

Introduction au calcul des probabilités

Philippe BRIAND

Bureau 335, Bâtiment 23

tel. : 02 23 23 58 86

email : philippe.briand@univ-rennes1.fr

<http://perso.univ-rennes1.fr/philippe.briand/>

Rennes, Septembre 2003

Espace probabilisé

Le calcul des probabilités est une branche de l'analyse dont le but est l'étude des phénomènes aléatoires c'est à dire des expériences dont le résultat ne peut être prédit avec certitude : par exemple, si on répète l'expérience plusieurs fois, on peut obtenir des résultats différents. Une forme d'indétermination apparaît dans l'issue de l'expérience. On peut interpréter cette forme de hasard comme notre incompetence à concevoir, expliciter et utiliser les phénomènes physiques considérés ou encore comme un manque d'information sur les conditions de l'expérience. Certains phénomènes sont peut-être par essence même sujets au hasard . . .

Les exemples de telles expériences sont nombreux : on peut penser au jeu de pile ou face, au lancer d'un dé, au sexe d'un enfant à naître, au temps d'attente d'un client à la poste, à la durée de vie d'une particule radioactive . . .

Pour étudier ces phénomènes, on doit tout d'abord en faire un modèle mathématique. Le modèle retenu pour décrire les expériences aléatoires est celui d'un triplet communément noté $(\Omega, \mathcal{F}, \mathbb{P})$ et appelé *espace probabilisé* ou *espace de probabilité*.

Un exemple. Il est 18 heures et vous devez impérativement vous rendre à l'autre bout de la ville pour faire une course importante. Votre petit(e) ami(e) doit passer chez vous vers 20 heures ce soir et d'autre part votre chère tante Pétunia doit vous rendre visite vers 18 heures 45.

À cause des embouteillages (on peut imaginer d'autres raisons), vous ne pouvez pas déterminer avec précision à quelle heure vous serez de retour de votre course en ville. Vous êtes alors intéressés par les trois événements suivants :

- A : retour avant 20 h.
- B : retour après 18 h. 45
- C : retour avant 20 h. 30

Les événements A et B parlent d'eux-mêmes et pour le C vous comptez sur la patience de votre ami(e).

En dehors de ces trois événements qui vous intéressent au plus haut point, on considère aussi la famille d'événements plus élémentaires formée par : « je suis de retour à la date t » où t varie continûment de 18 h. à 21 h. On appelle ces événements élémentaires des épreuves ou des réalisations. Considérons l'épreuve ω à savoir vous êtes de retour à 19 heures. Dans cette épreuve, les événements A et C sont réalisés tandis que l'événement B lui ne l'est pas.

Pour étudier ce phénomène, on est donc amené à considérer un espace fondamental Ω , l'espace de toutes les possibilités, et à l'identifier avec l'intervalle $[18, 21]$. L'épreuve ω considérée précédemment est alors identifiée à l'élément 19. L'événement A est représenté par $[18, 20]$, B par $[18.75, 21]$, C par $[18, 20.5]$. Dans cet exemple, on voit que l'appartenance « $\omega \in A$ » équivaut à « A est réalisé dans l'épreuve ω ». Cette équivalence permet de passer du vocabulaire probabiliste « A est réalisé dans l'épreuve ω » au vocabulaire ensembliste « $\omega \in A$ ».

Le triplet fondamental. Dans l'exemple précédent, on a pu expliciter complètement l'espace Ω mais ce n'est pas le cas en général ; on sait seulement qu'il fait partie d'un triplet $(\Omega, \mathcal{F}, \mathbb{P})$ possédant les propriétés suivantes :

– Ω est un ensemble non vide décrivant l'ensemble des réalisations possibles de l'expérience. Un point $\omega \in \Omega$ est une réalisation (ou épreuve) particulière. Par exemple, si on modélise le jeu de

pile ou face on peut prendre comme espace Ω l'ensemble $\{0, 1\}$ ou $\{P, F\}$ ou $\{pile, face, tranche\}$, pour le lancer d'un dé on peut prendre $\Omega = \{1, 2, 3, 4, 5, 6\}$.

– \mathcal{F} est une sous-classe de parties de Ω qui représente la classe des événements associés à l'expérience. Par exemple, si on observe la durée de vie d'une ampoule on peut s'intéresser aux événements suivants :

- S : la durée de vie est positive ; c'est l'événement certain.
- I : la durée de vie est strictement négative ; événement impossible.
- B : la durée de vie est supérieure à 100 heures.
- C : la durée de vie est inférieure à 200 heures.

Quelles opérations logiques peut-on effectuer sur les événements ?

- si A est un événement, alors sa négation est aussi un événement ;
- si A et B sont des événements alors « A ou B », « A et B » sont des événements ;
- si on a A_n une suite d'événements, « au moins un des A_n se réalise » est encore un événement.

Ces opérations logiques se traduisent au niveau ensembliste par les opérations classiques : passage au complémentaire pour la négation, union pour « ou », intersection pour « et ».

Dans les cas simples, on peut prendre pour \mathcal{F} l'ensemble de toutes les parties de Ω mais dans les cas un peu plus complexes, il faudra se limiter à une sous-classe.

– Il reste à mesurer les chances qu'un événement se réalise (les chances de devenir millionnaire si on joue au Loto par exemple). Cela consiste à donner une pondération à chacun des événements. Pour obtenir cela, on peut penser à utiliser une approche empirique. Fixons un événement $A \in \mathcal{F}$. On reproduit N fois l'expérience et on compte le nombre de fois N_A où A se réalise. On évalue alors les chances que A se réalise par sa fréquence $\text{Fr}(A) = N_A/N$. On peut remarquer que $0 \leq \text{Fr}(A) \leq 1$ pour tout événement et que si A et B sont deux événements incompatibles i.e. $A \cap B = \emptyset$, on a $\text{Fr}(A \cup B) = \text{Fr}(A) + \text{Fr}(B)$. Cette façon de procéder n'est pas totalement satisfaisante car le résultat dépend du nombre N d'expériences réalisées. De plus il n'est pas toujours très facile de répéter une expérience et cela peut s'avérer très coûteux. C'est pourquoi nous introduirons les mesures de probabilité qui sont des applications de \mathcal{F} dans $[0, 1]$ qui associe à chaque événement le nombre $\mathbb{P}(A)$ qui est la probabilité que A se réalise.

Nous voici donc un peu plus familier avec l'expression qui reviendra souvent dans ce cours : « soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé ». La connaissance explicite de ce triplet n'est importante que si l'on s'intéresse à la phase de modélisation ; dans ce cas, Ω est en général l'espace des états de l'expérience et (E, \mathcal{E}) remplace souvent (Ω, \mathcal{F}) . En dehors de cette phase, on ne spécifie pas l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ qui est un espace abstrait sur lequel on n'a aucune information.

Contexte habituel. En fait, la situation la plus fréquente que nous rencontrerons est la suivante : on fait une expérience aléatoire qui est traduite par l'espace $(\Omega, \mathcal{F}, \mathbb{P})$ mais on s'intéresse à certaines conséquences de l'expérience. Par exemple, on lance deux dés et on regarde si la somme est supérieure à huit. On peut prendre pour Ω l'ensemble de tous les résultats possibles i.e. l'ensemble $\{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$; si les dés ne sont pas pipés chacun de ces couples a les mêmes chances d'être obtenu et donc on définit \mathbb{P} par $\mathbb{P}((i, j)) = 1/36$. Comme on s'intéresse à la somme des points des dés, on introduit, pour tout couple $\omega = (i, j)$, $X(\omega) = i + j$. On a donc une application $X : \Omega \rightarrow \{2, \dots, 12\}$. Dans ce problème, la quantité que l'on cherche à déterminer est $\mathbb{P}(\{\omega \in \Omega, X(\omega) \in \{8, 9, 10, 11, 12\}\})$ que l'on note $\mathbb{P}_X(\{8, 9, 10, 11, 12\})$.

De façon générique, nous aurons une application définie sur Ω d'espace d'états E – nous dirons variable aléatoire – $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ et l'on cherchera à déterminer la mesure de

probabilité \mathbb{P}_X – la loi de X sous \mathbb{P} – définie par

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

La fin de ce premier cours consiste à donner les définitions mathématiques des objets que nous avons introduits.

1. Notion de tribu.

Définition. Soit Ω un ensemble non vide et \mathcal{F} une classe de parties de Ω . On dit que \mathcal{F} est une *tribu* sur Ω si

- (i) $\Omega \in \mathcal{F}$;
- (ii) \mathcal{F} est stable par passage au complémentaire : si A appartient à \mathcal{F} alors A^c appartient à \mathcal{F} ;
- (iii) \mathcal{F} est stable par union dénombrable : si $(A_n)_{\mathbb{N}} \subset \mathcal{F}$ alors $\cup_{\mathbb{N}} A_n$ appartient à \mathcal{F} .

Si \mathcal{F} est une tribu sur Ω , (Ω, \mathcal{F}) est appelé espace mesurable et les éléments de \mathcal{F} des événements.

Remarque. Si \mathcal{F} est une tribu sur Ω , alors $\emptyset \in \mathcal{F}$. De plus, \mathcal{F} est stable par union finie, par intersection finie et dénombrable, par différence ... Expliquons par exemple pourquoi \mathcal{F} est stable par intersection dénombrable. Soit donc $(A_n)_{\mathbb{N}} \subset \mathcal{F}$. On a $\cap_n A_n = (\cup_n A_n^c)^c$. Comme, pour tout n , $A_n \in \mathcal{F}$, le point (ii) de la définition implique que A_n^c appartient à \mathcal{F} pour tout n . Le point (iii) entraîne donc que $\cup_n A_n^c$ est élément de \mathcal{F} ; on peut alors appliquer à nouveau le point (ii) pour obtenir que $\cap_n A_n$ est un élément de \mathcal{F} .

Exemple. Soit Ω un ensemble non vide. $\mathcal{F}_1 = \{\emptyset, \Omega\}$ est une tribu appelée tribu grossière; $\mathcal{F}_2 = \mathcal{P}(\Omega)$, la collection des parties de Ω , est une tribu; enfin, si $A \subset \Omega$, $\mathcal{F}_3 = \{\emptyset, \Omega, A, A^c\}$ est une tribu sur Ω .

Exemple. On considère \mathcal{A} l'ensemble des parties $A \subset \mathbb{N}$ vérifiant A est fini ou A^c est fini. \mathcal{A} n'est pas une tribu sur \mathbb{N} . En effet, les deux premiers points de la définition sont satisfaits mais le troisième ne l'est pas. Pour s'en convaincre, considérons, pour $n \in \mathbb{N}$, $A_n = \{2n\}$ qui appartient à \mathcal{A} puisque son cardinal est 1. Par contre, $\cup_n A_n$ n'appartient pas à \mathcal{A} : ni $\cup_n A_n$ ni son complémentaire ne sont de cardinal fini puisqu'il s'agit respectivement des entiers pairs et impairs.

Comme déjà dit en introduction, on aimerait travailler avec la tribu $\mathcal{P}(\Omega)$. Ceci est possible par exemple si Ω est un ensemble fini ou dénombrable. Si $\Omega = \mathbb{R}$, il n'est pas possible de mesurer toutes les parties de \mathbb{R} sans aboutir à une contradiction et on se limite à une classe de parties appelée tribu borélienne.

Définition. On appelle tribu borélienne sur \mathbb{R} , notée $\mathcal{B}(\mathbb{R})$, la plus petite tribu, au sens de l'inclusion, contenant tous les intervalles de \mathbb{R} .

Cette définition appelle un commentaire : il n'est pas évident a priori que l'on puisse parler de « la plus petite tribu » contenant les intervalles. Toutefois, la définition a bien un sens car l'intersection d'une famille quelconque de tribus sur Ω est encore une tribu sur Ω . On peut donc considérer l'intersection de toutes les tribus sur \mathbb{R} contenant les intervalles qui devient par construction « la plus petite ».

2. Mesure de probabilité.

Définition. Soient (Ω, \mathcal{F}) un espace mesurable et \mathbb{P} une application de \mathcal{F} dans $[0, 1]$. \mathbb{P} est une (mesure de) probabilité sur (Ω, \mathcal{F}) si

- (i) $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$;
- (ii) pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'événements deux à deux disjoints, c'est à dire vérifiant $A_n \cap A_m = \emptyset$ si $n \neq m$, $\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n)$.

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ où (Ω, \mathcal{F}) est un espace mesurable et \mathbb{P} une probabilité sur cet espace s'appelle un espace probabilisé ou espace de probabilité.

Remarque. On peut voir facilement que la condition $\mathbb{P}(\emptyset) = 0$ est une conséquence des deux autres points. D'autre part, la deuxième propriété s'appelle la σ -additivité.

La définition conduit aux propriétés suivantes :

Proposition 1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Si A et B appartiennent à \mathcal{F} , alors

- additivité : si $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$;
- croissance : si $A \subset B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$;
- si $A \subset B$, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$; en particulier, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Pour obtenir l'additivité, il suffit de prendre dans la définition $A_0 = A$, $A_1 = B$ et $A_i = \emptyset$ si $i \geq 2$ et d'utiliser la σ -additivité. Pour les deux points suivants, si $A \subset B$, on écrit $B = A \cup (B \setminus A)$ et on utilise l'additivité : $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Ensuite, pour la dernière relation, on a - faites un dessin - $A \cup B = A \cup (B \setminus A)$ et $B = (B \setminus A) \cup (B \cap A)$. Par additivité,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A), \quad \mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(B \cap A) ;$$

Finalement, on obtient

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

Exemple. Soient Ω un ensemble non vide et ω un point de Ω . Pour $A \subset \Omega$, on pose $\delta_\omega(A) = 1$ si $\omega \in A$ et 0 sinon. δ_ω est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ appelé masse de Dirac au point ω .

En effet, $\delta_\omega(\Omega) = 1$ et $\delta_\omega(\emptyset) = 0$. Seule la σ -additivité n'est pas immédiate. Soit $(A_n)_{n \in \mathbb{N}}$ une suite de parties disjointes 2 à 2. Si $\omega \notin \cup_n A_n$, alors ω n'appartient à aucun des A_n et donc $\delta_\omega(\cup_n A_n) = 0$, $\delta_\omega(A_n) = 0$ pour tout n . Si ω appartient à $\cup_n A_n$, il existe un entier p tel que $\omega \in A_p$. Pour tout $n \neq p$, $\omega \notin A_n$ puisque $A_p \cap A_n = \emptyset$. Donc, dans ce cas,

$$\delta_\omega(\cup_n A_n) = 1, \quad \delta_\omega(A_p) = 1, \quad \delta_\omega(A_n) = 0, \quad \text{si } n \neq p.$$

Dans les deux cas, on a $\delta_\omega(\cup_n A_n) = \sum_{n \in \mathbb{N}} \delta_\omega(A_n)$, ce qui montre la σ -additivité.

Proposition 2. Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$. Alors,

- $\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) \leq \sum_{n \geq 0} \mathbb{P}(A_n)$;
- si la suite $(A_n)_{n \in \mathbb{N}}$ est croissante ($A_n \subset A_{n+1}$ pour tout n),

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \sup_{n \in \mathbb{N}} \mathbb{P}(A_n) ;$$

– si la suite $(A_n)_{n \in \mathbb{N}}$ est décroissante ($A_{n+1} \subset A_n$ pour tout n),

$$\mathbb{P}\left(\bigcap_{n \geq 0} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \inf_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

La démonstration de cette proposition repose sur la construction suivante : si $(A_n)_{n \in \mathbb{N}}$ est une suite de parties de \mathcal{F} on construit une suite $(B_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ de parties deux à deux disjointes telle que, pour tout n ,

$$B_n \subset A_n, \quad \bigcup_{i=0}^n B_i = \bigcup_{i=0}^n A_i, \quad \bigcup_{i \geq 0} B_i = \bigcup_{i \geq 0} A_i.$$

Il suffit de poser pour cela $B_0 = A_0$ et $B_n = A_n \setminus \bigcup_{i=0}^{n-1} A_i$, si $n \geq 1$. On a alors,

$$\mathbb{P}\left(\bigcup_n A_n\right) = \mathbb{P}\left(\bigcup_n B_n\right) = \sum_{n \geq 0} \mathbb{P}(B_n) \leq \sum_{n \geq 0} \mathbb{P}(A_n).$$

Quand la suite $(A_n)_{n \in \mathbb{N}}$ est croissante, on a, pour tout entier $n \geq 1$, $B_n = A_n \setminus A_{n-1}$ et par suite $\mathbb{P}(B_n) = \mathbb{P}(A_n) - \mathbb{P}(A_{n-1})$. Il vient alors,

$$\mathbb{P}\left(\bigcup_n A_n\right) = \mathbb{P}\left(\bigcup_n B_n\right) = \sum_{n \geq 0} \mathbb{P}(B_n) = \lim_{p \rightarrow \infty} \sum_{n=0}^p \mathbb{P}(B_n) = \lim_{p \rightarrow \infty} \mathbb{P}(B_p),$$

ce qui établit la seconde assertion. Le troisième point s'obtient par passage au complémentaire.

3. Indépendance, conditionnement.

Commençons par une notion très importante pour le calcul des probabilités, la notion d'indépendance pour des événements.

3.1. Événements indépendants. On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Rappelons tout d'abord que

Définition. Deux événements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

On obtient souvent des événements indépendants lorsqu'on reproduit une expérience sans que la première expérience n'interfère avec la seconde. C'est par exemple le cas lorsque l'on joue deux fois à pile ou face. Voici un autre exemple.

Exemple. Si on lance deux fois un dé équilibré, les deux événements A « le premier dé vaut 4 » et B « le deuxième dé vaut 5 » sont indépendants puisque

$$\mathbb{P}(A \cap B) = 1/36 = 1/6 \times 1/6 = \mathbb{P}(A)\mathbb{P}(B).$$

Par contre, si C désigne « la somme des dés vaut 5 », A et C ne sont pas indépendants. En effet, $\mathbb{P}(A) = 1/6$, $\mathbb{P}(C) = 1/9$ et $\mathbb{P}(A \cap C) = 1/36$.

Introduisons une notion plus délicate : l'indépendance de n événements.

Définition. Soit $(A_i)_{1 \leq i \leq n} \subset \mathcal{F}$ n événements. Ils sont indépendants si

$$\forall J \subset \{1, \dots, n\}, \quad \mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

Le piège est le suivant : si A, B, C sont trois événements indépendants alors (A, B) , (A, C) et (B, C) sont des couples d'événements indépendants mais la réciproque est fautive.

Remarque. La définition 3.1 est encore valable pour une suite d'événements ; J doit alors être une partie finie de \mathbb{N} .

Exemple. On lance deux fois un dé équilibré et on s'intéresse aux événements suivants : A = « le premier résultat est pair », B = « le second résultat est pair » et C = « la somme est paire ». A, B et C sont deux à deux indépendants mais ne sont pas indépendants dans leur ensemble.

En effet, $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$, $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = 1/4$, mais $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \cap B) = 1/4$.

3.2. Probabilité conditionnelle. Fixons $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$. Considérons l'application \mathbb{P}^B suivante :

$$\mathbb{P}^B : \mathcal{F} \longrightarrow [0, 1], \quad A \longmapsto \mathbb{P}(A \cap B) / \mathbb{P}(B).$$

On montre que \mathbb{P}^B est une probabilité sur (Ω, \mathcal{F}) ; en effet, $\mathbb{P}^B(\Omega) = \mathbb{P}(B) / \mathbb{P}(B) = 1$, $\mathbb{P}^B(\emptyset) = 0$ et si $(A_n)_n$ est une suite d'événements deux à deux disjoints, on a

$$\mathbb{P}^B(\cup_n A_n) = \mathbb{P}(B)^{-1} \mathbb{P}(B \cap (\cup_n A_n)) = \mathbb{P}(B)^{-1} \mathbb{P}(\cup_n (B \cap A_n)) ;$$

les événements $A_n \cap B$ sont eux-mêmes deux à deux disjoints et par suite

$$\mathbb{P}^B(\cup_n A_n) = \mathbb{P}(B)^{-1} \sum \mathbb{P}(B \cap A_n) = \sum \mathbb{P}^B(A_n).$$

Définition. Soit $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$. On appelle probabilité conditionnelle de A sachant B , notée $\mathbb{P}(A|B)$, la quantité $\mathbb{P}^B(A) = \mathbb{P}(A \cap B) / \mathbb{P}(B)$. Elle représente la probabilité que A se réalise sachant que B est réalisé.

On peut vérifier facilement que deux événements A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Exemple. Un couple possède deux enfants qui ne sont pas jumeaux. La probabilité que les deux enfants soient des filles vaut $1/4$; si on sait de plus que l'un des deux est une fille, cette probabilité vaut $(1/4)/(3/4) = 1/3$; si on sait à présent que l'aînée est une fille, la probabilité d'avoir deux filles est $(1/4)/(1/2) = 1/2$.

On peut remarquer que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \times \mathbb{P}(B) = \mathbb{P}(B|A) \times \mathbb{P}(A)$ – la dernière égalité n'étant vraie que si $\mathbb{P}(A) > 0$. On en déduit la

Formule de Bayes. Soient $(A_i)_{i \in \mathbb{N}} \subset \mathcal{F}$ une partition de Ω telle que $\mathbb{P}(A_i) > 0$ pour tout i et $B \in \mathcal{F}$ tel que $\mathbb{P}(B) > 0$. On a alors

$$\mathbb{P}(B) = \sum_{i \geq 0} \mathbb{P}(B \cap A_i) = \sum_{i \geq 0} \mathbb{P}(B|A_i) \times \mathbb{P}(A_i),$$

et

$$\forall j \in \mathbb{N}, \quad \mathbb{P}(A_j|B) = \mathbb{P}(B|A_j) \times \mathbb{P}(A_j) / \sum_{i \geq 0} \mathbb{P}(B|A_i) \mathbb{P}(A_i).$$

Exemple. Françoise, Bernard et Philippe vont se répartir les copies de PRBU à proportion de 50%, 25% et 25%. Françoise recevra 75% des étudiants dont elle corrige la copie, Bernard 60% et Philippe 70%. La probabilité qu'un étudiant choisi au hasard réussisse son examen s'obtient comme suit

$$\mathbb{P}(R) = \mathbb{P}(R|F) \mathbb{P}(F) + \mathbb{P}(R|B) \mathbb{P}(B) + \mathbb{P}(R|Ph) \mathbb{P}(Ph) = 0,7.$$

L'étudiant a réussi; la probabilité que sa copie ait été corrigée par Bernard est

$$\mathbb{P}(B|R) = \mathbb{P}(B \cap R)/\mathbb{P}(R) = \mathbb{P}(R|B) \times \mathbb{P}(B)/\mathbb{P}(R) = 0,6 \times 0,25/0,7 \simeq 0,2143.$$

Loi d'une variable aléatoire réelle

On introduit la notion de variable aléatoire dans le cas réel ainsi que la notion fondamentale de loi d'une variable aléatoire. La première chose à préciser est que dans toute la suite nous utiliserons l'abréviation « v.a.r. » pour « variable aléatoire réelle » au singulier comme au pluriel.

1. Définitions, vocabulaire.

Commençons par la définition fondamentale suivante :

Définition. Soient (Ω, \mathcal{F}) un espace mesurable. On appelle *variable aléatoire réelle* définie sur Ω toute application X de Ω dans \mathbb{R} vérifiant

$$\forall x \in \mathbb{R}, \quad X^{-1}(]-\infty, x]) = \{\omega \in \Omega, X(\omega) \leq x\} \in \mathcal{F}.$$

Une telle application est dite mesurable. En pratique on ne vérifie que très rarement la mesurabilité. De plus, la structure de la tribu borélienne (la plus petite tribu qui contient les intervalles) a pour conséquence qu'une variable aléatoire X vérifie

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{F}.$$

Remarque. Les probabilistes utilisent généralement la notation suivante : si B est un borélien alors l'ensemble $X^{-1}(B)$ est noté en abrégé $\{X \in B\}$ soit

$$X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} = \{X \in B\}.$$

De même, l'ensemble $X^{-1}(]-\infty, x])$ est noté $\{X \leq x\}$.

Introduisons à présent la loi d'une v.a.r. X .

Théorème 1. Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et X une v.a.r. définie sur Ω . L'application \mathbb{P}_X définie par

$$\mathbb{P}_X : \mathcal{B}(\mathbb{R}) \longrightarrow [0, 1], \quad B \longmapsto \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)),$$

est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

La mesure de probabilité \mathbb{P}_X s'appelle la loi de la variable aléatoire X sous \mathbb{P} ou plus simplement la loi de X .

Montrons rapidement que \mathbb{P}_X ainsi définie est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On a tout d'abord, $X^{-1}(\mathbb{R}) = \Omega$, $X^{-1}(\emptyset) = \emptyset$ et par suite,

$$\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1, \quad \mathbb{P}_X(\emptyset) = \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0.$$

Si $(B_n)_{n \in \mathbb{N}}$ est une suite de boréliens deux à deux disjoints, alors $(X^{-1}(B_n))_{n \in \mathbb{N}}$ est une suite de parties de \mathcal{F} deux à deux disjointes. Comme $X^{-1}(\cup_n B_n) = \cup_n X^{-1}(B_n)$, on a, \mathbb{P} étant une mesure de probabilité,

$$\mathbb{P}_X\left(\bigcup_n B_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_n B_n\right)\right) = \mathbb{P}\left(\bigcup_n X^{-1}(B_n)\right) = \sum_{n \geq 0} \mathbb{P}(X^{-1}(B_n)) = \sum_{n \geq 0} \mathbb{P}_X(B_n).$$

Ceci montre que \mathbb{P}_X est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

La question qui se pose alors est de savoir comment décrire la loi d'une variable aléatoire réelle.

2. Fonction de répartition.

Si X est une v.a.r. sa loi \mathbb{P}_X est déterminée par la valeur de $\mathbb{P}(X \in B)$ pour tout borélien B . Toutefois, comme nous allons le voir il suffit en fait de connaître la valeur de $\mathbb{P}(X \leq x)$ pour tout réel x pour caractériser la loi de X . Ceci justifie la définition suivante :

Définition. Soit X une variable aléatoire réelle. On appelle *fonction de répartition* de X , notée F (ou F_X si besoin), la fonction suivante :

$$F : \mathbb{R} \longrightarrow [0, 1], \quad x \longmapsto F(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \leq x).$$

La terminologie « fonction de répartition de X » est un peu impropre car en fait F dépend de la mesure \mathbb{P}_X et pas directement de X . Un des intérêts de la fonction de répartition est quelle permet d'exprimer $\mathbb{P}(X \in I)$ pour tout intervalle réel I . C'est l'objet du résultat qui suit.

Proposition 2. Soient X une v.a.r. et F sa fonction de répartition. Alors,

1. F est croissante ;
2. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$;
3. F est continue à droite et possède des limites à gauche ;
4. si a et b sont deux réels tels que $a < b$, on a notant $F(x^-) = \lim_{t \rightarrow x^-} F(t)$,
 - (a) $\mathbb{P}(a < X \leq b) = \mathbb{P}_X(]a, b]) = F(b) - F(a)$;
 - (b) $\mathbb{P}(X = b) = \mathbb{P}_X(\{b\}) = F(b) - F(b^-)$;
 - (c) $\mathbb{P}(a \leq X \leq b) = \mathbb{P}_X([a, b]) = F(b) - F(a^-)$;
 - (d) $\mathbb{P}(a < X < b) = \mathbb{P}_X(]a, b[) = F(b^-) - F(a)$;
 - (e) $\mathbb{P}(a \leq X < b) = \mathbb{P}_X([a, b[) = F(b^-) - F(a^-)$;
 - (f) $\mathbb{P}(X < a) = \mathbb{P}_X(]-\infty, a[) = F(a^-)$.

La preuve de ce résultat n'est pas très difficile et repose sur les propriétés de monotonie des mesures de probabilité ; faisons-la partiellement. La croissance de F est évidente. Il s'en suit – c'est un bon exercice d'analyse sur les bornes sup – que F possède des limites à droite et à gauche en tout point. La continuité à droite s'obtient en écrivant que $] - \infty, x] = \bigcap_{n \geq 1}] - \infty, x + 1/n]$ et en utilisant le fait que la probabilité de l'intersection d'une suite décroissante $(A_n)_{\mathbb{N}}$ est la limite des probabilités des A_n ; en effet,

$$F(x) = \mathbb{P}_X(\bigcap_{n \geq 1}] - \infty, x + 1/n]) = \lim_{n \rightarrow \infty} \mathbb{P}_X(] - \infty, x + 1/n]) = \lim_{n \rightarrow \infty} F(x + 1/n).$$

Les limites résultent respectivement de $\emptyset = \bigcap_{n \geq 1}] - \infty, -n]$ et $\mathbb{R} = \bigcup_{n \geq 1}] - \infty, n]$. On obtient $\mathbb{P}_X(]a, b]) = F(b) - F(a)$ en écrivant $] - \infty, b] =] - \infty, a] \cup]a, b]$. En remarquant que, pour tout réel x , $\{x\} = \bigcap_{n \geq 1}]x - 1/n, x]$, on a, si x est réel, $\mathbb{P}_X(\{x\}) = \lim_{n \rightarrow \infty} \mathbb{P}_X(]x - 1/n, x]) = F(x) - F(x^-)$. Les autres assertions sont alors immédiates : par exemple, si on écrit $]a, b] =]a, b[\cup \{b\}$, on obtient

$$\mathbb{P}_X(]a, b]) = \mathbb{P}_X(]a, b[) + \mathbb{P}_X(\{b\}), \quad \text{soit,} \quad F(b) - F(a) = \mathbb{P}_X(]a, b[) + F(b) - F(b^-),$$

d'où l'on déduit immédiatement que $\mathbb{P}(a < X < b) = \mathbb{P}_X(]a, b[) = F(b^-) - F(a)$.

Une conséquence de ce résultat est que, F est continue en un point x si et seulement si $\mathbb{P}_X(\{x\}) = 0$. Notons également que comme F est croissante, ses points de discontinuité forment un ensemble au plus dénombrable.

Exemple. Soit X une variable aléatoire réelle de fonction de répartition F donnée par, $[x]$ désignant la partie entière de x ,

$$F(x) = \frac{1}{5} \exp(x-1), \text{ si } x < 1, \quad F(x) = \frac{1}{4}[x], \text{ si } 1 \leq x < 3, \quad F(x) = 1 - \frac{1}{2x}, \text{ sinon.}$$

On obtient facilement les probabilités suivantes (un dessin de F aide beaucoup),

$$\mathbb{P}(X = 0) = 0, \quad \mathbb{P}(X = 2) = F(2) - F(2^-) = 1/2 - 1/4 = 1/4,$$

de même que, notant que $F(3) = 5/6$, $F(1) = 1/4$ et $F(1^-) = 1/5$,

$$\mathbb{P}(1 < X \leq 3) = F(3) - F(1) = 7/12, \quad \mathbb{P}(1 \leq X \leq 3) = F(3) - F(1^-) = 19/30.$$

Nous en venons à un résultat très profond, qui dépasse le cadre de ce cours. Il montre en fait que la loi d'une v.a.r. est entièrement caractérisée par sa fonction de répartition.

Théorème 3. *La loi \mathbb{P}_X d'une variable aléatoire réelle X est entièrement déterminée par sa fonction de répartition F .*

Plus précisément, si $F : \mathbb{R} \rightarrow [0, 1]$ est une fonction croissante, continue à droite vérifiant $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$, il existe une unique probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$\forall x \in \mathbb{R}, \quad \mu(]-\infty, x]) = F(x).$$

On déduit notamment de ce théorème qu'une probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est caractérisée par $\mu(I)$ pour I intervalle.

Exemple. Soit F la fonction définie par $F(x) = 0$ si $x < 0$, $F(x) = x$ si $0 \leq x < 1$, $F(x) = 1$ si $x \geq 1$. D'après le résultat précédent, il existe une unique mesure de probabilité sur \mathbb{R} de fonction de répartition F . On appelle cette probabilité la loi uniforme sur $[0, 1]$ et on la note parfois m . Elle modélise le fait de tirer un nombre au hasard dans $[0, 1]$.

Nous allons à présent nous intéresser à deux types particuliers de v.a.r. : les variables aléatoires discrètes d'une part et d'autre part les variables aléatoires absolument continues.

3. Variables discrètes.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle définie sur Ω .

Définition. X est une variable aléatoire discrète si l'ensemble $X(\Omega)$ est au plus dénombrable (fini ou dénombrable) : il existe une suite de réels $(x_i)_{i \in \mathbb{N}}$ distincts tels que $X(\Omega) = \{x_i, i \in \mathbb{N}\}$.

Remarque. Si on note, pour $i \in \mathbb{N}$, $p_i = \mathbb{P}(X = x_i)$ on doit avoir $\sum_{i \in \mathbb{N}} p_i = 1$. Certains des p_i peuvent être nuls.

Exemple. Imaginons que l'on lance un dé plusieurs fois de suite.

Si X représente le résultat du premier lancer, $X(\Omega) = \{1, \dots, 6\}$ et $p_1 = \dots = p_6 = 1/6$.

Si X représente à présent le premier lancer pour lequel on obtient 1, alors $X(\Omega) = \mathbb{N}^*$ et pour $k \in \mathbb{N}^*$, $p_k = \mathbb{P}(X = k) = (5/6)^{k-1}(1/6)$.

Nous allons voir que la loi d'une v.a.r. discrète est relativement simple à caractériser. Tout d'abord, pour x réel déterminons l'ensemble $\{X \leq x\}$. Comme $X(\Omega) = \{x_i, i \in \mathbb{N}\}$, on a

$$\{X \leq x\} = \bigcup_{i \in \mathbb{N}} \{X \leq x\} \cap \{X = x_i\} = \bigcup_{i \in \mathbb{N}: x_i \leq x} \{X = x_i\}.$$

On déduit de cette égalité ensembliste deux choses : tout d'abord, X est mesurable si et seulement si les ensembles $\{X = x_i\}$ appartiennent à \mathcal{F} pour tout $i \in \mathbb{N}$; et d'autre part, en utilisant la σ -additivité de la probabilité \mathbb{P} ,

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\bigcup_{i \in \mathbb{N}: x_i \leq x} \{X = x_i\}\right) = \sum_{i \in \mathbb{N}: x_i \leq x} \mathbb{P}(X = x_i) = \sum_{i \in \mathbb{N}} p_i \mathbf{1}_{x_i \leq x}.$$

On voit alors que la fonction de répartition de X – et par suite sa loi – est entièrement déterminée par les réels $p_i = \mathbb{P}(X = x_i)$. On obtient donc le résultat suivant :

Théorème 4. *Soit X une v.a.r. discrète prenant les valeurs $(x_i)_{i \in \mathbb{N}}$. La loi de X est complètement déterminée par les valeurs*

$$p_i = \mathbb{P}(X = x_i) = \mathbb{P}_X(\{x_i\}), \quad i \in \mathbb{N}.$$

Plus précisément, on a pour tout borélien B ,

$$\mathbb{P}(X \in B) = \sum_{i \in \mathbb{N}} p_i \mathbf{1}_B(x_i).$$

La dernière égalité signifie simplement que $\mathbb{P}(X \in B)$ s'obtient en faisant la somme des p_i sur les indices i tels que $x_i \in B$ on note $\mathbb{P}(X \in B) = \sum_{x_i \in B} p_i$.

Pour une variable discrète, quelle allure à le graphe de la fonction F ? Nous avons déjà dit, avec les notations du théorème précédent, que $F(x) = \sum_i p_i \mathbf{1}_{x_i \leq x}$. Si on suppose que les x_i sont classés par ordre croissant, on a pour $x < x_0$, $F(x) = 0$, puis pour $x \in [x_0, x_1[$, $F(x) = p_0$, puis sur $[x_1, x_2[$, $F(x) = p_0 + p_1 \dots F$ est donc une fonction constante par morceaux.

Exemple. Soit X une variable aléatoire prenant les valeurs $-2, 1, 2$ avec probabilité respectives $\mathbb{P}(X = -2) = 1/3$, $\mathbb{P}(X = 1) = 1/2$ et $\mathbb{P}(X = 2) = 1/6$. La fonction de répartition de X est donnée par

$$F(x) = 0 \text{ si } x < -2, \quad F(x) = \frac{1}{3} \text{ si } x \in [-2, 1[, \quad F(x) = \frac{5}{6} \text{ si } x \in [1, 2[, \quad F(x) = 1 \text{ sinon.}$$

Finissons ce paragraphe en précisant les lois de quelques variables aléatoires discrètes que nous utiliserons par la suite.

Loi uniforme sur $\{x_1, \dots, x_N\}$: pour tout $i = 1, \dots, N$, $\mathbb{P}(X = x_i) = 1/N$. Les expériences aléatoires correspondant à cette situation ne manquent pas : dé équilibré, tirer une carte au hasard, pile ou face avec une pièce non truquée ...

Loi de Bernoulli $\mathcal{B}(p)$: on se donne $p \in [0, 1]$ et $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = q = 1 - p$. Les exemples sont là encore nombreux : sexe d'un enfant à naître, pile ou face truqué, ...

Loi binomiale $\mathcal{B}(n, p)$: on fixe $n \in \mathbb{N}^*$ et $p \in [0, 1]$. X prend les valeurs $0, \dots, n$ et pour tout $k = 0, \dots, n$, on pose $\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k}$. L'expérience aléatoire correspondante est la suivante : on a une épreuve de base (par exemple obtenir pile) qui est réussie avec probabilité p ; on répète n fois l'expérience de façon indépendante et X représente le nombre de succès.

Loi géométrique $\mathcal{G}(p)$: $X(\Omega) = \mathbb{N}^*$, $p \in]0, 1[$. Pour $k \in \mathbb{N}^*$, $\mathbb{P}(X = k) = p(1 - p)^{k-1}$. On répète de façon indépendante, un pile ou face truqué qui donne pile avec probabilité p . X représente alors le premier instant où l'on obtient pile.

Loi de Poisson $\mathcal{P}(\lambda)$: $X(\Omega) = \mathbb{N}$, $\lambda > 0$. Pour tout $k \in \mathbb{N}$, $\mathbb{P}(X = k) = e^{-\lambda} \lambda^k / k!$.

Loi binomiale négative $\mathcal{B}_-(n, p)$: on a $p \in]0, 1[$ et $n \in \mathbb{N}^*$, $X(\Omega) = \{n, n + 1, n + 2, \dots\}$ et on définit pour $k \geq n$, $\mathbb{P}(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}$. L'expérience est la suivante : on lance une pièce donnant pile avec probabilité p de façon indépendante. X représente le premier instant où l'on obtient n fois pile.

Exercice. Montrer que les différentes lois \mathbb{P}_X ainsi définies sont bien des probabilités.

4. Variables absolument continues.

Dans ce paragraphe, on considère le cas des variables aléatoires réelles absolument continues ; commençons par une définition.

Définition. Soit $p : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. p est une densité de probabilité si

- (i) p est positive ;
- (ii) p est intégrable et $\int_{\mathbb{R}} p(t) dt = 1$.

En toute rigueur, il faudrait donner un sens précis au mot « intégrable ». Disons simplement, que cette notion d'intégrabilité – qui comprend entre autre le fait que p est borélienne c'est à dire que pour tout $a \in \mathbb{R}$, $\{x \in \mathbb{R}, p(x) \leq a\} \in \mathcal{B}(\mathbb{R})$ – correspond avec la notion classique d'intégrabilité lorsque p est, par exemple, continue par morceaux.

Exercice. Vérifier que la fonction $p(x) = \exp(-|x|)/2$ est une densité de probabilité.

Définition. Soit X une variable aléatoire réelle de fonction de répartition F . On dit que X est absolument continue s'il existe une densité de probabilité p telle que :

$$\forall x \in \mathbb{R}, \quad F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x p(t) dt.$$

Dans ce cas, on dit que X a pour densité p ou encore que p est une densité de X .

Si X est absolument continue de densité p , on a pour tout intervalle, toute réunion d'intervalles (et même tout borélien) B ,

$$\mathbb{P}(X \in B) = \int_B p(t) dt.$$

La question qui se pose à présent est la suivante : comment reconnaître qu'une v.a.r. X possède une densité ? Voici un résultat qui précise les liens entre fonction de répartition et densité de probabilité.

Théorème 5. Soit X une variable aléatoire réelle de fonction de répartition F .

- Si X possède une densité, alors F est continue et donc $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$.
- Si F est continue et C^1 par morceaux alors X possède une densité et la fonction p définie par $p(x) = F'(x)$ si F est dérivable au point x et $p(x) = 0$ sinon en est une.

On finit ce paragraphe en donnant les densités de quelques variables aléatoires absolument continues.

Loi uniforme sur $[a, b]$, $\mathcal{U}(a, b)$: X de densité $p(x) = (b - a)^{-1} \mathbf{1}_{[a, b]}(x)$;

Loi de Cauchy : X de densité $p(x) = 1/(\pi(1 + x^2))$;

Loi de Laplace : X de densité $p(x) = e^{-|x|}/2$;

Loi exponentielle, $\mathcal{Exp}(\lambda)$, $\lambda > 0$: X de densité $p(x) = \lambda \exp(-\lambda x) \mathbf{1}_{x \geq 0}$;

Loi de Gauss, $\mathcal{N}(\mu, \sigma^2)$: on dit aussi gaussienne de moyenne μ et de variance $\sigma^2 > 0$. X de densité $p(x) = (2\pi\sigma^2)^{-1/2} \exp(- (x - \mu)^2/(2\sigma^2))$. Lorsque $\mu = 0$ et $\sigma^2 = 1$, on dit loi normale centrée réduite.

Exercice. Vérifier que les fonctions précédentes sont des densités de probabilité et calculer les fonctions de répartition associées pour les quatre premières.

5. Calcul de loi.

Le problème est le suivant : X est une variable aléatoire réelle définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ de loi \mathbb{P}_X connue. On note $E = X(\Omega)$. Soit u une fonction (borélienne) de \mathbb{R} dans \mathbb{R} définie au moins sur E . On veut déterminer la loi \mathbb{P}_Y de la v.a. $Y = u(X)$. On note $E' = Y(\Omega)$.

Dans un tel problème, la première chose à faire est de déterminer la nature de la variable Y . Par nature, il faut entendre E' au plus dénombrable ce qui signifie que Y est discrète ou E' de type intervalle ou \mathbb{R} .

5.1. E' au plus dénombrable. Dans cette situation, on a $E' = \{y_i, i \in \mathbb{N}\}$. Pour caractériser \mathbb{P}_Y , on doit calculer $\mathbb{P}_Y(\{y_i\})$ pour tout i . Il faut donc trouver $u^{-1}(\{y_i\})$ pour tout i puisque

$$\forall i \in \mathbb{N}, \quad \mathbb{P}_Y(\{y_i\}) = \mathbb{P}(Y = y_i) = \mathbb{P}(u(X) = y_i) = \mathbb{P}(X \in u^{-1}(\{y_i\})) = \mathbb{P}_X(u^{-1}(\{y_i\})).$$

Exemple. Soit X une v.a. à valeurs dans \mathbb{Z}^* de loi donnée par $\mathbb{P}(X = n) = 2^{-(1+|n|)}$ pour $n \in \mathbb{Z}^*$. Quelle est la loi de la v.a. $Y = |X|$?

Y est à valeurs dans \mathbb{N}^* et de plus pour $n \in \mathbb{N}^*$,

$$\mathbb{P}(Y = n) = \mathbb{P}(X = \pm n) = \mathbb{P}(X = n) + \mathbb{P}(X = -n) = 2^{-n}.$$

Y suit une loi géométrique de paramètre $1/2$.

Exemple. Soit X une v.a. suivant la loi uniforme sur $[0, 1]$. Quelle est la loi de $Y = [10X]$ ($[x]$ désigne la partie entière de x) ?

Y est à valeurs dans $E' = \{0, 1, \dots, 9\}$. De plus, si $k \in E'$, on a

$$\mathbb{P}(Y = k) = \mathbb{P}(10X \in [k, k + 1]) = \mathbb{P}(X \in [k/10, k/10 + 1/10]) = 1/10.$$

On en déduit que Y suit la loi uniforme sur E' .

5.2. E' non dénombrable. On connaît la probabilité \mathbb{P}_X , au travers de sa fonction de répartition F_X ou d'une densité p_X , et on cherche la fonction de répartition de $Y = u(X)$. Pour y réel, on a

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(u(X) \in]-\infty, y]) = \mathbb{P}(X \in u^{-1}(]-\infty, y])) = \mathbb{P}_X(u^{-1}(]-\infty, y])) ;$$

on détermine ensuite l'ensemble $u^{-1}(]-\infty, y])$ et on obtient la probabilité requise.

Exemple. Soit X une v.a. de loi $\mathcal{Exp}(\lambda)$, $\lambda > 0$; rappelons que $F_X(x) = 1 - \exp(-\lambda x)$ si $x \geq 0$ et $F_X(x) = 0$ dès que $x < 0$. Considérons $Y = X \mathbf{1}_{X \leq 1}$ qui correspond à la fonction $u(x) = x \mathbf{1}_{]-\infty, 1]}}(x)$.

On a $u^{-1}(]-\infty, y]) =]-\infty, y]$ si $y < 0$ et $u^{-1}(]-\infty, y]) =]-\infty, y] \cup]1, +\infty[$ si $y \geq 0$. Il vient alors, $F_Y(y) = 0$ si $y < 0$, $F_Y(y) = 1$ si $y > 1$ et

$$\forall y \in [0, 1], \quad F_Y(y) = F_X(y) + 1 - F_X(1) = 1 - \exp(-\lambda y) + \exp(-\lambda).$$

On peut noter que Y n'est pas absolument continue puisque F_Y n'est pas continue au point 0.

Exemple. Soit X une v.a. uniforme sur $[-\pi/2, \pi/2]$. On considère $Y = \tan(X)$. \tan est une fonction strictement croissante de $]-\pi/2, \pi/2[$ dans \mathbb{R} . Si $y \in \mathbb{R}$, comme X est dans $]-\pi/2, \pi/2[$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\tan(X) \leq y) = \mathbb{P}(X \leq \arctan y) = F_X(\arctan y).$$

Il vient alors, pour tout réel y ,

$$F_Y(y) = \frac{1}{\pi} \int_{-\infty}^{\arctan y} \mathbf{1}_{[-\pi/2, \pi/2]}(x) dx = \frac{1}{2} + \frac{1}{\pi} \arctan y.$$

Y suit la loi de Cauchy.

6. Variables indépendantes.

Avant de donner une définition plus générale, regardons le cas de deux variables aléatoires. Soient X et Y deux v.a.r. définies sur le même espace probabilisé. X et Y sont indépendantes si

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

Cette définition se généralise comme suit :

Définition. Soient $(X_i)_{1 \leq i \leq n}$ n variables aléatoires définies sur le même espace probabilisé. Les v.a. $(X_i)_{1 \leq i \leq n}$ sont indépendantes si

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

Remarque. Dès que l'une des v.a. est discrète, disons X_1 , on peut remplacer dans la condition (1) « $X_1 \leq x_1$ » par « $X_1 = v$ » pour chacune des valeurs v que prend X_1 .

De plus, si $(X_i)_{1 \leq i \leq n}$ sont des variables indépendantes alors on a

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i)$$

pour B_i intervalle, réunion d'intervalles et même borélien.

Pour finir, signalons qu'une suite de v.a. $(X_n)_{n \in \mathbb{N}}$ est dite indépendante (ou une suite de v.a. indépendantes) si, pour tout n , $(X_i)_{0 \leq i \leq n}$ sont $n + 1$ v.a. indépendantes.

Espérance d'une variable aléatoire

L'objectif de ce paragraphe est de définir ce qu'est la valeur moyenne d'une variable aléatoire. Par exemple, quel est le gain moyen d'un joueur au Loto? En d'autres termes, quelle somme un jour de Loto peut-il espérer gagner en moyenne?

1. Définitions.

Soit X une variable aléatoire réelle définie sur $(\Omega, \mathcal{F}, \mathbb{P})$. On note $E = X(\Omega)$. La définition générale de l'espérance de X , que l'on note $\mathbb{E}[X]$, fait appel à la théorie abstraite de l'intégration : pour mémoire,

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x dF(x), \quad F \text{ fonction de répartition de } X.$$

Pour éviter cette théorie, nous allons définir l'espérance d'une variable aléatoire seulement dans deux cas particuliers : pour les variables discrètes d'une part et pour les variables absolument continues d'autre part.

1.1. X v.a.r. discrète. On suppose que X prend les valeurs $(x_i)_{\mathbb{N}}$ i.e. $E = \{x_i, i \in \mathbb{N}\}$. On note toujours, pour $i \in \mathbb{N}$, $p_i = \mathbb{P}(X = x_i)$. Considérons d'autre part f une fonction réelle.

Définition. On dit que la variable aléatoire $f(X)$ est intégrable quand $\sum_{i \geq 0} |f(x_i)| p_i < +\infty$. Dans ce cas, la série $\sum_{i \geq 0} f(x_i) p_i$ converge et on appelle cette quantité espérance ou moyenne de $f(X)$ soit

$$\mathbb{E}[f(X)] = \sum_{i \geq 0} f(x_i) p_i = \sum_{i \geq 0} f(x_i) \mathbb{P}(X = x_i).$$

Un cas particulier important de cette définition est celui de la fonction $f(x) = x$: la moyenne d'une v.a. discrète X est donnée par

$$\mathbb{E}[X] = \sum_{i \geq 0} x_i p_i = \sum_{i \geq 0} x_i \mathbb{P}(X = x_i),$$

lorsque $\sum_{i \geq 0} |x_i| p_i < +\infty$.

Remarque. Le cas E fini est comme à l'habitude inclus dans ce formalisme et dans ce cas les séries sont simplement des sommes finies.

Exemple. Calculs d'espérances de variables aléatoires discrètes.

1° Commençons par une v.a. X de loi uniforme sur $\{x_1, \dots, x_N\}$; pour tout $i = 1, \dots, N$, $\mathbb{P}(X = x_i) = 1/N$ et par suite

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

C'est la moyenne arithmétique des nombres x_1, \dots, x_N .

2° Si X est une v.a.r. de Poisson de paramètre $\lambda > 0$, on a comme $E = \mathbb{N}$,

$$\mathbb{E}[X] = \sum_{i \geq 0} i e^{-\lambda} \lambda^i / i! = \lambda e^{-\lambda} \sum_{i \geq 1} \lambda^{i-1} / (i-1)! = \lambda.$$

3° Prenons pour fonction f la fonction $\mathbf{1}_{\{x_p\}}$ pour $p \in \mathbb{N}$ fixé. On obtient dans ce cas

$$\mathbb{E}[\mathbf{1}_{\{x_p\}}(X)] = \sum_{i \geq 0} \mathbf{1}_{\{x_p\}}(x_i) \mathbb{P}(X = x_i).$$

Or, les valeurs x_i étant distinctes, $\mathbf{1}_{\{x_p\}}(x_i) = 0$ si $i \neq p$ et 1 si $i = p$; il vient alors

$$\mathbb{E}[\mathbf{1}_{\{x_p\}}(X)] = \mathbb{P}(X = x_p).$$

4° Soit X une v.a.r. suivant la loi de Poisson de paramètre $\lambda > 0$. Calculons l'espérance de la v.a. $1/(1+X)$ qui est bien définie puisque X est à valeurs dans \mathbb{N} . On a

$$\mathbb{E}\left[\frac{1}{X+1}\right] = \sum_{k \geq 0} \frac{1}{k+1} \mathbb{P}(X = k) = e^{-\lambda} \sum_{k \geq 0} \frac{\lambda^k}{(k+1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{k \geq 0} \frac{\lambda^{k+1}}{(k+1)!} = \frac{e^{-\lambda}}{\lambda} \sum_{i \geq 1} \frac{\lambda^i}{i!}.$$

On obtient finalement

$$\mathbb{E}\left[\frac{1}{X+1}\right] = \frac{e^{-\lambda}}{\lambda} \left(\sum_{i \geq 0} \frac{\lambda^i}{i!} - 1 \right) = \frac{e^{-\lambda}}{\lambda} (e^\lambda - 1) = (1 - e^{-\lambda}) / \lambda.$$

1.2. X absolument continue. Soit p une densité de la variable X et $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue par morceaux (ou borélienne).

Définition. La variable $f(X)$ est intégrable si l'intégrale $\int_{\mathbb{R}} |f(x)| p(x) dx$ est finie. Dans ce cas, $\int_{\mathbb{R}} f(x) p(x) dx$ est convergente et, la valeur de cette intégrale est appelée espérance de $f(X)$; on la note $\mathbb{E}[f(X)]$ soit

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

Exemple. Comme dans le cas discret l'espérance ou la moyenne d'une variable aléatoire X correspond au cas $f(x) = x$. Voici quelques exemples de calculs.

1° Si X suit la loi uniforme sur $[a, b]$ alors X est intégrable et

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x dx = (b+a)/2.$$

2° Si X suit la loi de Cauchy, X n'est pas intégrable car

$$\ln(1+n^2) = \int_{-n}^n |x|/(1+x^2) dx \leq \int_{\mathbb{R}} |x|/(1+x^2) dx.$$

3° Dans le cas où $f(x) = \mathbf{1}_I(x)$ avec I intervalle, on a

$$\mathbb{E}[f(X)] = \mathbb{E}[\mathbf{1}_I(X)] = \int_{\mathbb{R}} \mathbf{1}_I(x) p(x) dx = \int_I p(x) dx = \mathbb{P}_X(I) = \mathbb{P}(X \in I).$$

Le calcul précédent demeure vrai si on remplace l'intervalle I par un borélien.

4° Soit X une v.a. de loi $\mathcal{Exp}(\lambda)$, $\lambda > 1$. On a

$$\mathbb{E}[\exp(X)] = \int_0^\infty \exp(x) \lambda \exp(-\lambda x) dx = \lambda / (\lambda - 1).$$

Calcul de loi. Reprenons le problème du calcul de lois image à l'aide de cette nouvelle notion. Supposons que X soit une v.a.r. absolument continue de densité $p_X(x) = \mathbf{1}_S(x)p(x)$ où S est le support de p_X disons un intervalle fermé pour simplifier. D est l'intérieur de S c'est à dire l'intervalle ouvert associé à S (si $S = [0, 1]$, $D =]0, 1[$, si $S = \mathbb{R}_+$, $D = \mathbb{R}_+^*$). On cherche la loi de $Y = u(X)$ et on a l'intuition que Y est une v.a.r. absolument continue. Pour caractériser sa loi on cherche donc à déterminer une densité p_Y . Pour cela, on prend f une fonction continue par morceaux et bornée et on calcule $\mathbb{E}[f(Y)]$ (qui existe car f bornée) de deux façons différentes : tout d'abord, si Y a pour densité la fonction p_Y , on a

$$\mathbb{E}[f(Y)] = \int_{\mathbb{R}} f(y) p_Y(y) dy.$$

D'autre part, $Y = u(X)$ et comme X a pour densité $p_X(x) = \mathbf{1}_S(x)p(x)$,

$$\mathbb{E}[f(Y)] = \mathbb{E}[f(u(X))] = \int_{\mathbb{R}} f(u(x)) p_X(x) dx = \int_S f(u(x)) p(x) dx = \int_D f(u(x)) p(x) dx.$$

Si u est un C^1 -difféomorphisme de D dans $u(D)$, il vient via le changement de variables $y = u(x)$,

$$\mathbb{E}[f(Y)] = \int_{u(D)} f(y) p(u^{-1}(y)) |(u^{-1})'(y)| dy = \int_{\mathbb{R}} f(y) \mathbf{1}_{u(D)}(y) p(u^{-1}(y)) |(u^{-1})'(y)| dy.$$

Par identification,

$$p_Y(y) = \mathbf{1}_{u(D)}(y) p(u^{-1}(y)) |(u^{-1})'(y)|.$$

Exemple. Calculons la loi de la v.a. $Y = -\ln(X)/\lambda$ où X suit la loi uniforme sur $[0, 1]$ et $\lambda > 0$. Remarquons tout d'abord que Y est bien définie puisque $\mathbb{P}(X > 0) = 1$.

Soit donc f une fonction continue par morceaux et bornée ; comme X a pour densité $\mathbf{1}_{[0,1]}(x)$, on a

$$\mathbb{E}[f(Y)] = \mathbb{E}[f(-\ln(X)/\lambda)] = \int_0^1 f(-\ln(x)/\lambda) dx.$$

Effectuons le changement de variables $y = -\ln(x)/\lambda$ soit $x = \exp(-\lambda y)$; il vient

$$\mathbb{E}[f(Y)] = \int_{+\infty}^0 f(y) (-\lambda) \exp(-\lambda y) dy = \int_0^{+\infty} f(y) \lambda \exp(-\lambda y) dy,$$

soit encore

$$\mathbb{E}[f(Y)] = \int_{\mathbb{R}} f(y) \lambda \exp(-\lambda y) \mathbf{1}_{y \geq 0} dy.$$

On en déduit que Y a pour densité la fonction $y \mapsto \lambda \exp(-\lambda y) \mathbf{1}_{y \geq 0}$ c'est à dire que Y suit la loi exponentielle de paramètre λ .

Dans cet exemple, $u(x) = -\ln(x)/\lambda$ et $D = (0, 1)$. u est un C^∞ -difféomorphisme strictement décroissant de $]0, 1[$ dans $u(D) =]0, \infty[$. On a $u^{-1}(y) = \exp(-\lambda y)$ et donc, pour tout $y > 0$, $(u^{-1})'(y) = -\lambda \exp(-\lambda y)$. C'est donc bien $|(u^{-1})'(y)|$ qui apparaît dans la formule de p_Y : « on prend le $-$ pour remettre les bornes dans le bon sens ».

1.3. Vocabulaire. On dit aussi qu'une variable aléatoire X possède un moment d'ordre un pour dire que X est une variable aléatoire intégrable. Si tel est le cas l'espérance de X , $\mathbb{E}[X]$, est aussi appelée moyenne de X . Lorsque X est intégrable et de moyenne nulle i.e. $\mathbb{E}[X] = 0$ on dit que X est centrée.

Plus généralement, une variable aléatoire X , réelle ou complexe, possède un moment d'ordre $n \in \mathbb{N}^*$ si la variable aléatoire $|X|^n$ est intégrable. Dans ce cas, $\mathbb{E}[X^n]$ s'appelle le moment d'ordre n et $\mathbb{E}[(X - \mathbb{E}[X])^n]$ le moment centré d'ordre n .

2. Propriétés de l'espérance.

Passons aux propriétés de l'espérance. Comme déjà dit, les deux définitions de l'espérance données précédemment ne sont que deux cas particuliers d'une définition plus générale. Aussi, il n'est pas surprenant que l'on obtiennent les mêmes propriétés dans les deux cas.

Proposition 1. *Soient X et Y deux v.a.r. intégrables, a et b deux réels. Alors, les v.a.r. $X + Y$ et $aX + b$ sont intégrables et*

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \quad \mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

L'espérance est croissante : si $\mathbb{P}(X \leq Y) = 1$ alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$ et, dans ce cas, $\mathbb{E}[X] = \mathbb{E}[Y]$ si et seulement si $\mathbb{P}(X = Y) = 1$.

On déduit de cette proposition que si X est une v.a.r. positive, $\mathbb{P}(X \geq 0) = 1$, alors $\mathbb{E}[X] \geq 0$. Signalons que si, pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$ alors $\mathbb{P}(X \leq Y) = 1$. C'est pourquoi on retient souvent la croissance de l'espérance sous la forme : $X \leq Y \implies \mathbb{E}[X] \leq \mathbb{E}[Y]$.

Notons aussi que l'espérance d'une constante est égale à cette constante.

2.1. Espérance et indépendance. On considère X et Y deux variables aléatoires réelles et on aimerait par exemple calculer l'espérance du produit XY ; lorsque X et Y sont indépendantes, le résultat est facile à obtenir. En effet,

Théorème 2. *Soient X et Y deux v.a.r. indépendantes et g, h deux fonctions boréliennes.*

Si $g(X)$ et $h(Y)$ sont intégrables, alors $g(X)h(Y)$ est intégrable et

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

On déduit de ce résultat que si X et Y sont deux v.a. indépendantes alors les v.a. $u(X)$ et $v(Y)$ sont encore indépendantes quelque soit les fonctions boréliennes u et v .

Remarque. Les résultats précédents se généralisent au cas de n v.a. indépendantes.

Définition. Deux v.a. réelles X et Y – telles que X^2 et Y^2 sont intégrables – sont dites *non-corrélées* si elles vérifient $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Il faut noter que deux variables indépendantes sont non-corrélées mais la réciproque est fautive comme le montre l'exemple suivant : X et ε deux v.a. indépendantes ; X de loi gaussienne $\mathcal{N}(0, 1)$ et ε de loi donnée par $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. X et εX sont non-corrélées mais ne sont pas indépendantes.

2.2. Variance d'une variable aléatoire. Si X est une variable aléatoire réelle, la première information que l'on cherche est la valeur moyenne, $\mathbb{E}[X]$. Ensuite, on s'intéresse à la dispersion de X autour de cette valeur moyenne : c'est la notion de *variance*.

Définition. Soit X une v.a.r. de carré intégrable i.e. X^2 est intégrable. Alors, X est intégrable et on définit la variance de X , notée $\mathbb{V}[X]$, comme le moment centré d'ordre deux, à savoir,

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

La racine carrée de $\mathbb{V}[X]$ s'appelle l'écart type de X .

Il faut d'abord remarquer que si X^2 est intégrable alors X l'est ; cela résulte de l'inégalité $|X| \leq 1 + X^2$ et de la croissance de l'espérance. Ensuite, $\mathbb{V}[X]$ est une quantité positive puisque c'est l'espérance d'un carré et pour se convaincre de la dernière égalité de la définition, calculons, pour c constante arbitraire réelle, $\mathbb{E}[(X - c)^2]$. On a $(X - c)^2 = X^2 - 2cX + c^2$, et donc d'après les propriétés de l'espérance,

$$\mathbb{E}[(X - c)^2] = \mathbb{E}[X^2] + c^2 - 2c\mathbb{E}[X],$$

si on prend $c = \mathbb{E}[X]$ on obtient la formule annoncée. Ce petit calcul permet aussi de montrer le dernier point de la proposition suivante :

Proposition 3. *Soient X une variable aléatoire réelle de carré intégrable, a et b deux réels. On a : $\mathbb{V}[aX + b] = a^2\mathbb{V}[X]$. De plus, $\mathbb{V}[X] = 0$ si et seulement si X est constante.*

Enfin, la fonction $c \mapsto \mathbb{E}[(X - c)^2]$ possède un minimum au point $c = \mathbb{E}[X]$ et ce minimum vaut $\mathbb{V}[X]$.

Finissons, par une dernière propriété :

Proposition 4. *Soient X et Y deux v.a.r. de carré intégrable, indépendantes ou non-corrélées. Alors $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.*

Exemple. Calculons tout d'abord la variance d'une v.a. X suivant la loi de Poisson de paramètre λ . Nous avons déjà vu que $\mathbb{E}[X] = \lambda$. Il reste à déterminer $\mathbb{E}[X^2]$. Pour cela, notons que

$$\mathbb{E}[X^2] = \mathbb{E}[X(X - 1)] + \mathbb{E}[X] = \sum_{i \geq 0} i(i - 1) e^{-\lambda} \lambda^i / i! + \lambda = \lambda^2 e^{-\lambda} \sum_{i \geq 2} \lambda^{i-2} / (i - 2)! + \lambda.$$

Le changement d'indices $k = i - 2$, donne

$$\mathbb{E}[X^2] = \lambda^2 + \lambda, \quad \text{et donc, par suite,} \quad \mathbb{V}[X] = \lambda.$$

Calculons à présent la variance de X lorsque X suit la loi uniforme sur $[a, b]$. La moyenne de X est $(b + a)/2$. On a de plus

$$\mathbb{E}[X^2] = \frac{1}{b - a} \int_a^b x^2 dx = \frac{1}{3} \frac{b^3 - a^3}{b - a} = \frac{1}{3}(a^2 + ab + b^2).$$

Il vient alors

$$\mathbb{V}[X] = \frac{1}{3}(a^2 + ab + b^2) - \frac{1}{4}(a^2 + 2ab + b^2) = \frac{1}{12}(a^2 - 2ab + b^2) = \frac{(b - a)^2}{12}.$$

3. Fonction caractéristique.

3.1. Généralités. Nous avons déjà dit que la loi d'une v.a.r. X était caractérisée par sa fonction de répartition. On peut aussi déterminer \mathbb{P}_X en utilisant une fonction de $\mathbb{R} \rightarrow \mathbb{C}$: la fonction caractéristique.

Définition. Soit X un v.a.r. On appelle *fonction caractéristique* de X , notée φ (ou φ_X si besoin), la fonction suivante :

$$\varphi : \mathbb{R} \rightarrow \mathbb{C}, \quad t \mapsto \varphi(t) = \mathbb{E}[e^{itX}].$$

Rappelons, tout d'abord que $e^{itX} = \cos(tX) + i \sin(tX)$ est une v.a. de module un (donc intégrable) et que $\varphi(t) = \mathbb{E}[\cos(tX)] + i \mathbb{E}[\sin(tX)]$. On a $\varphi(0) = 1$ et la fonction φ est bornée par un. On peut montrer le résultat suivant :

Proposition 5. *Soit X une v.a.r. Alors φ est uniformément continue et bornée par un. Si de plus X possède un moment d'ordre $n \in \mathbb{N}^*$ alors φ est de classe C^n et*

$$\varphi^{(n)}(0) = i^n \mathbb{E}[X^n].$$

Mais la fonction caractéristique n'a vraiment d'intérêt que dans la mesure où elle détermine la loi de la v.a.r. X . C'est l'objet du théorème suivant :

Théorème 6. *Soient X et Y deux v.a.r. de fonction caractéristique φ_X et φ_Y . \mathbb{P}_X et \mathbb{P}_Y sont égales si et seulement si $\varphi_X(t) = \varphi_Y(t)$ pour tout réel t .*

Exemple. La fonction caractéristique d'une v.a. de Poisson de paramètre λ est :

$$\forall t \in \mathbb{R}, \quad \varphi(t) = e^{-\lambda} \sum_{k \geq 0} e^{itk} \lambda^k / k! = \exp(\lambda(e^{it} - 1)).$$

Une conséquence du Théorème 2 pour les fonctions caractéristiques est la suivante : la fonction caractéristique de la somme de deux v.a. indépendantes est égale au produit de leur fonction caractéristique ; plus précisément

Proposition 7. *Soient X et Y deux v.a. réelles indépendantes. Alors,*

$$\forall t \in \mathbb{R}, \quad \varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t).$$

Exemple. Soient X et Y deux v.a. indépendantes suivant respectivement une loi de Poisson de paramètre λ et une loi de Poisson de paramètre μ . Déterminons la loi de la v.a. $X + Y$ en calculant sa fonction caractéristique. Comme X et Y sont indépendantes, on a, pour $t \in \mathbb{R}$,

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t) = \exp(\lambda(e^{it} - 1)) \exp(\mu(e^{it} - 1)) = \exp((\lambda + \mu)(e^{it} - 1)).$$

On reconnaît la fonction caractéristique d'une loi de Poisson de paramètre $\lambda + \mu$. D'après le Théorème 6, $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$.

3.2. Variables gaussiennes. Commençons par quelques rappels concernant les variables aléatoires gaussiennes.

Définition. Une v.a.r. X est dite normale (ou gaussienne) centrée réduite si elle admet pour densité la fonction

$$\forall x \in \mathbb{R}, \quad p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Si X est une v.a.r. normale centrée réduite, on a $\mathbb{E}[X] = 0$, $\mathbb{V}[X] = 1$, et de plus la fonction caractéristique de X est donnée par

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

Définition. Une v.a.r. Y est dite gaussienne si $Y = \sigma X + \mu$ où σ et μ sont deux réels et X est une v.a.r. normale centrée réduite. On dit que Y suit la loi $\mathcal{N}(\mu, \sigma^2)$.

Les paramètres μ et σ^2 correspondent respectivement à la moyenne et à la variance de Y puisque $\mathbb{E}[Y] = \sigma\mathbb{E}[X] + \mu = \mu$ et $\mathbb{V}[Y] = \sigma^2\mathbb{V}[X] = \sigma^2$.

On montre sans peine que Y est une variable aléatoire gaussienne si et seulement si sa fonction caractéristique est donnée par

$$\forall t \in \mathbb{R}, \quad \varphi_Y(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right).$$

Cette formule montre entre autre que la loi d'une variable aléatoire gaussienne est entièrement caractérisée par sa moyenne et sa variance.

Dans le cas où $\sigma \neq 0$, Y possède pour densité

$$y \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

En effet, si f est borélienne et bornée, on a, via le changement de variables $y = \sigma x + \mu$ (séparez les cas $\sigma > 0$ et $\sigma < 0$)

$$\begin{aligned} \mathbb{E}[f(Y)] = \mathbb{E}[f(\sigma X + \mu)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(\sigma x + \mu) e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(y) \exp\left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma}\right)^2\right) \frac{dy}{|\sigma|}. \end{aligned}$$

Finissons ces rappels par la propriété suivante :

Proposition 8. *Soient X et Y deux gaussiennes indépendantes. Alors $S = X + Y$ est une v.a.r. gaussienne de moyenne $\mathbb{E}[X] + \mathbb{E}[Y]$ et de variance $\mathbb{V}[X] + \mathbb{V}[Y]$.*

Notons μ et σ^2 la moyenne et la variance de X , ν et τ^2 la moyenne et la variance de Y . Comme X et Y sont indépendantes, on a, pour tout réel t ,

$$\varphi_S(t) = \varphi_X(t) \varphi_Y(t) = \exp\left(it\mu - \frac{\sigma^2 t^2}{2}\right) \exp\left(it\nu - \frac{\tau^2 t^2}{2}\right),$$

et donc

$$\varphi_S(t) = \exp\left(it(\mu + \nu) - \frac{(\sigma^2 + \tau^2) t^2}{2}\right),$$

qui est la fonction caractéristique d'une gaussienne de moyenne $\mu + \nu$ et de variance $\sigma^2 + \tau^2$.

3.3. Table des lois connues. Les exemples précédents font apparaître une nouvelle manière de déterminer la loi d'une v.a. : on calcule sa fonction caractéristique et on l'identifie avec la fonction caractéristique d'une loi connue.

Rappelons d'abord, les définitions des lois que nous utilisons. Tout d'abord pour celles qui sont absolument continues, précisons les densités p et les fonctions de répartition F lorsqu'elles sont calculables.

Loi uniforme $\mathcal{U}(a, b)$: $p(x) = (b-a)^{-1} \mathbf{1}_{[a,b]}(x)$, $F(x) = (b-a)^{-1}(x-a)$ si $a \leq x \leq b$,
 $F(x) = 0$ si $x < a$, $F(x) = 1$ si $x > b$;

Loi de Cauchy : $p(x) = 1/(\pi(1+x^2))$, $F(x) = 1/2 + \arctan(x)/\pi$;

Loi de Laplace : $p(x) = e^{-|x|}/2$, $F(x) = e^x/2$ si $x < 0$, $F(x) = 1 - e^{-x}/2$ si $x \geq 0$;

Loi exponentielle, $\mathcal{Exp}(\lambda)$, $\lambda > 0$: $p(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\mathbb{R}_+}(x)$, si $x < 0$ $F(x) = 0$, si $x \geq 0$ $F(x) = 1 - \exp(-\lambda x)$;

Loi gaussienne $\mathcal{N}(\mu, \sigma^2)$: $p(x) = (2\pi\sigma^2)^{-1/2} \exp(-(x - \mu)^2/(2\sigma^2))$.

Dans le cas discret, rappelons les définitions ; si $p \in [0, 1]$, $q = 1 - p$.

Loi de Bernoulli, $\mathcal{B}(p)$, $0 \leq p \leq 1$: $\mathbb{P}(X = 1) = p$, $\mathbb{P}(X = 0) = q$;

Loi binomiale, $\mathcal{B}(n, p)$, $n \geq 1$, $0 \leq p \leq 1$: pour $k = 0, \dots, n$, $\mathbb{P}(X = k) = C_n^k p^k q^{n-k}$;

Loi géométrique, $\mathcal{G}(p)$, $0 < p < 1$: pour $k \in \mathbb{N}^*$, $\mathbb{P}(X = k) = pq^{k-1}$;

Loi binomiale négative, $\mathcal{B}_-(n, p)$, $n \geq 1$, $0 < p < 1$: $\forall k \geq n$, $\mathbb{P}(X = k) = C_{k-1}^{n-1} p^n q^{k-n}$;

Loi de Poisson, $\mathcal{P}(\lambda)$, $\lambda > 0$: pour $k \in \mathbb{N}$, $\mathbb{P}(X = k) = e^{-\lambda} \lambda^k / k!$

Le tableau ci-dessous rappelle la moyenne, la variance, la fonction caractéristique des lois les plus courantes.

Loi / v.a.	Notation	Espérance	Variance	$\varphi(t) = \mathbb{E}[e^{itX}]$
Bernoulli	$\mathcal{B}(p)$	p	pq	$q + pe^{it}$
Binomiale	$\mathcal{B}(n, p)$	np	npq	$(q + pe^{it})^n$
Géométrique	$\mathcal{G}(p)$	$1/p$	q/p^2	$pe^{it}/(1 - qe^{it})$
Binomiale négative	$\mathcal{B}_-(n, p)$	n/p	nq/p^2	$(pe^{it}/(1 - qe^{it}))^n$
Poisson	$\mathcal{P}(\lambda)$	λ	λ	$\exp(\lambda(e^{it} - 1))$
Uniforme	$\mathcal{U}(a, b)$	$(a + b)/2$	$(b - a)^2/12$	$(e^{itb} - e^{ita}) / (it(b - a))$
Cauchy		non	non	$e^{- t }$
Laplace		0	2	$(1 + t^2)^{-1}$
Exponentielle	$\mathcal{Exp}(\lambda)$	$1/\lambda$	$1/\lambda^2$	$\lambda/(\lambda - it)$
Gaussienne	$\mathcal{N}(\mu, \sigma^2)$	μ	σ^2	$\exp(it\mu - \sigma^2 t^2/2)$

Couples aléatoires

Au chapitre précédent, nous avons étudié les variables aléatoires réelles c'est à dire les variables aléatoires prenant leurs valeurs dans \mathbb{R} . Dans cette partie du cours, on va considérer le cas de vecteurs aléatoires c'est à dire des variables aléatoires à valeurs dans \mathbb{R}^d où $d \in \mathbb{N}^*$. Pour simplifier l'exposé, nous considérerons seulement le cas de la dimension $d = 2$. Un vecteur aléatoire Z de \mathbb{R}^2 sera décrit dans la suite par son abscisse X et son ordonnée Y i.e. $Z = (X, Y)$. On utilise aussi le terme « couple aléatoire » pour un vecteur aléatoire de dimension 2.

1. Loi d'un couple aléatoire.

On va s'intéresser à la loi \mathbb{P}_Z d'un couple de v.a.r. $Z = (X, Y)$. On pourrait penser que si l'on connaît la loi de chacune des v.a.r. X et Y alors on connaît la loi du couple ; mais la situation est plus complexe.

1.1. Un exemple. Considérons un couple aléatoire Z ne prenant que quatre valeurs $(0, 0)$, $(0, 1)$, $(1, 0)$ et $(1, 1)$. On a, alors,

$$\mathbb{P}(X = 0) = \mathbb{P}[Z = (0, 0)] + \mathbb{P}[Z = (0, 1)], \quad \mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0),$$

$$\mathbb{P}(Y = 0) = \mathbb{P}[Z = (0, 0)] + \mathbb{P}[Z = (1, 0)], \quad \mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0).$$

Supposons tout d'abord que Z suit la loi uniforme c'est à dire

$$\mathbb{P}[Z = (0, 0)] = \mathbb{P}[Z = (0, 1)] = \mathbb{P}[Z = (1, 0)] = \mathbb{P}[Z = (1, 1)] = 1/4.$$

On a, dans ce cas

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2, \quad \mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = 1/2.$$

Si à présent, on suppose que

$$\mathbb{P}[Z = (0, 0)] = 1/8, \quad \mathbb{P}[Z = (0, 1)] = 3/8, \quad \mathbb{P}[Z = (1, 0)] = 3/8, \quad \mathbb{P}[Z = (1, 1)] = 1/8,$$

on obtient toujours

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2, \quad \mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = 1/2.$$

Dans les deux cas que l'on vient de considérer, la loi de X et Y est la loi uniforme sur $\{0, 1\}$ et pourtant la loi de $Z = (X, Y)$ n'est pas la même dans le premier et dans le second cas.

Deux enseignements semblent se dégager de cet exemple : tout d'abord si on connaît la loi du vecteur $Z = (X, Y)$, on peut déterminer celle des v.a. X et Y . Par contre la connaissance de la loi de X et de la loi de Y ne permet pas de déterminer la loi du vecteur $Z = (X, Y)$ en général.

1.2. Vocabulaire. Donnons à présent les définitions précises relatives aux couples aléatoires ; ces définitions généralisent celles concernant les variables aléatoires réelles.

Définition. On appelle *couple aléatoire* – ou vecteur aléatoire de dimension deux – toute paire $Z = (X, Y)$ de variables aléatoires réelles. On a donc :

$$Z : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow \mathbb{R}^2, \quad \omega \longmapsto Z(\omega) = (X(\omega), Y(\omega)).$$

Sur \mathbb{R} , nous avons considéré la plus petite tribu engendrée par les intervalles i.e. la tribu borélienne. Pour travailler dans \mathbb{R}^2 , on introduit la *tribu borélienne de \mathbb{R}^2* , $\mathcal{B}(\mathbb{R}^2)$, qui est « la plus petite tribu » sur \mathbb{R}^2 contenant tous les pavés $I \times J$ où I et J sont deux intervalles de \mathbb{R} . Cette tribu contient tous les pavés du type $A \times B$ où A et B sont deux boréliens de \mathbb{R} mais elle contient également des ensembles beaucoup plus complexes qui ne sont pas des produits cartésiens.

Si Z est un couple aléatoire, on montre que pour tout borélien C de \mathbb{R}^2 , l'image réciproque de C par Z , $Z^{-1}(C) = \{\omega \in \Omega, Z(\omega) \in C\}$, est un élément de \mathcal{F} . Ceci donne un sens à la définition suivante.

Définition. Soit $Z : \Omega \longrightarrow \mathbb{R}^2$ un couple aléatoire. L'application \mathbb{P}_Z définie par

$$\mathbb{P}_Z : \mathcal{B}(\mathbb{R}^2) \longrightarrow [0, 1], \quad C \longmapsto \mathbb{P}_Z(C) = \mathbb{P}(Z^{-1}(C))$$

est une mesure de probabilité appelée loi de Z .

Remarque. Si C est un borélien de \mathbb{R}^2 et $Z = (X, Y)$ un couple aléatoire, l'ensemble $Z^{-1}(C)$, noté $\{Z \in C\}$ par les probabilistes, n'est pas toujours très facile à déterminer. Néanmoins, si C est le pavé $A \times B$ la situation est plus favorable ; en effet, on a dans ce cas

$$Z^{-1}(C) = \{\omega \in \Omega, (X(\omega), Y(\omega)) \in A \times B\} = X^{-1}(A) \cap Y^{-1}(B),$$

soit encore $\{Z \in A \times B\} = \{X \in A, Y \in B\}$.

La tribu des boréliens de \mathbb{R}^2 contient des parties complexes ; toutefois, elle est engendrée par la classe des produits d'intervalles réels, ce qui conduit au résultat suivant :

Proposition 1. Soit $Z = (X, Y)$ un couple aléatoire. La loi \mathbb{P}_Z est caractérisée par

$$\mathbb{P}_Z(I \times J) = \mathbb{P}(X \in I, Y \in J),$$

pour tout couple (I, J) d'intervalles réels.

Lois marginales du couple. Si $Z = (X, Y)$ est un couple aléatoire, les lois marginales de Z sont les lois des v.a.r. X et Y . Il est important de noter que si on connaît la loi du couple \mathbb{P}_Z on connaît les lois marginales. En effet, si A est un intervalle, une réunion d'intervalles ou un borélien de \mathbb{R} ,

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

de même, pour la v.a.r. Y ,

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(X \in \mathbb{R}, Y \in B) = \mathbb{P}_Z(\mathbb{R} \times B).$$

2. Le cas discret.

Nous examinons à présent le cas où Z est une couple aléatoire discret : X et Y sont donc deux v.a.r. discrètes prenant respectivement les valeurs $\{x_i, i \in \mathbb{N}\}$ et $\{y_n, n \in \mathbb{N}\}$. Remarquons que Z prend les valeurs $\{(x_i, y_n), (i, n) \in \mathbb{N}^2\}$ et que l'on note, si i et n sont deux entiers, $p_{i,n} = \mathbb{P}[Z = (x_i, y_n)]$.

On peut faire, dans le cas d'un couple Z , le même raisonnement que dans le cas d'une variable aléatoire réelle, et montrer que la loi de Z est entièrement déterminée par les valeurs $p_{i,n}$ pour i et n entiers naturels. En fait, si C est un borélien de \mathbb{R}^2 , on a :

$$\mathbb{P}(Z \in C) = \sum_{i,n \geq 0} \mathbb{P}[Z = (x_i, y_n)] \mathbf{1}_C(x_i, y_n) = \sum_{i,n \geq 0} p_{i,n} \mathbf{1}_C(x_i, y_n).$$

Ceci signifie que la probabilité que Z appartienne à C s'obtient en sommant les $p_{i,n}$ sur les indices i et n pour lesquels (x_i, y_n) appartient à C .

Pour déterminer les lois marginales de Z , il suffit de calculer $\mathbb{P}(X = x_i)$ pour tout entier i et $\mathbb{P}(Y = y_n)$ pour tout entier n . Mais on a, pour tout $i \in \mathbb{N}$,

$$\mathbb{P}(X = x_i) = \mathbb{P}(X = x_i, Y \in \mathbb{R}) = \sum_{n \geq 0} \mathbb{P}(X = x_i, Y = y_n) = \sum_{n \geq 0} p_{i,n},$$

et pour tout $n \in \mathbb{N}$,

$$\mathbb{P}(Y = y_n) = \mathbb{P}(X \in \mathbb{R}, Y = y_n) = \sum_{i \geq 0} \mathbb{P}(X = x_i, Y = y_n) = \sum_{i \geq 0} p_{i,n}.$$

Donnons un exemple de calcul de marginales à partir de la loi du couple.

Exemple. Soient $\lambda > 0$, $\mu > 0$, $p, q \in]0, 1[$ tels que $p + q = 1$. Soit $Z = (X, Y)$ le vecteur aléatoire à valeurs dans $\mathbb{N} \times \mathbb{N}^*$ de loi donnée par :

$$\forall (i, n) \in \mathbb{N} \times \mathbb{N}^*, \quad \mathbb{P}(X = i, Y = n) = \begin{cases} e^{-\lambda} \frac{\lambda^i}{i!} p^2 (1-p)^k & \text{si } n = 2k + 1, \\ e^{-\mu} \frac{\mu^i}{i!} q^2 (1-q)^k & \text{si } n = 2k + 2. \end{cases}$$

Déterminons les marginales de Z . Commençons par la loi de Y ; pour $k \in \mathbb{N}$, on a

$$\mathbb{P}(Y = 2k + 1) = \sum_{i \geq 0} \mathbb{P}(X = i, Y = 2k + 1) = \sum_{i \geq 0} e^{-\lambda} \frac{\lambda^i}{i!} p^2 (1-p)^k = p^2 (1-p)^k,$$

$$\mathbb{P}(Y = 2k + 2) = \sum_{i \geq 0} \mathbb{P}(X = i, Y = 2k + 2) = \sum_{i \geq 0} e^{-\mu} \frac{\mu^i}{i!} q^2 (1-q)^k = q^2 (1-q)^k.$$

Pour la loi de X , soit $i \in \mathbb{N}$,

$$\mathbb{P}(X = i) = \sum_{n \geq 1} \mathbb{P}(X = i, Y = n) = \sum_{k \geq 0} \mathbb{P}(X = i, Y = 2k + 1) + \sum_{k \geq 0} \mathbb{P}(X = i, Y = 2k + 2),$$

et l'on a, utilisant le fait que $\sum_{k \geq 0} (1-x)^k = 1/x$ si $0 < x < 1$,

$$\mathbb{P}(X = i) = p e^{-\lambda} \frac{\lambda^i}{i!} + q e^{-\mu} \frac{\mu^i}{i!}.$$

Indépendance. Comme on l'a vu dans l'exemple introductif, on ne peut pas en général, trouver la loi du couple si on connaît seulement les lois marginales. Toutefois, il y a un cas très particulier pour lequel on peut reconstruire la loi de $Z = (X, Y)$ à partir des lois de X et Y : celui de l'indépendance. Rappelons que deux variables aléatoires discrètes X et Y sont indépendantes si et seulement si

$$\forall (i, n) \in \mathbb{N}^2, \quad \mathbb{P}(X = x_i, Y = y_n) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_n).$$

La formule précédente peut se lire de la façon suivante : X et Y sont deux v.a.r. discrètes indépendantes si et seulement si la loi du couple $Z = (X, Y)$ est donnée par

$$\forall (i, n) \in \mathbb{N}^2, \quad p_{i,n} = \mathbb{P}(Z = (x_i, y_n)) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_n).$$

Remarque. En pratique, on connaît la loi du couple Z au travers des $p_{i,n}$ et on se demande si les v.a. X et Y sont indépendantes. Pour cela, il suffit de vérifier que l'on peut séparer les variables dans $p_{i,n}$ c'est à dire écrire $p_{i,n} = u_i v_n$ pour tout (i, n) ; on n'a pas besoin de déterminer les marginales.

La dernière chose à mentionner pour les couples discrets est la manière de calculer l'espérance de variables réelles construites à partir du couple Z .

Proposition 2. Soient $Z = (X, Y)$ un couple discret et $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction. Si la somme double $\sum_{i,n \geq 0} |f(x_i, y_n)| p_{i,n}$ est finie alors la v.a.r. $f(X, Y)$ est intégrable et

$$\mathbb{E}[f(X, Y)] = \sum_{i,n \geq 0} f(x_i, y_n) \mathbb{P}(Z = (x_i, y_n)) = \sum_{i,n \geq 0} f(x_i, y_n) p_{i,n}.$$

3. Couple possédant une densité.

Dans ce paragraphe, on se concentre sur le cas d'un vecteur aléatoire $Z = (X, Y)$ possédant une densité. Avant de définir ce qu'est une densité de probabilité dans le cas d'un couple, on présente deux résultats sur les intégrales doubles.

3.1. Intégrales sur \mathbb{R}^2 . Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction. On aimerait répondre aux questions suivantes : à quelles conditions peut-on dire que f est intégrable sur \mathbb{R}^2 ? Comment calculer dans ce cas l'intégrale? Les intégrales itérées sont-elles égales? Plus précisément, comment calculer

$$\iint_{\mathbb{R}^2} f(x, y) dx dy, \quad \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy ?$$

Nous verrons au paragraphe suivant que – comme dans le cas des séries doubles – la situation la plus favorable est celle où f est positive.

Comme dans le cas de fonctions réelles, l'intégrabilité de f sous-entend que, pour tout réel a , l'ensemble $\{x \in \mathbb{R}^2, f(x) \leq a\}$ appartient à la tribu borélienne de \mathbb{R}^2 : on dit que f est borélienne. En pratique, on ne s'attarde pas trop sur ce point et on utilise souvent le fait que si la fonction f est continue sauf sur un ensemble au plus dénombrable ou même en dehors d'un ensemble tel que le bord d'un cercle, une droite c'est à dire en dehors d'un ensemble « d'aire nulle » – on dit négligeable – alors f est borélienne.

Exemple. La fonction définie par $f(x, y) = 1$ si $x^2 + y^2 < 1$ et 0 sinon est continue sur \mathbb{R}^2 privé du cercle unité qui est négligeable. f est donc borélienne.

Passons à présent au théorème de Tonelli qui concerne les fonctions positives.

Théorème 3. Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$ une fonction borélienne et positive. Notons I_1 et I_2 les intégrales itérées i.e.

$$I_1 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy \quad \text{et} \quad I_2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx.$$

On a l'alternative suivante :

- ou bien I_1 et I_2 valent toutes les deux $+\infty$,
- ou bien I_1 et I_2 sont toutes les deux finies et égales.

On retient ce résultat sous la forme suivante : si f est une fonction positive alors on a toujours, en tolérant la valeur $+\infty$,

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx.$$

Le théorème de Tonelli conduit à la définition suivante :

Définition. Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction. f est intégrable sur \mathbb{R}^2 si f est borélienne et

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x, y)| dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x, y)| dx \right) dy < +\infty.$$

Nous passons à présent au théorème de Fubini, analogue du théorème de Tonelli pour les fonctions qui ne sont pas de signe constant.

Théorème 4. Si $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ est une fonction intégrable alors

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy.$$

Dans ce cas, on appelle intégrale de f sur \mathbb{R}^2 (ou intégrale double) la quantité

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy.$$

Si D est un borélien de \mathbb{R}^2 et f une fonction intégrable, on pose

$$\iint_D f(x, y) dx dy = \iint_{\mathbb{R}^2} \mathbf{1}_D(x, y) f(x, y) dx dy.$$

Exemple. Donnons un premier exemple. Soit D un borélien borné de \mathbb{R}^2 et $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction borélienne bornée sur D . Alors $g(x, y) = \mathbf{1}_D(x, y) f(x, y)$ est intégrable sur \mathbb{R}^2 . En effet, il existe $a > 0$, $M > 0$ tels que $D \subset [-a, a]^2$ et $|f(x)| \leq M$ si $x \in D$ et on a

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |g(x, y)| dy \right) dx \leq \int_{-a}^a \left(\int_{-a}^a M dy \right) dx = 4a^2 M,$$

qui est finie ; donc g est intégrable. Ce calcul s'applique en particulier au cas d'une fonction continue sur \mathbb{R}^2 .

3.2. Densité d'un couple aléatoire. Donnons maintenant la définition d'un couple aléatoire possédant une densité. Tout d'abord, comme dans le cas réel, on a la définition suivante.

Définition. Soit $p : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction. p est une densité de probabilité sur \mathbb{R}^2 si

- (i) p est positive ;
- (ii) p est intégrable sur \mathbb{R}^2 et $\iint_{\mathbb{R}^2} p(x, y) dx dy = 1$.

Définition. Soient $Z = (X, Y)$ un couple aléatoire et p une densité de probabilité sur \mathbb{R}^2 . Z a pour densité p si, pour tout couple (I, J) d'intervalles de \mathbb{R} ,

$$\mathbb{P}_Z(I \times J) = \mathbb{P}(X \in I, Y \in J) = \iint_{I \times J} p(x, y) dx dy = \iint_{\mathbb{R}^2} \mathbf{1}_I(x) \mathbf{1}_J(y) p(x, y) dx dy.$$

Remarquons que cette définition caractérise bien la loi du couple Z puisque d'après la Proposition 1 il suffit de spécifier $\mathbb{P}(Z \in I \times J)$ pour le faire.

Exemple. Soit p la fonction définie par

$$p(x, y) = c e^{-x} \mathbf{1}_{|y| \leq x}. \quad (*)$$

Calculons c de sorte que p soit une densité de probabilité. p est positive dès que c est positive ; on se place dans ce cas. On a de plus

$$\iint_{\mathbb{R}^2} p(x, y) dx dy = c \int_{y \in \mathbb{R}} \int_{x \in \mathbb{R}} e^{-x} \mathbf{1}_{|y| \leq x} dx dy = c \int_{y \in \mathbb{R}} \int_{|y|}^{+\infty} e^{-x} dx dy = c \int_{\mathbb{R}} e^{-|y|} dy = 2c.$$

Il faut donc prendre $c = 1/2$ pour que p soit une densité de probabilité.

Nous conserverons cet exemple tout au long de ce paragraphe.

Théorème 5. Soient $Z = (X, Y)$ un couple aléatoire de densité p , $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction borélienne. Si la fonction $f p$ est intégrable sur \mathbb{R}^2 i.e. $\iint_{\mathbb{R}^2} |f(x, y)| p(x, y) dx dy < +\infty$ alors la v.a. réelle $f(X, Y)$ est intégrable et dans ce cas,

$$\mathbb{E}[f(X, Y)] = \iint_{\mathbb{R}^2} f(x, y) p(x, y) dx dy.$$

En particulier, pour tout borélien D de \mathbb{R}^2 ,

$$\mathbb{P}[(X, Y) \in D] = \mathbb{E}[\mathbf{1}_D((X, Y))] = \iint_D p(x, y) dx dy.$$

Exemple. Reprenons l'exemple (*) et calculons l'espérance de Y .

$$\mathbb{E}[Y] = \iint_{\mathbb{R}^2} y p(x, y) dx dy = c \int_{y \in \mathbb{R}} y \int_{x \in \mathbb{R}} e^{-x} \mathbf{1}_{|y| \leq x} dx dy = c \int_{\mathbb{R}} y e^{-|y|} dy = 0$$

puisque la fonction est intégrable et impaire.

Lois marginales. Soit $Z = (X, Y)$ un vecteur aléatoire de densité p . On veut déterminer les lois marginales de Z , \mathbb{P}_X et \mathbb{P}_Y . On va montrer pour cela que X (respectivement Y) possède une densité p_X (respectivement p_Y) que l'on va calculer en fonction de p . Utilisons la méthode vue pour les variables aléatoires réelles. Soit donc $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue par morceaux et bornée. Calculons $\mathbb{E}[f(X)]$. D'après le Théorème 5 et le théorème de Fubini, on a

$$\mathbb{E}[f(X)] = \iint_{\mathbb{R}^2} f(x) p(x, y) dx dy = \int_{\mathbb{R}} f(x) \left(\int_{\mathbb{R}} p(x, y) dy \right) dx.$$

Mais si X a pour densité p_X ,

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x) p_X(x) dx,$$

et par identification on obtient

$$p_X(x) = \int_{\mathbb{R}} p(x, y) dy, \quad \text{respectivement,} \quad p_Y(y) = \int_{\mathbb{R}} p(x, y) dx.$$

Notons que p_X et p_Y sont bien des densités puisqu'elles sont intégrables, positives et, par définition

$$\int_{\mathbb{R}} p_X(x) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} p(x, y) dy \right) dx = \iint_{\mathbb{R}^2} p(x, y) dx dy = 1 ;$$

idem pour p_Y .

Nous venons de montrer le résultat suivant :

Proposition 6. *Soit $Z = (X, Y)$ un vecteur aléatoire de densité p . Alors les lois marginales \mathbb{P}_X et \mathbb{P}_Y admettent pour densités respectives*

$$p_X(x) = \int_{\mathbb{R}} p(x, y) dy, \quad \text{et,} \quad p_Y(y) = \int_{\mathbb{R}} p(x, y) dx.$$

Exemple. Calculons les lois marginales de l'exemple (*). Il vient immédiatement, pour $x \in \mathbb{R}$,

$$p_X(x) = ce^{-x} \int_{\mathbb{R}} \mathbf{1}_{|y| \leq x} dy = ce^{-x} \mathbf{1}_{x \geq 0} \int_{-x}^x dy = 2cxe^{-x} \mathbf{1}_{x \geq 0} = xe^{-x} \mathbf{1}_{x \geq 0}.$$

Le calcul de p_Y est plus simple. On a directement

$$p_Y(y) = c \int_{\mathbb{R}} e^{-x} \mathbf{1}_{|y| \leq x} dx = c \int_{|y|}^{+\infty} e^{-x} dx = ce^{-|y|} = \frac{1}{2} e^{-|y|}.$$

Indépendance. Soit $Z = (X, Y)$ de densité p . À quelle condition sur p X et Y sont-elles indépendantes? Peut-on trouver un critère simple?

Théorème 7. *Soient $Z = (X, Y)$ un couple aléatoire de densité p , p_X et p_Y les densités marginales. Alors X et Y sont indépendantes si et seulement si $p(x, y) = p_X(x)p_Y(y)$ pour tout (x, y) de \mathbb{R}^2 éventuellement privé d'une partie négligeable.*

Ce résultat n'est pas très difficile à montrer. Nous avons vu précédemment que X et Y sont indépendantes si et seulement si pour tout couple (g, h) de fonctions boréliennes bornées

de \mathbb{R} dans \mathbb{R} , on a $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$. D'après le Théorème 5, on a prenant $f(x, y) = g(x)h(y)$,

$$\mathbb{E}[g(X)h(Y)] = \iint_{\mathbb{R}^2} g(x)h(y)p(x, y) dx dy.$$

D'un autre côté,

$$\mathbb{E}[g(X)]\mathbb{E}[h(Y)] = \int_{\mathbb{R}} g(x)p_X(x) dx \times \int_{\mathbb{R}} h(y)p_Y(y) dy = \iint_{\mathbb{R}^2} g(x)h(y)p_X(x)p_Y(y) dx dy.$$

Le résultat s'en suit immédiatement.

Remarque. Il faut savoir utiliser le théorème précédent dans les deux sens.

Exemple. Si on reprend l'exemple de la page 6, on voit que X et Y ne sont pas indépendantes puisque la densité du couple n'est pas égale au produit des densités marginales.

4. Calcul de loi image.

Au chapitre précédent, nous avons considéré le problème suivant : si X est une variable aléatoire de loi connue et u une fonction réelle disons continue par morceaux quelle est la loi de la variable $u(X)$?

Dans bien des cas, on part d'un couple aléatoire (X, Y) et on s'intéresse à une variable aléatoire réelle définie à partir de ce couple : par exemple la somme $X + Y$, le produit XY ou encore le quotient X/Y .

Lorsque le couple possède une densité, on peut espérer que la variable construite à l'aide de celui-ci va également en posséder une. On peut alors essayer de la déterminer en utilisant la démarche du chapitre précédent qui consiste à calculer, si ξ désigne cette variable réelle, $\mathbb{E}[f(\xi)]$ pour toute fonction f continue par morceaux et bornée et d'identifier la densité p_ξ .

Exemple. Donnons quelques exemples illustrant cette technique.

1° Considérons deux variables aléatoires X et Y indépendantes, de loi exponentielle de paramètre λ pour X et μ pour Y . Cherchons la loi de $\xi = X/Y$ en déterminant sa densité. Soit f une fonction continue par morceaux et bornée de \mathbb{R} dans \mathbb{R} , calculons $\mathbb{E}[f(\xi)]$. On a :

$$\mathbb{E}[f(\xi)] = \iint_{]0, +\infty[^2} f(x/y) \lambda e^{-\lambda x} \mu e^{-\mu y} dx dy = \lambda \mu \int_0^{+\infty} e^{-\mu y} \left(\int_0^{+\infty} f(x/y) e^{-\lambda x} dx \right) dy.$$

Mais, posant $z = x/y$, on obtient, pour tout $y > 0$,

$$\int_0^{+\infty} f(x/y) e^{-\lambda x} dx = \int_0^{+\infty} f(z) y e^{-\lambda z y} dz,$$

ce qui conduit à, via le théorème de Fubini,

$$\mathbb{E}[f(\xi)] = \lambda \mu \int_0^{+\infty} e^{-\mu y} \left(\int_0^{+\infty} f(z) y e^{-\lambda z y} dz \right) dy = \lambda \mu \int_0^{+\infty} f(z) \left(\int_0^{+\infty} y e^{-(\lambda z + \mu)y} dy \right) dz.$$

Une intégration par parties donne, pour $z > 0$ fixé,

$$\int_0^{+\infty} y e^{-(\lambda z + \mu)y} dy = \frac{1}{(\lambda z + \mu)^2};$$

il s'en suit que

$$\mathbb{E}[f(\xi)] = \int_0^{+\infty} f(z) \frac{\lambda\mu}{(\lambda z + \mu)^2} dz = \int_{\mathbb{R}} f(z) \frac{\lambda\mu}{(\lambda z + \mu)^2} \mathbf{1}_{z>0} dz.$$

ξ a donc pour densité la fonction $z \mapsto \lambda\mu(\lambda z + \mu)^{-2} \mathbf{1}_{z>0}$.

2° Soient X et Y deux v.a. indépendantes; X de densité $3x^2 \mathbf{1}_{[0,1]}(x)$, Y de loi uniforme sur $[0, 1]$. Déterminons la loi du produit $\xi = XY$.

Si f est une fonction continue par morceaux et bornée sur \mathbb{R} , on a :

$$\mathbb{E}[f(\xi)] = \iint_{[0,1]^2} f(xy) 3x^2 dx dy = 3 \int_0^1 x^2 \left(\int_0^1 f(xy) dy \right) dx.$$

Pour $x \in]0, 1[$ fixé, le changement de variables $z = xy$ donne

$$\int_0^1 f(xy) dy = \frac{1}{x} \int_0^x f(z) dz = \frac{1}{x} \int_0^1 f(z) \mathbf{1}_{z<x} dz.$$

Par suite, il vient, d'après le théorème de Fubini,

$$\mathbb{E}[f(\xi)] = 3 \int_0^1 x \left(\int_0^1 f(z) \mathbf{1}_{z<x} dz \right) dx = 3 \int_0^1 f(z) \left(\int_0^1 x \mathbf{1}_{z<x} dx \right) dz,$$

et on obtient finalement,

$$\mathbb{E}[f(\xi)] = \frac{3}{2} \int_0^1 f(z) (1 - z^2) dz = \int_{\mathbb{R}} f(z) \frac{3}{2} (1 - z^2) \mathbf{1}_{[0,1]}(z) dz.$$

ξ a pour densité la fonction $z \mapsto \frac{3}{2} (1 - z^2) \mathbf{1}_{[0,1]}(z)$.

3° Pour finir considérons deux variables aléatoires indépendantes suivant la loi de Laplace i.e. de densité $e^{-|x|}/2$ et déterminons la loi de $\xi = 2X + Y$. Si f est continue par morceaux et bornée, on a, via le théorème de Fubini,

$$\mathbb{E}[f(\xi)] = \frac{1}{4} \iint_{\mathbb{R}^2} f(2x + y) e^{-|x|} e^{-|y|} dx dy = \frac{1}{4} \int_{\mathbb{R}} e^{-|x|} \left(\int_{\mathbb{R}} f(2x + y) e^{-|y|} dy \right) dx;$$

le changement de variables $z = y + 2x$ donne, pour tout x ,

$$\int_{\mathbb{R}} f(2x + y) e^{-|y|} dy = \int_{\mathbb{R}} f(z) e^{-|z-2x|} dz.$$

Par conséquent, appliquant encore le théorème de Fubini, on obtient

$$\mathbb{E}[f(\xi)] = \frac{1}{4} \int_{\mathbb{R}} e^{-|x|} \left(\int_{\mathbb{R}} f(z) e^{-|z-2x|} dz \right) dx = \frac{1}{4} \int_{\mathbb{R}} f(z) \left(\int_{\mathbb{R}} e^{-|x|} e^{-|z-2x|} dx \right) dz.$$

Pour $z > 0$, on a

$$\begin{aligned} \int_{\mathbb{R}} e^{-|x|} e^{-|z-2x|} dx &= \int_{-\infty}^0 e^x e^{-(z-2x)} dx + \int_0^{z/2} e^{-x} e^{-(z-2x)} dx + \int_{z/2}^{+\infty} e^{-x} e^{z-2x} dx, \\ &= e^{-z} \int_{-\infty}^0 e^{3x} dx + e^{-z} \int_0^{z/2} e^x dx + e^z \int_{z/2}^{+\infty} e^{-3x} dx \\ &= \frac{1}{3} e^{-z} + e^{-z} (e^{z/2} - 1) + \frac{1}{3} e^z e^{-3z/2} \\ &= \frac{4}{3} e^{-z/2} - \frac{2}{3} e^{-z}. \end{aligned}$$

De manière analogue, on a pour $z < 0$,

$$\begin{aligned}
 \int_{\mathbb{R}} e^{-|x|} e^{-|z-2x|} dx &= \int_{-\infty}^{z/2} e^x e^{-(z-2x)} dx + \int_{z/2}^0 e^x e^{z-2x} dx + \int_0^{+\infty} e^{-x} e^{z-2x} dx, \\
 &= e^{-z} \int_{-\infty}^{z/2} e^{3x} dx + e^z \int_{z/2}^0 e^{-x} dx + e^z \int_0^{+\infty} e^{-3x} dx \\
 &= \frac{1}{3} e^{-z} e^{3z/2} + e^z (e^{-z/2} - 1) + \frac{1}{3} e^z \\
 &= \frac{4}{3} e^{z/2} - \frac{2}{3} e^z.
 \end{aligned}$$

Par suite,

$$\forall z \in \mathbb{R}, \quad \int_{\mathbb{R}} e^{-|x|} e^{-|z-2x|} dx = \frac{4}{3} e^{-|z|/2} - \frac{2}{3} e^{-|z|}.$$

Il vient alors

$$\mathbb{E}[f(\xi)] = \frac{1}{6} \int_{\mathbb{R}} f(z) (2e^{-|z|/2} - e^{-|z|}) dz ;$$

ξ a pour densité $z \mapsto \frac{1}{6} (2e^{-|z|/2} - e^{-|z|})$.

Compléments sur les couples aléatoires

1. Couple image.

Dans ce paragraphe, on va s'intéresser à la loi d'un vecteur aléatoire (S, T) qui s'obtient comme image d'un vecteur (X, Y) par une application u . On connaît la loi du couple (X, Y) et on cherche la loi de $(S, T) = u(X, Y) = (u_1(X, Y), u_2(X, Y))$. On ne considèrera que des couples (S, T) et (X, Y) possédant une densité. La méthode est alors la même que dans le cas unidimensionnel : on cherche la densité du couple (S, T) en faisant un changement de variables mais dans \mathbb{R}^2 . Nous commençons par la formule de changement de variables dans ce contexte.

1.1. Changement de variables. Voici le cadre de l'étude. D est un « ouvert » de \mathbb{R}^2 et h est une application de D dans \mathbb{R} . On voudrait calculer l'intégrale $\iint_D h(x, y) dx dy$ en introduisant un nouveau jeu de coordonnées (s, t) c'est à dire en effectuant le changement de variables « $(s, t) = u(x, y)$ » où u est une application de D dans \mathbb{R}^2 . Comme dans le cas réel, la fonction u doit être régulière.

C^1 -difféomorphisme. La formule du changement de variables nécessite de la régularité sur la fonction qui permet de passer d'un jeu de coordonnées à l'autre : celle-ci doit être un C^1 -difféomorphisme.

Définition. Soient D et D' deux ouverts de \mathbb{R}^2 . $u : D \rightarrow \mathbb{R}^2$ est un C^1 -difféomorphisme de D sur D' si u est une bijection de D sur D' telle que u et u^{-1} soient de classe C^1 .

Si $u : D \rightarrow \mathbb{R}^2$ est différentiable, sa différentielle Du est représentée par la matrice suivante – appelée matrice jacobienne –

$$\forall (x, y) \in D, \quad Du(x, y) = \begin{pmatrix} \partial_x u_1(x, y) & \partial_y u_1(x, y) \\ \partial_x u_2(x, y) & \partial_y u_2(x, y) \end{pmatrix}.$$

Le jacobien de u est le déterminant de cette matrice, soit

$$\forall (x, y) \in D, \quad Ju(x, y) = \det(Du(x, y)) = \partial_x u_1(x, y) \partial_y u_2(x, y) - \partial_x u_2(x, y) \partial_y u_1(x, y).$$

On peut retenir le critère suivant :

Proposition 1. Soient D un ouvert de \mathbb{R}^2 et $u : D \rightarrow \mathbb{R}^2$ une application injective de classe C^1 . Si, pour tout $(x, y) \in D$, $Ju(x, y) \neq 0$ alors $u(D)$ est un ouvert de \mathbb{R}^2 et l'application u est un C^1 -difféomorphisme de D sur $u(D)$.

Si u est un C^1 -difféomorphisme de D sur D' alors u^{-1} est un C^1 -difféomorphisme de D' sur D . On a dans ce cas, comme dans le cas réel, $Du^{-1} = (Du \circ u^{-1})^{-1}$ et en particulier,

$$\forall (s, t) \in D', \quad Ju^{-1}(s, t) = \frac{1}{Ju(u^{-1}(s, t))}.$$

Un cas particulier important est celui des fonctions u linéaires :

$$\forall (x, y) \in \mathbb{R}^2, \quad u(x, y) = \begin{pmatrix} u_1(x, y) \\ u_2(x, y) \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix};$$

si M est la matrice qui apparaît ci-dessus, alors $Ju(x, y) = \det(M)$ et donc l'application u est un C^1 -difféomorphisme de \mathbb{R}^2 dans \mathbb{R}^2 si et seulement si $\det(M) \neq 0$.

Formule du changement de variables. Commençons par un énoncé précis.

Théorème 2. Soient D' un ouvert de \mathbb{R}^2 , v un C^1 -difféomorphisme sur D' . Considérons une fonction $h : v(D') \rightarrow \mathbb{R}$ borélienne. Si h est positive ou intégrable sur $v(D')$ alors

$$\iint_{v(D')} h(x, y) \, dx dy = \iint_{D'} h(v(s, t)) |Jv(s, t)| \, ds dt. \quad (1)$$

La signification de ce résultat est la suivante :

- si h est positive, les deux termes de l'égalité précédente sont soit tout deux infinis soit tout deux finis et égaux ;
- sinon h est intégrable sur $v(D')$ si et seulement si $h \circ v |Jv|$ est intégrable sur D' et dans ce cas on a l'égalité (1).

La formule (1) correspond à un changement de variables du type $(x, y) = v(s, t)$ où (x, y) sont les variables d'intégration de départ (dont on veut se débarrasser). En pratique, cette situation n'est pas très fréquente ; il faut tout de même se souvenir du passage en coordonnées polaires. On souhaite calculer $\iint h(x, y) \, dx dy$ en posant $x = r \cos \theta$, $y = r \sin \theta$. Ceci correspond à la fonction $v(r, \theta) = (r \cos \theta, r \sin \theta)$ qui est un C^1 -difféomorphisme de $]0, +\infty[\times]-\pi, \pi[$ sur $\mathbb{R}^2 \setminus \mathbb{R}_-$ (le plan privé de la demi-droite $y = 0, x \leq 0$). Notez que $v^{-1}(x, y) = (r, \theta)$ avec

$$r = \sqrt{x^2 + y^2}, \quad \text{et,} \quad \theta = 2 \arctan \left(\frac{y}{x + \sqrt{x^2 + y^2}} \right).$$

Exemple. Montrons que $I = \int_{\mathbb{R}} \exp(-x^2/2) \, dx = \sqrt{2\pi}$. Pour cela, calculons I^2 :

$$I^2 = \int_{\mathbb{R}} \exp(-x^2/2) \, dx \int_{\mathbb{R}} \exp(-y^2/2) \, dy = \iint_{\mathbb{R}^2} \exp(-(x^2 + y^2)/2) \, dx dy,$$

puis comme la demi-droite $y = 0, x \leq 0$ est négligeable,

$$I^2 = \iint_{\mathbb{R}^2} \exp(-(x^2 + y^2)/2) \, dx dy = \iint_{\mathbb{R}^2 \setminus \mathbb{R}_-} \exp(-(x^2 + y^2)/2) \, dx dy.$$

Effectuons le changement de variables $(x, y) = (r \cos \theta, r \sin \theta) = v(r, \theta)$. On a

$$\forall (r, \theta) \in]0, +\infty[\times]-\pi, \pi[, \quad Jv(r, \theta) = \begin{vmatrix} \partial_r x(r, \theta) & \partial_\theta x(r, \theta) \\ \partial_r y(r, \theta) & \partial_\theta y(r, \theta) \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r.$$

La formule du changement de variables donne alors

$$I^2 = \iint_{]0, +\infty[\times]-\pi, \pi[} \exp(-r^2/2) r \, dr d\theta = \int_{|\theta| < \pi} d\theta \int_0^{+\infty} r \exp(-r^2/2) \, dr = 2\pi.$$

I est l'intégrale d'une fonction positive, donc I est positive ; il vient $I = \sqrt{2\pi}$.

Changement de variables $(s, t) = u(x, y)$. En pratique, on veut calculer $\iint_D h(x, y) \, dx dy$ et on effectue le changement de variables $(s, t) = u(x, y)$ où u est un C^1 -difféomorphisme de D sur $D' := u(D)$. On se ramène au cadre d'application de la formule (1) en considérant $v = u^{-1}$ puisque $(x, y) = u^{-1}(s, t)$. On a alors

$$\iint_D h(x, y) \, dx dy = \iint_{u^{-1}(D')} h(x, y) \, dx dy = \iint_{D'} h(u^{-1}(s, t)) |Ju^{-1}(s, t)| \, ds dt,$$

et comme $D' = u(D)$ et $Ju^{-1} = (Ju \circ u^{-1})^{-1}$, on a finalement

$$\iint_D h(x, y) \, dx dy = \iint_{u(D)} \frac{h(u^{-1}(s, t))}{|Ju(u^{-1}(s, t))|} \, ds dt. \quad (2)$$

Pour appliquer cette formule, on doit calculer,

$$\forall (s, t) \in u(D), \quad \frac{h(u^{-1}(s, t))}{|Ju(u^{-1}(s, t))|}.$$

En pratique, on détermine le rapport

$$\forall (x, y) \in D, \quad \frac{h(x, y)}{|Ju(x, y)|}$$

que l'on exprime en fonction des coordonnées (s, t) . On a ainsi $h(x, y)/|Ju(x, y)| = g(s, t)$ et

$$\iint_D h(x, y) \, dx dy = \iint_{u(D)} g(s, t) \, ds dt.$$

Cela permet dans certains cas d'éviter d'inverser la fonction u .

1.2. Application aux couples aléatoires. Le point de départ est un vecteur aléatoire $Z = (X, Y)$ dont on connaît la loi \mathbb{P}_Z au travers de sa densité $p_Z(x, y) = \mathbf{1}_D(x, y) p(x, y)$ où D est un ouvert de \mathbb{R}^2 . On cherche la loi du vecteur aléatoire $U = (S, T)$ défini par $(S, T) = u(X, Y)$ où u est un C^1 -difféomorphisme sur D . On essaie de déterminer la densité p_U de (S, T) . Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction borélienne et bornée ; calculons $\mathbb{E}[f(S, T)]$ de deux façons différentes. Tout d'abord,

$$\mathbb{E}[f(S, T)] = \iint_{\mathbb{R}^2} f(s, t) p_U(s, t) \, ds dt ;$$

mais aussi, comme la densité de (X, Y) est $\mathbf{1}_D(x, y) p(x, y)$,

$$\mathbb{E}[f(S, T)] = \mathbb{E}[f(u(X, Y))] = \iint_D f(u(x, y)) p(x, y) \, dx dy.$$

Comme u est un C^1 -difféomorphisme, on fait le changement de variables $(s, t) = u(x, y)$ qui conduit à – cf. (2) avec $h(x, y) = f(u(x, y)) p(x, y)$ –

$$\mathbb{E}[f(S, T)] = \iint_{u(D)} f(s, t) \frac{p(u^{-1}(s, t))}{|Ju(u^{-1}(s, t))|} \, ds dt = \iint_{\mathbb{R}^2} f(s, t) \mathbf{1}_{u(D)}(s, t) \frac{p(u^{-1}(s, t))}{|Ju(u^{-1}(s, t))|} \, ds dt.$$

On obtient donc

$$\forall (s, t) \in \mathbb{R}^2, \quad p_U(s, t) = \mathbf{1}_{u(D)}(s, t) \frac{p(u^{-1}(s, t))}{|Ju(u^{-1}(s, t))|}.$$

En pratique, on peut suivre le plan suivant :

- on montre que u est C^1 et injective sur D ;
- on montre que, pour $(x, y) \in D$, $Ju(x, y) \neq 0$ en calculant le jacobien ;
- on détermine le domaine $u(D)$: très souvent on doit inverser u i.e. exprimer (x, y) en fonction de (s, t) ;
- on utilise la méthode ci-dessus : il reste à exprimer $p(x, y)/|Ju(x, y)|$ en fonction de (s, t) .

Exemple. Soient X et Y deux v.a. indépendantes, X de loi $\mathcal{E}xp(\lambda)$ et Y de loi $\mathcal{E}xp(\mu)$, $\lambda, \mu > 0$. Déterminons la loi du couple $(X/Y, Y)$.

Pour se ramener à la situation que l'on vient de décrire on introduit les variables $S = X/Y$ et $T = Y$. On cherche la loi de (S, T) . Soit f une fonction borélienne et bornée; calculons $\mathbb{E}[f(S, T)]$. On a

$$\mathbb{E}[f(S, T)] = \mathbb{E}[f(X/Y, Y)] = \iint_{]0, +\infty[^2} f(x/y, y) \lambda e^{-\lambda x} \mu e^{-\mu y} dx dy.$$

Considérons la fonction $u :]0, +\infty[^2 \rightarrow \mathbb{R}^2$ définie par $u(x, y) = (x/y, y)$. u est de classe C^∞ sur l'ouvert $D =]0, +\infty[^2$ et on a

$$\forall (x, y) \in]0, +\infty[^2, \quad \text{J}u(x, y) = \begin{vmatrix} 1/y & -x/y^2 \\ 0 & 1 \end{vmatrix} = \frac{1}{y} \neq 0.$$

u est injective : si $(x/y, y) = (x'/y', y')$ alors $x = x'$ et $y = y'$. Déterminons $u(]0, +\infty[^2)$. Trivialement, $u(]0, +\infty[^2) \subset]0, +\infty[^2$; remarquons que $(s, t) = u(x, y)$ signifie que $x = st$ et $y = t$ (c'est u^{-1}). Il vient alors $u(]0, +\infty[^2) =]0, +\infty[^2$. La formule du changement de variables donne

$$\mathbb{E}[f(S, T)] = \lambda \mu \iint_{]0, +\infty[^2} f(s, t) t e^{-\lambda st} e^{-\mu t} ds dt.$$

Le couple $U = (S, T)$ a donc pour densité $p_U(s, t) = \lambda \mu t \exp(-t(\lambda s + \mu)) \mathbf{1}_{s>0} \mathbf{1}_{t>0}$.

Remarquons que dans cet exemple (comme dans bien d'autres), on peut éviter de faire un changement de variables en dimension deux. En effet, on a, d'après le théorème de Fubini,

$$\mathbb{E}[f(S, T)] = \lambda \mu \iint_{]0, +\infty[^2} f\left(\frac{x}{y}, y\right) e^{-\lambda x} e^{-\mu y} dx dy = \lambda \mu \int_0^{+\infty} e^{-\mu y} \left(\int_0^{+\infty} f\left(\frac{x}{y}, y\right) e^{-\lambda x} dx \right) dy.$$

Pour $y > 0$ fixé, on fait le changement de variables réelles $s = x/y$, pour obtenir

$$\int_0^{+\infty} f(x/y, y) e^{-\lambda x} dx = \int_0^{+\infty} f(s, y) y e^{-\lambda sy} ds,$$

et par suite

$$\mathbb{E}[f(S, T)] = \lambda \mu \iint_{]0, +\infty[^2} f(s, y) y e^{-\lambda sy} e^{-\mu y} ds dy,$$

ce qui est le résultat que l'on avait trouvé.

2. Rudiments sur les vecteurs gaussiens.

Dans ce paragraphe, nous examinons très brièvement un cas particulier de couples aléatoires, celui des vecteurs gaussiens. Les vecteurs gaussiens sont souvent utilisés car, d'un point de vue pratique, ils conduisent à des calculs relativement simples.

Pour pouvoir utiliser certaines notations matricielles, convenons de représenter les couples aléatoires comme des vecteurs colonnes i.e. $X = (X_1, X_2)^t$ où t désigne la transposition. De plus, si x et y sont deux vecteurs de \mathbb{R}^2 on note $x \cdot y$ leur produit scalaire $x \cdot y = x^t y = x_1 y_1 + x_2 y_2$.

Définition. Un couple aléatoire $X = (X_1, X_2)^t$ suit la loi normale centrée réduite de dimension 2, $\mathcal{N}(0, I_2)$, si X_1 et X_2 sont deux v.a.r. normales centrées réduites indépendantes.

Si X est un tel couple alors X possède une densité qui est :

$$\forall x = (x_1, x_2) \in \mathbb{R}^2, \quad p(x) = p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) = \frac{1}{2\pi} \exp\left(-\frac{x^t x}{2}\right).$$

Comme dans le cas de variables aléatoires réelles, la loi d'un couple X est déterminée par sa fonction caractéristique qui est la fonction de \mathbb{R}^2 dans \mathbb{C} définie par

$$\forall \lambda \in \mathbb{R}^2, \quad \varphi_X(\lambda) = \mathbb{E}\left[e^{i\lambda^t X}\right] = \mathbb{E}\left[e^{i\lambda_1 X_1} e^{i\lambda_2 X_2}\right].$$

Si X suit la loi $\mathcal{N}(0, I_2)$, l'indépendance de X_1 et X_2 conduit immédiatement à :

$$\forall \lambda \in \mathbb{R}^2, \quad \varphi_X(\lambda) = \varphi_{X_1}(\lambda_1) \varphi_{X_2}(\lambda_2) = e^{-\lambda_1^2/2} e^{-\lambda_2^2/2} = \exp(-\lambda^t \lambda / 2).$$

Définition. Un vecteur aléatoire Y est gaussien si $Y = AX + \mu$ où μ est un vecteur de \mathbb{R}^2 , A est une matrice réelle de taille 2×2 et X un vecteur normal centré réduit.

Calculons la fonction caractéristique de Y . On a, pour $\lambda \in \mathbb{R}^2$,

$$\varphi_Y(\lambda) = \mathbb{E}\left[\exp(i\lambda^t (AX + \mu))\right] = e^{i\lambda^t \mu} \mathbb{E}\left[\exp(i\lambda^t AX)\right] = e^{i\lambda^t \mu} \mathbb{E}\left[\exp\left(i(A^t \lambda)^t X\right)\right],$$

et par suite,

$$\varphi_Y(\lambda) = e^{i\lambda^t \mu} \varphi_X(A^t \lambda) = \exp\left(i\lambda^t \mu - \frac{\lambda^t A A^t \lambda}{2}\right).$$

On note Γ la matrice $A A^t$. Γ est une matrice symétrique, semi-définie positive. On montre facilement que Γ est la matrice de variance-covariance de Y c'est à dire

$$\Gamma = \begin{pmatrix} \mathbb{V}[Y_1] & \text{Cov}[Y_1, Y_2] \\ \text{Cov}[Y_1, Y_2] & \mathbb{V}[Y_2] \end{pmatrix},$$

où $\text{Cov}[Y_1, Y_2] = \mathbb{E}[(Y_1 - \mathbb{E}[Y_1])(Y_2 - \mathbb{E}[Y_2])]$. D'autre part le vecteur μ est simplement le vecteur des moyennes de Y i.e. $\mu = (\mathbb{E}[Y_1], \mathbb{E}[Y_2])^t$.

Comme dans le cas réel, on voit que la loi d'un couple gaussien est déterminée par la moyenne μ et la matrice de variance-covariance Γ . On dit que Y suit la loi $\mathcal{N}(\mu, \Gamma)$.

Remarque. Si Y est un vecteur gaussien, alors les marginales sont des gaussiennes. En effet, si $t \in \mathbb{R}$, on a, posant $\lambda = (t, 0)^t$,

$$\varphi_Y(\lambda) = \varphi_{Y_1}(t) = \exp(it\mu_1 - \Gamma_{1,1}t^2/2),$$

ce qui montre que Y_1 suit la loi $\mathcal{N}(\mu_1, \Gamma_{1,1})$.

On aimerait à présent savoir si le vecteur Y possède une densité. En fait cela n'est vrai que si la matrice Γ est inversible, ce qui revient à dire que la matrice A est inversible puisque $\det \Gamma = (\det A)^2$. Supposons donc que $\det \Gamma > 0$. Soit f une fonction borélienne et bornée de \mathbb{R}^2 dans \mathbb{R} . On a

$$\mathbb{E}[f(Y)] = \mathbb{E}[f(AX + \mu)] = \frac{1}{2\pi} \iint_{\mathbb{R}^2} f(Ax + \mu) \exp\left(-\frac{x^t x}{2}\right) dx_1 dx_2.$$

Effectuons le changement de variables $y = Ax + \mu = u(x)$ soit $x = A^{-1}(y - \mu) = v(y)$, pour obtenir comme $Jv(y) = \det(A^{-1})$,

$$\mathbb{E}[f(Y)] = \frac{1}{2\pi} \iint_{\mathbb{R}^2} f(y) \exp\left(-\frac{(A^{-1}(y - \mu))^t (A^{-1}(y - \mu))}{2}\right) |\det(A^{-1})| dy_1 dy_2.$$

On a $(A^{-1})^t = (A^t)^{-1}$ et $(A^t)^{-1} A^{-1} = (AA^t)^{-1} = \Gamma^{-1}$; on en déduit tout d'abord que $(A^{-1}(y - \mu))^t (A^{-1}(y - \mu)) = (y - \mu)^t \Gamma^{-1} (y - \mu)$, puis notant que $|\det(A^{-1})| = (\det \Gamma)^{-1/2}$,

$$\mathbb{E}[f(Y)] = \frac{1}{2\pi\sqrt{\det \Gamma}} \iint_{\mathbb{R}^2} f(y) \exp\left(-\frac{(y - \mu)^t \Gamma^{-1} (y - \mu)}{2}\right) dy_1 dy_2.$$

Y admet donc pour densité la fonction p_Y définie par

$$\forall y \in \mathbb{R}^2, \quad p_Y(y) = \frac{1}{2\pi\sqrt{\det \Gamma}} \exp\left(-\frac{(y - \mu)^t \Gamma^{-1} (y - \mu)}{2}\right).$$

Théorèmes limites

Pour finir ce cours nous allons donner deux exemples de théorèmes limites pour des suites de v.a. réelles indépendantes : la loi faible des grands nombres et le théorème de la limite centrale ou théorème « central limit », TCL en abrégé dans la suite.

1. Loi des grands nombres.

Nous commençons par deux inégalités très classiques du calcul des probabilités ; elles sont d'usage fréquent.

1.1. Inégalités de Markov et Bienaymé–Tchebychev. Tout d'abord l'inégalité de Markov dont la démonstration est d'une simplicité enfantine.

Proposition 1. *Soit X une v.a.r. positive et intégrable. Alors,*

$$\forall a > 0, \quad a \mathbb{P}(X \geq a) \leq \mathbb{E}[X].$$

La démonstration de cette inégalité est élémentaire. Il suffit de remarquer que, comme X est une v.a.r. positive,

$$X = X \mathbf{1}_{X < a} + X \mathbf{1}_{X \geq a} \geq X \mathbf{1}_{X \geq a} \geq a \mathbf{1}_{X \geq a},$$

et d'utiliser la croissance de l'espérance, pour obtenir

$$\mathbb{E}[X] \geq a \mathbb{E}[\mathbf{1}_{X \geq a}] = a \mathbb{P}(X \geq a).$$

L'inégalité ci-dessus est vraie pour $a = 0$ mais ne présente dans ce cas aucun intérêt.

Remarque. Notons que si X est une v.a.r. positive et $r > 0$, alors $X(\omega) \geq a$ si et seulement si $X^r(\omega) \geq a^r$, et ce pour tout $a > 0$. En particulier, si X est une v.a.r. positive qui possède un moment d'ordre r , on a

$$\forall a > 0, \quad a^r \mathbb{P}(X \geq a) = a^r \mathbb{P}(X^r \geq a^r) \leq \mathbb{E}[X^r].$$

Cette remarque très simple conduit à une seconde inégalité connue sous le nom d'inégalité de Bienaymé–Tchebychev.

Proposition 2. *Soient $a > 0$ et X une v.a.r. de carré intégrable. Alors,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}[X]}{a^2}.$$

Il suffit d'appliquer la remarque précédente à la variable $Y = |X - \mathbb{E}[X]|$ avec $r = 2$. En effet, pour tout $a > 0$,

$$a^2 \mathbb{P}(Y \geq a) \leq \mathbb{E}[Y^2] = \mathbb{V}[X].$$

Remarque. Ces inégalités sont valables indépendamment de la loi de la v.a.r. X . Il n'est donc pas très surprenant qu'elles ne soient pas extrêmement précises comme on peut s'en convaincre sur l'exemple suivant. Soient X de loi uniforme sur $[0, 1]$ et $a = 1$; $\mathbb{E}[X] = 1/2$ et donc l'inégalité de Markov donne $0 = \mathbb{P}(X \geq 1) \leq 1/2$, qui n'est pas optimal !

1.2. Convergence en moyenne quadratique. Nous allons nous intéresser à la convergence d'une suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$. Nous devons préciser en quel sens on doit comprendre cette convergence.

Définition. Soient $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. de carré intégrable et X une v.a.r., toutes définies sur le même espace de probabilité. $(X_n)_{n \in \mathbb{N}}$ converge vers X en moyenne quadratique si :

$$\lim_{n \rightarrow +\infty} \mathbb{E} [|X_n - X|^2] = 0.$$

Il existe différentes notions de convergence pour les suite variables aléatoires réelles. Définissons un autre mode de convergence : la convergence en probabilité.

Définition. Soient $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r., X une v.a.r., toutes définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que $(X_n)_{n \in \mathbb{N}}$ converge vers X en probabilité si :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P} (|X_n - X| \geq \varepsilon) = 0.$$

Remarque. Un moyen d'établir la convergence en probabilité de $(X_n)_{n \in \mathbb{N}}$ vers X consiste à montrer que, pour un $r > 0$, $\mathbb{E} [|X_n - X|^r]$ converge vers 0 et d'utiliser l'inégalité de Markov puisque

$$\forall \varepsilon > 0, \quad \mathbb{P} (|X_n - X| \geq \varepsilon) \leq \mathbb{E} [|X_n - X|^r] / \varepsilon^r.$$

En particulier, la convergence en moyenne quadratique entraîne la convergence en probabilité.

Exemple. Pour illustrer la notion de convergence en probabilité, considérons X une v.a.r. uniforme sur $[0, 1]$ et, pour $n \in \mathbb{N}^*$, $X_n = X + n^2 \mathbf{1}_{X \leq 1/n}$. Alors, la suite $(X_n)_{n \in \mathbb{N}^*}$ converge vers X en probabilité. En effet, si $0 < \varepsilon < 1$, alors

$$\mathbb{P} (|X_n - X| \geq \varepsilon) = \mathbb{P} (X \leq 1/n) = 1/n.$$

Par contre, la suite $(X_n)_{n \in \mathbb{N}^*}$ ne converge pas vers X en moyenne quadratique puisque

$$\mathbb{E} [|X_n - X|^2] = n^4 \mathbb{P} (X \leq 1/n) = n^3.$$

1.3. Loi faible des grands nombres. Imaginons un instant que l'on lance une pièce non truquée un grand nombre de fois, disons n fois. On note P_n le nombre de fois où « pile » est apparu. Intuitivement, si n est grand la fréquence d'apparition de pile, P_n/n , va être proche de $1/2$. Le théorème suivant, la loi faible des grands nombres, confirme cette intuition.

Théorème 3. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de v.a.r. indépendantes, de même loi, de carré intégrable et définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On note, pour tout $n \geq 1$, $S_n = X_1 + \dots + X_n$. Alors, avec $\mu = \mathbb{E}[X_1]$,

$$\frac{S_n}{n} = \frac{1}{n} (X_1 + \dots + X_n) \longrightarrow \mu, \quad \text{en moyenne quadratique, quand } n \rightarrow \infty.$$

Remarque. Le résultat reste vrai si on suppose seulement que les v.a.r. sont 2 à 2 non-corrélées et possèdent la même moyenne et la même variance au lieu de les supposer indépendantes et de même loi.

La démonstration est relativement facile : comme les v.a.r. sont indépendantes et de même loi – c'est pareil sous les hypothèses de la remarque précédente – on a, notant σ^2 la variance de X_1 ,

$$\mathbb{E} [n^{-1} S_n] = \mu, \quad \mathbb{E} \left[|n^{-1} S_n - \mu|^2 \right] = \mathbb{V} [n^{-1} S_n] = \frac{\sigma^2}{n};$$

le résultat s'en suit immédiatement.

2. Théorème de la limite centrale.

Nous avons vu au paragraphe précédent deux types de convergence pour les suites de v.a.r. : la convergence en moyenne quadratique et la convergence en probabilité. Toutefois pour énoncer le théorème de la limite centrale nous aurons besoin d'une notion plus faible de convergence.

2.1. Convergence en loi. Pour ce type de convergence, les variables aléatoires peuvent être définies sur des espaces de probabilité différents.

Définition. Une suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une v.a.r. X si, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, continue et bornée,

$$\mathbb{E}[f(X_n)] \longrightarrow \mathbb{E}[f(X)], \quad \text{quand } n \rightarrow +\infty.$$

Remarque. Comme déjà dit, les v.a.r. X_n peuvent être définies sur des espaces de probabilité différents puisque les v.a. n'interviennent qu'au travers d'espérances.

Dans la définition, la continuité de la fonction f est importante. En effet, si X_n est une v.a. prenant les valeurs $1/n$ et 1 avec probabilité $1/2$ alors X_n converge en loi vers X où X prend les valeurs 0 et 1 avec probabilité $1/2$ puisque, si f est continue bornée,

$$\mathbb{E}[f(X_n)] = (f(1/n) + f(1)) / 2 \longrightarrow (f(0) + f(1)) / 2 = \mathbb{E}[f(X)].$$

Pourtant, si on prend $f = \mathbf{1}_{\{0\}}$ qui n'est pas continue au point 0 ,

$$\mathbb{E}[f(X_n)] = \mathbb{P}(X_n = 0) = 0, \quad \mathbb{E}[f(X)] = \mathbb{P}(X = 0) = 1/2.$$

En général, si X_n converge en loi vers X , on ne peut pas dire que $\mathbb{P}(X_n \in A)$ converge vers $\mathbb{P}(X \in A)$ car $x \mapsto \mathbf{1}_A(x)$ n'est pas une fonction continue.

Exemple. Soit X_n une v.a. de loi uniforme sur $\{i/n, i = 0, \dots, n-1\}$. Alors X_n converge en loi vers X de loi uniforme sur $[0, 1]$. En effet, si f est une fonction continue et bornée, on a

$$\mathbb{E}[f(X_n)] = \frac{1}{n} \sum_{i=0}^{n-1} f(i/n) \longrightarrow \int_0^1 f(x) dx = \mathbb{E}[f(X)],$$

puisque $\mathbb{E}[f(X_n)]$ est la somme de Riemann de f associée à la subdivision de pas $1/n$ sur $[0, 1]$.

Dans cet exemple, les v.a.r. X_n sont discrètes et la v.a.r. X est absolument continue. Si toutes les v.a.r. sont discrètes et à valeurs dans le même ensemble discret on peut obtenir une caractérisation simple de la convergence en loi.

Proposition 4. Soient $X_n, n \in \mathbb{N}$, et X des v.a.r. à valeurs dans \mathbb{N} . Alors X_n converge en loi vers X si et seulement si, pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \longrightarrow \mathbb{P}(X = k)$ quand $n \rightarrow +\infty$.

Rappelons que, dans ce cas, si f est continue bornée,

$$\mathbb{E}[f(X_n)] = \sum_{i \geq 0} f(i) \mathbb{P}(X_n = i), \quad \mathbb{E}[f(X)] = \sum_{i \geq 0} f(i) \mathbb{P}(X = i);$$

on montre que la condition est nécessaire en considérant, pour tout entier k , la fonction continue $f(x) = (1 - 2|x - k|)^+$ qui vérifie $f(p) = 0$ pour tout $p \in \mathbb{N}$ tel que $p \neq k$. La condition est

suffisante par passage à la limite dans $\mathbb{E}[f(X_n)]$, f étant bornée. En effet, si on fixe $0 < \varepsilon < 1/2$, il existe un entier m tel que $\sum_{i=0}^m \mathbb{P}(X = i) \geq 1 - \varepsilon$, puisque $\sum_{i \geq 0} \mathbb{P}(X = i) = 1$. Comme $\mathbb{P}(X_n = i) \rightarrow \mathbb{P}(X = i)$ pour tout i on en déduit que, si n est suffisamment grand disons $n \geq p$, $\sum_{i=0}^m \mathbb{P}(X_n = i) \geq 1 - 2\varepsilon$. On a donc, pour tout entier n ,

$$\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)] = \sum_{i>m} f(i) \{\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\} + \sum_{i=0}^m f(i) \{\mathbb{P}(X_n = i) - \mathbb{P}(X = i)\},$$

et, si $n \geq p$, l'inégalité triangulaire donne

$$\left| \mathbb{E}[f(X_n)] - \mathbb{E}[f(X)] \right| \leq 3\varepsilon \|f\|_\infty + \|f\|_\infty \max_{0 \leq i \leq m} |\mathbb{P}(X_n = i) - \mathbb{P}(X = i)|,$$

où $\|f\|_\infty = \sup_{i \in \mathbb{N}} |f(i)|$. Cette dernière inégalité permet de conclure puisque m est fini.

Exemple. On suppose que X_n suit la loi $\mathcal{B}(n, p_n)$. Si $np_n \rightarrow \lambda > 0$, alors X_n converge en loi vers une v.a. X suivant une loi de Poisson $\mathcal{P}(\lambda)$. Appliquons le critère précédent. On a, si $k \leq n$, $\mathbb{P}(X_n = k) = C_n^k (p_n)^k (1 - p_n)^{n-k}$. Tout d'abord $\mathbb{P}(X_n = 0) = (1 - p_n)^n = \exp\{n \ln(1 - p_n)\}$. Comme $\lim_{x \rightarrow 0} \ln(1 - x)/x = -1$, on a comme $p_n \rightarrow 0$, $\mathbb{P}(X_n = 0) \rightarrow e^{-\lambda} = \mathbb{P}(X = 0)$. Un calcul facile conduit à, pour $n \geq k + 1$,

$$\frac{\mathbb{P}(X_n = k + 1)}{\mathbb{P}(X_n = k)} = \frac{(n - k)p_n}{(k + 1)(1 - p_n)} = \frac{1 - kn^{-1}}{1 - p_n} \frac{np_n}{k + 1} \rightarrow \frac{\lambda}{k + 1}.$$

Il vient alors, pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \rightarrow e^{-\lambda} \lambda^k / k!$.

Nous admettrons le résultat suivant qui donne des caractérisations de la convergences en loi.

Théorème 5. Soient $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. et X une v.a.r. Alors les propositions suivantes sont équivalentes :

- $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X ;
- pour tout $t \in \mathbb{R}$, $\varphi_{X_n}(t) \rightarrow \varphi_X(t)$;
- pour tout $x \in \mathbb{R}$ où F_X est continue, $F_{X_n}(x) \rightarrow F_X(x)$.

De plus, si $(X_n)_{n \in \mathbb{N}}$ converge vers X en probabilité alors $(X_n)_{n \in \mathbb{N}}$ converge vers X en loi.

Exemple. Soit Y_n une v.a. de loi géométrique λ/n , $\lambda > 0$. Montrons que $X_n = Y_n/n$ converge en loi vers une v.a. X de loi $\mathcal{Exp}(\lambda)$. Soient φ_n la fonctions caractéristique de X_n et φ celle de X . On a

$$\forall t \in \mathbb{R}, \quad \varphi_n(t) = \frac{(\lambda/n) \exp(it/n)}{1 - (1 - \lambda/n) \exp(it/n)}, \quad \varphi(t) = \frac{\lambda}{\lambda - it}.$$

Comme $\lim_{z \rightarrow 0} (e^z - 1)/z = 1$, on a, pour tout $t \in \mathbb{R}$,

$$\varphi_n(t) = \frac{\lambda}{n(\exp(-it/n) - 1) + \lambda} \rightarrow \frac{\lambda}{\lambda - it} = \varphi(t).$$

Exercice. Reprendre l'exemple de la page 4 à l'aide du critère sur les fonctions caractéristiques.

2.2. Le TCL. Nous démontrons maintenant un résultat important de la théorie des probabilités : le théorème de la limite centrale.

Théorème 6. Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes, de même loi, de carré intégrable. On note $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \mathbb{V}[X_1]$; on suppose que $\sigma^2 > 0$. Alors,

$$\frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mu \right) \longrightarrow X, \quad \text{en loi, quand } n \rightarrow +\infty,$$

où X est une v.a.r. de loi $\mathcal{N}(0, 1)$ et $S_n = X_1 + \dots + X_n$.

Remarque. Comme la loi limite possède une fonction de répartition continue, on a, pour tout intervalle $]a, b[$,

$$\mathbb{P}(a < \sqrt{n}(n^{-1}S_n - \mu)/\sigma < b) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-u^2/2) du.$$

La loi des grands nombres nous dit que $n^{-1}S_n$ converge vers μ en moyenne quadratique. Le TCL montre que la vitesse de convergence est en $1/\sqrt{n}$. De plus, si n est suffisamment grand, $\sqrt{n}(n^{-1}S_n - \mu)/\sigma$ se comporte comme une v.a.r. normale centrée réduite.

On a en particulier, pour n assez grand,

$$\mathbb{P}(|n^{-1}S_n - \mu| \geq \alpha) = \mathbb{P}(|\sqrt{n}(n^{-1}S_n - \mu)/\sigma| \geq \alpha\sqrt{n}/\sigma) \simeq 2(1 - \Phi(\alpha\sqrt{n}/\sigma)),$$

où Φ est la fonction de répartition de la loi normale centrée réduite.

Connaissant deux des nombres n , α et ε , il est possible au moyen de cette approximation de déterminer le troisième pour que

$$\mathbb{P}(|n^{-1}S_n - \mu| \geq \alpha) \leq \varepsilon.$$

Le TCL se démontre en utilisant les fonctions caractéristiques. On note Y_k la variable aléatoire $(X_k - \mu)/\sigma$. On a donc $\mathbb{E}[Y_k] = 0$ et $\mathbb{V}[Y_k] = 1$ et les Y_k demeurent indépendantes. Avec ces notations, on doit montrer que

$$R_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \longrightarrow X, \quad \text{en loi quand } n \rightarrow +\infty$$

où X suit la loi normale centrée réduite. Déterminons la fonction caractéristique de la v.a. R_n , φ_n : par indépendance, on a, notant φ la fonction caractéristique de Y_1 ,

$$\forall t \in \mathbb{R}, \quad \varphi_n(t) = \varphi(t/\sqrt{n})^n ;$$

comme Y_1 est de carré intégrable, centrée réduite, sa fonction caractéristique est de classe C^2 et on a $\varphi'(0) = i\mathbb{E}[Y_1] = 0$ et $\varphi''(0) = i^2\mathbb{E}[Y_1^2] = -1$. φ possède donc un DL à l'ordre 2 en 0 de la forme

$$\varphi(t) = 1 - t^2/2 + t^2\varepsilon(t), \quad \text{avec } \lim_{t \rightarrow 0} \varepsilon(t) = 0.$$

Soit $t \in \mathbb{R}$ fixé. Puisque $|z^n - w^n| \leq n|z - w|$ si $|z| \leq 1$ et $|w| \leq 1$, pour n assez grand, $1 - t^2/(2n)$ appartient à $[0, 1]$, et on a comme $|\varphi(t/\sqrt{n})| \leq 1$,

$$\left| \varphi_n(t) - \left(1 - \frac{t^2}{2n}\right)^n \right| = \left| \left(1 - \frac{t^2}{2n} + \frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)\right)^n - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq n \frac{t^2}{n} \left| \varepsilon\left(\frac{t}{\sqrt{n}}\right) \right|.$$

Comme $\ln(1+z)/z \rightarrow 1$ si $z \rightarrow 0$, on obtient

$$\lim_{n \rightarrow +\infty} \left(1 - \frac{t^2}{2n}\right)^n = \lim_{n \rightarrow +\infty} \exp \left\{ n \ln \left(1 - \frac{t^2}{2n}\right) \right\} = \exp(-t^2/2),$$

ce qui entraîne, puisque $\varepsilon(t/\sqrt{n})$ tend vers 0 si $n \rightarrow +\infty$,

$$\forall t \in \mathbb{R}, \quad \lim_{n \rightarrow +\infty} \varphi_n(t) = \exp(-t^2/2),$$

ce qui démontre le TCL via le Théorème 5.