De la théorie de la mesure aux probabilités: une petite traduction

C. Vernier

2 février 2023

Pour ce petit texte, je me suis essentiellement basée sur l'excellent livre de W. Appel, *Probabilités pour les non-probabilistes*, qui est disponible à la BU...enfin, qui sera disponible quand j'aurai fini de le monopoliser.

1 Des probabilités intuitives à la mesure des ensembles

1.1 Minute philosophique

Qu'est-ce qu'une probabilité, au fond ? [1]

$\rightsquigarrow$ Les probabilités, c'est la modélisation de phénomènes aléatoires, autrement dit des expériences qui, même si on les répète dans des circonstances quasi-identiques, auront des résultats variables ; et dans ce cas, on ne peut pas espérer faire de prédiction absolue.

Pour modéliser ça, on va associer à chaque "futur possible" un élément ω d'un ensemble Ω.

Exemples : Souvent, on se contente de définir un "futur possible" par *résultat* de l'expérience qui nous intéresse :

1. Pour un lancer de dés, on peut par exemple utiliser l'ensemble $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. Pour une succession infinie de jeu de pile ou face (il faut vraiment beaucoup s'ennuyer) : l'ensemble des suites infinies du type $PFPPFPPF$... : $\Omega = \{P, F\}^{\mathbb{N}^*}$.
3. Pour l'arrivée d'un train à Bordeaux (phénomène aléatoire s'il en est) : $\Omega = \overline{\mathbb{R}}^+$

Définition 1

- *On appelle (pompeusement)* Ω *l'*univers des possibles*, ou juste l'*univers *(en anglais* sample space*)*
- *Un élément* $\omega \in \Omega$ *est appelé* épreuve *ou* réalisation du hasard *(en anglais* sample point*).*
- *Un sous-ensemble* $A \subset \Omega$ *est appelé un* évènement*.*

Remarque : On pourrait s'attendre à ce qu'un "évènement" soit un élément de Ω, mais l'avantage de plutôt s'intéresser aux sous-ensembles, c'est qu'on peut traiter plus facilement la probabilité d'évènements qui sont réalisés par plusieurs futurs possibles : par exemple, dans le cas du tirage d'un dé, modélisé par

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

on peut s'intéresser à l'évènement "le résultat est 3" $A = \{3\}$, mais aussi à des choses comme

- "Le résultat est pair" $\rightsquigarrow A = \{2, 4, 6\}$
- "Le résultat est au moins 5" $\rightsquigarrow A = \{5, 6\}$
- "Le résultat est plus petit que 42" $\rightsquigarrow A = \Omega$
- "Le résultat est un multiple de π " $\rightsquigarrow A = \emptyset$

1. (Non, vraiment)

Maintenant, pour faire des probabilités, on aimerait associer à chaque évènement une quantité qui mesure s'il est plus ou moins susceptible de se produire.

Par exemple, si $\Omega = \{\omega_1, \omega_2, ...\}$ est un ensemble fini ou dénombrable, on pourrait associer à chaque élément $\omega \in \Omega$ une probabilité $p(\omega) \in [0,1]$; et dans ce cas on pose

$$\forall A \subset \Omega, \mathbb{P}(A) = \sum_{a \in A} p(\omega).$$

Mais si A n'est pas dénombrable, on ne peut pas faire ça : supposons qu'on tire un réel au hasard dans l'intervalle $\Omega = [0,1]$. Quelle est la probabilité de tomber *précisément* sur votre préféré ?

$\rightsquigarrow$ Ne le prenez pas mal, chaque élément de Ω est associé à une probabilité 0. Du coup, faut-il poser $p(\omega) = 0$ pour tout ω ?

Ce n'est pas terrible, parce qu'on aimerait que l'évènement certain "Le résultat est compris entre 0 et 37", correspondant à $A = \Omega$, ait une probabilité de 1, et ça donnerait

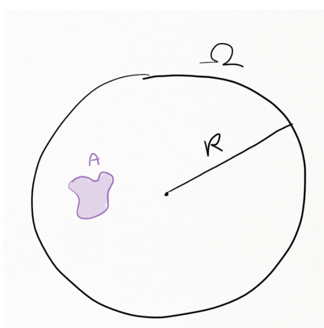
$$1 = \mathbb{P}(\Omega) = \text{“} \sum_{\omega \in [0,1]} p(\omega)\text{”} = \text{“} \sum_{\omega \in [0,12]} p(\omega) = 0\text{”}$$

ce qui est un problème.

En revanche, il semble raisonnable d'évaluer la probabilité d'obtenir un réel entre 0 et $\frac{1}{4}$ à $\frac{1}{4}$.

$\rightsquigarrow$ Plutôt que d'associer un nombre à chaque élément ω de Ω, on aimerait associer un nombre à chaque *évènement* de Ω : pour tout $A \subset \Omega$, $\mathbb{P}(A)$ est la probabilité de tomber sur un élément de A, et ici il semble naturel de choisir la longueur de l'intervalle, si A est un intervalle.

:



Dans le même ordre d'idée, si on joue aux fléchettes avec une cible de rayon R, alors on peut modéliser l'expérience en prenant pour Ω un disque de rayon R et il semble plutôt naturel de définir la probabilité d'un évènement $A \subset \Omega$ par

$$\mathbb{P}(A) = \frac{Aire(A)}{Aire(D)}$$

Ca a l'air super... sauf qu'en fait on ne peut pas faire ça. Reprenons notre expérience de tirage au sort d'un réel dans $[1,2]$. On a dit que, si $A \subset [1,2]$ est un intervalle, mettons $A = [a, b[$, il semble raisonnable de prendre

$$\mathbb{P}(A) = b - a$$

Mais évidemment, tous les sous-ensembles de $[0,1]$ ne sont pas des intervalles ! Comment évaluer, par exmeple, la probabilité de l'évènement A="le résultat est de la forme $\frac{\pi}{\sqrt{n}}$" ? Ou alors "A=le résultat n'a que des 5 et des 7 dans son développement décimal" ?

$\rightsquigarrow$ C'est pour ça qu'il nous faut la théorie de la mesure ! Elle nous permettra de mesurer des sous-ensembles dans des ensembles aussi bien finis que continus.

L'ennui, c'est qu'un italien contrariant du nom de Giuseppe Vitali a montré qu'il n'existait aucune application

$$m : \mathcal{P}(\mathbb{R}) \to [0, +\infty]$$

qui vérifie

- Si I est un intervalle, $m(I)$ est sa longueur $\sup I - \inf I$;
- Si A et B sont disjoints, $m(A \sqcup B) = m(A) + m(B)$, et plus généralement, pour toute suite de sous-ensembles disjoints $(A_k)_{k\in\mathbb{N}}$, $m(\bigsqcup A_k) = \sum_m (A_k)$
- Pour tout $A \subset \mathbb{R}$, $\mathbb{P}(A + t) = \mathbb{P}(A)$ (ça, c'est le côté "uniforme")

(autrement dit, il n'existe pas de mesure invariante par translation sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$.

Voir ici les ensembles qu'il a construits pour montrer ça :
http://carolinevernier.website/compl/non-lebesgue-mesurable.pdf
(ou commencer par la vidéo "How the Axiom of Choice Gives Sizeless Sets", qui illustre bien l'idée derrière leur construction :
https://youtu.be/hcRZadc5KpI

Et du coup, évidemment, il n'y a pas de fonction qui mesure raisonnablement l'aire de toute partie de $\mathbb{R}^2$, ou le volume de toute partie de $\mathbb{R}^3$: une sympathique application de ça est le *paradoxe de Banach-Tarski*. Voir cette vidéo d'El Jj pour une courte introduction, et celle-ci pour plus d'explications. Et terminer par celle-là pour toujours plus de fun.

Dans le même ordre d'idées, le polonais Stanislaw Ulam a montré que, sur un ensemble continu Ω, les seules probabilités définies sur $(\Omega, \mathcal{P}(\Omega))$ sont des probabilités "discrètes", autrement dit qu'il existe un sous-ensemble dénombrable $D \subset \Omega$ tel que $\mathbb{P}(\Omega \setminus D) = 0$ (la probabilité est nulle partout sauf sur l'ensemble D...ce qui ne nous permet pas d'étudier grand chose.)

On revoit donc nos ambitions un peu à la baisse : plutôt que de définir une probabilité comme une application

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$$

on commence par définir, parmi tous les sous-ensembles de Ω, une sous-famille $\mathscr{T}$ d'évènements, plus précisément une *tribu*, qu'on va pouvoir gérer [2].

Du coup, en fin de compte :

Définition 2

On appelle :

- espace probabilisable *un couple* $(\Omega, \mathscr{T})$*, où* Ω *est notre univers des possibles, et* $\mathscr{T}$ *une tribu sur* Ω*.*
- évènements *les sous-ensembles* $A \subset \Omega$ qui appartiennent à $\mathscr{T}$ *(pas forcément tous les sous-ensembles de* Ω*, donc).*
- probabilité *sur* $(\Omega, \mathscr{T})$ *une application*
$$\mathbb{P} : \mathscr{T} \to [0, 1]$$
telle que $\mathbb{P}(\Omega) = 1$ *et* $\mathbb{P}$ *est* σ*-additive : Ce sont les* axiomes de Kolmogorov *des probabilités, d'après le mathématicien Andrei Kolmogorov qui, le premier, a introduit le formalisme de la théorie de la mesure en probabilités. Maintenant, vous savez de qui c'est la faute !*
- espace probabilisé *un triplet* $(\Omega, \mathscr{T}, \mathbb{P})$*, où* $(\Omega, \mathscr{T})$ *est un espace probabilisable et* $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathscr{T})$*.*
- *On dit qu'un évènement* $A \in \mathscr{T}$ *est* négligeable *si* $\mathbb{P}(A) = 0$ *; et plus généralement, on dit qu'un sous-ensemble* $N \subset \Omega$ *est négligeable s'il existe un évènement négligeable* A *tel que* $N \subset A$*.*
- *On dit qu'un évènement est* presque sûr *si* $\mathbb{P}(A) = 1$*.*
- *On dit qu'une propriété* $\mathscr{P}(\omega)$ *est* vraie presque sûrement *si l'ensemble* $\{\omega \in \Omega, non(\mathscr{P}(\omega))\}$ *est négligeable (si cet ensemble est un évènement, ce qui est quasiment tojours le cas, cela revient à dire que l'évènement* $\{\omega \in \Omega, \mathscr{P}(\omega)\}$ *est presque sûr).*

$\rightsquigarrow$ Du coup, un espace probabilisable, c'est la même chose qu'un espace mesurable en théorie de la mesure, et une probabilité, c'est juste une mesure positive telle que $\mathbb{P}(\Omega) = 1$.

$\rightsquigarrow$ Avec ce cadre (merci Andrei), on va pouvoir utiliser toutes les propriétés qu'on connaît sur la théorie de la mesure (notamment, COMOSEQ) pour traiter les probabilités : et le même cadre va nous permettre de faire des probabilités discrètes et continues.

Et inversement, pour les notions purement probabilistes, comme les fonctions de répartition, le conditionnement ou l'indépendance, on va chercher quelle notion de théorie de la mesure permet de "dire ce qu'on veut dire".

2. Dans la pratique, tous les ensembles "intéressants" vont être gérables. Il faut se lever de bonne heure pour trouver un sous-ensemble dans $\mathbb{R}$ qui ne soit pas un borélien !
De plus, en fait, cette restriction à des tribus $\mathscr{T}$ s'avèrera très utile ssi on la manie bien : on verra un peu plus tard qu'on peut associer une tribu à toute variable aléatoire X, et cette tribu contient enquelque sorte "l'information portée par X" (je sais, c'est flou pour le moment).

Conditionnement - une première idée

Soit $(\Omega, \mathscr{T}, \mathbb{P})$ un espace probabilisé.

Question : Comment mathématiser l'idée de "probabilité qu'un évènement B soit réalisé sachant qu'un évènement A est réalisé" ?

L'idée, c'est qu'on étudie la probabilité de réalisation de B, non pas parmi toutes les réalisations $\omega \in \Omega$, mais seulement parmi celles qui réalisent A : et parmi celles là, B est réalisé si et seulement si $A \cap B$ est réalisé.

Donc, en quelque sorte, on travaille sur un nouvel univers $\Omega' = A$, muni d'une nouvelle probabilité P_A proportionnelle à $\mathbb{P}(A \cap B)$, mais pas *égale* à $\mathbb{P}(A \cap B)$, car on veut que P_A soit une probabilité sur A, donc on veut que $P_A(A) = 1$. Ce qui nous donne

Définition 3

Soit $A \in \mathscr{T}$ un évènement tel que $\mathbb{P}(A) > 0$. Alors l'application

$$P_A : B \in \mathscr{T} \mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \in [0, 1]$$

est une probabilité sur $(\Omega, \mathscr{T})$.

(On peut aussi voir P_A comme une probabilité sur A munie de la tribu-trace $\mathscr{T}_A = \{A \cap B, B \in \mathscr{T}\}$ qu'on appelle probabilité conditionnelle de A sachant B*).*

On note souvent, surtout de l'autre côté de la Manche et de l'Atlantique,

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

la probabilité conditionnelle de B sachant A (mais avec cette notation on voit moins bien le côté "définition d'une nouvelle mesure de probabilités").

Avec cette définition, les propriétés de base des mesures nous permettent de retrouver quelques formules classiques :

- **Probabilités composées :** Pour tous évènements $A, B \in \mathscr{T}$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B)$.

  Plus généralement, si $A_1, ..., A_n \in \mathscr{T}$ sont n évènements,

$$\mathbb{P}(A_1 \cap ... \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)...\mathbb{P}(A_n|A_1 \cap ... \cap A_{n-1})$$

- **Probabilités totales :** Si $(A_i)_i \in \mathscr{T}^{\mathbb{N}}$ est une partition finie ou dénombrable de Ω en évènements disjoints, et si pour tout i, $\mathbb{P}(A_i) > 0$, alors

$$\mathbb{P}(B) = \sum_i \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

  pour tout $B \in \mathscr{T}$.

- **Formule de Bayes :** Pour tous $T, D \in \mathscr{T}$ tels que $\mathbb{P}(T) > 0, \mathbb{P}(D) > 0$, $\mathbb{P}(T|D) = \frac{\mathbb{P}(T)\mathbb{P}(D|T)}{\mathbb{P}(D)}$, et plus généralement si $(T_i)_{i \in I} \in \mathscr{T}^{\mathbb{N}}$ est une partition finie ou dénombrable de Ω en évènements disjoints, et si pour tout i, $\mathbb{P}(T_i) > 0, \mathbb{P}(D) > 0$ alors

$$\forall p \in I, \mathbb{P}(T|D) = \frac{\mathbb{P}(T_p)\mathbb{P}(D|T_p)}{\mathbb{P}(D)} = \frac{\mathbb{P}(T_p)\mathbb{P}(D|T_p)}{\sum_i \mathbb{P}(D_i)\mathbb{P}(D|T_i)}$$

  $\rightsquigarrow$ L'idée est de répondre à la question suivante : supposons qu'on souhaite tester une théorie T, dont on estime la probabilité d'être vraie à $\mathbb{P}(T)$, et qu'on a obtenu les données D. Maintenant qu'on a ces données, comment devrait-on modifier notre estimation de la probabilité que la théorie soit vraie, (sachant les données D du coup) ?

  Et la formule de Bayes nous permet de calculer ça à partir de quantités plus faciles : la probabilité de D, qu'on peut exprimer avec la formule des probabilités totales, et la probabilité d'obtenir D *si* T était vraie.

2 Variables aléatoires

Jusqu'ici, on a considéré des cas assez simples où l'ensemble univers Ω a pour éléments les résultats possibles d'une seule expérience (lancer de dé, pile ou face, arrivée du Nantes-Bordeaux, etc.).

Mais dans beaucoup de cas intéressants, on peut avoir envie de réaliser plusieurs expériences un peu différentes sur un même ensemble : par exemple, dans un composé radioactif, la probabilité de décomposition de deux atomes différents ; à la gare de Bordeaux, l'arrivée du TGV Paris-Bordeaux d'une part et le départ du TER Bordeaux-Condat-le-Lardin ; aux fléchettes, la probabilité que Stephen Hawking marque et celle que Natasha Romanoff marque.

Dans ce cas, on peut changer de point de vue : on considère un univers des possibles absolument monstrueux Ω, dont chaque réalisation correspond à un cas particuliers de circonstances (températures, pression, météo, chute d'un arbre sur les voies à St-Médard-de-Guizière, hauteur des talons de Hawking et de Romanoff), et *on ne va surtout pas chercher à définir Ω ou à le comprendre*. On va simplement supposer qu'on sait mesurer certains sous-ensembles de Ω : les évènements contenus dans une tribu $\mathscr{T}$.

A la place de Ω, on va s'intéresser à des fonctions $X : \Omega \to \mathbb{R}$ qui mesurent le résultats de diverses expériences qu'on peut faire dans Ω.

Remarque : $X(\Omega)$ est donc un sous-ensemble de $\mathbb{R}$, mais ça n'a pas besoin d'être $\mathbb{R}$ tout entier : par exemple si $X(\omega)$ est un ensemble fini $\{x_1, \ldots, x_n\}$ ou dénombrable $\{x_k, k \in \mathbb{N}\}$, alors on dit X est une variable aléatoire *discrète*.

Ce qui ne nous empêche pas de dormir : du moment que les valeurs x_i que prend X sont des réels, X est bien une variable aléatoire réelle, et donc les v.a. discrètes rentrent dans notre cadre d'étude.

Exemple plus concret : Reprenons l'exemple ultra-classique d'un lancer de dé [3]. L'idée, plutôt que de travailler avec $\Omega = \{1, 2, 3, 4, 5, 6\}$, c'est de prendre un ensemble Ω qui prend en compte tous les facteurs mécaniques, atmosphériques, naturels et surnaturels qui influencent la trajectoire du dé, et de s'intéresser à une fonction $X : \Omega \to \{1, 2, 3, 4, 5, 6\}$ qui à chacune de ces circonstances associe le résultat final.

On a dit qu'on ne chercherait pas à comprendre Ω, et on ne supposera donc pas qu'il a une structure particulière : on ne sais pas, notamment, si c'est un espace vectoriel, si oui normé, etc...

Et du coup, on ne va pas s'intéresser aux propriétés des fonctions $X : \Omega \to \mathbb{R}$ qui dépendent de ça : la continuité, la différentiabilité, etc. Ce qui va nous occuper, c'est la *distribution* des valeurs de X :

- quelle est la probabilité que $X(\omega) = 0$?
- quelle est la probabilité que $X(\omega) \leq 207$?

$\rightsquigarrow$ C'est ce qu'on va appeler la *loi* de X.

Or, l'ensemble des réalisations du hasard ω telles que $X(\omega) = 0$, c'est l'ensemble

$$\{\omega \in \Omega | X(\omega) = 0\} = X^{-1}(\{0\})$$

et donc la probabilité que $X = 0$ est la mesure de probabilité de cet ensemble :

$$\mathbb{P}(X^{-1}(\{0\}))$$

et plus généralement, la probabilité que $X(\omega)$ soit dans un certain sous-ensemble $A \subset \mathbb{R}$ est

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

$\rightsquigarrow$ Il faut qu'on puisse *mesurer* $X^{-1}(A)$ pour les sous-ensembles de $\mathbb{R}$ qui nous intéressent (et dans tous les cas pratiques, ce seront des boréliens).

Définition 4

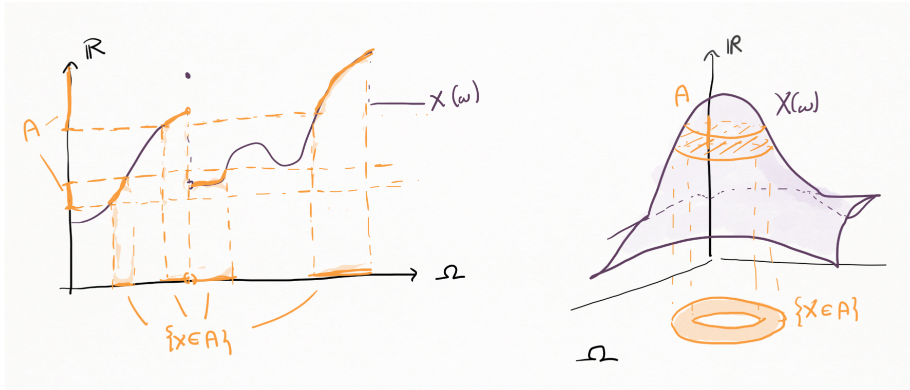
Soit $(\Omega, \mathscr{T})$ un espace probabilisable. Une variable aléatoire réelle *est une application mesurable*

$$X : (\Omega, \mathscr{T}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$$

Autrement dit, pour tout borélien $B \in \mathscr{B}(\mathbb{R})$, $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in A\}$ est un évènement de $\mathscr{T}$. On notera souvent $\{X \in B\}$ l'évènement $X^{-1}(B)$.

L'ensemble des variables aléatoires réelles sur $(\Omega, \mathscr{T})$ est donc $\mathcal{L}^0(((\Omega, \mathscr{T}); \mathbb{R})$.

3. Remarquez, non seulement c'est l'exemple classique en probabilités, mais les dés sont à la base de tout le vocabulaire probabiliste : "alea" signifie "dé" en latin, "chance" vient de "cadere" qui veut dire "tomber" en latin...comme des dés, et en arabe, "al-zahr" a donné l'espagnol "azar" qui nous a donné "hasard". Et les tous premiers probabilistes, à l'époque des Lumières, s'intéressaient essentiellement à des problèmes de casino : dés et cartes.



A nouveau, du coup, tous les résultats qu'on connaît sur les applications mesurables s'exportent donc sans douleur en théorie des probabilités :

- Une somme,produit, composée, max ou min de deux variables aléatoires est une variable aléatoire ;
- Le sup, l'inf, la limite simple (si elle existe) d'une suite de variables aléatoire est aussi une variable aléatoire ;
- Une fonction $X : \Omega \to \mathbb{R}$ est une variable aléatoire ssi pour tout $a \in \mathbb{R}$, $\{X \leq A\} = X^{-1}(]-\infty, a])$ appartient à $\mathscr{T}$.

Pour étudier la répartition des valeurs d'une v.a. réelle X selon les réalisations du hasard, on va introduire une nouvelle probabilité P_X sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$:

Définition 5

Soit X une variable aléatoire réelle sur $(\Omega, \mathscr{T}, \mathbb{P})$. On appelle loi de X *la probabilité image de $\mathbb{P}$ par X : c'est la probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ définie par*

$$P_X : B \in \mathscr{B}(\mathbb{R}) \mapsto \mathbb{P}(X^{-1}(B))$$

Autrement dit P_X est la mesure image *de $\mathbb{P}$ par X.*

Il se trouve qu'en fait, pour étudier la loi d'une variable aléatoire X, on peut se contenter de regarder ce que vaut, pour chaque $t \in \mathbb{R}$, la probabilité que $X \leq t$: cette information caractérise complètement la loi de X. On va donc lui donner un nom :

Définition 6

Soit $X : \Omega \to \mathbb{R}$ une v.a. réelle. On appelle fonction de répartition *de X la fonction*

$$F_X : t \in \mathbb{R} \mapsto \mathbb{P}(\{X \leq t\}) \in [0, 1]$$

$\rightsquigarrow$ *C'est une fonction croissante, continue à droite mais pas forcément continue, et telle que*

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to +\infty} F_X(x) = 1,$$

Remarque : Petit rappel d'analyse réelle Dire que F_X est *continue à droite* signifie que, pour tout $a \in \mathbb{R}, \lim_{x \to a, x > a} F_X(x) = F_X(a)$.

Notons $F_X(a^-) = \lim_{x \to a, x < a} F_X(x)$. Puisque F_X est continue à droite, mais pas forcément à gauche, il se peut que $F_X(a^-) \neq F(a)$. En fait, cela arrive quand X a une probabilité non nulle de valoir *exactement* a :

Proposition 1

Pour tout $a \in \mathbb{R}$, $P_X(\{a\}) = \mathbb{P}(\{X = a\}) = F(a) - F(a^-)$.

Preuve : Remarquons que

$$\{a\} = \bigcap_{n \geq 1} \left] a - \frac{1}{n}, a \right]$$

donc par COMOSEQ,

$$\begin{aligned}
P_X(\{a\}) &= \lim_{n\to\infty} P_X\left(\left]a-\frac{1}{n},a\right]\right)\\
&= \lim_{n\to\infty} \mathbb{P}\left(X\in\left]a-\frac{1}{n},a\right]\right)\\
&= \lim_{n\to\infty} \mathbb{P}\left(X\in]-\infty,a] \text{ et } X\notin\in\left]-\infty,a-\frac{1}{n}\right]\right)\\
&= \lim_{n\to\infty} F_X(a)-F_X(a-\frac{1}{n})\\
&= F(a)-F(a^-)
\end{aligned}$$

Remarque : Qu'entend-on par "la fonction de répartition caractérise la loi de la variable aléatoire" ?

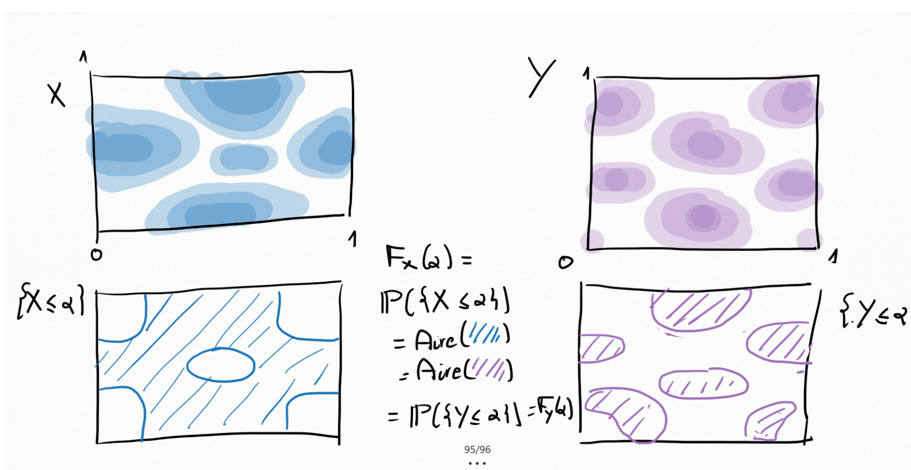
Un résultat pas si simple[4] de théorie de la mesure permet de montrer que

Proposition 2

Si $X, Y : \Omega \to \mathbb{R}$ sont deux v.a. réelles telles que $F_X = F_Y$, alors les lois de X sont les mêmes : $P_X = P_Y$.

Attention toutefois, ça ne veut pas dire que X et Y sont les mêmes !

Par exemple, sur $(\Omega, \mathscr{T}, \mathbb{P}) = ([0,1]^2, \mathscr{B}([0,1]^2, \mu_2)$ (où λ_2 est la mesure de Lebesgue sur $\mathbb{R}^2$), imaginons deux v.a. $X, Y : [0,1]^2 \to \mathbb{R}$ représentées ci-dessous (des couleurs plus foncées représentent de plus grandes valeurs de X et Y) :



Alors on peut très bien avoir, pour n'importe quel $\alpha \in \mathbb{R}$,

$$F_X(\alpha) = \lambda_2(\{X \leq \alpha\}) = \lambda_2(\{Y \leq \alpha\}) = F_Y(\alpha)$$

même si les ensembles $\{X \leq \alpha\}$ et $\{Y \leq \alpha\}$ sont différents.

Variables aléatoires discrètes

Un cas particulier intéressant est celui des probabilités et variables aléatoires discrètes. Considérons donc une v.a. réelle $X : \Omega \to \mathbb{R}$ telle que $X(\Omega)$ est dénombrable (X peut prendre seulement un nombre fini ou dénombrable de valeurs différentes, par exemple si X modélise, je ne sais pas, disons, le lancer d'un dé).

On note

$$X(\Omega) = \{x_k, k \in \mathbb{N}\}, \text{ et pour tout } k \in \mathbb{N}, p_k = \mathbb{P}(\{X = x_k\})$$

4. Il s'agit d'une application du "lemme des classes monotones", qui permet de montrer de nombreux résultats d'unicité sur les mesures, notamment l'unicité de la mesure de Borel sur $\mathbb{R}$.

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

$$\begin{aligned}\{X \in A\} &= \{\omega \in \Omega, X(\omega) \in A\} \\ &= \{\omega \in \Omega, X(\omega) \in A \cap X(\Omega)\} \\ &= \{\omega \in \Omega, \exists k \in \mathbb{N}, X(\omega) = x_k, x_k \in A\} \\ &= \bigsqcup_{k \mid x_k \in A} \{X = x_k\}\end{aligned}$$

donc, par σ-additivité de $\mathbb{P}$,

$$\mathbb{P}(\{X \in A\}) = \sum_{k \mid x_k \in A} \mathbb{P}(\{X = x_k\}) = \sum_{k \mid x_k \in A} p_k$$

Mais du coup, la loi de X est la probabilité sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ donnée par

$$P_X(A) = \sum_{k \mid x_k \in A} p_k = \sum_{k \in \mathbb{N}} p_k \mathbb{1}_A(x_k)$$

Or, si on note δ_{x_k} la *mesure de Dirac en* x_k, on a pour tout $A \subset \mathbb{R}$

$$\delta_{x_k}(A) = \begin{cases} 1 \text{ si } x_k \in A \\ 0 \text{ si } x_k \notin A \end{cases} = \mathbb{1}_A(x_k)$$

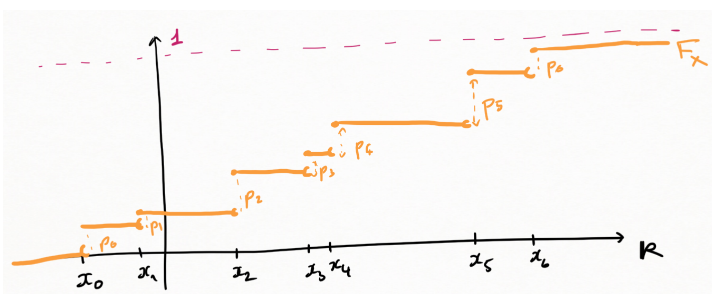
donc au total, la loi de X est une somme de mesures de Dirac :

$$P_X = \sum_{k \in \mathbb{N}} p_k \delta_{x_k}$$

Supposons de plus, quitte à les réarranger que la suite $(x_k)_k$ est croissante. Alors pour tout $\alpha \in \mathbb{R}$,

$$F_X(\alpha) = \mathbb{P}\{X \leq \alpha\} = \sum_{k \mid x_k \leq \alpha} p_k$$

donc F_X ressemble à une fonction "en escaliers" (prenant éventuellement un nombre infini, mais dénombrable, de valeurs), croissante, avec des "sauts" de taille p_k à chaque x_k. Autrement dit, $F_X(x_k) - F_X(x_k^-) = p_k = \mathbb{P}(\{x_k\})$.



Variables aléatoires à densité

Dans le cas de variables aléatoires tels que $X(\Omega)$ n'est pas dénombrable*On dit que $X(\Omega)$ a "la puissance du continu", ce qui en jette*, on va rarement s'intéresser à la probabilité $\mathbb{P}(\{X = a\}$ que X prenne une valeur spécifique : en général, cette probabilité sera 0.

A la place, on va s'intéresser à la probabilité que X soit très proche de a,mettons dans un intervalle $[\,a, a+\delta\,[$. A priori, plus δ est petit, plus cette propriété va être faible, et inversement. Il ne serait donc pas déraisonnable (en tout cas tant qu'on reste proche de a) qu'elle soit proportionnelle à δ :

$$\mathbb{P}(\{X \in [\,a, a+\delta\,[\,\}) \simeq f_X(a) \cdot \delta$$

avec un facteur de proportionnalité $f(a)$ qui, pour autant qu'on sache, dépend de a.

Mais dans ce cas, en utilisant la définition de la fonction de répartition F_X, la fonction f ainsi définie vérifierait

$$f_X(a) = \lim_{\delta \to 0} \frac{\mathbb{P}(\{X \in [\,a, a+\delta\,[\,\})}{\delta} = \lim_{\delta \to 0} \frac{F_X(a+\delta) - F_X(a)}{\delta}$$

En particulier, si F_X était dérivable en a, alors $f_X(a) = F_X'(a)$, et on aurait du coup, pour tous $a, b \in \mathbb{R}$

$$P_X([a,b])\mathbb{P}(\{X \in [\,a, b\,]\,\})) = F_X(b) - F_X(a) = \int_a^b f_X(t)dt, \text{ et } F_X(x) = \mathbb{P}(\,[-\infty, x\,[\,) = \int_{-\infty}^x f_X(t)dt$$

et, à grands coups de théorèmes de théorie de la mesure, on en déduit en fait que pour tout borélien $A \in \mathscr{B}(\mathbb{R})$

$$P_X(A) = \mathbb{P}(\{X \in A\}) = \int_A f_X(t)d\lambda_1(t)$$

où λ_1 est la mesure de Lebesgue sur $\mathbb{R}$.

$\rightsquigarrow$ Dans ce cas, la mesure de probabilité P_X est en fait la mesure de densité f_X. Ce cas où F_X est une sympathique fonction (presque partout) dérivable et $f_X = F_X'$ est (presque partout) continue donne le cadre de la plupart des lois continues usuelles : loi uniforme, exponentielle, de Cauchy, Gamma, bref tout le zoo.

Mais les lois à densité sont en fait plus générales :

Définition 7

— On dit qu'une variable aléatoire réelle $X : \Omega \to \mathbb{R}$ est à densité *(ou* absolument continue*) s'il existe une fonction borélienne $f : \mathbb{R} \to \mathbb{R}^+$ telle que, pour tout $A \in \mathscr{B}(\mathbb{R})$,*

$$P_X(A) = \int_A f_X(t)d\lambda_1(t)$$

(autrement dit, si P_X est la mesure de dnesité f_X par rapport à la mesure de Lebesgue) — La fonction f_X est alors appelée densité de X. *Elle vérifie*

$$\int_{\mathbb{R}} f_X \, d\lambda = \mathbb{P}(X \in \mathbb{R}) = 1.$$

Remarque : Dans ce cas, puisque les singletons sont λ_1-négligeables, on a, pour tout $a \in \mathbb{R}$,

$$\mathbb{P}(\{X = a\} = P_X(\{a\}) = \int_A f_X(t)\, d\lambda_1(t) = 0$$

$\rightsquigarrow$ Une v.a. réelle à densité a une probabilité nulle de prendre une valeur particulière : c'est un peu l'opposé des v.a. discrètes.

Exemples de variables aléatoires "mixtes"

Cela dit, toutes les variables aléatoires ne sont pas discrètes ou continues : certaines sont un peu des deux, avec à la fois des singletons de probabilité non nulle et des intervalles où F_X est continue (et éventuellement dérivable). Voyons quelques exemples

Exemple 1 : Comment pourrait-on modéliser le temps d'attente à un feu rouge ?

L'idée d'un feu rouge, c'est d'alterner entre une phase où il est au vert (de durée disons v) et une phase où il est au rouge (de durée, au hasard, r) [5] avec des unités choisies pour que $r + v = 1$ (ce qui va nous simplifier les calculs).

$\rightsquigarrow$ On définit une variable aléatoire $X : \Omega \to \mathbb{R}$ qui représente le temps d'attente d'une voiture au feu, en supposant que l'arrivée de la voiture tombe avec la même probabilité à n'importe quel moment du cycle : autrement dit, l'instant d'arrivée au feu suit une loi uniforme sur $[0, 1]$.

5. Nous sommes à Paris, le feu orange n'existe pas vraiment

Ainsi :

- On ne peut pas attendre un temps négatif :

$$\forall t < 0,\ \mathbb{P}(\{X \leq t\}) = 0 \text{ donc } \forall t < 0,\ F_X(t) = 0.$$

On en déduit en particulier que $F_X(0^-) = 0$

- Le temps d'attente est au maximum r :

$$\forall t \geq r,\ \mathbb{P}(\{X \leq t\}) = 1 \text{ donc } \forall t \geq r,\ F_X(t) = 1.$$

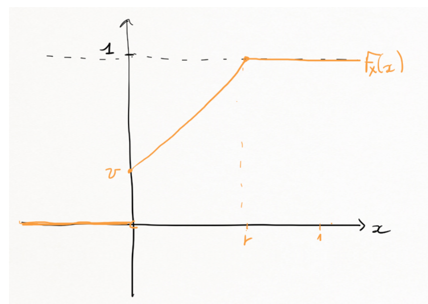
- La probabilité de ne pas attendre du tout est égale à la probabilité d'arriver pendant la phase où le feu est vert :

$$\begin{aligned} \mathbb{P}(\{X = 0\}) = v &\rightsquigarrow F_X(0) - F_X(0^-) = v \\ &\rightsquigarrow F_X(0) = v \end{aligned}$$

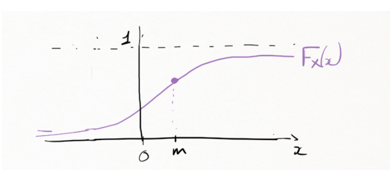
- Enfin, pour la probabilité d'attendre un temps inférieur ou égal à x, pour $x \in]0, r[$, notons V l'évènement "le feu est vert à l'instant d'arrivée" et R l'évènement "le feu est rouge à l'instant d'arrivée". Alors $V = R^c$ et, à coups de probabilités totales,

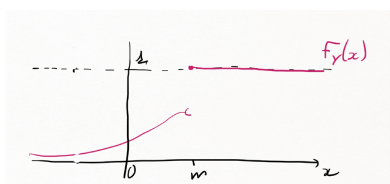
$$\begin{aligned} F_X(x) &= \mathbb{P}(\{X \leq x\}) \\ &= \mathbb{P}(X \leq x | R)\mathbb{P}(R) + \mathbb{P}(X \leq x | V)\mathbb{P}(V) \\ &= \frac{x}{r} \cdot r + 1 \cdot v = x + v \end{aligned}$$

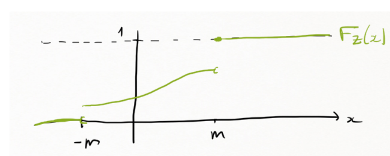
Ce qui nous donne une fonction de répartition ni discrète ni continue.



Exemple 2 : Plus généralement, on tombe facilement sur de telles variables aléatoires en tronquant une variable aléatoire continue X : par exemple, on peut définit $Y = \min(X, m)$ pour une certaine valeur prédéfinie m, ou encore $Z = X\mathbb{1}_{|X| \leq m}$. Ce qui nous donne :







3 Espérance d'une variable aléatoire

Au fond...qu'est-ce que l'espérance ?[6]

Cas des v.a. discrètes

Heuristique : L'espérance d'une v.a. discrète, ce serait a priori la "valeur moyenne" prise par une variable aléatoire pour toutes les réalisations du hasard, pondérée par la probabilité de chaque réalisation, quelque chose comme

$$m = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \qquad (*)$$

ce qui, en réunissant les termes[7] où $X(\omega)$ prend une même valeur x, devrai coïncider avec

$$m = \sum_{x \in X(\Omega)} x\mathbb{P}(\{X = x\}) \qquad (**)$$

6. (Non, vraiment)
7. Supposons qu'on en a le droit !

D'un autre côté, on s'attend à ce que cette moyenne théorique coïncide avec la moyenne empirique qu'on obtient "en répétant l'expérience une infinité" de fois. Autrement dit, si $X_1, X_2, ...$ sont des v.a. indépendantes ayant toutes la même loi que X, on s'attend à ce que

$$m(\omega) = \lim_{n\to\infty} \frac{1}{n}(X_1(\omega) + ... + X_n(\omega)) \qquad (***)$$

soit (presque partout) constante égale à m.

$\rightsquigarrow$ Mais encore faut-il que les sommes $(*)$ et $(**)$ existent !

- Evidemment, si Ω est fini (ou si $X(\Omega)$ l'est), ces sommes ne posent pas de problème particulier :

Exemple : Prenons $\Omega = \{1, 2, 3, 4, 5, 6\}$ avec sa probabilité uniforme $\mathbb{P}(\{\omega\}) = \frac{1}{6}$, et introduisons une v.a. X sur Ω, mettons

$$X : \omega \in \Omega \mapsto \begin{cases} \omega & \text{si } \omega \text{ est pair} \\ \frac{\pi^2}{6} & \text{si } \omega \text{ est impair} \end{cases}$$

Alors il semble raisonnable que l'espérance de X soit

$$\sum_{k=1}^{6} \frac{X(k)}{6} = 2 + \frac{\pi^2}{12}$$

- Si $X(\Omega) = \{x_k, k \in \mathbb{N}\}$, il s'agit donc de donner un sens à

$$\sum_{k\in\mathbb{N}} x_k \mathbb{P}(\{X = x_k\}) = \sum_{k\in\mathbb{N}} x_k p_k$$

A priori, pour ça, il suffit que la série de terme général $(x_k p_k)_{k\in\mathbb{N}}$ soit convergente. On pourrait donc dire que X admet une espérance dans ce cas.

Mais en fait, si on définit l'espérance ainsi, *elle ne coïncide pas toujours avec la moyenne empirique* $(***)$. A la section 2.3 du chapitre 2 du merveilleux livre de W. Appel, *Probabilités pour les non-probabilistes*, on trouve deux exemples de variables aléatoires discrètes, la première décrite par

$$\forall k \in \mathbb{N}, x_k = (-1)^{k+1} k,\ \ p_k = \frac{k}{(k+1)!}$$

la seconde par

$$\forall k \in \mathbb{N}, \tilde{x}_k = (-1)^{k+1} k,\ \ \tilde{p}_k = \frac{1}{k(k+1)}$$

et on constate avec dépit que, bien que les séries de terme général $(x_k p_k)_k$ et $(\tilde{x}_k \tilde{p}_k)_k$ soient toutes les deux convergentes, dans le premier cas les réalisations empiriques convergent bien vers la moyenne théorique, mais pas dans le deuxième.

Le problème, c'est que quand on a numéroté les valeurs prises par X : $x_0, x_1, ...,$ *on leur a arbitrairement attribué un ordre particulier*, qui ne correspond à rien dans l'expérience. Or, pour certaines sommes infinies, changer l'ordre des termes change la valeur de la somme (et même la nature de la série : pour plus de détails, voir ici). Et lorsqu'on fait l'expérience empiriquement, la réalité se fiche de quel ordre nous a semblé naturel en modélisant !

Pour ne pas avoir ce problème, on définira l'espérance lorsque la série est absolument convergente : dans ce cas, l'ordre de sommation ne change rien, comme l'a montré M. Riemann.

Définition 8

Soit X une v.a. réelle discrète. On note $X(\Omega) = \{x_k, k \in \mathbb{N}\}$ (éventuellement, $X(\Omega)$ est fini).

- *On dit que X* admet une espérance (finie) *si la série $\sum_k |x_k| \mathbb{P}(\{X = x_k\})$ converge.*
- *Dans ce cas, son* espérance (mathématique) *est définie par*

$$\mathbb{E}[X] = \sum_{k\in\mathbb{N}} x_k \mathbb{P}(\{X = x_k\})$$

Cas général

Sautons dans le grand bain et définissons l'espérance d'une v.a. mesurable $X \in \mathcal{L}^0((\Omega, \mathscr{T}); \mathbb{R})$.

Lorsque Ω n'est pas supposé fini ou dénombrable, l'analogue de la somme de ($*$) ce qui nous permet de faire une "somme (éventuellement) continue" des valeurs $X(\Omega)$ "pondérées" par la mesure $\mathbb{P}(\{\omega\}$, c'est l'intégrale de Lebesgue par rapport à la mesure $\mathbb{P}$. Ce qui nous amène à

Définition 9

Comme toujours, $(\Omega, \mathscr{T}, \mathbb{P})$ est un espace probabilisé et $X \in\in \mathcal{L}^0((\Omega, \mathscr{T}); \mathbb{R})$ une variable aléatoire.

- *On dit que X est* intégrable *si*

$$\int_\Omega |X| d\mathbb{P} < \infty \iff \left(\int_\Omega X^+ d\mathbb{P} < \infty \text{ et } \int_\Omega X^- d\mathbb{P} < \infty \right) \tag{1}$$

- *On note $\mathcal{L}^1(\Omega, \mathscr{T}, \mathbb{P}; \mathbb{R})$ l'ensemble des variables aléatoires intégrables (ou plus simplement $\mathcal{L}^1$, quand on sait où on est).*
- *Si X est intégrable alors on définit* l'espérance *de X par*

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

Très bien, mais comment calculer cette intégrale, alors qu'on ne comprend même pas à quoi ressemble Ω ?

$\rightsquigarrow$ Ce qu'on connaît mieux, c'est la loi de X, P_X, qui est lé mesure image de $\mathbb{P}$ par X. Or, on dispose d'un théorème très pratique, le *théorème du transfert* pour les mesures images :

Théorème 10

Soit $(X, \mathscr{T}, \mu)$ un espace mesuré, et $(Y, \mathscr{A})$ un espace mesurable. Soit de plus $f : X \to Y$ une fonction mesurable de $(X, \mathscr{T})$ dans $(Y, \mathscr{A})$. On note μ_f la mesure image[8] de μ par f.

Soit $\varphi : Y \to \mathbb{R}$ une fonction mesurable. Alors φ est intégrable par rapport à μ_f ssi $\varphi \circ f$ est μ-intégrable, et dans ce cas

$$\int_X (\varphi \circ f)(x)\, d\mu(x) = \int_Y \varphi(y) d\mu_f(y)$$

(Ce théorème permet donc de réaliser une sorte de "changement de variable mesurable").

Traduisons ceci dans le cas qui nous intéresse : on applique le théorème du transfert avec

$$\begin{array}{ccc|ccc} (X, \mathscr{T}) & = & (\Omega, \mathscr{T}) & f & = & X \\ (Y, \mathscr{A}) & = & (\mathbb{R}, \mathscr{B}(\mathbb{R})) & \mu_f & = & P_X \\ \mu & = & \mathbb{P} & \varphi & = & id_{\mathbb{R}} \end{array}$$

ce qui nous donne

Proposition 3

Soit $(\Omega, \mathscr{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. Alors X est intégrable ssi

$$\int_\mathbb{R} |x|\, dP_X(x) < \infty$$

et dans ce cas

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_\mathbb{R} x\, dP_X(x)$$

Et plus généralement, d'ailleurs, si X est une v.a. et $\varphi \in \mathcal{L}^0(\mathbb{R})$ une fonction borélienne, alors la variable aléatoire $\varphi(X)$ est intégrable ssi φ est P_X-intégrable, et dans ce cas

$$\mathbb{E}[\varphi(X)] = \int_\mathbb{R} \varphi(x)\, dP_X(x)$$

$\rightsquigarrow$ Ce qui nous permet, pour $\varphi(x) = x^k$, d'étudier les *moments* de X, ou encore, avec $\varphi(x) = (x - \mathbb{E}[X])^2$, de définir la *variance* d'une v.a.

Revenons au cas discret un instant

Une v.a. discrète, avait-on constaté, ce n'est jamais qu'une v.a. dont la loi est une somme (finie ou dénombrable) de mesures de Dirac :

$$P_X = \sum_k p_k \delta_{x_k}$$

où $X(\Omega) = \{x_k, k \in \mathbb{N}\}$ et $p_k = \mathbb{P}\{x_k\}$.

Mais du coup, X est intégrable [9] ssi

$$\int_{\mathbb{R}} |x| dP_X(x) < \infty \iff \int_{\mathbb{R}} x\, d\left(\sum_k p_k \delta_{x_k}\right) < \infty \iff \sum_k p_k \int_{\mathbb{R}} |x| d\delta_{x_k} < \infty$$
$$\iff \sum_k p_k |x_k| < \infty$$

et dans ce cas

$$\mathbb{E}[X] = \int_{\mathbb{R}} x dP_X(x) = \sum_k p_k x_k$$

$\rightsquigarrow$ On retrouve la définition 8, ce qui est rassurant !

Variables aléatoires à densité

Dans le cas où X admet une densité $f_X : \Omega \to \mathbb{R}^+$, P_X est la mesure de densité f_X par rapport à la mesure de Lebesgue λ_1. Or, on a un autre théorème pour intégrer les fonctions par rapport à une mesure à densité :

Théorème 11

Soit $(Y, \mathscr{A}, \mu)$ un espace mesuré et $f \in \mathcal{L}^1((Y, \mathscr{A}, \mu), \mathbb{R}^+)$. On note ν la mesure de densité f par rapport à μ. Soit $\varphi \in \mathcal{L}^0(Y, \mathscr{A}; \mathbb{R})$. Alors φ est ν-intégrable sur X ssi φf est μ-intégrable sur X, et dans ce cas

$$\int_X \varphi\, d\nu = \int_X \varphi f\, d\mu$$

Utilisons ceci avec

$$(Y, \mathscr{A}, \mu) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda_1), \quad f = f_X, \nu = P_X, \quad \varphi = id_{\mathbb{R}}$$

On obtient

Théorème 12

Soit $(\Omega, \mathscr{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \to \mathbb{R}$ une variable aléatoire à densité f_X. Alors X est intégrable ssi $x f_X(x)$ est intégrable par rapport à la mesure de Lebesque sur $\mathbb{R}$, et dans ce cas

$$\mathbb{E}[X] = \int_{\mathbb{R}} x\, dP_X(x) = \int_{\mathbb{R}} x f_X(x)\, d\lambda_1(x)$$

4 Vecteurs aléatoires

On l'a dit, l'intérêt des variables aléatoires, c'est d'en avoir plusieurs sur le même Ω, pour représenter différents phénomènes.

Si on a deux variables aléatoires, on peut les réunir en un couple (X, Y) qui est alors une fonction $\Omega \to \mathbb{R}^2$. Et cette fonction est mesurable, puisque X et Y le sont : ça nous donne donc une variable aléatoire à valeurs dans $\mathbb{R}^2$. Et plus généralement,

9. Pour le calcul de l'intégrale par rapport à une somme de mesures de Dirac, on peut adapter le raisonnement expliqué ici : http://carolinevernier.website/pretext_exemples_integration/sec_int_pos.html

Définition 13

Soit $(\Omega, \mathscr{T}, \mathbb{P})$ *un espace probabilisé. On appelle* vecteur aléatoire *une fonction mesurable* $V : (\Omega, \mathscr{T}) \to (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.

On a alors $V = (X_1, \ldots, X_n)$ *où, pour chaque* $i = 1, ..., n$, X_i *est une variable aléatoire réelle sur* Ω.

On peut alors étendre nos définitions aux vecteurs aléatoires :

- La *loi* $P_V = P_{(X_1,\ldots,X_n)}$ de V est la mesure image de $\mathbb{P}$ par V sur $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$:

$$\forall B \in \mathscr{B}(\mathbb{R}^n),\ P_{(X_1,\ldots,X_n)}(B) = \mathbb{P}(\{\omega \in \Omega | (X_1(\omega), \ldots, X_n(\omega)) \in B\}) \tag{2}$$

- Pour chaque $i = 1, ...n$, la loi de X_i (autrement dit la mesure image de $\mathbb{P}$ par X_i sur $\mathbb{R}$) est appelé i-ième *loi marginale* de X_i.

- Si toutes les v.a. X_i sont intégrables, on définit le *vecteur espérance* de V par

$$\mathbb{E}[V] = (\mathbb{E}(X_1), \ldots, \mathbb{E}[X_n]) \in \mathbb{R}^n$$

- Par le théorème du transfert, si $\varphi : \mathbb{R}^n \to \mathbb{R}$ est une fonction mesurable, alors la v.a. réelle $\omega \in \Omega \mapsto \varphi(V(\omega)) \in \mathbb{R}$ est intégrable ssi la fonction φ est P_V-intégrable sur $\mathbb{R}^n$, et on a alors

$$\mathbb{E}[\varphi(V)] = \int_{\mathbb{R}^n} \varphi(x_1, \ldots, x_n)\, dP_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n)$$

$\leadsto$ On définit la *covariance* d'un couple (X, Y) de v.a. intégrables par l'espérance de la fonction $\varphi(x, y) = (x - \mathbb{E}[X])(y - \mathbb{E}[Y])$

$$Cov(X, Y) = \int_{\mathbb{R}^2} (x - \mathbb{E}[X])(y - \mathbb{E}[Y]) dP_{(X,Y)}(x, y)$$

(si elle existe, bien sûr : il faut que XY soit intégrable).

$\leadsto$ On définit la *matrice de covariance* de $V = (X_1, \ldots, X_n)$ comme la matrice $n \times n$ dont les coefficients sont $C_{ij} = Cov(X_i, X_j)$.

- On dit que V admet une densité s'il existe une fonction $f_V : \mathbb{R}^n \to \mathbb{R}^+$ telle que, pour tout $A \in \mathscr{B}(\mathbb{R}^n)$

$$P_V(A) = \int_A f_V(x_1, \ldots, x_n) d\lambda_n(x_1, \ldots x_n) = \int_A f_V(x_1, \ldots, x_n) d\lambda_1(x_1) \otimes \ldots \otimes d\lambda_1(x_n)$$

où λ_n est la mesure de Lebesgue sur $\mathbb{R}^n$, égale à la mesure produit $\lambda_1 \otimes ... \otimes \lambda_1$.

Mais alors, par le théorème de Fubini, la loi de X_1 est décrite par

$$\begin{aligned} P_{X_1}(A) &= \mathbb{P}(\{X_1 \in A\}) = \mathbb{P}(V \in A \times \mathbb{R} \ldots \times \mathbb{R}) \\ &= \int_{A \times \mathbb{R} \times \ldots \times \mathbb{R}} f_V(x_1, \ldots, x_n) d\lambda_1(x_1) \otimes \ldots \otimes d\lambda_1(x_n) \\ &= \int_A \left(\int_{\mathbb{R}} ... \int_{\mathbb{R}} f_V(x_1, \ldots, x_n) d\lambda_{n-1}(x_2, \ldots, x_n) \right) d\lambda_1(x_1) \end{aligned}$$

donc X_1 est une variable à densité, de densité

$$f_{X_1}(t) = \int_{\mathbb{R}^{n-1}} f_V(t, u) d\lambda_{n-1}(u)$$

(et ça se passe de la même façon pour $X_2, \ldots, X_n$).

Les f_{X_i} s'appellent alors *densités marginales*.

Indépendance

On a envie de dire que deux évènements $A, B \in \mathscr{T}$ sont indépendants si la réalisation ou non de A n'a aucun effet sur la probabilité que B se réalise : autrement dit "savoir que A" ne nous dit rien sur B. Ce qui s'écrirait (si $0 < \mathbb{P}(A) < 1$)

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

et c'est cette dernière égalité qu'on va utiliser comme définition (elle a le mérite de marcher même si $\mathbb{P}(A) = 0$).

Définition 14

- *On dit que deux évènements $A, B \in \mathscr{T}$ sont* indépendants *si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*
- *On dit que deux tribus $\mathscr{T}_1$ et $\mathscr{T}_2$* indépendantess *si leurs évènements respectifs sont tous indépendants :*

$$\forall A_1 \in \mathscr{T}_1, \forall A_2 \in \mathscr{T}_2, \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

- *Si X est une variable aléatoire sur $(\Omega, \mathscr{T}, \mathbb{P})$, on définit la* tribu engendrée par X *par*

$$\sigma(X) = \{X^{-1}(A), A \in \mathscr{B}(\mathbb{R})\}$$

$\rightsquigarrow$ *C'est une tribu sur Ω, incluse dans $\mathscr{T}$ (puisque $X \in \mathcal{L}^0$). Et de là, on dit que deux variables aléatoires X, Y sur Ω sont* indépendantes *si les tribus $\sigma(X)$ et $\sigma(Y)$ le sont.*

Mais du coup,

$$\begin{aligned} X, Y \text{ indépendantes } &\iff \forall B_1, B_2 \in \mathscr{B}(\mathbb{R}), \mathbb{P}(\{X \in B_1\} \cap \{Y \in \mathscr{B}_2\}) = \mathbb{P}(\{X \in B_1\})\mathbb{P}(\{Y \in B_2\}) \\ &\iff \forall B_1, B_2 \in \mathscr{B}(\mathbb{R}), \mathbb{P}(\{(X,Y) \in B_1 \times B_2\}) = \mathbb{P}(\{X \in B_1\})\mathbb{P}(\{Y \in B_2\}) \\ &\iff \forall B_1, B_2 \in \mathscr{B}(\mathbb{R}), P_{(X,Y)}(B_1 \times B_2) = P_X(B_1)P_Y(B_2) \\ &\iff P_{(X,Y)} = P_X \otimes P_Y \end{aligned}$$

$\rightsquigarrow$ Deux variables aléatoires X et Y sont indépendantes si, et seulement si, la loi du couple (X, Y) est le produit (au sens de mesure produit) des lois de X et de Y.

Sources (d'information et de fun)

- W. Appel,*Probabilités pour les non-probabilistes*, Editions H&K, 2015
- A. Kolmogorov, *Foundations of the Theory of Probability*, Chelsea Publishing Company, 1933
- Vidéo par PBS Infinite Series, How the Axiom of Choice Gives Sizeless Sets | Infinite Series, en anglais
- Vidéos par El Jj, Deux (deux !) minutes pour le théorème de Banach-Tarski (version courte) et Deux (deux ?) minutes pour... Le théorème de Banach-Tarski (version longue)
- Vidéo de VSauce sur le même sujet : The Banach–Tarski Paradox
- Quelques rappels de théorie de la mesure : `http://carolinevernier.website/pretest/compl_int.html` et quelques exemples de tribus, mesures, etc : `http://carolinevernier.website/pretext_exemples_integration/ex_int.html`
- Le point de vue de Poincaré : Certitudes, bon sens et probabilités, lu par Øljen