indiqué plus haut que $\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = F(x) - F(x-)$. Comme \mathbb{P}_X est une probabilité sur \mathbb{R} elle ne peut charger qu'un nombre dénombrable de points (la somme des masses fait 1, donc le nombre des masses non nulles est au plus dénombrable). Ceci implique que F_X ne possède au plus qu'un nombre dénombrable de sauts, correspondant aux points chargés par \mathbb{P}_X . Un théorème (que l'on évoque juste ici) dit que l'on peut écrire F de manière unique sous la forme $F = F_c + F_S$ avec F_S constante par morceaux, continue à droite, croissante, et F_c continue. Il s'agit de la traduction de la décomposition de \mathbb{P}_X sous forme d'une somme de Dirac et d'une mesure ne chargeant pas les points, sous forme de fonctions de répartition (voir annexe). Supposons que $F = F_c$ (c'est-à-dire que F ne présente pas de saut). On ne peut pas pour autant en déduire que X (où plus rigoureusement \mathbb{P}_X) admet une densité (se souvenir de l'escalier du diable, fonction de répartition de la mesure construite en complément qui ne charge que le Cantor). Il faut, pour cela que F , bien que continue, ne croisse pas sur un ensemble négligeable ! Autrement dit, il faut que F soit l'intégrale de sa dérivée !

Propriété 7 Soit X une variable aléatoire admettant f comme densité. La fonction de répartition de X , F satisfait à

$$F(x) = \int_{-\infty}^x f(t) dt \text{ pour tout } x.$$

et donc

$$F'(x) = f(x) \quad \lambda \text{ presque partout.}$$

Réciproquement,

Lemme 2 soit F la fonction de répartition d'une variable aléatoire X ; si F est continue partout et dérivable sauf au plus sur un ensemble dénombrable de points, alors X admet une densité f . De plus $f(x) = F'(x)$ là où F est dérivable.

Remarquons que puisque F est dérivable p.p., f est définie seulement p.p.. Mais ce n'est pas grave car on cherche seulement la classe de f dans L_1 .

Éléments de preuve : on se raccroche à la décomposition des mesures (voir annexe). On cherche à prouver que \mathbb{P}_X est absolument continue, ce qui implique l'existence d'une densité (par Radon-NiKodym). On est donc amené à prouver que \mathbb{P}_X ne charge pas les Lebesgue-négligeables. On voit d'abord que F ne charge pas les points puisqu'elle est continue. Ensuite, on exclut une mesure de type singulière en remarquant qu'une telle mesure entraîne la non dérivabilité de F sur un ensemble négligeable non dénombrable. \square

7. Calcul de la loi d'une variable aléatoire

a. Grâce à la fonction de répartition

Supposons que X possède une densité f . Soit h une fonction mesurable. Quelle est la loi de $Y = h(X)$? Il existe principalement deux méthodes permettant de calculer la loi de Y lorsque la fonction h n'est pas trop compliquée :

a) On calcule la fonction de répartition de Y : (ceci est possible lorsque la fonction h est monotone).
On écrit

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(h(X) \leq x) = \mathbb{P}(X \leq h^{-1}(x)) = F_X(h^{-1}(x)).$$

La troisième égalité est légitime dans le cas où h est inversible et croissante. Comme application, on a

Propriété 8 Soit X une v.a. de densité f_X et Y définie par $Y = aX + b$ (pour un certain $a \geq 0$ et $b \in \mathbb{R}$). Y a pour densité

$$y \longmapsto f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Preuve : Exprimons la fonction de répartition de Y en fonction de celle de X :

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(aX + b \leq x) = \begin{cases} \mathbb{P}(X \leq \frac{x-b}{a}) = F_X(\frac{x-b}{a}) & \text{si } a > 0 \\ \mathbb{P}(X \geq \frac{x-b}{a}) = 1 - F_X(\frac{x-b}{a}) & \text{si } a < 0 \end{cases}$$

Par dérivation, on obtient le résultat annoncé et $f_Y(y) = -\frac{1}{a} f_X(\frac{y-b}{a})$ dans le cas où a est négatif. \square
Remarquons que cette méthode s'applique également au cas où F_X n'est pas simple puisqu'il n'est pas nécessaire de la connaître.

Propriété 9 Soit $X \sim \mathcal{N}(m, \sigma^2)$ et $Y \sim \mathcal{N}(0, 1)$. Les deux v.a. X et $\sigma Y + m$ ont même loi.

Preuve : Pour prouver ce deuxième point on montre que les fonctions de répartition des deux variables X et $\sigma Y + m$ sont égales puisque les fonctions de répartition caractérisent la loi.

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x \frac{e^{-\frac{1}{2}\left(\frac{u-m}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}} du.$$

$$\begin{aligned} F_{\sigma Y + m}(x) &= \mathbb{P}(\sigma Y + m \leq x) = \mathbb{P}\left(Y \leq \frac{x-m}{\sigma}\right) \\ &= \int_{-\infty}^{\frac{x-m}{\sigma}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \end{aligned}$$

Le changement de variable $t = \frac{u-m}{\sigma}$ dans cette dernière intégrale donne le résultat escompté, que l'on exprime souvent sous la forme :

Si $X \sim \mathcal{N}(m, \sigma^2)$ alors :

$$\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1). \tag{13}$$

b. Grâce au calcul de “l'espérance”

Soit X une variable aléatoire de densité f . On a vu plus haut, que si h est mesurable (et intégrable contre $f\lambda$) on a

$$\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)f(x)dx.$$

Cette formule est donc valable pour toute fonction f continue bornée. Il existe une “réciproque” à cette propriété :

Proposition 6 Soit X une variable aléatoire réelle. Supposons qu'il existe une fonction f positive mesurable, telle que pour toute fonction g continue bornée,

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x)f(x)dx, \quad (14)$$

alors f est une densité pour la variable X .

Rappelons que f est unique dans L_1 .

Preuve : Tout d'abord, en prenant g définie par $g(x) = 1$ pour tout $x \in \mathbb{R}$, on voit que $\int_{\mathbb{R}} f(x)dx = 1$ et que donc f est une densité (on peut avec quelques efforts se passer de l'hypothèse f positive).

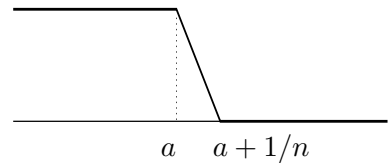
Il reste à montrer que $\mathbb{P}_X = f\lambda$. Pour cela on montre que pour tout a , $\mathbb{P}_X(]-\infty, a]) = \int_{-\infty}^a f(x)dx$. Ceci permettra de conclure car les ensembles du type $]-\infty, a]$ (union avec \mathbb{R}) forment une classe déterminante pour la mesure (un π -système de Dynkin).

Soit a fixé dans \mathbb{R} ; on considère C_a la fonction définie sur \mathbb{R} par $C_a(x) = \mathbb{1}_{]-\infty, a]}(x)$. On s'intéresse à

$$\mathbb{E}(C_a(X)) = \mathbb{P}(X \leq a) = F_X(a).$$

Pour $n \geq 1$, on définit la fonction g_n par

$$g_n(x) = \begin{cases} 1 & \text{si } x \leq a \\ n(a-x) + 1 & \text{si } x \in [a, a + 1/n] \\ 0 & \text{si } x \geq a + 1/n \end{cases} .$$



La suite $(g_n)_n$ est majorée par 1 et converge simplement vers la fonction C_a , or

- la fonction 1 est intégrable sous la mesure $f\lambda$ (λ est la mesure de Lebesgue).

Ainsi, par le théorème de Lebesgue (de convergence dominée) on a :

$$\int_{\mathbb{R}} C_a(x)f(x)dx = \lim_n \int_{\mathbb{R}} g_n(x)f(x)dx. \quad (15)$$

- D'autre part, la fonction 1 est intégrable sous la mesure \mathbb{P}_X . Donc par le théorème de Lebesgue

$$\mathbb{E}(C_a(X)) = \lim_n \mathbb{E}(g_n(X)) \quad (16)$$

(en effet : $\mathbb{E}(C_a(X)) = \int C_a(x)d\mathbb{P}_X(x)$ et $\mathbb{E}(g_n(X)) = \int g_n(x)d\mathbb{P}_X(x)$; on peut donc appliquer le théorème de Lebesgue).

Or par hypothèse $\mathbb{E}(g_n(X)) = \int_{\mathbb{R}} g_n(x)f(x)dx$ puisque g_n est continue bornée. On peut donc identifier les membres de gauche dans (15) et (16) : on a donc obtenu, ce que l'on cherchait :

$$\mathbb{E}(C_a(X)) = \mathbb{P}(X \leq a) = \int_{\mathbb{R}} C_a(x)f(x)dx = \int_{-\infty}^a f(x)dx. \quad \square$$

Calcul de la loi d'une fonction d'une variable aléatoire : Supposons X admettant comme densité f_X , connue. Soit $Y = h(X)$, avec h une fonction C^1 par morceaux. Cherchons la loi de Y : On écrit

$$\mathbb{E}(g(Y)) = \int g(y)f_Y(y)dy = \mathbb{E}(g(h(X))) = \int g(h(x))f_X(x)dx.$$

Maintenant, pour identifier f_Y , on fait un changement de variable dans la deuxième intégrale : on pose $y = h(x)$. Cette première intégrale se réécrit comme suit :

$$\int g(h(x))f_X(x)dx = \int g(y)f_X(h^{-1}(y))(h^{-1}(y))'dy \quad (17)$$

Ceci, d'après le lemme nous assure que

$$f_Y(y) = f_X(h^{-1}(y))|h^{-1}(y)'|$$

Remarque 8 si h est un C^1 difféomorphisme de \mathbb{R} dans \mathbb{R} , le changement de variable dans (17) ne pose pas de problème. Dès que h n'est plus injective, $h^{-1}(y)$ n'est plus un nombre mais un ensemble, et (17) n'est plus valide. Pour $Y = X^2$ par exemple, il faut scinder les intégrales en deux parties... On peut remarquer que la méthode par la fonction de répartition fonctionne dans à peu près les mêmes cas.

Un exemple : Soit X une variable exponentielle de paramètre 1 (donc de densité $f(x) = \mathbf{1}_{\mathbb{R}^+}(x)e^{-x}$). Cherchons la loi de $Y = \sqrt{X}$. On écrit :

$$\mathbb{E}(h(Y)) = \int_{\mathbb{R}} h(y)f_Y(y)dy = \mathbb{E}(h(\sqrt{X})) = \int_{\mathbb{R}} h(\sqrt{x})\mathbf{1}_{\mathbb{R}^+}(x)e^{-x}dx$$

On pose $y = \sqrt{x}$, d'où $x = y^2$, $dx = 2ydy$, ainsi on a

$$\mathbb{E}(h(Y)) = \int_{\mathbb{R}} h(y)2ye^{-y^2}\mathbf{1}_{\mathbb{R}^+}(y)dy;$$

y comme racine carrée devant être positif. Il s'ensuit que la densité de Y est

$$f_Y(y) = 2ye^{-y^2}\mathbf{1}_{\mathbb{R}^+}(y).$$

8. Lois jointes

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soient X_1, \dots, X_k des variables aléatoires réelles.

La loi image de \mathbb{P} par (X_1, \dots, X_k) que l'on peut encore noter $\mathbb{P}_{(X_1, \dots, X_k)}$ est une mesure borélienne sur \mathbb{R}^k . Elle est définie pour tout borélien B de \mathbb{R}^k par

$$\mathbb{P}_{(X_1, \dots, X_k)}(B) = \mathbb{P}((X_1, \dots, X_k) \in B) = \mathbb{P}(\{\omega, (X_1(\omega), \dots, X_k(\omega)) \in B\})$$

Elle peut, comme dans le cas unidimensionnel, être décrite par la fonction de répartition multidimensionnelle $F : \mathbb{R}^k \rightarrow [0, 1]$ définie par

$$F(x_1, \dots, x_k) = \mathbb{P}(X_1 \leq x_1, \dots, X_k \leq x_k) = \mathbb{P}_{(X_1, \dots, X_k)}\left(\prod_{i=1}^k]-\infty, x_i]\right).$$

En effet, la mesure $\mathbb{P}_{(X_1, \dots, X_k)}$ sur \mathbb{R}^k est bien déterminée par la fonction de répartition généralisée car les ensembles $\prod_{i=1}^k]-\infty, x_i]$ engendrent bien les boréliens de \mathbb{R}^k (et forment un π -système de Dynkin).

La classification des lois grâce à leur fonction de répartition est moins aisée ici. Le cas discret multidimensionnel a été traité plus tôt, où on a établi entre autres les faits suivants (nous raisonnons

dans ce qui suit sur les couples de variables aléatoires, mais on peut généraliser sans peine tout ce qui est dit ici à une famille finie de v.a.) :

Cas discret : la loi d'un couple de v.a. (X, Y) est déterminée par les valeurs de $\mathbb{P}((X, Y) = (x, y))$ pour $(x, y) \in \Omega_X \times \Omega_Y$. On a alors les formules suivantes

$$\mathbb{E}(f(X, Y)) = \sum_x \sum_y f(x, y) \mathbb{P}((X, Y) = (x, y)),$$

On appelle *marginales* ou lois marginales, les lois de X et de Y . Pour les calculer, on utilise la formule triviale suivante :

$$\mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R})$$

pour tout A borélien. Dans le cas discret, on écrit

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}((X, Y) = (x, y)) = \sum_y \mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y);$$

Si X et Y sont indépendantes, pour tout (x, y)

$$\mathbb{P}((X, Y) = (x, y)) = \mathbb{P}(X = x) \mathbb{P}(Y = y).$$

Voyons maintenant ce que deviennent ces formules dans le cas où les variables X et/ou Y ne sont pas discrètes.

Cas continu : Supposons que $\mathbb{P}_{(X, Y)}$ soit absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^2 . Il existe, d'après Radon-Nikodym, une fonction $f_{(X, Y)}$ positive, de \mathbb{R}^2 dans \mathbb{R}^2 , d'intégrale 1, telle que

$$\mathbb{P}_{(X, Y)} = f_{(X, Y)} \lambda_2$$

où λ_2 est la mesure de Lebesgue sur \mathbb{R}^2 ; autrement dit, pour tout B borélien de \mathbb{R}^2 ,

$$\mathbb{P}((X, Y) \in B) = \int_B f_{(X, Y)} d\lambda_2 = \int \int_B f_{(X, Y)}(x, y) dx dy.$$

$f_{(X, Y)}$ est appelée densité jointe du couple (X, Y) .

Proposition 7 *Supposons que le couple (X, Y) admette une densité $f_{(X, Y)}$ sur \mathbb{R}^2 alors X et Y admettent les densités suivantes sur \mathbb{R} :*

$$f_X(x) = \int_{\mathbb{R}} f_{(X, Y)}(x, y) dy, \quad f_Y(y) = \int_{\mathbb{R}} f_{(X, Y)}(x, y) dx.$$

Remarquons que la réciproque est fautive. X et Y peuvent avoir une densité alors que (X, Y) non. Par exemple, supposons que $X \sim \mathcal{U}[0, 1]$ et $Y = X$. Alors (X, Y) ne charge qu'une diagonale dans \mathbb{R}^2 et ne possède donc pas de densité (une droite est de mesure de Lebesgue 0 dans \mathbb{R}^2).

Preuve de la proposition : On a pour tout A borélien

$$\mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) = \int_A \int_{\mathbb{R}} f_{(X, Y)}(x, y) dy dx$$

ceci nous assure que

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dy \quad (18)$$

car $\mathbb{P}_X(A) = \int_A d\mathbb{P}_X(x)$ donc $d\mathbb{P}_X(x) = \left(\int_{\mathbb{R}} f_{(X,Y)}(x,y)dy \right) dx$; on voit alors que \mathbb{P}_X admet pour densité ce qui a été indiqué plus haut. Par symétrie, on trouve

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dx. \quad \square$$

Proposition 8 soit $f_{(X,Y)|Y=y}$ la fonction définie pour y t.q. $f_Y(y) > 0$ par

$$f_{(X,Y)|Y=y}(x) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)},$$

est une densité (de la loi conditionnelle de (X,Y) sachant $Y = y$).

Preuve :Le fait que $f_{(X,Y)|Y=y}$ soit une densité est évident. Pour sa dénomination, écrivons

$$\mathbb{P}((X,Y) \in A \times B) = \int_{y \in B} \left(\int_{x \in A} \frac{f_{(X,Y)}(x,y)}{f_Y(y)} dx \right) f_Y(y) dy$$

Calcul des espérances : Pour g mesurable de \mathbb{R}^2 dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\mathbb{E}(g(X,Y)) = \int \int g(x,y) f_{(X,Y)}(x,y) dx dy,$$

(ceci est bien défini si $\int \int |g(x,y)| f_{(X,Y)}(x,y) dx dy$ converge).

9. Indépendances

Définition : Soient $(X_i)_{i=1,\dots,k}$ des variables aléatoires à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La famille de v.a. X_i est dite indépendante si, pour tout $(A_1, \dots, A_k) \in (\mathcal{B}(\mathbb{R}))^k$

$$\mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k) = \prod_{i=1}^k \mathbb{P}(X_i \in A_i). \quad (19)$$

On dit aussi que les v.a. X_i sont indépendantes.

On dit qu'une famille infinie de v.a. $(X_i)_{i \in I}$ est indépendante, si toute sous famille extraite de cardinal fini est une famille de v.a. indépendante.

Si X et Y sont indépendants, on a

$$\mathbb{P}_{(X,Y)}(A \times B) = \mathbb{P}((X,Y) \in A \times B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B) = \mathbb{P}_X(A) \mathbb{P}_Y(B).$$

On voit que la mesure $\mathbb{P}_{(X,Y)}$ sur \mathbb{R}^2 est la mesure produit $\mathbb{P}_X \times \mathbb{P}_Y$. Comme, pour tout A et B intervalles de \mathbb{R} , on a

$$\mathbb{P}((X,Y) \in A \times B) = \int_{A \times B} d\mathbb{P}_{(X,Y)}(x,y)$$

mais aussi

$$\begin{aligned}\mathbb{P}((X, Y) \in A \times B) &= \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \\ &= \int_A f_X(x)dx \int_B f_Y(y)dy = \int_A \int_B f_X(x)f_Y(y)dydx\end{aligned}$$

on en déduit que $d\mathbb{P}_{(X,Y)}(x, y) = f_X(x)f_Y(y)dxdy$, ainsi $\mathbb{P}_{(X,Y)}$ possède une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^2 . On la note $f_{(X,Y)}$. Puisque les deux mesures $f_X(x)f_Y(y)dxdy$ et $f_{(X,Y)}(x, y)dxdy$ coïncident sur les pavés, on en conclut que

Proposition 9 Si X et Y sont indépendantes, si X et Y admettent respectivement f_X et f_Y comme densité, alors le couple (X, Y) a pour densité $f_{(X,Y)}$ sur \mathbb{R}^2 avec

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) \text{ p.p..}$$

La réciproque est également vraie :

$$\text{si } f_{(X,Y)}(x, y) = f_X(x)f_Y(y) \text{ } \lambda_2 \text{ p.p..}$$

alors X et Y sont indépendants. La preuve est immédiate (application de Fubini).

Remarque 9 (Important!) Si $X \perp\!\!\!\perp Y$, on a $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \times \mathbb{P}_Y$ (la mesure $\mathbb{P}_{(X,Y)}$ est la mesure produit $\mathbb{P}_X \times \mathbb{P}_Y$. Ceci est au coeur de la démonstration des propositions 9 et 10.

Deux exemples

- Supposons que (X, Y) ait pour densité

$$f_{(X,Y)}(x, y) = \frac{e^{-\frac{x^2}{2}-y}}{\sqrt{2\pi}} \mathbf{1}_{\mathbb{R}^+}(y)$$

on voit que

$$f_{(X,Y)}(x, y) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \times e^{-y} \mathbf{1}_{\mathbb{R}^+}(y)$$

donc si on prend $f_X(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ et $f_Y(y) = e^{-y} \mathbf{1}_{y \geq 0}$, $f_{(X,Y)} = f_X f_Y$. Il s'ensuit que X et Y sont indépendantes, X étant gaussienne $\mathcal{N}(0, 1)$ et Y exponentielle $\mathcal{E}(1)$.

- Supposons que

$$f_{(X,Y)}(x, y) = xe^{-xy} \mathbf{1}_{[0,1]}(x) \mathbf{1}_{\mathbb{R}^+}(y). \quad (20)$$

Cherchons les marginales :

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)dy = 1 \cdot \mathbf{1}_{[0,1]}(x), \quad (21)$$

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)dx = \frac{1 - ye^{-y} - e^{-y}}{y^2} \cdot \mathbf{1}_{\mathbb{R}^+}(y), \quad (22)$$

Remarquons que $f_{(X,Y)} \neq f_X f_Y$: les variables X et Y ne sont pas indépendantes !

Exercice 6 : 1) Soit $f_{(X,Y)}$ la densité d'un couple de v.a. (X, Y) (sur \mathbb{R}^2). Supposons qu'il existe une fonction positive g_1 et une fonction g_2 telles que

$$f_{(X,Y)}(x, y) = g_1(x)g_2(y).$$

Montrer qu'il existe une constante c tel que cg_1 soit une densité. Montrer que g_2/c est une densité; en déduire que les variables X et Y sont indépendantes.

2) On a admis rapidement, au vu de (20), (21) et (22), que $f_{(X,Y)} \neq f_X f_Y$. Or on doit prouver que $f_{(X,Y)} \neq f_X f_Y$ sur un non Lebesgue-négligeable. Le faire.

Proposition 10 Soient (X_1, \dots, X_n) une famille de v.a. indépendantes :

i) Toute sous famille est indépendante.

ii) Pour toutes fonctions boréliennes f_1, \dots, f_n de \mathbb{R} à valeurs dans \mathbb{R} , la famille $(f_1(X_1), \dots, f_n(X_n))$ est une famille de v.a. indépendantes.

iii) Soit $1 \leq k \leq n-1$. Soit f (resp. g) une fonction borélienne de \mathbb{R}^k (resp. \mathbb{R}^{n-k}) à valeurs dans \mathbb{R} , alors

$$f(X_1, \dots, X_k) \perp\!\!\!\perp g(X_{k+1}, \dots, X_n).$$

iv) Soit $1 \leq k \leq n-1$. Soit f (resp. g) une fonction borélienne de \mathbb{R}^k (resp. \mathbb{R}^{n-k}) à valeurs dans \mathbb{R} telles que

$$\mathbb{E}(|f(X_1, \dots, X_k)|) < +\infty \quad \text{et} \quad \mathbb{E}(|g(X_{k+1}, \dots, X_n)|) < +\infty$$

alors

$$\mathbb{E}(f(X_1, \dots, X_k)g(X_{k+1}, \dots, X_n)) = \mathbb{E}(f(X_1, \dots, X_k))\mathbb{E}(g(X_{k+1}, \dots, X_n))$$

Preuve : (i) Il suffit de prendre $A_i = \Omega$ dans (19) pour les variables manquantes. Pour prouver (ii) on écrit,

$$\begin{aligned} \mathbb{P}(f_1(X_1) \in A_1, \dots, f_n(X_n) \in A_n) &= \mathbb{P}(X_1 \in f_1^{-1}(A_1), \dots, X_n \in f_n^{-1}(A_n)) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \in f_i^{-1}(A_i)) = \prod_{i=1}^n \mathbb{P}(f_i(X_i) \in A_i) \end{aligned}$$

la deuxième égalité étant une conséquence de l'indépendance des X_i .

Pour (iii), on doit prouver que pour A et B boréliens de \mathbb{R} ,

$$\mathbb{P}(f(X_1, \dots, X_k) \in A, g(X_{k+1}, \dots, X_n) \in B) = \mathbb{P}(f(X_1, \dots, X_k) \in A)\mathbb{P}(g(X_{k+1}, \dots, X_n) \in B)$$

Ce qui se réécrit

$$\begin{aligned} &\mathbb{P}((X_1, \dots, X_k) \in f^{-1}(A), (X_{k+1}, \dots, X_n) \in g^{-1}(B)) \\ &= \mathbb{P}((X_1, \dots, X_k) \in f^{-1}(A))\mathbb{P}((X_{k+1}, \dots, X_n) \in g^{-1}(B)) \end{aligned} \quad (23)$$

Or, cette dernière égalité est une conséquence de l'indépendance des X_i dans le cas où $f^{-1}(A)$ et $g^{-1}(B)$ sont des pavés respectivement dans \mathbb{R}^k et dans \mathbb{R}^{n-k} . De cela, on déduit que la mesure $\mathbb{P}_{(Y_1, Y_2)}$ où $Y_1 = (X_1, \dots, X_k)$ et $Y_2 = (X_{k+1}, \dots, X_n)$ sur $\mathbb{R}^k \times \mathbb{R}^{n-k}$ est la mesure produit $\mathbb{P}_{Y_1} \times \mathbb{P}_{Y_2}$ (car c'est une mesure produit sur les pavés). Il s'ensuit que (23) est valable.

Pour (iv), on voit que l'on a en plus de (iii) (qui nous garantit que la mesure $\mathbb{P}_{(Y_1, Y_2)} = \mathbb{P}_{Y_1} \times \mathbb{P}_{Y_2}$) toutes les hypothèses pour utiliser le théorème de Fubini. \square .

10. Changement de variables

Soit (X, Y) un couple de variables aléatoires de densité jointe $f_{(X,Y)}$ et h un C_1 difféomorphisme entre deux ouverts de \mathbb{R}^2 (lorsque les domaines qui nous intéressent ne sont pas ouverts, on peut souvent, quitte à modifier les domaines sur un négligeable, se ramener à un ouvert).

Notons $(U, V) = h(X, Y)$. Notre but est d'expliquer comment on peut calculer la loi du couple (U, V) . L'idée est la même que dans le cas unidimensionnel. On donne un analogue de la proposition 6 que l'on admet :

Proposition 11 *Soit (X, Y) un couple de variables aléatoires réelles. Supposons qu'il existe une fonction f sur \mathbb{R}^2 , positive et mesurable, telle que pour toute fonction g de \mathbb{R}^2 à valeurs dans \mathbb{R} , continue bornée,*

$$\mathbb{E}(g(X, Y)) = \int_{\mathbb{R}^2} g(x, y) f(x, y) d\lambda_2(x, y) = \int \int g(x, y) f(x, y) dx dy, \quad (24)$$

alors f est une densité pour le couple (X, Y) .

Revenons à notre problème de la détermination de la loi de (U, V) . Pour k une fonction continue bornée de \mathbb{R}^2 dans \mathbb{R}^2 ,

$$\begin{aligned} \mathbb{E}(k(U, V)) &= \int k(u, v) d\mathbb{P}_{(U,V)}(u, v) \\ &= \mathbb{E}(k(h(X, Y))) = \int k(h(x, y)) f_{(X,Y)}(x, y) dx dy \end{aligned}$$

on pose $(u, v) = h(x, y)$ et on obtient

$$\mathbb{E}(k(h(X, Y))) = \int k(u, v) f_{(X,Y)}(h^{-1}(u, v)) |J(h^{-1}(u, v))| du dv$$

de sorte que

$$f_{(U,V)}(u, v) = f_{(X,Y)}(h^{-1}(u, v)) |J(h^{-1}(u, v))|$$

De nouveau, ceci est légal si la fonction h est un C^1 -difféomorphisme.

Exemple : Supposons que X et Y soient gaussiennes $\mathcal{N}(0, 1)$ indépendantes, de sorte que

$$f_{(X,Y)}(x, y) = \frac{\exp(-\frac{x^2+y^2}{2})}{2\pi}.$$

Soit $(U, V) = (X, XY)$; autrement dit, $h(X, Y) = (U(X, Y), V(X, Y))$ avec $U(X, Y) = X, V(X, Y) = XY$ et $h(x, y) = (x, xy)$. L'inverse de h est $h^{-1}(x, y) = (x, y/x)$. On a

$$\begin{aligned} f_{(U,V)}(u, v) &= f_{(X,Y)}(h^{-1}(u, v)) |J(h^{-1}(u, v))| \\ &= \frac{\exp(-\frac{u^2+(y/u)^2}{2})}{2\pi} \frac{1}{|u|} \end{aligned}$$

Si on le souhaite, on peut maintenant trouver la loi de V en intégrant selon u .

11. Simulations de variables aléatoires

De nombreux phénomènes aléatoires sont trop compliqués pour être étudiés frontalement. Souvent, on abandonne l'idée de pouvoir "évaluer" exactement la probabilité d'un événement trop complexe. On se contente de "simulations". Par exemple, si vous pensez avoir trouvé une stratégie pour gagner à un jeu (aléatoire), vous pouvez l'essayer, et la répéter des centaines de milliers de fois en quelques secondes sur un ordinateur. Vous possédez alors une base de données sur laquelle vous pouvez commencer à voir si votre stratégie marche sans avoir à risquer vos deniers (et même, comme on le verra plus tard, savoir dans quelle mesure on peut faire confiance à cet échantillon).

L'esprit dans lequel on fait les simulations est donc celui là : les résultats obtenus par simulation permettent souvent de conjecturer la "vérité", et ont une "valeur statistique", c'est-à-dire qu'ils sont suffisamment valables pour être utilisés dans toutes les sciences.

Les simulations de variables aléatoires sont notamment très utilisées en physique atomique, génétique, météorologie, études des phénomènes de circulation, files d'attente, mais aussi en mathématiques (pour calculer des intégrales ou chercher des solutions à des équations)...

Le point de départ : simulation de la variable uniforme $U \sim [0, 1]$.

La plupart des logiciels de programmation met à notre disposition une variable "uniforme" U sur $[0,1]$. (Appelée souvent par l'instruction "random" ou "ran").

Voici un exemple de générateur de "nombres aléatoires" : on pose

$$g(u) = 16807u \text{ modulo } 2147483647,$$

et u_0 un nombre dans $\{1, \dots, 2147483647\}$, et pour $n \geq 1$, $u_n = g(u_{n-1})$. Les variables $(U_n)_n$ successivement renvoyées par l'ordinateur sont :

$$U_n = u_n/2147483647.$$

Les limites de ce générateur sont claires : les nombres U_i ne sont pas aléatoires, ne sont pas indépendants, ne sont pas uniformes sur $[0,1]$ (seuls les nombres du type $a/2147483647$ avec a entiers sont atteints). Néanmoins, ce générateur est suffisant pour de nombreuses applications. Il existe maintenant des générateurs plus efficaces mais qui ont tous les mêmes défauts (à d'autres échelles).

Dans la suite, on suppose que l'on dispose d'une suite de v.a. indépendantes $(U_i)_{i \geq 1}$, uniformes sur $[0,1]$. Ce qui est remarquable, indépendamment des simulations, c'est qu'avec les variables U_i , on peut simuler toutes les lois uni et même multi-dimensionnelles (avec une seule variable U_i on peut même simuler toute une famille dénombrable de variables aléatoires dans \mathbb{R}).

Simulation de v.a. discrètes Pour simplifier (mais ça ne change absolument rien), supposons que la variable X à simuler prenne la valeur x_k avec probabilité p_k .

On tire une v.a. uniforme U :

Si $U \leq p_0$ on prend $X = x_0$

Si $p_0 < U \leq p_0 + p_1$ on prend $X = x_1$

Si $p_0 + \dots + p_k < U \leq p_0 + \dots + p_k + p_{k+1}$ on prend $X = x_{k+1}$ (pour $k \geq 0$). Il est très facile de remarquer qu'on a pris $X = x_{k+1}$ avec $\mathbb{P}(p_0 + \dots + p_k < U \leq p_0 + \dots + p_k + p_{k+1}) = p_{k+1}$ puisque U est uniforme.

Simulation de v.a. admettant une densité Si on note F_X la fonction de répartition de X alors $F_X(X)$ suit la loi uniforme. En effet, montrons que $F_X(X)$ (qui est une v.a.) a même fonction de répartition que U :

$$\mathbb{P}(F_X(X) \leq x) = \mathbb{P}(X \leq F_X^{-1}(x)) = F_X(F_X^{-1}(x)) = x,$$

où F_X^{-1} est la réciproque généralisée de F_X , c'est-à-dire :

$$F_X^{-1}(x) = \inf\{y \mid F_X(y) \geq x\}.$$

Lemme 3 $F_X^{-1}(U)$ a même distribution que X .

Preuve : Pour tout $x \in \mathbb{R}$, on a

$$\begin{aligned} \mathbb{P}(F_X^{-1}(U) \leq x) &= \mathbb{P}(\inf\{y, F(y) \geq U\} \leq x) \\ &= \mathbb{P}(U \leq F(x)) = F(x) \quad \square \end{aligned}$$

Ainsi, si on sait inverser la fonction de répartition d'une v.a., on sait la simuler.

12. La fonction caractéristique

Définition : La fonction caractéristique d'une variable aléatoire réelle X est la transformée de Fourier de sa mesure, à savoir :

$$\begin{aligned} \Phi_X : \mathbb{R} &\longrightarrow \mathbb{C} \\ t &\longmapsto \Phi_X(t) = \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x) \end{aligned} \quad (25)$$

La fonction $t \longmapsto \Phi_X(t)$ a de nombreuses propriétés dont voici les plus directes :

Proposition 12 – *i*) $t \mapsto \Phi_X(t)$ est définie pour tout $t \in \mathbb{R}$.
– *ii*) Φ_X dépend uniquement de la loi de X .
– *iii*) Φ_X est une fonction de module inférieur ou égal à 1, continue et $\Phi_X(0) = 1$.

Preuve : (ii) est évident. Montrons (i). Il suffit de voir pourquoi l'intégrale de (25) converge pour toute valeur de t . Or $|e^{itx}| \leq 1$. Donc $\int_{\mathbb{R}} |e^{itx}| d\mathbb{P}_X(x)$ converge et donc $\int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x)$ aussi. On a obtenu au passage que Φ_X était de module inférieur ou égal à 1 ; le fait que $\Phi_X(0) = 1$ est clair. La continuité provient des deux faits suivants :

- pour tout x , l'application $t \longmapsto e^{itx}$ est continue.
- Pour tout t , $x \longmapsto |e^{itx}|$ est dominée par la fonction constante à 1 qui est intégrable par rapport à \mathbb{P}_X . Le théorème de Lebesgue quant à la continuité des intégrales paramétrées permet de conclure. \square

Lemme 4 Soit a et b deux réels et X une variable aléatoire.

$$\forall t \in \mathbb{R}, \quad \Phi_{aX+b}(t) = e^{itb} \Phi_X(at).$$

Preuve :

$$\forall t \in \mathbb{R}, \quad \mathbb{E}(e^{it(aX+b)}) = e^{itb} \mathbb{E}(e^{itaX}). \quad \square$$

Proposition 13 Si X et Y sont deux variables réelles indépendantes, la fonction caractéristique de la somme est donnée par :

$$\forall t \in \mathbb{R}, \quad \Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t).$$

Preuve : On a besoin pour prouver cela d'une proposition similaire à la Proposition 3 (voir page 17) :

Proposition 14 Soient $(X_i)_{i=1,\dots,k}$ une famille de variables aléatoires réelles. Les deux assertions suivantes sont équivalentes :

- (i) les variables X_i sont indépendantes
- (ii) pour toutes fonctions f_i mesurables bornées,

$$\mathbb{E}\left(\prod_{i=1}^k f_i(X_i)\right) = \prod_{i=1}^k \mathbb{E}(f_i(X_i))$$

Admettons pour un instant la Proposition 14 et montrons la Proposition 13. La fonction $x \mapsto e^{itx}$ est mesurable et bornée. Donc, pour X et Y indépendantes,

$$\mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX} e^{itY}) = \mathbb{E}(e^{itX})\mathbb{E}(e^{itY}).$$

Ceci conclut la preuve de Proposition 13. \square

Preuve de la Proposition 14 : Supposons (ii) réalisée. Prenons $f_i = \mathbb{1}_{A_i}$ où A_i est un borélien. On a

$$\mathbb{E}\left(\prod_{i=1}^k f_i(X_i)\right) = \mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k)$$

et d'autre part

$$\prod_{i=1}^k \mathbb{E}(f_i(X_i)) = \prod_{i=1}^k \mathbb{P}(X_i \in A_i)$$

et ceci pour tout $(A_i)_{i \in \{1,\dots,k\}}$ boréliens. L'indépendance est établie.

Réciproquement, supposons les X_i indépendants. Alors la formule de (ii) est valable si les f_i sont des indicatrices de boréliens. Soit f_i une fonction mesurable positive ; il existe une suite croissante de fonctions étagées, positives, $f_i^{(n)}$ telle que $f_i^{(n)}$ converge simplement vers f_i . Par Beppo Levi $\mathbb{E}(f_i^{(n)}(X_i))$ converge vers $\mathbb{E}(f_i(X_i))$. De même, la fonction $(x_1, \dots, x_k) \mapsto \prod_{i=1}^k f_i^{(n)}(x_i)$ (vu comme fonction de k variables) est aussi mesurable et croissante en n . Par Beppo Levi $\mathbb{E}(\prod_{i=1}^k f_i^{(n)}(X_i))$ converge vers $\mathbb{E}(\prod_{i=1}^k f_i(X_i))$. D'autre part, par la linéarité de l'espérance (les f_i sont des combinaisons linéaires de fonctions indicatrices de boréliens),

$$\mathbb{E}\left(\prod_{i=1}^k f_i^{(n)}(X_i)\right) = \prod_{i=1}^k \mathbb{E}(f_i^{(n)}(X_i)).$$

Ainsi, par unicité de la limite,

$$\mathbb{E}\left(\prod_{i=1}^k f_i(X_i)\right) = \prod_{i=1}^k \mathbb{E}(f_i(X_i)).$$

On étend maintenant le résultat aux fonctions f_i signées comme habituellement en traitant à part f_{i+} et f_{i-} . \square

Note : Pour bien comprendre l'utilisation de Beppo Levi ici, il faut voir l'espérance mathématique comme une intégrale contre la mesure image. Autrement dit

$$\mathbb{E}(f(X)) = \int f(x)d\mathbb{P}_X(x).$$

Il est alors simple de voir que l'on peut appliquer les résultats d'intégration à l'espérance.

Exemples de fonctions caractéristiques

Tout ces calculs sont immédiats.

- Bernoulli. Si $X \sim B(p)$, $\Phi_X(t) = \mathbb{E}(e^{itX}) = pe^{it} + (1-p)$
- Binomiale. Si $X \sim B(n, p)$, $\mathbb{E}(e^{itX}) = (pe^{it} + (1-p))^n$ (c'est une conséquence de la Proposition 3).
- Poisson. Si $X \sim \mathcal{P}(\lambda)$, $\mathbb{E}(e^{itX}) = e^{-\lambda + \lambda e^{it}}$
- X Uniforme sur $[0,1]$. $\mathbb{E}(e^{itX}) = \frac{e^{it} - 1}{it}$
- X Uniforme sur $\{0, \dots, n\}$. $\mathbb{E}(e^{itX}) = \frac{1}{n+1} \frac{e^{it(n+1)} - 1}{e^{it} - 1}$
- Exponentielle $X \sim \mathcal{E}(1)$. $\mathbb{E}(e^{itX}) = (1 - it)^{-1}$
- Le cas de la loi normale est plus intéressant : Soit $X \sim \mathcal{N}(0, 1)$. On cherche à calculer

$$\Phi_X(t) = \int_{-\infty}^{+\infty} \frac{\exp(itx - x^2/2)}{\sqrt{2\pi}} dx.$$

On commence par remarquer que $t \mapsto \Phi_X(t)$ est holomorphe en t . Pour cela, on veut dériver en t sous le signe somme. Or $\frac{\partial}{\partial t} \frac{\exp(itx - x^2/2)}{\sqrt{2\pi}} = ix \frac{\exp(itx - x^2/2)}{\sqrt{2\pi}}$ est bien intégrable en x (et surtout, $x \mapsto |ix \frac{\exp(itx - x^2/2)}{\sqrt{2\pi}}|$ est dominée par $|x \exp(-x^2/2)|$ pour t dans la boule $C(0, c)$ de \mathbb{C} . Ainsi, un théorème de Lebesgue nous assure que Φ_X est dérivable. On prend alors t imaginaire pur $t = i\alpha$. On obtient

$$\Phi_X(i\alpha) = \int_{-\infty}^{+\infty} \frac{\exp(-\alpha x - x^2/2)}{\sqrt{2\pi}} dx = e^{\alpha^2/2} \int_{-\infty}^{+\infty} \frac{\exp(-\frac{1}{2}(x + \alpha)^2)}{\sqrt{2\pi}} dx = e^{\alpha^2/2}.$$

Cette dernière intégrale fait 1, puisqu'il s'agit de l'intégrale de la densité de la loi $\mathcal{N}(-\alpha, 1)$. On a donc établi, sur $i\mathbb{R}$, $\Phi_X(t) = e^{-t^2/2}$. Par la formule du prolongement analytique, cette formule est donc valable sur tout \mathbb{C} et donc en particulier sur \mathbb{R} . Donc

$$\forall t \in \mathbb{R}, \quad \Phi_X(t) = e^{-t^2/2}.$$

- Si $Y \sim \mathcal{N}(m, \sigma^2)$,

$$\Phi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it(\sigma X + m)}) = e^{itm} \Phi_X(t\sigma) = e^{itm} e^{-(\sigma t)^2/2}$$

Proposition 15 Soit X et X' deux variables réelles. Les deux assertions suivantes sont équivalentes :

(i) X et X' ont même loi.

(ii) Pour toute fonction f continue, à support compact, $\mathbb{E}(f(X)) = \mathbb{E}(f(X'))$.

Preuve : (i) \implies (ii) est clair. Passons à la preuve de (ii) \implies (i). Soit x fixé et g la fonction $g(t) = \mathbb{1}_{]-\infty, x]}(t)$, et g_n la fonction dessinée sur la figure ci-dessous. g_n converge simplement vers



g et $|g_n| \leq h \equiv 1$ qui est intégrable sous \mathbb{P}_{X_n} . Ainsi, d'après le théorème de convergence dominée de Lebesgue, $\mathbb{E}(g_n(X)) \longrightarrow \mathbb{E}(g(X)) = \mathbb{P}(X \leq x) = F_X(x)$. Le même calcul appliqué à X' donne $\mathbb{E}(g_n(X')) \longrightarrow F_{X'}(x)$. Puisque $\mathbb{E}(g_n(X)) = \mathbb{E}(g_n(X'))$ on en déduit que $F_{X'}(x) = F_X(x)$ pour tout x et donc que les variables X et X' ont la même loi. \square

Théorème 1 Soit X une variable réelle. La fonction caractéristique Φ_X caractérise la loi de X .

Preuve : Si deux variables X et X' ont même loi, il est clair que $\Phi_X = \Phi_{X'}$. Montrons la réciproque. On suppose maintenant que $\Phi_X(t) = \Phi_{X'}(t)$ pour tout t . Soit f continue à support compact, on va montrer que $\mathbb{E}(f(X)) = \mathbb{E}(f(X'))$ ce qui permettra de conclure d'après la Proposition 15. Pour tout $\varepsilon > 0$, il existe un compact $K = [-r, r]$ tel que

$$\mathbb{P}_X(K^c) \leq \varepsilon, \text{ et } \mathbb{P}_{X'}(K^c) \leq \varepsilon,$$

car \mathbb{P}_X et $\mathbb{P}_{X'}$ sont des mesures finies sur \mathbb{R} . Prenons r suffisamment grand tel que, de plus, $f(x) = 0$ sur K^c . Sur le compact K , l'algèbre A des fonctions engendrées par $t \mapsto e^{il\pi t/r}$ (pour t décrivant K et l entier) est dense dans l'ensemble des fonctions continues sur K pour la norme uniforme (d'après le théorème de Stone-Weierstrass). Ainsi, pour tout $\varepsilon > 0$, il existe un entier k et des constantes réelles $(\lambda_i)_{0 \leq i \leq k}$ et des constantes entières $(l_i)_{0 \leq i \leq k}$ telles que

$$\sup_{x \in K} |f(x) - \sum_{j=0}^k \lambda_j e^{il_j \pi x/r}| \leq \varepsilon.$$

Notons $\phi(x) = \sum_{j=0}^k \lambda_j e^{il_j \pi x/r}$. La fonction ϕ vue cette fois comme fonction sur \mathbb{R} , est $2r$ périodique. Ceci nous assure que $\|\phi\|_\infty$ est inférieure à $\|f\|_\infty + \varepsilon$ sur K^c (car c'est le cas sur la période $[-r, r]$). Ainsi, on a

$$\begin{aligned} |\mathbb{E}(f(X)) - \mathbb{E}(\phi(X))| &\leq \int_K |f(x) - \phi(x)| d\mathbb{P}_X(x) + \int_{K^c} |f(x) - \phi(x)| d\mathbb{P}_X(x) \\ &\leq \int_K |f(x) - \phi(x)| d\mathbb{P}_X(x) + \int_{K^c} |\phi(x)| d\mathbb{P}_X(x) \\ &\leq \varepsilon + (\|f\|_\infty + \varepsilon)\mathbb{P}(K^c) \leq \varepsilon + (\|f\|_\infty + \varepsilon)\varepsilon \end{aligned}$$

et de même, on a

$$|\mathbb{E}(f(X')) - \mathbb{E}(\phi(X'))| \leq \epsilon + (\|f\|_\infty + \epsilon)\epsilon.$$

On note maintenant que

$$\mathbb{E}(\phi(X)) = \sum_j \lambda_j \Phi_X(l_j \pi / r)$$

de sorte que $\mathbb{E}(\phi(X)) = \mathbb{E}(\phi(X'))$ car $\Phi_X = \Phi_{X'}$. Pour conclure, on écrit,

$$|\mathbb{E}(f(X)) - \mathbb{E}(f(X'))| \leq |\mathbb{E}(f(X)) - \mathbb{E}(\phi(X))| + |\mathbb{E}(\phi(X)) - \mathbb{E}(\phi(X'))| + |\mathbb{E}(\phi(X')) - \mathbb{E}(f(X'))|.$$

Les deux termes extrêmes dans le membre de droite sont inférieurs à $\epsilon + (\|f\|_\infty + \epsilon)\epsilon$, celui du centre vaut 0. Ainsi, $\mathbb{E}(f(X)) = \mathbb{E}(f(X'))$. \square

Proposition 16 *Supposons que X admette un moment d'ordre $n \in \mathbb{N}$. Alors Φ_X est de classe C^n et*

$$\mathbb{E}(X^n) = \frac{\Phi_X^{(n)}(0)}{i^n}$$

Preuve : Pour $n = 1$, on a $\frac{\partial}{\partial t} e^{itx} = ix e^{itx}$ existe pour tout t et de plus, $|ix e^{itx}| \leq |x|$ qui est intégrable par hypothèse contre \mathbb{P}_X . Donc, d'après le théorème de Lebesgue quant à la dérivation des intégrales à paramètres, Φ_X est dérivable et on a

$$\Phi_X'(t) = \int_{\mathbb{R}} ix e^{itx} d\mathbb{P}_X(x) = \mathbb{E}(iX e^{itX}),$$

dans laquelle il suffit de prendre $t = 0$. On itère le procédé; on obtient $(\frac{\partial}{\partial t})^n e^{itx} = (ix)^n e^{itx}$. Le reste de la preuve est identique au cas $n = 1$. \square

a. Convolution

Proposition 17 *Soit (X, Y) un couple de variable aléatoire réelle de densité jointe $f_{(X,Y)}$. La variable aléatoire $Z = X + Y$ admet pour densité*

$$f_Z(t) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x, t-x) dx.$$

Remarque : Si X et Y sont indépendantes

$$f_Z(t) = \int_{-\infty}^{+\infty} f_X(x) f_Y(t-x) dx = f_X * f_Y(t).$$

Preuve : On a

$$\begin{aligned} \Phi_{(X+Y)}(t) &= \mathbb{E}(e^{it(X+Y)}) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{it(x+y)} f_{(X,Y)}(x, y) dx dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{itu} f_{(X,Y)}(x, u-x) dx du \\ &= \int_{\mathbb{R}} e^{itu} \left(\int_{\mathbb{R}} f_{(X,Y)}(x, u-x) dx \right) du \end{aligned}$$

On reconnaît la fonction caractéristique d'une variable aléatoire de densité $u \mapsto \left(\int_{\mathbb{R}} f_{(X,Y)}(x, u-x) dx \right)$. Cette densité est donc la densité de $X + Y$ (d'après le Théorème 1). \square

Proposition 18 Soit $X = (X_1, \dots, X_n)$ une variable aléatoire à valeurs dans \mathbb{R}^n . Les variables X_i sont indépendantes si et seulement si, quel que soit $u = (u_1, \dots, u_n)$ de \mathbb{R}^n ,

$$\mathbb{E}(e^{i\langle u, X \rangle}) = \prod_{j=1}^n \Phi_{X_j}(u_j). \quad (26)$$

On pourra écrire le membre de gauche sous la forme $\Phi_X(u)$.

Preuve : Si les X_i sont indépendants, alors (26) est vraie d'après la Proposition 14. Réciproquement, si (26) est vraie, alors on peut établir en suivant le raisonnement utilisé dans la preuve du Théorème 1 que $\mathbb{E}(\prod_{i=1}^k f_i(X_i)) = \prod_{i=1}^k \mathbb{E}(f_i(X_i))$ pour f_i continue bornée, puis en suivant la preuve de la Proposition 15, on établit que $\mathbb{E}(\prod_{i=1}^k f_i(X_i)) = \prod_{i=1}^k \mathbb{E}(f_i(X_i))$ est valable pour les fonctions mesurables bornées, ce qui nous permet de prendre les indicatrices d'ensemble et on conclut par les arguments du début de la preuve de la Proposition 11. \square

b. Transformée de Laplace

Définition : Soit X une variable aléatoire réelle. On appelle transformée de Laplace de X (ou de la loi de X), la fonction Ψ définie par

$$\begin{aligned} \Psi_X : \mathbb{R} &\longrightarrow \mathbb{R} \\ t &\longmapsto \mathbb{E}(e^{-tX}) \end{aligned}$$

Remarque 10 - Parfois on appelle transformée de Laplace la fonction $t \mapsto \mathbb{E}(e^{tX})$ (bien sûr, ça ne change pas grand chose).

- Contrairement à la transformée de Fourier, la transformée de Laplace n'est en général pas définie pour tout t . Il existe des variables (Cauchy par exemple) pour laquelle Ψ ne converge qu'en 0. Si X est une variable positive, Ψ converge sur \mathbb{R}^+ au moins.

- La transformée de Laplace, si elle converge sur un intervalle différent de $\{0\}$, caractérise la loi.

IV. Théorèmes limites en probabilité

Les théorèmes limites en probabilité sont d'une importance cruciale ; ils décrivent de diverses manières les phénomènes de régularisation qui apparaissent lorsque l'on répète une expérience aléatoire. Par exemple, de la complète incertitude lors du jet d'une pièce sur le fait qu'elle va tomber sur pile ou face, on passe à la certitude quant à la fréquence asymptotique de "piles" obtenus lors d'une suite de lancers (indépendants, de même loi) : c'est la loi des grands nombres. Cette régularisation a permis aux physiciens de modéliser de nombreux phénomènes : par exemple, en mécanique statistique, on essaie de modéliser les mouvements de diverses particules comme les atomes dans un gaz, ou les divers types de particules prenant part aux réactions nucléaires. Ainsi, on peut expliquer le comportement macroscopique d'un système (souvent non aléatoire) par une "somme" de contributions aléatoires des particules.

1. Lemmes de Borel-Cantelli

Lemme 5 (Borel-Cantelli) : Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(A_n)_n$ une suite d'événements de \mathcal{A} telle que la somme $\sum_n \mathbb{P}(A_n)$ converge. Alors $\mathbb{P}(\limsup_n A_n) = 0$.

Preuve : On rappelle tout d'abord que $\limsup_n A_n$ est définie par

$$\limsup_n A_n = \bigcap_{n \geq 0} \bigcup_{k \geq n} A_k$$

et donc que c'est bien un élément de la tribu. Si $\sum_n \mathbb{P}(A_n)$ converge, $\forall \epsilon > 0$, il existe N t.q.

$\sum_{n \geq N} \mathbb{P}(A_n) \leq \epsilon$. Donc $\mathbb{P}(\bigcup_{k \geq N} A_k) < \epsilon$ et donc il en va de même pour $\limsup_n A_n$. \square

Comme habituellement, $\limsup_n A_n$ est l'ensemble des éléments de Ω présents dans une infinité de A_n . Le lemme de Borel Cantelli dit que si $\sum \mathbb{P}(A_n)$ converge, un nombre fini seulement de A_n sont réalisés avec probabilité 1.

Application : On lance successivement des dés justes. Le premier a une face, le deuxième 4, ... le n^{eme} , n^2 faces. Les faces sont numérotées de 1 à n^2 . (On ne suppose pas les lancers indépendants!!). Notons A_n l'événement : le n^{eme} dé tombe sur 1. On a $\mathbb{P}(A_n) = 1/n^2$ et donc $\sum_n \mathbb{P}(A_n)$ converge.

Le lemme de Borel Cantelli affirme que les dés tomberont sur 1 qu'un nombre fini de fois avec probabilité 1.

Le Lemme de Borel-Cantelli connaît une sorte de réciproque :

Lemme 6 (Borel-Cantelli 2) : Soit $(A_n)_n$ une suite d'événements indépendants de \mathcal{A} telle que la somme $\sum_n \mathbb{P}(A_n)$ diverge. Alors $\mathbb{P}(\limsup_n A_n) = 1$.

Preuve : Rappelons que $(A_n)_n$ est une suite d'événements indépendants, si toute famille finie extraite des A_i est une famille indépendante. On a

$$\mathbb{P}\left(\bigcup_{j=k}^m A_j\right) = 1 - \mathbb{P}\left(\bigcap_{j=k}^m \complement A_j\right) = 1 - \prod_{j=k}^m \mathbb{P}(\complement A_j) = 1 - \exp\left(\sum_{j=k}^m \log(1 - \mathbb{P}(A_j))\right)$$

Grâce à l'inégalité $\log(1-x) \leq -x$ (pour $x \in [0, 1[$) et en faisant tendre m vers $+\infty$, on voit que

$$-\exp\left(\sum_{j=k}^{+\infty} \log(1 - \mathbb{P}(A_j))\right) \geq -\exp\left(\sum_{j=k}^{+\infty} -\mathbb{P}(A_j)\right) = 0$$

et donc $\mathbb{P}\left(\bigcup_{j=k}^{+\infty} A_j\right) = 1$. Une intersection dénombrable d'événements de mesure 1 est également de mesure 1. \square

Application : On lance successivement des dés justes. Le premier a une face, le deuxième 2, ... le n^{eme} , n faces. Les faces sont numérotées de 1 à n . On suppose les lancers indépendants!! Notons A_n l'événement : le n^{eme} dé tombe sur 1. On a $\mathbb{P}(A_n) = 1/n$ et donc $\sum_n \mathbb{P}(A_n)$ diverge. Le lemme de Borel Cantelli 2 affirme que les dés tomberont sur 1 un nombre infini de fois avec probabilité 1.

Remarque 11 *L'hypothèse d'indépendance est ici nécessaire. Elle induit un "mélange" des événements. Sans cette hypothèse, il est aisé de trouver des contre-exemples en prenant par exemple tous les A_i égaux. Dans ce cas, $\limsup A_n = A_1$ et il n'y a pas de raison que cet événement soit de mesure 1.*

Les différents types de convergence

Il y a deux manières de converger pour une suite de variables aléatoires $(X_n)_n$:

- la première est la suivante : si les variables X_n sont définies sur le même espace Ω à valeurs dans \mathbb{R} (ou un espace métrique). On peut se demander, si pour ω fixé, $\lim_n X_n(\omega)$ existe.
 - Si la limite existe \mathbb{P} presque partout, on parlera de convergence presque sûre (il s'agit de la convergence simple sur une sous partie de Ω de masse 1).
 - S'il existe une variable X telle que, pour tout $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \rightarrow 0$, on parle de convergence en probabilité.
 - On dit qu'il y a convergence dans L^p si $E(|X_n - X|^p) = \int_{\Omega} |X_n(\omega) - X(\omega)|^p d\mathbb{P}(\omega) \rightarrow 0$
- le deuxième concerne une convergence plus faible ; les variables ne sont même plus supposées être définies sur le même espace. Il s'agit de la convergence en loi. Elle ne concerne plus directement les variables X_n mais leur loi image \mathbb{P}_{X_n} .

On précise tout cela ainsi que les rapports entre ces différents types de convergence dans la suite du chapitre.

2. Convergence en probabilité

Définition : Soit $(X, X_1, X_2, X_3, \dots)$ une suite de variables aléatoires définies sur le même espace $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que la suite $(X_n)_n$ converge vers X en probabilité si, pour tout $\epsilon > 0$,

$$\mathbb{P}(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0.$$

On note $X_n \xrightarrow{\text{proba}} X$.

Exemple : Soit $X_n = \sum_{i=1}^n Y_i$ où les Y_i sont indépendantes, de loi Bernoulli $B(1/i^2)$. La v.a. X_n est à valeurs entières et la suite $(X_n)_n$ est croissante : elle converge dans $\overline{\mathbb{R}}$. Soit X sa limite, on a, pour $0 < \epsilon < 1$,

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(\exists k, k \geq n+1, Y_k = 1) \leq \sum_{k=n+1}^{+\infty} 1/k^2;$$

puisque la série de terme général $1/k^2$ converge, le terme de droite dans la majoration tend vers 0 comme reste d'une série convergente. Ainsi X_n converge vers X en probabilité. On peut en déduire également que X est p.s. fini (en fait, X compte dans l'application de Borel-Cantelli (1), le nombre de dé tombant sur 1).

Remarque 12 On a vu plus haut et dans les exercices que deux variables peuvent avoir la même loi, même si elles sont définies sur des espaces de probabilité différents (penser au jeu de pile ou face qui peut être "réalisé" avec un dé par exemple). Pour que la suite $(X_n)_n$ converge en probabilité vers une variable limite X , il faut que ces variables soient définies sur le même espace.

a. Loi faible des grands nombres

Proposition 19 (loi faible des grands nombres) Soit $(X_i)_i$ une suite de variables aléatoires indépendantes, de même loi, et de moyenne $m < +\infty$. On note

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$$

la moyenne empirique des n premières valeurs des X_i . On a

$$\bar{X}_n \xrightarrow[n]{\text{proba}} m.$$

Tout d'abord, remarquons qu'on a ici fait un petit abus : on devrait écrire, $\bar{X}_n \xrightarrow[n]{\text{proba}} X$ où X est la v.a. définie par $X(\omega) = m$ pour tout $\omega \in \Omega$. On a identifié en fait la constante m avec la fonction constante (ou la v.a. constante) X .

Preuve : on donne une preuve de cette proposition sous l'hypothèse supplémentaire de l'existence d'un moment d'ordre 2, c'est-à-dire, on suppose que $\mathbb{E}(|X_1|^2) < +\infty$ converge. Ceci nous assure que $\text{Var}(X_1) = \sigma^2 < +\infty$. D'après Bienaymé-Tchebichev, pour tout $\epsilon > 0$, (et puisque $\mathbb{E}(\bar{X}_n) = m$),

$$\mathbb{P}(|\bar{X}_n - m| \geq \epsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\epsilon^2} = \frac{n\sigma^2}{n^2\epsilon^2}$$

ce qui tend bien vers 0 lorsque n tend vers $+\infty$. \square

3. Convergence presque sûre

Définition : Soit $(X_n)_n$ une suite de v.a. définie sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que la suite $(X_n)_n$ converge presque sûrement si

$$\mathbb{P}(\lim_n X_n(\omega) \text{ existe}) = 1.$$

Notons $X(\omega)$ la limite de $X_n(\omega)$ lorsque celle-ci existe. La limite peut ne pas exister sur une partie de Ω négligeable. On note

$$X_n \xrightarrow[n]{p.s.} X.$$

Théorème 2 (*loi forte des grands nombres*) : Soit $(X_i)_i$ une suite de v.a. de même loi, de moyenne m alors

$$\bar{X}_n \xrightarrow{p.s.} m.$$

Preuve : La conclusion de ce théorème est valable sous la seule hypothèse de l'existence de la moyenne pour X_1 . La preuve que l'on donne ici utilise l'hypothèse supplémentaire de l'existence de moment d'ordre 4. On suppose donc dans ce qui suit que

$$\mathbb{E}(X_1^4) < +\infty.$$

On suppose aussi que $\mathbb{E}(X_1) = 0$, ce que l'on peut faire sans restreindre la généralité de la preuve (on pourrait aussi poser $Y_i = X_i - m$, prouver le résultat pour \bar{Y}_n et le retraduire en terme de \bar{X}_n).

On a

$$\mathbb{P}(|\bar{X}_n| \geq \epsilon) = \mathbb{P}(|\bar{X}_n|^4 \geq \epsilon^4) \stackrel{Markov}{\leq} \frac{1}{n^4 \epsilon^4} \mathbb{E}((X_1 + \dots + X_n)^4).$$

En développant la dernière parenthèse, et comme les variables X_i sont indépendantes et centrées et ont même loi, on obtient

$$\mathbb{E}((X_1 + \dots + X_n)^4) = n\mathbb{E}(X_1^4) + C_n^2 \mathbb{E}(X_1^2) \mathbb{E}(X_2^2)$$

puisque tout terme contenant un X_i à la puissance 1 est nul. On peut aussi écrire $\mathbb{E}(X_1^2) \mathbb{E}(X_2^2)$ sous la forme $\mathbb{E}(X_1^2)^2$. En tout cas, on a

$$\mathbb{P}(|\bar{X}_n| \geq \epsilon) = O(1/n^2).$$

D'après Borel-Cantelli, si on note par $A_n^\epsilon = \{\omega, |\bar{X}_n(\omega)| \geq \epsilon\}$, on voit que puisque $\sum \mathbb{P}(A_n^\epsilon)$ converge, $\mathbb{P}(\limsup A_n^\epsilon) = 0$. On a donc pour tout $k \in \mathbb{N}^*$, $\mathbb{P}(\limsup A_n^{1/k}) = 0$. Ainsi, $\mathbb{P}(\cup_k \limsup A_n^{1/k}) = 0$. Or

$$\{\omega, \bar{X}_n(\omega) \not\rightarrow 0\} = \{\omega, \omega \in \cup_k \limsup A_n^{1/k}\}.$$

Il s'ensuit que

$$\mathbb{P}(\{\omega, \bar{X}_n(\omega) \rightarrow 0\}) = 1. \quad \square$$

Remarque 13 On peut montrer que si $\mathbb{E}(|X_1|) = +\infty$ alors la suite \bar{X}_n ne converge pas.

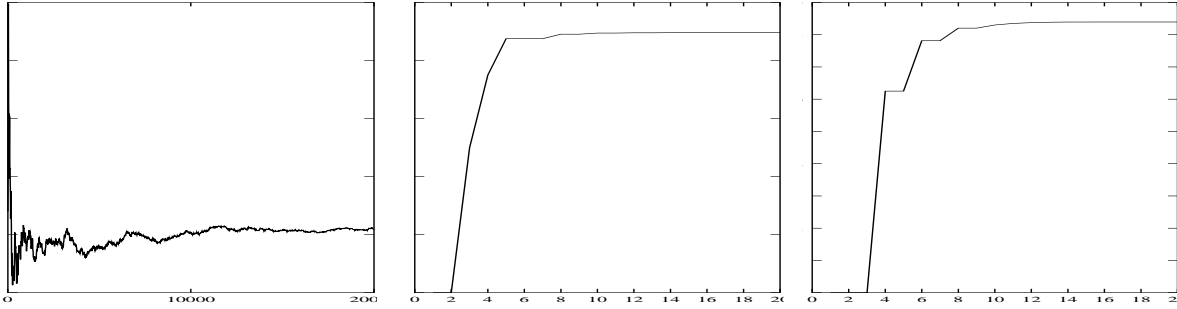
Le théorème 2 est relativement intuitif. Il affirme que la moyenne empirique (observée par l'expérience) converge vers la moyenne (espérance mathématique). Si on l'applique à une suite de v.a. de Bernoulli indépendantes $B(p)$. On obtient

$$\bar{X}_n \xrightarrow{p.s.} p.$$

Ainsi, c'est la loi des grands nombres qui affirme que la proportion de *pile* dans une suite de *pile* ou *face* tend vers p . Remarquons une dernière fois que ce résultat est la conséquence d'un théorème et non pas la définition de la probabilité.

Deux exemples : Dans le premier exemple (premier dessin ci-dessous), on illustre la loi des grands nombres par une simulation. Sur le graphique, on représente \bar{X}_n en fonction de n où les variables X_i sont indépendantes et sont de loi $\mathcal{N}(0, 1)$. La loi des grands nombres nous assure que la suite \bar{X}_n converge p.s. vers 0 (par ailleurs on peut facilement montrer que \bar{X}_n est de loi $\mathcal{N}(0, 1/n)$).

Dans le deuxième exemple, on prend $X_n = \sum_{i=1}^n 2^{-i} b_i$ où la suite X_n est construite à l'aide d'une suite de variables b_i indépendantes, de loi Bernoulli 1/2. Deux simulations de la suite $(X_n)_n$ sont présentées (voir page suivante, 2ème et 3ème figure). On peut voir que la limite n'est pas la même sur ces deux simulations.



4. Convergence dans L^p

Définition : Soit X, X_1, X_2, \dots des variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$; on dit que X_n converge vers X dans L^p et on note

$$X_n \xrightarrow[n]{L^p} X$$

si $X \in L^p$ (c'est-à-dire si $\mathbb{E}(|X|^p) < +\infty$) et si

$$\mathbb{E}(|X_n - X|^p) \longrightarrow 0.$$

5. Convergence en loi. Caractérisations

Définition : On considère des variables aléatoires $(X_n)_{n \geq 0}$ et X , à valeurs dans \mathbb{R} (ou \mathbb{R}^d) non nécessairement définies sur le même espace de probabilité. On dit que la suite $(X_n)_n$ converge en loi vers X si pour toute fonction f continue bornée sur \mathbb{R} (ou \mathbb{R}^d),

$$\mathbb{E}(f(X_n)) \xrightarrow[n]{} \mathbb{E}(f(X)); \quad (27)$$

on note

$$X_n \xrightarrow[n]{loi} X.$$

Remarquer que comme les variables X_n peuvent être définies sur des espaces différents, la quantité $\mathbb{E}(f(X_n) - f(X))$ n'est pas définie a priori : on a besoin de \mathbb{P} pour définir l'espérance, et le \mathbb{P} permettant de calculer $\mathbb{E}(f(X_n))$ et celui permettant de calculer $\mathbb{E}(f(X))$ ne sont pas les mêmes. Ainsi, dans le cas où les espaces sont différents, la définition alternative suivante, met plus en exergue les objets :

Définition : On considère des variables aléatoires X, X_1, X_2, \dots , définies sur des espaces de probabilités $(\Omega, \mathcal{A}, \mathbb{P}), (\Omega_1, \mathcal{A}_1, \mathbb{P}^{(1)}), (\Omega_2, \mathcal{A}_2, \mathbb{P}^{(2)}), \dots$, à valeurs dans \mathbb{R} (ou \mathbb{R}^d). On note \mathbb{E}_k l'espérance sur l'espace $(\Omega_k, \mathcal{A}_k, \mathbb{P}^{(k)})$.

On dit que la suite $(X_n)_n$ converge en loi vers X si pour toute fonction f continue bornée sur \mathbb{R} (ou \mathbb{R}^d),

$$\mathbb{E}_n(f(X_n)) \xrightarrow[n]{} \mathbb{E}(f(X)); \quad (28)$$

Cette convergence est très différente de celles vues jusqu'ici. Dans les convergences vues plus haut (proba, p.s., L^p), les variables aléatoires sont vues comme des fonctions de Ω dans \mathbb{R} . Ainsi,

la convergence p.s. est vraiment une convergence ponctuelle (\mathbb{P} p.p.) sur Ω , c'est-à-dire, \mathbb{P} presque sûrement, $(X_n(\omega))_n$ converge.

Dans la convergence en loi, ce n'est plus du tout le cas. Les variables ne sont plus définies (a priori) sur le même espace : la variable X_n est définie sur un espace Ω_n muni d'une mesure \mathbb{P}_n . La convergence ponctuelle n'a plus de sens donc. La convergence en loi ne décrit pas du tout la convergence de la suite X_n ; elle décrit la convergence de la suite de mesures \mathbb{P}_{X_n} (la convergence des lois).

Remarque 14 *La convergence en loi de X_n vers X n'implique pas la convergence de $\mathbb{E}(X_n)$ vers $\mathbb{E}(X)$ (car la fonction $x \mapsto x$ n'est pas bornée) ; elle n'implique pas non plus la convergence d'aucun moment de X_n vers ceux de X . Le fait que f soit bornée (et mesurable puisque continue) implique que $\mathbb{E}(f(X_n))$ aussi bien que $\mathbb{E}(f(X))$ sont bien définies et finies.*

Exemples :

(a) Soit X_i une suite de variables indépendantes de loi $\mathcal{N}(0, 1)$. Cette suite converge en loi vers une variable de loi $\mathcal{N}(0, 1)$; il n'y a évidemment pas dans ce cas de convergence de la suite numérique $(X_i)_i$ (en effet, en utilisant le lemme de Borel-Cantelli (2), on peut aisément montrer, qu'avec probabilité 1, un nombre infini de $X_i(\omega)$ sont supérieures à 1, mais aussi, un nombre infini de $X_i(\omega)$ sont inférieures à -1 . Une telle suite ne converge pas.)

(b) Si X_i est une suite de variables aléatoires de loi de Poisson $\mathcal{P}(1/i)$. La suite $(X_i)_i$ converge en loi vers la variable X dégénérée en 0 (c'est-à-dire la variable de loi image $\delta_0 : \mathbb{P}(X = 0) = 1$). En effet

$$\mathbb{E}(f(X_i)) - \mathbb{E}(f(X)) = \sum_{k=0}^{+\infty} \left(f(k) \frac{e^{-1/i}}{k!i^k} \right) - f(0) = f(0)(e^{-1/i} - 1) + \sum_{k=1}^{+\infty} \left(f(k) \frac{e^{-1/i}}{k!i^k} \right).$$

Le membre de droite tend vers 0 lorsque i tend vers $+\infty$: c'est clair pour $f(0)(e^{-1/i} - 1)$; pour $\sum_{k=1}^{+\infty} \left(f(k) \frac{e^{-1/i}}{k!i^k} \right)$ on utilise le fait que chaque terme tend vers 0 (lorsque $i \rightarrow +\infty$) et le théorème de

convergence dominé : $|f(k)| \frac{e^{-1/i}}{k!i^k} \leq \frac{|f(k)|}{k!}$ qui est le terme général d'une série convergente puisque f est bornée.

Proposition 20 *Si $X_n \xrightarrow[n]{loi} X$ en loi alors, pour toute fonction f continue, $f(X_n) \xrightarrow[n]{loi} f(X)$.*

Preuve : On doit prouver que pour g continue bornée, $\mathbb{E}(g(f(X_n))) \rightarrow \mathbb{E}(g(f(X)))$. Or, comme $g \circ f$ est continue bornée, la convergence $X_n \xrightarrow[n]{loi} X$ entraîne que $\mathbb{E}(g(f(X_n))) \rightarrow \mathbb{E}(g(f(X)))$. \square

Remarque 15 *(Convergence en loi pour des couples de v.a.) La définition 5. dans le cadre de la convergence en loi des v.a. bidimensionnelles peut être reformulée comme suit : Soit $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$ des couples de variables aléatoires (à valeurs réelles). On dit que la suite (X_n, Y_n) converge en loi vers (X, Y) , si, pour toute fonction H , continue bornée de \mathbb{R}^2 dans \mathbb{R} , $\mathbb{E}(H(X_n, Y_n)) \rightarrow \mathbb{E}(H(X, Y))$.*

Proposition 21 *Si la suite (X_n, Y_n) converge en loi vers (X, Y) , pour toutes fonctions continues f et g (de \mathbb{R} dans \mathbb{R}) $(f(X_n), g(Y_n)) \xrightarrow[n]{loi} (f(X), g(Y))$. De plus si X_n et Y_n sont indépendants, alors X et Y aussi.*

Preuve : On doit prouver que pour H continue borné de \mathbb{R}^2 dans \mathbb{R} , $\mathbb{E}(H(f(X_n), g(Y_n))) \longrightarrow \mathbb{E}(H(f(X), g(Y)))$. Soit L l'application de \mathbb{R}^2 dans \mathbb{R} définie par $L(x, y) = (f(x), g(y))$. L'application L est continue. Donc $H \circ L$ est continue bornée. Il suit que $\mathbb{E}(H(L(X_n, Y_n)))$ tend vers $\mathbb{E}(H(L(X, Y)))$; ainsi $(f(X_n), g(Y_n)) \xrightarrow[n]{loi} (f(X), g(Y))$.

Montrons maintenant que si la suite $(X_n, Y_n) \xrightarrow[n]{loi} (X, Y)$, et si, de plus, X_n et Y_n sont indépendants, alors X et Y aussi. Pour f et g continues bornées, l'indépendance de X_n et Y_n donne

$$\mathbb{E}(f(X_n)g(Y_n)) = \mathbb{E}(f(X_n))\mathbb{E}(g(Y_n)). \quad (29)$$

Maintenant, on a, puisque l'application $(x, y) \longrightarrow f(x)g(y)$ est continue bornée, puisque $(X_n, Y_n) \xrightarrow[n]{loi} (X, Y)$,

$$\mathbb{E}(f(X_n)g(Y_n)) \longrightarrow \mathbb{E}(f(X)g(Y)). \quad (30)$$

D'autre part, si $(X_n, Y_n) \xrightarrow[n]{loi} (X, Y)$, alors $X_n \xrightarrow[n]{loi} X$ (et $Y_n \xrightarrow[n]{loi} Y$). En effet, il suffit de prendre $g \equiv 1$ dans (30) pour s'en convaincre. Ainsi $\mathbb{E}(f(X_n))\mathbb{E}(g(Y_n)) \longrightarrow \mathbb{E}(f(X))\mathbb{E}(g(Y))$; par unicité de la limite dans (29), on a

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

Ceci implique que la mesure $\mathbb{P}_{X,Y}$ est la mesure produit $\mathbb{P}_X \times \mathbb{P}_Y$, c'est-à-dire que X et Y sont indépendants; en effet, on a obtenu

$$\int \int f(x)g(y)d\mathbb{P}_{X,Y}(x, y) = \int f(x)d\mathbb{P}_X(x) \int g(y)d\mathbb{P}_Y(y)$$

pour toute fonction continue bornée. On peut maintenant, montrer que ceci est vrai pour $f = \mathbb{1}_A$ et $g = \mathbb{1}_B$ pour A et B boréliens par densité. On obtient alors que

$$\int \int f(x)g(y)d\mathbb{P}_{X,Y}(x, y) = \mathbb{P}((X, Y) \in A \times B) = \int f(x)d\mathbb{P}_X(x) \int g(y)d\mathbb{P}_Y(y) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B). \quad \square$$

6. Critères généraux de convergence en loi

Proposition 22 *Soient $(X_n)_n$ et X des variables aléatoires réelles de fonctions de répartition F_{X_n} et F . Les deux propositions suivantes sont équivalentes :*

- (i) $X_n \xrightarrow[n]{loi} X$
- (ii) $F_{X_n}(x) \xrightarrow[n]{} F(x)$ pour tout x , point de continuité de F .

Il s'agit de la convergence simple de la suite de fonctions F_{X_n} vers F sur les points de continuité de F ; puisque F est croissante, l'ensemble des points de discontinuité de F est au plus dénombrable. Si a et b sont deux points de continuité de F , on a

$$\mathbb{P}(X_n \in]a, b]) = F_{X_n}(b) - F_{X_n}(a) \longrightarrow F(b) - F(a) = \mathbb{P}(X \in]a, b]). \quad (31)$$

Si a et b sont des points de continuité de F ceci signifie que X ne charge ni a ni b , sans quoi F présenterait un saut en ces points. On voit bien sur (31) la signification de la convergence en loi : la probabilité de présence de X_n dans un ensemble (non chargé au bord par X) converge vers la probabilité que X soit dans cet ensemble.

Preuve : Montrons $(ii) \implies (i)$. Prenons une fonction g de classe C_1 à support compact K . Sa dérivée g' est bornée à support compact K ; on a

$$\mathbb{E}(g(X_n)) = \int_{\mathbb{R}} g(x) d\mathbb{P}_{X_n}(x) = \int_{\mathbb{R}} \left(\int_{-\infty}^x g'(t) dt \right) d\mathbb{P}_{X_n}(x) = \int_{\mathbb{R}} (1 - F_{X_n}(x)) g'(x) dx;$$

(en effet $1 - F_{X_n}(x) = \int_x^{+\infty} d\mathbb{P}_{X_n}(t)$). Comme la mesure $\mathbf{1}_K d\lambda$ est finie, et comme $(1 - F_{X_n})g'$ converge simplement vers $(1 - F_X)g'$ sauf sur un ensemble négligeable, le théorème de convergence dominée de Lebesgue ($|(1 - F_{X_n})g'| \leq \|g'\|_{\infty}$) implique que

$$\lim_n \mathbb{E}(g(X_n)) = \mathbb{E}(g(X)).$$

Maintenant, comme les fonctions C_1 à support compact sont denses parmi les fonctions continues à support compact (pour la norme uniforme) on peut étendre le résultat aux fonctions continues à support compact. Prenons maintenant une fonction f continue et bornée par 1 ; montrons que

$$\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)).$$

par linéarité de l'espérance, ceci implique que ceci est aussi vrai pour toute fonction continue bornée, et donc, par définition que $X_n \xrightarrow{\text{loi}} X$.

Pour tout $\epsilon > 0$, il existe un intervalle compact $A = [-M, M]$ tel que, pour n assez grand, $\mathbb{P}(X_n \in A^C) < \epsilon$ (ceci provient du fait que la suite $F_{X_n}(M) - F_{X_n}(-M)$ converge pour tous points M et $-M$ de continuité de F_X) ; on a de plus pour le choix d'un M adéquat, $\mathbb{P}(X \in A^C) \leq \epsilon$.

On approche maintenant f sur A par une fonction g_{ϵ} à support compact comme suit : g_{ϵ} coïncide avec f sur A ; sur $[M, M + \epsilon]$, g_{ϵ} est linéaire par morceaux et vaut 0 en $M + \epsilon$ (même construction au voisinage de $-M$). g_{ϵ} est nulle ailleurs.

On a alors,

$$\begin{aligned} |\mathbb{E}(f(X_n)) - \mathbb{E}(g_{\epsilon}(X_n))| &\leq \mathbb{E}(|f(X_n) - g_{\epsilon}(X_n)|) = \mathbb{E}(|f(X_n) - g_{\epsilon}(X_n)| \mathbf{1}_{A^C}(X_n)) \\ &\leq 2\|f\|_{\infty} \mathbb{P}(X_n \in A^C) \leq 2\epsilon. \end{aligned}$$

et la même formule est valable en remplaçant partout X_n par X . Or, $\mathbb{E}(g_{\epsilon}(X_n))$ converge vers $\mathbb{E}(g_{\epsilon}(X))$ puisque g_{ϵ} est continue à support compact (et donc, pour n assez grand,

$$|\mathbb{E}(g_{\epsilon}(X_n)) - \mathbb{E}(g_{\epsilon}(X))| \leq \epsilon.$$

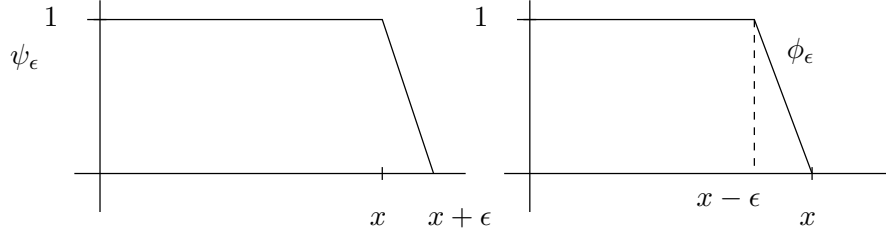
Ainsi, pour n assez grand, on a $|\mathbb{E}(f(X_n)) - \mathbb{E}(f(X))|$

$$\begin{aligned} &\leq |\mathbb{E}(f(X_n)) - \mathbb{E}(g_{\epsilon}(X_n))| + |\mathbb{E}(g_{\epsilon}(X_n)) - \mathbb{E}(g_{\epsilon}(X))| + |\mathbb{E}(f(X)) - \mathbb{E}(g_{\epsilon}(X))| \\ &\leq 2\epsilon + \epsilon + 2\epsilon; \end{aligned}$$

ceci prouve que $\mathbb{E}(f(X_n))$ converge vers $\mathbb{E}(f(X))$.

Montrons maintenant $(i) \implies (ii)$. Soit x point de continuité de F . Soient ϕ_{ϵ} et ψ_{ϵ} à valeurs dans $[0, 1]$ les fonctions continues ci-dessous : on a, pour tout ϵ , par la convergence en loi,

$$\lim_n \mathbb{E}(\phi_{\epsilon}(X_n)) = \mathbb{E}(\phi_{\epsilon}(X)) \text{ et } \lim_n \mathbb{E}(\psi_{\epsilon}(X_n)) = \mathbb{E}(\psi_{\epsilon}(X)). \quad (32)$$



On a, pour tout n ,

$$\mathbb{E}(\phi_\epsilon(X_n)) \leq F_{X_n}(x) \leq \mathbb{E}(\psi_\epsilon(X_n)). \quad (33)$$

Compte tenu que

$$F(x - \epsilon) \leq \mathbb{E}(\phi_\epsilon(X)) \leq \mathbb{E}(\psi_\epsilon(X)) \leq F(x + \epsilon) \quad (34)$$

et que F est continue en x , en faisant tendre ϵ vers 0 dans (34) on voit que

$$\lim_{\epsilon \rightarrow 0} \mathbb{E}(\phi_\epsilon(X)) = \lim_{\epsilon \rightarrow 0} \mathbb{E}(\psi_\epsilon(X)) = F(x). \quad (35)$$

Combinant (32) et (33), on a

$$\mathbb{E}(\phi_\epsilon(X)) \leq \liminf_n F_{X_n}(x) \leq \limsup_n F_{X_n}(x) \leq \mathbb{E}(\psi_\epsilon(X)).$$

On fait tendre ϵ vers 0 et (35) permet alors de conclure. \square

Convergence en loi pour des variables aléatoires sur \mathbb{N}

Proposition 23 Soient X_1, X_2, \dots et X des variables aléatoires à valeurs dans \mathbb{N} . Pour que $X_n \xrightarrow{\text{loi}} X$ il faut et il suffit que pour tout $k \in \mathbb{N}$

$$\mathbb{P}(X_n = k) \xrightarrow{n} \mathbb{P}(X = k). \quad (36)$$

Remarque 16 On suppose que la limite est une loi de probabilité ; ceci n'est pas assuré par la seule supposition que $(\mathbb{P}(X_n = k))_n$ converge.

Preuve : Si (36) a lieu, alors la $F_{X_n}(k)$ converge vers $F_X(k)$ pour tout k entier ($F_{X_n}(k) = \mathbb{P}(X_n = 0) + \dots + \mathbb{P}(X_n = k)$ et la limite d'une somme finie et la somme des limites (lorsqu'elles existent, et c'est le cas ici). Soit x réel, $F_{X_n}(x) = F_{X_n}([x])$ converge vers $F_X(x) = F_X([x])$; ainsi, sous l'hypothèse (36), F_{X_n} converge simplement vers F_X ; on conclut grâce à la proposition 22. Réciproquement, supposons que F_{X_n} converge ponctuellement vers F_X ; comme $F_{X_n}(x)$ est constante entre les entiers, F_X aussi. D'autre part, par le même argument que si dessus, on voit facilement, par itération que la convergence de $F_{X_n}(k)$ entraîne la convergence des $\mathbb{P}(X_n = j)$ (prendre d'abord $k = 0$ puis $k = 1$ et établir les convergences des $\mathbb{P}(X_n = j)$ successivement). \square

Proposition 24 (i) Si $(X_n)_n$ est une suite de variables aléatoires à valeurs dans \mathbb{N} convergeant en loi vers X alors G_{X_n} converge simplement vers G_X sur $[0, 1]$.

(ii) Si la suite des fonctions génératrices $(G_{X_n})_n$ converge simplement sur $[0, 1]$ vers une fonction G_X série génératrice d'une variable X , alors X_n converge en loi vers X .

Preuve : (i) Notons $p_k^{(n)} = \mathbb{P}(X_n = k)$ et $p_k = \mathbb{P}(X = k)$. On va utiliser la proposition 23 et le théorème de convergence dominée (avec la mesure de comptage). Tout d'abord remarquons que la convergence au point 1 est évidente (car $G_{X_n}(1) = G_X(1) = 1$). Soit $x \in [0, 1[$ fixé. Pour tout k , on a les deux points suivants : $|p_k^{(n)} x^k| \leq x^k$ et $p_k^{(n)} x^k \rightarrow p_k x^k$ lorsque $n \rightarrow +\infty$.

Si M désigne la mesure de comptage (c'est-à-dire, $M = \sum_{k \geq 0} \delta_k$)

$$\sum_{k \geq 0} x^k = \int_{k \in \mathbb{N}} x^k dM(k) = \frac{1}{1-x} < +\infty.$$

on a, d'après le théorème de Lebesgue

$$G_{X_n}(x) = \sum_{k \geq 0} p_k^{(n)} x^k = \int_{k \in \mathbb{N}} p_k^{(n)} x^k dM(k) \rightarrow \int_{k \in \mathbb{N}} p_k x^k dM(k) = \sum_{k \geq 0} p_k x^k = G_X(x).$$

(ii) La preuve de (ii) se fait par induction. Tout d'abord, on voit que $p_0^{(n)}$ converge vers p_0 (en prenant $x = 0$). On note $\mathcal{P}(k)$ la propriété “ $(p_i^{(n)})$ converge vers p_i pour tout $i \leq k$ ”, de sorte que l'on sait déjà que $\mathcal{P}(0)$ est vraie. Pour $k \geq 0$, montrons que $\mathcal{P}(k)$ vraie implique $\mathcal{P}(k+1)$ vraie. Supposons que $\mathcal{P}(k)$ soit vraie. Pour $x \in]0, 1[$, on a alors

$$\left| \frac{G_{X_n}(x) - \sum_{i=0}^k p_i^{(n)} x^i}{x^{k+1}} - \frac{G_X(x) - \sum_{i=0}^k p_i x^i}{x^{k+1}} \right| \xrightarrow{n} 0.$$

Le terme de gauche se réécrit $g_n(x) = |p_{k+1}^{(n)} - p_{k+1} + \sum_{i \geq k+2} (p_i^{(n)} - p_i) x^{i-(k+1)}|$. La valeur absolue de la somme est bornée par $x/(1-x)$. La convergence de $g_n(x)$ vers 0 implique que $p_{k+1}^{(n)}$ converge vers p_{k+1} (en effet, car on doit avoir $\limsup |p_{k+1}^{(n)} - p_{k+1}| \leq x/(1-x)$). Ainsi $\mathcal{P}(k+1)$ est vraie, et par récurrence on voit que $\mathcal{P}(k)$ est vraie pour tout k . Il s'ensuit que (ii) est vraie. \square

Théorème 3 (Lévy) : Soit $(X_n)_n$ une suite de variables aléatoire à valeurs dans \mathbb{R} . On a :

i) Si $X_n \xrightarrow{\text{loi}} X$ alors Φ_{X_n} converge simplement vers Φ_X .

ii) Si la suite fonction Φ_{X_n} converge simplement vers une fonction Φ continue en 0, alors Φ est la fonction caractéristique d'une variable aléatoire X ; de plus $X_n \xrightarrow{\text{loi}} X$

Preuve : (i) \implies (ii) est une conséquence du fait que pour tout $t, x \mapsto e^{itx}$ est continue et bornée. (ii) \implies (i) admis (une preuve classique utilise la transformée de Fourier inverse et se trouve dans de nombreux ouvrages).

7. Théorème de la limite centrale

Théorème 4 Soit $(X_n)_n$ une suite de variables aléatoires réelles, indépendantes et de même loi, de moyenne m et de variance finie et non nulle σ^2 . Pour tout n , on note

$$S_n = \sum_{i=1}^n X_i.$$

On a

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

Remarque 17 *L'indépendance des variables est primordiale. On peut s'en rendre compte en prenant un cas extrême : si on prend $X_1 = X_2 = \dots$, (avec moyenne 0 et variance 1, par exemple) alors $\frac{S_n - nm}{\sqrt{n}} = \sqrt{n}X_1$, et ceci ne converge pas vers une $\mathcal{N}(0, 1)$.*

Preuve du théorème de la limite centrale :

Quitte à remplacer les X_i par $X_i - \mathbb{E}(X_i)$ on peut supposer que $m = 0$. Puisque la fonction caractéristique d'une somme de v.a. indépendantes est le produit des fonctions caractéristiques de ces variables, on a :

$$\Phi_{S_n/\sigma\sqrt{n}}(t) = \mathbb{E}(e^{it\frac{S_n}{\sigma\sqrt{n}}}) = \Phi_X\left(\frac{t}{\sigma\sqrt{n}}\right)^n.$$

Le fait que X ait une variance nous assure que Φ_X est de classe \mathcal{C}^2 . Φ admet donc un développement de Taylor d'ordre 2, en 0 :

$$\Phi_X(t) = 1 + t\Phi'_X(0) + \frac{t^2}{2}\Phi''_X(0) + o(t^2).$$

$m = 0$ et $\text{Var}(X) = \sigma^2$ implique que $\Phi'_X(0) = 0$ et $\Phi''_X(0) = -\sigma^2$. Ainsi, au voisinage de 0,

$$\Phi_X(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

Or, lorsque n tend vers $+\infty$, (à t fixé), $t/\sigma\sqrt{n}$ tend vers 0. On a donc

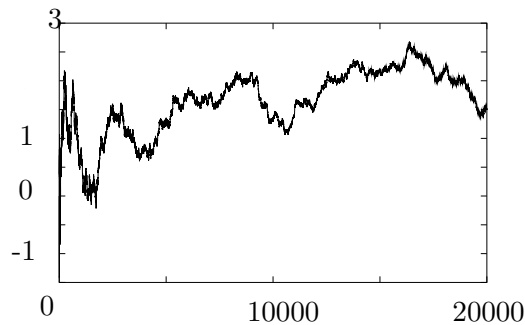
$$n \log \Phi_X(t/\sigma\sqrt{n}) = -n \frac{\sigma^2 t^2}{2n\sigma^2} + o(1) = \frac{-t^2}{2} + o(1).$$

En prenant l'exponentielle, on voit que

$$\Phi_{S_n/\sigma\sqrt{n}}(t) \xrightarrow{n} \exp(-t^2/2),$$

qui est la fonction caractéristique d'une $\mathcal{N}(0, 1)$. Il s'ensuit que $\frac{S_n}{\sigma\sqrt{n}}$ converge en loi vers une $\mathcal{N}(0, 1)$. \square

Le théorème de la limite centrale est seulement une convergence en loi comme le montre la figure suivante où est représenté la suite $(S(i)/\sqrt{i})_i$ où $S(i) = \sum_{j=1}^i X_j$ et où les X_j sont des variables valant ± 1 avec proba 1/2. $S(i)/\sqrt{i/4}$ converge en loi vers une $\mathcal{N}(0, 1)$. "On voit" que $S(i)/\sqrt{i/4}$ ne converge pas en tant que suite. C'est la loi de $S(i)/\sqrt{i/4}$ qui converge, d'après le théorème de la limite centrale. La convergence en loi n'est pas visible sur ce genre de dessin. Pour voir quelque



chose, il faudrait, par exemple, tracer la fonction de répartition de $S(i)/\sqrt{i/4}$. On verrait alors, que $F_{S(i)/\sqrt{i/4}}$ converge simplement vers $F_{\mathcal{N}(0,1)}$.

8. Hiérarchie des convergences

On a quatre convergences à classer : convergence en proba, p.s., L^p et en loi. Les trois premières sont du même type comme déjà dit et sont donc naturellement comparables. La quatrième est un peu à part, et d'ailleurs c'est la plus faible des convergences.

En effet, prenons par exemple une suite $(X_n)_n$ de v.a. gaussiennes $\mathcal{N}(0, 1)$ indépendantes. Cette suite converge en loi. Par contre, la suite $(X_n)_n$ ne converge dans aucun des trois autres sens. L'indépendance des X_i empêche la suite $X_i(\omega)$ de converger.

Proposition 25 Soit $X, X_1, \dots, X_n, \dots$ des variables aléatoires définies sur le même espace $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans \mathbb{R}

a) on a pour $p > q$,

$$(X_n \xrightarrow[n]{L_p} X) \implies (X_n \xrightarrow[n]{L_q} X) \implies (X_n \xrightarrow[n]{lo} X)$$

$$(X_n \xrightarrow[n]{p.s.} X) \implies (X_n \xrightarrow[n]{proba} X) \implies (X_n \xrightarrow[n]{lo} X)$$

$$(X_n \xrightarrow[n]{L_p} X) \implies (X_n \xrightarrow[n]{L_q} X) \nearrow$$

b) Si $(X_n \xrightarrow[n]{proba} X)$ il existe une suite extraite X_{n_k} telle que $(X_{n_k} \xrightarrow[k]{p.s.} X)$

c) Si $(X_n \xrightarrow[n]{p.s.} X)$ et s'il existe une variable aléatoire Y dominant la suite $|X_n|$ et appartenant à L_p alors $(X_n \xrightarrow[n]{L_p} X)$

d) Si $(X_n \xrightarrow[n]{L_p} X)$ il existe une sous suite extraite X_{n_k} telle que $(X_{n_k} \xrightarrow[k]{p.s.} X)$

Preuve : (a). Montrons d'abord que la convergence en probabilité implique la convergence en loi grâce à la convergence des fonctions caractéristiques (théorème de Lévy) :

$$|\mathbb{E}(e^{itX_n}) - \mathbb{E}(e^{itX})| \leq \mathbb{E}(|e^{itX_n} - e^{itX}| \mathbf{1}_{|X_n - X| \leq \eta}) + \mathbb{E}(|e^{itX_n} - e^{itX}| \mathbf{1}_{|X_n - X| > \eta}).$$

Soit $\epsilon > 0$ fixé. Pour tout t , le premier terme dans le membre de droite peut être rendu inférieur à ϵ grâce à l'uniforme continuité de $x \mapsto e^{itx}$ (en prenant η suffisamment petit) ; fixons maintenant un tel η . Le deuxième terme du membre de droite tend vers 0 car majoré par $2\mathbb{P}(|X_n - X| > \eta)$; pour n assez grand, il est donc inférieur à ϵ .

– La convergence L_p implique la convergence L_q est un résultat classique sur les espaces de mesure finie et un corollaire d'Hölder ; on écrit

$$\mathbb{E}(|X_n - X|^q) = \int_{\Omega} |X_n - X|^q d\mathbb{P} = \| |X_n - X|^q \|_1,$$

(il s'agit de la norme 1 contre la mesure \mathbb{P} ; en utilisant cette interprétation, on a

$$\begin{aligned} \mathbb{E}(|X_n - X|^q) &= \mathbb{E}(|X_n - X|^q \cdot 1) \leq \left(\mathbb{E}(|X_n - X|^{q \frac{p}{p-q}}) \right)^{\frac{q}{p}} \left(\mathbb{E}(1^{\frac{p}{p-q}}) \right)^{\frac{p-q}{p}} \\ &\leq \left(\mathbb{E}(|X_n - X|^p) \right)^{\frac{q}{p}} \longrightarrow 0. \end{aligned}$$

– La convergence L_p implique la convergence en proba : d'après Markov, pour tout $\epsilon > 0$, on a

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \mathbb{E}(|X_n - X|^p) / \epsilon^p \longrightarrow 0.$$

– la convergence p.s. implique la convergence en probabilité : On a $\mathbb{P}(\lim |X_n - X| = 0) = 1$, donc, $\mathbb{P}(\limsup_n (|X_n - X| \geq \epsilon)) = 0$. On a $\limsup_n () = \bigcap_n B_n$ avec $B_n = \bigcup_{k \geq n} \{\omega, |X_k(\omega) - X(\omega)| \geq \epsilon\}$; les B_n forment une suite décroissante d'événements. On a donc

$$\mathbb{P}(\limsup_n |X_n - X| \geq \epsilon) = \lim_n \mathbb{P}(B_n) = 0;$$

on conclut en utilisant le fait que $\mathbb{P}(|X_n - X| \geq \epsilon) \leq \mathbb{P}(B_n)$.

(b). La convergence en probabilité implique, que pour tout $i \geq 1$, il existe N_i tel que

$$\forall n \geq N_i, \quad \mathbb{P}(|X_n - X| \geq 1/i) \leq \frac{1}{i^2};$$

ceci est une conséquence du fait que $\mathbb{P}(|X_n - X| \geq 1/i) \xrightarrow{n \rightarrow +\infty} 0$. On peut de plus choisir une suite N_i strictement croissante (dans \mathbb{N}).

Prenons

$$A_i = \{\omega \mid |X_{N_i}(\omega) - X(\omega)| \geq 1/i\}.$$

Les ensembles A_i sont des éléments de la tribu \mathcal{A} . D'après le lemme de Borel-Cantelli (1), comme $\sum \mathbb{P}(A_i) \leq \sum 1/i^2 = \pi^2/6 < +\infty$, $\mathbb{P}(\limsup A_i) = 0$. Ceci implique que \mathbb{P} p.s. ω n'appartient à au plus qu'un nombre fini de A_i . Mais si ω appartient à un nombre fini de A_i ceci signifie que $X_{N_i}(\omega)$ converge vers $X(\omega)$. On a donc établi que $X_{N_i}(\omega)$ converge vers $X(\omega)$ \mathbb{P} p.s..

(c) on suppose qu'il existe Y telle que, pour tout ω , $|X_n(\omega)| \leq Y(\omega)$ et $\mathbb{E}(Y^p) < +\infty$; d'après le théorème de convergence dominée, comme $|X_n - X|^p$ tend vers 0 et est dominé par $2|Y|^p$, $\mathbb{E}(|X_n - X|^p) = \int_{\Omega} |X_n(\omega) - X(\omega)|^p d\mathbb{P}(\omega) \rightarrow 0$. Le fait que $X \in L_p$ vient du fait que si $|X_n| \leq Y$ alors $|X| \leq Y$

(d) La convergence dans L_p implique la convergence en proba et donc l'existence d'une sous suite convergeant p.s. d'après (b). \square

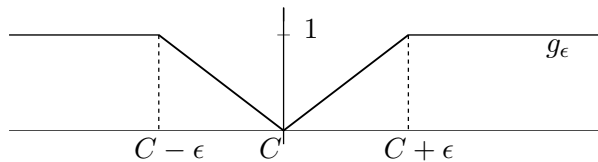
Proposition 26 Soit C une constante; les deux assertions suivantes sont équivalentes :

- a) $X_n \xrightarrow[n]{proba} C$.
- b) $X_n \xrightarrow[n]{loi} C$.

(On devrait encore écrire, $X_n \rightarrow X$ avec X constante, égale à C , p.s..)

Preuve : (a) \implies (b) Ceci est une conséquence du (a) de la proposition précédente.

(b) \implies (a) En prenant g_ϵ la fonctions continue et bornée dessinée ci-dessous on a,



$$\mathbb{P}(|X_n - C| \geq \epsilon) \leq \mathbb{E}(g_\epsilon(X_n)) \rightarrow \mathbb{E}(g_\epsilon(X)) = g_\epsilon(C) = 0.$$

Ceci implique bien sûr la convergence en probabilité de X_n vers $X \equiv C$. \square

V. ANNEXES : Rappels d'intégration et de théorie des ensembles

1. Tribus et mesures

Tribu : - Soit Ω un ensemble. On appelle tribu (ou σ -algèbre) de Ω une partie de $\mathcal{P}(\Omega)$ contenant l'ensemble vide, stable par passage au complémentaire, stable par union (ou intersection) dénombrable.

Si \mathcal{A} est une tribu, le couple (Ω, \mathcal{A}) est appelé *espace mesurable*.

Pour $\mathcal{C} \subset \mathcal{P}(\Omega)$, on appelle tribu engendrée par \mathcal{C} la plus petite tribu de Ω contenant \mathcal{C} .

Mesure : On appelle mesure sur un espace mesurable (Ω, \mathcal{A}) une application $\mu : \mathcal{A} \rightarrow \mathbb{R}^+$ telle que :

- $\mu(\emptyset) = 0$
- Si les A_n constituent une suite d'éléments deux à deux disjoints de la tribu \mathcal{A} , $\mu(\cup_n A_n) = \sum_n \mu(A_n)$

Probabilité : Une probabilité sur \mathcal{A} (tribu de Ω) est une mesure de masse totale 1 ($\mu(\Omega) = 1$).

Fonction mesurable : Soient (Ω, \mathcal{A}) et (E, \mathcal{E}) deux espaces mesurables. Une fonction f de Ω dans E est dite mesurable si pour tout $b \in \mathcal{E}$, $f^{-1}(b) \in \mathcal{A}$.

Propriété des probabilités : (Dans ce qui suit \mathbb{P} est une probabilité sur \mathcal{A} , tribu de Ω)

- Si les A_n forment une suite croissante de \mathcal{A} , $\mathbb{P}(\cup_n A_n) = \lim \mathbb{P}(A_n)$
- Si les A_n forment une suite décroissante de \mathcal{A} , $\mathbb{P}(\cap_n A_n) = \lim \mathbb{P}(A_n)$

2. Intégration

Théorème de Beppo Levi : Si $(f_n)_n$ est une suite croissante de fonctions positives (et mesurables) convergeant simplement alors

$$\lim_n \int f_n(x) d\mu(x) = \int \lim_n f_n(x) d\mu(x).$$

Théorème de convergence dominée de Lebesgue : Si $(f_n)_n$ est une suite de fonctions (mesurables) convergeant simplement, telle que les fonctions $|f_n|$ sont dominées par une fonction $g \in L_1$, alors

$$\lim_n \int f_n(x) d\mu(x) = \int \lim_n f_n(x) d\mu(x).$$

Théorème de Fubini : Si f est une fonction de $E \times F$ à valeur dans \mathbb{R} , mesurable par rapport à la tribu produit (des tribus sur E et F), alors, si $f \in L_1(\mu \otimes \lambda)$, on a,

$$\int_{E \times F} f d\mu \otimes \lambda = \int_F \int_E f(x, y) d\mu(x) d\lambda(y) = \int_E \int_F f(x, y) d\lambda(y) d\mu(x).$$

(On a bien sûr, pour presque tout y , $x \mapsto f(x, y)$ mesurable et intégrable, ainsi que $x \mapsto \int f(x, y) dy$ (et de même en échangeant x et y). Le théorème de Fubini-Tonelli assure que l'interversion est légitime dans le cas où f est positive et mesurable.

Espaces L_p : L'espace L_p est l'ensemble des fonctions de puissance p ème intégrable. Lorsque la mesure est finie, (par exemple lorsqu'il s'agit d'une mesure de probabilité), on a :

$$L_p \subset L_q \text{ si } p \geq q.$$

3. Quelques éléments sur les mesures sur \mathbb{R}

a. Décompositions des mesures réelles

Les théorèmes qui suivent (comme les précédents d'ailleurs) sont donnés sans preuves. Celles-ci peuvent être trouvées dans tout bon livre de théorie de la mesure dans un cadre plus général (voir par exemple Rudin "Analyse réelle et complexe" p 117-). Dans la suite, on se pose la question suivante : soit μ une mesure sur $L(\Omega)$ (Ω étant un sous ensemble de \mathbb{R}) ; que peut-on dire de cette mesure par rapport à celle de Lebesgue ?

Soient μ et η deux mesures positives sur une tribu \mathfrak{M} .

Définition : On dit que η est absolument continue par rapport à μ , et on écrit

$$\eta \ll \mu,$$

si $\eta(E) = 0$ pour tout $E \in \mathfrak{M}$ tel que $\mu(E) = 0$.

Définition : S'il existe un ensemble $A \in \mathfrak{M}$ tel que $\eta(E) = \eta(E \cap A)$ pour tout $E \in \mathfrak{M}$, on dit que η est portée par A .

Définition : Deux mesures μ_1 et μ_2 sur une tribu \mathfrak{M} sont dites *mutuellement singulières* (ou étrangères) s'il existe deux ensembles A et B disjoints de \mathfrak{M} tels que A porte μ_1 et B porte μ_2 . On note

$$\mu_1 \perp \mu_2.$$

On note dans toute la suite λ la mesure de Lebesgue, $L(\Omega)$ la tribu des Lebesgue mesurables.

Théorème de décomposition de Lebesgue : Soit μ une probabilité sur $L(\Omega)$.

a) Il existe un unique couple μ_a et μ_s sur $L(\Omega)$ tel que :

$$\mu = \mu_a + \mu_s, \quad \mu_a \ll \lambda \quad \text{et} \quad \mu_s \perp \lambda$$

Les mesures μ_a et μ_s sont positives et $\mu_a \perp \mu_s$

b) Il existe une unique fonction de $L_1(\lambda)$ telle que

$$\mu_a(E) = \int_E h d\lambda \quad \text{pour tout } E \in \tau$$

Le point (b) très important, s'appelle le théorème de Radon-Nikodym.

b. Application aux mesures de probabilité

Prenons tout d'abord des cas particuliers pour bien voir ce qui se passe.

• Mesure absolument continue par rapport à la mesure de Lebesgue. (Prenons $\Omega = \mathbb{R}$).

Si μ est abs. cont. par rapport à λ , d'après le théorème de décomposition de Lebesgue μ_s est la mesure nulle, et on a l'existence d'une fonction h (intégrable et même d'intégrale 1) telle que, pour tout A Lebesgue mesurable,

$$\mu(A) = \int_A h(x) d\lambda(x) = \int_A h(x) dx. \quad (37)$$

On dit que h est la *densité* de μ par rapport à la mesure de Lebesgue. Réciproquement, toute fonction f , Lebesgue-mesurable, positive et d'intégrale 1 permet de construire une mesure μ absolument continue par rapport à la mesure de Lebesgue grâce à la formule (37).

- Mesure singulière. Si μ est singulière elle ne charge qu'un ensemble A Lebesgue négligeable. Ainsi, μ_a est la mesure nulle. Supposons que A puisse être choisi dénombrable (A n'est pas unique dans la décomposition ; il est défini à un μ -négligeable près). Si A est dénombrable, μ est une mesure sur un dénombrable. Elle peut s'écrire comme somme pondérée de Dirac. Si A n'est pas dénombrable, retirons les éventuelles Dirac. Il reste une mesure qui ne charge pas les points et qui n'est pas absolument continue par rapport à la mesure de Lebesgue.

Voyons, sur un exemple, à quoi ressemble une telle mesure : prenons C l'ensemble de Cantor (il est négligeable et non dénombrable). C'est l'ensemble des nombres ne s'écrivant en base 3 qu'avec des 0 et des 2. Notons $x^{(3)} = 0, x_1 x_2 x_3 \dots$ pour le développement de x en base 3.

Considérons maintenant l'écriture de x en base 2 des nombres de $[0, 1[$ que l'on note $x^{(2)}$. Les dyadiques (les décimaux de la base 2), c'est-à-dire, les nombres possédant une écriture de la forme $x^{(2)} = 0, x_1 x_2 \dots x_k 00 \dots$ (pour un certain k et $x_i = 0$ ou 1) possèdent deux écritures (une avec un nombre infini de 1, l'autre avec un nombre infini de 0). On choisit l'écriture avec des 0.

On définit f de $[0, 1]$ et à valeurs sur $[0, 1]$ par

$$f(x) = y$$

où

$$y^{(3)} = 2x^{(2)} :$$

On a par exemple, si $x^{(2)} = 0, 1101011\dots$, $(f(x))^{(3)} = 0, 2202022\dots$ (on a multiplié les décimales par 2 et changé de base). f est évidemment croissante et est donc borélienne. De plus l'image de f est contenue dans C . On peut donc construire la mesure μ sur $\mathcal{B}[0, 1]$ suivante :

$$\mu(A) = \mu(A \cap C) = \lambda(f^{-1}(A \cap C)).$$

Cette mesure est bien définie car, f étant croissante, elle est borélienne. Cette mesure charge évidemment uniquement le Cantor, mais elle ne charge pas les points car λ ne les charge pas (il est de plus aisé de voir que cette mesure n'est pas nulle).

- Le théorème de décomposition de Lebesgue affirme donc que toute mesure de probabilité peut s'écrire de manière unique sous la forme de deux mesures. En fait, comme on l'a vu dans le deuxième exemple, la mesure singulière peut encore être décomposée comme une somme de Dirac et une mesure ne chargeant pas les points. La partie "somme de Dirac" constitue la partie probabilité discrète de la mesure (étudiée dans le premier chapitre). L'autre partie singulière sera laissée de côté dans ce cours, mais il est important parfois de se souvenir qu'elle existe.

Références

- [1] J. Bertoin (2000), *Probabilités : Cours de licence de mathématiques appliquées*, Paris 6.
- [2] L. Breiman (1968) *Probability.*, SIAM.
- [3] P. Flajolet et Sedgewick *The average case analysis of algorithm*
- [4] J. Jacod (1999) *Probabilités*, Cours de licence, Paris 6.