

# VARIABLES ALÉATOIRES

## DÉNOMBRABLES

« C'était une foule immense que nul ne pouvait dénombrer,  
de toutes nations, tribus, peuples et langues. »  
Livre de l'Apocalypse, 7,9

### 1 ENSEMBLE DÉNOMBRABLE

**Avis au lecteur pressé :** Cette partie est assez développée. L'étudiant, pas très à l'aise avec le formalisme mathématique ou n'ayant pas beaucoup de temps à consacrer à la théorie, pourra se contenter de la première explication ci-dessous et sauter les définitions et propriétés qui suivent.

#### Explications

Un ensemble **dénombrable** est un ensemble infini, dont on peut « numéroter » les éléments sous la forme d'une suite  $x_0, x_1, \dots, x_n, \dots$  indicée par  $\mathbf{N}$  ou éventuellement par  $\mathbf{Z} : \dots, x_{-2}, x_{-1}, x_0, x_1, \dots$ .

Il existe des ensembles infinis non dénombrables. Intuitivement, ces ensembles non dénombrables ont « beaucoup plus d'éléments » que les ensembles dénombrables : ils sont trop touffus pour que l'on puisse compter les éléments un à un.  $\mathbf{R}$  est un ensemble non dénombrable.

**Pour ceux qui veulent approfondir :**

#### — Définition 1.1 (Ensemble dénombrable) —

Un ensemble  $E$  est **dénombrable** s'il peut être mis en bijection avec  $\mathbf{N}$ .  
Une ensemble est **au plus dénombrable**, s'il est fini ou dénombrable.

#### — Propriété 1.2 (Ensembles dénombrables de référence) —

1. Un ensemble fini n'est pas dénombrable.
2.  $\mathbf{N}$  et tous ses sous-ensembles infinis sont dénombrables.
3.  $\mathbf{Z}$  et  $\mathbf{Q}$  sont dénombrables.

#### Preuve

1. Soit  $E$  un ensemble fini. Si par l'absurde, il était dénombrable, alors, on pourrait construire une bijection  $\varphi : E \rightarrow \mathbf{N}$ . Donc  $\varphi(E)$  est une partie non vide de  $\mathbf{N}$ .  
 $E$  étant finie,  $\varphi(E)$  aussi et elle contient un plus grand élément  $n$ .  
Donc  $n + 1 \notin \varphi(E)$ . Donc  $\varphi$  n'est pas surjective. Absurde.  
Donc  $E$  n'est pas dénombrable.
2. Si  $E$  est un sous-ensemble de  $\mathbf{N}$  infini, alors, on peut construire une bijection de  $\mathbf{N} \rightarrow E$ , en posant  $\varphi(0) = \min E$  et  $\forall n \in \mathbf{N}, \varphi(n + 1) = \min \{k \in E, k > \varphi(n)\}$ .  
*Cette construction consiste simplement à énumérer les éléments de  $E$  un à un dans l'ordre du plus petit au plus grand : à chaque fois, on rajoute le plus petit élément que l'on n'a pas encore compté.*  
 $E$  étant infini,  $\varphi$  est bien définie.  
 $\varphi$  est injective (strictement croissante).  
Montrons par l'absurde qu'elle est surjective. *Cette preuve est très simple à condition de bien la visualiser.*  
L'idée est donc de supposer par l'absurde qu'elle n'est pas surjective et de prendre le plus petit élément  $p$  qui n'appartient pas à son image :
  - Soit  $c$  est le premier élément de  $E$  : impossible car on l'a utilisé pour  $\varphi(0)$ .

- Sinon, l'élément qui le précède dans  $E$  peut s'écrire  $k = \varphi(n)$  et par construction,  $p$  devrait être  $\varphi(n + 1)$  puisque c'est le suivant.

Donc  $\varphi$  est surjective. Ainsi c'est une bijection entre  $\mathbf{N}$  et  $E$  :  $E$  est dénombrable. Il en est de même pour tout sous-ensemble infini d'un ensemble dénombrable.

3. Pour  $\mathbf{Z}$ , on pose  $\varphi$  tel que  $\varphi(n) = n/2$  si  $n$  est pair, et  $\varphi(n) = -\frac{n+1}{2}$  si  $n$  est impair. On montre (exercice facile) que  $\varphi$  est bijective de  $\mathbf{N}$  dans  $\mathbf{Z}$  :  $\mathbf{Z}$  est dénombrable. La dénombrabilité de  $\mathbf{Q}$  s'obtient avec le théorème suivant (produit cartésien) car  $\mathbf{Q}$  est en bijection avec  $\mathbf{Z} \times \mathbf{N}^*$ .

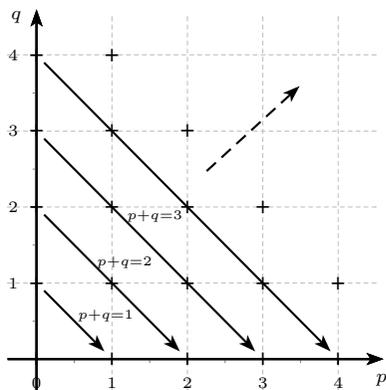
**Propriété 1.3 (Produit cartésien)**

Si  $E$  et  $F$  sont dénombrables, alors  $E \times F$  est dénombrable.  
 Si  $E$  est un ensemble dénombrable, alors  $\forall n \in \mathbf{N}^*$ ,  $E^n$  est dénombrable.

Remarque : c'est faux en général pour un produit cartésien infini.

**Preuve**

On le montre déjà dans le cas particulier où  $E = F = \mathbf{N}$ . La généralisation sera très simple ensuite. L'idée est de compter tous les couples  $(p, q)$  du carré infini  $\mathbf{N}^2$ . Cela suppose de définir un ordre dans lequel on les compte tous, une et une fois. Une bonne méthode est de les compter en diagonale depuis  $(0, 0)$  et en s'écartant de plus en plus.



Chaque diagonale correspond à  $p + q = n = \text{cste}$  (quand on augmente de 1 la valeur de  $p$  vers la droite, on diminue de 1 celle de  $q$  vers le bas). Et la diagonale  $n$  contient exactement  $n + 1$  points (aller de  $p = 0, q = n$  à  $p = n$  et  $q = 0$ ). Ainsi, pour que la numérotation d'une diagonale à l'autre se suive sans trou ni chevauchement, on commence la diagonale  $n$  par la valeur  $d_n = d_{n-1} + n$ . Comme la numérotation commence à 0, on pose  $d_0 = 0$ . Ainsi,  $\forall n \in \mathbf{N}$ ,  $d_n = \frac{n(n+1)}{2} = \frac{(p+q)(p+q+1)}{2}$ . Ensuite, quand on avance sur la diagonale, on part de  $p = 0$  à  $p = n$  : on ajoute la valeur de  $p$  pour indiquer le numéro du point sur la diagonale.

$$\frac{(p + q)(p + q + 1)}{2} + p$$

On définit donc l'application  $\psi : \begin{cases} \mathbf{N}^2 & \rightarrow \mathbf{N} \\ (p, q) & \mapsto \frac{(p+q)(p+q+1)}{2} + p \end{cases}$  qui est bijective par construction. Ainsi  $\mathbf{N}^2$  est dénombrable. Passons au cas général de  $E$  et  $F$  dénombrables. Il existe deux bijections  $\varphi_1 : \mathbf{N} \rightarrow E$  et  $\varphi_2 : \mathbf{N} \rightarrow F$ . Ainsi, l'application  $\Phi : (p, q) \mapsto (\varphi_1(p), \varphi_2(q))$  est une bijection de  $\mathbf{N}^2$  dans  $E \times F$ .  $\mathbf{N}^2$  étant dénombrable,  $E \times F$  est dénombrable. Pour  $E^n$ , on procède par récurrence.

**Propriété 1.4**

Si  $E$  contient un sous-ensemble infini non dénombrable, alors  $E$  n'est pas dénombrable.

**Explications**

Un ensemble n'est pas dénombrable, s'il est soit « trop petit » (c'est-à-dire fini) ou s'il est « trop gros » (on ne peut pas énumérer ses éléments un à un). On comprend donc que si  $E$  contient un ensemble « trop gros », alors il est lui-même trop gros pour être dénombré.

**Preuve**

C'est la contraposée de : « si  $E$  est dénombrable, alors tout sous ensemble infini de  $E$  est dénombrable ». On l'a démontré pour  $E = \mathbf{N}$ , mais par bijectivité entre  $E$  dénombrable et  $\mathbf{N}$ , la preuve précédente peut aussi s'appliquer à tout autre ensemble dénombrable.

**Propriété 1.5**

Une union dénombrable d'ensembles dénombrables est dénombrable.

**Preuve**

On note  $(A_i)_{i \in \mathbf{N}}$  une famille dénombrable d'ensembles dénombrables, et  $A = \bigcup_{i \in \mathbf{N}} A_i$ . Dans un premier temps, on suppose que tous les  $A_i$  sont disjoints. Dans ce cas, pour chaque  $i \in \mathbf{N}$ , on considère la bijection  $\varphi_i : \mathbf{N} \rightarrow A_i$ . On peut donc construire une bijection :  $\varphi : \begin{cases} \mathbf{N}^2 & \rightarrow A \\ (i, j) & \mapsto \varphi_i(j) \end{cases}$ .  $\varphi$  est évidemment surjective par surjectivité des  $\varphi_i$  et elle est injective car les applications  $\varphi_i$  le sont et les  $A_i$  disjoints. Donc  $A$  est en bijection avec  $\mathbf{N}^2$  qui est dénombrable (propriété 1.3), donc  $A$  est dénombrable. Si l'union n'est pas disjointe, alors on sait tout d'abord que  $A$  ne peut pas être fini car il contient un ensemble  $A_0$  qui est dénombrable. Ainsi  $A$  est au moins dénombrable. On pose alors  $B_0 = A_0$  et pour chaque  $i \geq 1$ ,  $B_i = A_i \setminus B_{i-1}$ . On vérifie ainsi facilement par récurrence que  $A = \bigcup_{i \in \mathbf{N}} B_i$  et que les  $B_i$  sont deux à deux disjoints. Quitte à rajouter des éléments dans chaque  $B_i$  qui serait fini, on peut définir une union disjointe :  $\tilde{A} = \bigcup_{i \in \mathbf{N}} \tilde{B}_i$  avec  $\forall i \in \mathbf{N}$ ,  $B_i \subset \tilde{B}_i$  et  $\tilde{B}_i$  dénombrable. Donc  $A \subset \tilde{A}$  qui est lui-même dénombrable d'après la première étude. Donc  $A$  est au plus dénombrable (propriété 1.4), donc  $A$  est dénombrable.

**Propriété 1.6**

Tout intervalle de  $\mathbf{R}$  est non dénombrable.  
En particulier,  $\mathbf{R}$  n'est pas dénombrable.

**Explications**

Sur  $\mathbf{R}$ , les nombres sont trop « serrés » les uns contre les autres pour que l'on puisse les dénombrer.

$\mathbf{Q}$  peut être dénombré, mais l'ajout des irrationnels rend l'ensemble trop « touffu » pour qu'il reste dénombrable.

Ce caractère « touffu » de  $\mathbf{R}$  ne dépend pas de son étendue : même dans un petit intervalle, il y a trop de nombres pour pouvoir les isoler un à un en les dénombrant.

**Preuve**

On commence par montrer que  $[0, 1[$  n'est pas dénombrable, puis que tout intervalle contenant au moins de deux points peut être mis en bijection avec  $[0, 1[$ .

*Cette deuxième étape revient simplement à considérer  $[0, 1[$  comme un élastique que l'on comprimerait ou étirerait pour obtenir n'importe quel autre intervalle de la même forme.*

*Remarque :* Si  $I$  ne contient pas au moins deux points distincts, alors il est fini, donc non dénombrable.

1. La non dénombrabilité de  $[0, 1[$  peut être prouvée avec l'argument de la diagonale de Cantor. Le voici :

$[0, 1[$  contient tous les nombres qui s'écrivent sous forme décimale  $0, x x x x \dots$  où la suite des décimales n'est pas stationnaire à 9.

Si par l'absurde  $[0, 1[$  était dénombrable, alors il existerait une bijection  $\varphi : \mathbf{N} \rightarrow \mathbf{R}$ . et on pourrait écrire les nombres dans l'ordre donné par la bijection  $\varphi(0), \varphi(1) \dots$

On construit alors le nombre  $x$  dont la  $n$ -ième décimale est égale à 1 la  $n$ -ième décimale de  $\varphi(n)$  est différente de 1 et 2 sinon.

Ainsi  $x \in [0, 1[$ , montrons que  $x \notin \varphi(\mathbf{N})$ .

Si par l'absurde  $x \in \varphi(\mathbf{N})$ , alors il existerait  $k \in \mathbf{N}$  tel que  $x = \varphi(k)$ .

En particulier, la  $k$ -ième décimale de  $x$  serait celle de  $\varphi(k)$  ce qui est impossible par construction de  $x$ . Donc  $x \notin \varphi(\mathbf{N})$ .

Donc  $\varphi$  n'est pas surjective, donc non bijective : c'est absurde.  $[0, 1[$  n'est pas dénombrable.

2. L'idée est de se ramener à un intervalle de la même forme que  $[0, 1[$  (semi-ouvert à droite) pour construire la bijection.

Soit  $I$  un intervalle de  $\mathbf{R}$  contenant au moins 2 points  $a < b$ . Il contient donc  $[a, b[$ .

On va montrer que  $[a, b[$  n'est pas dénombrable en le mettant en bijection avec  $[0, 1[$ .

On construit

$$f : \begin{cases} [0, 1[ & \mapsto [a, b[ \\ x & \mapsto (b-a)x + a \end{cases}$$

Comme  $b > a$ ,  $f$  réalise bien une bijection de  $[0, 1[$  dans  $[a, b[$ .

Ainsi  $[a, b[$  n'est pas dénombrable.

(sinon, on composerait  $\varphi : [a, b[ \rightarrow \mathbf{N}$  avec  $f$  et on obtiendrait une bijection entre  $[0, 1[$  et  $\mathbf{N}$ , ce qui n'est pas possible car  $[0, 1[$  n'est pas bijectif).

Donc  $I$  contient un sous-ensemble infini non dénombrable.

Donc  $I$  n'est pas dénombrable. ■

**2 ESPACE PROBABILISÉ**

Le fait de se trouver avec un univers dénombrable rend les définitions du cas fini insuffisantes pour s'assurer du bon fonctionnement de la théorie.

Nous allons donc reprendre les différentes définitions pour les adapter au cas dénombrable.

Il s'agit d'une généralisation : les nouvelles définitions que nous allons voir « engloberont » celles du cas fini qui n'en étaient que des cas particuliers.

Ainsi, tout ce que nous avons vu dans le cas fini reste bien sûr vrai... dans le cas fini.

Mais ce que nous voyons ici, sera vrai à la fois pour le cas dénombrable, mais aussi pour le cas fini.

L'an prochain, l'étude des probabilités sur  $\mathbf{R}$  (qui n'est pas dénombrable) demandera de passer au niveau encore supérieur de généralisation.

**A Événements****Notation**

Dans ce chapitre  $\Omega$  désigne un ensemble appelé **univers**.

**Avis au lecteur pressé :** comme pour la dénombrabilité, il est possible de sauter cette section en sachant simplement que :

- on note  $\mathcal{E}$  l'ensemble des événements sur  $\Omega$ .
- « tout ce qu'on pourra raisonnablement imaginer » sera un événement de  $\mathcal{E}$ .
- les unions et intersections (dénombrables) d'événements sont des événements.
- le complémentaire d'un événement est aussi un événement.

**Pour ceux qui veulent approfondir :**

Avant de définir une fonction de probabilité, il faut savoir ce qu'est un événement.

Dans le cas fini, on a dit que les événements étaient toutes les parties de  $\Omega$ . Il est possible de conserver une telle définition pour un univers dénombrable.

Par contre, lorsque l'univers n'est plus dénombrable, une telle définition ne permet plus de définir des probabilités telles que le souhaiterait l'intuition<sup>1</sup>.

Il faut donc repartir de zéro et comme à chaque fois que l'on veut construire une théorie mathématique qui doit rendre compte d'une intuition préalable, on est amené à se poser la question des propriétés minimales essentielles que doit vérifier notre objet.

1. Ainsi, avec cette définition un peu simpliste des événements, il ne serait plus possible de définir la loi normale sur  $\mathbf{R}$  ou même simplement la loi uniforme sur  $[0, 1[$  (pour des raisons théoriques difficiles).

Dans notre cas, on souhaite bien entendu

1. que  $\Omega$  soit un événement, ainsi que l'ensemble vide.
2. que l'événement contraire existe : que le complémentaire d'un événement soit encore un événement.
3. que l'union ou l'intersection de deux événements en soit encore un. Et plus généralement, que l'on puisse réaliser une union ou une intersection dénombrable d'événements pour travailler avec des suites d'événements.

Nous n'avons besoin de rien d'autre.

On appellera **tribu** tout ensemble de partie de  $\Omega$  qui vérifie ces propriétés et on exigera que l'ensemble des événements  $\mathcal{E}$  forme une tribu sur  $\Omega$ .

**Définition 2.1** (*Tribu*)

Une **tribu**  $\mathcal{E}$  sur  $\Omega$  est un ensemble de parties de  $\Omega$  qui vérifie :

1.  $\Omega \in \mathcal{E}$ .
2.  $A \in \mathcal{E} \Rightarrow \bar{A} \in \mathcal{E}$ .
3.  $\forall (A_n)_{n \in \mathbf{N}} \in \mathcal{E}^{\mathbf{N}}, \bigcup_{n \in \mathbf{N}} A_n \in \mathcal{E}$ .

Il n'est pas utile d'exiger que  $\emptyset \in \mathcal{E}$  et que  $\mathcal{E}$  soit stable par intersection car cela découle directement du passage au complémentaire :

**Propriété 2.2**

Si  $\mathcal{E}$  est une tribu sur  $\Omega$ , alors

4.  $\emptyset \in \mathcal{E}$ .
5.  $\forall (A_n)_{n \in \mathbf{N}} \in \mathcal{E}^{\mathbf{N}}, \bigcap_{n \in \mathbf{N}} A_n \in \mathcal{E}$ .

**Preuve**

4.  $\Omega \in \mathcal{E}$  donc  $\bar{\Omega} \in \mathcal{E}$ , donc  $\emptyset \in \mathcal{E}$ .
5. Soit  $(A_n)_{n \in \mathbf{N}}$  une suite d'événements de  $\mathcal{E}$ .  
Par passage au complémentaire :  $\forall n \in \mathbf{N}, \bar{A}_n \in \mathcal{E}$ ,  
Donc, par stabilité par l'union dénombrable :  $\bigcup_{n \in \mathbf{N}} \bar{A}_n \in \mathcal{E}$ .  
Or  $\mathcal{E}$  est stable par passage au complémentaire, donc  $\bigcap_{n \in \mathbf{N}} A_n = \overline{\bigcup_{n \in \mathbf{N}} \bar{A}_n} \in \mathcal{E}$ .

**Théorème 2.3**

$\mathcal{P}(\Omega)$  est une tribu sur  $\Omega$ . On l'appelle la **tribu triviale** sur  $\Omega$ .

**Preuve**

On vérifie la définition : sans la moindre difficulté. ■

C'est la tribu qu'on utilise pour un univers fini et qu'on peut également prendre pour un univers dénombrable.

Par contre, bien que ce soit également une tribu pour  $\mathbf{R}$ , son usage pose des problèmes avec les probabilités continues et on utilise alors d'autres tribus un peu moins riches (avec moins d'événements) mais qui contiennent tous les événements « intéressants » tout en offrant davantage de souplesse pour définir des probabilités.

On laissera ces difficultés théoriques pour les probabilistes.

**Définition 2.4** (*Espace probabilisable*)

Un ensemble  $\Omega$  muni d'une tribu  $\mathcal{E}$  est un **espace probabilisable**.

En d'autres termes, maintenant que nous savons ce qu'est un événement, nous sommes prêts pour parler des probabilités.

**B Probabilité**

**Définition 2.5** (*Probabilité sur  $\Omega$* )

Soit  $(\Omega, \mathcal{E})$  un espace probabilisable (quelconque).

Une probabilité  $\mathbf{P}$  sur  $(\Omega, \mathcal{E})$  est une application  $\mathbf{P} : \mathcal{E} \rightarrow \mathbf{R}$  telle que

1.  $\forall A \in \mathcal{E}, \mathbf{P}(A) \in [0, 1]$ .
2.  $\mathbf{P}(\Omega) = 1$ .
3.  $\forall (A_n)_{n \in \mathbf{N}} \in \mathcal{E}^{\mathbf{N}}$ , deux à deux **disjoints**,

$$\mathbf{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathbf{P}(A_n).$$

Ces axiomes correspondent exactement à ceux des probabilités finies, si ce n'est que l'on a rajouté remplacé l'union finie par l'union dénombrable.

Ce dernier point mérite notre attention.

En effet, on remarque que  $\bigcup_{n=0}^{+\infty} A_n$  est un événement (comme union dénombrable d'événements), et qu'il possède donc à ce titre une probabilité dans  $[0, 1]$ .

Ainsi, la série à termes positifs  $\sum \mathbf{P}(A_n)$  converge et admet une limite dans  $[0, 1]$  :

Toute série de probabilités d'événements deux à deux disjoints converge dans  $[0, 1]$ .

Toutes les propriétés vues dans le cas fini restent vraies (croissance, inégalité entre la probabilité de l'union et les sommes des probabilités...).

**Théorème 2.6** (*Théorèmes des limites monotones*)

Le troisième axiome de la définition 2.5 est équivalent à chacun des énoncés suivants :

3'. *Théorème de la limite monotone croissante :*

$\forall (B_n)_{n \in \mathbf{N}} \in \mathcal{E}^{\mathbf{N}}$ , tel que  $\forall k \in \mathbf{N}$ ,  $B_k \subset B_{k+1}$  (suite croissante),

$$\mathbf{P} \left( \bigcup_{n=0}^{+\infty} B_n \right) = \lim_{n \rightarrow +\infty} \mathbf{P}(B_n).$$

3''. *Théorème de la limite monotone décroissante :*

$\forall (B_n)_{n \in \mathbf{N}} \in \mathcal{E}^{\mathbf{N}}$ , tel que  $\forall k \in \mathbf{N}$ ,  $B_{k+1} \subset B_k$  (suite décroissante),

$$\mathbf{P} \left( \bigcap_{n=0}^{+\infty} B_n \right) = \lim_{n \rightarrow +\infty} \mathbf{P}(B_n).$$

Pour 3', la suite des événements est croissante donc  $\forall n \in \mathbf{N}$ ,  $\bigcup_{k=0}^n B_k = B_n$ .

De même, pour 3'',  $\forall n \in \mathbf{N}$ ,  $\bigcap_{k=0}^n B_k = B_n$ .

**Preuve**

*La preuve peut être sautée en première lecture.*

(3  $\Rightarrow$  3') On définit la suite d'événements  $A_0 = B_0$  et pour tout  $k \geq 1$ ,  $A_k = B_k \setminus B_{k-1}$ .

Ainsi, on peut montrer par récurrence que  $\forall n \in \mathbf{N}$ ,  $\bigcup_{k=0}^n A_k = \bigcup_{k=0}^n B_k$  et que les  $A_k$  sont deux à deux incompatibles. Donc l'axiome 3 s'applique à la famille  $(A_k)_{k \in \mathbf{N}}$  et on trouve que

$$\forall n \in \mathbf{N}, \mathbf{P} \left( \bigcup_{n=0}^{+\infty} B_n \right) = \sum_{n=0}^{+\infty} \mathbf{P}(A_n).$$

$$\text{Ainsi } \sum_{n=0}^{+\infty} \mathbf{P}(A_n) = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \mathbf{P}(A_k) = \lim_{n \rightarrow +\infty} \mathbf{P} \left( \bigcup_{k=0}^n A_k \right) = \lim_{n \rightarrow +\infty} \mathbf{P}(B_n).$$

(3'  $\Rightarrow$  3) Soit  $(A_n)_{n \in \mathbf{N}}$  est une famille d'événements deux à deux disjoints, On peut construire la famille  $(B_n)_{n \in \mathbf{N}}$ , avec  $B_0 = A_0$  et  $\forall n \geq 1$ ,  $B_n = B_{n-1} \cup A_n$ .

Ainsi, on démontre immédiatement que la suite  $(B_n)_{n \in \mathbf{N}}$  est croissante et que

$\forall n \in \mathbf{N}$ ,  $\bigcup_{k=0}^n B_k = \bigcup_{k=0}^n A_k$ . Donc, l'axiome 3' donne :

$$\mathbf{P} \left( \bigcup_{n=0}^{+\infty} A_n \right) = \mathbf{P} \left( \bigcup_{n=0}^{+\infty} B_n \right) = \lim_{n \rightarrow +\infty} \mathbf{P}(B_n) = \lim_{n \rightarrow +\infty} \mathbf{P} \left( \bigcup_{k=0}^n A_k \right)$$

$$= \lim_{n \rightarrow +\infty} \sum_{k=0}^n \mathbf{P}(A_k) = \sum_{n=0}^{+\infty} \mathbf{P}(A_n).$$

Pour 3'', on passe par le complémentaire. ■

**C Probabilités conditionnelles**

Les probabilités conditionnelles vues dans le cas fini restent valables et s'expriment de la même façon. Seule la formule des probabilités totales doit être reformulée pour prendre en compte le cas d'une famille complète contenant un nombre infini dénombrable d'événements.

**Définition 2.7** (*Famille quasi-complète d'événements*)

Une **famille complète d'événements** dans  $\Omega$  est une famille finie ou dénombrables d'événements  $(B_i)_{i \in I}$  telle que

1. Les événements  $B_i$  sont deux à deux incompatibles,
2.  $\bigcup_{i \in I} B_i = \Omega$ .

Une **famille quasi-complète d'événements** dans  $\Omega$  est une famille finie ou dénombrables d'événements  $(B_i)_{i \in I}$  telle que

1. Les événements  $B_i$  sont deux à deux incompatibles,
2.  $\sum_{i \in I} \mathbf{P}(B_i) = 1$ .

Concrètement, on pourra travailler de la même façon avec les familles quasi-complètes qu'avec les familles complètes. Les familles quasi-complètes se permettent simplement de ne pas considérer une partie négligeable de  $\Omega$ .

**Théorème 2.8** (*Formule des probabilités totales*)

Soit  $A$  un événement de  $(\Omega, \mathcal{E}, \mathbf{P})$ , et  $(B_i)_{i \in \mathbf{N}}$  une famille (quasi-)complète d'événements (de probabilités non nulles)

$$\mathbf{P}(A) = \sum_{i=0}^{+\infty} \mathbf{P}_{B_i}(A) \mathbf{P}(B_i).$$

(en particulier, la série converge)

**Théorème 2.9** (*Formule de Bayes*)

Soit  $A$  un événement de  $(\Omega, \mathcal{E}, \mathbf{P})$  de probabilité non nulle, et  $(B_i)_{i \in \mathbf{N}}$  une famille (quasi-)complète d'événements (de probabilités non nulles)

$$\mathbf{P}_A(B_k) = \frac{\mathbf{P}_{B_k}(A) \mathbf{P}(B_k)}{\mathbf{P}(A)} = \frac{\mathbf{P}_{B_k}(A) \mathbf{P}(B_k)}{\sum_{i=0}^{+\infty} \mathbf{P}_{B_i}(A) \mathbf{P}(B_i)}.$$

**Preuve**

Ce sont (presque) les mêmes preuves que dans le cas fini. Pour les systèmes quasi-complets, on rajoute l'événement négligeable qui donne une probabilité nulle. ■

### 3 VARIABLES ALÉATOIRES DISCRÈTES

Les variables aléatoires discrètes désignent à la fois les variables aléatoires finies et dénombrables.

**Définition 3.1** (*Variable aléatoire*)

Une **variable aléatoire réelle discrète** est une application  $X : \mathcal{E} \rightarrow \mathcal{X}$  telle que  $\mathcal{X}$  soit une partie finie ou dénombrable de  $\mathbf{R}$ .  
On note  $\mathcal{X} = \{x_i, i \in I\}$  avec  $I = \llbracket 1, n \rrbracket$  pour une variable aléatoire finie, et  $I = \mathbf{N}$  ou  $I = \mathbf{Z}$  pour une variable aléatoire dénombrable.

*Remarque hors programme* : Une variable aléatoire discrète ne peut pas être infinie indénombrable. En effet, si on considère  $A_n = \{x \in \mathbf{R}, \mathbf{P}(x) \geq \frac{1}{n}\}$ , alors  $A_n$  contient au plus  $n$  éléments (pour que la somme des probabilités ne dépasse pas 1).  
Or,  $\mathcal{X} = \bigcup_{n \in \mathbf{N}} A_n$  est une union dénombrable d'ensembles au plus dénombrables :  $\mathcal{X}$  est donc fini ou dénombrable.

**Définition 3.2** (*Loi d'une variable*)

Soit  $X$  une variable aléatoire réelle discrète, avec  $\mathcal{X} = \{x_i, i \in I\}$ .  
Soit  $\mathbf{P}$  une probabilité sur  $\Omega$ ,  
  
La **loi de  $X$**  est la fonction  $\mathbf{P}_X : x \mapsto \mathbf{P}(X = x)$ .

On utilise les mêmes notations pour les événements qu'avec les variables aléatoires finies. On note :

**Propriété 3.3**

Soit  $X$  une variable aléatoire réelle discrète, avec  $\mathcal{X} = \{x_i, i \in I\}$ .  
  
 $\{[X = x_i], i \in I\}$  forme une famille complète d'événements.  
  
On a noté  $[X = x_i] = X^{-1}(\{x_i\})$ .

En particulier, on peut utiliser cette famille pour la formule des probabilités totales.

**Corollaire 3.4**

Soit  $X$  une variable aléatoire réelle discrète, avec  $\mathcal{X} = \{x_i, i \in I\}$ .  
Soit  $\mathbf{P}$  une probabilité sur  $\Omega$ ,  
  
$$\sum_{i \in I} \mathbf{P}(X = x_i) = 1.$$
  
  
Pour toute partie  $\mathcal{X}' \subset \mathcal{X}$ ,  
  
$$\mathbf{P}(X \in \mathcal{X}') = \sum_{x \in \mathcal{X}'} \mathbf{P}(X = x).$$

### 4 LOIS USUELLES

#### A Loi géométrique

**Exemple** (*Exemple type : rang du premier succès*)

On lance un dé (équilibré) à 6 faces une infinité de fois de façon indépendante.  
On note  $T$  le numéro du lancer correspondant au premier « six ».

$T$  suit alors la loi géométrique de paramètre  $\frac{1}{6}$ .

Le paramètre  $\frac{1}{6}$  est la probabilité de l'événement «succès» : avoir un « six ».

**Déterminons la loi de  $T$ .**

Le premier « six » peut apparaître à n'importe quel rang à partir du premier lancer. Ainsi  $T(\Omega) = \mathbf{N}^*$  (voir remarque plus loin).  
Si on note  $S_k$  l'événement : « le  $k$ -ième lancer est un « six », alors

$$\forall k \geq 1, [T = k] = \overline{S_1} \cap \overline{S_2} \cap \dots \cap \overline{S_{k-1}} \cap S_k$$

Les  $k - 1$  premiers lancers ont donné un échec : entre 1 et 5, et le  $k$ -ième lancer a donné un succès.

Les lancers sont indépendants donc :

$$\begin{aligned} \mathbf{P}([T = k]) &= \mathbf{P}(\overline{S_1} \cap \overline{S_2} \cap \dots \cap \overline{S_{k-1}} \cap S_k) \\ &= \mathbf{P}(\overline{S_1}) \times \mathbf{P}(\overline{S_2}) \times \dots \times \mathbf{P}(\overline{S_{k-1}}) \times \mathbf{P}(S_k) \\ &= \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6}. \end{aligned}$$

*Remarque* : Il y a un passage un peu rapide lors de la détermination de l'univers image : Est-on sûr que  $T(\Omega) = \mathbf{N}^*$  ?

On pourrait tout à fait imaginer que l'événement  $[T = +\infty]$  soit possible : que l'on n'obtienne jamais « six » dans la série de lancers.

Pour ne pas considérer l'événement  $[T = +\infty]$ , il faut montrer que cet événement est **négligeable**, c'est-à-dire qu'il est de probabilité nulle. Cela revient à montrer que l'événement contraire :  $[T \in \mathbf{N}^*]$  est **quasi-certain**, c'est-à-dire de probabilité 1.

On calcule donc cette probabilité :

$$\begin{aligned} \mathbf{P}(T \in \mathbf{N}^*) &= \sum_{k \in \mathbf{N}^*} \mathbf{P}(T = k) && \text{(événements incompatibles)} \\ &= \sum_{k=1}^{+\infty} \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} = \frac{1}{6} \sum_{k=0}^{+\infty} \left(\frac{5}{6}\right)^k && \text{(changement d'indice)} \\ &= \frac{1}{6} \frac{1}{1 - \frac{5}{6}} && \text{(somme géométrique)} \\ &= 1. \end{aligned}$$

Donc l'événement  $[T \in \mathbf{N}^*]$  est quasi-certain et on peut donc choisir  $\mathcal{X} = \mathbf{N}^*$ .

On a noté la variable aléatoire  $T$  car elle représente un temps : le temps d'attente pour obtenir le premier succès.

**Définition 4.1** (*Loi géométrique*)

Soit  $T$  une variable aléatoire définie sur  $\Omega$  à valeurs dans  $\mathbf{N}^*$ , on dit que  $T$  suit une **loi géométrique** de paramètre  $p \in ]0, 1[$  si

$$\forall k \in \mathbf{N}^*, \mathbf{P}(T = k) = p(1 - p)^{k-1}.$$

On note  $T \sim \mathcal{G}(p)$  ou  $T \hookrightarrow \mathcal{G}(p)$ .

**Propriété 4.2**

Si  $T$  suit une loi géométrique de paramètre  $p$ , alors

$$\forall k \in \mathbf{N}, \mathbf{P}(T > k) = (1 - p)^k.$$

**Explications**

Cela représente la probabilité d'attendre strictement plus de  $k$  essais avant d'obtenir le succès. On remarque que cette probabilité est exactement égale à la probabilité d'obtenir exactement  $k$  échecs.

**Preuve**

$$\begin{aligned} \forall k \in \mathbf{N}, \mathbf{P}(T > k) &= 1 - \mathbf{P}(T \leq k) = 1 - \sum_{k=1}^k p(1 - p)^{k-1} = 1 - p \sum_{k=0}^{k-1} (1 - p)^k \\ &= 1 - p \frac{1 - (1 - p)^k}{1 - (1 - p)} = 1 - p \frac{1 - (1 - p)^k}{p} = (1 - p)^k. \end{aligned}$$

**Explications**

Il faut se souvenir que la loi géométrique correspond au rang du premier succès lors de la répétition indépendantes d'une même expérience de Bernoulli.

Il est donc important

- que l'expérience soit répétée une infinité de fois<sup>2</sup> de façon indépendante,
- que ce soit la même expérience de Bernoulli : le paramètre ne change pas.

Ces deux conditions sont capitales, et elles justifient une caractéristique essentielle de la loi géométrique : l'absence de mémoire.

En effet, chaque expérience étant indépendante des précédentes, si le succès n'est toujours pas arrivé à un certain rang  $j$ , alors la probabilité qu'il arrive au rang  $j + 1$  est la même que celle qu'il soit arrivé au rang 1.

En d'autres termes, ce n'est pas parce que l'expérience a déjà été réalisée de nombreuses fois auparavant, que l'événement a davantage de chances de se produire.

On parle aussi de non vieillissement.

**Propriété 4.3** (*Non vieillissement*)

Si  $T$  suit une loi géométrique, alors

$$\forall j \in \mathbf{N}, \forall k \in \mathbf{N}, \mathbf{P}(T > j + k | T > j) = \mathbf{P}(T > k).$$

Ou, écrit avec l'autre notation pour les probabilités conditionnelles

$$\forall j \in \mathbf{N}, \forall k \in \mathbf{N}, \mathbf{P}_{[T > j]}(T > j + k) = \mathbf{P}(T > k).$$

**Preuve**

Soient  $j \in \mathbf{N}$ , et  $k \in \mathbf{N}$ ,

$$\begin{aligned} \mathbf{P}(T > j + k | T > j) &= \frac{\mathbf{P}([T > j + k] \cap [T > j])}{\mathbf{P}(T > j)} \quad (\text{on sait que } \mathbf{P}(T > j) > 0) \\ &= \frac{\mathbf{P}(T > j + k)}{\mathbf{P}(T > j)} \end{aligned}$$

En effet,  $[T > j + k] \subset [T > j]$ , donc  $[T > j + k] \cap [T > j] = [T > j + k]$ .

$$\text{Donc, } \mathbf{P}(T > j + k | T > j) = \frac{(1 - p)^{j+k}}{(1 - p)^j} = (1 - p)^k = \mathbf{P}(T > k). \quad \blacksquare$$

On peut démontrer que c'est même une caractérisation des lois géométriques : si une variable aléatoire dénombrable est sans mémoire, alors c'est une loi géométrique.

**Exemple**

Les Shadoks essaient d'envoyer une fusée vers la Terre (à moins que ce ne soit une casserole).

Les différents essais sont indépendants entre eux, et on sait que la probabilité qu'ils réussissent au cours du premier million d'essais est de 63%.

Sachant que les Shadoks pompent toute la journée et qu'ils ont déjà échoué un million de lancers, quelle est la probabilité qu'ils ratent encore le million de lancers suivants ?

**Solution :**

Si on note  $T$  la variable aléatoire correspondant au numéro du premier lancer réussi, alors  $T$  suit une loi géométrique.

La loi étant sans mémoire, la probabilité qu'il ratent le million suivant est la même qu'ils aient raté le premier million :  $1 - 0,63 = 0,37$ .

Ils ont donc 37% de chances de rater le million d'essais suivants.

2. Au moins virtuellement, car dans l'exercice concret, on autorise le chargé d'expérience à aller se reposer dès que le premier succès est arrivé.

## B Loi de Poisson

### Exemple (tirage dans une grande urne)

Soit  $n$  un entier naturel « très grand ». Une urne contient  $n$  boules, dont 5 sont noires et les autres blanches.

On réalise  $n$  tirages avec remise et on note  $X$  le nombre de boules noires obtenues au cours de ces tirages. Ainsi,  $X$  suit une loi binomiale de paramètres  $n$  et  $\frac{5}{n}$ .

Le nombre moyen de succès est  $np_n = 5$ .

On va montrer que pour  $n \rightarrow +\infty$ , on peut réaliser l'approximation :

$$\forall k \in \mathbf{N}, \mathbf{P}(X = k) = \frac{5^k}{k!} e^{-5}.$$

C'est ainsi que l'on définit la loi de Poisson.

### Explications

La loi de Poisson modélise le nombre d'occurrences d'un événement *rare* sur un certain intervalle. Le paramètre de la loi correspond au nombre moyen de succès dans l'intervalle considéré.

#### Définition 4.4 (Loi de Poisson)

$X$  suit une loi de Poisson de paramètre  $\lambda > 0$  si

$$\forall k \in \mathbf{N}, \mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

En particulier,  $\mathcal{X} = X(\Omega) = \mathbf{N}$ .

On note  $X \sim \mathcal{P}(p)$  ou  $X \hookrightarrow \mathcal{P}(p)$ .

### Preuve

$\forall k \in \mathbf{N}$ , on a  $\frac{\lambda^k}{k!} e^{-\lambda} > 0$ . Or, on sait que  $\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^\lambda$ , donc  $\sum_{k=0}^{+\infty} \mathbf{P}(X = k) = 1$ . ■

La propriété qui suit justifie la formule pour la loi de Poisson à partir du premier exemple. Cette loi est obtenue comme une approximation de la loi binomiale pour  $n$  très grand, lorsque le nombre moyen de succès converge :  $np_n \rightarrow \lambda$ .

#### Propriété 4.5 (Approximation de la loi binomiale)

Soit  $(p_n)_{n \in \mathbf{N}} \in ]0, 1[^\mathbf{N}$  une suite telle que  $np_n \xrightarrow{n \rightarrow +\infty} \lambda > 0$ .

Si, pour tout  $n \in \mathbf{N}$ , on note  $X_n$  une variable aléatoire de loi  $\mathcal{B}(n, p_n)$ , alors,

$$\forall k \in \mathbf{N}, \lim_{n \rightarrow +\infty} \mathbf{P}(X_n = k) = \mathbf{P}(X = k)$$

avec  $X \hookrightarrow \mathcal{P}(\lambda)$ .

Dans la pratique, on peut utiliser cette approximation<sup>3</sup> pour  $n \geq 30$ ,  $p \leq 0,1$  et  $np \leq 15$ .

3. On trouvera d'autres valeurs dans d'autres cours... chacun y va de sa petite règle. Aucun critère précis n'est donné dans le programme, c'est donc indicatif.

### Exemple

On considère un ouvrage 300 pages, et on estime à 1%, la probabilité qu'il y ait des erreurs de typographie sur une page donnée. Les justesses des pages sont indépendantes entre elles.

1. Estimer la probabilité qu'il n'y ait pas d'erreurs dans l'ouvrage.
2. Estimer la probabilité qu'il y ait moins de 5 pages avec des erreurs dans l'ouvrage.

### Solution :

Si on considère que la probabilité qu'une page ait des erreurs est indépendante de la justesse des autres pages. On peut alors considérer que le nombre de pages avec erreur dans le livre suit une loi binomiale de paramètres  $n = 300$  et  $p = 1\%$ .

On trouve alors  $\mathbf{P}(X = 0) = \binom{n}{0} p^0 (1-p)^n \approx 4,9\%$ .

Et  $\mathbf{P}(X \leq 5) = \sum_{k=0}^5 \binom{n}{k} p^k (1-p)^{n-k}$  : avec un tableur, on calcule 91,7%.

Comme  $n$  est grand, et  $p$  petit, on peut approximer la loi du nombre de pages erronées avec une loi de Poisson de paramètre  $\lambda = np = 300 \times 1\% = 3$ .

On trouve alors  $\mathbf{P}(X = 0) = e^{-3} \approx 4,97\%$ .

De même,  $\mathbf{P}(X \leq 5) = \sum_{k=0}^5 \frac{\lambda^k}{k!} e^{-\lambda} = 91,6\%$ .

### Preuve

Soit  $k \in \mathbf{N}$  et  $n \geq 1$ ,  $\mathbf{P}(X = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{1}{n^k} \binom{n}{k} \lambda^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$ .

Or  $\frac{1}{n^k} \binom{n}{k} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k} \frac{1}{k!} = \frac{n}{n} \frac{n-1}{n} \frac{n-2}{n} \cdots \frac{n-k+1}{n} \frac{1}{k!}$ .

Donc par produit de  $k$  limites qui valent toutes 1 ( $k$  est un nombre fixé)

$$\frac{1}{n^k} \binom{n}{k} \xrightarrow{n \rightarrow +\infty} 1 \cdot \frac{1}{k!} = \frac{1}{k!}.$$

Pour  $n \geq 1$ ,  $\left(1 - \frac{\lambda}{n}\right)^{n-k} = \exp\left((n-k) \ln\left(1 - \frac{\lambda}{n}\right)\right)$ .

Or,  $(n-k) \ln\left(1 - \frac{\lambda}{n}\right) \underset{n \rightarrow +\infty}{\sim} -\lambda \frac{(n-k)}{n} \underset{n \rightarrow +\infty}{\sim} -\lambda$ .

Par composition avec la fonction exponentielle, continue en  $\lambda$  :

$$\lim_{n \rightarrow +\infty} \exp\left(n \ln\left(1 - \frac{\lambda}{n}\right) - k \ln\left(1 - \frac{\lambda}{n}\right)\right) = e^{-\lambda}.$$

Ainsi, par produit des deux limites précédentes,

$$\lim_{n \rightarrow +\infty} \mathbf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

■

**Exemple** (*Source radioactive*)

On cherche à étudier le nombre de désintégrations d'une source radioactive pendant un temps  $T$  fixé.

On admet que la probabilité qu'il y ait une désintégration au cours d'un certain temps  $\Delta t$  est proportionnelle à  $\Delta t$ .

→ la source « ne vieillit pas » : les noyaux ont autant de probabilité de se désintégrer quelque soit leur âge.

On peut donc poser  $\alpha$  tel que la probabilité d'une désintégration sur un temps  $\Delta t$  soit  $\alpha\Delta t$ . On découpe l'intervalle de temps  $[0, T[$  avec des pas de temps  $\Delta t = \frac{T}{n}$  très petits de telle sorte que l'on puisse raisonnablement supposer qu'entre  $t$  et  $t + \Delta t$ , il puisse se produire au plus une seule désintégration.

Si on note  $A_k$  : « il y a une désintégration dans l'intervalle  $[k\Delta t, (k + 1)\Delta t[$  » alors,  $\forall n \in \llbracket 0, n - 1 \rrbracket$ ,  $\mathbf{P}(A_k) = p_n = \alpha\Delta t = \alpha\frac{T}{n}$ .

On suppose enfin que les événements  $(A_k)_k$  sont indépendants.

→ la source contient suffisamment de noyaux pour que, sur l'intervalle de temps  $[0, T]$ , on puisse considérer que le nombre total de noyau est constant (la quantité de noyaux désintégrés est négligeable par rapport au nombre total de noyaux).

Si on note  $X_n$  le nombre de désintégrations pendant l'intervalle  $[0, T]$ , alors  $X_n$  est la somme de variables aléatoires de Bernoulli, indépendantes et de même loi.

Ainsi

$$\forall k \in \llbracket 0, n \rrbracket, \mathbf{P}(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k}.$$

Lorsque  $n \rightarrow +\infty$  (un pas de temps infiniment précis) on peut approximer

$$\forall k \in \mathbf{N}, \mathbf{P}(X = k) = \frac{(\alpha T)^k}{k!} e^{-\alpha T} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

On a posé  $\lambda = \alpha\Delta t$  qui correspond, au nombre moyen de désintégration au cours du temps  $T$ .  $X$  suit alors une loi de Poisson de paramètre  $\lambda > 0$ .

**5 ESPÉRANCE ET ÉCART TYPE**

On définit l'espérance et l'écart-type de variables aléatoires dénombrables de la même manière que dans le cas fini, à une grosse exception près : ces quantités sont des limites qui peuvent être infinies.

**A Espérance**

**Définition 5.1** (*Espérance*)

Soit  $X$  une variable aléatoire dénombrable à valeurs dans  $\mathcal{X} = \{x_i, i \in I\} \subset \mathbf{R}$ .

$X$  admet une espérance si la série  $\sum_{i \in I} x_i \mathbf{P}(X = x_i)$  est **absolument convergente**.

Dans ce cas, on définit l'espérance de  $X$  par

$$\mathbf{E}(X) = \sum_{i \in I} x_i \mathbf{P}(X = x_i).$$

Cette définition coïncide avec celle du cas fini lorsque la variable aléatoire est finie.

⚠ La convergence absolue est nécessaire. En effet, elle permet de sommer les termes dans n'importe quel ordre. Au contraire, pour une série semi-convergente, le changement d'ordre de sommation peut changer totalement la valeur de la limite !

**Propriété 5.2**

Soient  $X$  et  $Y$  deux variables aléatoires qui *admettent une espérance*.

L'espérance est **linéaire** :

$\forall (\alpha, \beta) \in \mathbf{R}^2$ ,  $\alpha X + \beta Y$  admet une espérance et

$$\mathbf{E}(\alpha X + \beta Y) = \alpha \mathbf{E}(X) + \beta \mathbf{E}(Y).$$

L'espérance est **croissante** :

Si  $X \leq Y$ , alors

$$\mathbf{E}(X) \leq \mathbf{E}(Y).$$

**Preuve**

Admis. ■

**Théorème 5.3** (*Espérance de la loi géométrique*)

Soit  $p \in ]0, 1[$ , et  $X$  une variable aléatoire telle que  $X \hookrightarrow \mathcal{G}(p)$ .  
 $X$  admet une espérance :

$$\mathbf{E}(X) = \frac{1}{p}.$$

**Explications**

Intuitivement, si on a une chance sur 5 de réussir à chaque essai, alors on réussira en moyenne au cinquième essai. Ainsi, pour  $p = \frac{1}{N}$ ,  $\mathbf{E}(X) = N$ .

**Preuve**

La série  $\sum k(1-p)^{k-1}$  est une série géométrique dérivée : elle converge.  
 Donc la série  $\sum kp(1-p)^{k-1}$  converge (et elle est à termes positifs, donc il s'agit d'une convergence absolue). Donc  $X$  admet une espérance :

$$\mathbf{E}(X) = \sum_{k=1}^{+\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{+\infty} k(1-p)^{k-1} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}. \quad \blacksquare$$

**Théorème 5.4** (*Espérance de la loi de Poisson*)

Soit  $\lambda \in ]0, 1[$ , et  $X$  une variable aléatoire telle que  $X \hookrightarrow \mathcal{P}(\lambda)$ .  
 $X$  admet une espérance finie

$$\mathbf{E}(X) = \lambda.$$

**Explications**

On se souvient que pour construire la loi de Poisson à partir de la loi binomiale, on avait pris  $\lambda = \lim_{n \rightarrow +\infty} np_n$ . Et  $np_n$  correspond exactement à l'espérance de la loi binomiale considérée. On n'est donc pas étonné de retrouver  $\lambda$  comme espérance de la loi limite.

**Preuve** (*À savoir refaire*)

$$\begin{aligned} \mathbf{E}(X) &= \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= \sum_{k=0}^{+\infty} \frac{\lambda^{k+1}}{k!} e^{-\lambda} \quad (\text{changement d'indice}) \\ &= \lambda \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^\lambda e^{-\lambda} = \lambda. \quad \blacksquare \end{aligned}$$

**B Moments****Théorème 5.5** (*Théorème de transfert*)

Soit  $u : \mathbf{R} \rightarrow \mathbf{R}$  une fonction réelle, et  $X$  une variable aléatoire à valeurs dans  $\mathcal{X}$ .  
 $u(X)$  admet une espérance si et seulement si  $\sum u(x)\mathbf{P}(X=x)$  converge absolument pour  $x \in \mathcal{X}$  et dans ce cas,

$$\mathbf{E}(u(X)) = \sum_{x \in \mathcal{X}} u(x)\mathbf{P}(X=x).$$

**Définition 5.6** (*Moment d'ordre k*)

Soit  $X$  une variable aléatoire, on définit le moment d'ordre  $k$  de  $X$  (sous réserve d'existence)

$$\mathbf{E}(X^k) = \sum_{x \in \mathcal{X}} x^k \mathbf{P}(X=x).$$

**Propriété 5.7**

Si  $X$  admet un moment d'ordre  $p$ , alors pour tout  $k \in \llbracket 0, p \rrbracket$ ,  $X$  admet un moment d'ordre  $k$ .

**Preuve**

$$|X|^k \leq |X|^p \mathbf{1}_{|X|>1} + \mathbf{1}_{|X|\leq 1} \leq |X|^p + \mathbf{1}_{|X|\leq 1}.$$

En effet, si  $|X(\omega)| = x > 1$ , alors on a  $x^k \leq x^p$ , donc  $|X(\omega)|^k = x \leq |X(\omega)|^p + 0$ .

De même, si  $|X(\omega)| = x \leq 1$ , alors on a  $x^k \leq 1$ , donc  $|X(\omega)|^k = x \leq |X(\omega)|^p \times 0 + 1 \leq |X(\omega)|^p + 1$ .

Or  $\sum |x|^p \mathbf{P}(X=x)$  converge, par hypothèse, et  $\sum \mathbf{1}_{|x|\leq 1}(x)\mathbf{P}(X=x)$  converge également, donc  $\sum |x|^k \mathbf{P}(X=x)$  converge :  $X$  admet un moment d'ordre  $k$ .  $\blacksquare$

**Définition 5.8** (*Variance*)

Si  $X$  est une variable aléatoire qui admet un moment d'ordre 2, alors, on définit sa variance comme le moment d'ordre 2 de sa variable centrée :

$$\mathbf{V}(X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right).$$

**Preuve**

Puisque  $X$  admet un moment d'ordre 2, on sait que  $X$  admet une espérance d'après la propriété précédente.

Montrons à présent que  $X - \mathbf{E}(X)$  admet un moment d'ordre 2.

$(X - \mathbf{E}(X))^2 = X^2 + \mathbf{E}(X)^2 - 2X\mathbf{E}(X)$ . Or  $X^2$  admet une espérance (par hypothèse),  $\mathbf{E}(X^2)$  est une variable certaine, donc elle admet une espérance.

Et  $X\mathbf{E}(X)$  admet une espérance (car  $X$  admet une espérance, et  $\mathbf{E}(X)$  est une constante).

Donc,  $(X - \mathbf{E}(X))^2$  admet une espérance.  $\blacksquare$

**Propriété 5.9** (*Propriétés de la variance*)

Soit  $X$  une variable aléatoire admettant un moment d'ordre 2

1. (*positivité*)  $\mathbf{V}(X) \geq 0$ .

Et  $\mathbf{V}(X) = 0$  si et seulement si  $X$  est presque certaine.

2. (*Formule de Koëniq-Huygens*)  $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$ .

3. Si  $(a, b) \in \mathbf{R}^2$ , alors  $\mathbf{V}(aX + b) = a^2\mathbf{V}(X)$ .

**Définition 5.10** (*Écart type*)

Soit  $X$  une variable aléatoire positive de variance finie, on définit son écart type par

$$\sigma(X) = \sqrt{\mathbf{V}(X)}.$$

**Théorème 5.11** (*Variance de la loi géométrique*)

Si  $p \in ]0, 1[$  et  $X \hookrightarrow \mathcal{G}(p)$ , alors

$$\mathbf{V}(X) = \frac{1-p}{p^2}.$$

**Preuve**

On sait que la série  $\sum k(k-1)q^{k-2}$  converge pour  $q \in ]0, 1[$  (série géométrique dérivée), donc  $\sum k(k-1)pq^{k-1}$  converge par linéarité.

Ainsi  $\mathbf{E}(X(X-1))$  est bien défini et  $\mathbf{E}(X(X-1)) = pq \frac{2}{(1-q)^3} = \frac{2(1-p)}{p^2}$ .

Il est également immédiat que  $X$  admet un moment d'ordre 2, car  $X^2 = X(X-1) + X$ ,  $\mathbf{V}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = \mathbf{E}(X(X-1)) + \mathbf{E}(X) - (\mathbf{E}(X))^2$  ■

$$= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

**Théorème 5.12** (*Variance de la loi de Poisson*)

Si  $\lambda > 0$  et  $X \hookrightarrow \mathcal{P}(\lambda)$ , alors

$$\mathbf{V}(X) = \lambda.$$

*Remarque* : Pour une loi de Poisson, l'espérance et la variance sont égales.

**Preuve**

À faire en exercice avec un changement d'indice. ■

**C Inégalités remarquables**

Les inégalités de Markov et de Bienaymé-Tchebychev vues dans le cas fini restent valables dans le cas dénombrable.

**Théorème 5.13** (*Inégalité de Markov*)

Si  $X$  est une variable aléatoire **positive** et  $a > 0$ , alors

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a}.$$

*Remarque* : Cette inégalité est vraie si l'espérance est infinie, mais elle n'a alors pas beaucoup d'intérêt.

**Théorème 5.14** (*Inégalité de Bienaymé-Tchebychev*)

Si  $X$  est une variable aléatoire admettant un ordre 2 et  $a > 0$ , alors

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

**Preuve**

Ce sont les mêmes preuves que pour une variable aléatoire finie. ■

## 6 COUPLES ET FAMILLES

Les théorèmes et définitions vus pour les variables aléatoires finies restent vrais. Cette dernière partie du cours, n'est donc qu'une redite rapide de ce qui avait été exposé dans le chapitre sur les couples et vecteurs de variables aléatoires.

### A Couples

**Loi conjointe de  $Z = (X, Y)$  :**  $\mathbf{P}((X, Y) = (x, y)) = \mathbf{P}(X = x \text{ et } Y = y)$ .

**Lois marginales :** la loi  $\mathbf{P}_X$  de la variable  $X$  est la première loi marginale de  $(X, Y)$ , et  $\mathbf{P}_Y$  est la seconde loi marginale.

**La loi conjointe donne les lois marginales :**

$$\mathbf{P}_X(x) = \sum_{y \in Y(\Omega)} \mathbf{P}(X = x \text{ et } Y = y) \quad \text{et} \quad \mathbf{P}_Y(y) = \sum_{x \in X(\Omega)} \mathbf{P}(X = x \text{ et } Y = y).$$

**Covariance :** Si  $X$  et  $Y$  admettent toutes les deux un moment d'ordre 2 :

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

**⚠** Dans la suite, lorsque l'on parlera de variance ou de covariance dans une formule, il sera sous-entendu que les variables admettent toutes un moment d'ordre 2.

**Formule de Kœnig-Huygens :**  $\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ .

**Propriétés de la covariance :**

1.  $\mathbf{Cov}(X, Y) = \mathbf{Cov}(Y, X)$ . (symétrie)
2.  $\mathbf{Cov}(aX + b, Y) = a\mathbf{Cov}(X, Y)$ .
3.  $\mathbf{Cov}(aX_1 + X_2, Y) = a\mathbf{Cov}(X_1, Y) + \mathbf{Cov}(X_2, Y)$ . (linéarité)
4.  $\mathbf{Cov}(X, X) = \mathbf{V}(X) \geq 0$ . (positive)

**Lien avec la variance :**  $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\mathbf{Cov}(X, Y)$ .

**Coefficient de corrélation :** Si  $X$  et  $Y$  ne sont pas presque certaines, alors

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

**Propriétés du coefficient de corrélation :**  $-1 \leq \rho(X, Y) \leq 1$ .

$$\rho(X, Y) \in \{-1, 1\} \iff \exists(a, b) \in \mathbf{R}^2, \text{ tel que } \mathbf{P}(X = aY + b) = 1.$$

$$\rho(X, Y) = 1 \Rightarrow \mathbf{P}(Y^* = X^*) = 1, \quad \text{et} \quad \rho(X, Y) = -1 \Rightarrow \mathbf{P}(Y^* = -X^*) = 1.$$

avec  $X^*$  et  $Y^*$  les variables centrées réduites.

$$\forall(a, b) \in \mathbf{R}^2, \forall(c, d) \in \mathbf{R}^2, \text{ avec } a > 0 \text{ et } c > 0 \quad \rho(aX + b, cY + d) = \rho(X, Y).$$

## B Indépendance

**Indépendance :**

- (Deux variables)

$$\forall(x, y) \in X(\Omega) \times Y(\Omega), \mathbf{P}([X = x] \cap [Y = y]) = \mathbf{P}(X = x)\mathbf{P}(Y = y).$$

- ( $n$  variables mutuellement indépendantes)

$$\forall(x_i)_{i \in [1, n]} \in \left(\prod_{i=1}^n X_i\right)(\Omega), \mathbf{P}\left(\bigcap_{i \in I} X_i = x_i\right) = \prod_{i \in I} \mathbf{P}(X_i = x_i).$$

**Propriétés liées à l'indépendance :**

Si  $(X_1, X_2, \dots, X_n)$  sont des variables mutuellement indépendantes, alors

1. toute sous-famille l'est aussi.
2. pour  $(A_1, A_2, \dots, A_n) \in (X_1 \times X_2 \times \dots \times X_n)(\Omega)$ , les événements  $(X_i \in A_i)$  sont mutuellement indépendants.
3. (passage à l'image)  $(f_1(X_1), f_2(X_2), \dots, f_n(X_n))$  sont mutuellement indépendantes.
4. (lemme des coalitions)  $g(X_1, X_2, \dots, X_p)$  et  $g(X_{p+1}, \dots, X_n)$  sont indépendantes.

**Espérance, écart type :**

Si  $(X_1, X_2, \dots, X_n)$  sont des variables mutuellement indépendantes, alors

$$\mathbf{E}(X_1 X_2 \dots X_n) = \mathbf{E}(X_1)\mathbf{E}(X_2) \dots \mathbf{E}(X_n).$$

$$\mathbf{V}(X_1 + X_2 + \dots + X_n) = \mathbf{V}(X_1) + \mathbf{E}(X_2) + \dots + \mathbf{E}(X_n).$$

Pour  $i \neq j$ ,  $\mathbf{Cov}(X_i, X_j) = 0$  et  $\rho(X_i, X_j) = 0$ .