



# Probabilités

« L'estomac, contrairement au cerveau, prévient quand il est vide. » – proverbe bouddhiste

## Table des matières

<b>1</b>	<b>Préliminaire technique : dénombrabilité et familles sommables</b>	<b>2</b>
1.1	Dénombrabilité . . . . .	2
1.2	Familles sommables . . . . .	2
<b>2</b>	<b>Espaces probabilisés</b>	<b>4</b>
2.1	Dans les univers finis (rappels de première année) . . . . .	4
2.2	Univers infinis et tribus . . . . .	5
2.3	Probabilités sur $(\Omega, \mathcal{T})$ . . . . .	5
2.4	Continuité croissante (et décroissante) . . . . .	6
<b>3</b>	<b>Conditionnement</b>	<b>7</b>
3.1	Probabilités conditionnelles . . . . .	7
3.2	Manipulation des probabilités conditionnelles . . . . .	9
3.2.1	Probabilités totales . . . . .	9
3.2.2	Formule de Bayes . . . . .	9
3.2.3	Probabilités composées . . . . .	10
3.3	Indépendance . . . . .	11
<b>4</b>	<b>Concept de variable aléatoire</b>	<b>12</b>
4.1	Seul le résultat compte! . . . . .	12
4.2	Loi, fonction de répartition . . . . .	13
4.3	Lois usuelles . . . . .	15
4.4	Couples de variables aléatoires . . . . .	16
4.5	Variables indépendantes . . . . .	18
<b>5</b>	<b>Espérance et variance</b>	<b>20</b>
5.1	Espérance, théorème de transfert . . . . .	20
5.2	Contrôle de l'écart à la moyenne ; variance . . . . .	23
5.3	Bienaymé-Tchebychev . . . . .	24
5.4	Covariance . . . . .	25
5.5	Fonctions génératrices . . . . .	26
5.6	Retour sur les lois usuelles . . . . .	27

```
int getRandomNumber()
{
    return 4; // chosen by fair dice roll.
             // guaranteed to be random.
}
```

XKCD 221 – RFC 1149.5 specifies 4 as the standard IEEE-vetted random number.

## ⚠ Attention

Faites confiance au hasard, mais ne croyez pas trop à la chance !

(ces petits encadrés me plaisent bien ; je vais en faire un usage absurde au moins pour ce poly)

# 1 Préliminaire technique : dénombrabilité et familles sommables

## 1.1 Dénombrabilité

DÉFINITION 1 — *Dénombrabilité*

- Un ensemble est déclaré
- *fini* lorsqu'il peut être mis en bijection avec un ensemble de la forme  $\llbracket 1, n \rrbracket$  ;
  - *dénombrable* lorsqu'il peut être mis en bijection avec  $\mathbb{N}$ .

On dit parfois « au plus dénombrable » pour « fini ou dénombrable »

- $\mathbb{N}$ ,  $\mathbb{Z}$  et  $\mathbb{Q}$  sont dénombrables.
- $\mathcal{P}(\mathbb{N})$  et  $\mathbb{R}$  ne sont pas dénombrables.
- $\{0, 1\}^{\mathbb{N}}$  (i.e. : l'ensemble des suites à valeurs dans  $\{0, 1\}$ , qui est en bijection avec  $\mathcal{P}(\mathbb{N})$ ) n'est pas dénombrable. On pense ici à une suite infinie de tirages à pile ou face.
- Si  $A \subset B$  et  $B$  est dénombrable, alors  $A$  est fini ou dénombrable.
- Si  $E$  et  $F$  sont dénombrables, alors  $E \times F$  l'est aussi.
- Si  $E_1, \dots, E_n$  sont dénombrables, alors  $\bigcup_{k=1}^n E_k$  l'est aussi (c'est également vrai pour une intersection, mais c'est moins spectaculaire!).
- Plus fort : si pour tout  $n \in \mathbb{N}$ ,  $E_n$  est dénombrable, alors  $\bigcup_{n \in \mathbb{N}} E_n$  l'est aussi (« une réunion dénombrable d'ensembles dénombrables est dénombrable »).

On verra souvent de telles réunions ou intersections, en particulier « croissantes ou décroissantes ».

- Dans le cas général, que signifie « appartenir à  $\bigcup_{k \in \mathbb{N}} E_k$  » ? et « appartenir à  $\bigcap_{k \in \mathbb{N}} E_k$  » ?
- Que vaut  $\bigcup_{n \in \mathbb{N}^*} [1/n, 3 - 1/n]$  ? et  $\bigcap_{n \in \mathbb{N}^*} [1/n, 3 - 1/n]$  ?
- Que vaut  $\bigcup_{n \in \mathbb{N}^*} ] - 1/n, 1 + 1/n[$  ? et  $\bigcap_{n \in \mathbb{N}^*} ] - 1/n, 1 + 1/n[$  ?
- Si pour tout  $n \in \mathbb{N}$ ,  $E_{n+1} \subset E_n$ , que désignent  $\bigcup_{n \in \mathbb{N}} E_n$  et  $\bigcap_{n \in \mathbb{N}} E_n$  ?
- Si pour tout  $n \in \mathbb{N}$ ,  $E_n \subset E_{n+1}$ , que désignent  $\bigcap_{k \in \mathbb{N}} E_k$  et  $\bigcup_{k \in \mathbb{N}} E_k$  ?

## 1.2 Familles sommables

Il s'agit ici, sans trop regarder les choses au microscope, de s'autoriser à sommer les familles dénombrables de réels positifs, quoi qu'il arrive : au pire la somme est déclarée égale à  $+\infty$ . Cela permettra toujours de sommer dans tous les sens, par paquets, en travaillant dans  $\mathbb{R}^+ \cup \{+\infty\} = [0, +\infty]$ ... Dans le cas de familles de réels pas particulièrement positifs, voire de complexes, il s'agira (comme pour la convergence absolue des séries) de considérer les valeurs absolues/modules pour déclarer la famille sommable ou non.

DÉFINITION 2 — *Familles sommables de réels positifs*

- Si  $I$  est dénombrable et  $(x_i)_{i \in I}$  est une famille de réels **positifs**, elle est déclarée *sommable* lorsque les sommes finies  $\sum_{j \in J} x_j$  (avec  $J$  fini inclus dans  $I$ ) sont majorées par une constante.
- Lorsque c'est le cas, on note  $\sum_{i \in I} x_i$  la borne supérieure (finie, donc) de ces sommes. Si ce n'est pas le cas, on note  $\sum_{i \in I} x_i = +\infty$ .

- Pour une série de réels positifs, il y a équivalence entre convergence et sommabilité, et le cas échéant la somme de la série est bien entendu égale à la somme de la famille.
- Si les  $x_n$  sont tous positifs, on pourra donc toujours parler de  $\sum_{n \in \mathbb{N}} x_n$ , qui vaut  $+\infty$  en cas de non-sommabilité. Ça change des séries pour lesquelles on ne s'autorise à parler de  $\sum_{n=0}^{+\infty} x_n$  que si la série est convergente.
- On admet le caractère licite de toute opération « par paquets » : si  $I$  est la réunion disjointe finie ou dénombrable d'ensemble  $I_n$ , disons pour  $n \in K$ , alors  $(x_i)_{i \in I}$  est sommable si et seulement si chaque  $(x_i)_{i \in I_n}$  est sommable, ainsi que  $\left( \sum_{i \in I_n} x_i \right)_{n \in K}$ . Et si c'est le cas, alors  $\sum_{i \in I} x_i = \sum_{n \in K} \left( \sum_{i \in I_n} x_i \right)$ .
- Une conséquence de ce dernier point est (si si!) que dans le cas d'une famille indexées par  $\mathbb{N}$  (une série...), si elle est sommable (convergence absolue) alors la somme peut être calculée en permutant les termes :  $\sum_{n \in \mathbb{N}} u_n = \sum_{n \in \mathbb{N}} u_{\varphi(n)}$ .

### 💡 Point Clé

Si les  $x_i$  sont positifs on peut toujours parler de  $\sum_{i \in I} x_i$ , quitte à ce que cette valeur soit  $+\infty$ . Et on peut calculer cette « somme » en regroupant les termes comme on veut.

DÉFINITION 3 — Familles sommables de complexes

|| Si  $I$  est dénombrable et  $(y_i)_{i \in I}$  est une famille de complexes, elle est déclarée *sommable* lorsque  $(|y_i|)_{i \in I}$  l'est.

- Il s'agit bien entendu de l'analogue « convergence absolue » pour les séries.
- On ne parlera de  $\sum_{i \in I} y_i$  que si la famille est sommable.
- Cette somme se calculera avec tout outil raisonnable : séparer partie réelle et partie imaginaire si on a des complexes. Les positifs et les négatifs si ce sont des réels. En sommant par paquets...

Quand  $I = \mathbb{N}$ , les familles sommables nous permettent de sommer « par paquets » en partitionnant  $\mathbb{N}$ . Mais le cas  $I = \mathbb{N}^2$  est particulièrement intéressant pour sommer au choix par lignes ou par colonnes, en ayant bien noté que

$$\mathbb{N}^2 = \bigsqcup_{n \in \mathbb{N}} \{(n, p) \mid p \in \mathbb{N}\} = \bigsqcup_{p \in \mathbb{N}} \{(n, p) \mid n \in \mathbb{N}\}$$

- Si pour tout  $n \in \mathbb{N}$  la série  $\sum_p u_{n,p}$  est convergente et qu'en plus la série des sommes  $\sum_n \left( \sum_{p=0}^{+\infty} u_{n,p} \right)$  est convergente, alors pour tout  $p \in \mathbb{N}$  la série  $\sum_n u_{n,p}$  est convergente, avec de plus la série des sommes  $\sum_p \left( \sum_{n=0}^{+\infty} u_{n,p} \right)$  convergente, et comme on peut s'y attendre :

$$\sum_{n=0}^{+\infty} \left( \sum_{p=0}^{+\infty} u_{n,p} \right) = \sum_{p=0}^{+\infty} \left( \sum_{n=0}^{+\infty} u_{n,p} \right)$$

- On verra « en particulier »<sup>1</sup> qu'on peut justifier ainsi (si l'un des deux membres converge) :

$$\sum_{n=0}^{+\infty} \left( \sum_{p=0}^n u_{n,p} \right) = \sum_{p=0}^{+\infty} \left( \sum_{n=p}^{+\infty} u_{n,p} \right)$$

1. Ce point sera précisé...

EXEMPLES :

1. Montrer que si  $|q| < 1$ , alors  $(q^{|n|})_{n \in \mathbb{Z}}$  est sommable, et calculer la somme.
2. Calculer  $\sum_{(p,q) \in \mathbb{N} \times \mathbb{N}^*} \frac{1}{(p+q^2)(p+q^2+1)}$ .
3. Montrer que pour  $n \geq 2$ ,  $\left(\frac{1}{(np)^k}\right)_{p,k \in \mathbb{N}^*}$  est sommable.

## 2 Espaces probabilisés

### 2.1 Dans les univers finis (rappels de première année)

DÉFINITION 4 — *Univers, événements*

- L'*univers* désigne en probabilité un ensemble (provisoirement FINI).
- Un *événement* est (provisoirement) une partie de l'univers.
- Deux événements sont dits *incompatibles* lorsque leur intersection est vide.

EXEMPLES :

1. Pile ou face :  $\Omega = \{P, F\}$  ; pas beaucoup d'événements !
2. Le fait de gagner ou non le gros lot du loto :  $\Omega = \{G, P\}$ .
3. Un dé :  $\Omega = \{1, 2, \dots, 6\}$  ;  $2^6$  événements, dont

$$\{R \geq 3\}, \{R \leq 2\}, \{R \text{ est pair}\}, \{R = 1 \text{ ou } R = 3\}.$$

On préférera rapidement voir un lancer de dé comme une variable aléatoire.

4. Un couple de dés :  $\Omega = \{(1, 1); (1, 2); \dots; (6, 6)\} = \llbracket 1, 6 \rrbracket^2$ . Événements :  $2^{36}$ , dont :  $\{x = y\}$ ,  $\{x \text{ est pair}\}$ ,  $\{x + y \leq 4\}$ ,  $\{(x, y) = (2, 3)\}$ ...

Informellement, on veut quantifier les événements par un réel entre 0 (impossible,... ou presque sûrement impossible!) et 1 (certain, ou presque!). On va seulement demander deux propriétés raisonnables :

DÉFINITION 5 — *Probabilité – univers fini*

- Si  $\Omega$  est un ensemble fini, une *probabilité* sur  $\Omega$  est une application de  $\mathcal{P}(\Omega)$  dans  $[0, 1]$  telle que :
  - $\mathbb{P}(\Omega) = 1$  ;
  - si  $A, B \in \mathcal{P}(\Omega)$  sont d'intersection vide, alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .

EXEMPLE : On peut définir, pour  $X \subset \Omega$ ,  $\mathbb{P}(X) = \frac{|X|}{|\Omega|}$ , et on obtient bien ainsi une probabilité (probabilité de dénombrement). Bien entendu, si  $\Omega$  est constitué de deux éléments modélisant le fait de gagner ou non au loto, cette probabilité modélise assez mal la réalité. Par contre, si  $\Omega$  est constitué de l'ensemble des résultats possibles pour un tirage « honnête » (d'un dé, d'une pièce...) alors c'est une plutôt bonne idée.

REMARQUE : La donnée de  $\mathbb{P}$  impose la probabilité des événements élémentaires  $\mathbb{P}(\{x\})$  pour  $x \in \Omega$ ... et puisque  $\Omega$  est fini, la réciproque est vraie<sup>2</sup>.

PROPRIÉTÉS : Soit  $\mathbb{P}$  est une probabilité sur  $\Omega$ .

- *Passage au complémentaire* : si  $A \subset \Omega$ , alors  $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$ .
- *Réunion (cas général)* : si  $A, B \subset \Omega$ , alors  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .
- *Croissance* : si  $A \subset B \subset \Omega$ , alors  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

---

2. Ce sera faux lorsque  $\Omega$  est infini non dénombrable.

## 2.2 Univers infinis et tribus

On va maintenant modéliser des phénomènes réclamant des univers infinis : typiquement  $\{0, 1\}^{\mathbb{N}}$  pour une suite infinie de tirages à pile ou face, ou bien une suite de positions pour une grenouille sautant de façon aléatoire entre plusieurs points aux temps 1, 2, etc.

On souhaite conserver la propriété sur les réunions disjointes : si  $E_1, \dots, E_n$  sont d'intersections (deux à deux) vides, alors

$$\mathbb{P}\left(\bigcup_{k=1}^n E_k\right) = \sum_{k=1}^n \mathbb{P}(E_k)$$

et même mieux (union dénombrable).

On souhaite aussi pouvoir « mesurer » (évaluer leur probabilité) une classe raisonnable d'événements : cette classe doit en général contenir les événements élémentaires (les singletons), et doit être stable par union dénombrable et par passage au complémentaire.

Quand  $\Omega$  est infini non dénombrable, définir une probabilité non triviale sur  $\mathcal{P}(\Omega)$  sera compliqué (et n'a pas forcément grand sens), donc il faut accepter de ne mesurer que certaines parties de  $\Omega$  : ce sont seulement celles-ci qu'on appellera « événements ».

DÉFINITION 6 — *Tribus, ou  $\sigma$ -algèbres*

- Soit  $\Omega$  un ensemble. Une *tribu*, ou  $\sigma$ -*algèbre* sur  $\Omega$  est un ensemble  $\mathcal{T}$  de parties de  $\Omega$  tel que :
- $\emptyset \in \mathcal{T}$  ;
  - $\mathcal{T}$  est stable par réunion (finie ou) dénombrable ;
  - $\mathcal{T}$  est stable par passage au complémentaire.

Bien entendu, une tribu est stable par *intersection* fini ou dénombrable, et contient  $\Omega$ .

EXEMPLES :

- $\mathcal{T} = \{\emptyset, \Omega\}$  est la plus petite tribu imaginable... et a peu d'intérêt !
- $\mathcal{T} = \mathcal{P}(\Omega)$  est une grosse tribu... mais a parfois peu d'intérêt, car on aura du mal à définir une probabilité dessus – sauf dans le cas fini (ou éventuellement dénombrable) pour lequel on prend en général cette « tribu complète ».
- Si  $A \subset \Omega$ ,  $\mathcal{T} = \{\emptyset, A, \bar{A}, \Omega\}$  est la « tribu engendrée par  $A$  ».

DÉFINITION 7 — *Système (quasi-)complet*

- Si pour tout  $i \in I$ ,  $A_i \in \mathcal{T}$ , pour tous  $i, j \in I$  distincts  $A_i \cap A_j = \emptyset$  et  $\bigcup_{i \in I} A_i = \Omega$ , on dit que la famille  $(A_i)_{i \in I}$  constitue un *système complet* d'événements.
- Si on n'a pas  $\bigcup_{i \in I} A_i = \Omega$  mais seulement  $\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = 1$ , on parlera de système *quasi-complet*.

REMARQUE : La plupart du temps, IRL,  $\Omega$  n'est pas explicité, et encore moins  $\mathcal{T}$  ! En pratique, la modélisation d'un problème nous fournit une probabilité qui est définie sur les parties dont on a besoin... et il est même difficile de construire une partie qui n'est pas dans  $\mathcal{T}$  !

## 2.3 Probabilités sur $(\Omega, \mathcal{T})$

DÉFINITION 8 — *Espace probabilisé*

- Un *espace probabilisé* est un triplet  $(\Omega, \mathcal{T}, \mathbb{P})$ , avec  $\mathcal{T}$  une tribu de  $\Omega$ , et  $\mathbb{P}$  une application de  $\mathcal{T}$  dans  $[0, 1]$  vérifiant :
- $\mathbb{P}(\Omega) = 1$  ;
  - si  $(A_n)_{n \in \mathbb{N}}$  est une famille d'événements **incompatibles deux à deux**, alors  $\sum \mathbb{P}(A_n)$  converge, et
- $$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(A_n)$$
- (propriété de  $\sigma$ -additivité).

On notera que sans l'hypothèse d'incompatibilité (deux à deux), on conserve l'inégalité générale :

$$\mathbb{P}\left(\bigcup_{n=1}^N A_n\right) \leq \sum_{n=1}^N \mathbb{P}(A_n)$$

dans le cas fini, et *en cas de convergence du membre de droite*, pour une réunion infinie :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

On parle de *sous-additivité*. Avec le vocabulaire des familles sommables, on peut se passer des précautions de type « en cas de convergence » : il suffit de donner cette inégalité dans  $[0, +\infty]$ .

Les propriétés suivantes sont d'autres conséquences de la définition d'une probabilité.

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

- $\mathbb{P}(\emptyset) = 0$ .
- Si  $A \in \mathcal{T}$  alors  $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$ .
- Si  $A, B \in \mathcal{T}$  et  $A \subset B$ , alors  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

REMARQUE : si  $\Omega$  est dénombrable, alors on peut écrire  $\Omega = \{x_0, x_1, \dots, x_n, \dots\}$ . Si de plus  $\mathcal{T} = \mathcal{P}(\Omega)$ , on a alors par  $\sigma$ -additivité  $\sum \mathbb{P}(\{x_n\})$  convergente et de somme égale à 1. Et même, si  $A \subset \Omega$  est dans  $\mathcal{T}$ , alors  $\sum_{x \in A} \mathbb{P}(\{x\})$  convergente et de somme égale à  $\mathbb{P}(A)$  : on est assez proche de la situation où  $\Omega$  est fini.

Réciproquement, si  $\Omega = \{x_n \mid n \in \mathbb{N}\}$  et  $\sum p_n$  est une série de réels positifs convergente et de somme égale à 1, alors on peut définir une probabilité sur  $\Omega$  définie sur la tribu complète  $\mathcal{T} = \mathcal{P}(\Omega)$  en posant, pour  $A \subset \Omega$  :  $\mathbb{P}(A) = \sum_{x_n \in A} p_n$ . Je vous sens en panique : « Mais mais mais, ça dépend de l'ordre de sommation ! ». En fait non (cf les familles sommables)...

DÉFINITION 9 — *Événements presque sûrs, négligeables*

- || Un événement  $A \in \mathcal{T}$  est dit :
- *presque sûr* lorsque sa probabilité vaut 1 ;
  - *négligeable* lorsque sa probabilité est nulle.

Deux exemples typiques : dans une série de lancers indépendants d'une pièce non biaisée, l'événement « il y aura un moins un PILE » est presque sûr (mais pas égal à  $\Omega$ ) et « il n'y aura que des piles » est négligeable (mais pas vide).

## 2.4 Continuité croissante (et décroissante)

**Exercice 1.** Dans une série de lancers indépendants d'une pièce équilibrée, quelle est la probabilité pour qu'il y ait au moins un pile ?

Si  $(A_n)_{n \in \mathbb{N}}$  est une suite *croissante* d'événements (pour tout  $n \in \mathbb{N}$ ,  $A_n \subset A_{n+1}$ ), alors  $E = \bigcup_{k \in \mathbb{N}} A_k$  est également dans  $\mathcal{T}$ , et on a envie de dire (faire un dessin !) que « les  $A_n$  s'approchent de  $E$  ». Même chose pour  $F = \bigcap_{k \in \mathbb{N}} B_k$  lorsque  $(B_n)_{n \in \mathbb{N}}$  est une suite *décroissante* d'événements.

EXEMPLES :

- $A_n$  : « PILE est tombé sur l'un des  $n$  premiers tirages ».
- $A_n$  : « Sur les  $n$  premières colles, j'ai eu au moins un 13 ».
- $A_n$  : « Après  $n$  sauts, la grenouille est passée au moins une fois par tel point ».
- $B_n$  : « Sur les  $n$  premiers tirages, la moyenne du nombre de PILE a toujours été supérieure ou égale à  $1/3$  ».
- $B_n$  : « Sur les  $n$  premières notes, j'ai toujours eu la moyenne ».
- $B_n$  : « Après  $n$  sauts, la grenouille n'est jamais passée par tel point ».

Que dire de  $\mathbb{P}(A_n)$  vis-à-vis de  $\mathbb{P}(E)$  ; et de  $\mathbb{P}(B_n)$  vis-à-vis de  $\mathbb{P}(F)$  ? Je suis certain qu'en réfléchissant sur le dessin vous avez la bonne intuition ; peut-être même que le résultat vous semble *évident* ! Bon, il est intuitif, mais il y a quand même quelque chose à montrer !

### THÉORÈME 1 — Continuité (dé)croissante

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

— Si  $(A_n)_{n \in \mathbb{N}}$  est une suite croissante d'événements, alors :

$$\mathbb{P}(A_n) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right)$$

— Si  $(B_n)_{n \in \mathbb{N}}$  est une suite décroissante d'événements, alors :

$$\mathbb{P}(B_n) \xrightarrow{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k \in \mathbb{N}} B_k\right)$$

Ou encore, l'existence des limites faisant partie de la conclusion :

$$\mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) \quad \text{et} \quad \mathbb{P}\left(\bigcap_{k \in \mathbb{N}} B_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n)$$

PREUVE : Le deuxième s'obtient par complémentation du premier... lequel s'obtient en définissant par exemple  $C_0 = A_0$  et, pour tout  $n \in \mathbb{N}^*$  :  $C_n = A_n \setminus A_{n-1}$ , de sorte que  $E$  est la réunion **disjointe** des  $C_n$ . On peut alors écrire :

$$P(E) = \sum_{n=0}^{\infty} \mathbb{P}(C_n) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \mathbb{P}(C_k).$$

Mais  $\sum_{k=0}^n \mathbb{P}(C_k) = \mathbb{P}\left(\bigcup_{k=0}^n C_k\right) = \mathbb{P}(A_n) \dots$

En colle, on vous demandera évidemment de justifier cette dernière égalité ensembliste. ■

EXEMPLES :

- Quelle est la probabilité pour que la séquence « PFFP » apparaisse lors d'une série infinie de lancers d'une pièce équilibrée ?
- Une grenouille saute de façon aléatoire (au sens : uniforme) entre trois points. Quelle est la probabilité pour qu'elle ne passe jamais par l'un des trois points ?

REMARQUE : On avait envie de dire que « les  $A_n$  s'approchent de  $E$  », ou encore : «  $\lim A_n = E$  » ; on peut alors énoncer modulo cette notation :

$$\mathbb{P}(\lim A_n) = \lim \mathbb{P}(A_n),$$

et on comprend mieux le nom de ce résultat ! *N'est-ce pas ?*

Il arrive qu'on s'intéresse à une réunion *non croissante* ou une intersection *non décroissante* ; il faut alors adapter les énoncés précédents. Les énoncés qui suivent sont clairs si on a compris les précédents. Dans le cas contraire, ils resteront mystérieux... et confondus avec les précédents, malgré les efforts auxquels vous consentirez pour les apprendre !

### THÉORÈME 2 — Continuité (dé)croissante, bis

Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace probabilisé. Si  $(C_n)_{n \in \mathbb{N}}$  est une suite d'événements, alors :

$$\mathbb{P}\left(\bigcup_{k \in \mathbb{N}} C_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcup_{k=0}^n C_k\right) \quad \text{et} \quad \mathbb{P}\left(\bigcap_{k \in \mathbb{N}} C_k\right) = \lim_{n \rightarrow +\infty} \mathbb{P}\left(\bigcap_{k=0}^n C_k\right)$$

## 3 Conditionnement

### 3.1 Probabilités conditionnelles

Voici trois exemples typiques (le premier était dans ce cours avant le COVID!) :

1. Une maladie touche 0,01% de la population. On dispose d'un test :
  - pour les personnes atteintes, il est positif avec probabilité 99% (il y a 1% de faux-négatifs) ;
  - pour les personnes saines, il est négatif avec probabilité 99% (il y a 1% de faux-positifs) ;

- (a) Informellement, qu'est-ce qui est plus grave : les faux-positifs ou les faux-négatifs ?
  - (b) On suppose qu'une personne est testée positive. Quelle est la probabilité qu'elle soit effectivement malade ?
2. (CCINP) On dispose de 100 dés, dont 25 sont pipés. Pour chaque dé pipé, la probabilité d'obtenir 6 est de  $1/2$  à chaque tirage.  
On tire un dé au hasard, on le lance, et on obtient un 6. Quelle est la probabilité pour qu'il soit pipé ?
  3. Dans un lycée, il y a trois fois plus de taupins que de khâgneux. Chez les taupins (resp. khâgneux) on trouve 5% (resp. 25%) de gauchers.  
On choisit un élève au hasard parmi les taupins et les khâgneux, et il s'avère être gaucher. Quelle est la probabilité pour qu'il soit khâgneux ?

Dans les trois situations, les problèmes sont de même nature :

1. On a une information de la forme « probabilité pour que le test soit positif sachant que le patient est malade (resp. sain) » et on s'intéresse à la probabilité pour qu'il soit malade sachant que le test est positif.
2. On a une information de la forme « probabilité pour que le résultat soit 6 sachant que le dé est pipé (resp. équilibré) » et on s'intéresse à la probabilité pour que le dé soit pipé sachant que le résultat est 6.
3. On a une information de la forme « probabilité pour que l'élève soit gaucher sachant que c'est un khâgneux (resp. taupin) » et on s'intéresse à la probabilité pour qu'il soit khâgneux sachant qu'il est gaucher.

Plus formellement dans chaque cas on connaît  $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$  et on s'intéresse à  $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ . C'est a priori très différent, n'est-ce pas ?

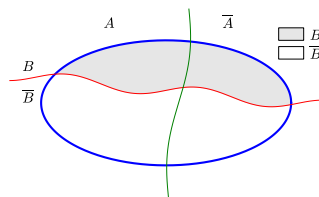


FIGURE 1 – Probabilités conditionnelles

1.  $A$  : « le patient est malade » et  $B$  : « le test est positif ».
2.  $A$  : « le dé est pipé » et  $B$  : « on tire 6 ».
3.  $A$  : « l'élève est un khâgneux » et  $B$  : « l'élève est gaucher ».

DÉFINITION 10 — *Probabilité conditionnelle*

Si  $A, B \in \mathcal{T}$  et  $\mathbb{P}(A) > 0$ , on définit la *probabilité de  $B$  sachant  $A$* , et on note  $\mathbb{P}_A(B)$ , ou  $\mathbb{P}(B|A)$  :

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

**Exercice 2.** Sur les deux dessins suivants, évaluer à la louche  $\mathbb{P}(A)$ ,  $\mathbb{P}(A|B)$  et  $\mathbb{P}(\bar{A}|B)$  (les probabilités étant données par les aires, normalisées à 1 pour  $\Omega$ ) :

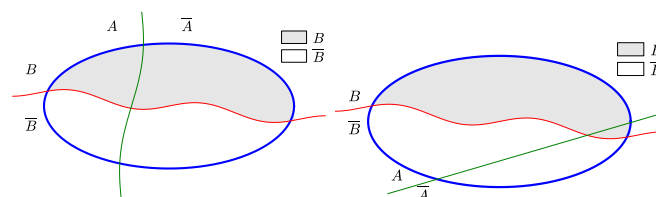


FIGURE 2 – IRL on vous donne les probabilités et vous devez faire les dessins !



PROPOSITION 1 — Les probabilités conditionnelles... sont des probabilités !

| Si  $\mathbb{P}(B) > 0$ , alors  $\mathbb{P}_B$  constitue une probabilité sur  $(\Omega, \mathcal{T})$ .

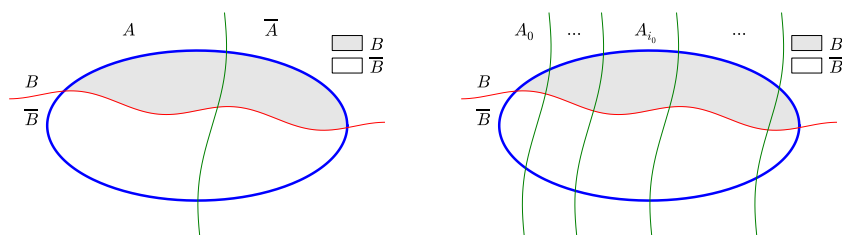
PREUVE : Left to the reader. ■

## 3.2 Manipulation des probabilités conditionnelles

### 3.2.1 Probabilités totales

**Exercice 3.** *Quelle est la proportion de gauchers dans l'ensemble des élèves de taupe et de khâgne ?*

Le problème est ici de déterminer  $\mathbb{P}(B)$ , quand on connaît  $\mathbb{P}(A)$ ,  $\mathbb{P}(B|A)$  et  $\mathbb{P}(B|\bar{A})$ , ou plus généralement les  $\mathbb{P}(A_i)$  et  $\mathbb{P}(B|A_i)$ , avec  $(A_i)_{i \in I}$  un système complet d'événements.



THÉORÈME 3 — *Formule des probabilités totales*

Si  $(A_i)_{i \in I}$  est un système (quasi-)complet d'événements (c'est-à-dire pour rappel : les intersections deux à deux sont vides, et la réunion vaut  $\Omega$ , ou simplement est de probabilité 1) et que  $B$  est un autre événement, alors  $\sum_{i \in I} \mathbb{P}_{A_i}(B)\mathbb{P}(A_i)$  est convergente, et :

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}_{A_i}(B)\mathbb{P}(A_i)$$

PREUVE : Écrire :

$$B = \bigcup_{i \in I} (B \cap A_i)$$

et noter que cette réunion dénombrable est disjointe.

Dans le cas où la réunion des  $A_i$  ne vaut pas  $\Omega$ , considérer  $C$  le complémentaire de cette réunion, et écrire

$$B = (B \cap C) \cup \bigcup_{i \in I} (B \cap A_i)$$

(réunion qui reste disjointe) sachant que  $\mathbb{P}(B \cap C) \leq \mathbb{P}(C) = 0$ . ■

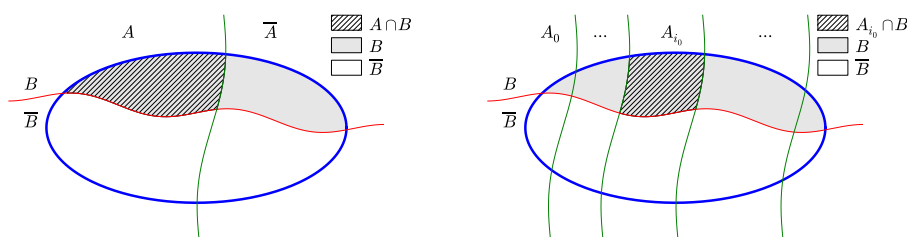
REMARQUE : Dans le cas fréquent de la famille totale  $\{A, \bar{A}\}$ , la formule devient :

$$\mathbb{P}(B) = \mathbb{P}_A(B)\mathbb{P}(A) + \mathbb{P}_{\bar{A}}(B)\mathbb{P}(\bar{A}).$$

C'est ce que vous obtenez informellement avec l'arbre de terminale. Cet arbre n'est pas interdit s'il vous aide, mais vous **devez** être capable de formaliser la preuve en termes de probabilités totales.

### 3.2.2 Formule de Bayes

Si maintenant on cherche  $\mathbb{P}_B(A) = \mathbb{P}(A|B)$ , la formule de Bayes est là pour ça :



THÉORÈME 4 — *Formule de Bayes*

Si  $A$  et  $B$  sont des événements de probabilités strictement positives, alors :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

PREUVE : Tout simplement :

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

■

Reprenons les trois exemples vus plus haut (test et malades ; dés pipés ou non ; khâgneux et taupins). Dans ce type de problème, on évalue le plus souvent la probabilité au dénominateur via la formule des probabilités totales, soit ici (trois fois) :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\mathbb{P}(A)}{\mathbb{P}_A(B)\mathbb{P}(A) + \mathbb{P}_{\bar{A}}(B)\mathbb{P}(\bar{A})}.$$

Potentiellement (voir l'exercice « veaux, vaches, cochons, couvée » de la feuille de TD) on pourra écrire :

$$\mathbb{P}_B(A_{i_0}) = \frac{\mathbb{P}_{A_{i_0}}(B)\mathbb{P}(A_{i_0})}{\sum_{i \in I} \mathbb{P}_{A_i}(B)\mathbb{P}(A_i)}$$

avec  $(A_i)_{i \in I}$  une famille complète d'événements.

1. La probabilité pour que quelqu'un testé positif soit malade vaut :

$$\mathbb{P}_B(A) = \frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 0,01 \times (1 - 10^{-4})} \simeq \frac{10^{-4}}{10^{-2}} = 10^{-2}$$

(surpris?)

2. La probabilité pour que le dé ayant donné un six soit pipé vaut :

$$\mathbb{P}_B(A) = \frac{\frac{1}{2} \frac{1}{4}}{\frac{1}{2} \frac{1}{4} + \frac{1}{6} \frac{3}{4}} = \frac{1}{2}.$$

3. La probabilité pour qu'un gaucher soit khâgneux vaut :

$$\mathbb{P}_B(A) = \frac{\frac{1}{4} \frac{1}{4}}{\frac{1}{4} \frac{1}{4} + \frac{1}{20} \frac{3}{4}} = \frac{5}{8}.$$

### 3.2.3 Probabilités composées

Il est parfois naturel d'évaluer « dans l'ordre chronologique » certains événements (exemples : « les  $n$  premières boules tirées sans remise sont rouges » ; « les  $k$  premiers entiers choisis au hasard sont distincts », ...) pour lesquelles la modélisation nous donne des informations de la forme  $\mathbb{P}(A_{n+1}|A_0 \cap A_1 \cap \dots \cap A_n)$  ou encore  $\mathbb{P}(B_{n+1}|B_n)$  (avec  $(B_n)$  décroissante) ; on dispose alors du résultat suivant :

THÉORÈME 5 — *Formule des probabilités composées*

Si  $A_0, \dots, A_n$  sont des événements tels que  $\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_{n-1}) \neq 0$  alors :

$$\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_n) = \mathbb{P}(A_0)\mathbb{P}_{A_0}(A_1)\mathbb{P}_{A_0 \cap A_1}(A_2) \cdots \mathbb{P}_{A_0 \cap \dots \cap A_{n-1}}(A_n).$$

PREUVE : On aura noté que l'hypothèse implique le fait que toutes les probabilités  $\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_k)$  sont non nulles, pour  $0 \leq k \leq n-1$ , ce qui légitime les conditionnements réalisés. On regarde ensuite droit dans les yeux le produit :

$$\mathbb{P}(A_0) \frac{\mathbb{P}(A_0 \cap A_1)}{\mathbb{P}(A_0)} \frac{\mathbb{P}(A_0 \cap A_1 \cap A_2)}{\mathbb{P}(A_0 \cap A_1)} \cdots \frac{\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_n)}{\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_{n-1})}.$$

■

EXEMPLE : On tire successivement et avec remise  $k$  boules numérotées de 1 à  $n$ . Quelle est la probabilité pour que les  $k$  numéros tirés soient distincts ?

Ici, on peut prendre pour événement  $A_i$  : « la  $i$ -ème boule tirée est distincte de toutes les précédentes ». Le point crucial est que  $\mathbb{P}(A_i)$  est délicate à évaluer (si les  $i-1$  premières boules ont donné le même numéro, ce n'est pas la même chose que s'ils étaient distincts). Par contre, il est facile d'évaluer  $\mathbb{P}_{A_1 \cap A_2 \cap \dots \cap A_{i-1}}(A_i)$  : c'est  $\frac{n-(i-1)}{n}$  (exactement  $i-1$  numéros sont interdits), et on trouve finalement que la probabilité pour qu'une application « prise au hasard » soit injective vaut <sup>3</sup> :

$$\frac{n}{n} \times \frac{n-1}{n} \times \dots \times \frac{n-(k-1)}{n} = \frac{n!}{(n-k)!n^k}.$$

REMARQUE : On préfère parfois prendre une suite **décroissante** d'événements et le résultat précédent se reformule : Si  $(B_k)_{0 \leq k \leq n}$  est une suite décroissante d'événements, alors :

$$\mathbb{P}(B_n) = \mathbb{P}(B_0) \times \mathbb{P}_{B_0}(B_1) \times \dots \times \mathbb{P}_{B_{n-1}}(B_n)$$

EXEMPLE : Dans l'exemple précédent on pouvait considérer  $B_k$  l'événement « les  $k$  premiers tirages sont distincts »...

### 3.3 Indépendance

Informellement : deux événements sont indépendants lorsque l'appartenance à l'un ne donne aucune information sur l'appartenance à l'autre. La plupart du temps c'est la modélisation du problème qui nous assure les indépendances, qu'on peut répercuter dans les calculs... et non l'inverse !

**Exercice 4.** Avant toute formalisation, diriez-vous que ces événements sont indépendants ?

- « Être taupin » et « être gaucher » (avec la modélisation vue plus haut).
- « Être testé positif » et « être malade » (toujours en suivant l'exemple vu plus haut).
- « Obtenir un résultat pair » et « obtenir un résultat  $\leq 4$  » (dans un dé équilibré à 8 faces numérotées de 1 à 8).
- $E_1$  : «  $D$  est pair » ;  $E_2$  : «  $G$  est pair » ;  $E_3$  : «  $D + G$  est pair », avec  $D$  et  $G$  les résultats obtenus en lançant indépendamment deux dés (équilibrés à 6 faces).

En pensant au dernier exemple du paragraphe précédent, on distinguera bien :

- l'indépendance de DEUX événements, ou encore l'indépendance deux à deux d'une famille d'événements ;
- l'indépendance mutuelle (ou globale) d'une famille d'événements (possiblement infinie).

DÉFINITION 11 — Indépendance de deux événements

Deux événements  $A$  et  $B$  sont déclarés *indépendants* lorsque :

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

Ce qui revient, si  $\mathbb{P}(A) > 0$ , à  $\mathbb{P}_A(B) = \mathbb{P}(B)$ .

On note dès maintenant que l'indépendance n'a rien à voir avec l'incompatibilité :

#### Attention

Deux événements indépendants ne sont pas forcément (et même très rarement) incompatibles !

**Exercice 5.** Avec cette définition, reprendre les exemples vus plus haut. Dans le dernier cas, il y a trois couples d'événements à considérer.

**Exercice 6.** Montrer que si  $A$  et  $B$  sont indépendants, alors  $A$  et  $\bar{B}$  le sont aussi.

Et pour plus de deux événements ?

3. Sans surprise, mais on voulait « oublier » ces résultats mystérieux sur les arrangements.

DÉFINITION 12 — *Indépendance mutuelle d'une famille (finie ou non) d'événements*

Une famille d'événements  $(A_i)_{i \in I}$  est déclarée mutuellement indépendante lorsque, pour toute sous-famille finie  $(A_{i_k})_{1 \leq k \leq N}$ , on a :

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_N}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_N})$$

Bien entendu, l'indépendance mutuelle entraîne l'indépendance deux à deux mais<sup>4</sup> la réciproque est fautive : voir l'exemple des deux dés.

*Les événements  $E_1$ ,  $E_2$  et  $E_3$  sont indépendants deux à deux mais pas mutuellement.*

## 4 Concept de variable aléatoire

### 4.1 Seul le résultat compte !

Les probabilités aspirent à modéliser des tas de choses issues de processus complexes, mais dont le résultat est simple, et doit être régi par des lois simples.

EXEMPLES :

- Dans un **lancer de pièce**, le résultat est un élément de l'ensemble  $\{P, F\}$  à deux éléments, mais le processus est complexe, quoique vaguement déterministe (les lois de la dynamique et les conditions initiales vont imposer le résultat).
- Dans un **lancer de dé réel**, il est difficile d'obtenir un résultat **réellement** équiprobable, mais on souhaite que le nombre obtenu soit **idéalement** l'un des entiers entre 1 et 6, chacun avec probabilité  $1/6$ .
- Quand on tire une **boule numérotée dans une urne**, le résultat souhaité est une quantité entre 1 et  $N$ , a priori de façon équiprobable.
- Si on s'intéresse, dans une **suite infinie de lancers de pièce**, au premier rang d'apparition de PILE, le résultat attendu est cette fois un entier strictement positif, dont la **loi** semble cette fois plus délicate<sup>5</sup>, mais surtout, il faut faire le pari que ce premier PILE va arriver. L'objet mathématique sous-jacent est alors une application définie pas tout à fait sur  $\Omega$  (ensemble des suites de PILE/FACE). Il faut enlever un événement atomique<sup>6</sup>, ce qui n'est pas trop grave, sa probabilité étant nulle ; on peut aussi décider que pour cet événement, le rang du premier PILE est 0, ou 42 : c'est a priori peu pertinent, mais si le formalisme est bien fait, cette valeur ne doit pas intervenir dans toute quantité pertinente qui sera calculée ensuite.
- Dans une file d'attente à un guichet ou une caisse, le **temps d'attente** est cette fois un réel positif (le nombre de secondes d'attente, par exemple).
- Si on s'intéresse à la **lettre que va taper un singe devant un clavier**, les méandres de la réflexion du bel animal sont tortueux, mais on fait le pari que le résultat obtenu sera l'une des 26 lettres de l'alphabet<sup>7</sup>, chacune avec la même probabilité.
- Un ivrogne de retour du pub revient en faisant des pas dans une rue... de façon aléatoire. Après  $n$  pas, où se trouve-t-il ? Dans l'ensemble  $\llbracket -n, n \rrbracket$ , mais encore ?

Dans chaque cas, la réalité est finalement abstraite en une application partant de  $\Omega$ , *ensemble mystérieux et non explicite en général*, et arrivant dans un ensemble assez simple.

DÉFINITION 13 — *Variable aléatoire discrète*

Une **variable aléatoire discrète** est une application d'un espace probabilisé  $(\Omega, \mathcal{T}, \mathbb{P})$  dans un ensemble  $E$  telle que :

- l'ensemble  $X(\Omega)$  des valeurs prises est fini ou dénombrable (« au plus dénombrable ») ;
- pour tout  $x \in X(\Omega)$ , l'ensemble  $X^{-1}(\{x\}) = \{\omega \in \Omega; X(\omega) = x\}$  des antécédents de  $y$  est dans  $\mathcal{T}$  (c'est un « événement » : il possède une probabilité).

Qui dit application dit patates : une façon pas trop mauvaise de représenter une variable aléatoire consiste à regrouper les éléments de  $\Omega$  en fonction de leur image par  $X$ .

4. Ai-je été assez lourd ?

5. Rien de bien méchant tout de même : ce sera une loi géométrique.

6. La suite constante de FACE.

7. Pour le singe, on a simplifié le clavier.

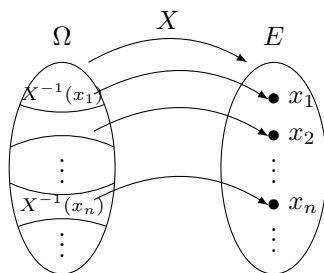


FIGURE 3 – Variable, application, patates...

REMARQUES :

- Pour des raisons d'efficacité, on s'autorisera des notations raccourcies pour les événements :

$$\{X = x_0\} = X^{-1}(x_0) = X^{-1}(\{x_0\}) = \{\omega \in \Omega; X(\omega) = x_0\}$$

- d'où des écritures telles que  $\mathbb{P}(X = x_0)$ ,  $\mathbb{P}(X \geq x_0)$  ou encore  $\mathbb{P}(X \in A)$ .
- On se fiche de  $\Omega$  (et même des  $X(\omega)$ , d'une certaine façon!). Ce qui compte, ce sont les probabilités d'obtenir tel ou tel résultat, c'est-à-dire les  $\mathbb{P}(X = x_i) = \mathbb{P}(X^{-1}(\{x_i\}))$ , ou encore les  $\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$  pour  $A \subset E$ .
- La condition d'appartenance à  $\mathcal{T}$  dit que les événements associés à  $X$  sont « mesurables » : si cette condition n'est pas vérifiée, c'est que la modélisation pose problème! Qu'on se rassure : ça n'arrivera pas...
- Cette condition « pour tout  $x \in X(\Omega)$ ,  $X^{-1}(\{x\}) \in \mathcal{T}$  » est équivalente à : « pour tout  $A \subset E$ ,  $X^{-1}(A) \in \mathcal{T}$  », grâce à la stabilité de  $\mathcal{T}$  par réunion dénombrable et au caractère au plus dénombrable de  $X(\Omega)$ .
- Si l'ensemble image est infini non dénombrable, c'est bien plus pénible : on veut à nouveau pouvoir mesurer des images réciproques, mais les  $X^{-1}(\{x\})$  ne suffisent plus, et il faut cette fois considérer dans  $E$  une tribu... laissons ça pour après la taupe (ou avant le bac, où la richesse et la puissance des programmes permet visiblement un traitement fin de ces questions).
- C'est  $X(\Omega)$  plus que  $E$  qui compte : pour un lancer de dé, on peut prendre  $E = \mathbb{R}...$  Dans les premiers exemples, seul le cas de l'attente à un guichet n'est pas dénombrable ( $X(\Omega) = \mathbb{R}_+$ ).
- La modélisation d'un problème consiste en général, sous des hypothèses vaguement réalistes, à conclure qu'on dispose d'une variable aléatoire vérifiant telle loi (sans expliciter les  $X(\omega)$ , ni même  $\Omega$ ).
- On dira souvent « dénombrable » au sens de « au plus dénombrable ». Si vraiment on veut parler d'un vrai dénombrable (infini, donc), on précisera!

## 4.2 Loi, fonction de répartition

La loi d'une variable aléatoire est définie formellement comme une fonction de probabilité sur l'ensemble d'arrivée. Concrètement, c'est quelque chose qui raconte quelle est la probabilité pour que  $X(\omega)$  soit égal à telle ou telle chose... ou appartienne à telle partie de  $E$ .

DÉFINITION 14 — *Loi d'une variable aléatoire*

|| La **loi** d'une variable aléatoire  $X : \Omega \rightarrow E$  est la loi  $\mathbb{P}_X$  définie sur  $E$  par :  $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$ .

REMARQUES :

- Oui, c'est bien une probabilité; left to the reader!
- $\mathbb{P}_X$  est définie sur la tribu complète  $\mathcal{P}(E)$ .
- La donnée de la loi  $\mathbb{P}_X$  fournit les probabilités  $\mathbb{P}(X = x)$  pour  $x \in E$ , mais réciproquement, si ces probabilités d'événements élémentaires sont connues, alors celle de toute partie  $A$  de  $E$  l'est aussi, grâce au caractère au plus dénombrable de  $X(\Omega)$  et à la  $\sigma$ -additivité :

$$\mathbb{P}_X(A) = \sum_{x \in X(\Omega) \cap A} \mathbb{P}(X = x).$$

- En pratique, pour décrire la loi d'une variable aléatoire discrète, on donne donc en général la probabilité des événements élémentaires.
- Quand deux variables aléatoires  $X$  et  $Y$  ont la même loi, on note :  $X \sim Y$ .

On peut représenter la loi d'une variable aléatoire discrète à valeurs dans  $\mathbb{R}$  par un diagramme en bâtons (« histogramme »), la hauteur de chaque bâton représentant la probabilité de l'événement élémentaire.

EXEMPLE : Pour le lancer de dés, on peut représenter la loi par le schéma suivant :

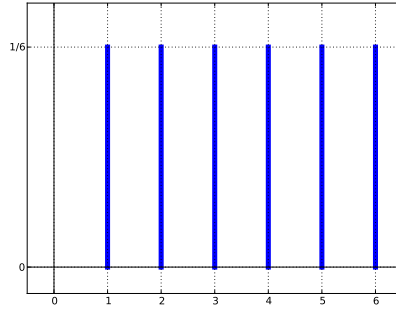


FIGURE 4 – Loi pour un dé honnête

Pour décrire la loi d'une variable aléatoire discrète à valeurs réelles, on peut s'intéresser également aux probabilités  $\mathbb{P}(X \leq t)$ . On note d'abord que  $\{X \leq t\}$  est bien un événement puisque c'est la réunion des événements  $\{X = x\}$ , avec  $x$  décrivant l'ensemble (au plus) **dénombrable**  $X(\Omega) \cap ]-\infty, t]$ .

DÉFINITION 15 — *Fonction de répartition*

Si  $X$  est une variable aléatoire réelle, sa **fonction de répartition** est la fonction  $F_X$  définie par :

$$\forall t \in \mathbb{R}, \quad F_X(t) = \mathbb{P}(X \leq t).$$

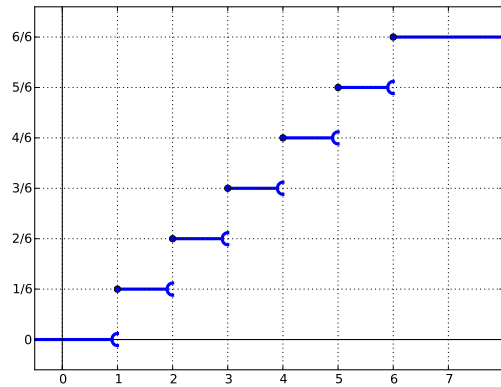


FIGURE 5 – Fonction de répartition pour un dé honnête

REMARQUE : il y a une situation notable où la fonction de répartition d'une variable aléatoire est intéressante : c'est lorsqu'on s'intéresse au maximum de  $k$  variables **indépendantes**  $X_1, \dots, X_k$ . Si on note  $Y$  ce maximum, on a en effet :

$$\{Y \leq n\} = \bigcap_{i=1}^k \{X_i \leq n\}$$

et donc :  $F_Y = F_{X_1} F_{X_2} \dots F_{X_k}$ , ce qui permet ensuite de reconstituer la loi de  $Y$ .

Toutes les fonctions de répartition partagent un certain nombre de propriétés :

PROPOSITION 2 — Propriétés des fonctions de répartition

Soit  $X$  une variable aléatoire réelle discrète.

- $F_X$  est croissante, tend respectivement vers 0 et 1 en  $-\infty$  et  $+\infty$ .
- $F_X$  est continue en tout point de  $\mathbb{R} \setminus X(\Omega)$ .
- $F_X$  est continue à droite en tout point.
- $F_X$  est continue à gauche en  $t$  si et seulement si  $\mathbb{P}(X = t) = 0$ .

**Exercice 7.** Soit  $X$  une variable aléatoire réelle discrète de fonction de répartition  $F_X$ . Montrer :

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) \quad \text{et} \quad \mathbb{P}(a \leq X \leq b) = F_X(b) - \lim_{t \rightarrow a^-} F_X(t).$$

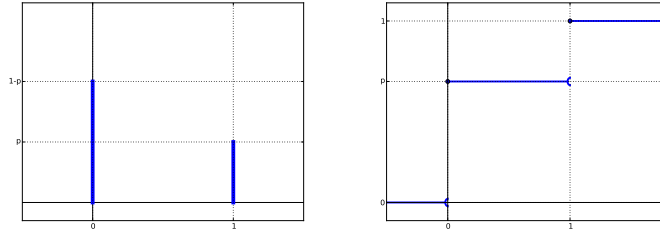


FIGURE 6 – Loi et fonction de répartition pour une variable de Bernoulli  $\mathcal{B}(1/3)$

Parfois, la modélisation réclame qu’il existe un espace probabilisé et une variable aléatoire  $X$  sur cet espace tels que les événements «  $X = x_0$  » aient une probabilité imposée.

**THÉORÈME 6 — Probabilités imposées**

Si  $X$  prend ses valeurs dans  $\{x_i \mid i \in \mathbb{N}\}$  (les  $x_i$  étant distincts) et  $(p_i)_{i \in \mathbb{N}}$  est une suite de réels positifs tels que la série  $\sum p_i$  est convergente et de somme égale à 1, alors il existe une probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{T})$  telle que pour tout  $i \in \mathbb{N}$ ,  $\mathbb{P}(X = x_i) = p_i$ .

Ce théorème est aussi utile que sa preuve est sans intérêt (pour nous). Bien entendu, il s’étend au cas où les  $x_i$  sont en nombre fini. C’est via ce procédé qu’on définit les grandes lois classiques dans la section suivante. Une loi  $\mathcal{L}$  sera donnée par l’ensemble  $\{x_i \mid i \in \mathbb{N}\}$  des valeurs atteintes et les probabilités  $\mathbb{P}(X = x_i)$ . On notera :  $X \hookrightarrow \mathcal{L}$ .

### 4.3 Lois usuelles

Les tirages à pile ou face (équiprobables ou non) vont fournir un support naturel pour les trois premières lois présentées.

- **Loi de Bernoulli :** Soit  $p \in ]0, 1[$ . Une variable aléatoire  $X$  suit la loi de Bernoulli de paramètre  $p$  – et on note  $X \hookrightarrow \mathcal{B}(p)$  – lorsque  $X(\Omega) = \{0, 1\}$ ,  $\mathbb{P}(X = 1) = p$  et  $\mathbb{P}(X = 0) = 1 - p$ .

EXEMPLES :

- Un jeu de pile ou face, biaisé ou non, est naturellement modélisé par une variable aléatoire suivant une loi de Bernoulli (en associant par exemple PILE à 1).

- Si  $A$  est un événement, sa fonction indicatrice  $\mathbf{1}_A : x \in \Omega \mapsto \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$  suit une loi de Bernoulli de paramètre  $\mathbb{P}(A)$ .

- **Loi binomiale :** Si on réalise  $n > 0$  tirages à pile ou face indépendants (modélisés par une Bernoulli  $\mathcal{B}(p)$ ), le nombre  $X$  de PILE (qui peut être vu comme la somme de  $n$  variables de Bernoulli indépendantes ; nous y reviendrons plus tard) prend ses valeurs dans  $\llbracket 0, n \rrbracket$ , avec pour tout  $k \in \llbracket 0, n \rrbracket$ ,  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$  : c’est la loi binomiale, et on note  $X \hookrightarrow \mathcal{B}(n, p)$ .

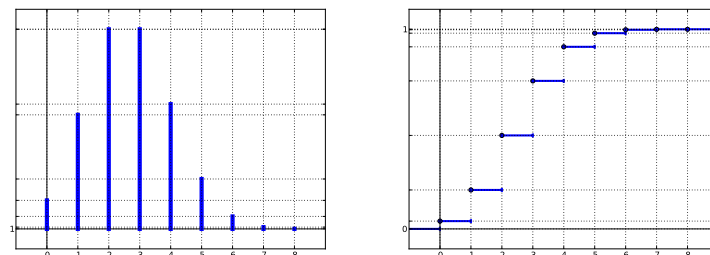


FIGURE 7 – Loi et fonction de répartition pour une variable binomiale  $\mathcal{B}(8, 1/3)$

**Exercice 8.** Si  $X \hookrightarrow \mathcal{B}(n, p)$ , pour quel(s)  $k$  a-t-on  $\mathbb{P}(X = k)$  maximale ? (On pourra évaluer le rapport entre deux termes consécutifs.)

- **Loi géométrique :** Toujours avec notre pièce biaisée, on réalise une série de lancers indépendants tant qu’on n’a pas obtenu PILE, et on s’intéresse au rang d’apparition du premier pile (on travaille

donc sur l'espace des suites de PILE/FACE privé de la suite constituée uniquement de FACE – événement de probabilité nulle). Ce rang vaut  $k \geq 1$  si et seulement si les  $k - 1$  premiers tirages ont fourni FACE et le  $k$ -ième a fourni PILE :

$$X(\Omega) = \mathbb{N}^* \quad \text{et} \quad \forall k \in \mathbb{N}^*, \mathbb{P}(X = k) = p(1 - p)^{k-1}.$$

Il s'agit de la loi géométrique de paramètre  $p : X \hookrightarrow \mathcal{G}(p)$ .

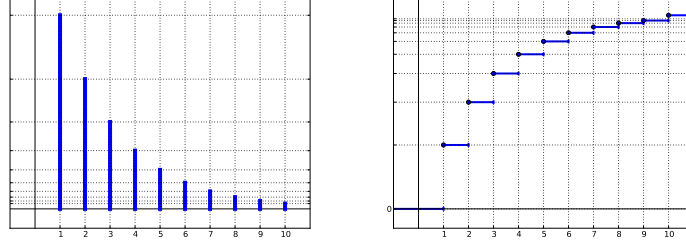


FIGURE 8 – Loi et fonction de répartition pour la loi géométrique  $\mathcal{G}(1/3)$

- **Loi de Poisson :** Cette loi est d'interprétation plus délicate. Elle modélisera des événements « rares » (du point de vue microscopique) qu'on observe d'un point de vue macroscopique ; détails plus tard dans ce chapitre !

$$X(\Omega) = \mathbb{N} \quad \text{et} \quad \forall k \in \mathbb{N}, \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

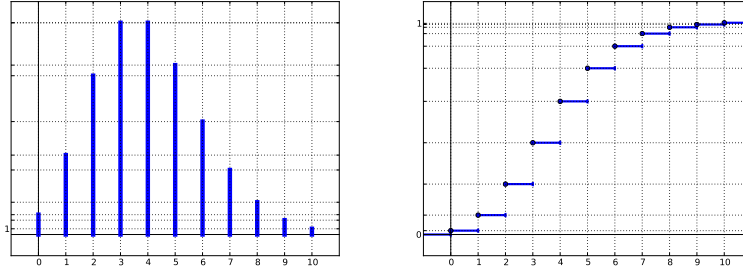


FIGURE 9 – Loi et fonction de répartition pour la loi de Poisson  $\mathcal{P}(4)$

**Exercice 9.** Si  $X \hookrightarrow \mathcal{P}(\lambda)$ , pour quel(s)  $k$  a-t-on  $\mathbb{P}(X = k)$  maximale ?

#### 4.4 Couples de variables aléatoires

Si  $X$  et  $Y$  sont deux variables aléatoires discrètes sur un même espace probabilisé  $(\Omega, \mathcal{T}, \mathbb{P})$ , alors le couple  $(X, Y)$  est lui-même une variable aléatoire discrète<sup>8</sup> sur le même espace (et l'ensemble d'arrivée est le produit cartésien des ensembles d'arrivée de  $X$  et  $Y$ ).

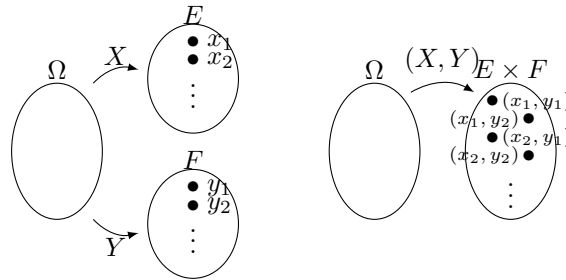


FIGURE 10 – Un couple de variables aléatoires

8. Le produit cartésien de deux ensembles (au plus) dénombrables est lui-même (au plus) dénombrable.



EXEMPLES :

- Alice et Bob passent un examen (une suite de questions à réponses binaires 0/1). Alice répond au hasard, et Bob aussi (de façon indépendante d'Alice). Leurs réponses à chaque question fournissent un couple de variables aléatoires à valeurs dans  $\{0, 1\}^2$ . Les deux composantes  $A$  et  $B$  suivent une loi de Bernoulli  $\mathcal{B}(1/2)$ .
- Dans le même cadre, Alice répond au hasard, et Bob recopie sur Alice. On récupère un nouveau couple de variables aléatoires qui – individuellement – ont le même comportement que les composantes  $A$  et  $B$  précédentes : tout le monde suit une loi de Bernoulli  $\mathcal{B}(1/2)$ .
- Toujours dans le même cadre, c'est Bob qui répond au hasard, et Alice qui copie sur Bob. Les lois sont les mêmes que dans le cas précédent.
- Enfin, Alice répond au hasard, et Bob décide de prendre systématiquement la réponse opposée.

La loi de  $(X, Y)$  s'appelle la « loi (con)jointe » de  $(X, Y)$ . Si, dans le couple, on ne regarde qu'une composante, on obtient les lois marginales.

DÉFINITION 16 — *Loi conjointe, lois marginales*

Si  $(X, Y)$  est un couple de variable aléatoires à valeurs dans  $E \times F$ , leur **loi conjointe** est leur loi au sens usuel : il s'agit de la probabilité  $\mathbb{P}_{(X,Y)}$  définie sur  $E \times F$  par

$$\forall C \in E \times F, \quad \mathbb{P}_{(X,Y)}(C) = \mathbb{P}((X, Y) \in C).$$

Les **lois marginales** du couple  $(X, Y)$  sont celles de  $X$  et  $Y$  au sens suivant :

$$\forall A \subset E \quad \forall B \subset F, \quad \mathbb{P}_X(A) = \mathbb{P}((X, Y) \in A \times F) \quad \text{et} \quad \mathbb{P}_Y(B) = \mathbb{P}((X, Y) \in E \times B)$$

Si les valeurs prises sont dans un ensemble fini, on peut représenter dans un tableau la loi du couple. En sommant sur les lignes ou colonnes, on obtient les lois marginales. Dans le cas où  $X$  ou  $Y$  prend une infinité de valeurs, on peut continuer à imaginer une espèce de matrice infinie.

$A \backslash B$	0	1	$\mathbb{P}_A$
0	1/4	1/4	1/2
1	1/4	1/4	1/2
$\mathbb{P}_B$	1/2	1/2	1

$A \backslash B$	0	1	$\mathbb{P}_A$
0	1/2	0	1/2
1	0	1/2	1/2
$\mathbb{P}_B$	1/2	1/2	1

$A \backslash B$	0	1	$\mathbb{P}_A$
0	0	1/2	1/2
1	1/2	0	1/2
$\mathbb{P}_B$	1/2	1/2	1

FIGURE 11 – Alice et Bob : mêmes lois marginales, mais trois lois conjointes distinctes

### ⚠ Attention

On voit dans les tableaux précédents que la connaissance des lois des lois marginales de  $A$  et  $B$  ne dit donc rien sur la loi du couple  $(A, B)$ .

REMARQUE : On voit aussi que même en cas de forte corrélation («  $X$  et  $Y$  se ressemblent de façon suspecte » – ce sera précisé plus tard) **on ne peut absolument pas en déduire de lien de causalité** : est-ce Alice qui a recopié sur Bob, ou le contraire ? Il est également possible qu'ils aient l'un et l'autre recopié sur un troisième.

**Exercice 10.** On suppose que  $X$  et  $Y$  décrivent le résultat du lancer indépendant de deux dés : le premier honnête, et le deuxième biaisé :  $\mathbb{P}(Y = 1) = 1/3$ . Décrire par un tableau la loi de  $(X, Y)$  ainsi que les lois marginales.

On peut s'intéresser aux colonnes/lignes des tableaux précédents, c'est-à-dire à la loi de  $X$  (respectivement  $Y$ ) sous condition sur  $Y$  (respectivement  $X$ ).

DÉFINITION 17 — *Loi conditionnelle de  $Y$  sous condition  $X = x$*

Si  $(X, Y)$  est un couple de variables aléatoires à valeurs dans  $E \times F$ , et  $b \in F$  est tel que  $\mathbb{P}(Y = b) > 0$ , la probabilité sous condition  $Y = b$ , notée  $\mathbb{P}_{Y=b}$  est la loi de probabilité définie sur  $E$  par :

$$\forall A \subset E, \quad \mathbb{P}_{Y=b}(A) = \mathbb{P}(X \in A | Y = b) = \frac{\mathbb{P}((X, Y) \in A \times \{b\})}{\mathbb{P}(Y = b)}.$$

On définit bien entendu la notion symétrique, avec condition sur  $X$ . On peut également conditionner par  $Y \leq b$ , etc. Pour calculer ces lois sous conditions, il suffit essentiellement de normaliser les colonnes/lignes du tableau décrivant la loi conjointe.

$b$	0	1
$\mathbb{P}_{B=b}(A=0)$	1/2	1/2
$\mathbb{P}_{B=b}(A=1)$	1/2	1/2

$b$	0	1
$\mathbb{P}_{B=b}(A=0)$	1	0
$\mathbb{P}_{B=b}(A=1)$	0	1

$b$	0	1
$\mathbb{P}_{B=b}(A=0)$	0	1
$\mathbb{P}_{B=b}(A=1)$	1	0

FIGURE 12 – Alice et Bob : les lois de  $A$  sous condition sur  $B$  dans les trois scénarios

L'exercice suivant (dont seul le début est faisable en attendant quelques paragraphes) est assez typique de l'utilisation des lois sous conditions : ce sont elles que la modélisation naturelle du problème impose.

**Exercice 11.** CCP 2015 et 2016 (deux fois)

$N$  désigne le nombre d'électrons produits dans une réaction :  $N \hookrightarrow \mathcal{P}(\lambda)$ . Les électrons sont efficaces<sup>9</sup> dans une proportion<sup>10</sup>  $p \in ]0, 1[$ , et  $X$  (respectivement  $Y$ ) désigne le nombre d'électrons efficaces (respectivement inefficaces).

1. Donner la loi de  $X$  sous la condition  $N = j$ .
2. Donner la loi conjointe de  $(X, N)$ .
3. Donner la loi de  $X$ , son expérience et sa variance.
4. Les variables  $X$  et  $Y$  sont-elles indépendantes ?

## 4.5 Variables indépendantes

On rappelle que dans un espace probabilisé  $(\Omega, \mathcal{T}, \mathbb{P})$ , deux événements  $A, B \in \mathcal{T}$  sont déclarés indépendants lorsque  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Ensuite,  $n$  événements  $A_1, \dots, A_n$  sont **mutuellement** indépendants lorsque dès qu'on en prend  $k$ , disons  $A_{i_1}, \dots, A_{i_k}$ , on a  $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$ , notion qui est plus forte que l'indépendance deux à deux. Pour des variables aléatoires  $X$  et  $Y$ , c'est essentiellement la même chose : elles sont déclarées indépendantes lorsque les événements relatifs à  $X$  (i.e. définis à l'aide de  $X(\omega)$ ) sont indépendants de ceux relatifs à  $Y$ .

DÉFINITION 18 — *Indépendance de deux variables aléatoires*

Deux variables aléatoires discrètes  $X$  et  $Y$  définies sur  $\Omega$  sont dites **indépendantes** lorsque :

$$\forall x \in X(\Omega) \forall y \in Y(\Omega), \quad \mathbb{P}(X = x \text{ et } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

On notera :  $X \perp\!\!\!\perp Y$ .

REMARQUES :

- On peut aussi prendre  $x$  dans  $E$  plutôt que  $X(\Omega)$ ...
- Par  $\sigma$ -additivité et caractère discret des lois, cette relation s'étend à des couples de parties :  $X$  et  $Y$  sont indépendantes si et seulement si :

$$\forall A \subset E \forall B \subset F, \quad \mathbb{P}(X \in A \text{ et } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

- Dire que  $X$  et  $Y$  sont indépendantes revient à dire que toutes les probabilités sous condition  $\mathbb{P}_{X=x}$  suivent la même loi.
- En pratique, c'est la modélisation qui assure l'indépendance de variables aléatoires (lancers indépendants de dés ou de pièces, désintégration de particules ne se connaissant pas personnellement, ...).

EXEMPLE : Pour Alice et Bob, sans surprise, les variables  $A$  et  $B$  sont indépendantes seulement dans le premier des scénarios.

**Exercice 12.** On suppose que  $X_1$  et  $X_2$  sont deux variables aléatoires suivant respectivement des lois de Poisson  $\mathcal{P}(\lambda_1)$  et  $\mathcal{P}(\lambda_2)$ . Montrer que leur somme  $S = X_1 + X_2$  suit une loi de Poisson  $\mathcal{P}(\lambda_1 + \lambda_2)$ .

La notion d'indépendance s'étend évidemment à une famille (finie ou non) de variables aléatoires.

9. Quelle formulation misérable...

10. J'imagine qu'il faut comprendre « avec une probabilité ».

DÉFINITION 19 — Variables mutuellement indépendantes

Des variables aléatoires  $X_1, \dots, X_n$  sont dites **mutuellement indépendantes** lorsque pour tout  $k \in \llbracket 1, n \rrbracket$  et tout  $x_1, \dots, x_k \in \mathbb{R}$ , on a :

$$1 \leq i_1 < i_2 < \dots < i_k \leq n \implies \mathbb{P}(X_{i_1} = x_1 \text{ et } \dots \text{ et } X_{i_k} = x_k) = \mathbb{P}(X_{i_1} = x_1) \cdots \mathbb{P}(X_{i_k} = x_k).$$

Pour une famille **infinie** de variables aléatoires, il y a indépendance mutuelle lorsque toute sous-famille finie est mutuellement indépendante.

EXEMPLES :

- Si on somme  $n$  variables mutuellement indépendantes qui suivent une loi de Bernoulli  $\mathcal{B}(p)$ , on obtient une nouvelle variable suivant une loi binomiale  $\mathcal{B}(n, p)$  (typiquement : le nombre de PILE obtenus sur  $n$  lancers indépendants de pièces biaisées ou non).
- Comme pour les événements indépendants, l'indépendance deux à deux de variables aléatoires n'entraîne pas leur indépendance mutuelle : prendre  $X \hookrightarrow \mathcal{B}(1/2)$ ,  $Y \hookrightarrow \mathcal{B}(1/2)$  indépendant de  $X$ , et  $Z = X + Y$  modulo 2 (au sens du reste dans la division euclidienne : «  $1 + 1 = 0$  »). Il y a bien indépendance deux à deux, mais :

$$\frac{1}{4} = \mathbb{P}(X = 0 \text{ et } Y = 0 \text{ et } Z = 0) \neq \mathbb{P}(X = 0)\mathbb{P}(Y = 0)\mathbb{P}(Z = 0) = \frac{1}{8},$$

ou encore :

$$0 = \mathbb{P}(X = 0 \text{ et } Y = 0 \text{ et } Z = 1) \neq \mathbb{P}(X = 0)\mathbb{P}(Y = 0)\mathbb{P}(Z = 1) = \frac{1}{8}.$$

- Étant donnée une loi  $\mathcal{L}$ , on admet l'existence d'un espace probabilisé et d'une famille  $(X_n)_{n \in \mathbb{N}}$  de variables mutuellement indépendantes suivant toutes la loi  $\mathcal{L}$ . C'est ce type de résultat qu'on utilise pour modéliser une suite infinie de lancers indépendants de pièce ou dé.

La proposition suivante dit que si Alice et Bob calculent chacun une quantité à l'aide de données indépendantes, alors les résultats qu'ils obtiennent sont indépendants. L'usage probabiliste veut que les composées  $f \circ X$  soient notées  $f(X)$ , comme en physique (la variable aléatoire  $X$  étant plus vue comme une *quantité* que comme une *fonction*, d'où son nom, accessoirement).

PROPOSITION 3 — Indépendance et composition

Si  $X$  et  $Y$  sont deux variables aléatoires discrètes indépendantes à valeurs dans respectivement  $E$  et  $F$ , et  $f$  et  $g$  sont deux fonctions définies respectivement sur  $E$  et  $F$ , alors les variables aléatoires discrètes  $f(X)$  et  $g(Y)$  sont indépendantes.

PREUVE (HORS PROGRAMME) : Formellement (sachant que ces sommations sont licites du fait de la  $\sigma$ -additivité et de la dénombrabilité des pré-images) :

$$\begin{aligned} \mathbb{P}(f(X) = u \text{ et } g(Y) = v) &= \sum_{\substack{x \in f^{-1}(u) \\ y \in g^{-1}(v)}} \mathbb{P}(X = x \text{ et } Y = y) = \sum_{\substack{x \in f^{-1}(u) \\ y \in g^{-1}(v)}} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &= \left( \sum_{x \in f^{-1}(u)} \mathbb{P}(X = x) \right) \left( \sum_{y \in g^{-1}(v)} \mathbb{P}(Y = y) \right) \\ &= \mathbb{P}(f(X) = u)\mathbb{P}(g(Y) = v) \end{aligned}$$

■

Dans la même famille :

PROPOSITION 4 — Théorème des coalitions – vaguement H.-P.

Si  $X_1, \dots, X_n$  sont des variables aléatoires mutuellement indépendantes,  $k \in \llbracket 1, n-1 \rrbracket$ ,  $Y = f(X_1, \dots, X_k)$  et  $Z = g(X_{k+1}, \dots, X_n)$ , alors  $Y$  et  $Z$  sont indépendantes.

EXEMPLE : Prenons  $n+k$  variables mutuellement indépendantes  $X_1, \dots, X_{n+k}$  suivant des lois de Bernoulli  $\mathcal{B}(p)$ , notons  $Y = X_1 + \dots + X_n$  et  $Z = X_{n+1} + \dots + X_{n+k}$ .

- D'une part,  $Y + Z$  est la somme de  $n + k$  Bernoulli indépendantes, donc suit une loi binomiale  $\mathcal{B}(n+k, p)$ .

- D'autre part, le théorème des coalitions nous dit que  $Y$  et  $Z$  sont indépendantes, avec  $Y \hookrightarrow \mathcal{B}(n, p)$  et  $Z \hookrightarrow \mathcal{B}(k, p)$ .

Si  $Y \hookrightarrow \mathcal{B}(n, p)$  et  $Z \hookrightarrow \mathcal{B}(k, p)$  avec  $Y$  et  $Z$  indépendantes, alors  $Y + Z \hookrightarrow \mathcal{B}(n + k, p)$ .

**Exercice 13.** Détecter l'arnaque (fine) dans la preuve précédente ; puis la réparer !

DÉFINITION 20 — Variables indépendantes identiquement distribuées

On parle d'une « suite de variables indépendantes identiquement distribuées », abrégé en « i.i.d » pour une famille  $(X_n)_{n \in \mathbb{N}}$  où toutes les  $X_n$  sont des variables aléatoires de même loi, et indépendantes. On admet (c'est raisonnable et pratique, sans être trivial) que pour toute loi de probabilité, il existe effectivement de telles suites avec cette loi...

## 5 Espérance et variance

### 5.1 Espérance, théorème de transfert

Moralement, l'espérance d'une variable aléatoire est sa moyenne :

- Le joueur de loto sait qu'il perd souvent (peu) et qu'il gagne rarement (gros). Mais encore ? Il intuite vaguement<sup>11</sup> que « en moyenne, il est perdant », mais comment quantifier cela ?
- On tire à pile ou face tant qu'on n'a pas eu PILE. Le premier succès peut arriver au temps 1, mais aussi au temps 10 (vraiment pas de chance) ou même au temps 1000 (belle obstination, mais il faut probablement mettre en doute l'honnêteté de la pièce !). Comment définir la valeur *moyenne* du temps du premier succès ?
- L'ivrogne qui part du bar en marchant de façon aléatoire a peu de chances ( $\mathbb{P} = 1/2^n$ ) de faire les  $n$  pas qui le ramèneront à la maison du premier coup... mais se dit qu'il va bien finir par arriver devant chez lui. Oui, mais en combien de temps en moyenne ?

La définition de l'espérance peut se faire en termes de familles sommables... mais privilégions le vieux point de vue « séries ». Le premier point de vue est pertinent... pour prouver les propriétés de manipulation des espérances... propriétés qui seront admises !

DÉFINITION 21 — Espérance

Une variable aléatoire discrète  $X$  à valeurs dans  $\mathbb{R}$  est dite d'**espérance finie** lorsque la série  $\sum x_n \mathbb{P}(X = x_n)$  est absolument convergente (avec  $X(\Omega) = \{x_n \mid n \in \mathbb{N}\}$ ). Lorsque c'est le cas, son espérance est :

$$\mathbb{E}(X) := \sum_{n=0}^{+\infty} x_n \mathbb{P}(x_n).$$

On admet (conséquence de l'absolue convergence/sommabilité) que la valeur de cette somme est indépendante de la façon d'énumérer  $X(\Omega)$ .

$X$  est dite *centrée* lorsqu'elle possède une espérance, et que celle-ci est nulle.

En termes de familles sommables, l'existence de l'espérance est synonyme de sommabilité de  $(x \mathbb{P}(X = x))_{x \in X(\Omega)}$ .

EXEMPLES :

- Si une variable aléatoire prend un nombre fini de valeurs, alors elle est d'espérance finie (et la série devient une somme usuelle).
- On rappelle (et surtout, on invite le lecteur à faire le calcul) que :
  - si  $X \hookrightarrow \mathcal{B}(p)$  alors  $\mathbb{E}(X) = p$  ;
  - si  $X \hookrightarrow \mathcal{B}(n, p)$  alors  $\mathbb{E}(X) = np$  ;
  - si  $X \hookrightarrow \mathcal{U}([a, b])$  alors  $\mathbb{E}(X) = \frac{a+b}{2}$ .
- Si  $X$  suit une loi géométrique  $\mathcal{G}(p)$ , alors  $X$  est d'espérance finie, avec  $\mathbb{E}(X) = 1/p$ .
- Si  $X$  suit une loi de Poisson  $\mathcal{P}(\lambda)$ , alors  $X$  est d'espérance finie, avec  $\mathbb{E}(X) = \lambda$ .

Notons que pour les variables à valeurs dans  $\mathbb{N}$ , on a une formule étrange de prime abord :

<sup>11</sup>. En le disant parfois de façon plus rustique.

PROPOSITION 5 — Espérance d'une variable à valeurs dans  $\mathbb{N}$

Si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{N}$  avec une espérance finie, alors :

$$\mathbb{E}(X) = \sum_{k=1}^{+\infty} \mathbb{P}(X \geq k)$$

PREUVE : En admettant la sauvegarde intervention de sommes (encore une fois, le point crucial est le fait que les convergences sont absolues, et que dans ces conditions, c'est la fête ; penser au théorème «  $\sum \int |f_n| \dots$  ») :

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} n \mathbb{P}(X = n) = \sum_{n=1}^{\infty} \sum_{k=1}^n \mathbb{P}(X = n) = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbb{P}(X = n) = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k)$$

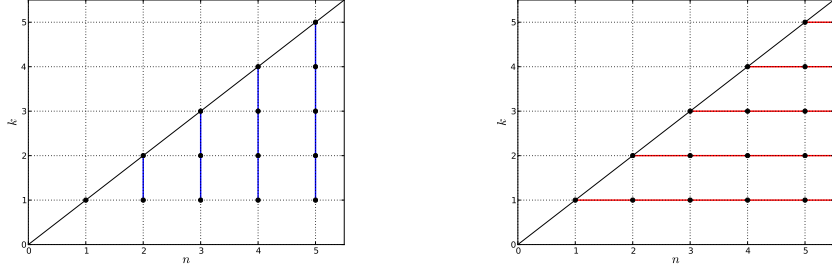


FIGURE 13 – Pour comprendre l'intervention  $\sum_{n=1}^{\infty} \sum_{k=1}^n = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty}$

REMARQUE : En écrivant

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} n (\mathbb{P}(X \geq n) - \mathbb{P}(X \geq n+1))$$

et en continuant comme on imagine, on obtient une seconde preuve que vous pourrez juger meilleure ou moins bonne... mais en vous relisant vous trouverez probablement une erreur/arnaque (réparable) sur une question de convergence.

En pratique, on veut souvent calculer l'espérance de « compositions »  $f \circ X$ , ou encore  $f(X)$ , avec l'abus de langage usuel. Le théorème suivant dit qu'il n'est pas nécessaire de savoir quels sont les  $\omega$  tels que  $f(X(\omega)) = y$  (bref, connaître la loi de  $f(X)$ ) : celle de  $X$  est suffisante.

THÉORÈME 7 — Dit « de transfert »

Soient  $X : \Omega \rightarrow E$  une variable aléatoire discrète et  $f$  une application de  $X(\Omega) = \{x_n \mid n \in \mathbb{N}\}$  dans  $\mathbb{R}$ . La variable aléatoire  $f(X)$  est d'espérance finie si et seulement si  $\sum f(x_n) \mathbb{P}(X = x_n)$  est absolument convergente. Lorsque c'est le cas, on a :

$$\mathbb{E}(f(X)) = \sum_{n=0}^{+\infty} f(x_n) \mathbb{P}(X = x_n).$$

PREUVE (HORS PROGRAMME) : En jetant un voile pudique sur les interversions de sommes : on a d'une part  $f(X(\Omega)) = \{y_k \mid k \in \mathbb{N}\}$  et d'autre part, on peut partitionner  $X(\Omega)$  à l'aide des images par  $f$  :

$$X(\Omega) = \{x_n, n \in \mathbb{N}\} = \bigsqcup_{k \in \mathbb{N}} \underbrace{f^{(-1)}(\{y_k\})}_{E_k}.$$

On a alors :

$$\begin{aligned} \mathbb{E}(f(X)) &= \sum_{y \in f(X(\Omega))} y \mathbb{P}(f(X) = y) = \sum_{k=0}^{\infty} y_k \mathbb{P}(f(X) = y_k) = \sum_{k=0}^{\infty} y_k \mathbb{P}(X \in E_k) \\ &= \sum_{k=0}^{\infty} \sum_{x \in E_k} f(x) \mathbb{P}(X = x) = \sum_{n \in \mathbb{N}} f(x_n) \mathbb{P}(X = x_n). \end{aligned}$$

Sous le voile « familles sommables », il y a aussi le fait que le résultat ne dépend pas de la façon d'énumérer  $X(\Omega)$  ou  $f(X(\Omega))$ ...

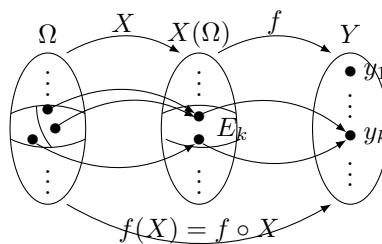


FIGURE 14 – Le transfert et les patates

REMARQUE : Dans l'énoncé précédent, on a bien entendu supposé que  $(x_n)_{n \in \mathbb{N}}$  était une énumération (injective) de  $X(\Omega)$ . Si  $X(\Omega)$  est non pas dénombrable mais fini, la sommation est finie, et il n'est plus question de série.

On admettra les résultats suivants, aussi raisonnables que casse-pieds à montrer :

PROPOSITION 6 — Quelques propriétés de l'espérance

— *Linéarité* : si  $X_1$  et  $X_2$  sont d'espérance finie et  $\alpha, \beta \in \mathbb{R}$ , alors  $\alpha X_1 + \beta X_2$  aussi, avec

$$\mathbb{E}(\alpha X_1 + \beta X_2) = \alpha \mathbb{E}(X_1) + \beta \mathbb{E}(X_2).$$

— *Positivité* : si  $X$  est à valeurs dans  $\mathbb{R}^+$  et d'espérance finie, alors (incroyable !)  $\mathbb{E}(X) \geq 0$ .

— *Croissance* : si  $X_1$  et  $X_2$  sont deux variables aléatoires d'espérance finie avec  $X_1 \leq X_2$ , alors  $\mathbb{E}(X_1) \leq \mathbb{E}(X_2)$ .

REMARQUE : Les résultats précédents restent valides si les hypothèses sont faites « presque sûrement » : par exemple, si  $\mathbb{P}(X_1 \leq X_2) = 1$ , alors  $\mathbb{E}(X_1) \leq \mathbb{E}(X_2)$ ...

Ainsi, l'espérance d'une somme est la somme des espérances. Bien entendu, il n'y a a priori aucune raison pour que la moyenne d'un produit soit le produit des moyennes. Mais lorsque des variables sont indépendantes, ce résultat est tout de même vrai !

PROPOSITION 7 — Espérance d'un produit de variables indépendantes

Si  $X$  et  $Y$  sont d'espérance finie et indépendantes, alors  $XY$  est d'espérance finie, avec :

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

PREUVE : Hors programme! ■

Bien entendu, cette propriété s'étend à tout produit fini de variables aléatoires indépendantes.

**Exercice 14.** À un devoir surveillé de thermo, la moyenne est de 6. Que dire de la proportion d'élèves ayant eu la moyenne ? :-)

PROPOSITION 8 — Inégalité de Markov

Soit  $X$  une variable aléatoire discrète à valeurs dans  $\mathbb{R}^+$  et d'espérance finie. Si  $\alpha > 0$ , alors :

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

PREUVE : Dans le calcul de  $\mathbb{E}(X)$ , on isole les valeurs de  $X$  minorées par  $\alpha$  :

$$\mathbb{E}(X) = \sum_{x_i < \alpha} \underbrace{x_i \mathbb{P}(X = x_i)}_{\geq 0} + \sum_{x_i \geq \alpha} x_i \mathbb{P}(X = x_i) \geq \sum_{x_i \geq \alpha} \underbrace{x_i}_{\geq \alpha} \mathbb{P}(X = x_i) \geq \alpha \sum_{x_i \geq \alpha} \mathbb{P}(X = x_i) = \alpha \mathbb{P}(X \geq \alpha).$$

Dans le paragraphe suivant, on va voir des moyens sensiblement plus efficaces pour contrôler ce type d'écart. ■

## 5.2 Contrôle de l'écart à la moyenne ; variance

On souhaite contrôler l'écart entre  $X$  et sa « moyenne »  $\mathbb{E}(X)$ . La première quantité à laquelle on peut penser est  $|X - \mathbb{E}(X)|$ . C'est fondamentalement cette chose dont on aimerait maîtriser la loi. En pratique, son carré sera bien plus facilement manipulable (et connaître la loi du carré, c'est connaître la loi de la valeur absolue...). La **variance** de  $X$  sera l'espérance de  $(X - \mathbb{E}(X))^2$ ... lorsqu'elle existe !  
Notons tout d'abord :

*Si  $X^2$  a une espérance finie, alors  $X$  aussi.*

C'est une conséquence de l'inégalité de Cauchy-Schwarz qui viendra plus tard, mais on peut aussi noter que dans la série  $\sum |x_n| \mathbb{P}(X = x_n)$ , il y a les termes pour lesquels  $|x_n| \leq 1$  ; les sommes partielles sont alors majorées par 1. Quant aux termes pour lesquels  $|x_n| \geq 1$ , on a  $|x_n| \leq |x_n^2|$ , donc les sommes partielles sont majorées par  $\mathbb{E}(X^2)$ .

**Exercice 15.** *Bricoler une variable aléatoire  $X$  d'espérance finie, mais dont le carré n'est pas d'espérance finie.*

Ainsi, puisque  $(X - \mathbb{E}(X))^2 = X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2$ ,  $X^2$  est d'espérance finie si et seulement si  $(X - \mathbb{E}(X))^2$  l'est.

DÉFINITION 22 — *Variance et écart type d'une variable aléatoire*

Une variable aléatoire discrète à valeurs réelles  $X$  possède une **variance** lorsque  $X^2$  est d'espérance finie. Sa variance vaut alors :

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

et son **écart-type** vaut :  $\sigma = \sqrt{\text{Var}(X)}$ . Une variable aléatoire possédant (une espérance et) une variance est dite réduite lorsque la variance vaut 1.

Toujours grâce à la remarque précédente, on a même :

PROPOSITION 9 — Calcul pratique de la variance – « formule de Huygens »

Si  $X$  est une variable aléatoire possédant une variance, alors :

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

REMARQUE : On peut d'ailleurs en déduire :  $\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)$ , qui pourra également être vu comme conséquence de l'inégalité de Cauchy-Schwarz.

EXEMPLE : Si  $X \hookrightarrow \mathcal{B}(p)$ , alors  $X = X^2$ , donc  $\mathbb{E}(X) = \mathbb{E}(X^2) = p$  puis :

$$\text{Var}(X) = p - p^2 = p(1 - p) = pq,$$

avec la notation habituelle  $q = 1 - p$ .

PROPOSITION 10 — Quelques propriétés

Si  $X$  possède une variance et  $a, b \in \mathbb{R}$ , alors  $aX + b$  possède une variance, avec :

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

*En d'autres termes, les translations ne modifient pas la variance (c'est moral, pour des raisons d'homogénéité), et les multiplications par une constante multiplient la moyenne du carré de l'écart à la moyenne par le carré de cette constante, ce qui est à nouveau moral.*

PREUVE : Simple calcul utilisant les propriétés de linéarité de l'espérance. ■

Quand une variable aléatoire  $X$  possède une variance, alors  $Y = \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}}$  est réduite centrée.

Que dire de la variance d'une somme ?

**Exercice 16.** *On suppose que  $X$  est une variable aléatoire possédant une variance, et que  $Y = X$  (on ne saurait faire moins indépendantes !). Comparer  $\text{Var}(X + Y)$  et  $\text{Var}(X) + \text{Var}(Y)$ .*

Si les variables aléatoires sont indépendantes, la donne est différente.

PROPOSITION 11 — Variance d'une somme de variables indépendantes

Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes possédant une variance, alors  $X + Y$  également, avec de plus :

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

PREUVE : Simple calcul, qu'il serait bon que vous sachiez faire, puis que vous sachiez faire vite... ■

**Exercice 17.** Grâce au résultat précédent, retrouver la variance d'une variable aléatoire suivant une loi binomiale (passer par des variables de Bernoulli).

**Exercice 18.** Un ivrogne qui habite à un kilomètre du bar part dans une grande rue en titubant avec le processus habituel (un mètre à droite ou à gauche, de façon aléatoire). Après 1000 pas, quelle est la probabilité qu'il soit arrivé au but ? Et quelle est la probabilité pour qu'il se soit éloigné de plus de 200 mètres du point de départ ?

**Exercice 19.** Pour établir  $\text{Var}(X + Y + Z) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)$ , de quelle type d'indépendance a-t-on besoin ?

### 5.3 Bienaymé-Tchebychev

Le résultat suivant précise bien le rôle de la variance, qui permet de contrôler explicitement la probabilité de s'éloigner de la moyenne.

THÉORÈME 8 — Inégalité de Bienaymé-Tchebychev

Soit  $X$  une variable aléatoire possédant une variance  $\sigma^2$ , et donc une espérance  $m = \mathbb{E}(X)$ . Si  $\lambda \geq 1$ , alors :

$$\mathbb{P}(|X - m| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Ou encore, si  $\varepsilon \geq \sigma$  :

$$\mathbb{P}(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

PREUVE : Même principe que dans la preuve de l'inégalité de Markov, mais en isolant cette fois dans le calcul de  $\mathbb{E}(|X - m|^2)$  les valeurs de  $X$  telles que  $|X - m| \geq \lambda\sigma$  :

$$\sigma^2 = \mathbb{E}(|X - m|^2) \geq \sum_{|x_i - m| \geq \lambda\sigma} \underbrace{|x_i - m|^2}_{\geq \lambda^2 \sigma^2} \mathbb{P}(X = x_i) \geq \lambda^2 \sigma^2 \sum_{|x_i - m| \geq \lambda\sigma} \mathbb{P}(X = x_i) = \lambda^2 \sigma^2 \mathbb{P}(|X - m| \geq \lambda\sigma).$$

REMARQUE : On peut même prouver Bienaymé-Tchebychev en appliquant directement l'inégalité de Markov à  $(X - \mathbb{E}(X))^2$ ... mais je passe mon temps à oublier Markov, justement. #Alzheimer

**Exercice 20.** Soit  $X$  une variable aléatoire suivant une loi binomiale  $\mathcal{B}(n, p)$  (dont l'espérance et la variance sont respectivement égaux à  $np$  et  $np(1-p)$ ). Avec l'inégalité de Bienaymé-Tchebychev, qu'obtient-on comme majoration de  $\mathbb{P}(|X - np| \geq \frac{np}{2})$  ? Et de  $\mathbb{P}(X \leq \frac{np}{2})$  ? Conclusion ?

Le dernier point de ce paragraphe est dans la zone grise du programme...

**Exercice 21.** CCP 2015

1. Soit  $(Y_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires mutuellement indépendantes, de même loi et admettant un moment d'ordre 2. On pose  $S_n = \sum_{k=1}^n Y_k$ . Prouver :

$$\forall a > 0, \quad \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}(Y_1)\right| \geq a\right) \leq \frac{\text{Var}(Y_1)}{na^2}.$$



2. **Application** : On effectue des tirages successifs – avec remise – d’une boule dans une urne contenant deux boules rouges et trois boules noires.  
 À partir de quel nombre de tirages peut-on garantir à plus de 95% que la proportion de boules rouges obtenues sera comprise entre 0,35 et 0,45 ?

Ainsi, l’inégalité de Bienaymé-Tchebychev nous assure qu’avec une bonne probabilité la moyenne observée va être proche de l’espérance. C’est ce que précise le résultat suivant.

THÉORÈME 9 — *Loi faible des grands nombres*

Si  $(X_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires deux à deux indépendantes, de même loi et admettant un moment d’ordre 2, alors, si  $S_n = \sum_{i=1}^n X_i$  et  $m = \mathbb{E}(X_1)$ , on a :

$$\forall \varepsilon > 0, \quad \mathbb{P} \left( \left| \frac{S_n}{n} - m \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow +\infty} 0.$$

PREUVE : Simple application de Bienaymé-Tchebychev. On notera que pour pouvoir écrire  $\text{Var}(S_n) = n\text{Var}(X_1)$  (et donc  $\text{Var}(S_n/n) = \text{Var}(X_1)/n$ ), on a seulement besoin de l’indépendance **deux à deux**, plus faible que l’indépendance **mutuelle** donnée dans l’exercice précédent. ■

REMARQUES :

- À défaut d’être très performante/optimale, l’inégalité de Bienaymé-Tchebychev nous fournit tout de même un contrôle explicite de cette convergence :

$$\mathbb{P} \left( \left| \frac{1}{n} S_n - m \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

- La *loi forte des grands nombres* (hors programme) nous dit quelque chose de plus fort : avec probabilité 1, la moyenne observée va tendre vers l’espérance.

## 5.4 Covariance

Pour deux variables aléatoires, une façon à la fois élémentaire et forte d’être (informellement) « dépendantes » consiste à ce qu’elles vérifient une relation de la forme  $Y = aX + b$ ; bref, une relation de dépendance affine. Dans ce paragraphe, on va voir un outil permettant de « détecter » une telle dépendance sans connaître les lois des variables, mais simplement quelques espérances. Tout d’abord, on peut centrer les variables en considérant  $Y' = Y - \mathbb{E}(Y)$  et  $X' = X - \mathbb{E}(X)$ , si bien que la relation  $Y = aX + b$  devient  $Y' = aX'$ .

### Exercice 22.

- Montrer que si  $X$  et  $Y$  sont deux variables aléatoires réelles indépendantes, alors :

$$\mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = 0$$

- Que dire de cette quantité si  $Y = aX + b$  ?

L’exercice précédent a mis les deux (trois !) cas extrêmes ; c’est grosso modo ce que dit le résultat suivant pour des variables centrées (d’espérances nulles).

PROPOSITION 12 — *Inégalité de Cauchy-Schwarz*

Soient  $X$  et  $Y$  deux variables aléatoires admettant des variances. Alors  $XY$  possède une espérance finie, et :

$$|\mathbb{E}(XY)| \leq (\mathbb{E}(X^2))^{1/2} (\mathbb{E}(Y^2))^{1/2}.$$

De plus, il y a égalité si et seulement si  $X$  et  $Y$  sont proportionnelles avec probabilité 1, au sens :  $\mathbb{P}(X = 0) = 1$ , ou bien il existe  $a$  tel que  $\mathbb{P}(Y = aX) = 1$ .

PREUVE : Considérer  $\varphi(\lambda) = \mathbb{E}((\lambda X + Y)^2)$ ...

On mesure d’une certaine façon les dépendances affines via le rapport des deux quantités dans l’inégalité de Cauchy-Schwarz (après centrage).

### DÉFINITION 23 — Covariance

Soient  $X$  et  $Y$  deux variables aléatoires admettant des variances. Leur **covariance** est

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

EXEMPLES :

— Si  $X$  et  $Y$  sont indépendantes, alors leur covariance est nulle :

$$\text{Cov}(X, Y) = 0.$$

— Considérons le couple  $(X, Y)$  de loi conjointe donnée par la tableau suivant

$X \backslash Y$	-1	0	1
-1	1/5	0	1/5
1	1/5	1/5	1/5

FIGURE 15 – L'absence de corrélation (covariance nulle) n'exclut pas la dépendance

On vérifie sans mal que  $X$  et  $Y$  ne sont pas indépendantes, mais leur covariance est nulle.

On peut parfois être amené à calculer la variance d'une somme. C'est la somme des variances dans le cas où les variables sont (deux à deux) indépendantes, mais en cas de dépendance, on conserve une formule raisonnablement simple :

### PROPOSITION 13 — Variance d'une somme

Si  $X_1, \dots, X_n$  sont des variables aléatoires possédant des variances finies, alors leur somme aussi, et plus précisément :

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)$$

La preuve est directe en utilisant  $\text{Var}(X) = \text{Cov}(X, X)$ , et la bilinéarité de  $(X, Y) \mapsto \text{Cov}(X, Y)$ .

## 5.5 Fonctions génératrices

On se souvient vaguement (chapitre sur les séries entières ...) que si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{N}$ , alors la série entière  $\sum \mathbb{P}(X = n)t^n$  est de rayon de convergence au moins égal à 1. De plus, pour  $t \in [-1, 1]$ , on peut voir la somme de cette série absolument convergente comme l'espérance de la variable aléatoire  $\omega \mapsto t^{X(\omega)}$ , d'où le raccourci  $\mathbb{E}(t^X)$ .

### DÉFINITION 24 — Fonction génératrice

Si  $X$  est une variable aléatoire à valeurs dans  $\mathbb{N}$ , sa **fonction génératrice** est définie (au moins sur  $[-1, 1]$ ) par :

$$G_X(t) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n = \mathbb{E}(t^X).$$

EXEMPLES :

- Pour les variables à valeurs dans une partie finie de  $\mathbb{N}$ , la série devient un polynôme (et réciproquement, si  $G_X$  est un polynôme, alors il existe un ensemble fini dans lequel  $X$  prend ses valeurs presque sûrement).
- Si  $X \hookrightarrow \mathcal{B}(p)$ , alors  $G_X(t) = (1 - p) + pt$ .
- Si  $X \hookrightarrow \mathcal{B}(n, p)$ , alors  $G_X(t) = ((1 - p) + pt)^n$  (binômisation) ; le lien avec la fonction génératrice d'une Bernoulli va bientôt s'éclaircir...
- Si  $X \hookrightarrow \mathcal{G}(p)$ , alors  $G_X(t) = \frac{pt}{1 - (1 - p)t}$ .
- Si  $X \hookrightarrow \mathcal{P}(\lambda)$ , alors  $G_X(t) = e^{\lambda(t-1)}$ .

Commençons pas régler les questions analytiques ; certaines propriétés ont déjà été montrées ; les autres (bien que vaguement accessibles) seront admises.

PROPOSITION 14 — Propriétés analytiques

Soit  $X$  une variable aléatoire. La fonction génératrice  $G_X$  :

- possède un rayon de convergence  $R \geq 1$  ;
- converge absolument sur  $[-1, 1]$  ;
- est dérivable en 1 si et seulement si  $X$  possède une espérance ; on a alors

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

- est dérivable deux fois en 1 si et seulement si  $X$  possède une variance ; on a alors

$$G''_X(1) = \mathbb{E}(X(X - 1)).$$

REMARQUE : D'une manière générale,  $X$  « possède un moment d'ordre  $k$  » – i.e. :  $X^k$  est d'espérance finie – si et seulement si  $G_X$  est dérivable  $k$  fois en 1.

Les aspects précédents permettent déjà de calculer des espérances ou variances grâce aux outils sur les séries entières, ce qui n'est déjà pas si mal ; mais ce n'est pas le principal ! Le point crucial va résider dans le résultat suivant :

PROPOSITION 15 — Fonction génératrice d'une somme

Si  $X$  et  $Y$  sont deux variables aléatoires réelles **indépendantes** à valeurs dans  $\mathbb{N}$ , alors la fonction génératrice de leur somme est le produit de leurs fonctions génératrices.

$$G_{X+Y} = G_X G_Y.$$

PREUVE : Déjà faite dans le cours sur les séries entières. N'allez pas chercher ce fichu cours : faites la preuve vous-même ! ■

Bien entendu, ce résultat s'étend à la somme de  $n$  variables aléatoires **mutuellement indépendantes** (pas seulement deux à deux).

**Exercice 23.** *Montrer que la somme de deux variables aléatoires suivant des lois de Poisson... suit elle-même une loi de Poisson.*

SOLUTION : Le point trivial est le calcul : il a déjà été fait ! Une fois qu'on a  $G_{X+Y}(t) = e^{(\lambda_1 + \lambda_2)(t-1)}$ , il reste à remonter à la loi de  $Z = X + Y$ . Et là, on utilise la récupération des coefficients d'une série entière : la donnée de  $G_Z(t) = \sum_{n=0}^{+\infty} \alpha_n t^n$  pour  $t$  au voisinage de 0 impose les valeurs des  $\alpha_n$ . Puisque  $G_Z(t) = G_W(t)$ , où  $W$  suit une loi de Poisson  $\mathbb{P}(\lambda_1 + \lambda_2)$ , on a bien pour tout  $n \in \mathbb{N}$  :  $\mathbb{P}(Z = n) = \mathbb{P}(W = n) = \frac{e^{-\lambda} \lambda^n}{n!}$ .

## 5.6 Retour sur les lois usuelles

Quelles lois correspondent à quelles modélisation de la vie réelle ?

- **Les lois de Bernoulli** modélisent des expériences binaires (typiquement : pile/face), à résultat dans  $\{0, 1\}$ . En les sommant, on peut compter le nombre de « succès » de telles expériences. Il y a une petite variante :  $X$  suit une **loi de Rademacher** lorsque  $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = 1/2$ . C'est typiquement ce qu'on utilise pour modéliser une marche d'ivrogne. On pourra alors voir  $X$  comme égale à  $2Y - 1$ , avec  $Y \hookrightarrow \mathcal{B}(1/2)$ . Ce n'est pas fulgurant, mais après sommation, ce genre de bricolage peut devenir intéressant et nous épargner quelques calculs (typiquement pour des espérances et variances).
- **La loi binomiale** est celle qu'on obtient qu'on on répète une expérience de Bernoulli et qu'on compte le nombre de succès : nombre de PILE obtenus lors de  $n$  lancers **indépendants**.
- **La loi uniforme** sera typiquement celle qui s'appliquera à un lancer de dé honnête, ou un choix de boule dans une urne.

- **La loi géométrique** est naturelle dans des contextes de processus sans mémoire, où on s'intéresse au premier temps d'arrivée d'un événement : à l'instant  $t = 1$ , l'événement peut se produire. Et s'il ne se produit pas, « les compteurs sont remis à zéros » :  $\mathbb{P}(X = 2|X > 1) = \mathbb{P}(X = 1)$  ; ou encore (en passant aux complémentaires) :  $\mathbb{P}(X > 2|X > 1) = \mathbb{P}(X > 1)$ . Voir l'exercice et la proposition qui suivent.
- **La loi de Poisson** est fondamentalement un cas limite de loi binomiale. Elle modélise des événements affectant une grande population : chaque individu est rarement affecté, mais à l'échelle macroscopique, l'événement n'est pas si rare. Exemples typiques :
  - le nombre (par seconde) de désintégrations radioactives dans une grande population d'atomes ;
  - le nombre d'accès journalier à un site où on trouve d'excellents polys, chez une population assez grande, mais dont chaque individu se connecte rarement ;
  - le nombre journalier de pannes de processeurs plutôt fiables dans l'ensemble des google-centers de la planète ;
  - le nombre journalier d'accidents induisant de la tôle froissée déclarés à une compagnie d'assurance ;
  - etc.

On parle parfois de « loi des événements rares ». La formule a priori étrange  $\mathbb{P}(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$  sera expliquée dans la suite.

**Exercice 24.** On considère un dé biaisé renvoyant *PILE* avec probabilité  $p \in ]0, 1[$ .

1. Quelle est la probabilité pour que le premier *PILE* arrive au troisième lancer ? Et au  $k$ -ième ?
2. Quelle est la probabilité pour qu'il n'y ait pas de *PILE* lors des  $k$  premiers lancers ? Et celle pour qu'il n'y ait pas de *PILE* lors des  $n + k$  premiers lancers sachant qu'il n'y en a pas eu lors des  $n$  premiers lancers ?

C'est ce phénomène qu'on appelle « processus sans mémoire » – et qui est contre-intuitif pour les gens de la rue (non, le fait qu'un nombre ne soit pas sorti à la roulette depuis 100 tirages n'augmente ni ne diminue ses chances de sortie au tirage suivant...)!

On résume ici différents résultats. Il est crucial de connaître les lois, espérances... et si possible la variance !

Loi	Notation	$X(\Omega)$	$\mathbb{P}(X = k)$	Espérance	Variance	$G_X(t)$
Uniforme	$\mathcal{U}([1, n])$	$[1, n]$	$1/n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{t(1-t^n)}{n(1-t)}$
Bernoulli	$\mathcal{B}(p)$	$\{0, 1\}$	$\begin{cases} p & \text{si } k = 1 \\ 1 - p & \text{si } k = 0 \end{cases}$	$p$	$p(1 - p)$	$1 - p + pt$
Binomiale	$\mathcal{B}(n, p)$	$[0, n]$	$\binom{n}{k} p^k (1 - p)^{n-k}$	$np$	$np(1 - p)$	$(1 - p + pt)^n$
Géométrique	$\mathcal{G}(p)$	$\mathbb{N}^*$	$p(1 - p)^{k-1} \ (k \geq 1)$	$1/p$	$(1 - p)/p^2$	$\frac{pt}{1 - (1 - p)t}$
Poisson	$\mathcal{P}(\lambda)$	$\mathbb{N}$	$\frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda$	$\lambda$	$e^{\lambda(t-1)}$

**Exercice 25.** Déterminer l'espérance et la variance d'une variable suivant une loi uniforme  $\mathcal{U}([a, b])$ .