

---

# Notes de cours de Probabilités

Institut Supérieur de l'Aéronautique et de l'Espace (ISAE)

---

Jérémie Bigot

18 mars 2014

# Remerciements

Ces notes de cours sont très largement inspirées des photocopiés de probabilités préparés par Claudie Chabriac qui sont disponibles à l'adresse :

<https://sites.google.com/site/cchabriac/>

Mes remerciements vont donc tout d'abord à Claudie qui m'a autorisé à reprendre une grande partie de ces notes de cours pour préparer ce photocopié. Ce document est également inspiré des cours de probabilités dont la responsabilité a été assurée à l'ENSICA et à SUPAERO par Yves Caumel et Manuel Samuelides depuis de nombreuses années. Je voudrais les remercier tous les deux pour leurs conseils pour la préparation de ces notes de cours. Les cours de probabilités à l'ISAE s'appuient sur les enseignements dispensés par de nombreux vacataires que je voudrais également remercier très chaleureusement pour leur investissement. Le lecteur intéressé par un cours plus approfondi de probabilités pour l'ingénieur pourra consulter les ouvrages suivants :

- Caumel Yves (2011), Probabilites et Processus Stochastiques, Springer Editions, Serie Statistique et Probabilites Appliquees.
- Samuelides Manuel (1995), Modèles aléatoires, SUPAERO - ISAE.
- François Olivier (2004), Notes de cours de Probabilités, ENSIMAG.  
<http://membres-timc.imag.fr/Olivier.Francois/>.

Pour un panorama beaucoup plus complet de la théorie des probabilités, nous renvoyons également au livre suivant :

- Barbé Philippe & Ledoux Michel (2007), Probabilité (L3M1), EDP Sciences.

# Table des matières

<b>1</b>	<b>CONCEPTS DE BASE</b>	<b>7</b>
1.1	Introduction de la notion d'événement . . . . .	7
1.2	Tribus . . . . .	9
1.2.1	Définition, premières propriétés . . . . .	9
1.2.2	Tribu engendrée par une famille de parties . . . . .	10
1.3	Introduction de la notion de probabilité . . . . .	11
1.4	Événements indépendants . . . . .	14
1.5	Rappels d'analyse combinatoire . . . . .	15
1.5.1	Tirages ordonnés avec remise (ou applications) . . . . .	15
1.5.2	Tirages ordonnés sans remise (ou injections) . . . . .	15
1.5.3	Tirages non ordonnés sans remise (ou combinaisons) . . . . .	16
1.5.4	Tirages non ordonnés avec remise . . . . .	17
1.5.5	Partages d'ensembles . . . . .	17
<b>2</b>	<b>VARIABLES ALÉATOIRES REELLES, LOIS CLASSIQUES SUR <math>\mathbb{R}</math></b>	<b>19</b>
2.1	Généralités . . . . .	19
2.1.1	Tribu engendrée par une application . . . . .	19
2.1.2	Variables aléatoires . . . . .	20
2.1.3	Loi de probabilité d'une variable aléatoire . . . . .	21
2.1.4	Variables aléatoires réelles . . . . .	21
2.2	Fonctions de répartition . . . . .	22
2.3	Variables aléatoires réelles discrètes . . . . .	23
2.3.1	Définition, fonction de répartition et loi d'une v.a.r. discrète . . . . .	23
2.3.2	Fonction d'une v.a.r. discrète . . . . .	24
2.3.3	Lois discrètes classiques . . . . .	24
2.4	Variables aléatoires réelles absolument continues . . . . .	26
2.4.1	Définition, fonction de répartition, densité . . . . .	26
2.4.2	Loi d'une fonction d'une v.a.r. absolument continue . . . . .	27
2.4.3	Lois absolument continues classiques . . . . .	29
<b>3</b>	<b>ESPERANCE, VARIANCE ET MOMENTS DE VARIABLES ALÉATOIRES REELLES</b>	<b>33</b>
3.1	Variables aléatoires réelles discrètes . . . . .	33
3.1.1	Espérance . . . . .	33
3.1.2	Exemples d'espérances pour des lois discrètes classiques . . . . .	33
3.1.3	Variance et moments d'ordre $r$ . . . . .	35
3.1.4	Calcul de $\mathbb{E}(\varphi(X))$ . . . . .	35
3.1.5	Fonctions génératrices . . . . .	36

3.2	Variables aléatoires réelles absolument continues . . . . .	38
3.2.1	Espérance d'une v.a.r. absolument continue . . . . .	38
3.2.2	Variance et moments d'une v.a.r. absolument continue . . . . .	39
3.2.3	Calcul de $\mathbb{E}(\varphi(X))$ . . . . .	40
3.2.4	Fonction caractéristique . . . . .	41
3.2.5	Application aux lois classiques . . . . .	41
<b>4</b>	<b>INTEGRALE DE LEBESGUE ET VARIABLES ALEATOIRES</b>	<b>43</b>
4.1	Construction de l'intégrale de Lebesgue . . . . .	44
4.1.1	Espace mesuré . . . . .	44
4.1.2	Intégrale d'une fonction au sens de Lebesgue . . . . .	45
4.1.3	Mesures définies par des densités . . . . .	47
4.2	Loi de probabilité d'une variable aléatoire . . . . .	47
4.2.1	Loi de probabilité d'une v.a. réelle discrète . . . . .	48
4.2.2	Loi de probabilité d'une v.a. réelle continue . . . . .	48
4.3	Espérance d'une variable aléatoire . . . . .	49
<b>5</b>	<b>COUPLES ALÉATOIRES DISCRETS</b>	<b>51</b>
5.1	Généralités . . . . .	51
5.1.1	Loi de probabilité d'un couple $(X, Y)$ . . . . .	51
5.1.2	Lois marginales . . . . .	51
5.1.3	Indépendance de deux v.a.r. discrètes . . . . .	53
5.1.4	Somme de deux v.a.r. discrètes . . . . .	53
5.2	Opérateurs classiques . . . . .	54
5.2.1	Espérance . . . . .	54
5.2.2	Variance et covariance . . . . .	56
<b>6</b>	<b>COUPLES ALÉATOIRES A DENSITÉ</b>	<b>59</b>
6.1	Loi de probabilité d'un couple de variables aléatoires réelles . . . . .	59
6.2	Indépendance . . . . .	61
6.3	Opérateurs classiques . . . . .	62
6.3.1	Espérance . . . . .	62
6.3.2	Variance et covariance . . . . .	62
6.4	Changement de variables . . . . .	63
6.5	Sommes de deux v.a.r. absolument continues . . . . .	64
<b>7</b>	<b>CONVERGENCE DE SUITES DE VARIABLES ALÉATOIRES</b>	<b>67</b>
7.1	Inégalités . . . . .	67
7.2	Convergence presque sûre . . . . .	68
7.3	Convergence en moyenne quadratique . . . . .	69
7.4	Convergence en probabilité . . . . .	69
7.5	Convergence en loi . . . . .	70
7.6	Loi des grands nombres . . . . .	72
7.7	Théorème central limite - convergence vers la loi normale . . . . .	72
<b>8</b>	<b>CONDITIONNEMENT</b>	<b>75</b>
8.1	Probabilités conditionnelles . . . . .	75
8.1.1	Introduction . . . . .	75
8.1.2	Formule des probabilités totales . . . . .	76

8.1.3	Théorème de Bayes . . . . .	77
8.2	Conditionnement dans le cas discret . . . . .	77
8.2.1	Lois conditionnelles pour un couple discret . . . . .	77
8.2.2	Espérance conditionnelle dans le cas discret . . . . .	78
8.3	Conditionnement dans le cas continu . . . . .	79
8.3.1	Lois conditionnelles dans le cas continu . . . . .	79
8.3.2	Espérance conditionnelle dans le cas continu . . . . .	80
8.4	Complément sur les lois conditionnelles . . . . .	81
8.4.1	Loi d'une variable absolument continue $Y$ conditionnée par une variable discrète $X$ . . . . .	81
8.4.2	Loi d'une variable discrète conditionnée par une variable absolument continue . . . . .	82
8.5	Compléments sur l'espérance conditionnelle . . . . .	82
8.6	Régression linéaire de $Y$ par rapport à $X$ . . . . .	85
8.7	Variance conditionnelle . . . . .	86
<b>9</b>	<b>VECTEURS ALÉATOIRES - VECTEURS GAUSSIENS</b>	<b>87</b>
9.1	Vecteurs aléatoires . . . . .	87
9.1.1	Loi d'un vecteur aléatoire à densité . . . . .	88
9.1.2	Matrice de covariance . . . . .	88
9.1.3	Changement de variables . . . . .	89
9.1.4	Fonction caractéristique . . . . .	89
9.2	Vecteurs gaussiens . . . . .	90
9.2.1	Caractérisation des vecteurs gaussiens . . . . .	90
9.2.2	Quelques propriétés des vecteurs gaussiens . . . . .	91
9.2.3	Espérance conditionnelle et projection orthogonale . . . . .	92
9.2.4	Lois conditionnelles et prédiction . . . . .	94
<b>10</b>	<b>PROCESSUS DE MARKOV</b>	<b>97</b>
10.1	Introduction aux Processus Aléatoires . . . . .	97
10.2	Généralités sur les Processus de Markov . . . . .	99
10.3	Chaînes de Markov à temps discret . . . . .	100
10.3.1	Matrice de transition et graphe d'une chaîne de Markov . . . . .	100
10.3.2	Classification des états . . . . .	100
10.3.3	Formules de récurrence . . . . .	102
10.3.4	Caractérisation de la nature d'un état . . . . .	103
10.3.5	Distribution stationnaire . . . . .	104
10.3.6	Comportement asymptotique . . . . .	108
10.4	Processus de Markov continus . . . . .	109
10.4.1	Régime transitoire . . . . .	109
10.4.2	Régime permanent . . . . .	111
<b>11</b>	<b>PROCESSUS DE POISSON</b>	<b>113</b>
11.1	Introduction . . . . .	113
11.2	Définition et description du processus . . . . .	114
11.3	Caractérisation d'un processus de Poisson par ses temps d'arrivée . . . . .	117
11.4	Propriétés supplémentaires . . . . .	119
11.4.1	Processus de Poisson et loi binomiale . . . . .	119
11.4.2	Processus non homogènes . . . . .	119



# Chapitre 1

## CONCEPTS DE BASE

On fait appel aux probabilités pour décrire une expérience dont le résultat est impossible à prévoir avec certitude, mais dont on connaît quand-même l'ensemble des résultats possibles. A titre d'exemple, considérons l'expérience suivante : on jette un dé et on lit le numéro apparu sur la face supérieure. Lors de cette expérience, on peut s'intéresser au numéro de la face supérieure (il y a alors 6 résultats possibles), ou seulement à la parité de cette face (il y a alors 2 résultats possibles). La notion de résultat d'une expérience n'est donc pas toujours claire : c'est l'expérimentateur qui décide de ce qui mérite le nom de résultat en fonction de ses propres motivations. Ce chapitre a pour but d'introduire les premières notions de probabilités afin de modéliser une expérience dont l'issue est aléatoire, et de se familiariser avec les outils qui seront utilisés tout au long du cours.

### 1.1 Introduction de la notion d'événement

Si, lorsqu'on répète l'expérience dans des conditions identiques, le résultat observé est susceptible de changer, l'expérience est dite *aléatoire*. L'ensemble de tous les résultats possibles ou *états* est appelé *univers* de l'expérience : on le notera  $\Omega$ .

*Expérience 1* : lancé d'un dé à 6 faces  $\Omega = \{1, 2, 3, 4, 5, 6\}$  et  $\text{card}\Omega = 6$  (où  $\text{card}$  désigne le nombre d'éléments).

*Expérience 2* : lancé de deux dés à 6 faces

$$\Omega = \{(1, 1), \dots, (1, 6), (2, 1), \dots, (2, 6), \dots, (6, 1), \dots, (6, 6)\}$$

et  $\text{card}\Omega = 36$ .

L'ensemble  $\Omega$  peut être :

- fini (ensemble des 6 faces d'un dé, des 32 cartes d'un jeu,...) ;
- infini dénombrable (ensemble des entiers naturels, ou d'états que l'on peut numéroter) ;
- infini non dénombrable (position d'une particule dans un liquide, poids, taille,...).

*Exemple* : On lance une fléchette en direction d'une cible. On peut convenir d'appeler résultat :  
→ le gain correspondant à la zone du point d'impact :  $\Omega = \{0, 100, 200, 500, 1000\}$  ;  
→ la distance du point d'impact au centre de la cible, mesurée à 1cm près par défaut :  $\Omega = \mathbb{N}$  ;  
→ le point d'impact :  $\Omega$  partie de  $\mathbb{R}^2$  correspondant au mur.

Lorsqu'on effectue une expérience aléatoire, certains faits liés à cette expérience peuvent se produire ou non : on les appelle événements.

Exemples d'événements :

Pour l'expérience 1,  $A_1$  "le numéro obtenu est pair" ;  $A_2$  "le numéro obtenu est  $\geq 4$ "

Pour l'expérience 2,  $B$  "la somme des deux numéros obtenus est 6".

Chaque résultat possible d'une expérience aléatoire est appelé événement simple, mais les événements susceptibles d'intéresser ne sont pas seulement les événements simples.

Exemple : l'évènement "le numéro obtenu est pair" lors du lancé d'un dé n'est pas un évènement simple.

Un événement est lié à une expérience associée à  $\Omega$  si, pour tout résultat  $\omega \in \Omega$ , on sait dire si cet événement a lieu ou non. On convient d'identifier un tel événement à l'ensemble des  $\omega \in \Omega$  pour lequel il a lieu. Un événement sera donc identifié à une partie de  $\Omega$ .

Exemple :

Pour l'expérience 1,  $A_1$  est réalisé si et seulement si  $\omega \in \{2, 4, 6\}$ . On notera  $A_1 = \{2, 4, 6\}$ .

Pour l'expérience 2,  $B$  est réalisé si et seulement si  $(\omega_1, \omega_2)$  vérifie  $\omega_1 + \omega_2 = 6$ .

On notera  $B = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ .

Plus généralement, à chaque expérience, on peut associer un ensemble  $\Omega$  tel que chaque événement puisse être représenté par une partie de  $\Omega$ .

### Rappel sur le vocabulaire ensembliste :

1- Soit  $A$  une partie de  $\Omega$ .

On note  $\bar{A}$  le complémentaire de  $A$  : c'est l'ensemble de tous les états qui ne sont pas dans  $A$ .

**Propriété 1.1.**  $\overline{\bar{A}} = A$ .

2- Soient  $A$  et  $B$  deux parties de  $\Omega$ .

On note  $A \cap B$  l'intersection de  $A$  et de  $B$  : c'est l'ensemble des états qui sont à la fois dans  $A$  et dans  $B$ .

On note  $A \cup B$  la réunion de  $A$  et de  $B$  : c'est l'ensemble des états qui sont dans  $A$  ou dans  $B$  (ils peuvent être dans les deux).

On note  $A \subset B$  et on dit que  $A$  est inclus dans  $B$  si tous les états de  $A$  sont dans  $B$ .

**Propriété 1.2.**

1)  $\cap$  et  $\cup$  sont commutatives et associatives.

2) Si  $A \subset B$ , alors  $A \cap B = A$ ,  $A \cup B = B$  et  $\bar{B} \subset \bar{A}$ .

3- Si  $A_1, \dots, A_n, \dots$  sont une infinité dénombrable de parties de  $\Omega$ , on note  $\bigcup_n A_n$  (ou, de façon

plus précise,  $\bigcup_{n=1}^{+\infty} A_n$  la réunion dénombrable des  $A_n$  : c'est l'ensemble des états qui sont au moins dans l'un des  $A_n$  et on note  $\bigcap_n A_n$  l'intersection dénombrable des  $A_n$  : c'est l'ensemble des états qui sont dans tous les  $A_n$  à la fois.

**Propriété 1.3.** 1)  $\overline{\bigcup_n A_n} = \bigcap_n \overline{A_n}$ ;  $\overline{\bigcap_n A_n} = \bigcup_n \overline{A_n}$ .

2)  $A \cap \left( \bigcup_n A_n \right) = \bigcup_n (A \cap A_n)$  : *distributivité de l'intersection par rapport à la réunion* ;  
 $A \cup \left( \bigcap_n A_n \right) = \bigcap_n (A \cup A_n)$  : *distributivité de la réunion par rapport à l'intersection*.

On appelle  $\emptyset$  l'événement impossible car il n'est jamais réalisé et  $\Omega$  l'événement certain car il est toujours réalisé.

Si  $A \cap B = \emptyset$ , on dit que  $A$  et  $B$  sont incompatibles car ils ne peuvent avoir lieu en même temps.

## 1.2 Tribus

### 1.2.1 Définition, premières propriétés

**Définition 1.1.** On appelle tribu  $\mathcal{A}$  sur  $\Omega$ , tout sous-ensemble de parties de  $\Omega$  tel que :

i)  $\Omega \in \mathcal{A}$  ;

ii) si  $A \in \mathcal{A}$ , alors  $\overline{A} \in \mathcal{A}$  ;

iii) si pour tout  $n \in \mathbb{N}$ ,  $A_n \in \mathcal{A}$ , alors  $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{A}$ .

Le couple  $(\Omega, \mathcal{A})$  est appelé espace probabilisable.

**Propriété 1.4.** 1)  $\emptyset \in \mathcal{A}$  ;

2) si, pour tout  $n \in \mathbb{N}$ ,  $A_n \in \mathcal{A}$ , alors  $\bigcap_{n=0}^{+\infty} A_n \in \mathcal{A}$ .

*Démonstration.* 1) Par i)  $\Omega \in \mathcal{A}$  et par ii),  $\overline{\Omega} \in \mathcal{A}$  ; or  $\overline{\Omega} = \emptyset$ .

2) Par ii)  $\overline{A_i} = B_i \in \mathcal{A}$  et par iii)  $\bigcup_{n=0}^{+\infty} B_n \in \mathcal{A}$ . Enfin, de nouveau par ii),  $\overline{\bigcup_{n=0}^{+\infty} B_n} \in \mathcal{A}$ .

Or  $\overline{\bigcup_{n=0}^{+\infty} B_n} = \bigcap_{n=0}^{+\infty} \overline{B_n} = \bigcap_{n=0}^{+\infty} A_n$ . □

**Propriété 1.5.** L'intersection d'une famille quelconque (dénombrable ou non) de tribus de  $\Omega$  est une tribu de  $\Omega$ .

Exemples de tribus :

- $\{\emptyset, \Omega\}$  : la plus petite ;
- $\mathcal{P}(\Omega)$  : la plus grande ;
- $\{\emptyset, \Omega, A, \overline{A}\}$ .

**Cas particulier très important :** lorsque  $\Omega$  est fini ou infini dénombrable, on prend toujours  $\mathcal{A} = \mathcal{P}(\Omega)$ . (Ce ne sera pas le cas lorsque  $\Omega = \mathbb{R}$  : la tribu considérée dans ce cas sera en général la tribu des boréliens  $\mathcal{B}_{\mathbb{R}}$  qui est la plus petite tribu contenant les ouverts de  $\mathbb{R}$ ).

**Système complet d'événements :**

On appelle système complet d'événements de  $\Omega$ , toute famille finie ou dénombrable  $(A_i)_{i \in I}$  telle que :

i)  $A_i \in \mathcal{A}$  pour tout  $i \in I$  ;

ii)  $\bigcup_{i \in I} A_i = \Omega$  ;

iii) si  $i \neq j$ , alors  $A_i \cap A_j = \emptyset$ .

(On dit aussi que les  $A_i$  forment une partition dénombrable de  $\Omega$ ).

### 1.2.2 Tribu engendrée par une famille de parties

Soit  $\mathcal{C}$  une famille de parties de  $\Omega$ . On souhaite définir  $\sigma(\mathcal{C})$  la “plus petite” tribu de parties de  $\Omega$  contenant  $\mathcal{C}$ , c'est-à-dire que  $\sigma(\mathcal{C})$  doit être une tribu et que, si  $\mathcal{D}$  est une tribu contenant  $\mathcal{C}$ , alors  $\sigma(\mathcal{C}) \subset \mathcal{D}$ .

On considère l'ensemble des tribus contenant  $\mathcal{C}$ . Cet ensemble n'est pas vide, puisqu'il contient  $\mathcal{P}(\Omega)$ , ensemble des parties de  $\Omega$ . D'après la propriété 1.5, l'intersection de toutes les tribus contenant  $\mathcal{C}$  est une tribu. Cette intersection contient  $\mathcal{C}$  et, si  $\mathcal{D}$  est une tribu contenant  $\mathcal{C}$ , alors  $\mathcal{D}$  contient l'intersection des tribus contenant  $\mathcal{C}$ . Ceci prouve que  $\sigma(\mathcal{C})$  existe, et est unique : c'est l'intersection de toutes les tribus contenant  $\mathcal{C}$ .

**Définition 1.2.** On appelle tribu engendrée par une famille non vide  $\mathcal{C}$  de parties de  $\Omega$ , l'intersection de toutes les tribus contenant  $\mathcal{C}$  : on la note  $\sigma(\mathcal{C})$ .

Exemples :

1. Si  $\mathcal{C} = \{A\}$  où  $A \subset \Omega$ ,  $A \neq \emptyset$  et  $A \neq \Omega$ ,  $\sigma(\mathcal{C}) = \{\emptyset, A, \overline{A}, \Omega\}$ .

2. Si  $\mathcal{C} = \{A, B, C\}$  est une partition de  $\Omega$ ,  $\sigma(\mathcal{C}) = \{\emptyset, A, B, C, A \cup B, A \cup C, B \cup C, \Omega\}$  et  $\text{card}\sigma(\mathcal{C}) = 2^3 = 8$ . (Plus généralement, si  $\mathcal{C}$  est une partition telle que  $\text{card}\mathcal{C} = n$ , alors  $\text{card}\sigma(\mathcal{C}) = 2^n$ ).

3. Plus généralement, si  $\mathcal{C} = \{A_i ; i \in I\}$  est une partition finie ou dénombrable de  $\Omega$  ( $I \subset \mathbb{N}$ ), alors  $\sigma(\mathcal{C}) = \left\{ \bigcup_{i \in J} A_i ; J \subset I \right\}$ . En effet, par définition même d'une tribu d'une part, et de  $\sigma(\mathcal{C})$

d'autre part, on a  $\mathcal{C} \subset \left\{ \bigcup_{i \in J} A_i ; J \subset I \right\} \subset \sigma(\mathcal{C})$ . On montre ensuite que  $\left\{ \bigcup_{i \in J} A_i ; J \subset I \right\}$  est une tribu en vérifiant les 3 axiomes de la définition.

4. Si  $\Omega = \mathbb{R}^d$ , on utilise le plus souvent la tribu des boréliens de  $\mathbb{R}^d$ . Il s'agit de la tribu engendrée par les pavés de  $\mathbb{R}^d$ , c'est-à-dire :

$$\mathcal{B}_{\mathbb{R}^d} = \sigma \left( \left\{ \prod_{i=1}^d [a_i, b_i[ ; a_i, b_i \in \mathbb{R} ; 1 \leq i \leq d \right\} \right).$$

Dans le cas très important où  $d = 1$ ,  $\mathcal{B}_{\mathbb{R}}$  contient en particulier  $\mathbb{N}$ ,  $\mathbb{Z}$ ,  $\mathbb{Q}$ , tous les intervalles de  $\mathbb{R}$  (ouverts, fermés, semi-ouverts, bornés ou non),... Cependant  $\mathcal{B}_{\mathbb{R}} \neq \mathcal{P}(\mathbb{R})$  (la démonstration de ce dernier point sort du cadre de ce cours).

### 1.3 Introduction de la notion de probabilité

Une probabilité  $P$  est une mesure qui permet d'évaluer les chances de réalisation des événements. Si  $\mathcal{A}$  est une tribu sur  $\Omega$ , à chaque événement  $A$  de  $\mathcal{A}$ , on associe un réel  $P(A)$  compris entre 0 et 1, appelé probabilité de l'événement  $A$ .

**Définition 1.3.** Soit  $(\Omega, \mathcal{A})$  un espace probablisable. On appelle **probabilité** (ou mesure de probabilité) sur  $(\Omega, \mathcal{A})$  toute application  $P$  de  $\mathcal{A}$  vers  $[0, 1]$  telle que :

i)  $P(\Omega) = 1$  ;

ii) pour toute suite d'événements  $A_n \in \mathcal{A}$ , incompatibles deux à deux, on a :

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

Le triplet  $(\Omega, \mathcal{A}, P)$  est appelé espace probablisé.

**Propriété 1.6. 1)** Si  $A \subset B$ , alors  $P(A) \leq P(B)$  ;

2)  $P(\bar{A}) = 1 - P(A)$  ;  $P(\emptyset) = 0$  ;

3)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  ;

4)  $P \left( \bigcup_{i=1}^n A_i \right) = \sum_{k=1}^n (-1)^{k+1} \left( \sum_{1 \leq i_1 \dots \leq i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}) \right)$  (identité de Poincaré) ;

5) si  $(A_n)_n$  est une suite croissante de  $\mathcal{A}$  ( $A_n \subset A_{n+1}$ ), alors  $P \left( \bigcup_n A_n \right) = \lim_{n \rightarrow +\infty} P(A_n)$  ;

6) si  $(B_n)_n$  est une suite décroissante de  $\mathcal{A}$  ( $B_{n+1} \subset B_n$ ), alors  $P \left( \bigcap_n B_n \right) = \lim_{n \rightarrow +\infty} P(B_n)$ .

*Démonstration.* 1)  $B = (B \cap A) \cup (B \cap \bar{A}) = A \cup (B \cap \bar{A})$  et  $A \cap (B \cap \bar{A}) = \emptyset$  donc

$$P(B) = P(A) + P(B \cap \bar{A}) \geq P(A).$$

2)  $P(\Omega) = P(A) + P(\bar{A})$  car  $A \cup \bar{A} = \Omega$  et  $A \cap \bar{A} = \emptyset$ . Or  $P(\Omega) = 1$  donc  $P(\bar{A}) = 1 - P(A)$  et, en particulier pour  $A = \Omega$ ,  $P(\emptyset) = 0$ .

3) On décompose  $A \cup B$  en 3 événements incompatibles 2 à 2 :

$$A \cup B = (A \cap \bar{B}) \cup (A \cap B) \cup (\bar{A} \cap B).$$

De même,  $A = (A \cap B) \cup (A \cap \bar{B})$  et  $B = (A \cap B) \cup (\bar{A} \cap B)$ .

En passant aux probabilités, on obtient :

$$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B),$$

avec  $P(A) = P(A \cap B) + P(A \cap \bar{B})$  et  $P(B) = P(A \cap B) + P(\bar{A} \cap B)$ . On a donc bien :

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

4) Pour  $n = 3$ , on a  $P(A_1 \cup A_2 \cup A_3) = P((A_1 \cup A_2) \cup A_3) = P(A_1 \cup A_2) + P(A_3) - P((A_1 \cup A_2) \cap A_3)$  d'après 3). Mais, toujours d'après 3),

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

et  $(A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$  union à laquelle on applique encore 3) :

$$P((A_1 \cup A_2) \cap A_3) = P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3)$$

et en regroupant tout, on a bien le résultat.

On procède ensuite par récurrence sur  $n$  (on suppose la propriété vraie au rang  $n$  et on applique 3) à  $A = \bigcup_{i=1}^n A_i$  et à  $B = A_{n+1}$ , puis l'hypothèse de récurrence pour  $A$ , la distributivité, et de nouveau l'hypothèse de récurrence aux  $A_i \cap A_{n+1} \dots$ )

5) Posons  $C_0 = A_0$  et pour  $n \geq 1$ ,  $C_n = A_n \cap \bar{A}_{n-1}$  (on a alors  $A_n = A_{n-1} \cup C_n$ ). On décompose la preuve en plusieurs étapes :

→ Les  $C_i$  sont 2 à 2 disjoints

En effet, si, par exemple,  $i > j$ , supposons qu'il existe  $\omega \in C_i \cap C_j$ . On a alors  $\omega \in C_i = A_i \cap \bar{A}_{i-1} \subset \bar{A}_{i-1}$  et  $\omega \in C_j \subset A_j \subset A_{i-1}$  car  $j \leq i-1$ , ce qui est contradictoire.

→  $A_N = \bigcup_{n=0}^N C_n$  : démonstration par récurrence sur  $N$  ;

C'est vrai pour  $N = 0$  et si  $A_N = \bigcup_{n=0}^N C_n$ , alors

$$\bigcup_{n=0}^{N+1} C_n = A_N \cup C_{N+1} = A_N \cup (A_{N+1} \cap \bar{A}_N) = A_{N+1}.$$

→  $\bigcup_n A_n = \bigcup_n C_n$  : en effet, d'abord  $C_n \subset A_n$  donc  $\bigcup_n C_n \subset \bigcup_n A_n$  ; d'autre part  $A_N = \bigcup_{n=0}^N C_n \subset \bigcup_{n \geq 0} C_n$

pour tout  $N \geq 0$  donc  $\bigcup_{N \geq 0} A_N \subset \bigcup_{n \geq 0} C_n$ .

→ On en déduit  $P\left(\bigcup_n A_n\right) = P\left(\bigcup_n C_n\right) = \sum_{n=0}^{+\infty} P(C_n)$  car les  $C_n$  sont 2 à 2 disjoints. Or, par définition

d'une série,  $\sum_{n=0}^{+\infty} P(C_n) = \lim_{N \rightarrow +\infty} \sum_{n=0}^N P(C_n)$  et comme

$$P(A_N) = P\left(\bigcup_{n=0}^N C_n\right) = \sum_{n=0}^N P(C_n),$$

on a bien  $P\left(\bigcup_{n \geq 0} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n)$ .

6) Si  $(B_n)$  est décroissante ( $B_{n+1} \subset B_n$ ) et si  $A_n = \overline{B}_n$ , alors  $(A_n)$  est croissante et, d'après 5), on a  $P\left(\bigcup_{n \geq 0} \overline{B}_n\right) = \lim_{n \rightarrow +\infty} P(\overline{B}_n) = 1 - \lim_{n \rightarrow +\infty} P(B_n)$ . Or

$$P\left(\bigcup_{n \geq 0} \overline{B}_n\right) = 1 - P\left(\overline{\bigcap_{n \geq 0} B_n}\right) = 1 - P\left(\bigcap_{n \geq 0} B_n\right).$$

Donc  $P\left(\bigcap_{n \geq 0} B_n\right) = \lim_{n \rightarrow +\infty} P(B_n)$ . □

### Cas particuliers très importants :

**1-  $\Omega$  fini :**  $\Omega = \{\omega_1, \dots, \omega_n\}$  et  $\mathcal{A} = \mathcal{P}(\Omega)$ .

Soient  $p_1, \dots, p_n$ ,  $n$  nombres réels. Il existe une probabilité  $P$  sur  $(\Omega, \mathcal{A})$  telle que, pour tout  $i \in \{1, \dots, n\}$ ,  $P(\{\omega_i\}) = p_i$  si et seulement si, pour tout  $i \in \{1, \dots, n\}$ ,  $p_i \geq 0$  et  $\sum_{i=1}^n p_i = 1$ .

La probabilité  $P$  est alors unique et, pour tout  $A \in \mathcal{A}$ ,  $P(A) = \sum_{i; \omega_i \in A} p_i$ .

**Définition 1.4.** On dit qu'il y a équiprobabilité lorsque les probabilités de tous les événements simples sont égales.

**Théorème 1.1.** S'il y a équiprobabilité, alors, pour tout événement  $A$ , on a  $P(A) = \frac{\text{card}A}{\text{card}\Omega}$ .

*Démonstration.* Supposons que  $\text{card}\Omega = n$ . Il existe  $\lambda$  tel que  $p_i = \lambda$  pour tout  $i$ . On a alors  $\sum_{i=1}^n p_i = n\lambda = 1$ , d'où  $\lambda = \frac{1}{n}$  et, pour tout  $A \in \mathcal{A}$ ,  $P(A) = \sum_{i; \omega_i \in A} p_i = (\text{card}A) \times \lambda = \frac{\text{card}A}{\text{card}\Omega}$ . □

*Exemple de calculs de probabilités dans le cas où  $\Omega$  est fini :* on lance un dé non truqué et on considère les événements  $A$  "le résultat est pair" et  $B$  "le résultat est multiple de 3".

$\Omega = \{1, 2, 3, 4, 5, 6\}$ ,  $\mathcal{A} = \mathcal{P}(\Omega)$  et, pour  $1 \leq i \leq 6$ ,  $P(\{i\}) = \frac{1}{6}$  (le dé étant non truqué, il y a équiprobabilité).

$$A = \{2, 4, 6\}; B = \{3, 6\}; A \cap B = \{6\}; A \cup B = \{2, 3, 4, 6\}.$$

$$P(A) = \frac{\text{card}A}{\text{card}\Omega} = \frac{3}{6} = \frac{1}{2} \text{ et de même } P(B) = \frac{2}{6} = \frac{1}{3}; P(A \cap B) = \frac{1}{6}; P(A \cup B) = \frac{4}{6} = \frac{2}{3}.$$

**2-  $\Omega$  infini dénombrable :**  $\Omega = \{\omega_i; i \in \mathbb{N}\}$  et  $\mathcal{A} = \mathcal{P}(\Omega)$ .

Il existe une probabilité  $P$  sur  $(\Omega, \mathcal{A})$  telle que, pour tout  $i \in \mathbb{N}$ ,  $P(\{\omega_i\}) = p_i$  si et seulement si, pour tout  $i \in \mathbb{N}$ ,  $p_i \geq 0$  et  $\sum_{i=0}^{+\infty} p_i = 1$ .

La probabilité  $P$  est alors unique et, pour tout  $A \in \mathcal{P}$ ,  $P(A) = \sum_{i; \omega_i \in A} p_i$ .

**Remarque :** On ne peut pas avoir équi-probabilité dans le cas infini dénombrable car, pour qu'une série converge, il est nécessaire que son terme général tende vers 0 et si  $p_i = \lambda$  pour tout  $i$ , on doit avoir  $\lambda = 0$ ; mais alors  $\sum_i p_i = 0 \neq 1$ .

## 1.4 Événements indépendants

**Approche intuitive :** 2 événements sont indépendants si la réalisation de l'un est sans effet sur la réalisation de l'autre. En termes de probabilités, on a la définition suivante :

**Définition 1.5.** Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé.

1) Deux événements  $A$  et  $B$  de  $\mathcal{A}$  sont indépendants si  $P(A \cap B) = P(A) \times P(B)$ .

2) Une famille d'événements  $(A_n)_n$  est dite famille d'événements indépendants si, pour tout  $p \in \mathbb{N}^*$  et pour tout  $\{i_1, \dots, i_p\} \subset \mathbb{N}$ ,

$$P(A_{i_1} \cap \dots \cap A_{i_p}) = P(A_{i_1}) \times \dots \times P(A_{i_p}). \quad (*)$$

**Mises en garde :**

1- Ne pas confondre événements incompatibles ( $A \cap B = \emptyset$ ) et événements indépendants ( $P(A \cap B) = P(A)P(B)$ ).

2- Il faut vérifier (\*) pour toutes les sous-familles : en particulier, l'indépendance 2 à 2 d'événements n'implique pas leur indépendance.

Le concept d'indépendance est donc très délicat : ainsi, il sera parfois difficile de deviner ou de pressentir l'indépendance, qui devra donc être vérifiée par le calcul.

**Théorème 1.2.** Si  $(A, B)$  est un couple d'événements indépendants, il en est de même des couples  $(\bar{A}, B)$ ,  $(A, \bar{B})$  et  $(\bar{A}, \bar{B})$ .

*Démonstration.*

$$P(\bar{A})P(B) - P(\bar{A} \cap B) = (1 - P(A))P(B) - (P(B) - P(A \cap B)) = P(A \cap B) - P(A)P(B)$$

donc, si  $P(A \cap B) = P(A)P(B)$ , alors  $P(\bar{A} \cap B) = P(\bar{A})P(B)$ .

Comme  $A$  et  $B$  jouent des rôles symétriques, on a le même résultat pour  $(A, \bar{B})$ , puis, en remplaçant  $B$  par  $\bar{B}$ , pour  $(\bar{A}, \bar{B})$ .  $\square$

On peut généraliser ce résultat :

**Théorème 1.3.** Si  $(A_n)_n$  est une famille d'événements indépendants, il en est de même de  $(A'_n)_n$ , avec  $A'_n = A_n$  ou bien  $A'_n = \bar{A}_n$ .

*Démonstration.* Elle se fait par récurrence sur le nombre de complémentaires.  $\square$

## 1.5 Rappels d'analyse combinatoire

Lorsque l'univers  $\Omega$  d'une expérience est fini, on utilise l'équiprobabilité sur  $(\Omega, \mathcal{P}(\Omega))$  chaque fois qu'aucun événement simple n'a de privilège sur les autres. Dans ce cas, le calcul des probabilités se ramène donc au calcul du nombre d'éléments de  $\Omega$  et de ses sous-ensembles. L'analyse combinatoire est précisément l'ensemble des méthodes permettant de compter les éléments d'un ensemble.

**Définition 1.6.** Soit  $N_1, \dots, N_p$  des ensembles de cardinal fini. L'ensemble des  $p$ -uplets  $(y_1, \dots, y_p)$  où  $y_i \in N_i$  pour tout  $i \in \{1, \dots, p\}$  est appelé produit cartésien des  $N_i$ . Il est noté  $N_1 \times \dots \times N_p$ , ou bien  $\prod_{i=1}^p N_i$ .

Si  $N_1 = \dots = N_p = N$ , alors  $\prod_{i=1}^p N_i$  est noté  $N^p$ .

**Propriété 1.7.**  $\text{card}(N_1 \times \dots \times N_p) = \text{card}N_1 \times \dots \times \text{card}N_p$  et  $\text{card}(N^p) = (\text{card}N)^p$ .

On notera  $A_p, B_p, \dots$  des ensembles de cardinal  $p$ .

### 1.5.1 Tirages ordonnés avec remise (ou applications)

On note  $\mathcal{F}(E_p, F_n)$  l'ensemble des applications de  $E_p$  vers  $F_n$ .

**Théorème 1.4.**  $\text{card}\mathcal{F}(E_p, F_n) = n^p$ .

*Démonstration.* Soit  $E_p = \{e_1, \dots, e_p\}$ . A tout  $e_i$  de  $E_p$ , on associe, de façon unique  $f(e_i)$  dans  $F_n$ . La donnée de  $f$  équivaut à celle du  $p$ -uplet  $(f(e_1), \dots, f(e_p))$  de  $(F_n)^p$  et, d'après la propriété 2.1,  $\text{card}\mathcal{F}(E_p, F_n) = (\text{card}F_n)^p = n^p$ .  $\square$

*Exemple 2.1 :* Combien de “mots” (ayant un sens ou pas) de 5 lettres peut-on former ?

On prend 26 boules gravées de  $A$  à  $Z$  que l'on place dans une urne. On fait 5 tirages **successifs** (on veut un mot donc l'ordre des lettres a de l'importance), **avec remise** de la boule après chaque tirage (une lettre peut figurer plusieurs fois dans un mot).

L'ensemble des résultats possibles est l'ensemble des 5-uplets  $(u_1, \dots, u_5)$  avec  $u_i \in \{A, \dots, Z\}$ . On a alors  $\text{card}\Omega = 26^5$ .

### 1.5.2 Tirages ordonnés sans remise (ou injections)

On note  $\mathcal{I}(E_p, F_n)$  l'ensemble des injections de  $E_p$  vers  $F_n$ , lorsque  $n \geq p$ .

**Théorème 1.5.**  $\text{card}\mathcal{I}(E_p, F_n) = n(n-1) \dots (n-(p-1))$ .

*Démonstration.* Il est nécessaire que  $n \geq p$  car chaque élément de  $E_p$  doit avoir une image distincte.

La donnée de  $j$  dans  $\mathcal{I}(E_p, F_n)$  équivaut à celle du  $p$ -uplet  $(j(e_1), \dots, j(e_p))$  formé de  $p$  éléments **distincts** de  $F_n$  : il y a  $n$  choix possibles pour  $j(e_1)$  ; le choix de  $j(e_1)$  étant fait, il ne reste plus que  $n - 1$  choix possibles pour  $j(e_2), \dots$ , et, pour  $j(e_p)$ , il ne reste plus que  $n - (p - 1)$  choix ( $p - 1$  éléments de  $F_n$  ayant déjà été utilisés).

Ainsi, on a bien  $\text{card}\mathcal{I}(E_p, F_n) = n \times (n - 1) \times \dots \times (n - (p - 1))$ .  $\square$

*Exemple 2.2* : Combien de mots de 5 lettres peut-on former sans utiliser 2 fois la même lettre ?

On reprend les boules de l'exemple 2.1 : on fait 5 tirages **successifs** (il s'agit toujours d'un mot, donc l'ordre a toujours de l'importance) mais **on ne remet pas** les boules déjà tirées dans l'urne (afin de ne pas retomber sur la même lettre).

On a alors  $\text{card}\Omega = 26 \times 25 \times 24 \times 23 \times 22$ .

**Remarque** : Si  $n = p$ , les injections sont en fait des bijections et on a alors :

$$\text{card}\mathcal{I}(E_n, F_n) = n(n - 1) \times \dots \times 2 \times 1 = n!$$

Dans le cas général, ( $n \geq p$ ),  $\text{card}\mathcal{I}(E_p, F_n) = \frac{n!}{(n - p)!}$  que l'on note  $A_n^p$ .

### 1.5.3 Tirages non ordonnés sans remise (ou combinaisons)

**Définition 1.7.** Si  $n \geq p$ , on appelle coefficient binomial  $C_n^p$  (ou parfois  $\binom{n}{p}$ ), le nombre

$$\frac{n!}{p!(n - p)!}.$$

**Propriété 1.8.**

- 1)  $C_n^p = C_n^{n-p}$  ;
- 2)  $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$  (formule de Pascal)

*Démonstration.* 1)  $C_n^{n-p} = \frac{n!}{(n - p)!(n - (n - p))!} = \frac{n!}{(n - p)!p!} = C_n^p$ .

2)  $C_{n-1}^p + C_{n-1}^{p-1} = \frac{(n - 1)!}{p!(n - 1 - p)!} + \frac{(n - 1)!}{(p - 1)!(n - p)!} = \frac{(n - 1)!}{p!(n - p)!} (n - p + p) = \frac{n!}{p!(n - p)!} = C_n^p$ .

$\square$

**Définition 1.8.** Une  $p$ -combinaison de  $F_n$  est une partie de  $F_n$  à  $p$  éléments.

**Théorème 1.6.** Le nombre de  $p$ -combinaisons de  $F_n$  est  $C_n^p$ .

*Démonstration.* A toute injection  $j$  de  $\mathcal{I}(E_p, F_n)$ , on associe la partie  $\{j(e_1), \dots, j(e_p)\}$ . Il y a  $p!$  injections qui donnent la même partie  $\{j(e_1), \dots, j(e_p)\}$  (autant de façons que de classer les  $j(e_i)$  pour  $1 \leq i \leq p$ ). Pour  $n \geq p$ , le nombre de parties est donc  $\frac{A_n^p}{p!} = \frac{n!}{p!(n - p)!} = C_n^p$ .  $\square$

*Exemple 2.3 :* Combien de “mains” de 5 cartes différentes existe-t-il dans un jeu de 32 ?

Il y en a  $C_{32}^5$ , qui correspond au nombre de façons de choisir 5 cartes parmi 32 (l'ordre des cartes dans la main n'a pas d'importance!)

#### 1.5.4 Tirages non ordonnés avec remise (ou combinaisons avec répétitions)

**Définition 1.9.** Une *p-combinaison avec répétition* de  $F_n$  est une liste de  $p$  éléments de  $F_n$ , les répétitions étant autorisées et l'ordre dans la liste n'intervenant pas.

Par exemple  $[a, a, b, c, c, f]$  est une 6-c.a.r.

**Théorème 1.7.** Le nombre de *p-combinaisons avec répétitions* de  $F_n$  est  $C_{p+n-1}^{n-1}$ .

*Démonstration.* Il revient au même de considérer toutes les répartitions possibles de  $p$  boules identiques dans  $n$  tiroirs  $x_1, \dots, x_n$ . En effet, à chaque tiroir, on associe le nombre de boules qu'il contient. On va utiliser cette dernière modélisation pour démontrer le théorème, mais on va procéder à l'envers, c'est-à-dire aligner d'abord les  $p$  boules et placer ensuite les  $n - 1$  cloisons des tiroirs.

Pour placer la première cloison, on a  $p + 1$  choix possibles car les  $p$  boules délimitent  $p + 1$  espaces libres. Une fois placée la première cloison, on a  $p + 2$  choix pour la deuxième car la première cloison a partagé un espace libre en 2,  $\dots$ . Ainsi, pour placer la  $(n - 1)$ -ième et dernière cloison, on a  $p + n - 1$  choix.

On a donc  $(p + 1) \times \dots \times (p + n - 1)$  façons de placer les cloisons. Mais les  $n - 1$  cloisons étant identiques, l'ordre de placement n'intervient pas et une même configuration peut être obtenue de  $(n - 1)!$  façons différentes. Le nombre de configurations est donc :

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

Une autre façon plus rapide de procéder est de considérer que l'on a au total  $n - 1 + p$  objets à placer dont  $n - 1$  cloisons identiques et  $p$  boules identiques. Les places occupées par les  $n - 1$  cloisons dans un tel  $n - 1 + p$ -uplet d'objets déterminent une unique configuration et il y en a en tout  $C_{n-1+p}^{n-1} = C_{n-1+p}^p$ .  $\square$

*Exemple 2.4 :* Au jeu des “chiffres et des lettres”, combien y-a-t-il de tirages possibles pour les 9 lettres ?

Il s'agit de 9-c.a.r. de 26 éléments donc ce nombre de choix est  $C_{26+9-1}^9 = C_{34}^9$ .

#### 1.5.5 Partages d'ensembles

**Problème :** Etant donné  $p$  entiers positifs ou nuls,  $n_1, \dots, n_p$ , vérifiant  $n_1 + \dots + n_p = n$ , on cherche le nombre de partages de  $F_n$  en  $p$  parties  $A_1, \dots, A_p$  telles que  $\text{card}A_i = n_i$ .

**Théorème 1.8.** Le nombre de partages de  $F_n$  en  $p$  parties  $A_1, \dots, A_p$  telles que  $\text{card}A_i = n_i$  est  $\frac{n!}{n_1! \dots n_p!}$ .

*Démonstration.* On commence par choisir les éléments de  $A_1$  : il faut en choisir  $n_1$  parmi  $n$ , soit  $C_n^{n_1}$  choix. Puis, on choisit, parmi les  $n - n_1$  qui restent, les  $n_2$  éléments de  $A_2$ , soit  $C_{n-n_1}^{n_2}$  choix...

Une fois choisis les éléments de  $A_1, \dots, A_{p-1}$ , il ne reste plus que  $C_{n-n_1-\dots-n_{p-1}}^{n_p} = C_{n_p}^{n_p} = 1$  choix pour ceux de  $A_p$  (tous ceux qui restent).

Le nombre de partages est donc :

$$\begin{aligned} C_n^{n_1} C_{n-n_1}^{n_2} \cdots C_{n-n_1-\dots-n_{p-1}}^{n_p} &= \frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdots \frac{(n-n_1-\dots-n_{p-1})!}{n_p!0!} \\ &= \frac{n!}{n_1! \cdots n_p!} \end{aligned}$$

□

Exemple 2.5 : Au jeu “des chiffres et des lettres”, combien y-a-t-il de tirages de 9 lettres possibles contenant  $n_A$  fois la lettre  $A, \dots, n_Z$  fois la lettre  $Z$ , avec  $n_A + \dots + n_Z = 9$ ?

Le nombre de choix est donc  $\frac{9!}{n_A! \cdots n_Z!}$ .

## Chapitre 2

# VARIABLES ALÉATOIRES REELLES, LOIS CLASSIQUES SUR $\mathbb{R}$

Soit  $\mathcal{E}$  une expérience aléatoire et  $(\Omega, \mathcal{A}, P)$  l'espace probabilisé qui en rend compte. Il arrive très souvent qu'à chaque résultat de  $\mathcal{E}$ , on associe une valeur numérique (par exemple la somme des chiffres obtenus après le lancer de 2 dés); on définit ainsi une application  $X$  de  $\Omega$  vers  $\mathbb{R}$ . On sera amené à considérer l'ensemble des résultats  $\omega \in \Omega$  tel que  $X(\omega) = a$ , ou encore  $X(\omega) < a$ , ou encore  $X(\omega) \in [a, b[...$  On voudrait parler de la probabilité de telle ou telle situation. Or, dans un espace probabilisé, seuls les événements ont une probabilité (et la tribu n'est pas toujours formée de toutes les parties de  $\Omega$ ). Donc, pour pouvoir considérer la probabilité de ces ensembles, il faut que ceux-ci soient dans la tribu  $\mathcal{A}$ . Dans ce chapitre, après avoir présenté quelques généralités sur les variables aléatoires et leurs lois, on s'intéressera plus particulièrement aux variables aléatoires réelles.

### 2.1 Généralités

#### 2.1.1 Tribu engendrée par une application

Rappelons que si  $X$  est une application d'un espace  $\Omega$  dans un espace  $\Omega'$  et si  $\mathcal{A}' \subset \mathcal{P}(\Omega')$ , l'ensemble  $X^{-1}(\mathcal{A}')$  est le sous-ensemble de  $\Omega$ , appelé image réciproque de  $\mathcal{A}'$  par  $X$ , défini par :

$$X^{-1}(A') = \{\omega \in \Omega ; X(\omega) \in A'\} \text{ noté aussi } [X \in A']$$

En d'autres termes, il s'agit de l'ensemble des éléments de  $\Omega$  qui ont leur image par  $X$  dans  $A'$ .

**Propriété 2.1.** *Si  $X$  est une application de  $\Omega$  vers  $\Omega'$  et si  $\mathcal{A}'$  est une tribu de parties de  $\Omega'$ , alors*

$$\mathcal{B}_X = X^{-1}(\mathcal{A}') = \{X^{-1}(A') ; A' \in \mathcal{A}'\}$$

*est une tribu de parties de  $\Omega$ .*

*Démonstration.* •  $X^{-1}(\Omega') = \{\omega \in \Omega ; X(\omega) \in \Omega'\} = \Omega \in \mathcal{B}_X$ .

- Si  $A' \in \mathcal{A}'$ , alors

$$\overline{X^{-1}(A')} = \{\omega \in \Omega ; X(\omega) \notin A'\} = \{\omega \in \Omega ; X(\omega) \in \overline{A'}\} = X^{-1}(\overline{A'}) \in \mathcal{B}_X$$

car  $\overline{A'} \in \mathcal{A}'$ .

- Si  $(A'_i)_{i \in I}$  est une famille dénombrable d'événements de  $\mathcal{A}'$ , alors

$$\bigcup_{i \in I} X^{-1}(A'_i) = X^{-1}\left(\bigcup_{i \in I} A'_i\right) \in \mathcal{B}_X$$

car  $\bigcup_{i \in I} A'_i \in \mathcal{A}'$ . □

**Définition 2.1.** La tribu  $\mathcal{B}_X$  est appelée tribu engendrée par  $X$ .

**Propriété 2.2.** Si  $X$  est une application de  $\Omega$  vers  $\Omega'$ , si  $\mathcal{C}'$  est une famille de parties de  $\Omega'$  et si  $\mathcal{A}' = \sigma(\mathcal{C}')$  est la tribu engendrée par  $\mathcal{C}'$ , alors

$$\mathcal{B}_X = \sigma(X^{-1}(\mathcal{C}'))$$

c'est-à-dire  $X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}'))$ .

*Démonstration.* Admise (l'égalité s'obtient en fait par double inclusion). □

### 2.1.2 Variables aléatoires

**Définition 2.2.** Soit  $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ . On dit que  $X$  est une variable aléatoire (en abrégé v.a.), ou encore application mesurable sur  $(\Omega, \mathcal{A})$  et à valeurs dans  $(\Omega', \mathcal{A}')$ , si

$$\mathcal{B}_X \subset \mathcal{A}$$

c'est-à-dire si, pour tout  $A' \in \mathcal{A}'$ ,  $X^{-1}(A') \in \mathcal{A}$ .

**Propriété 2.3.** Si  $X$  est une application de  $\Omega$  vers  $\Omega'$ , et si  $\mathcal{A}' = \sigma(\mathcal{C}')$  est la tribu engendrée par  $\mathcal{C}'$ , où  $\mathcal{C}'$  est une famille de parties de  $\Omega'$ , alors

$$X \text{ est une variable aléatoire} \iff \text{pour tout } A' \in \mathcal{C}', X^{-1}(A') \in \mathcal{A}$$

c'est-à-dire qu'il suffit de considérer les images réciproques des éléments d'une partie engendrant la tribu  $\mathcal{A}'$ .

*Démonstration.*  $\Rightarrow$  Évident, puisque  $\mathcal{C}' \subset \mathcal{A}'$

$\Leftarrow$  D'après la propriété 2.2, on a  $\mathcal{B}_X = \sigma(X^{-1}(\mathcal{C}'))$ . Or  $\sigma(X^{-1}(\mathcal{C}')) \subset \mathcal{A}$  puisque la famille de parties  $X^{-1}(\mathcal{C}')$  est contenue dans la tribu  $\mathcal{A}$  par hypothèse, et que la tribu engendrée par cette famille (plus petite tribu la contenant) est ainsi contenue dans  $\mathcal{A}$ . On a donc bien  $\mathcal{B}_X \subset \mathcal{A}$  donc  $X$  est bien une variable aléatoire. □

**Propriété 2.4.** Si  $X$  est une v.a. de  $(\Omega, \mathcal{A})$  dans  $(\Omega', \mathcal{A}')$  et si  $Y$  est une v.a. de  $(\Omega', \mathcal{A}')$  dans  $(\Omega'', \mathcal{A}'')$ , alors  $Y \circ X$  est une v.a. de  $(\Omega, \mathcal{A})$  dans  $(\Omega'', \mathcal{A}'')$ .

*Démonstration.* Soit  $A'' \in \mathcal{A}''$ . On a :

$$\begin{aligned} (Y \circ X)^{-1}(A'') &= \{\omega \in \Omega ; Y \circ X(\omega) = Y(X(\omega)) \in A''\} \\ &= \{\omega \in \Omega ; X(\omega) \in A' = Y^{-1}(A'')\} = X^{-1}(A') = A \in \mathcal{A} \end{aligned}$$

car  $A' \in \mathcal{A}'$  ( $Y$  v.a.) et  $A = X^{-1}(A') \in \mathcal{A}$  ( $X$  v.a.), ce qui prouve bien que  $Y \circ X$  est une v.a. □

### 2.1.3 Loi de probabilité d'une variable aléatoire

**Propriété 2.5.** Si  $X$  est une v.a. définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ , à valeurs dans  $(\Omega', \mathcal{A}')$ , alors l'application  $P_X : A' \in \mathcal{A}' \mapsto P_X(A') = P([X \in A'])$  est une probabilité sur  $(\Omega', \mathcal{A}')$ .

*Démonstration.* • On a  $[X \in \Omega'] = \{\omega \in \Omega ; X(\omega) \in \Omega'\} = \Omega$  donc

$$P_X(\Omega') = P([X \in \Omega']) = P(\Omega) = 1.$$

• Soit  $(A'_i)_{i \in I}$  une famille dénombrable d'événements deux à deux disjoints de  $\mathcal{A}'$ . Alors  $([X \in A'_i])_{i \in I}$  est une famille dénombrable d'événements deux à deux disjoints de  $\mathcal{A}$  car  $[X \in A'_i] \cap [X \in A'_j] = [X \in A'_i \cap A'_j] = [X \in \emptyset] = \emptyset$ . On a donc

$$\begin{aligned} P_X\left(\bigcup_{i \in I} A'_i\right) &= P\left(X \in \bigcup_{i \in I} A'_i\right) = P\left(\bigcup_{i \in I} [X \in A'_i]\right) \\ &= \sum_{i \in I} P([X \in A'_i]) = \sum_{i \in I} P_X(A'_i) \end{aligned}$$

ce qui prouve bien que  $P_X$  est une probabilité. □

**Définition 2.3.** La probabilité  $P_X : A' \in \mathcal{A}' \mapsto P([X \in A'])$  est appelée loi de probabilité de  $X$  (on dit aussi image de  $P$  par  $X$ ).

### 2.1.4 Variables aléatoires réelles

On rappelle que, si  $X$  est une application de  $\Omega$  vers  $\Omega'$ , on désigne par  $X(\Omega)$  l'ensemble  $\{X(\omega) ; \omega \in \Omega\}$ ; c'est l'ensemble des valeurs susceptibles d'être prises par  $X$ .

Une variable aléatoire de  $(\Omega, \mathcal{A})$  dans  $(\Omega', \mathcal{A}')$  où  $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  (avec  $\mathcal{B}_{\mathbb{R}}$  tribu des boréliens de  $\mathbb{R}$ ) est dite réelle. On rappelle que  $\mathcal{B}_{\mathbb{R}} = \sigma(\{[a, b[ ; a, b \in \mathbb{R}\})$ .

En fait, on a aussi  $\mathcal{B}_{\mathbb{R}} = \sigma(\{I ; I \text{ intervalle de } \mathbb{R}\})$  puisque

$$\{[a, b[ ; a, b \in \mathbb{R}\} \subset \{I ; I \text{ intervalle de } \mathbb{R}\} \subset \mathcal{B}_{\mathbb{R}}.$$

On peut ainsi écrire de manière équivalente :

**Définition 2.4.** On appelle variable aléatoire réelle (en abrégé v.a.r.) toute application  $X$  de  $(\Omega, \mathcal{A})$  dans  $\mathbb{R}$  telle que, pour tout intervalle  $I$  de  $\mathbb{R}$ , on ait  $X^{-1}(I) \in \mathcal{A}$ .

On rappelle que  $X^{-1}(I) = \{\omega \in \Omega ; X(\omega) \in I\}$  est noté par commodité  $[X \in I]$ . On écrira ainsi  $[a < X \leq b]$  pour  $X^{-1}(]a, b])$ ;  $[X \leq x]$  pour  $X^{-1}(]-\infty, x])$  et  $[X = x]$  pour  $X^{-1}(\{x\})$ .

**Cas particulier important :** Si  $\mathcal{A} = \mathcal{P}(\Omega)$ , toute application de  $\Omega$  vers  $\mathbb{R}$  est une v.a.r.

**Théorème 2.1.** Une application  $X$  de  $(\Omega, \mathcal{A})$  dans  $\mathbb{R}$  est une variable aléatoire réelle si et seulement si, pour tout  $x \in \mathbb{R}$ , on a  $[X \leq x] \in \mathcal{A}$ .

*Démonstration.* On sait, par définition que  $X$  est une v.a.r. si  $[X \in I] \in \mathcal{A}$  pour tout  $I$  intervalle de  $\mathbb{R}$  donc, si  $X$  est une v.a.r., en appliquant la définition pour  $I_x = ]-\infty, x]$ , on a bien  $[X \in I_x] = [X \leq x] \in \mathcal{A}$ .

Réciproquement, tout intervalle  $I$  de  $\mathbb{R}$  s'écrit comme réunion, intersection, complémentaire, d'une suite d'intervalles de la forme  $] - \infty, x]$ .

Par exemple, si  $a < b$ ,  $]a, b[ = \left( \bigcup_{n \geq 1} \left] - \infty, b - \frac{1}{n} \right] \right) \cap \overline{] - \infty, a]}$ . On a donc aussi :

$$\begin{aligned} X^{-1}(]a, b[) &= X^{-1} \left( \left( \bigcup_{n \geq 1} \left] - \infty, b - \frac{1}{n} \right] \right) \cap \overline{] - \infty, a]} \right) \\ &= \left( \bigcup_{n \geq 1} X^{-1} \left( \left] - \infty, b - \frac{1}{n} \right] \right) \right) \cap \overline{X^{-1}(] - \infty, a])}. \end{aligned}$$

On a alors  $\bigcup_{n \geq 1} X^{-1} \left( \left] - \infty, b - \frac{1}{n} \right] \right) \in \mathcal{A}$  car  $\mathcal{A}$  est stable par union dénombrable ( $\mathcal{A}$  tribu) ;

$\overline{X^{-1}(] - \infty, a])} \in \mathcal{A}$  car  $\mathcal{A}$  est stable par complémentation ; et finalement  $X^{-1}(]a, b[) \in \mathcal{A}$  car  $\mathcal{A}$  est stable par intersection. Les autres cas se traiteraient de manière analogue.  $\square$

## 2.2 Fonctions de répartition

**Définition 2.5.** Soit  $X$  une v.a.r. définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On appelle fonction de répartition de  $X$  l'application  $F_X$  de  $\mathbb{R}$  sur  $\mathbb{R}$  définie par

$$F_X(x) = P([X \leq x]).$$

**Remarque :**  $F_X$  est bien définie sur  $\mathbb{R}$  car  $[X \leq x] \in \mathcal{A}$  pour tout  $x \in \mathbb{R}$ .

**Propriété 2.6.**

- 1) Pour tout  $x \in \mathbb{R}$ ,  $F_X(x) \in [0, 1]$  ;
- 2)  $F_X$  est croissante ;
- 3)  $F_X$  est continue à droite ;
- 4)  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  ;  $\lim_{x \rightarrow +\infty} F_X(x) = 1$  ;
- 5) pour tout  $(a, b) \in \mathbb{R}^2$ , si  $a < b$ , alors  $P([a < X \leq b]) = F_X(b) - F_X(a)$  ;
- 6)  $P([X = x]) = F_X(x) - F_X(x_-)$  ( $= 0$  si  $F_X$  est continue en  $x$ ).

*Démonstration.* **1)** Évident :  $F_X(x) = P([X \leq x]) \in [0, 1]$  car c'est la probabilité d'un événement.

**2) et 5)** Si  $x' \leq x$ , alors  $[X \leq x'] = [X \leq x] \cup [x < X \leq x']$  (réunion de 2 ensembles disjoints); donc  $P([X \leq x']) = P([X \leq x]) + P([x < X \leq x'])$ , c'est-à-dire

$$F_X(x') = F_X(x) + P([x < X \leq x']).$$

Comme  $P([x < X \leq x']) \geq 0$ , on a bien  $F_X(x) \leq F_X(x')$ , c'est-à-dire  $F_X$  croissante, et on obtient 5) en prenant  $x = a$  et  $x' = b$ .

**3)** On utilise le résultat d'analyse suivant :

Toute fonction monotone  $F$  admet en tout point  $x$  des limites à gauche (notée  $F(x_-)$ ) et à droite (notée  $F(x_+)$ );

si  $\lim_{n, x_n < x} x_n = x$ , alors  $\lim_{n \rightarrow +\infty} F(x_n) = F(x_-)$

si  $\lim_{n, x_n > x} x_n = x$ , alors  $\lim_{n \rightarrow +\infty} F(x_n) = F(x_+)$ .

Ici, on écrit  $]-\infty, x] = \bigcap_n \left] -\infty, x + \frac{1}{n} \right]$ . On a alors  $[X \leq x] = \bigcap_n D_n$  avec  $D_n = \left[ X \leq x + \frac{1}{n} \right]$ .

$(D_n)$  est une suite décroissante d'événements, donc, d'après la propriété 1.6 6), on a alors  $P\left(\bigcap_n D_n\right) = \lim_n P(D_n)$ , c'est-à-dire  $P([X \leq x]) = \lim_n P\left(\left[X \leq x + \frac{1}{n}\right]\right)$ , soit  $F_X(x) = \lim_n F_X\left(x + \frac{1}{n}\right) = F_X(x_+)$ , d'après le résultat d'analyse que l'on vient d'évoquer.

On a donc  $F_X(x) = F_X(x_+)$  et  $F_X$  est bien continue à droite.

**4)**  $\Omega = \bigcup_n [X \leq n]$  donc  $P(\Omega) = 1 = \lim_n P([X \leq n]) = \lim_n F_X(n) = \lim_{x \rightarrow +\infty} F_X(x)$  par limite croissante.

De même,  $\emptyset = \bigcap_n [X \leq -n]$  donc  $P(\emptyset) = 0 = \lim_n F_X(-n) = \lim_{x \rightarrow -\infty} F_X(x)$  par limite décroissante.

**6)**  $[X \leq x] = [X < x] \cup [X = x]$ . Or  $[X < x] = \bigcup_n \left[X \leq x - \frac{1}{n}\right]$  et, par propriété de limite croissante,

$$P([X < x]) = \lim_n P\left(\left[X \leq x - \frac{1}{n}\right]\right) = \lim_n F_X\left(x - \frac{1}{n}\right) = F_X(x_-).$$

Donc,  $F_X(x) = F_X(x_-) + P([X = x])$ . □

## 2.3 Variables aléatoires réelles discrètes

### 2.3.1 Définition, fonction de répartition et loi d'une v.a.r. discrète

**Définition 2.6.** Une v.a.r.  $X$  sur  $(\Omega, \mathcal{A}, P)$  est dite discrète si  $X(\Omega)$  est fini ou dénombrable.

On note  $X(\Omega) = \{x_1, \dots, x_n\}$  ou bien  $X(\Omega) = \{x_i; i \in \mathbb{N}\}$  et  $p_i = P([X = x_i])$ .

**Remarque :** On a, pour tout  $x \in \mathbb{R}$ ,  $[X \leq x] = \bigcup_{i: x_i \leq x} [X = x_i]$  et une application  $X$  de  $\Omega$  vers  $\mathbb{R}$  est donc une v.a.r. discrète à valeurs dans  $\{x_i; i \in \mathbb{N}\}$  si, pour tout  $i \in \mathbb{N}$ , on a  $[X = x_i] \in \mathcal{A}$ . Sa loi  $P_X$  est alors entièrement déterminée par les  $P([X = x_i]) = p_i$ . Ainsi, on peut reformuler dans ce cas :

**Définition 2.7.** On appelle loi de probabilité d'une v.a.r. discrète  $X$ , l'ensemble des couples  $(x_i, p_i)$ .

**Remarque :** On a  $p_i \geq 0$  et  $\sum_i p_i = 1$ . On notera également  $p_i = P_X(\{x_i\})$ .

**Fonction de répartition :**  $F_X(x) = \sum_{i; x_i \leq x} p_i$ .

Sur  $[x_{i-1}, x_i[$ , on a  $F_X(x) = F_X(x_{i-1})$  et  $F_X(x_i) - F_X(x_{i-1}) = P([X = x_i]) = p_i$  :

La fonction de répartition  $F_X$  d'une v.a.r. discrète  $X$  est ainsi une fonction en escalier croissante, présentant des sauts de  $p_i$  en chaque  $x_i$ .

### 2.3.2 Fonction d'une v.a.r. discrète

**Théorème 2.2.** Soit  $X$  une v.a.r. discrète et  $\varphi$  une fonction mesurable à valeurs réelles définie sur  $X(\Omega)$ ; alors  $Y = \varphi(X)$  est une v.a.r. discrète vérifiant  $Y(\Omega) = \varphi(X(\Omega))$  et, pour tout  $y \in Y(\Omega)$ , on a :

$$P([Y = y]) = \sum_{x_i; \varphi(x_i) = y} P([X = x_i]).$$

*Démonstration.* Soit, pour  $y \in Y(\Omega)$ ,  $I_y = \{x_i \in X(\Omega) ; \varphi(x_i) = y\}$ . On a alors

$$[Y = y] = \{\omega \in \Omega ; \varphi(X(\omega)) = y\} = \bigcup_{x_i \in I_y} \{\omega \in \Omega ; X(\omega) = x_i\} = \bigcup_{x_i \in I_y} [X = x_i].$$

D'où  $P([Y = y]) = \sum_{x_i \in I_y} P([X = x_i]) = \sum_{x_i; \varphi(x_i) = y} P([X = x_i])$ . □

### 2.3.3 Lois discrètes classiques

#### 1) Lois discrètes finies.

a) **Loi de Bernoulli  $\mathcal{B}(p)$  :**

$$X(\Omega) = \{0, 1\} ; P([X = 1]) = p ; P([X = 0]) = 1 - p.$$

Schéma théorique : On considère une expérience  $\mathcal{E}$  et un événement  $A$  lié à  $\mathcal{E}$  tel que  $P(A) = p$ . On effectue une fois  $\mathcal{E}$  et on appelle  $X$  le nombre de réalisation de  $A$ . On a donc  $X = 1$  si  $A$  est réalisé et  $X = 0$  si  $A$  n'est pas réalisé. d'où  $[X = 1] = A$  et  $[X = 0] = \bar{A}$ .  $X$  est bien une v.a.r. car  $A \in \mathcal{A}$  : on dit que  $X$  est la variable indicatrice de l'événement  $A$  et on note  $X = \mathbb{I}_A$ . On a bien  $X(\Omega) = \{0, 1\}$  ;  $P([X = 1]) = p$  ;  $P([X = 0]) = 1 - p$ .

b) **Loi binomiale  $\mathcal{B}(n, p)$  :**

$$X(\Omega) = \{0, \dots, n\} ; P([X = k]) = C_n^k p^k (1 - p)^{n-k} \text{ pour tout } k \in X(\Omega).$$

Schéma théorique : On considère l'expérience  $\mathcal{E}$  du a) et l'événement  $A$  lié à  $\mathcal{E}$  tel que  $P(A) = p$ . On effectue  $n$  fois  $\mathcal{E}$  et on appelle  $X$  le nombre de réalisations de  $A$  au cours des  $n$  épreuves. On a bien  $X(\Omega) = \{0, 1, \dots, n\}$ .

Du point de vue de  $A$ , le résultat des  $n$  expériences peut être représenté par un  $n$ -uplet formé de 0 et de 1. On a alors  $X = k$  si et seulement si le  $n$ -uplet est formé de  $k$  "1" et de  $n - k$  "0". Il y a  $C_n^k$  tels  $n$ -uplets (autant que de façons de placer exactement  $k$  "1" dans un  $n$ -uplet), et ils ont tous la même probabilité  $p^k(1 - p)^{n-k}$  (les expériences étant effectuées de façon indépendante, les probabilités se multiplient).

On a donc bien  $X(\Omega) = \{0, 1, \dots, n\}$  et, pour tout  $k \in X(\Omega)$ ,

$$P([X = k]) = C_n^k p^k (1 - p)^{n-k}.$$

**Remarque :** On vérifie que  $\sum_{k=0}^n P([X = k]) = 1$  : c'est la formule du binôme.

De plus, pour  $n = 1$ ,  $\mathcal{B}(1, p) = \mathcal{B}(p)$ .

**c) Loi hypergéométrique  $\mathcal{H}(n, M, N)$  :**

$$X(\Omega) = \{\max(0, n - N + M), \dots, \min(n, M)\}; P([X = k]) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \text{ pour } k \in X(\Omega).$$

Schéma théorique : Soit  $E$  un ensemble constitué de  $N$  objets dont  $M$  de type 1 et  $N - M$  de type 2. On effectue  $n$  tirages sans remise dans  $E$  ( $n \leq N$ ). Soit  $X$  le nombre d'objets de type 1 obtenus.

On ne peut avoir  $X = k$  que si  $0 \leq k \leq M$  et  $0 \leq n - k \leq N - M$ , c'est-à-dire  $0 \leq k \leq M$  et  $n - N + M \leq k \leq n$ . Les valeurs prises par  $X$  sont donc les entiers  $k$  tels que  $\max(0, n - N + M) \leq k \leq \min(n, M)$ .

Un élément  $\omega$  de  $[X = k]$  est constitué de  $k$  objets de type 1 parmi  $M$  et de  $n - k$  objets de type 2 parmi  $N - M$ . Il y a  $C_M^k$  façons de choisir les objets de type 1 et  $C_{N-M}^{n-k}$  objets de type 2. Or il y a en tout  $C_N^n$  façons de choisir  $n$  objets parmi  $N$ .

On a donc bien  $P([X = k]) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$  pour tout  $k \in \{\max(0, n - N + M), \dots, \min(n, M)\}$ .

**Remarque :** Si  $M \geq n$  et  $N - M \geq n$  (cas le plus fréquent), alors  $X(\Omega) = \{0, \dots, n\}$ .

**Attention :** Selon les ouvrages, l'ordre des paramètres dans la dénomination de la loi peut changer.

**d) Loi équiprobable  $\mathcal{U}(\{x_1, \dots, x_n\})$  :**

$$X(\Omega) = \{x_1, \dots, x_n\}; P([X = x_k]) = \frac{1}{n} \text{ pour tout } x_k \in X(\Omega).$$

## 2) Lois discrètes infinies.

**a) Loi géométrique sur  $\mathbb{N}^*$   $\mathcal{G}(p)$  (ou  $\mathcal{P}(1, p)$ ) :**

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

Schéma théorique : On considère toujours l'expérience  $\mathcal{E}$  du 1)a) et l'événement  $A$  lié à  $\mathcal{E}$  tel que  $P(A) = p$ . On répète maintenant  $\mathcal{E}$  dans les mêmes conditions jusqu'à ce que  $A$  soit réalisé. On note  $X$  le nombre d'épreuves effectuées.

On a bien  $X(\Omega) = \mathbb{N}^*$  et, si l'on note  $A_i$  l'événement "A est réalisé à la  $i$ -ème épreuve", on a alors  $[X = 1] = A_1$  d'où  $P([X = 1]) = p$  et, pour  $k \geq 2$ ,  $[X = k] = \overline{A_1} \cap \dots \cap \overline{A_{k-1}} \cap A_k$  d'où

$$P([X = k]) = P(\overline{A_1} \cap \dots \cap \overline{A_{k-1}} \cap A_k) = p(1-p)^{k-1}.$$

D'où, finalement,  $X(\Omega) = \mathbb{N}^*$  et, pour tout  $k \in \mathbb{N}^*$ ,  $P([X = k]) = p(1-p)^{k-1}$ .

**b) Loi de Poisson  $\mathcal{P}(\lambda)$  :**

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

**Remarque :** On a bien  $\sum_{k=0}^{+\infty} P([X = k]) = 1$ .

*Exemple :* Le nombre  $X$  d'événements aléatoires sur un intervalle  $T$  tels que des appels téléphoniques sur un central, où les arrivées des voitures à un péage d'autoroute suit une loi de Poisson.

On verra que  $\lambda$  représente le nombre moyen d'appels où d'arrivées pendant  $T$ .

## 2.4 Variables aléatoires réelles absolument continues

### 2.4.1 Définition, fonction de répartition, densité

On a vu que, si  $X$  est une v.a.r. sa fonction de répartition  $F_X : x \mapsto P([X \leq x])$  est une fonction croissante, continue à droite et présente un saut en  $x$  si  $P([X = x]) \neq 0$ . Elle vérifie de plus  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

On va s'intéresser ici à des v.a.r.  $X$  telles que  $P([X = x]) = 0$  pour tout  $x \in \mathbb{R}$ , c'est-à-dire telles que  $F_X$  soit continue sur  $\mathbb{R}$ .

**Remarque :** Dans ce cas,  $X(\Omega)$  ne peut pas être dénombrable. En effet, on aurait alors  $P(\Omega) = \sum_{x \in X(\Omega)} P([X = x]) = 0$ , ce qui contredit  $P(\Omega) = 1$ .

On peut commencer par rappeler un résultat d'analyse qui sera utile par la suite :

**Rappel :** Si  $f$  est intégrable sur  $[a, b]$ , alors, pour tout  $x \in [a, b]$ ,  $f$  est intégrable sur  $[a, x]$  et si on pose  $F(x) = \int_a^x f(t)dt$ , on a :

- i)  $F$  est continue sur  $[a, b]$  ;
- ii)  $F$  est dérivable à gauche (resp. à droite) en tout point  $x_0 \in ]a, b[$ , (resp.  $[a, b]$ ) où  $f$  admet une limite à gauche (resp. à droite) et  $F'_g(x_0) = f(x_{0-})$  (resp.  $F'_d(x_0) = f(x_{0+})$ ).

**Définition 2.8.** Une v.a.r.  $X$  de fonction de répartition  $F_X$  est dite absolument continue s'il existe une fonction à valeurs réelles  $f$  définie sur  $\mathbb{R}$ , appelée densité de  $X$  telle que :

- i)  $f(x) \geq 0$  pour tout  $x \in \mathbb{R}$  ;
- ii) pour tout  $B \in \mathcal{B}_{\mathbb{R}}$ ,  $f^{-1}(B) = \{x \in \mathbb{R} ; f(x) \in B\} \in \mathcal{B}_{\mathbb{R}}$  ;
- iii)  $\int_{-\infty}^{+\infty} f(t)dt$  existe et vaut 1 ;
- iv)  $F_X(x) = \int_{-\infty}^x f(t)dt$ .

**Remarque :** Une fonction vérifiant ii) est appelée *fonction borélienne*. En particulier, toute fonction continue ou “suffisamment continue” (par exemple, continue par morceaux) de  $\mathbb{R}$  dans  $\mathbb{R}$  est borélienne. Dans les exercices ou exemples, on ne considèrera que des fonctions  $f$  continues ou continues par morceaux : ainsi, le point ii) sera automatiquement vérifié.

**Conséquences très importante de la définition 2.8 :** Si  $X$  est absolument continue, on a, si  $a < b$  :

$$\begin{aligned} P([a < X < b]) &= P([a < X \leq b]) = P([a \leq X < b]) = P([a \leq X \leq b]) \\ &= F_X(b) - F_X(a) = \int_a^b f(t)dt. \end{aligned}$$

En effet,  $P([X = a]) = P([X = b]) = 0$  car  $F_X$  est continue, et  $F_X(b) - F_X(a) = \int_{-\infty}^b f(t)dt - \int_{-\infty}^a f(t)dt = \int_a^b f(t)dt$ .

De plus, en tout point  $t_0$  où la densité  $f$  est continue, on a que

$$F'_X(t_0) = f(t_0).$$

**Remarques :**

1) La connaissance de  $f$  détermine entièrement  $F_X$  mais la connaissance de  $F_X$  ne détermine pas  $f$  de façon unique : on peut modifier à son gré  $f$  sur un ensemble fini sans changer  $\int_{-\infty}^x f(t)dt$  (plus de précisions sur ce point sont données dans le chapitre sur le lien entre variables aléatoires et la théorie de l'intégration de Lebesgue). Dans la pratique,  $f$  sera choisie la moins discontinue possible. Par abus de langage, on appellera parfois “loi de  $X$ ” la densité de  $X$ .

2) Si  $X$  est une v.a.r. absolument continue, alors  $F_X$  est continue, mais la réciproque est fautive.

3) Il existe des v.a.r. qui ne sont ni discrètes, ni absolument continues (celles, par exemple, dont la fonction de répartition est strictement croissante et présente des sauts en certains points).

**Comment reconnaître la fonction de répartition d'une v.a.r. absolument continue ?**

Une fonction de répartition  $F$  est celle d'une v.a.r.  $X$  absolument continue si  $F$  est continue sur  $\mathbb{R}$ , dérivable sur  $\mathbb{R} \setminus I$ , où  $I$  est un sous-ensemble fini (ou dénombrable) de  $\mathbb{R}$ , si  $F'$  est continue sur  $\mathbb{R} \setminus I$  et si  $F'$  admet des limites à droite et à gauche en tout point de  $I$ .

### 2.4.2 Loi d'une fonction d'une v.a.r. absolument continue

Soit  $X$  une v.a.r. absolument continue (de densité  $f_X$  et de fonction de répartition  $F_X$ ) et  $\varphi$  une fonction réelle définie au moins sur un intervalle  $I$  contenant  $X(\Omega)$ . Pour déterminer la loi de  $Y = \varphi(X)$ , la méthode la plus générale consiste à exprimer sa fonction de répartition.

$$F_Y(y) = P([Y \leq y]) = P([\varphi(X) \leq y]) = P([X \in \varphi^{-1}(]-\infty, y])])$$

où  $\varphi^{-1}(]-\infty, y]) = \{x \in \mathbb{R} ; \varphi(x) \leq y\}$  que l'on note ici  $B(y)$ .

Le plus souvent,  $B(y)$  sera un intervalle ou bien une réunion d'intervalles disjoints et, comme  $(X \in \bigcup_k I_k) = \bigcup_k [X \in I_k]$ , il sera alors aisé d'exprimer  $F_Y$  à l'aide de  $F_X$ .

*Exemples :*

1) Si  $Y = aX + b$ , avec  $a \neq 0$ ,  $F_Y(y) = P([Y \leq y]) = P([aX + b \leq y]) = P([aX \leq y - b])$ .

• Si  $a > 0$ ,  $F_Y(y) = P\left(\left[X \leq \frac{y-b}{a}\right]\right) = F_X\left(\frac{y-b}{a}\right)$ . Si  $F_X$  est dérivable sauf éventuellement en quelques points, il en est de même pour  $F_Y$  par composition, et on a alors  $f_Y(y) = F'_Y(y) = \frac{1}{a}F'_X\left(\frac{y-b}{a}\right) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right)$ .

• Si  $a < 0$ ,  $F_Y(y) = P\left(\left[X \geq \frac{y-b}{a}\right]\right) = 1 - F_X\left(\frac{y-b}{a}\right)$  car  $F_X$  est continue. Ici aussi,  $F_Y$  est dérivable (sauf en quelques points) et, en utilisant  $(F \circ g)' = (F' \circ g) \times g'$ , on a alors  $f_Y(y) = F'_Y(y) = -\frac{1}{a}F'_X\left(\frac{y-b}{a}\right) = -\frac{1}{a}f_X\left(\frac{y-b}{a}\right)$ .

Finalement,  $f_{aX+b}(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$ .

2) Si  $Y = X^2$ , alors  $F_Y(y) = P([Y \leq y]) = P([X^2 \leq y])$ .

• Si  $y < 0$ , alors  $F_Y(y) = 0$  et  $f_Y = F'_Y = 0$  sur  $\mathbb{R}_-^*$ .  
 • Si  $y > 0$ , alors  $F_Y(y) = P([-\sqrt{y} \leq X \leq \sqrt{y}]) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$  et  $f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y}))$  dans ce cas.

Finalement,  $f_{X^2}(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})) \mathbb{1}_{]0,+\infty[}(y)$ .

3) Si  $Y = e^X$ , alors  $F_Y(y) = P([Y \leq y]) = P([e^X \leq y])$ .

• Si  $y \leq 0$ , alors  $F_Y(y) = 0$  et  $f_Y = F'_Y = 0$ .  
 • Si  $y > 0$ , alors  $F_Y(y) = P([X \leq \ln y]) = F_X(\ln y)$  car  $t \mapsto \ln t$  est une bijection strictement croissante de  $\mathbb{R}_+^*$  sur  $\mathbb{R}$ , puis  $f_Y(y) = F'_Y(y) = \frac{1}{y}f_X(\ln y)$ .

Finalement,  $f_{e^X}(y) = \frac{1}{y}f_X(\ln y) \mathbb{1}_{]0,+\infty[}(y)$ .

De façon plus générale, on pourra utiliser, lorsque  $\varphi$  est bijective, le résultat suivant :

**Théorème 2.3.** Soit  $X$  une v.a.r. absolument continue de densité  $f_X$  et  $\varphi$  une fonction mesurable à valeurs réelles de classe  $C^1$  (continue et dérivable de dérivée continue) sur un intervalle  $I$  contenant  $X(\Omega)$ , telle que  $\varphi'$  ne s'annule en aucun point de  $I$ ; alors  $Y = \varphi(X)$  est une v.a.r. absolument continue et elle admet la densité  $f_Y$  définie par :

$$f_Y(y) = \frac{f_X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|} \mathbb{1}_{\varphi(I)}(y).$$

*Démonstration.*  $\varphi'$  est continue et ne s'annule pas sur  $I$ ; elle y garde donc un signe constant.

On part de  $F_Y(y) = P([Y \leq y]) = P([\varphi(X) \leq y])$ .

- Si  $y \geq \sup \varphi(X(\Omega))$ , alors  $P([\varphi(X) \leq y])$  est toujours vérifié et  $F_Y(y) = 1$ ;
- Si  $y < \inf \varphi(X(\Omega))$ , alors  $P([\varphi(X) \leq y])$  n'est jamais vérifié et  $F_Y(y) = 0$ ;
- Si  $y \in [\inf \varphi(X(\Omega)), \sup \varphi(X(\Omega))]$ , alors  $y \in \varphi(I)$  car  $\varphi(X(\Omega)) \subset \varphi(I)$ .  
 → Si  $\varphi' > 0$ , alors  $\varphi$  et aussi  $\varphi^{-1}$  sont strictement croissantes sur  $I$ . Il vient donc

$$F_Y(y) = P([\varphi(X) \leq y]) = P([X \leq \varphi^{-1}(y)]) = F_X(\varphi^{-1}(y)).$$

$F_X$  est continue, et si elle est dérivable sauf éventuellement en un nombre fini ou dénombrable de points, comme  $\varphi^{-1}$  est dérivable, par composition,  $F_Y = F_X \circ \varphi^{-1}$  est dérivable sur  $\varphi(I)$ , sauf peut-être en un

nombre fini ou dénombrable de points et :

$$F'_Y(y) = (F_X \circ \varphi^{-1})'(y) = F'_X(\varphi^{-1}(y))(\varphi^{-1})'(y) = \frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

→ Si  $\varphi' < 0$ , alors  $\varphi$  et aussi  $\varphi^{-1}$  sont strictement décroissantes sur  $I$ . Il vient donc

$$F_Y(y) = P([\varphi(X) \leq y]) = P([X \geq \varphi^{-1}(y)]) = 1 - F_X(\varphi^{-1}(y)).$$

$F_X$  est continue, dérivable sauf éventuellement en un nombre fini ou dénombrable de points, et  $\varphi^{-1}$  est dérivable. D'où  $F_Y = 1 - F_X \circ \varphi^{-1}$  est dérivable sur  $\varphi(I)$ , sauf peut-être en un nombre fini ou dénombrable de points et :

$$F'_Y(y) = -(F_X \circ \varphi^{-1})'(y) = -F'_X(\varphi^{-1}(y))(\varphi^{-1})'(y) = -\frac{f_X(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}.$$

Donc, dans les deux cas,  $f_Y(y) = \frac{f_X(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|}$ . □

### 2.4.3 Lois absolument continues classiques

**Rappel :** Si  $\Delta$  est un sous-ensemble de  $\mathbb{R}$ , alors  $\mathbb{I}_\Delta(x) = \begin{cases} 1 & \text{si } x \in \Delta \\ 0 & \text{si } x \notin \Delta \end{cases}$ . La fonction  $\mathbb{I}_\Delta$  est appelée fonction indicatrice de  $\Delta$ .

**a) Loi uniforme  $\mathcal{U}([a, b])$  :**

loi de densité  $f$  définie par  $f(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$ .

On a dans ce cas,  $F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b[ \text{ , c'est-à-dire} \\ 1 & \text{si } x \geq b. \end{cases}$

$$F(x) = \frac{x-a}{b-a} \mathbb{I}_{[a,b[}(x) + \mathbb{I}_{[b,+\infty[}(x).$$

**b) Loi exponentielle  $\mathcal{E}(\lambda)$  :**

loi de densité  $f$  définie par  $f(x) = \lambda \exp(-\lambda x) \mathbb{I}_{]0,+\infty[}(x)$ .

On a dans ce cas,  $F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^x \lambda e^{-\lambda t} dt & \text{si } x \geq 0 \end{cases}$ ,

où  $\int_0^x \lambda e^{-\lambda t} dt = [-e^{-\lambda t}]_0^x = 1 - e^{-\lambda x}$ , c'est-à-dire

$$F(x) = (1 - e^{-\lambda x}) \mathbb{I}_{]0,+\infty[}(x).$$

**c) Loi Gamma  $\gamma(\lambda, a)$  (ou bien  $\Gamma(a, \lambda)$ ) :**

loi de densité  $f$  définie par  $f(x) = \frac{\lambda^a}{\Gamma(a)} \exp(-\lambda x) x^{a-1} \mathbb{I}_{]0,+\infty[}(x)$  où  $\Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} dt$ .

**d) Loi Beta  $B(a, b)$  :**

loi de densité  $f$  définie par  $f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{]0,1[}(x)$  où  $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ .

e) **Loi de Cauchy**  $\mathcal{C}(\alpha)$  :

loi de densité  $f$  définie par  $f(x) = \frac{\alpha}{\pi(\alpha^2 + x^2)}$ .

f) **Loi Normale**  $\mathcal{N}(m, \sigma^2)$  :

loi de densité  $f$  définie par  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$  ( $m$  est appelé la moyenne et  $\sigma$  l'écart-type).

**Théorème 2.4.** Une v.a.r.  $X$  a pour loi la loi normale  $\mathcal{N}(m, \sigma^2)$  si et seulement si  $X^* = \frac{X-m}{\sigma}$  est une v.a.r. de loi normale  $\mathcal{N}(0, 1)$ .

*Démonstration.* On a vu que  $f_{aX+b}(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$ .

Si  $X$  a pour loi la loi normale  $\mathcal{N}(m, \sigma^2)$ , alors  $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$  et alors, comme  $X^* = \frac{1}{\sigma}X - \frac{m}{\sigma}$ , on applique la formule avec  $a = \frac{1}{\sigma}$  et  $b = -\frac{m}{\sigma}$ . On a alors :

$$f_{X^*}(y) = \sigma f_X\left(\sigma\left(y + \frac{m}{\sigma}\right)\right) = \sigma f_X(\sigma y + m) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2},$$

et  $X^*$  suit bien la loi  $\mathcal{N}(0, 1)$ .

Réciproquement, si  $f_{X^*}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ , comme  $X = \sigma X^* + m$ , on a alors :

$$f_X(y) = \frac{1}{\sigma} f_{X^*}\left(\frac{y-m}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-m)^2}{2\sigma^2}}$$

et  $X$  suit bien la loi  $\mathcal{N}(m, \sigma^2)$ . □

**Conséquence :** Tout calcul de probabilité faisant intervenir la loi normale  $\mathcal{N}(m, \sigma^2)$  se ramène à un calcul faisant intervenir la loi normale  $\mathcal{N}(0, 1)$ .

On note  $\Phi$  la fonction de répartition d'une v.a.r. de loi  $\mathcal{N}(0, 1)$  :

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Comme on ne sait pas calculer  $\Phi$ , il existe des tables qui donnent des valeurs approchées de  $\Phi(x)$  pour tout  $x \in \mathbb{R}$ .

**Théorème 2.5.** Soit  $X$  une v.a.r. de loi normale  $\mathcal{N}(0, 1)$ , de fonction de répartition  $\Phi$ . Alors :

- 1) pour tout  $x \in \mathbb{R}$ ,  $\Phi(x) = 1 - \Phi(-x)$  ; en particulier  $\Phi(0) = \frac{1}{2}$  ;
- 2) pour tout  $x \in \mathbb{R}$ ,  $P(|X| \leq x) = 2\Phi(x) - 1$  et  $P(|X| \geq x) = 2(1 - \Phi(x))$ .

*Démonstration.* **1)** En faisant le changement de variable  $u = -t$ , on obtient :

$$\begin{aligned}\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt &= - \int_{+\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = \int_{-x}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du - \int_{-\infty}^{-x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du\end{aligned}$$

et on a bien  $\Phi(x) = 1 - \Phi(-x)$ . En particulier  $\Phi(0) = 1 - \Phi(-0) = 1 - \Phi(0)$ , d'où  $2\Phi(0) = 1$ .

**2)**  $P(|X| \leq x) = P(-x \leq X \leq x) = \Phi(x) - \Phi(-x) = 2\Phi(x) - 1$ .

De plus,  $P(|X| \geq x) = 1 - P(|X| < x) = 1 - (2\Phi(x) - 1) = 2(1 - \Phi(x))$ . □



## Chapitre 3

# ESPERANCE, VARIANCE ET MOMENTS DE VARIABLES ALÉATOIRES REELLES

### 3.1 Variables aléatoires réelles discrètes

Soit  $X$  une v.a.r. discrète définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On pose  $X(\Omega) = \{x_i ; i \in \mathbb{N}\}$  (le cas fini est similaire au cas infini dénombrable et présente moins de difficultés ; on traitera donc uniquement le cas dénombrable).

#### 3.1.1 Espérance

**Définition 3.1.** On dit que  $X$  possède une espérance si la série  $\sum_{n \geq 0} |x_n| P([X = x_n])$  converge ; on appelle alors espérance de  $X$  et on note  $\mathbb{E}(X)$  le nombre défini par :

$$\mathbb{E}(X) = \sum_{n \geq 0} x_n P([X = x_n]).$$

#### 3.1.2 Exemples d'espérances pour des lois discrètes classiques

On reprend ici les lois évoquées au Chapitre 2.

a) **Loi de Bernoulli**  $\mathcal{B}(p)$  (ou  $\mathcal{B}(1, p)$ ) :

$$X(\Omega) = \{0, 1\} ; P([X = 1]) = p ; P([X = 0]) = 1 - p ;$$

$$\boxed{\mathbb{E}(X) = p.}$$

En effet,  $\mathbb{E}(X) = 0 \times P([X = 0]) + 1 \times P([X = 1]) = p$ .

b) **Loi binomiale**  $\mathcal{B}(n, p)$  :

$$X(\Omega) = \{0, \dots, n\} \text{ et, pour } k \in X(\Omega), P([X = k]) = C_n^k p^k (1 - p)^{n-k} ;$$

$$\boxed{\mathbb{E}(X) = np.}$$

c) **Loi hypergéométrique**  $\mathcal{H}(n, M, N)$  :

$$X(\Omega) = \{\max(0, n - N + M), \dots, \min(n, M)\}; \text{ pour } k \in X(\Omega), P([X = k]) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$

$$\boxed{\mathbb{E}(X) = \frac{Mn}{N}.}$$

En effet,  $\mathbb{E}(X) = \sum_{k \in X(\Omega)} k \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$  avec, pour  $k \geq 1$ ,  $k C_M^k = \frac{M!}{(k-1)!(M-k)!} = M C_{M-1}^{k-1}$ ,  
 donc  $\mathbb{E}(X) = \frac{M}{C_N^n} \sum_{k \in X(\Omega) \setminus \{0\}} C_{M-1}^{k-1} C_{N-1-(M-1)}^{n-1-(k-1)} = \frac{M}{C_N^n} C_{N-1}^{n-1} = \frac{Mn!(N-n)!(N-1)!}{N!(n-1)!(N-n)!} = \frac{Mn}{N}$ .

(pour trouver  $\sum_{k \in X(\Omega) \setminus \{0\}} C_{M-1}^{k-1} C_{N-1-(M-1)}^{n-1-(k-1)} = C_{N-1}^{n-1}$ , on s'est servi de  $\sum_{k \in X(\Omega)} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} = 1$ ,  
 soit  $\sum_{k \in X(\Omega)} C_M^k C_{N-M}^{n-k} = C_N^n$  appliqué à  $k' = k - 1$ ,  $N' = N - 1$ ,  $M' = M - 1$  et  $n' = n - 1$ .)

d) **Loi géométrique sur  $\mathbb{N}^*$**   $\mathcal{G}(p)$  (ou  $\mathcal{P}(1, p)$ ) :

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

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

En effet, en posant  $q = 1 - p$ ,

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k \geq 1} k P([X = k]) = p \sum_{k \geq 1} k q^{k-1} \\ &= p \sum_{k \geq 1} \frac{d}{dq} (q^k) = p \frac{d}{dq} \left( \sum_{k \geq 0} q^k \right) = p \frac{d}{dq} \left( \frac{1}{1-q} \right) = \frac{p}{(1-q)^2} = \frac{1}{p}. \end{aligned}$$

e) **Loi de Poisson**  $\mathcal{P}(\lambda)$  :

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

$$\boxed{\mathbb{E}(X) = \lambda.}$$

En effet,

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k \geq 0} k P([X = k]) = \sum_{k \geq 1} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^{j+1}}{j!} = \lambda e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^\lambda = \lambda. \end{aligned}$$

3.1.3 Variance et moments d'ordre  $r$ 

**Définition 3.2.** On appelle moment d'ordre  $r$  de  $X$  le nombre  $m_r(X)$  défini par  $m_r(X) = \sum_{n \geq 0} x_n^r P([X = x_n])$  pourvu que cette série converge absolument.

ii) On appelle moment centré d'ordre  $r$  de  $X$  le nombre

$$\mu_r(X) = \sum_{n \geq 0} (x_n - \mathbb{E}(X))^r P([X = x_n]),$$

pourvu que cette série converge absolument.

iii) On appelle variance de  $X$  et on note  $\text{var}(X)$  le moment centré d'ordre 2 :  $\mu_2(X)$  i.e.

$$\text{var}(X) = \sum_{n \geq 0} (x_n - \mathbb{E}(X))^2 P([X = x_n])$$

iv) Si  $X$  admet une variance, on appelle écart-type de  $X$  le nombre  $\sigma_X$  défini par

$$\sigma_X = \sqrt{\text{var}(X)}.$$

3.1.4 Calcul de  $\mathbb{E}(\varphi(X))$ 

Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  une fonction mesurable à valeurs réelles quelconque. L'application  $Y = \varphi(X) : \Omega \rightarrow \mathbb{R}$  est encore une v.a.r. discrète. En effet,  $\varphi(X)(\Omega) = \varphi(X(\Omega))$  est dénombrable car c'est l'image par  $\varphi$  d'un ensemble dénombrable et, pour tout  $y_n \in \varphi(X(\Omega))$ , si  $I_n = \{i ; \varphi(x_i) = y_n \text{ où } x_i \in X(\Omega)\}$ , alors  $[\varphi(X) = y_n] = \bigcup_{i \in I_n} [X = x_i]$  est dans la tribu  $\mathcal{A}$ , comme réunion dénombrable d'éléments de  $\mathcal{A}$ .

**Théorème 3.1.** Soit  $\varphi$  une fonction continue par morceaux de  $\mathbb{R}$  dans  $\mathbb{R}$  telle que  $\varphi(X)$  admette une espérance. Alors :

$$\mathbb{E}(\varphi(X)) = \sum_{n \geq 0} \varphi(x_n) P([X = x_n]).$$

**Remarque :** Ce théorème est l'un des plus importants : il permet en effet de calculer  $\mathbb{E}(\varphi(X))$  sans connaître la loi de  $\varphi(X)$  mais en connaissant simplement la loi de  $X$ .

*Démonstration.* Par définition de  $\mathbb{E}(\varphi(X))$ , on a :

$$\begin{aligned} \mathbb{E}(\varphi(X)) &= \sum_n y_n P([\varphi(X) = y_n]) = \sum_n y_n P\left(\bigcup_{i \in I_n} [X = x_i]\right) \\ &= \sum_n y_n \sum_{i \in I_n} P([X = x_i]) = \sum_n \sum_{i \in I_n} \varphi(x_i) P([X = x_i]) \\ &= \sum_i \varphi(x_i) P([X = x_i]) \end{aligned}$$

car  $(I_n)_n$  est une partition de  $\mathbb{N}$  et  $\{x_i ; i \in \mathbb{N}\} = \bigcup_n \{x_i ; i \in I_n\}$ . □

Conséquences :

1)  $\boxed{\mathbb{E}(aX + b) = a\mathbb{E}(X) + b}$  si  $\mathbb{E}(X)$  existe. En effet, si  $X(\Omega) = \{x_n\}$ ,

$$\begin{aligned}\mathbb{E}(aX + b) &= \sum_n (ax_n + b)P([X = x_n]) = a \sum_n x_n P([X = x_n]) + b \sum_n P([X = x_n]) \\ &= a\mathbb{E}(X) + b.\end{aligned}$$

2)  $\boxed{\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2}$ . En effet, en posant  $m = \mathbb{E}(X)$ ,

$$\begin{aligned}\text{var}(X) &= \mathbb{E}((X - m)^2) = \sum_n (x_n - m)^2 P([X = x_n]) \\ &= \mathbb{E}(X^2 - 2mX + m^2) = \sum_n (x_n^2 - 2mx_n + m^2)P([X = x_n]) \\ &= \sum_n x_n^2 P([X = x_n]) - 2m \sum_n x_n P([X = x_n]) + m^2 \sum_n P([X = x_n]) \\ &= \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}$$

3)  $\boxed{\text{var}(aX + b) = a^2 \text{var}(X)}$ . En effet,

$$\begin{aligned}\text{var}(aX + b) &= \mathbb{E} \left( ((aX + b) - \mathbb{E}(aX + b))^2 \right) \\ &= \mathbb{E} \left( (aX + b - (a\mathbb{E}(X) + b))^2 \right) = \mathbb{E} \left( (aX + b - a\mathbb{E}(X) - b)^2 \right) \\ &= \mathbb{E} \left( a^2(X - \mathbb{E}(X))^2 \right) = a^2 \mathbb{E} \left( (X - \mathbb{E}(X))^2 \right) = a^2 \text{var}(X)\end{aligned}$$

### 3.1.5 Fonctions génératrices

La fonction génératrice est un outil de calcul permettant de calculer les moments d'une v.a.r. discrète à valeurs dans  $\mathbb{N}$ , et plus particulièrement l'espérance et la variance. On pose  $G_X(t) = \sum_{n \geq 0} t^n P([X = n])$ . En appliquant le théorème 3.1 à  $\varphi_t : x \mapsto t^x$ , on a donc  $G_X(t) = \mathbb{E}(t^X)$ .

Pour  $t = 1$ ,  $G_X(1) = 1$ , et, pour  $|t| \leq 1$ ,  $\sum_n |t|^n P([X = n]) \leq \sum_n P([X = n]) = 1$ . La série  $\sum_n t^n P([X = n])$  converge donc absolument pour  $|t| \leq 1$  et  $G_X(t)$  existe pour  $t \in [-1, 1]$ .

**Théorème 3.2.** Soit  $G_X(t) = \mathbb{E}(t^X) = \sum_{n \geq 0} t^n P([X = n])$ . Alors :

$$\mathbb{E}(X) = \lim_{t \rightarrow 1^-} G'_X(t) \text{ et } \text{var}(X) = \lim_{t \rightarrow 1^-} (G''_X(t) + G'_X(t) - (G'_X(t))^2).$$

*Démonstration.* A l'intérieur du domaine de convergence de la série  $\sum_n t^n P([X = n])$ , on peut dériver terme à terme et on a que pour  $|t| < 1$  :

$$G'_X(t) = \sum_{n \geq 1} nt^{n-1} P([X = n])$$

$$\begin{aligned}
G_X''(t) &= \sum_{n \geq 2} n(n-1)t^{n-2}P([X = n]) \\
&\vdots \\
G_X^{(k)}(t) &= \sum_{n \geq k} n(n-1)\cdots(n-k+1)t^{n-k}P([X = n]) \\
&\vdots
\end{aligned}$$

Sous réserve de convergence absolue, on peut passer à la limite dans la somme lorsque  $t$  tend vers 1 par valeurs inférieures (c'est-à-dire intervertir  $\lim$  et  $\sum$ ). On a alors, en appliquant le théorème 3.1 aux fonctions  $\varphi_k : x \mapsto x(x-1)\cdots(x-k+1)$  :

$$\begin{aligned}
\lim_{t \rightarrow 1} G_X'(t) &= \sum_{n \geq 1} nP([X = n]) = \mathbb{E}(X) \\
\lim_{t \rightarrow 1} G_X''(t) &= \sum_{n \geq 2} n(n-1)P([X = n]) = \mathbb{E}(X(X-1)) \\
&\vdots \\
\lim_{t \rightarrow 1} G_X^{(k)}(t) &= \sum_{n \geq k} n(n-1)\cdots(n-k+1)P([X = n]) = \mathbb{E}(X(X-1)\cdots(X-k+1)) \\
&\vdots
\end{aligned}$$

En particulier,

$$\begin{aligned}
\text{var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\
&= G_X''(1_-) + G_X'(1_-) - G_X'(1_-)^2.
\end{aligned}$$

□

### Application aux lois classiques :

a) **Loi binomiale**  $\mathcal{B}(n, p)$  :

$$X(\Omega) = \{0, \dots, n\} \text{ et, pour } k \in X(\Omega), P([X = k]) = C_n^k p^k (1-p)^{n-k}.$$

$$\boxed{G_X(t) = (pt + 1 - p)^n; \mathbb{E}(X) = np; \text{var}(X) = np(1-p).}$$

En effet,

$$G_X(t) = \sum_{k=0}^n t^k P([X = k]) = \sum_{k=0}^n t^k C_n^k p^k (1-p)^{n-k} = (pt + 1 - p)^n$$

$$G_X'(t) = np(pt + 1 - p)^{n-1} \quad ; \quad G_X''(t) = n(n-1)p^2(pt + 1 - p)^{n-2}.$$

$$G_X'(1_-) = np \quad ; \quad G_X''(1_-) = n(n-1)p^2$$

$$\text{et donc } \text{var}(X) = G_X''(1) + G_X'(1_-) - G_X'(1_-)^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p).$$

b) **Loi géométrique sur  $\mathbb{N}^*$**   $\mathcal{G}(p)$  :

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

$$G_X(t) = \frac{p}{1-p} \left( \frac{1}{1-t(1-p)} - 1 \right); \mathbb{E}(X) = \frac{1}{p}; \text{var}(X) = \frac{1-p}{p^2}.$$

En effet, en posant  $q = 1 - p$ , on a :

$$\begin{aligned} G_X(t) &= \sum_{k \geq 1} t^k P([X = k]) = \sum_{k \geq 1} t^k p q^{k-1} \\ &= \frac{p}{q} \left( \sum_{k \geq 1} (tq)^k \right) = \frac{p}{q} \left( \frac{1}{1-tq} - 1 \right) \end{aligned}$$

$$G'_X(t) = \frac{p}{q} \frac{q}{(1-tq)^2} = \frac{p}{(1-tq)^2} \text{ et } G''_X(t) = \frac{2pq}{(1-tq)^3}.$$

On a donc  $G'_X(1_-) = \frac{p}{(1-q)^2} = \frac{1}{p}$  et  $G''_X(1_-) = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}$  puis finalement

$$G''_X(1_-) + G'_X(1_-) - G'_X(1_-)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q + p - 1}{p^2} = \frac{q}{p^2}.$$

**c) Loi de Poisson  $\mathcal{P}(\lambda)$  :**

$X(\Omega) = \mathbb{N}$ , et, pour tout  $k \in X(\Omega)$ ,  $P([X = k]) = e^{-\lambda} \frac{\lambda^k}{k!}$ .

$$G_X(t) = e^{-\lambda} e^{\lambda t}; \mathbb{E}(X) = \lambda; \text{var}(X) = \lambda.$$

En effet,  $G_X(t) = \sum_{k \geq 0} t^k P([X = k]) = e^{-\lambda} \sum_{k \geq 0} t^k \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda t}$ .

$G'_X(t) = e^{-\lambda} \lambda e^{\lambda t}$ ,  $G''_X(t) = e^{-\lambda} \lambda^2 e^{\lambda t}$ , donc  $G'_X(1_-) = \lambda$ ,  $G''_X(1_-) = \lambda^2$  et

$$G''_X(1_-) + G'_X(1_-) - G'_X(1_-)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

## 3.2 Variables aléatoires réelles absolument continues

### 3.2.1 Espérance d'une v.a.r. absolument continue

**Définition 3.3.** Soit  $X$  une v.a.r. absolument continue, de densité  $f_X$ . Si

$$\int_{-\infty}^{+\infty} |t| f_X(t) dt < +\infty$$

alors  $X$  admet une espérance définie par

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt.$$

## 3.2.2 Variance et moments d'une v.a.r. absolument continue

**Définition 3.4.** Soit  $X$  une v.a.r. absolument continue, de densité  $f_X$ .

i) Soit  $r \in \mathbb{N}^*$ . On appelle moment d'ordre  $r$  de  $X$  et on note  $m_r(X)$  le réel défini par :

$$m_r(X) = \int_{-\infty}^{+\infty} t^r f_X(t) dt ; \text{ pourvu que cette intégrale converge absolument.}$$

ii) En particulier, le moment d'ordre 1 de  $X$  (s'il existe) est l'espérance de  $X$ ,

iii) Si  $X$  admet une espérance, on appelle moment centré d'ordre  $r$  de  $X$  le moment d'ordre  $r$ , s'il existe, de  $X - \mathbb{E}(X)$ . On le note  $\mu_r(X)$ .

iv) En particulier, si  $\mu_2(X)$  existe,  $\mu_2(X)$  est appelé variance de  $X$  et noté  $\text{var}(X)$  i.e.

$$\text{var}(X) = \int_{-\infty}^{+\infty} (t - \mathbb{E}(X))^2 f_X(t) dt$$

v) Si  $X$  admet une variance, on appelle écart-type de  $X$  le nombre  $\sigma_X$  défini par

$$\sigma_X = \sqrt{\text{var}(X)}.$$

**Propriété 3.1.**

1) Soit  $X$  une v.a.r. de densité  $f_X$  admettant une espérance et soit  $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ . Alors  $aX + b$  admet une espérance et  $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ .

2) Si  $\int_{-\infty}^{+\infty} x^2 f_X(x) dx$  converge, alors  $X^2$  admet une espérance et

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx.$$

*Démonstration.* 1) Dans le Chapitre 2, nous avons montré que si  $X$  a pour densité  $f_X$ , alors  $aX + b$  a pour densité  $f_{aX+b}$  définie par  $f_{aX+b}(t) = \frac{1}{|a|} f_X\left(\frac{t-b}{a}\right)$ .

$$\text{Donc, sous réserve d'existence, } \mathbb{E}(aX + b) = \int_{-\infty}^{+\infty} t f_{aX+b}(t) dt = \frac{1}{|a|} \int_{-\infty}^{+\infty} t f_X\left(\frac{t-b}{a}\right) dt.$$

On fait, dans la dernière intégrale, le changement de variable  $u = \frac{t-b}{a}$ .

$$\bullet \text{ Si } a > 0, \mathbb{E}(aX + b) = \frac{1}{a} \int_{-\infty}^{+\infty} (au + b) f_X(u) a du = a \int_{-\infty}^{+\infty} u f_X(u) du + b \int_{-\infty}^{+\infty} f_X(u) du = a\mathbb{E}(X) + b;$$

$$\bullet \text{ Si } a < 0, \mathbb{E}(aX + b) = \frac{1}{-a} \int_{+\infty}^{-\infty} (au + b) f_X(u) a du = a \int_{-\infty}^{+\infty} u f_X(u) du + b \int_{-\infty}^{+\infty} f_X(u) du = a\mathbb{E}(X) + b.$$

Comme  $\mathbb{E}(X)$  existe,  $\mathbb{E}(aX + b)$  existe aussi et  $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ .

2) Dans le Chapitre 2, nous avons montré que si  $X$  a pour densité  $f_X$ , alors  $X^2$  a pour densité  $f_{X^2}$  définie par  $f_{X^2}(t) = \frac{1}{2\sqrt{t}} (f_X(\sqrt{t}) + f_X(-\sqrt{t})) \mathbb{I}_{]0, +\infty[}(t)$ .

Donc, sous réserve d'existence,

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{+\infty} t f_{X^2}(t) dt = \int_0^{+\infty} \frac{\sqrt{t}}{2} (f_X(\sqrt{t}) + f_X(-\sqrt{t})) dt \\ &= \int_0^{+\infty} \frac{\sqrt{t}}{2} f_X(\sqrt{t}) dt + \int_0^{+\infty} \frac{\sqrt{t}}{2} f_X(-\sqrt{t}) dt.\end{aligned}$$

Dans la première intégrale, on fait le changement de variable  $u = \sqrt{t}$  et dans la deuxième intégrale, on fait le changement de variable  $u = -\sqrt{t}$  ( $t = u^2$ ,  $dt = 2udu$ ).

$$\mathbb{E}(X^2) = \int_0^{+\infty} u^2 f_X(u) du + \int_0^{-\infty} (-u^2) f_X(u) du = \int_{-\infty}^{+\infty} u^2 f_X(u) du.$$

□

### 3.2.3 Calcul de $\mathbb{E}(\varphi(X))$

Plus généralement, on admet le théorème suivant :

**Théorème 3.3.** *Soit  $X$  une v.a.r. absolument continue de densité  $f_X$  et  $\varphi$  une fonction à valeurs réelles continue par morceaux sur  $X(\Omega)$ , alors  $\varphi(X)$  est une v.a.r. absolument continue et, si elle admet une espérance, alors*

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x) dx.$$

**Conséquences :**

1)  $m_r(X) = \mathbb{E}(X^r)$

2)  $F_X(x) = \mathbb{E}(\mathbb{I}_{]_{-\infty}, x]}(X))$  (en effet,  $\mathbb{E}(\mathbb{I}_{]_{-\infty}, x]}(X)) = \int \mathbb{I}_{]_{-\infty}, x]}(t) f_X(t) dt = \int_{-\infty}^x f_X(t) dt$ ).

3) On retrouve  $\mathbb{E}(aX + b) = \int_{-\infty}^{+\infty} (at + b) f_X(t) dt = a\mathbb{E}(X) + b$ .

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

En effet, en posant  $\mathbb{E}(X) = m$ , on a

$$\begin{aligned}\text{var}(X) &= \mathbb{E}((X - m)^2) = \mathbb{E}(X^2 - 2mX + m^2) \\ &= \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.\end{aligned}$$

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

En effet,

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

### 3.2.4 Fonction caractéristique

**Définition 3.5.** Soit  $X$  une v.a.r. réelle. La fonction caractéristique de  $X$  est la fonction à valeurs complexes définie par

$$\forall t \in \mathbb{R}, \phi_X(t) = \mathbb{E}(e^{itX}).$$

Notons que si  $X$  admet pour densité  $f_X$ , alors la fonction caractéristique  $\phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx$  est la transformée de Fourier de  $f_X$ . Par la formule d'inversion de Fourier, on a donc que

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t) dt.$$

La loi d'une v.a.r. réelle à densité est donc entièrement déterminée par sa fonction caractéristique. Cette fonction a des propriétés analogues à celles des fonctions génératrices de v.a. discrètes, comme le montre (par exemple) le résultat suivant.

**Propriété 3.2** (admis). Soit  $X$  une v.a.r. réelle. Alors  $X$  admet un moment d'ordre  $r \in \mathbb{N}$  si et seulement si sa fonction caractéristique  $\phi_X$  est dérivable  $r$  fois en  $t = 0$ , et on a que

$$\mathbb{E}(X^r) = (-i)^r \frac{d^r \phi_X(t)}{dt^r} \Big|_{t=0}.$$

### 3.2.5 Application aux lois classiques

a) **Loi uniforme**  $\mathcal{U}([a, b])$  :

Loi de densité  $f_X$  définie par  $f_X(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x)$ .

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt = \int_a^b \frac{1}{b-a} t dt = \frac{1}{b-a} \left[ \frac{t^2}{2} \right]_a^b = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{b+a}{2}.$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt = \int_a^b \frac{1}{b-a} t^2 dt = \frac{1}{b-a} \left[ \frac{t^3}{3} \right]_a^b = \frac{1}{3} \frac{b^3 - a^3}{b-a} = \frac{b^2 + ab + a^2}{3}.$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{b^2 + ab + a^2}{3} - \frac{b^2 + 2ab + a^2}{4} = \frac{b^2 - 2ab + a^2}{12} = \frac{(b-a)^2}{12}.$$

$$\mathbb{E}(X) = \frac{a+b}{2} \text{ et } \text{var}(X) = \frac{(a-b)^2}{12}.$$

b) **Loi exponentielle**  $\mathcal{E}(\lambda)$  :

Loi de densité  $f_X$  définie par  $f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{]0,+\infty[}(x)$ .

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt = \int_0^{+\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda} \int_0^{+\infty} u e^{-u} du = \frac{1}{\lambda} \Gamma(2) = \frac{1}{\lambda};$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt = \int_0^{+\infty} t^2 \lambda e^{-\lambda t} dt = \frac{1}{\lambda^2} \int_0^{+\infty} u^2 e^{-u} du = \frac{1}{\lambda^2} \Gamma(3) = \frac{2}{\lambda^2};$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

$$\mathbb{E}(X) = \frac{1}{\lambda} \text{ et } \text{var}(X) = \frac{1}{\lambda^2}.$$

c) **Loi Gamma**  $\gamma(\lambda, a)$  :

Loi de densité  $f_X$  définie par  $f_X(x) = \frac{\lambda^a}{\Gamma(a)} e^{-\lambda x} x^{a-1} \mathbb{I}_{]0, +\infty[}(x)$ , où  $\Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} dt$ .

$X$  admet des moments de tous ordres car  $\lim_{t \rightarrow +\infty} t^2 (t^r f_X(t)) = 0$  donc  $t^r f_X(t) \leq \frac{1}{t^2}$  pour  $t \geq t_0$ .

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f_X(t) dt = \int_0^{+\infty} t \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1} dt = \frac{1}{\lambda \Gamma(a)} \int_0^{+\infty} u^a e^{-u} du = \frac{1}{\lambda} \frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\lambda};$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f_X(t) dt = \int_0^{+\infty} t^2 \frac{\lambda^a}{\Gamma(a)} e^{-\lambda t} t^{a-1} dt = \frac{1}{\lambda^2 \Gamma(a)} \int_0^{+\infty} u^{a+1} e^{-u} du = \frac{1}{\lambda^2} \frac{\Gamma(a+2)}{\Gamma(a)} = \frac{a(a+1)}{\lambda};$$

$$\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{a(a+1)}{\lambda^2} - \frac{a^2}{\lambda^2} = \frac{a}{\lambda^2}.$$

$$\mathbb{E}(X) = \frac{a}{\lambda} \text{ et } \text{var}(X) = \frac{a}{\lambda^2}.$$

d) **Loi Normale**  $\mathcal{N}(m, \sigma^2)$  :

Loi de densité  $f_X$  définie par  $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{\sigma^2}\right)$ .

$X$  admet des moments de tous ordres car  $\lim_{t \rightarrow +\infty} t^2 (t^r f_X(t)) = 0$  donc  $t^r f_X(t) \leq \frac{1}{t^2}$  pour  $t \geq t_0$ .

On a vu au Chapitre 3 que, si une v.a.r.  $X$  a pour loi  $\mathcal{N}(m, \sigma^2)$ , alors  $X^* = \frac{X-m}{\sigma}$  est une v.a.r. de loi normale  $\mathcal{N}(0, 1)$ .

$$\mathbb{E}(X^*) = \int_{-\infty}^{+\infty} t f_{X^*}(t) dt = \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0 \text{ (intégrale d'une fonction impaire sur } \mathbb{R} \text{)}$$

$$\mathbb{E}(X^{*2}) = \int_{-\infty}^{+\infty} t^2 f_{X^*}(t) dt = 2 \int_0^{+\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \text{ (intégrale d'une fonction paire sur } \mathbb{R} \text{)}.$$

On fait une intégration par parties en posant  $u' = t e^{-\frac{t^2}{2}}$  et  $v = t$  ( $u = -e^{-\frac{t^2}{2}}$  et  $v' = 1$ ).

$$2 \int_0^A t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{2}{\sqrt{2\pi}} \left( \left[ -t e^{-\frac{t^2}{2}} \right]_0^A + \int_0^A e^{-\frac{t^2}{2}} dt \right) = -\frac{2}{\sqrt{2\pi}} A e^{-\frac{A^2}{2}} + \frac{2}{\sqrt{2\pi}} \int_0^A e^{-\frac{t^2}{2}} dt \rightarrow$$

1 quand  $A \rightarrow +\infty$  car  $t e^{-\frac{t^2}{2}} \rightarrow 0$  et  $\frac{2}{\sqrt{2\pi}} \int_0^A e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-A}^A e^{-\frac{t^2}{2}} dt \rightarrow 1$ .

On a donc  $\mathbb{E}(X^*) = 0$ ,  $\mathbb{E}(X^{*2}) = 1$  et  $\text{var}(X^*) = \mathbb{E}(X^{*2}) - \mathbb{E}(X^*)^2 = 1$ .

Or  $X = \sigma X^* + m$ , donc  $\mathbb{E}(X) = \sigma \mathbb{E}(X^*) + m$  et  $\text{var}(X) = \sigma^2 \text{var}(X^*)$  d'où :

$$\mathbb{E}(X) = m \text{ et } \text{var}(X) = \sigma^2.$$

## Chapitre 4

# INTEGRALE DE LEBESGUE ET VARIABLES ALEATOIRES

Ce chapitre propose de présenter de façon succincte comment la théorie des probabilités s'appuie sur une notion d'intégration développée par Henri Lebesgue au début du 20ème siècle. L'intégrale de Lebesgue est un outil plus puissant que l'intégrale de Riemann qui s'appuie sur la théorie de la mesure. L'objectif de ce chapitre est d'illustrer comment le formalisme de l'intégrale de Lebesgue permet de traiter dans un cadre unifié à la fois les variables aléatoires discrètes et les variables aléatoires continues (à densité). Pour un cours beaucoup plus détaillé sur la théorie des probabilités et ses liens avec l'intégrale de Lebesgue et la théorie de la mesure, nous renvoyons le lecteur intéressé vers le livre suivant :

- Barbé Philippe & Ledoux Michel (2007), Probabilité (L3M1), EDP Sciences.

Dans tout le chapitre, le triplet  $(\Omega, \mathcal{A}, P)$  désigne un espace probabilisé, et  $X$  une variable aléatoire réelle (v.a.r) c'est à dire une application mesurable de  $(\Omega, \mathcal{A})$  à valeurs dans  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . Rappelons que la loi de probabilité de  $X$  (appelée aussi mesure image) est la probabilité  $P_X : \mathcal{B}_{\mathbb{R}} \rightarrow [0, 1]$  définie par  $P_X(B) = P(X \in B)$  pour tout  $B \in \mathcal{B}_{\mathbb{R}}$ .

Dans les chapitres précédents, la distinction entre variables aléatoires discrètes et continues a été faite en particulier au travers de l'écriture et des propriétés de leur fonction de répartition  $F_X(x) = P_X([-\infty, x])$ . Les notations que nous avons utilisées sont rappelées dans le tableau ci-dessous :

v.a.r. discrète	v.a.r. à densité $f_X$
$X(\Omega) = \{x_n\}_{n \in \mathbb{N}}$	$X(\Omega) \subset \mathbb{R}$
$F_X(x) = P([X \leq x]) = \sum_{x_n \leq x} P([X = x_n])$	$F_X(x) = P([X \leq x]) = \int_{-\infty}^x f_X(t) dt$
$\mathbb{E}(X) = \sum_{n \in \mathbb{N}} x_n P([X = x_n])$	$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$

Jusqu'à présent, dans le cas d'une variable aléatoire à densité, les intégrales  $\int_{-\infty}^x f_X(t) dt$  et  $\int_{-\infty}^{+\infty} t f_X(t) dt$  ont été (implicitement) définie par rapport à la notion classique d'intégration (au

sens de Riemann). Soit  $n \in \mathbb{N}$ , et introduisons la mesure de dirac  $\delta_{x_n} : \mathcal{B}_{\mathbb{R}} \rightarrow [0, 1]$  définie par

$$\delta_{x_n}(B) = \begin{cases} 1 & \text{si } x_n \in B \\ 0 & \text{si } x_n \notin B, \end{cases}$$

pour tout  $B \in \mathcal{B}_{\mathbb{R}}$ . On peut alors remarquer que si  $X$  est une v.a. discrète telle que  $X(\Omega) = \{x_n\}_{n \in \mathbb{N}}$  alors sa fonction de répartition peut s'écrire sous la forme

$$F_X(x) = \sum_{x_n \leq x} p_n = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}([-\infty, x]),$$

avec  $p_n = P([X = x_n])$ . Dans ce chapitre, nous allons voir (en particulier) comment le formalisme de l'intégrale de Lebesgue permet d'écrire de façon unifiée la fonction de répartition d'une v.a. qu'elle soit à densité ou bien discrète.

## 4.1 Construction de l'intégrale de Lebesgue

Nous donnons ici une présentation rapide de la construction de l'intégrale de Lebesgue.

### 4.1.1 Espace mesuré

**Définition 4.1.** Soit  $\Omega$  un espace muni d'une tribu  $\mathcal{A}$ . On appelle **mesure (positive)** sur  $(\Omega, \mathcal{A})$  toute application  $\mu$  de  $\mathcal{A}$  vers  $\mathbb{R}_+$  telle que :

i)  $\mu(\emptyset) = 0$  ;

ii) pour toute suite d'événements  $A_n \in \mathcal{A}$ , incompatibles deux à deux, on a :

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mu(A_n) \left( = \lim_{N \rightarrow +\infty} \sum_{n=0}^N \mu(A_n) \right).$$

Le triplet  $(\Omega, \mathcal{A}, \mu)$  est appelé *espace mesuré*. Notons qu'une probabilité  $P$  est le cas particulier d'une mesure positive prenant ses valeurs dans  $[0, 1]$  et qui vérifie  $P(\Omega) = 1$ . Des exemples importants de mesures sont les suivants.

**Définition 4.2.** Soit  $w_0 \in \Omega$  un élément fixé. La **mesure de Dirac** (ou *masse de Dirac*) au point  $w_0$  est la mesure définie par

$$\forall A \in \mathcal{A}, \delta_{w_0}(A) = \begin{cases} 1 & \text{si } w_0 \in A \\ 0 & \text{si } w_0 \notin A. \end{cases}$$

**Définition 4.3.** Soit  $(w_n)_{n \geq 1}$  une suite points de  $\Omega$ . La **mesure de dénombrement** associée à cette suite est la mesure définie par

$$\mu = \sum_{n \geq 1} \delta_{w_n}.$$

Nous verrons par la suite que les mesures de Dirac et les mesures de dénombrement permettent de caractériser les lois de variables aléatoires discrètes.

Un autre exemple est celui de la construction d'une mesure  $\mu$  sur  $\Omega = \mathbb{R}^d$  muni de la tribu des boréliens  $\mathcal{A} = \mathcal{B}_{\mathbb{R}^d}$ . L'exemple essentiel pour ce cours est celui de la mesure de Lebesgue  $\lambda$  qui est définie à partir de ses valeurs sur l'ensemble des pavés  $\otimes_{j=1}^d ]a_j, b_j]$  de  $\mathbb{R}^d$  par

$$\lambda \left( \otimes_{j=1}^d ]a_j, b_j] \right) = \prod_{j=1}^d (b_j - a_j).$$

Dans le cas particulier de  $d = 1$ , on a que  $\lambda(]a, b]) = b - a$  est la longueur de l'intervalle  $]a, b]$ . La mesure de Lebesgue restreinte à un intervalle permet de caractériser les variables aléatoires continues de loi uniforme.

#### 4.1.2 Intégrale d'une fonction au sens de Lebesgue

Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré, et  $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une application mesure. Nous souhaitons pouvoir donner un sens à l'intégrale de  $h$  par rapport à la mesure  $\mu$  qui sera notée

$$\int_{\Omega} h(w) d\mu(w).$$

La théorie de Lebesgue peut bien sûr se généraliser à l'intégration de fonctions  $h$  à valeurs dans d'autres espaces que  $\mathbb{R}$ . Toutefois, pour des raisons de simplicité, cette (rapide) présentation sera restreinte au cas de fonctions à valeurs réelles.

Soit  $A \in \mathcal{A}$  et  $\mathbb{1}_A : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  l'application définie par  $\mathbb{1}_A(w) = 1$  si  $w \in A$  et  $\mathbb{1}_A(w) = 0$  si  $w \notin A$ , appelée fonction indicatrice de l'ensemble  $A$ . Pour construire l'intégrale de Lebesgue, on définit tout d'abord l'intégrale de la somme d'indicatrices d'événements.

**Définition 4.4.** *L'intégrale de la fonction indicatrice  $\mathbb{1}_A$  par rapport à la mesure  $\mu$  est*

$$\int_{\Omega} \mathbb{1}_A(w) d\mu(w) = \mu(A).$$

**Définition 4.5.** *Soit  $\alpha_1, \dots, \alpha_n$  des réels positifs,  $A_1, \dots, A_n$  des événements de  $\mathcal{A}$  et posons*

$$h_n(w) = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}(w).$$

*L'application  $h_n$  s'appelle une **fonction étagée positive**, et l'intégrale de la fonction  $h_n$  par rapport à la mesure  $\mu$  est définie par*

$$\int_{\Omega} h_n(w) d\mu(w) = \sum_{j=1}^n \alpha_j \mu(A_j).$$

Ensuite pour définir l'intégrale d'une fonction mesurable  $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  à **valeurs positive**, on se sert de la propriété que  $h$  est la limite d'une suite croissante  $(h_n)_{n \geq 1}$  de fonctions

étagées positives telles que

$$\forall w \in \Omega, h(w) = \lim_{n \rightarrow +\infty} h_n(w) \text{ (convergence simple).}$$

**Définition 4.6.** Soit  $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction mesurable à valeurs positives. L'intégrale de la fonction  $h$  par rapport à la mesure  $\mu$  est définie par

$$\int_{\Omega} h(w) d\mu(w) = \lim_{n \rightarrow +\infty} \int_{\Omega} h_n(w) d\mu(w),$$

où  $(h_n)_{n \geq 1}$  est une suite croissante de fonctions convergeant simplement vers  $f$ .

On peut maintenant définir l'intégrale d'une fonction  $h$  à valeurs positives ou négatives.

**Définition 4.7.** Soit  $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction mesurable. La fonction  $h$  est dite intégrable si

$$\int_{\Omega} |h(w)| d\mu(w) < +\infty,$$

et dans ce cas l'intégrale de la fonction  $h$  par rapport à la mesure  $\mu$  est définie par

$$\int_{\Omega} h(w) d\mu(w) = \int_{\Omega} \max(h(w), 0) d\mu(w) - \int_{\Omega} (-\min(h(w), 0)) d\mu(w).$$

En général, il peut être très compliqué de calculer l'intégrale de Lebesgue d'une fonction mesurable. Toutefois, dans le cas où  $\mu = \lambda$  est la mesure de Lebesgue et si  $h : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction continue sur un intervalle  $[a, b]$  telle que  $h(w) = 0$  si  $w \notin [a, b]$ , alors l'intégrale de Lebesgue de  $h$  par rapport à la mesure  $\lambda$  coïncide avec l'intégrale classique de Riemann i.e.

$$\int_{\Omega} h(w) d\mu(w) = \int_a^b f(x) dx.$$

Il est également aisé de calculer l'intégrale d'une fonction par rapport à une mesure de Dirac ou à une mesure de dénombrement.

**Propriété 4.1** (admis). Soit  $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction mesurable. Soit  $w_0 \in \mathcal{A}$  et  $\delta_{w_0}$  la masse de Dirac au point  $w_0$ . Alors,

$$\int_{\Omega} h(w) d\delta_{w_0}(w) = h(w_0).$$

**Propriété 4.2** (admis). Soit  $h : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction mesurable. Soit  $(w_n)_{n \geq 1}$  une suite points de  $\Omega$ ,  $(\alpha_n)_{n \geq 1}$  une suite de réels positifs, et  $\mu_n$  la **mesure discrète**

$$\mu_n = \sum_{n \geq 1} \alpha_n \delta_{w_n}$$

Alors,

$$\int_{\Omega} h(w) d\mu_n(w) = \sum_{n \geq 1} \alpha_n h(w_n).$$

On peut donc remarquer qu'intégrer une fonction mesurable  $h$  par rapport à la mesure discrète  $\mu_n$  consiste simplement à effectuer la somme pondérée des valeurs de la fonction  $h$  aux points  $w_n$  affectés des poids  $\alpha_n$ . L'intérêt très pratique de la théorie de l'intégration de Lebesgue est de proposer un formalisme unifié pour traiter à la fois de sommes et d'intégrales.

### 4.1.3 Mesures définies par des densités

**Propriété 4.3** (admis). Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré et  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction à valeurs positives. L'application  $\nu : \mathcal{A} \rightarrow \mathbb{R}_+$  définie par

$$\forall A \in \mathcal{A}, \nu(A) = \int_A f(w) d\mu(w)$$

est une mesure (positive) dite **mesure de densité  $f$  par rapport à  $\mu$** .

Si  $\nu$  est une mesure de densité  $f$  par rapport à  $\mu$ , alors pour toute fonction  $h$  intégrable par rapport à  $\nu$ , on a que

$$\int h(w) d\nu(w) = \int h(w) f(w) d\mu(w).$$

Notons que si  $\nu_n = \sum_{n \geq 1} \alpha_n \delta_{w_n}$  est une mesure discrète, alors

$$\int h(w) d\nu_n(w) = \sum_{n \geq 1} \alpha_n h(w_n) = \int h(w) f(w) d\mu_n(w).$$

où  $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  est une application mesurable telles que  $f(w_n) = \alpha_n$  pour tout  $n \geq 1$  et  $\mu_n = \sum_{n \geq 1} \delta_{w_n}$ . Ainsi, **toute mesure discrète est une mesure à densité par rapport à la mesure de dénombrement  $\mu_n$** .

## 4.2 Loi de probabilité d'une variable aléatoire

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. On rappelle qu'une v.a. réelle (continue ou discrète) est une application mesurable de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , et que sa loi de probabilité est la mesure image définie par  $P_X(A) = P(X \in A)$  pour tout  $A \in \mathcal{B}_{\mathbb{R}}$ .

### 4.2.1 Loi de probabilité d'une v.a. réelle discrète

Par définition, une variable aléatoire réelle  $X$  est de loi discrète si elle prend ses valeurs dans une partie finie ou dénombrable de  $\mathbb{R}$  que l'on notera  $X(\Omega) = \{x_n\}_{n \in \mathbb{N}}$ . La loi d'une v.a. discrète est caractérisée par la donnée des probabilités

$$p_n = P(X = x_n), \quad n \in \mathbb{N}.$$

Pour tout  $A \in \mathcal{B}_{\mathbb{R}}$ , on a que

$$P(X \in B) = \sum_{n \in B \cap \mathbb{N}} p_n.$$

Ainsi, la loi de probabilité de  $X$  est la mesure discrète

$$P_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n},$$

et on a que

$$P(X \in B) = P_X(B) = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}(B) = \sum_{n \in \mathbb{N}} p_n \mathbb{1}_B(n) = \sum_{n \in B} p_n$$

### 4.2.2 Loi de probabilité d'une v.a. réelle continue

**Définition 4.8.** Soit  $f : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une fonction mesurable positive telle que son intégrale par rapport à la mesure de Lebesgue soit égale à 1 i.e.

$$\int f(x) d\lambda(x) = 1,$$

où  $\lambda$  est la mesure de Lebesgue sur  $\mathbb{R}$ . La fonction  $f$  est dite **densité de probabilité** par rapport à la mesure de Lebesgue.

**Définition 4.9.** Soit  $f$  une densité de probabilité par rapport à la mesure de Lebesgue. Une variable aléatoire réelle  $X$  est dite continue et de densité  $f$  (ou plus simplement à densité) si sa loi de probabilité est telle que

$$\forall A \in \mathcal{B}_{\mathbb{R}}, \quad P_X(A) = \int_A f(x) d\lambda(x).$$

Il est important de remarquer que la mesure de Lebesgue d'un ensemble dénombrable de réels est nulle. De ce fait, on peut changer en nombre dénombrable de points la valeur d'une densité de probabilité  $f$  (par rapport à la mesure de Lebesgue) sans changer la valeur de  $P_X(A)$ . Dans la suite du cours on notera également par  $dx$  l'intégration par rapport mesure de Lebesgue i.e.

$$P_X(A) = \int_A f(x) d\lambda(x) = \int_A f(x) dx.$$

### 4.3 Espérance d'une variable aléatoire

**Définition 4.10.** Soit  $X$  une variable aléatoire réelle telle que

$$\int_{\Omega} |X(w)| dP(w) < +\infty.$$

L'espérance de  $X$  est l'intégrale de la fonction mesurable  $X : \Omega \rightarrow \mathbb{R}$  par rapport à la mesure de probabilité  $P$  notée

$$\mathbb{E}(X) = \int_{\Omega} X(w) dP(w).$$

L'espérance d'une v.a.r.  $X$  représente donc la valeur moyenne que prend cette variable. Dans la pratique, on n'effectue aucun calcul d'intégrale par rapport à la mesure de probabilité  $P$ . Les calculs se feront en fait à partir d'intégrales sur  $\mathbb{R}$  ou de sommes discrètes. Ceci se justifie à l'aide du résultat suivant connu sous le nom de théorème de la mesure image.

**Théorème 4.1.** (admis) Soit  $X$  une variable aléatoire réelle, et  $\phi : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une application mesurable, telle que  $\phi(X)$  soit intégrable par rapport à la mesure de probabilité  $P$ . Alors

$$\mathbb{E}(\phi(X)) = \int_{\Omega} \phi(X(w)) dP(w) = \int_{\mathbb{R}} \phi(x) dP_X(x).$$

Un corollaire immédiat du théorème de la mesure image est que l'espérance d'une v.a.r.  $X$  de loi  $P_X$  se calcule à l'aide la formule suivante

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dP_X(x).$$

Ainsi, dans le cas où  $X$  est une v.a.r. continue de densité  $f$  (par rapport à la mesure de Lebesgue  $dx$ ) on a que

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

et dans le cas où  $X$  est une v.a.r. discrète de loi  $P_X = \sum_{n \in \mathbb{N}} p_n \delta_{x_n}$ , on obtient que

$$\mathbb{E}(X) = \sum_{n \in \mathbb{N}} x_n p_n.$$



## Chapitre 5

# COUPLES ALÉATOIRES DISCRETS

### 5.1 Généralités

Soit  $X$  et  $Y$  deux v.a.r. discrètes, définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ . On pose  $X(\Omega) = \{x_i ; i \in \mathbb{N}\}$  et  $Y(\Omega) = \{y_j ; j \in \mathbb{N}\}$ .

#### 5.1.1 Loi de probabilité d'un couple $(X, Y)$

**Définition 5.1.** On appelle loi de probabilité du couple  $(X, Y)$  l'ensemble des triplets  $(x_i, y_j, p_{ij})$  avec

$$p_{ij} = P([X = x_i] \cap [Y = y_j]) \text{ pour } i \in \mathbb{N} \text{ et } j \in \mathbb{N}.$$

On a alors  $\sum_i \sum_j p_{ij} = \sum_j \sum_i p_{ij} = 1$ .

Dans le cas où  $X(\Omega)$  et  $Y(\Omega)$  sont finis avec peu d'éléments, les résultats peuvent être donnés dans un tableau à doubles entrées.

Exemple : on considère un lancer de 2 dés à 6 faces. On note pas  $X_1$  et  $X_2$  les variables aléatoires qui représentent le résultat obtenu avec chaque dé. On pose  $Y = \max(X_1, X_2)$ . Le tableau des probabilités  $p_{ij}$  du couple aléatoire discret  $(X, Y) = (X_1, Y)$  est alors donné dans la Table 5.1.

#### 5.1.2 Lois marginales

**Définition 5.2.** Les v.a.r.  $X$  et  $Y$  sont appelées v.a.r. marginales du couple  $(X, Y)$  ; les lois des v.a.r.  $X$  et  $Y$  sont appelées lois marginales du couple  $(X, Y)$ .

On pose  $p_{i.} = \sum_j p_{ij}$  et  $p_{.j} = \sum_i p_{ij}$ .

TABLE 5.1 – Lancer de 2 dés : tableau à double entrée de la loi de probabilités du couple  $(X_1, Y)$ , et des lois marginales de  $X_1$  et  $Y$ .

$X_1 / Y$	1	2	3	4	5	6	$P(X_1 = x_i) = p_i$
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	0	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	0	0	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	0	0	0	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	0	0	0	0	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	0	0	0	0	0	$\frac{6}{36}$	$\frac{1}{6}$
$P(Y = y_j) = p_j$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	

**Théorème 5.1.** La loi de  $X$  est définie par l'ensemble des couples  $(x_i, p_i)$  pour  $i \in \mathbb{N}$  et la loi de  $Y$  est définie par l'ensemble des couples  $(y_j, p_j)$  pour  $j \in \mathbb{N}$ .

*Démonstration.*  $Y(\Omega) = \{y_j ; j \in \mathbb{N}\}$  donc  $\Omega = \bigcup_j [Y = y_j]$  réunion dénombrable d'événements disjoints. Les événements  $[Y = y_j]$  forment un système complet d'événements, donc :

$$\begin{aligned} [X = x_i] &= [X = x_i] \cap \Omega = [X = x_i] \cap \left( \bigcup_j [Y = y_j] \right) \\ &= \bigcup_j ([X = x_i] \cap [Y = y_j]) \end{aligned}$$

d'après la distributivité de  $\cap$  par rapport à  $\bigcup$ . On a alors

$$P([X = x_i]) = \sum_j P([X = x_i] \cap [Y = y_j]) = \sum_j p_{ij} = p_i.$$

et de même

$$P([Y = y_j]) = \sum_i P([X = x_i] \cap [Y = y_j]) = \sum_i p_{ij} = p_j.$$

□

Lorsque l'on représente la loi du couple dans un tableau, la loi de  $X$  est obtenue en faisant la somme sur les lignes (c'est-à-dire que  $p_i$  est la somme des  $p_{ij}$  sur la  $i$ -ème ligne) et la loi de  $Y$  est obtenue en faisant la somme sur les colonnes (c'est-à-dire que  $p_j$  est la somme des  $p_{ij}$  sur la  $j$ -ème colonne). On peut se reporter à la Table 5.1 pour l'exemple du lancé de 2 dés.

### 5.1.3 Indépendance de deux v.a.r. discrètes

**Définition 5.3.** Deux v.a.r. discrètes  $X$  et  $Y$  sont dites indépendantes si, pour tout  $(i, j) \in \mathbb{N}^2$ , on a :

$$P([X = x_i] \cap [Y = y_j]) = P([X = x_i])P([Y = y_j]) \quad (\text{c'est-à-dire } p_{ij} = p_i.p_j).$$

**Remarque :** C'est le seul cas où la loi du couple est entièrement déterminée par les lois marginales.

### 5.1.4 Somme de deux v.a.r. discrètes

**Théorème 5.2.** Soit  $X$  et  $Y$  deux v.a.r. discrètes et soit  $Z = X + Y$ .

Pour  $z \in Z(\Omega)$ , on pose  $I_z = \{(i, j) \in \mathbb{N}^2 ; x_i + y_j = z\}$ ; alors  $P([Z = z]) = \sum_{(i,j) \in I_z} p_{ij}$ .

*Démonstration.* Pour calculer  $P([Z = z])$ , on considère l'ensemble  $I_z = \{(i, j) \in \mathbb{N}^2 ; x_i + y_j = z\}$  qui décrit toutes les situations possibles pour lesquelles  $X + Y$  prend la valeur  $z$ .

On a alors  $[Z = z] = \bigcup_{(i,j) \in I_z} ([X = x_i] \cap [Y = y_j])$  réunion d'événements disjoints, donc

$$P([Z = z]) = \sum_{(i,j) \in I_z} P([X = x_i] \cap [Y = y_j]) = \sum_{(i,j) \in I_z} p_{ij}.$$

□

Exemple : On lance un dé deux fois de suite et on s'intéresse à la somme des points.

$$Z(\Omega) = \{2, \dots, 12\},$$

$$I_6 = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}, \quad I_7 = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$

$$P([X = x_i] \cap [Y = y_j]) = \frac{1}{36} \text{ pour tout couple } (x_i, y_j) \in \{1, \dots, 6\}^2 \text{ donc}$$

$$P([Z = 6]) = \frac{5}{36} \text{ et } P([Z = 7]) = \frac{6}{36}.$$

**Cas particulier important si  $X$  et  $Y$  sont indépendantes et à valeurs dans  $\mathbb{N}$  :**

**Théorème 5.3.** Si  $X$  et  $Y$  sont indépendantes à valeurs dans  $\mathbb{N}$ , de fonctions génératrices respectives  $G_X$  et  $G_Y$ , alors  $P([X + Y = n]) = \sum_{i+j=n} p_i.p_j$  et  $G_{X+Y}(t) = G_X(t)G_Y(t)$  pour tout  $t \in [0, 1]$ .

*Démonstration.*

$$G_X(t)G_Y(t) = \left( \sum_{i \geq 0} t^i P([X = i]) \right) \left( \sum_{j \geq 0} t^j P([Y = j]) \right) = \sum_{i \geq 0} \sum_{j \geq 0} t^{i+j} P([X = i])P([Y = j])$$

(On peut échanger l'ordre des termes car les deux séries sont à termes positifs.)

On a, d'après l'indépendance de  $X$  et de  $Y$ ,  $P([X = i])P([Y = j]) = P([X = i] \cap [Y = j])$  et, si  $I_k = \{(i, j) \in \mathbb{N}^2 ; i + j = k\}$ , pour  $k \in \mathbb{N}$ , les  $I_k$  forment une partition de  $\mathbb{N}^2$ . D'où :

$$G_X(t)G_Y(t) = \sum_{k \geq 0} \sum_{(i,j) \in I_k} t^{i+j} P([X = i] \cap [Y = j]) = \sum_{k \geq 0} t^k P([X + Y = k]) = G_{X+Y}(t).$$

□

### Application :

**Propriété 5.1.** Si  $X$  et  $Y$  sont deux v.a.r. indépendantes :

- 1) si  $X$  suit la loi binomiale  $\mathcal{B}(n, p)$  et  $Y$  la loi  $\mathcal{B}(m, p)$ , alors  $X + Y$  suit la loi  $\mathcal{B}(n + m, p)$  ;
- 2) si  $X$  suit la loi de Poisson  $\mathcal{P}(\lambda)$  et  $Y$  la loi  $\mathcal{P}(\mu)$ , alors  $X + Y$  suit la loi  $\mathcal{P}(\lambda + \mu)$ .

*Démonstration.* 1)  $G_X(t) = \sum_{k=0}^n t^k C_n^k p^k (1-p)^{n-k} = (pt + 1 - p)^n$  et de même  $G_Y(t) = (pt + 1 - p)^m$   
donc

$$G_{X+Y}(t) = G_X(t)G_Y(t) = (pt + 1 - p)^{n+m}.$$

On reconnaît là la fonction génératrice de la loi binomiale  $\mathcal{B}(n + m, p)$  donc  $X + Y$  suit la loi  $\mathcal{B}(n + m, p)$ .

2) De même,  $G_X(t) = \sum_{k=0}^{+\infty} t^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda t}$  et  $G_Y(t) = e^{-\mu} e^{\mu t}$  donc

$$G_{X+Y}(t) = G_X(t)G_Y(t) = e^{-(\lambda+\mu)} e^{(\lambda+\mu)t}.$$

On reconnaît là la fonction génératrice de la loi de Poisson  $\mathcal{P}(\lambda + \mu)$  donc  $X + Y$  suit la loi  $\mathcal{P}(\lambda + \mu)$ . □

## 5.2 Opérateurs classiques

### 5.2.1 Espérance

**Propriété 5.2.**

- 1) Si  $X$  et  $Y$  possèdent une espérance, alors  $\mathbb{E}(X + Y)$  existe et

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \text{ (linéarité de l'espérance).}$$

- 2) Si  $X$  et  $Y$  possèdent un moment d'ordre 2, alors  $\mathbb{E}(XY)$  existe.

*Démonstration.* 1) Soit  $Z = X + Y$ .

Alors  $[Z = z] = \bigcup_{(i,j) \in I_z} ([X = x_i] \cap [Y = y_j])$  où  $I_z = \{(i, j) \in \mathbb{N}^2 ; x_i + y_j = z\}$ .

$$\begin{aligned}
\sum_{z \in Z(\Omega)} zP([Z = z]) &= \sum_{z \in Z(\Omega)} z \sum_{(i,j) \in I_z} P([X = x_i] \cap [Y = y_j]) \\
&= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in I_z} (x_i + y_j)P([X = x_i] \cap [Y = y_j]) = \sum_{(i,j)} (x_i + y_j)p_{ij} \\
&= \sum_{(i,j)} x_i p_{ij} + \sum_j y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} \\
&= \sum_i x_i p_{i.} + \sum_j y_j p_{.j}
\end{aligned}$$

Les deux séries obtenues sont absolument convergentes car  $\mathbb{E}(X)$  et  $\mathbb{E}(Y)$  existent donc, en reprenant la démarche précédente avec  $\sum_{z \in Z(\Omega)} |z|P([Z = z])$  et en utilisant les inégalités du type  $|z| \leq |x_i| + |y_j|$ , on obtient l'absolue convergence de  $\sum_{z \in Z(\Omega)} zP([Z = z])$ , ce qui légitime ce qui a été fait ci-dessus (*i.e.* intervertir les sommations).

On a donc bien  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ .

**2)** Si  $Z = XY$  et si, pour  $z \in Z(\Omega)$ , on pose  $J_z = \{(i, j) \in \mathbb{N}^2 ; x_i y_j = z\}$ , alors  $[Z = z] = \bigcup_{(i,j) \in J_z} ([X = x_i] \cap [Y = y_j])$  et, sous réserve d'absolue convergence pour pouvoir intervertir les sommes, on a :

$$\begin{aligned}
\sum_{z \in Z(\Omega)} zP([Z = z]) &= \sum_{z \in Z(\Omega)} z \sum_{(i,j) \in I_z} P([X = x_i] \cap [Y = y_j]) \\
&= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in J_z} x_i y_j P([X = x_i] \cap [Y = y_j]) \\
&= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in J_z} x_i y_j p_{ij} = \sum_{(i,j)} x_i y_j p_{ij}.
\end{aligned}$$

Or  $|x_i y_j| \leq \frac{1}{2}(x_i^2 + y_j^2)$  car  $\frac{1}{2}(x_i^2 + y_j^2) - |x_i y_j| = \frac{1}{2}(|x_i| - |y_j|)^2 \geq 0$  et

$$\begin{aligned}
\sum_{(i,j)} |x_i y_j| p_{ij} &\leq \frac{1}{2} \sum_{(i,j)} (x_i^2 + y_j^2) p_{ij} = \frac{1}{2} \left( \sum_{(i,j)} x_i^2 p_{ij} + \sum_{(i,j)} y_j^2 p_{ij} \right) \\
&= \frac{1}{2} \sum_i x_i^2 \sum_j p_{ij} + \frac{1}{2} \sum_j y_j^2 \sum_i p_{ij} = \frac{1}{2} \left( \sum_i x_i^2 p_{i.} + \sum_j y_j^2 p_{.j} \right) \\
&= \frac{1}{2} \mathbb{E}(X^2) + \frac{1}{2} \mathbb{E}(Y^2) < +\infty
\end{aligned}$$

car  $X$  et  $Y$  sont d'ordre 2. La convergence absolue est donc bien vérifiée et on a alors

$$\mathbb{E}(XY) = \sum_{z \in Z(\Omega)} zP([Z = z]) = \sum_{(i,j)} x_i y_j p_{ij} \text{ et } \mathbb{E}(XY) \leq \frac{1}{2} \mathbb{E}(X^2) + \frac{1}{2} \mathbb{E}(Y^2).$$

□

**Théorème 5.4.** Soit  $\varphi$  une application de  $\mathbb{R}^2$  dans  $\mathbb{R}$ , telle que  $\varphi(X, Y)$  admette une espérance. Alors :

$$\mathbb{E}(\varphi(X, Y)) = \sum_{(i,j)} \varphi(x_i, y_j) P([X = x_i] \cap [Y = y_j]) = \sum_{i,j} \varphi(x_i, y_j) p_{ij}.$$

*Démonstration.* La démonstration est identique à celle des propriétés précédentes.

On pose  $Z = \varphi(X, Y)$  et, pour  $z \in Z(\Omega)$ ,  $K_z = \{(i, j) \in \mathbb{N}^2 ; \varphi(x_i, y_j) = z\}$   
 $Z(\Omega)$  est fini ou dénombrable car  $X(\Omega)$  et  $Y(\Omega)$  le sont. On a alors

$$\begin{aligned} \sum_{z \in Z(\Omega)} z P([Z = z]) &= \sum_{z \in Z(\Omega)} z \sum_{(i,j) \in K_z} P([X = x_i] \cap [Y = y_j]) \\ &= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in K_z} \varphi(x_i, y_j) P([X = x_i] \cap [Y = y_j]) \\ &= \sum_{z \in Z(\Omega)} \sum_{(i,j) \in K_z} \varphi(x_i, y_j) p_{ij} = \sum_{(i,j)} \varphi(x_i, y_j) p_{ij}. \end{aligned}$$

Les interversions des sommes sont légitimes car  $Z$  est d'ordre 1 donc  $\sum_{z \in Z(\Omega)} z P([Z = z])$  est absolument convergente. On a donc

$$\mathbb{E}(\varphi(X, Y)) = \sum_{(i,j)} \varphi(x_i, y_j) p_{ij}$$

□

**Application :**

**Propriété 5.3.** Si  $X$  et  $Y$  sont indépendantes, alors  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ .

*Démonstration.* En effet, on a alors  $p_{ij} = p_i.p_j$  et, d'après ce qui précède,

$$\mathbb{E}(XY) = \sum_{(i,j)} x_i y_j p_{ij} = \sum_{(i,j)} x_i y_j p_i.p_j = \left( \sum_i x_i p_i \right) \left( \sum_j y_j p_j \right) = \mathbb{E}(X)\mathbb{E}(Y).$$

□

**Remarque :** La réciproque est en général fausse.

### 5.2.2 Variance et covariance

**Définition 5.4.** Si  $X$  et  $Y$  sont d'ordre 2, on appelle :

1) covariance de  $X$  et de  $Y$ , le réel  $\text{cov}(X, Y)$  défini par

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

2) coefficient de corrélation linéaire de  $X$  et de  $Y$ , le réel  $\rho(X, Y)$  défini par

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} ;$$

3) matrice de covariance de  $X$  et de  $Y$ , la matrice  $\Gamma(X, Y)$  définie par :

$$\Gamma(X, Y) = \begin{pmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{var}(Y) \end{pmatrix} = \begin{pmatrix} \text{cov}(X, X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) \end{pmatrix}.$$

**Propriété 5.4.**

**1)i)**  $\text{cov}(X, Y) = \text{cov}(Y, X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$  ;

**1)ii)**  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$  ;

**1)iii)** si  $X$  et  $Y$  sont **indépendantes**, alors  $\text{cov}(X, Y) = 0$  et

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

**1)iv)**  $\text{cov}(aX + b, cY + d) = ac \text{cov}(X, Y)$ .

**2)**  $\rho(X, Y) \in [-1, 1]$  et  $\rho(aX + b, cY + d) = \frac{ac}{|ac|} \rho(X, Y)$ .

**3)**  $\Gamma(X, Y)$  est une matrice réelle symétrique et, pour tout  $(u, v) \in \mathbb{R}^2$ ,

$$(u, v)\Gamma(X, Y) \begin{pmatrix} u \\ v \end{pmatrix} \geq 0.$$

*Démonstration.* **1)i)**  $\text{cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY - \mathbb{E}(X)Y - X\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y))$   
 puis, en utilisant la linéarité de  $\mathbb{E}$ ,

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

**1)ii)** Toujours en utilisant la linéarité de  $\mathbb{E}$ , on a :

$$\begin{aligned} \text{var}(X + Y) &= \text{cov}(X + Y, X + Y) = \mathbb{E}((X + Y)^2) - \mathbb{E}(X + Y)^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\ &= \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \end{aligned}$$

**1)iii)** Découle de 1)i) et de la propriété 5.4.

**1)iv)**

$$\begin{aligned} \text{cov}(aX + b, cY + d) &= \mathbb{E}((aX + b)(cY + d)) - \mathbb{E}(aX + b)\mathbb{E}(cY + d) \\ &= \mathbb{E}(acXY + adX + bcY + bd) - (a\mathbb{E}(X) + b)(c\mathbb{E}(Y) + d) \\ &= ac\mathbb{E}(XY) + ad\mathbb{E}(X) + bc\mathbb{E}(Y) + bd \\ &\quad - ac\mathbb{E}(X)\mathbb{E}(Y) - ad\mathbb{E}(X) - bc\mathbb{E}(Y) - bd \\ &= a\text{ccov}(X, Y) \end{aligned}$$

**2)i)** Pour tout  $\lambda \in \mathbb{R}$ , on a  $\text{var}(X + \lambda Y) \geq 0$ .

Or  $\text{var}(X + \lambda Y) = \text{var}(X) + 2\lambda\text{cov}(X, Y) + \lambda^2\text{var}(Y)$ , qui est un trinôme du second degré en  $\lambda$  toujours positif : son discriminant réduit est donc négatif, c'est-à-dire que l'on a :

$$(\text{cov}(X, Y))^2 - \text{var}(X)\text{var}(Y) \leq 0.$$

Il s'ensuit alors que  $\rho^2(X, Y) - 1 \leq 0$ .

**3)i)** Immédiat car  $\text{cov}(X, Y) = \text{cov}(Y, X)$ .

**3)ii)**

$$\begin{aligned} \begin{pmatrix} u & v \end{pmatrix} \Gamma(X, Y) \begin{pmatrix} u \\ v \end{pmatrix} &= \begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(X, Y) & \text{var}(Y) \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \\ &= \begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} u\text{var}(X) + v\text{cov}(X, Y) \\ u\text{cov}(X, Y) + v\text{var}(Y) \end{pmatrix} \\ &= u^2\text{var}(X) + 2uv\text{cov}(X, Y) + v^2\text{var}(Y) = \text{var}(uX + vY) \geq 0. \end{aligned}$$

□

## Chapitre 6

# COUPLES ALÉATOIRES A DENSITÉ

### 6.1 Loi de probabilité d'un couple de variables aléatoires réelles

Soit  $(\Omega, \mathcal{A}, P)$  un espace de probabilité et  $X$  et  $Y$  deux v.a.r. définies sur  $(\Omega, \mathcal{A}, P)$ .

**Définition 6.1.** On appelle fonction de répartition du couple  $(X, Y)$  la fonction  $F_{X,Y}$  définie sur  $\mathbb{R}^2$  par :

$$F_{X,Y}(x, y) = P([X \leq x] \cap [Y \leq y]).$$

**Propriété 6.1.** Les fonctions de répartition des v.a.r.  $X$  et  $Y$  vérifient

$$F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) \text{ et } F_Y(y) = \lim_{x \rightarrow +\infty} F_{X,Y}(x, y).$$

*Démonstration.*  $\Omega = \bigcup_n [Y \leq n]$  réunion croissante d'événements, donc, par distributivité,

$$[X \leq x] = [X \leq x] \cap \left( \bigcup_n [Y \leq n] \right) = \bigcup_n ([X \leq x] \cap [Y \leq n])$$

réunion croissante d'événements. On a donc, d'après la propriété 1.6 5),

$$P([X \leq x]) = \lim_{n \rightarrow +\infty} P([X \leq x] \cap [Y \leq n]).$$

Il vient alors

$$F_X(x) = \lim_{n \rightarrow +\infty} F_{X,Y}(x, n) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y)$$

et,  $X$  et  $Y$  jouant des rôles symétriques, la deuxième assertion en découle de la même façon.  $\square$

**Définition 6.2.** La loi du couple  $(X, Y)$  est dite absolument continue s'il existe une application  $f_{X,Y}$  de  $\mathbb{R}^2$  sur  $\mathbb{R}$ , positive et borélienne (c'est-à-dire que, pour tout  $B \in \mathcal{B}_{\mathbb{R}}$ ,  $f_{X,Y}^{-1}(B) \in \mathcal{B}_{\mathbb{R}^2}$ ), appelée densité du couple  $(X, Y)$ , telle que, pour tout  $(x, y) \in \mathbb{R}^2$  :

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv.$$

**Remarques :** toutes les fonctions “suffisamment continues” sont boréliennes. En pratique, on prendra souvent  $f_{X,Y}$  continue sur l'intérieur d'un sous-ensemble  $D$  de  $\mathbb{R}^2$  et nulle sur son complémentaire. Il existe des couples  $(X, Y)$  non discrets n'admettant pas non plus de densité. C'est le cas si, par exemple,  $X$  est une v.a.r. discrète et  $Y$  une v.a.r. absolument continue.

**Propriété 6.2.**

1)  $\int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f_{X,Y}(u, v) du \right) dv = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv \right) du = 1.$

2) Les lois marginales de  $X$  et de  $Y$  admettent les densités  $f_X$  et  $f_Y$  définies par

$$f_X(u) = \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv \text{ et } f_Y(v) = \int_{-\infty}^{+\infty} f_{X,Y}(u, v) du.$$

3) En tout  $(x_0, y_0)$  où  $f_{X,Y}$  est continue, on a  $f_{X,Y}(x_0, y_0) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x_0, y_0).$

*Démonstration.* 3) Découle de la définition de  $f_{X,Y}$  (définition 6.2).

2)  $F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y)$  avec

$$F_{X,Y}(x, y) = \int_{-\infty}^x \left( \int_{-\infty}^y f_{X,Y}(u, v) dv \right) du = \int_{-\infty}^y \left( \int_{-\infty}^x f_{X,Y}(u, v) du \right) dv$$

donc  $F_X(x) = \int_{-\infty}^{+\infty} \left( \int_{-\infty}^x f_{X,Y}(u, v) du \right) dv = \int_{-\infty}^x \left( \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv \right) du$  et

$$f_X(x) = F'_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, v) dv.$$

Comme  $X$  et  $Y$  jouent des rôles symétriques, la deuxième assertion en découle de la même façon.

1) Comme  $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ , on déduit de 2) que  $\int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f_{X,Y}(x, v) dv \right) dx = 1.$  □

**Parallèle avec les couples discrets :**

1) est l'analogue de  $\sum_i \left( \sum_j p_{ij} \right) = \sum_j \left( \sum_i p_{ij} \right) = 1;$

2) est l'analogue de  $p_{i.} = \sum_j p_{ij}$  et  $p_{.j} = \sum_i p_{ij}.$

*Exemple :* Soit  $(X, Y)$  un couple de loi uniforme sur le disque  $D(O, R)$ . Déterminer sa densité ainsi que les lois marginales de  $X$  et de  $Y$ .

$(X, Y)$  est un couple de loi uniforme sur le disque  $D(O, R)$  donc sa densité est constante sur  $D(O, R)$  et nulle ailleurs.

$f_{X,Y}(x, y) = C \mathbb{I}_{D(O,R)}(x, y)$  donc  $C \int \int_{D(O,R)} dx dy = 1$ . Or  $\int \int_{D(O,R)} dx dy$  est l'aire de  $D(O, R)$ , c'est-à-dire  $\pi R^2$  et  $C\pi R^2 = 1$  donne  $C = \frac{1}{\pi R^2}$  d'où

$$f_{X,Y}(x, y) = \frac{1}{\pi R^2} \mathbb{I}_{D(O,R)}(x, y).$$

Ainsi,  $f_{X,Y}(x, y) = \frac{1}{\pi R^2}$  si  $x^2 + y^2 \leq R^2$  et 0 sinon.

- Si  $|x| > R$ , alors  $x^2 + y^2 > R^2$  et  $f_{X,Y}(x, y) = 0$  pour tout  $y \in \mathbb{R}$ , d'où  $f_X(x) = 0$ .
- Si  $|x| \leq R$ , alors  $x^2 + y^2 \leq R^2$  pour  $y^2 \leq R^2 - x^2$ , c'est-à-dire pour  $y \in [-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}]$

et

$$f_X(x) = \int_{-\infty}^{+\infty} \frac{1}{\pi R^2} \mathbb{I}_{[-\sqrt{R^2 - x^2}, \sqrt{R^2 - x^2}]}(y) dy = \frac{2\sqrt{R^2 - x^2}}{\pi R^2}.$$

Finalement,  $f_X(x) = \frac{2\sqrt{R^2 - x^2}}{\pi R^2} \mathbb{I}_{[-R,R]}(x)$  et  $f_Y(y) = \frac{2\sqrt{R^2 - y^2}}{\pi R^2} \mathbb{I}_{[-R,R]}(y)$  car  $X$  et  $Y$  jouent le même rôle.

## 6.2 Indépendance

**Définition 6.3.** Soit  $X$  et  $Y$  deux v.a.r. définies sur  $(\Omega, \mathcal{A}, P)$ . Les variables aléatoires  $X$  et  $Y$  sont dites **indépendantes** si

$$P([X \in A] \cap [Y \in B]) = P([X \in A]) P([Y \in B])$$

pour tout  $A \in \mathcal{B}_{\mathbb{R}}$  et  $B \in \mathcal{B}_{\mathbb{R}}$ .

L'indépendance de deux variables aléatoires peut se caractériser plus simplement à l'aide de la fonction de répartition ou de la densité du couple  $(X, Y)$ .

**Propriété 6.3** (Admis). Deux v.a.r.  $X$  et  $Y$  sont indépendantes si, pour tout  $(x, y) \in \mathbb{R}^2$ , on a :

$$F_{X,Y}(x, y) = F_X(x)F_Y(y).$$

**Propriété 6.4.** Soit  $(X, Y)$  un couple de variables aléatoires de densité  $f_{X,Y}$  admettant pour densités marginales  $f_X$  et  $f_Y$ . Les v.a.r.  $X$  et  $Y$  sont ssi, pour tout  $(x, y) \in \mathbb{R}^2$ , on a :

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

On a également que l'indépendance de  $X$  et  $Y$  équivaut à  $\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X))\mathbb{E}(h(Y))$  pour toutes fonctions  $g$  et  $h$  pourvu que ces espérances existent.

**Attention :**  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$  n'est pas suffisant pour conclure sur l'indépendance de  $X$  et de  $Y$ .

## 6.3 Opérateurs classiques

### 6.3.1 Espérance

On a les mêmes propriétés que dans le cas des couples discrets. Les résultats démontrés pour les couples discrets au sujet des opérateurs restent valables : il suffit ici de remplacer les sommes par des intégrales et les probabilités par des densités. En particulier, on a les résultats suivants.

**Propriété 6.5.**

1) Si  $X$  et  $Y$  possèdent une espérance, alors  $\mathbb{E}(X + Y)$  existe et

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \text{ (linéarité de l'espérance).}$$

2) Si  $X$  et  $Y$  possèdent un moment d'ordre 2, alors  $\mathbb{E}(XY)$  existe.

**Propriété 6.6.** Si  $\varphi$  est une fonction de  $\mathbb{R}^2$  sur  $\mathbb{R}$  telle que  $\varphi(X, Y)$  admette une espérance, alors :

$$\mathbb{E}(\varphi(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y) f_{X,Y}(x, y) dx dy.$$

En particulier,  $\mathbb{E}(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f_{X,Y}(x, y) dx dy$  si cette espérance existe.

### 6.3.2 Variance et covariance

**Définition 6.4.** Si  $X$  et  $Y$  sont d'ordre 2, on appelle :

1) covariance de  $X$  et de  $Y$ , le réel  $\text{cov}(X, Y)$  défini par

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

2) coefficient de corrélation linéaire de  $X$  et de  $Y$ , le réel  $\rho(X, Y)$  défini par

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)} ;$$

3) matrice de covariance de  $X$  et de  $Y$ , la matrice  $\Gamma(X, Y)$  définie par :

$$\Gamma(X, Y) = \begin{pmatrix} \text{var}(X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{var}(Y) \end{pmatrix} = \begin{pmatrix} \text{cov}(X, X) & \text{cov}(X, Y) \\ \text{cov}(Y, X) & \text{cov}(Y, Y) \end{pmatrix}.$$

**Propriété 6.7.**

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

1)ii)  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$  ;

1)iii) si  $X$  et  $Y$  sont **indépendantes**, alors  $\text{cov}(X, Y) = 0$  et

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

1)iv)  $\text{cov}(aX + b, cY + d) = ac \text{cov}(X, Y)$ .

2)  $\rho(X, Y) \in [-1, 1]$  et  $\rho(aX + b, cY + d) = \frac{ac}{|ac|} \rho(X, Y)$ .

3)  $\Gamma(X, Y)$  est une matrice réelle symétrique et, pour tout  $(u, v) \in \mathbb{R}^2$ ,

$$(u, v)\Gamma(X, Y) \begin{pmatrix} u \\ v \end{pmatrix} \geq 0.$$

## 6.4 Changement de variables

Soit  $(X, Y)$  un couple aléatoire de densité  $f_{X,Y}$  et  $\psi$  une fonction (borélienne) de  $\mathbb{R}^2$  sur  $\mathbb{R}^2$ . On souhaite connaître la loi du couple  $(U, V) = \psi(X, Y)$ . On va d'abord énoncer l'analogie de  $f_{\varphi(X)}(y) = f_X(\varphi^{-1}(y))|(\varphi^{-1})'(y)|$  sur  $\mathbb{R}$ .

**Théorème 6.1.** Soit  $(X, Y)$  un couple aléatoire de densité  $f_{X,Y}$  et  $\psi$  une fonction de  $\mathbb{R}^2$  sur  $\mathbb{R}^2$ . Si  $f_{X,Y}$  est continue sur l'intérieur d'un ensemble  $D$  et nulle sur son complémentaire, si  $\psi$  est une bijection de  $D$  sur  $E = \psi(D)$  telle que les dérivées partielles de  $\psi$  et de  $\psi^{-1}$  existent

et soient continues, et si, de plus,  $J(\psi^{-1}) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \neq 0$  sur  $E$ , alors le couple aléatoire  $(U, V) = \psi(X, Y)$  admet pour densité la fonction  $f_{U,V}$  définie par :

$$f_{U,V}(u, v) = f_{X,Y}(\psi^{-1}(u, v)) |J(\psi^{-1})(u, v)| \text{ si } (u, v) \in E.$$

**Remarque :** Comme dans le cas réel, il peut arriver que  $\psi$  ne soit pas bijective sur  $D$  tout entier mais que sa restriction  $\psi_i$  à chacun des sous-ensembles  $D_i$ ,  $i \in \{1, 2, \dots, q\}$  d'une partition de  $D$ , le soit. On a alors

$$f_{U,V}(u, v) = \sum_{i=1}^q f_{X,Y}(\psi_i^{-1}(u, v)) |J(\psi_i^{-1})(u, v)| \mathbb{1}_{\psi(D_i)}(u, v).$$

Exemple : Soit  $(X, Y)$  un couple aléatoire de densité  $f_{X,Y}$  où  $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$ . Quelle est la loi de  $U = \frac{X}{Y}$  ?

**Méthode :** On calcule d'abord la loi de  $(U, V) = \left(\frac{X}{Y}, Y\right)$ , puis on déduit la loi de  $U$  comme loi marginale du couple  $(U, V)$ .

On pose  $\psi(x, y) = (u, v) = \left(\frac{x}{y}, y\right)$  :  $y = v$  et  $x = uv$  conduit à

$$\psi^{-1}(u, v) = (x, y) = (uv, v).$$

Le domaine  $\mathbb{R} \times \mathbb{R}^*$  se transforme en  $\mathbb{R} \times \mathbb{R}^*$ . Pour  $(u, v) \in \mathbb{R} \times \mathbb{R}^*$  :

$$J(\psi^{-1})(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v \neq 0,$$

donc  $J(\psi^{-1}) \neq 0$  sur  $\mathbb{R} \times ]0, +\infty[$  et sur  $\mathbb{R} \times ]-\infty, 0[$ .

On a alors, pour  $(u, v) \in \mathbb{R} \times \mathbb{R}^*$  :

$$f_{U,V}(u, v) = f_{X,Y}(\psi^{-1}(u, v)) |J(\psi^{-1}(u, v))| = \frac{1}{2\pi} e^{-\frac{1}{2}((uv)^2 + v^2)} |v|.$$

Il vient alors

$$\begin{aligned} f_U(u) &= \int_{-\infty}^{+\infty} f_{U,V}(u, v) dv = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}v^2(u^2+1)} |v| dv = 2 \times \frac{1}{2\pi} \int_0^{+\infty} e^{-\frac{1}{2}v^2(u^2+1)} v dv \\ &= \frac{1}{\pi} \frac{1}{u^2+1} \left[ -e^{-\frac{1}{2}v^2(u^2+1)} \right]_0^{+\infty} = \frac{1}{\pi} \frac{1}{1+u^2}. \end{aligned}$$

Donc, si  $(X, Y)$  suit la loi de densité  $f_{X,Y}$  définie par  $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$  alors la loi de  $U = \frac{X}{Y}$  est la loi de Cauchy  $\mathcal{C}(1)$ .

## 6.5 Sommes de deux v.a.r. absolument continues

**Théorème 6.2.** Soit  $(X, Y)$  un couple aléatoire de densité  $f_{X,Y}$ . Alors :

1) la v.a.r.  $X + Y$  a pour densité la fonction  $f_{X+Y}$  définie par :

$$f_{X+Y}(w) = \int_{-\infty}^{+\infty} f_{X,Y}(w-v, v) dv = \int_{-\infty}^{+\infty} f_{X,Y}(u, w-u) du.$$

2) Si  $X$  et  $Y$  sont indépendantes, alors :

$$f_{X+Y}(w) = \int_{-\infty}^{+\infty} f_X(w-v) f_Y(v) dv = \int_{-\infty}^{+\infty} f_X(u) f_Y(w-u) du.$$

*Démonstration.* On pose  $(U, V) = (X + Y, Y)$  et  $\psi(x, y) = (u, v) = (x + y, y)$ . Alors  $\psi$  est une bijection de  $\mathbb{R}^2$  sur lui-même et  $\psi^{-1}(u, v) = (x, y) = (u - v, v)$ .

$$J(\psi^{-1})(u, v) = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1, \text{ donc } f_{U,V}(u, v) = f_{X,Y}(u - v, v) \text{ et}$$

$$f_U(u) = \int_{-\infty}^{+\infty} f_{U,V}(u, v) dv = \int_{-\infty}^{+\infty} f_{X,Y}(u - v, v) dv.$$

De plus, si  $X$  et  $Y$  sont indépendantes,  $f_U(u) = \int_{-\infty}^{+\infty} f_X(u - v) f_Y(v) dv.$  □

**Remarque :** Si on prend  $(U, V) = (X, X + Y)$ , on obtient

$$f_{X+Y}(v) = \int_{-\infty}^{+\infty} f_{X,Y}(u, v-u) du.$$

*Exemple :* Soit  $(X, Y)$  un couple de densité  $f$  définie par  $f(x, y) = k(x^2 + y^2) \mathbb{I}_{[-1,1]^2}(x, y)$ . Déterminer  $k$  et calculer  $\text{cov}(X, Y)$ . Les v.a.r.  $X$  et  $Y$  sont-elles indépendantes? Déterminer la loi de  $X + Y$ .

**Réponse :**

- $k = \frac{3}{8}$ ,  $f_X(x) = \frac{3}{4}(x^2 + 1) \mathbb{I}_{[-1,1]}(x) = f_Y(x)$ .
- $\text{cov}(X, Y) = 0$  mais  $X$  et  $Y$  ne sont pas indépendantes.
- $f_{X+Y}(u) = \frac{1}{4}(1 + (1 - |u|)^3) \mathbb{I}_{[-2,2]}(u)$ .

**Application aux lois normales :**

**Théorème 6.3.** Si  $X$  et  $Y$  sont deux v.a.r. indépendantes de lois respectives  $\mathcal{N}(m_1, \sigma_1^2)$  et  $\mathcal{N}(m_2, \sigma_2^2)$ , alors la v.a.r.  $X + Y$  suit la loi normale  $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .

*Démonstration.* On a  $f_{X+Y}(u) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\left(\frac{(u-v-m_1)^2}{\sigma_1^2} + \frac{(v-m_2)^2}{\sigma_2^2}\right)\right) dv$ .

On va mettre  $\frac{(u-v-m_1)^2}{\sigma_1^2} + \frac{(v-m_2)^2}{\sigma_2^2}$  sous la forme  $\frac{(v-M)^2}{A} + B$ , où  $A, B$  et  $M$  sont dépendantes de  $u$ . On aura alors :

$$\begin{aligned} f_{X+Y}(u) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}B\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\left(\frac{(v-M)^2}{A}\right)\right) dv \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2}B\right) \sqrt{2\pi A}. \end{aligned}$$

Il reste donc à déterminer  $A, B$  et  $M$ .

$$\begin{aligned} \frac{(u-v-m_1)^2}{\sigma_1^2} + \frac{(v-m_2)^2}{\sigma_2^2} &= \frac{1}{\sigma_1^2\sigma_2^2} (\sigma_2^2(v-(u-m_1))^2 + \sigma_1^2(v-m_2)^2) \\ &= \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\sigma_2^2} \left( v^2 - 2v \frac{(u-m_1)\sigma_2^2 + m_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2} + \frac{(u-m_1)^2\sigma_2^2 + m_2^2\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \right) \end{aligned}$$

Or, on a :

$$\begin{aligned} &\frac{(u-m_1)^2\sigma_2^2 + m_2^2\sigma_1^2}{\sigma_1^2 + \sigma_2^2} - \left(\frac{(u-m_1)\sigma_2^2 + m_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)^2 \\ &= \frac{((u-m_1)^2\sigma_2^2 + m_2^2\sigma_1^2)(\sigma_1^2 + \sigma_2^2) - ((u-m_1)\sigma_2^2 + m_2\sigma_1^2)^2}{(\sigma_1^2 + \sigma_2^2)^2} \\ &= \frac{(u-m_1)^2\sigma_1^2\sigma_2^2 + m_2^2\sigma_1^2\sigma_2^2 - 2m_2(u-m_1)\sigma_1^2\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)^2} \\ &= \frac{\sigma_1^2\sigma_2^2(u-m_1-m_2)^2}{(\sigma_1^2 + \sigma_2^2)^2} \end{aligned}$$

donc  $A = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$  et  $B = \frac{(u - m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}$  et, comme  $f_{X+Y}(u) = \frac{\sqrt{A}}{\sqrt{2\pi}\sigma_1\sigma_2} \exp\left(-\frac{1}{2}B\right)$ ,

$$f_{X_1+X_2}(u) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{1}{2} \frac{(u - m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}\right).$$

On reconnaît là la loi normale  $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ .

□

## Chapitre 7

# CONVERGENCE DE SUITES DE VARIABLES ALÉATOIRES

Les variables aléatoires  $X_n, X$  utilisées dans ce chapitre sont toutes définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, P)$ , à valeurs dans  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , de fonctions de répartition respectives  $F_n$  et  $F$ . Dans ce chapitre, on va aborder les notions de convergence en moyenne, en moyenne quadratique et en probabilité mais c'est surtout la convergence en loi qui retiendra notre attention, ainsi que les questions d'approximations de lois.

### 7.1 Inégalités

**Propriété 7.1. (Inégalité de Markov)**

Soit  $\varepsilon$  et  $\alpha$  deux réels strictement positifs. Si  $X$  admet un moment d'ordre  $\alpha$  ( $\mathbb{E}(|X|^\alpha)$  existe et est fini), alors

$$P(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^\alpha} \mathbb{E}(|X|^\alpha).$$

*Démonstration.* On a :

$$|X|^\alpha = |X|^\alpha \mathbb{1}_{|X| \geq \varepsilon} + |X|^\alpha \mathbb{1}_{|X| < \varepsilon}.$$

Il est clair que, puisque  $|X|^\alpha$  possède une espérance, les v.a.r.  $|X|^\alpha \mathbb{1}_{|X| \geq \varepsilon}$  et  $|X|^\alpha \mathbb{1}_{|X| < \varepsilon}$  possèdent également une espérance. On a, par ailleurs :

$$|X|^\alpha \mathbb{1}_{|X| \geq \varepsilon} \geq \varepsilon^\alpha \mathbb{1}_{|X| \geq \varepsilon}.$$

Cette inégalité implique que :

$$\begin{aligned} \mathbb{E}(|X|^\alpha) &= \mathbb{E}(|X|^\alpha \mathbb{1}_{|X| \geq \varepsilon}) + \mathbb{E}(|X|^\alpha \mathbb{1}_{|X| < \varepsilon}) \\ &\geq \mathbb{E}(|X|^\alpha \mathbb{1}_{|X| \geq \varepsilon}) \geq \varepsilon^\alpha \mathbb{E}(\mathbb{1}_{|X| \geq \varepsilon}) = \varepsilon^\alpha P(|X| \geq \varepsilon) \end{aligned}$$

ce qui donne le résultat. □

**Corollaire 7.1 : (Inégalité de Bienaymé-Tchebicheff)**

Soit  $\varepsilon$  un réel strictement positif. Si  $X$  admet un moment d'ordre 2, alors

$$P(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{var}(X).$$

*Démonstration.* On a  $\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) < +\infty$ , c'est-à-dire la v.a.r.  $Y = X - \mathbb{E}(X)$  admet un moment d'ordre 2. Il suffit d'appliquer l'inégalité de Markov à  $Y$  avec  $\alpha = 2$ .  $\square$

On termine cette partie par deux autres inégalités classiques :

**Inégalité de Jensen :** Soit  $X$  une v.a.r. et  $\varphi$  une fonction convexe de  $\mathbb{R}$  dans  $\mathbb{R}$ . Alors, si  $X$  et  $\varphi(X)$  possèdent une espérance, on a :

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X)).$$

**Inégalité de Cauchy-Schwarz :** Si  $X$  et  $Y$  sont des v.a.r. ayant un moment d'ordre 2, alors :

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

## 7.2 Convergence presque sûre

**Définition 7.1.** Soit  $(X_n)_{n \in \mathbb{N}}$  et  $X$  des variables aléatoires réelles. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge presque sûrement vers  $X$  si et seulement si  $P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$ . On note alors  $X_n \xrightarrow{p.s.} X$ .

Exemple : Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires bornées

$$|X_n| \leq u_n,$$

où  $u_n$  est une suite de réels positifs qui tend vers zéro. Alors  $X_n \xrightarrow{p.s.} 0$ .

**Propriété 7.2.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires qui converge presque sûrement vers  $X$  et  $f : \mathbb{R} \rightarrow \mathbb{R}$  une fonction continue, alors  $f(X_n) \xrightarrow{p.s.} f(X)$ .

*Démonstration.* Comme  $f$  est continue sur  $\mathbb{R}$ , on a que  $\lim_{n \rightarrow +\infty} f(X_n(\omega)) = f(X(\omega))$  pour tout  $\omega \in \Omega$  tel que  $\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)$ . Par conséquent,

$$\left\{\omega \in \Omega : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)\right\} \subset \left\{\omega \in \Omega : \lim_{n \rightarrow +\infty} f(X_n(\omega)) = f(X(\omega))\right\},$$

et donc

$$1 \leq P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow +\infty} f(X_n(\omega)) = f(X(\omega))\right\}\right) \leq 1,$$

car  $X_n \xrightarrow{p.s.} X$ .  $\square$

Un critère important de convergence presque sûre est donné par le théorème suivant appelé Lemme de Borel-Cantelli (un résultat classique en théorie des probabilités).

**Théorème 7.1** (admis). Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires réelles et  $X$  une variable aléatoire réelle. Si  $\sum_{n \in \mathbb{N}} P(|X_n - X| > \epsilon) < +\infty$  pour tout  $\epsilon > 0$ , alors  $X_n \xrightarrow{p.s.} X$ .

### 7.3 Convergence en moyenne quadratique

**Définition 7.2.** Soit  $(X_n)_{n \in \mathbb{N}}$  et  $X$  des variables aléatoires réelles admettant un moment d'ordre 2. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en moyenne quadratique vers  $X$  si et seulement si  $\lim_{n \rightarrow +\infty} \mathbb{E}((X_n - X)^2) = 0$ . On note alors  $X_n \xrightarrow{M.Q.} X$ .

On va maintenant énoncer quelques propriétés relatives à ces modes de convergence.

**Propriété 7.3.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. admettant une variance. Si  $\mathbb{E}(X_n) \xrightarrow{n \rightarrow +\infty} \mu$  et  $\text{var}(X_n) \xrightarrow{n \rightarrow +\infty} 0$ , alors  $X_n \xrightarrow{M.Q.} \mu$ .

*Démonstration.* On a

$$\begin{aligned} \mathbb{E}((X_n - \mu)^2) &= \mathbb{E}((X_n - \mathbb{E}(X_n) + \mathbb{E}(X_n) - \mu)^2) \\ &= \mathbb{E}((X_n - \mathbb{E}(X_n))^2) + \mathbb{E}((\mathbb{E}(X_n) - \mu)^2) = \text{var}(X_n) + (\mathbb{E}(X_n) - \mu)^2 \xrightarrow{n \rightarrow +\infty} 0 \end{aligned}$$

□

**Propriété 7.4.** Soit  $(X_n)_{n \in \mathbb{N}}$  et  $X$  des v.a.r. Si  $X_n \xrightarrow{M.Q.} X$ , alors  $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$  et  $\lim_{n \rightarrow +\infty} \text{var}(X_n) = \text{var}(X)$ .

*Démonstration.* Les résultats se déduisent directement de l'inégalité de Jensen appliquée à la fonction convexe  $\varphi(x) = |x|$  :

$$|\mathbb{E}(X_n) - \mathbb{E}(X)| = |\mathbb{E}(X_n - X)| \leq \mathbb{E}(|X_n - X|) \leq \sqrt{\mathbb{E}((X_n - X)^2)}$$

et d'autre part de l'inégalité :

$$\left| \sqrt{\mathbb{E}(X_n^2)} - \sqrt{\mathbb{E}(X^2)} \right| \leq \sqrt{\mathbb{E}((X_n - X)^2)}.$$

□

### 7.4 Convergence en probabilité

**Définition 7.3.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. et soit  $X$  une v.a.r. On dit que  $(X_n)_{n \in \mathbb{N}}$  converge en probabilité vers  $X$  si et seulement si, pour tout  $\varepsilon > 0$ ,  $\lim_{n \rightarrow +\infty} P(|X_n - X| < \varepsilon) = 1$ . On note alors  $X_n \xrightarrow{P} X$ .

On a alors les propriétés suivantes :

**Propriété 7.5** (Admis). Soit  $(X_n)_{n \in \mathbb{N}}$  et  $(Y_n)_{n \in \mathbb{N}}$  deux suites de v.a.r. et  $X$  et  $Y$  deux v.a.r. Si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$  et si  $f$  est une fonction continue de  $\mathbb{R}$  dans  $\mathbb{R}$  et  $g$  une fonction continue de  $\mathbb{R}^2$  dans  $\mathbb{R}$ , alors :

- 1)  $f(X_n) \xrightarrow{P} f(X)$
- 2)  $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$ .

**Corollaire :** Soit  $(X_n)_{n \in \mathbb{N}}$  et  $(Y_n)_{n \in \mathbb{N}}$  deux suites de v.a.r. et  $X$  et  $Y$  deux v.a.r. Alors, si  $X_n \xrightarrow{P} X$  et  $Y_n \xrightarrow{P} Y$ , on a :

- 1)  $|X_n| \xrightarrow{P} |X|$
- 2) Pour tout  $(\lambda, \mu) \in \mathbb{R}^2$ ,  $\lambda X + \mu Y_n \xrightarrow{P} \lambda X + \mu Y$
- 3)  $X_n Y_n \xrightarrow{P} XY$ .

**Théorème 7.2.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. qui converge presque sûrement vers  $X$ . Alors  $X_n \xrightarrow{P} X$ .

*Démonstration.* Soit  $\epsilon > 0$ , et posons  $B_n^\epsilon = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}$ . La réalisation de l'évènement  $\left\{ w : \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \right\}$  implique, pour tout  $\epsilon > 0$ , celle de l'évènement

$$\{w : \exists n_0(\omega) \text{ t.q. } \forall n \geq n_0(\omega) w \notin B_n^\epsilon\} = \bigcup_{n_0 \geq 1} \bigcap_{n \geq n_0} \overline{B_n^\epsilon}$$

Comme  $X_n \xrightarrow{p.s.} X$ , on a donc que  $P\left(\bigcup_{n_0 \geq 1} \bigcap_{n \geq n_0} \overline{B_n^\epsilon}\right) = 1$ , ce qui s'écrit encore  $P\left(\bigcap_{n_0 \geq 1} \bigcup_{n \geq n_0} B_n^\epsilon\right) = 0$ . Les

évènements  $A_{n_0} = \bigcup_{n \geq n_0} B_n^\epsilon$  étant une suite décroissante, on obtient donc que  $\lim_{n_0 \rightarrow +\infty} P\left(\bigcup_{n \geq n_0} B_n^\epsilon\right) = 0$

ce qui implique que  $\lim_{n_0 \rightarrow +\infty} P(B_{n_0}^\epsilon) = 0$  car  $B_{n_0}^\epsilon \subset \bigcup_{n \geq n_0} B_n^\epsilon$  et donc  $X_n \xrightarrow{P} X$  (par définition).  $\square$

**Théorème 7.3.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. admettant un moment d'ordre 2 et  $X$  une v.a.r. admettant un moment d'ordre 2. Si  $X_n \xrightarrow{M.Q.} X$ , alors  $X_n \xrightarrow{P} X$ .

*Démonstration.* L'implication découle de l'inégalité de Markov :

$$\text{pour tout } \epsilon > 0, P(|X_n - X| > \epsilon) \leq \frac{1}{\epsilon^2} \mathbb{E}((X_n - X)^2) \xrightarrow{n \rightarrow +\infty} 0.$$

$\square$

## 7.5 Convergence en loi

Souvent, en statistique, on ne connaît pas les lois : on observe seulement une certaine distribution. Les théorèmes de convergence en loi permettent de justifier certaines approximations de distributions observées par des lois théoriques connues.

**Définition 7.4.** On dit que la suite de variables aléatoires  $(X_n)$  converge en loi vers  $X$  si, pour tout  $x$  en lequel  $F$  est continue,  $\lim_{n \rightarrow +\infty} F_n(x) = F(x)$ . On écrit  $X_n \xrightarrow{L} X$ .

Exemple 1 : soit  $X_n$  une v.a. de loi uniforme sur  $[0, \frac{1}{n}]$ . On a que

$$F_{X_n}(t) = \begin{cases} 1 & \text{si } t \geq \frac{1}{n} \\ nt & \text{si } 0 \leq t < \frac{1}{n} \\ 0 & \text{si } t < 0 \end{cases}, \text{ et donc}$$

- si  $t > 0$ , alors  $\lim_{n \rightarrow +\infty} F_{X_n}(t) = 1$

- si  $t < 0$ , alors  $\lim_{n \rightarrow +\infty} F_{X_n}(t) = 0$ .

Donc  $X_n$  converge en loi vers la v.a. constante  $X = 0$  dont la fonction de répartition est  $F_X(t) = \mathbb{1}_{[0, +\infty[}(t)$ .

Exemple 2 : si  $X_n$  suit la loi binomiale  $\mathcal{B}\left(n, \frac{\lambda}{n}\right)$  et si  $X$  suit la loi de Poisson  $\mathcal{P}(\lambda)$ , alors  $X_n \xrightarrow{L} X$ .

*Démonstration.* La fonction de répartition  $F$  de  $X$  est continue sur  $\mathbb{R} \setminus \mathbb{N}$ . Pour tout  $x \in \mathbb{R} \setminus \mathbb{N}$ , on a :

$$F_n(x) = \sum_{k \leq x} P(X_n = k) = \sum_{k=0}^{\min(E(x), n)} C_n^k p_n^k (1-p_n)^{n-k}$$

où  $E(x)$  désigne la partie entière de  $x$ . Pour  $n \geq E(x)$ , on a alors

$$F_n(x) = \sum_{k=0}^{E(x)} \frac{(np_n)^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k} (1-p_n)^{-k} (1-p_n)^n.$$

Or, pour tout  $k \in \{0, \dots, E(x)\}$ ,

$$\lim_{n \rightarrow +\infty} \frac{n(n-1) \cdots (n-k+1)}{n^k} (1-p_n)^{-k} = 1$$

car  $p_n \sim \frac{\lambda}{n} \rightarrow 0$  quand  $n \rightarrow +\infty$ , et

$$\frac{n(n-1) \cdots (n-k+1)}{n^k} = \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \rightarrow 1.$$

D'autre part,  $\lim_{n \rightarrow +\infty} (1-p_n)^n = \lim_{n \rightarrow +\infty} e^{n \ln(1-p_n)} = \lim_{n \rightarrow +\infty} e^{-np_n} = e^{-\lambda}$  donc finalement

$$\lim_{n \rightarrow +\infty} F_n(x) = \sum_{k=0}^{E(x)} \frac{\lambda^k}{k!} e^{-\lambda} = F(x).$$

□

Le résultat suivant donne des caractérisation de la convergence en loi à l'aide des fonctions caractéristique ou de la notion de convergence faible.

**Propriété 7.6.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. et soit  $X$  une v.a.r.

Les propriétés suivantes sont équivalentes :

– Convergence en loi :

$$X_n \xrightarrow{L} X$$

– Convergence des fonctions caractéristiques :

$$\phi_{X_n}(t) = \mathbb{E}(e^{itX_n}) \xrightarrow{n \rightarrow +\infty} \phi_X(t) = \mathbb{E}(e^{itX}), \forall t \in \mathbb{R}$$

– Convergence faible :

$$\mathbb{E}(f(X_n)) = \int_{\mathbb{R}} f(x) dP_{X_n}(x) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(f(X)) = \int_{\mathbb{R}} f(x) dP_X(x)$$

pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  continue bornée.

## 7.6 Loi des grands nombres

Un des résultats les plus importants en théorie des probabilités est la loi forte des grands nombres.

**Théorème 7.4** (admis). Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de variables aléatoires indépendantes, d'ordre 1, de même loi admettant une espérance  $m$ , et  $Z_n = \frac{X_1 + \dots + X_n}{n}$ . Alors  $Z_n \xrightarrow{p.s.} m$ .

La loi forte des grands nombres justifie l'intuition fréquentiste que la moyenne de variables aléatoires calculée à l'issue de  $n$  expériences indépendantes réalisées dans les mêmes conditions converge vers l'espérance commune des variables.

## 7.7 Théorème central limite - convergence vers la loi normale

Le **théorème central limite** (TCL), partie fondamentale de ce chapitre, sera admis.

**Théorème 7.5.** Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r. indépendantes d'ordre deux, de même loi, d'espérance  $m$  et de variance  $\sigma^2$ .

Soit  $S_n = X_1 + \dots + X_n$  et  $T_n = \frac{S_n - nm}{\sqrt{n}\sigma}$ . Alors, on a que  $T_n \xrightarrow{L} T$  où  $T$  est une v.a.r. de loi normale  $\mathcal{N}(0, 1)$ .

**Remarque :**  $\mathbb{E}(S_n) = nm$ ,  $\text{var}(S_n) = n\sigma^2$  et  $T_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}}$ .

*Démonstration.* sans perte de généralité, on peut supposer que  $m = 0$  et  $\sigma^2 = 1$ . On veut donc montrer que

$$\frac{S_n}{\sqrt{n}} \xrightarrow{L} T \text{ où } T \sim \mathcal{N}(0, 1)$$

Pour cela, on va montrer que

$$\phi_{\frac{S_n}{\sqrt{n}}}(t) \xrightarrow{n \rightarrow +\infty} \phi_T(t) = e^{-t^2/2} \text{ pour tout } t \in \mathbb{R}.$$

Par indépendance de  $X_1, \dots, X_n$ , on a que

$$\begin{aligned} \phi_{\frac{S_n}{\sqrt{n}}}(t) &= \mathbb{E}(e^{it\frac{S_n}{\sqrt{n}}}) = \mathbb{E}(e^{i\frac{t}{\sqrt{n}}(X_1 + \dots + X_n)}) \\ &= \prod_{k=1}^n \mathbb{E}(e^{i\frac{t}{\sqrt{n}}X_k}) = \prod_{k=1}^n \phi_{X_k}\left(\frac{t}{\sqrt{n}}\right) = \left(\phi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n. \end{aligned}$$

Etant donné que  $\mathbb{E}(X^2) < +\infty$ , la fonction  $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$  est deux fois dérivable avec

$$\phi'_X(0) = \mathbb{E}(X) = 0 \text{ et } \phi''_X(0) = -\mathbb{E}(X^2) = -1.$$

Donc, pour  $u \rightarrow 0$ , on a que

$$\begin{aligned} \phi_X(u) &= \phi_X(0) + \phi'_X(0)u + \frac{\phi''_X(0)}{2}u^2 + o(u^2) \\ &= 1 - \frac{u^2}{2} + o(u^2) \end{aligned}$$

Donc, pour  $n \rightarrow +\infty$

$$\begin{aligned} \phi_{\frac{S_n}{\sqrt{n}}}(t) &= \left(\phi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = e^{-t^2/2} + o(1) = \phi_T(t) + o(1), \end{aligned}$$

ce qui termine la démonstration. □

**Application :** Quand  $n$  est grand (il est classiquement admis, qu'en pratique, le TCL est valide lorsque  $n \geq 30$ ),  $S_n$  suit approximativement la loi normale  $\mathcal{N}(\mathbb{E}(S_n), \text{var}(S_n))$ , c'est-à-dire la loi  $\mathcal{N}(nm, n\sigma^2)$  et  $Z_n = \frac{S_n}{n}$  la loi normale  $\mathcal{N}(\mathbb{E}(Z_n), \text{var}(Z_n))$ , c'est-à-dire la loi  $\mathcal{N}(m, \frac{\sigma^2}{n})$ .

*Exemple :* on cherche à savoir combien de lancers on doit effectuer dans un jeu de pile ou face équilibré pour que la fréquence d'apparition de pile soit proche de  $1/2$  à  $\epsilon > 0$  près avec une probabilité supérieure à 95 %.

Soit  $\alpha = 0.05$ . Les résultats des lancers peuvent se modéliser par une suite de v.a.r.  $(X_n)_{n \in \mathbb{N}}$  indépendantes de loi de Bernoulli  $B(1/2)$ . L'espérance de ces variables est  $p = 1/2$  et leur variance est  $\sigma^2 = 1/4$ . On cherche donc un entier  $n$  à partir du quel

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - p\right| > \epsilon\right) \leq \alpha,$$

ce qui peut se ré-écrire sous la forme

$$P\left(|T_n| > \sqrt{n}\frac{\epsilon}{\sigma}\right) \leq \alpha,$$

avec  $T_n = \frac{X_1 + \dots + X_n - np}{\sqrt{n}\sigma}$ . Le théorème central limite peut alors s'utiliser pour l'approximation suivante

$$P\left(|T_n| > \sqrt{n}\frac{\epsilon}{\sigma}\right) \approx P\left(|Z| > \sqrt{n}\frac{\epsilon}{\sigma}\right) = 2\left(1 - \Phi\left(\sqrt{n}\frac{\epsilon}{\sigma}\right)\right),$$

où  $Z$  désigne une v.a.r. de loi  $\mathcal{N}(0, 1)$  et  $\Phi$  sa fonction de répartition. Il suffit donc de choisir

$$n \geq \frac{(\Phi^{-1}(1 - \alpha/2))^2}{\epsilon^2/\sigma^2}.$$

On a alors que  $\Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) \approx 1.96$ . Si  $\epsilon = 0.01$ , on obtient que  $n \geq 10000$  lancers suffisent à approcher la fréquence  $1/2$  au centième près avec 95% de chance de succès.

# Chapitre 8

## CONDITIONNEMENT

La notion de probabilité conditionnelle est la plus importante, mais aussi la plus délicate de la théorie des probabilités. Elle est introduite en particulier chaque fois que, pendant le déroulement d'une expérience aléatoire, une information partielle est fournie à l'expérimentateur. Un événement en conditionne un autre, si la réalisation de ce dernier dépend de la réalisation du premier. Les notions d'indépendance et de conditionnement sont donc étroitement liées.

### 8.1 Probabilités conditionnelles

#### 8.1.1 Introduction

Envisageons, l'exemple suivant :

- 1) On lance un dé. Quelle est la probabilité d'obtenir 3 ?
- 2) On lance un dé et on obtient un chiffre impair. Quelle est la probabilité que ce soit le chiffre 3 ?

Dans le cas 1), on répond  $1/6$ , mais dans le cas 2), on répond  $1/3$ .

Pour modéliser le cas 2), on peut prendre  $\Omega' = \{1, 3, 5\}$ , muni de l'équiprobabilité (probabilité  $P'$  telle que  $P'(\{1\}) = P'(\{3\}) = P'(\{5\}) = \frac{1}{3}$ ).

Mais on peut prendre aussi  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , muni de la probabilité  $Q$  telle que :

$$Q(\{1\}) = Q(\{3\}) = Q(\{5\}) = \frac{1}{3} \text{ et } Q(\{2\}) = Q(\{4\}) = Q(\{6\}) = 0.$$

Cette façon de procéder à l'avantage de conserver le même univers que dans le cas 1) et de montrer que la différence entre 1) et 2) est liée à un changement de probabilités.

**Définition 8.1.** Soit  $B \in \mathcal{A}$  tel que  $P(B) > 0$ . On appelle probabilité de  $A$  sachant  $B$  le nombre  $P^B(A)$  (ou  $P(A/B)$ ) défini par  $P^B(A) = \frac{P(A \cap B)}{P(B)}$ .

*Retour sur l'exemple précédent :*

On note  $A$  "on obtient le chiffre 3" et  $B$  "on obtient un chiffre impair". On a

$$A \cap B = A ; P(A \cap B) = \frac{\text{card}(A \cap B)}{\text{card}\Omega} = \frac{1}{6} ; P(B) = \frac{\text{card}B}{\text{card}\Omega} = \frac{3}{6} = \frac{1}{2}.$$

$$\text{D'où } P^B(A) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}.$$

**Propriété 8.1.** L'application  $P^B : A \mapsto P^B(A)$  est une probabilité sur  $(\Omega, \mathcal{A})$ .

*Démonstration.* i)  $P^B(\Omega) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$  et  $P^B(A) = \frac{P(A \cap B)}{P(B)} \in [0, 1]$  car  $A \cap B \subset B$ .  
 ii) Pour toute suite d'événements  $A_n$ , incompatibles 2 à 2, les événements  $A_n \cap B$  sont aussi incompatibles 2 à 2 et on a, en utilisant la distributivité de l'intersection par rapport à la réunion dénombrable, puis le fait que  $P$  est une probabilité :

$$\begin{aligned} P^B\left(\bigcup_n A_n\right) &= \frac{P\left(\left(\bigcup_n A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_n (A_n \cap B)\right)}{P(B)} \\ &= \frac{\sum_n P(A_n \cap B)}{P(B)} = \sum_n P^B(A_n) \end{aligned}$$

□

**Conséquence importante :**  $P^B(\bar{A}) = 1 - P^B(A)$ .

### Probabilités conditionnelles et indépendance :

**Théorème 8.1.** Si  $A$  et  $B$  sont deux événements indépendants de probabilité non nulle, alors

$$P^B(A) = P(A) \text{ et } P^A(B) = P(B).$$

*Démonstration.* Si  $A$  et  $B$  sont indépendants, alors  $P(A \cap B) = P(A)P(B)$  donc  $\frac{P(A \cap B)}{P(B)} = P(A)$  c'est-à-dire  $P^B(A) = P(A)$  et de même  $P^A(B) = \frac{P(A \cap B)}{P(A)} = P(B)$ . □

### 8.1.2 Formule des probabilités totales

**Théorème 8.2.** Soit  $(B_i)_{i \in I}$  un système complet d'événements de probabilités non nulles. Alors, pour tout  $A \in \mathcal{A}$ ,

$$P(A) = \sum_{i \in I} P^{B_i}(A)P(B_i).$$

*Démonstration.* Les  $B_i$  sont 2 à 2 disjoints et  $\Omega = \bigcup_{i \in I} B_i$  donc, par distributivité de l'intersection par rapport à l'union dénombrable :

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in I} B_i\right) = \bigcup_{i \in I} (A \cap B_i)$$

Or, comme les  $A \cap B_i$  sont aussi 2 à 2 disjoints, on a  $P(A) = \sum_{i \in I} P(A \cap B_i)$ .

On applique alors la formule des probabilités composées :  $P(A \cap B_i) = P^{B_i}(A)P(B_i)$ .  $\square$

**Cas particulier très utile :**  $P(A) = P^B(A)P(B) + P^{\bar{B}}(A)P(\bar{B})$ .

En effet,  $(B, \bar{B})$  forme un système complet d'événements de  $\Omega$ .

### 8.1.3 Théorème de Bayes

**Théorème 8.3.** Soit  $(B_i)_{i \in I}$  un système complet d'événements de probabilités non nulles et  $A$  un événement de probabilité non nulle. Alors, pour tout  $i_0 \in I$  :

$$P^A(B_{i_0}) = \frac{P^{B_{i_0}}(A)P(B_{i_0})}{\sum_{i \in I} P^{B_i}(A)P(B_i)}.$$

*Démonstration.* On a que  $P^A(B_{i_0}) = \frac{P(A \cap B_{i_0})}{P(A)}$  ainsi que  $P(A \cap B_{i_0}) = P^{B_{i_0}}(A)P(B_{i_0})$ . Ainsi, d'après la formule des probabilités totales (Théorème 8.2),  $P(A) = \sum_{i \in I} P^{B_i}(A)P(B_i)$ .  $\square$

**Cas particulier très important :**  $P^A(B) = \frac{P^B(A)P(B)}{P(A)} = \frac{P^B(A)P(B)}{P^B(A)P(B) + P^{\bar{B}}(A)P(\bar{B})}$ .

## 8.2 Conditionnement dans le cas discret

### 8.2.1 Lois conditionnelles pour un couple discret

On reprend les notations du chapitre 5 sur les couples de variables aléatoires discrètes. Soit  $X$  et  $Y$  deux v.a.r. discrètes, définies sur  $(\Omega, \mathcal{A}, P)$ , avec

$$X(\Omega) = \{x_i ; i \in I\} \text{ et } Y(\Omega) = \{y_j ; j \in J\}, \text{ } I, J \text{ dénombrables.}$$

On rappelle que la loi de probabilité du couple  $(X, Y)$  est déterminée par

$$p_{ij} = P([X = x_i] \cap [Y = y_j]) \text{ pour } i \in I \text{ et } j \in J.$$

et que les lois marginales de  $X$  et  $Y$  sont

$$P(X = x_i) = p_{i.} = \sum_j p_{ij} \quad \text{et} \quad P(Y = y_j) = p_{.j} = \sum_i p_{ij}.$$

**Définition 8.2.** Si  $p_{i.} \neq 0$ , la loi conditionnelle de  $Y$  sachant  $[X = x_i]$  est définie par :

$$P^{[X=x_i]}(Y = y_j) = P(Y = y_j | X = x_i) = \frac{P([X = x_i] \cap [Y = y_j])}{P(X = x_i)} = \frac{p_{ij}}{p_{i.}}, \quad j \in J.$$

Si  $p_{i.} = 0$ , alors, par convention,  $P^{[X=x_i]}(Y = y_j) = 0$ . On peut de même définir la loi conditionnelle de  $X$  sachant  $[Y = y_j]$ .

Lorsque l'on représente la loi du couple dans un tableau, la loi conditionnelle de  $X$  sachant  $[Y = y_j]$  est obtenue en prenant la  $j$ -ème colonne du tableau, divisée par la somme des  $p_{ij}$  sur la  $j$ -ème colonne et la loi conditionnelle de  $Y$  sachant  $[X = x_i]$  est obtenue en prenant la  $i$ -ème ligne, divisée par la somme des  $p_{ij}$  sur la  $i$ -ème ligne.

### 8.2.2 Espérance conditionnelle dans le cas discret

Soit  $X$  une v.a.r discrète définie sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $Y$  une v.a.r. discrète définie sur le même espace probabilisé et admettant une espérance ( $\mathbb{E}(|Y|)$  existe et est fini). Soit  $x_i \in X(\Omega)$  tel que  $P([X = x_i]) > 0$ . Montrons que l'espérance d'une v.a.r. discrète  $Z$  ayant pour loi la loi conditionnelle de  $Y$  à l'événement  $[X = x_i]$  (c'est-à-dire  $P([Z = y_j] = P^{[X=x_i]}([Y = y_j])$ ) pour tout  $j \in J$ ) existe. On a en effet :

$$\begin{aligned} \sum_{j \in J} |y_j| P^{[X=x_i]}([Y = y_j]) &= \sum_{j \in J} |y_j| \frac{P([X = x_i] \cap [Y = y_j])}{P([X = x_i])} \\ &= \frac{1}{P([X = x_i])} \sum_{j \in J} |y_j| P([X = x_i] \cap [Y = y_j]) \end{aligned}$$

avec  $\sum_{j \in J} |y_j| P([X = x_i] \cap [Y = y_j]) \leq \sum_{j \in J} |y_j| P([Y = y_j]) < +\infty$ . Ainsi,

**Définition 8.3.** L'espérance d'une v.a.r. dont la loi est la loi conditionnelle de  $Y$  à l'événement  $[X = x_i]$  est appelée espérance conditionnelle de  $Y$  à l'événement  $[X = x_i]$ . Elle est notée  $\mathbb{E}^{[X=x_i]}(Y)$ . On a donc

$$\mathbb{E}^{[X=x_i]}(Y) = \sum_{j \in J} y_j P^{[X=x_i]}([Y = y_j]).$$

*Exemple* : Soit  $X$  une v.a.r. de Poisson de paramètre  $\lambda > 0$  et  $Y$  une v.a.r. de Poisson de paramètre  $\mu > 0$  avec  $X$  et  $Y$  indépendantes. On a que  $X + Y$  est de loi de Poisson de paramètre  $\lambda + \mu$  (cf. exercices en PC). On a alors, si  $0 \leq i \leq n$ ,

$$\begin{aligned} P^{[X+Y=n]}([X = i]) &= \frac{P([X = i] \cap [X + Y = n])}{P([X + Y = n])} = \frac{P([X = i] \cap [Y = n - i])}{P([X + Y = n])} \\ &= \frac{P([X = i])P([Y = n - i])}{P([X + Y = n])} = \frac{e^{-\lambda} \frac{\lambda^i}{i!} e^{-\mu} \frac{\mu^{n-i}}{(n-i)!}}{e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^n}{n!}} \\ &= \frac{n!}{i!(n-i)!} \left( \frac{\lambda}{\lambda + \mu} \right)^i \left( \frac{\mu}{\lambda + \mu} \right)^{n-i} = C_n^i \left( \frac{\lambda}{\lambda + \mu} \right)^i \left( 1 - \frac{\lambda}{\lambda + \mu} \right)^{n-i} = P_Z(\{i\}) \end{aligned}$$

où  $Z$  est de loi binomiale  $\mathcal{B}\left(n, \frac{\lambda}{\lambda + \mu}\right)$ . Ainsi, la loi conditionnelle de  $X$  à  $[X + Y = n]$  est la loi binomiale  $\mathcal{B}\left(n, \frac{\lambda}{\lambda + \mu}\right)$ , d'espérance  $\frac{n\lambda}{\lambda + \mu}$  et on a donc

$$\mathbb{E}^{[X+Y=n]}(X) = \frac{n\lambda}{\lambda + \mu}.$$

**Définition 8.4.** Soit  $X$  une v.a.r. discrète sur  $(\Omega, \mathcal{A}, P)$  telle que, pour tout  $x \in X(\Omega)$ ,  $P([X = x]) \neq 0$  et soit  $Y$  une v.a.r. discrète sur  $(\Omega, \mathcal{A}, P)$  admettant une espérance. On appelle espérance conditionnelle de  $Y$  sachant  $X$ , notée  $\mathbb{E}^X(Y)$ , la v.a.r. discrète :

$$\mathbb{E}^X(Y) = h(X),$$

avec  $h(x) = \mathbb{E}^{[X=x]}(Y)$  pour tout  $x \in X(\Omega)$ .

Exemple : Dans l'exemple précédent ( $P_X = \mathcal{P}(\lambda)$ ,  $P_Y = \mathcal{P}(\mu)$  avec  $X$  et  $Y$  indépendantes), on a

$$\mathbb{E}^{X+Y}(X) = h(X+Y) = \frac{(X+Y)\lambda}{\lambda+\mu}$$

avec  $h(n) = \mathbb{E}^{[X+Y=n]}(X) = \frac{n\mu}{\lambda+\mu}$ .

**Théorème 8.4. (Théorème de l'espérance totale)**

Soit  $X$  et  $Y$  deux v.a.r. discrètes définies sur le même espace telles que  $\mathbb{E}(Y)$  existe. Alors la v.a.r. discrète  $\mathbb{E}^X(Y)$  admet une espérance et

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

*Démonstration.* On rappelle que  $\mathbb{E}(\varphi(X)) = \sum_{x \in X(\Omega)} \varphi(x)P([X = x])$  (sous réserve d'existence, c'est-à-dire de convergence absolue) que l'on applique ici à  $\varphi(X) = \mathbb{E}^X(Y)$ , c'est-à-dire que

$$\varphi(x) = \mathbb{E}^{[X=x]}(Y) = \sum_{y \in Y(\Omega)} yP^{[X=x]}([Y = y])$$

pour tout  $x \in X(\Omega)$ . On a alors que

$$\begin{aligned} \mathbb{E}(\mathbb{E}^X(Y)) &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} yP^{[X=x]}([Y = y])P([X = x]) \\ &= \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} yP([X = x] \cap [Y = y]) \\ &= \sum_{y \in Y(\Omega)} y \left( \sum_{x \in X(\Omega)} P([X = x] \cap [Y = y]) \right) = \sum_{y \in Y(\Omega)} yP([Y = y]) = \mathbb{E}(Y) \end{aligned}$$

les interversions de sommes étant légitimes puisque

$$\sum_{y \in Y(\Omega)} |y|P([Y = y]) = \sum_{y \in Y(\Omega)} |y| \sum_{x \in X(\Omega)} P([X = x] \cap [Y = y]) < +\infty.$$

□

## 8.3 Conditionnement dans le cas continu

### 8.3.1 Loix conditionnelles dans le cas continu

On reprend les notations du chapitre 6 sur les couples de variables aléatoires à densité.

**Définition 8.5.** Pour tout  $x \in \mathbb{R}$  tel que  $f_X(x) \neq 0$ , la densité conditionnelle de  $Y$  sachant  $[X = x]$  est la fonction  $f_Y^{X=x}$  définie par  $f_Y^{X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ .

**Remarque :** la fonction  $f_Y^{X=x}$  est bien une densité car elle est positive et  $\int_{-\infty}^{+\infty} f_Y^{X=x}(y) dy = 1$ .

**Parallèle avec les couples discrets :**

Ceci est l'analogie de  $\frac{p_{ij}}{p_i} = \frac{p_{ij}}{\sum_k p_{ik}}$ .

**Propriété 8.2.** Si  $X$  et  $Y$  sont indépendantes, alors  $f_Y^{X=x} = f_Y$  pour tout  $x$  tel que  $f_X(x) \neq 0$  et  $f_X^{Y=y} = f_X$  pour tout  $y$  tel que  $f_Y(y) \neq 0$ .

*Démonstration.*  $f_Y^{X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$  donc, si  $X$  et  $Y$  sont indépendantes,  $f_{X,Y}(x,y) = f_X(x)f_Y(y)$  et

$$f_Y^{X=x}(y) = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y).$$

La deuxième assertion se démontre de la même façon,  $X$  et  $Y$  jouant des rôles symétriques.  $\square$

### 8.3.2 Espérance conditionnelle dans le cas continu

Soit  $(X, Y)$  un couple de v.a.r. absolument continu, de densité  $f_{X,Y}$  et de densités marginales  $f_X$  et  $f_Y$  tel que  $\mathbb{E}(Y)$  existe ( $\mathbb{E}(|Y|)$  existe et est fini). On a vu que la loi conditionnelle  $P_Y^{[X=x]}$  admet dans ce cas la densité définie par  $f_Y^{X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$  pour tout  $x \in \mathbb{R}$  tel que  $f_X(x) > 0$ . On a alors :

$$\int_{\mathbb{R}} |y| f_{X,Y}(x,y) dy < +\infty.$$

**Définition 8.6.** L'espérance conditionnelle de  $Y$  à l'événement  $[X = x]$  est le réel :

$$\mathbb{E}^{X=x}(Y) = \int_{\mathbb{R}} y f_Y^{X=x}(y) dy.$$

L'espérance conditionnelle de  $Y$  sachant  $X$  est la v.a.r.  $\mathbb{E}^X(Y) = h(X)$  avec :

$$h(x) = \mathbb{E}^{X=x}(Y).$$

On a ici, tout comme dans le cas discret, le théorème de l'espérance totale si  $\mathbb{E}(Y)$  existe :

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

(La démonstration est la même en remplaçant les sommes par des intégrales et les probabilités par des densités).

Exemple : Soit  $(X, Y)$  de densité  $\lambda^2 e^{-\lambda y}$  et de support  $D = \{(x, y) ; 0 \leq x \leq y\}$ .

On a alors  $f_X(x) = \int_x^{+\infty} \lambda^2 e^{-\lambda y} dy \mathbb{1}_{[0,+\infty[}(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,+\infty[}(x)$  ( $X$  est de loi exponentielle  $\mathcal{E}(\lambda)$ ) et

$$f_Y^{X=x}(y) = \lambda e^{-\lambda(y-x)} \mathbb{1}_{[x,+\infty[}(y) ;$$

l'espérance de  $Y$  conditionnée par  $X = x$  est égale à :

$$\mathbb{E}^{X=x}(Y) = \int_x^{+\infty} y\lambda e^{-\lambda(y-x)} dy = \int_0^{+\infty} (x+u)\lambda e^{-\lambda u} du = x + \frac{1}{\lambda}$$

(en faisant le changement de variable  $u = y - x$ ). D'où il vient :

$$\mathbb{E}^X(Y) = X + \frac{1}{\lambda}.$$

## 8.4 Complément sur les lois conditionnelles

Dans les sections précédentes, on a défini la loi conditionnelle dans le cas d'un couple de v.a.r. discret et dans le cas d'un couple de v.a.r. absolument continu. On va donner ici rapidement les analogues dans le cas où l'une des deux v.a.r. est discrète et l'autre absolument continue. On reprendra également les notations des chapitres 5 et 6.

### 8.4.1 Loi d'une variable absolument continue $Y$ conditionnée par une variable discrète $X$

On considère des couples aléatoires mixtes dont l'une des composantes  $X$  est discrète tandis que l'autre  $Y$  est continue.

**Définition 8.7.** Un couple aléatoire  $(X, Y)$  défini sur  $D \times \mathbb{R}$  où  $D$  est un espace au plus dénombrable est dit mixte si sa loi  $P_{X,Y}$  est définie par :

$$\text{pour tout } (x_k, B) \in D \times \mathcal{B}_{\mathbb{R}}, P_{X,Y}(\{x_k\}, B) = \int_B f_{X,Y}(x_k, y) dy$$

où  $f_{X,Y}$  est une fonction borélienne positive vérifiant  $\sum_k \int_{\mathbb{R}} f_{X,Y}(x_k, y) dy = 1$ .

Les lois marginales sont ici :

$$f_Y(y) = \sum_k f_{X,Y}(x_k, y) \text{ et } P(X = x_k) = P_X(\{x_k\}) = \int_{\mathbb{R}} f_{X,Y}(x_k, y) dy.$$

Exemple : Soit le couple  $(X, Y)$  défini sur  $\mathbb{N} \times \mathbb{R}_+$  pour lequel :

$$f_{X,Y}(k, y) = e^{-2y} \frac{y^k}{k!}.$$

Il est facile de montrer que :  $\int_{\mathbb{R}_+} \sum_k f_{X,Y}(k, y) dy = 1$ .

**Rappel** :  $B(u, v) = \int_0^1 t^{u-1}(1-t)^{v-1} dt = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$  où  $\Gamma(z) = \int_0^{+\infty} e^{-t} t^{z-1} dt$  et  $\Gamma(n) = (n-1)!$  si  $n \in \mathbb{N}^*$ .

$f_Y(y) = \sum_{k=0}^{+\infty} f_{X,Y}(k, y) = \sum_{k=0}^{+\infty} e^{-2y} \frac{y^k}{k!} \mathbb{I}_{[0, +\infty[}(y) = e^{-y} \mathbb{I}_{[0, +\infty[}(y)$  :  $Y$  est une v.a. exponentielle  $\mathcal{E}(1)$ .

$P_X(\{k\}) = \int_{\mathbb{R}_+} f_{X,Y}(k, y) dy = \int_0^{+\infty} e^{-2y} \frac{y^k}{k!} dy = \left(\frac{1}{2}\right)^{k+1}$  ;  $X$  est de loi géométrique  $\mathcal{G}$  sur  $\mathbb{N}$  de paramètre  $\frac{1}{2}$ .

**Théorème 8.5** (Admis). *Étant donné le couple mixte  $(X, Y)$ , la densité conditionnelle de  $Y$  sachant  $[X = x_k]$  est définie par :*

$$f_Y^{[X=x_k]}(y) = \frac{f_{X,Y}(x_k, y)}{P_X(\{x_k\})}.$$

#### 8.4.2 Loi d'une variable discrète conditionnée par une variable absolument continue

**Théorème 8.6** (Admis). *La loi de probabilité conditionnelle de la v.a.r. discrète  $X$  sachant  $[Y = y]$  est définie, pour tout  $y$  tel que  $f_Y(y) > 0$  par :*

$$\text{pour tout } x_k, P^{Y=y}([X = x_k]) = \frac{f_{X,Y}(x_k, y)}{f_Y(y)}.$$

### 8.5 Compléments sur l'espérance conditionnelle

Dans cette partie, on va donner une autre approche de l'espérance conditionnelle, très utile dans le cas des variables ayant des moments d'ordre 2. En effet, dans ce cas, il est possible de travailler avec une structure hilbertienne pour la représentation de variables aléatoires.

**Définition 8.8.** *Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, et  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une v.a.r. (discrète ou continue) i.e. une application mesurable. On note par  $L^2(\Omega, \mathcal{A}, P)$  l'espace des v.a.r.  $X$  de carré intégrable i.e. telles que  $\mathbb{E}(X^2) < +\infty$ .*

**Théorème 8.7.** *L'espace  $L^2(\Omega, \mathcal{A}, P)$  est un espace de Hilbert lorsqu'il est muni du produit scalaire*

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

*et de la norme quadratique associée  $\|X\|_2 = \sqrt{\mathbb{E}(X^2)}$ .*

Le cadre de la géométrie hilbertienne est particulièrement adapté à la modélisation et à la compréhension des propriétés de l'espérance conditionnelle. On vient de définir l'espace  $L^2(\Omega, \mathcal{A}, P)$  des variables aléatoires  $X$  de carré intégrable (i.e.  $\mathbb{E}(X^2)$  fini) dans lequel on définit le produit scalaire  $\langle X, Y \rangle = \mathbb{E}(XY)$  des v.a.r.  $X$  et  $Y$ . En analyse hilbertienne, on définit la norme quadratique  $\|X\|_2 = \sqrt{\mathbb{E}(X^2)}$ , pour laquelle l'espace  $L^2(\Omega, \mathcal{A}, P)$  est un espace de Hilbert. Afin d'introduire un autre caractérisation de l'espérance conditionnelle, rappelons tout d'abord quelques propriétés de la projection orthogonale dans les espace de Hilbert.

**Propriété 8.3.** Soit  $(V, \langle \cdot, \cdot \rangle, \|\cdot\|)$  un espace de Hilbert et  $V_0 \subset V$  un sous-espace vectoriel fermé de  $V$ . Soit  $v \in V$ . La projection orthogonale de  $v$  sur  $V_0$  est le vecteur de  $v_0 \in V_0$  tel que

$$\|v - v_0\|^2 \leq \|v - u\|^2, \text{ pour tout } u \in V_0.$$

Il est aussi caractérisé par la propriété

$$\langle v - v_0, u \rangle = 0, \text{ pour tout } u \in V_0.$$

Pour donner un peu d'intuition sur le lien entre espérance conditionnelle et projection orthogonale, considérons le cas où  $X$  et  $Y$  deux v.a.r. discrètes, définies sur  $(\Omega, \mathcal{A}, P)$ , avec

$$X(\Omega) = \{x_i ; i \in I\} \text{ et } Y(\Omega) = \{y_j ; j \in J\}, \text{ } I, J \text{ dénombrables .}$$

On rappelle que  $\mathbb{E}(Y|X = x_i) = \sum_j y_j P(Y = y_j | X = x_i)$ . Soit  $h : \mathbb{R} \rightarrow \mathbb{R}$  une fonction déterministe. On a alors que

$$\begin{aligned} \mathbb{E}(\mathbb{E}(Y|X)h(X)) &= \sum_i \mathbb{E}(Y|X = x_i)h(x_i)P(X = x_i) \\ &= \sum_i \left( \sum_j y_j P(Y = y_j | X = x_i) \right) h(x_i)P(X = x_i) \\ &= \sum_i \sum_j y_j h(x_i)P([Y = y_j] \cap [X = x_i]) = \mathbb{E}(Yh(X)) \end{aligned}$$

On a donc montré que pour toute fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathbb{E} \left( \underbrace{(Y - \mathbb{E}(Y|X))}_{v-v_0} \underbrace{h(X)}_u \right) = 0 \text{ avec } v = Y, v_0 = \mathbb{E}(Y|X),$$

ce qui permet donc d'interpréter  $v_0 = \mathbb{E}(Y|X)$  comme la projection de  $v = Y$  dans l'espace des fonctions de la forme  $h(X)$  où  $h : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction déterministe. Les v.a.r de la forme  $h(X)$  s'apparentent à l'ensemble des fonctions mesurables par rapport à la tribu engendrée par  $X$  et peuvent s'interpréter comme la classe des v.a.r que l'on peut construire à partir de la connaissance de  $X$ . Plus précisément, on a les propriétés suivantes.

**Définition 8.9** (Information disponible par la connaissance de  $X$ ). Soit  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une v.a.r. On note par

$$\mathcal{B}_X = X^{-1}(\mathcal{B}_{\mathbb{R}}) = \{X^{-1}(B) ; B \in \mathcal{B}_{\mathbb{R}}\} \subset \mathcal{A}$$

la tribu engendrée par  $X$ .

**Théorème 8.8** (Doob). Soit  $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  et  $Y : \Omega \rightarrow \mathbb{R}$  deux v.a.r. Alors  $Y$  est mesurable par rapport à la tribu  $\mathcal{B}_X$  ssi il existe une fonction  $h : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  mesurable telle que

$$Y = h(X).$$

Le théorème suivant est la caractérisation hilbertienne de l'espérance conditionnelle.

**Définition 8.10.** Soit  $L^2(\Omega, \mathcal{B}_X, P)$  l'espace des v.a.r.  $Z : \Omega \rightarrow \mathbb{R}$  de carré intégrable et mesurables par rapport à  $\mathcal{B}_X$ . Alors  $L^2(\Omega, \mathcal{B}_X, P)$  est un sous-espace vectoriel fermé de  $L^2(\Omega, \mathcal{A}, P)$ .

**Théorème 8.9** (et Définition). Soit  $X$  et  $Y$  deux v.a. quelconques de  $L^2(\Omega, \mathcal{A}, P)$ . Alors,

$\mathbb{E}(Y|X)$  est l'espérance conditionnelle de  $Y$  sachant  $X$

$\Updownarrow$

$\mathbb{E}(Y|X)$  est la **projection orthogonale** de  $Y$  sur  $L^2(\Omega, \mathcal{B}_X, P)$ .

On a alors les propriétés suivantes :

**Propriétés 8.1 :**

(1) Linéarité :  $\mathbb{E}^Z(aX + bY) = a\mathbb{E}^Z(X) + b\mathbb{E}^Z(Y)$  avec  $a$  et  $b$  réels,  $X, Y, Z$  des v.a.r. quelconques.

(2) Croissance : si  $Y \geq X$  alors pour toute v.a.r.  $Z$ ,  $\mathbb{E}^Z(Y) \geq \mathbb{E}^Z(X)$

(3) Si  $Y$  est indépendante de  $X$ , alors  $\mathbb{E}^X(Y) = \mathbb{E}(Y)$

(4) Pour toute v.a.r.  $Y$  intégrable et toute fonction  $h$  bornée :

$$\mathbb{E}^X(h(X)Y) = h(X)\mathbb{E}^X(Y)$$

(5) Propriété de l'espérance totale : pour tout couple aléatoire  $(X, Y)$  :

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

(6)  $\mathbb{E}(h(X)Y) = \mathbb{E}(h(X)\mathbb{E}^X(Y))$  qui généralise la propriété précédente.

(7) Majoration de la variance :

$$\text{var}(\mathbb{E}^X(Y)) \leq \text{var}(Y)$$

(Si  $X$  conditionne  $Y$ , toute observation de  $X$  apporte une information sur  $Y$  et en réduit donc la variance).

Le théorème suivant établit que l'espérance  $\mathbb{E}(X)$  est la meilleure approximation (quadratique) de  $X$  par une constante.

**Théorème 8.10.** Pour toute v.a.r.  $X$  de  $L^2(\Omega, \mathcal{A}, P)$ ,  $\mathbb{E}(X)$  s'identifie à la projection de  $X$  sur le sous-espace des v.a.r. constantes.

## 8.6 Régression linéaire de Y par rapport à X

**Théorème 8.11.** *La meilleure approximation quadratique de Y par une fonction affine de X (supposée de variance non nulle) est la variable aléatoire  $\hat{Y}$  définie par :*

$$\hat{Y} = \left( \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \right) X + \left( \mathbb{E}(Y) - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \mathbb{E}(X) \right)$$

La droite de régression de Y par rapport à X, est définie par l'équation :

$$y = ax + b \text{ où } a = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \text{ et } b = \mathbb{E}(Y) - a\mathbb{E}(X).$$

*Démonstration.* Il s'agit de minimiser  $\phi(a, b) = \mathbb{E}((Y - (aX + b))^2)_{(a,b) \in \mathbb{R}^2}$ .

La fonction  $\phi(a, b)$  est convexe ; son minimum est atteint pour les valeurs de  $a$  et  $b$  qui annulent ses dérivées partielles :

$$\begin{cases} \frac{\partial \phi}{\partial a} = 0 & \Leftrightarrow \mathbb{E}[X(Y - (aX + b))] = 0 \\ \frac{\partial \phi}{\partial b} = 0 & \Leftrightarrow \mathbb{E}[Y - (aX + b)] = 0 \end{cases} \longrightarrow \begin{cases} \mathbb{E}(X^2)a + \mathbb{E}(X)b & = \mathbb{E}(XY) \\ \mathbb{E}(X)a + b & = \mathbb{E}(Y) \end{cases}$$

d'où, sachant que  $\sigma_X \neq 0$  :

$$\begin{cases} a = \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{\sigma_X^2} = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \\ b = \mathbb{E}(Y) - a\mathbb{E}(X) \end{cases} . \quad \square$$

On rassemble ici les deux types d'approximation quadratique d'une v.a.r. Y en fonction d'une v.a.r. X :

**Théorème 8.12.** *Si  $(X, Y)$  est un vecteur aléatoire de carrés intégrables, les propriétés suivantes sont vérifiées :*

(1) *La meilleure approximation quadratique de Y par une fonction affine dépendant de X est la v.a.  $\hat{Y}$  :*

$$\hat{Y} = \left( \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \right) X + \left( \mathbb{E}(Y) - \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} \mathbb{E}(X) \right)$$

(2)  *$\mathbb{E}^X(Y)$  est la meilleure approximation quadratique de Y par une fonction de X i.e.*

$$\mathbb{E}((Y - \mathbb{E}^X(Y))^2) \leq \mathbb{E}((Y - h(X))^2)$$

*pour toute fonction mesurable  $h : \mathbb{R} \rightarrow \mathbb{R}$ .*

## 8.7 Variance conditionnelle

**Définition 8.11.** *Étant données deux v.a.r.  $X$  et  $Y$ , la variance conditionnelle de  $Y$  sachant  $X$  est la variable aléatoire :*

$$\mathbb{E}^X ((Y - \mathbb{E}^X(Y))^2) =_{(\text{noté})} \text{var}^X(Y).$$

$\text{var}^{X=x}(Y)$  exprime la dispersion quadratique de  $Y$  autour de l'espérance conditionnelle  $\mathbb{E}^{X=x}(Y)$ .

**Théorème 8.13. (Propriété de la variance totale)**

*Pour tout couple de variables aléatoires  $(X, Y)$  :*

$$\text{var}(Y) = \mathbb{E}(\text{var}^X(Y)) + \text{var}(\mathbb{E}^X(Y))$$

*Dans le cadre de la géométrie hilbertienne de  $L^2(\Omega, \mathcal{A}, P)$ , la propriété de la variance totale équivaut au théorème de Pythagore appliqué au triangle rectangle de côtés  $Y$ ,  $\mathbb{E}(Y)$ ,  $\mathbb{E}^X(Y)$ .*

*Démonstration.*

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

Le dernier terme est égal à  $\text{var}(\mathbb{E}^X(Y))$ , le premier terme est égal à  $\mathbb{E}(\text{var}^X(Y))$ ; quant au terme médian, il est nul.  $\square$

## Chapitre 9

# VECTEURS ALÉATOIRES - VECTEURS GAUSSIENS

Ce chapitre est une généralisation en dimension  $n > 2$  des notions vues précédemment sur les couples de variables aléatoires. En dimension supérieure à 2, on parle plutôt de vecteurs aléatoires, et nous insisterons sur la notion de vecteurs gaussiens. Par convention, un vecteur aléatoire  $X$  de dimension  $n$  sera considéré comme un vecteur colonne de  $\mathbb{R}^n$  et ses composantes sont des variables aléatoires réelles qui seront notées  $X_1, \dots, X_n$ .

### 9.1 Vecteurs aléatoires

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé qui rend compte d'une expérience aléatoire. Un vecteur aléatoire (de dimension  $n$ ) est une application mesurable de  $(\Omega, \mathcal{A}, P)$  dans  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ . La loi  $P_X$  du vecteur aléatoire  $X$  est une mesure de probabilité sur  $\mathbb{R}^n$  muni de sa tribu borélienne  $\mathcal{B}_{\mathbb{R}^n}$ . La loi de  $X$ , dite mesure image, est définie par

$$\forall A \in \mathcal{B}_{\mathbb{R}^n}, P_X(A) = P(X \in A).$$

**Définition 9.1.** Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire de loi  $P$ . Les coordonnées de  $X$  sont dites indépendantes si, pour tout  $1 \leq p \leq n$  et pour tout  $\{i_1, \dots, i_p\} \subset \{1, \dots, n\}$ , on a que

$$P(\{X_{i_1} \in A_{i_1}\} \cap \dots \cap \{X_{i_p} \in A_{i_p}\}) = P(\{X_{i_1} \in A_{i_1}\}) \times \dots \times P(\{X_{i_p} \in A_{i_p}\}),$$

pour tout  $A_{i_1}, \dots, A_{i_p} \in \mathcal{B}_{\mathbb{R}}$ .

**Définition 9.2.** L'espérance d'un vecteur aléatoire  $X$  de dimension  $n$  est le vecteur de  $\mathbb{R}^n$  constitué des espérances de chacune des coordonnées i.e.

$$\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))'.$$

**Propriété 9.1.** Soit  $X$  un vecteur aléatoire de dimension  $n$ . Soit  $m \geq 1$  un entier positif,  $b$  un vecteur de  $\mathbb{R}^m$  et  $A$  une matrice de dimension  $m \times n$ . L'espérance du vecteur aléatoire  $Y = AX + b$  est

$$\mathbb{E}(Y) = A\mathbb{E}(X) + b.$$

*Démonstration.* Le résultat découle directement de la définition de l'espérance d'un vecteur aléatoire et de la propriété de linéarité de l'espérance vue précédemment pour des v.a. réelles.  $\square$

Dans la suite de ce chapitre, on va s'intéresser aux vecteurs aléatoires à densité (par rapport à la mesure de Lebesgue). Le cas des vecteurs aléatoires prenant un nombre fini ou dénombrable de valeurs est une généralisation immédiate des notions vues précédemment sur les couples de v.a. discrètes.

### 9.1.1 Loi d'un vecteur aléatoire à densité

**Définition 9.3.** La loi du vecteur aléatoire  $X$  de dimension  $n$  est dite absolument continue s'il existe une fonction mesurable  $f : (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , appelée densité de  $X$ , telle que

i)  $f(x) \geq 0$  pour tout  $x \in \mathbb{R}^n$  ;

ii)  $\int_{\mathbb{R}^n} f(x) dx$  existe et vaut 1 ;

iii)  $\forall A \in \mathcal{B}_{\mathbb{R}^n}, P_X(A) = \int_A f(x) dx$ .

Dans le cas où  $P_X$  est absolument continue, on dira (plus simplement) que le vecteur aléatoire  $X$  admet pour densité  $f$ .

**Propriété 9.2** (admis). Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire de dimension  $n$  admettant pour densité  $f$ . La coordonnée  $X_i$  du vecteur  $X$  est une variable aléatoire réelle qui admet pour densité la fonction  $f_{X_i}$  (dite loi marginale) définie par

$$f_{X_i}(x_i) = \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \text{ pour tout } x_i \in \mathbb{R}.$$

**Propriété 9.3** (admis). Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire de dimension  $n$  et de densité  $f$ . Les coordonnées de  $X$  sont indépendantes si et seulement si

$$f(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$$

pour tout  $(x_1, \dots, x_n) \in \mathbb{R}^n$ .

### 9.1.2 Matrice de covariance

**Définition 9.4.** Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire de dimension  $n$ . On appelle matrice de covariance de  $X$ , la matrice carrée de taille  $n \times n$ , notée  $\Sigma_X$ , dont les coefficients  $(\Sigma_{i,j})_{1 \leq i,j \leq n}$  sont donnés par

$$\Sigma_{i,j} = \text{cov}(X_i, X_j) = \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))).$$

En notation matricielle, la matrice de covariance d'un vecteur aléatoire  $X$  est

$$\Sigma_X = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))').$$

**Propriété 9.4** (admis). Soit  $X$  un vecteur aléatoire de dimension  $n$  et de matrice de covariance  $\Sigma_X$ . Soit  $A$  une matrice de dimension  $m \times n$  avec  $m$  un entier positif. La matrice de covariance  $\Sigma_Y$  du vecteur aléatoire  $Y = AX$  (de dimension  $m$ ) est

$$\Sigma_Y = A\Sigma_X A'.$$

**Propriété 9.5** (admis). Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire de dimension  $n$ . La matrice de covariance  $\Sigma_X = (\Sigma_{i,j})_{1 \leq i,j \leq n}$  de  $X$  est une matrice symétrique semi-définie positive i.e.

$$\sum_{1 \leq i,j \leq n} x_i \Sigma_{i,j} x_j \geq 0, \quad \forall x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Si les coordonnées de  $X$  sont indépendantes, alors  $\Sigma_X$  est une matrice diagonale.

**Attention :** la réciproque de la seconde assertion de la Proposition 9.5 n'est pas vraie. Si la matrice de covariance d'un vecteur aléatoire  $X$  est diagonale, cela n'implique pas que les coordonnées de  $X$  sont indépendantes.

### 9.1.3 Changement de variables

Soit  $X$  un vecteur aléatoire de densité  $f_X$ . Le support de  $f$  est le sous-ensemble  $D$  de  $\mathbb{R}^n$  défini par

$$D = \{x \in \mathbb{R}^n; f(x) > 0\}.$$

Dans la suite, on supposera que  $D$  est un ouvert de  $\mathbb{R}^n$ .

**Propriété 9.6** (admis). Soit  $X$  un vecteur aléatoire de densité  $f_X$  de support  $D$  ouvert. Soit  $\varphi$  une application mesurable de  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$  dans  $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ , et le vecteur aléatoire  $Z = \varphi(X)$ . Supposons que  $\varphi$  soit un  $C^1$  difféomorphisme de  $D$  dans son image  $\varphi(D)$ . Alors, le vecteur aléatoire  $Z$  admet pour densité la fonction  $f_Z$  donnée par

$$\forall z \in \varphi(D), f_Z(z) = |\text{Jac}(\varphi^{-1})(z)| f_X(\varphi^{-1}(z)),$$

où

$$\text{Jac}(\varphi^{-1})(z) = \begin{vmatrix} \frac{\partial \varphi_1^{-1}(z)}{\partial z_1} & \dots & \frac{\partial \varphi_1^{-1}(z)}{\partial z_n} \\ \vdots & \dots & \vdots \\ \frac{\partial \varphi_n^{-1}(z)}{\partial z_1} & \dots & \frac{\partial \varphi_n^{-1}(z)}{\partial z_n} \end{vmatrix}$$

avec  $\varphi^{-1}(z) = (\varphi_1^{-1}(z), \dots, \varphi_n^{-1}(z))'$ .

### 9.1.4 Fonction caractéristique

**Définition 9.5.** Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire de dimension  $n$ . La fonction caractéristique de  $X$  est la fonction à valeurs complexes définie par

$$\forall t = (t_1, \dots, t_n)' \in \mathbb{R}^n, \phi_X(t) = \mathbb{IE} \left( e^{i \sum_{k=1}^n t_k X_k} \right).$$

Si  $X$  admet pour densité  $f_X$ , alors la fonction caractéristique  $\phi_X(t)$  est la transformée de Fourier de  $f_X$ . Par la formule d'inversion de Fourier, on a donc que

$$\forall x = (x_1, \dots, x_n)' \in \mathbb{R}^n \in \mathbb{R}, f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i \sum_{k=1}^n t_k x_k} \phi_X(t) dt.$$

La loi d'un vecteur aléatoire (à densité) est donc entièrement déterminée par sa fonction caractéristique. On a en particulier les résultats suivants.

**Propriété 9.7** (admis). *Deux vecteurs aléatoires  $X$  et  $Y$  ont même loi si et seulement si  $\phi_X = \phi_Y$ .*

**Propriété 9.8** (admis). *Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire de dimension  $n$ . Les coordonnées de  $X$  sont indépendantes si et seulement si la fonction caractéristique de  $X$  est le produit des fonctions caractéristiques de ses coordonnées i.e.*

$$\forall t = (t_1, \dots, t_n)' \in \mathbb{R}^n, \phi_X(t) = \prod_{k=1}^n \phi_{X_k}(t_k)$$

Avant d'aborder la partie suivante sur les vecteurs gaussiens, il est utile d'énoncer également le résultat suivant qui donne la fonction caractéristique d'une v.a.r. de loi normale.

**Propriété 9.9** (admis). *Soit  $X$  une variable aléatoire de loi normale, d'espérance  $m$  et de variance  $\sigma^2$ . Alors*

$$\forall t \in \mathbb{R}, \phi_X(t) = \exp\left(itm - \frac{1}{2}t^2\sigma^2\right).$$

## 9.2 Vecteurs gaussiens

### 9.2.1 Caractérisation des vecteurs gaussiens

**Définition 9.6.** *On dit que  $X = (X_1, \dots, X_n)'$  est un vecteur aléatoire gaussien, si toute combinaison linéaire à coefficients réels des coordonnées  $X_1, \dots, X_n$  de  $X$  suit une loi normale de moyenne  $m$  et de variance  $\sigma^2$  (avec éventuellement  $\sigma^2 = 0$ ).*

Ainsi, si  $X = (X_1, \dots, X_n)'$  est un vecteur gaussien on a donc en particulier que chacune de ses coordonnées  $X_i$  suit une loi normale. Toutefois, la réciproque n'est pas vraie car on peut construire un vecteur aléatoire non-gaussien dont toutes les coordonnées suivent une loi normale. Dans le cas de coordonnées indépendantes, on néanmoins le résultat suivant.

**Propriété 9.10.** *Soit  $X_1, \dots, X_n$  des variables aléatoires réelles indépendantes de loi normale centrée réduite. Alors, le vecteur  $X = (X_1, \dots, X_n)'$  est un vecteur aléatoire gaussien.*

*Démonstration.* Soit  $u_1, \dots, u_n$  des réels, et  $Y = \sum_{k=1}^n u_k X_k$  une combinaison linéaire (quelconque) des coordonnées du vecteur aléatoire  $X = (X_1, \dots, X_n)'$ . En utilisant le Proposition 9.9 et le fait que les  $X_k$

sont des v.a. indépendantes de loi normale centrée réduite, on a que la fonction caractéristique de  $Y$  est

$$\begin{aligned} \forall t \in \mathbb{R}, \phi_Y(t) &= \mathbb{E} \left( e^{it \sum_{k=1}^n u_k X_k} \right) = \prod_{k=1}^n \mathbb{E} \left( e^{it u_k X_k} \right) \\ &= \prod_{k=1}^n \phi_{X_k}(t u_k) = \prod_{k=1}^n \exp \left( -\frac{t^2 u_k^2}{2} \right) \\ &= \exp \left( -\left( \sum_{k=1}^n u_k^2 \right) \frac{t^2}{2} \right). \end{aligned}$$

Par la Proposition 9.9, la v.a.  $Y$  suit donc une loi normale d'espérance nulle et de variance  $\sum_{k=1}^n u_k^2$ , ce qui montre que  $X$  est un vecteur gaussien.  $\square$

**Propriété 9.11.** Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire gaussien de dimension  $n$ , d'espérance  $m = (m_1, \dots, m_n) \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma$ . Alors, la fonction caractéristique de  $X$  est

$$\forall t \in \mathbb{R}^n, \phi_X(t) = \exp \left( it'm - \frac{1}{2} t' \Sigma t \right).$$

*Démonstration.* Soit  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$  et posons  $Y = \sum_{k=1}^n t_k X_k$ . Comme  $Y$  est une combinaison linéaire des coordonnées du vecteur gaussien  $X$ , on a que  $Y$  suit une loi normale d'espérance  $\mathbb{E}(Y) = \sum_{k=1}^n t_k m_k = t'm$ , et de variance  $\text{var}(Y) = t' \Sigma t$ . Par la Proposition 9.9, on a donc que la fonction caractéristique de  $Y$  est

$$\forall s \in \mathbb{R}, \phi_Y(s) = \exp \left( is(t'm) - \frac{1}{2} s^2 t' \Sigma t \right),$$

et le résultat s'obtient finalement en remarquant que  $\phi_X(t) = \phi_Y(1)$ .  $\square$

La loi d'un vecteur gaussien est donc caractérisée par la donnée de son espérance et de sa matrice de covariance. En particulier, si on effectue une transformation linéaire d'un vecteur aléatoire gaussien, on obtient de nouveau un vecteur aléatoire gaussien, dont il suffit de calculer l'espérance et la matrice de covariance pour le caractériser.

### 9.2.2 Quelques propriétés des vecteurs gaussiens

En général, deux variables aléatoires réelles non corrélées ne sont pas nécessairement indépendantes. Pour des vecteurs gaussiens, indépendance des coordonnées et absence de corrélation sont deux notions équivalentes.

**Propriété 9.12.** Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire gaussien de matrice de covariance  $\Sigma_X$ . Les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes si et seulement si la matrice  $\Sigma_X$  est diagonale.

*Démonstration.* Supposons que la matrice  $\Sigma_X$  soit diagonale. Notons  $m_k$  et  $\sigma_k^2$  l'espérance et la variance de  $X_k$ . Par la Proposition 9.11, on a que la fonction caractéristique de  $X$  s'écrit

$$\forall t = (t_1, \dots, t_n) \in \mathbb{R}^n, \phi_X(t) = \exp\left(i \sum_{k=1}^n t_k m_k - \frac{1}{2} \sum_{k=1}^n t_k^2 \sigma_k^2\right) = \prod_{k=1}^n \exp\left(it_k m_k - \frac{1}{2} t_k^2 \sigma_k^2\right) = \prod_{k=1}^n \phi_{X_k}(t_k),$$

et donc par la Proposition 9.8, on en déduit que les v.a.  $X_k$  sont indépendantes. La réciproque est immédiate.  $\square$

**Propriété 9.13.** Soit  $X = (X_1, \dots, X_n)'$  un vecteur aléatoire gaussien d'espérance  $m \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma$ . Si  $\Sigma$  est définie positive (i.e. de déterminant  $\det(\Sigma) > 0$  non nul) alors  $X$  admet pour densité la fonction  $f_X$  définie par

$$\forall x \in \mathbb{R}^n, f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x-m)'\Sigma^{-1}(x-m)\right).$$

*Démonstration.* La matrice  $\Sigma$  (symétrique à valeurs réelles et définie positive) est diagonalisable sous la forme

$$\Sigma = Q\Lambda Q',$$

où  $Q$  est une matrice orthogonale (i.e.  $Q'Q = I_n$ ), et  $\Lambda$  est une matrice diagonale, dont les termes diagonaux  $\lambda_i, i = 1, \dots, n$  sont des réels strictement positifs (valeurs propres de  $\Sigma$ ). Posons

$$Y = Q'(X - m).$$

Etant donné que  $X$  est un vecteur gaussien,  $Y$  est donc également un vecteur gaussien d'espérance  $0 \in \mathbb{R}^n$  et de matrice de covariance  $\Sigma_Y = Q'\Sigma Q = \Lambda$ . La matrice de covariance de  $Y$  étant diagonale on a donc que les coordonnées  $Y_i, i = 1, \dots, n$  de  $Y$  sont indépendantes. De plus,  $Y_i$  est de loi normale de moyenne nulle et de variance  $\lambda_i$ . Par conséquent la densité du vecteur aléatoire  $Y$  est

$$\begin{aligned} \forall y = (y_1, \dots, y_n) \in \mathbb{R}^n, f_Y(y) &= \prod_{i=1}^n f_{Y_i}(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\lambda_i^{1/2}} \exp\left(-\frac{y_i^2}{2\lambda_i}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\prod_{i=1}^n \lambda_i}} \exp\left(-\frac{1}{2}y'\Lambda^{-1}y\right). \end{aligned}$$

Soit  $\varphi$  l'application de  $\mathbb{R}^n$  dans  $\mathbb{R}^n$  définie par  $\varphi(y) = Qy + m$ . On a donc que  $X = \varphi(Y)$ . Etant donnée que  $\varphi^{-1}(x) = Q'(x - m)$  est de jacobien  $\pm 1$ , la Proposition 9.6 implique que la densité de  $X$  est donnée par

$$\forall x \in \mathbb{R}^n, f_X(x) = f_Y(Q'(x - m)) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - m)'Q\Lambda^{-1}Q'(x - m)\right),$$

où l'on a utilisé le fait que  $\det(\Sigma) = \prod_{i=1}^n \lambda_i$  et que  $\Sigma^{-1} = Q\Lambda^{-1}Q'$ .  $\square$

### 9.2.3 Espérance conditionnelle et projection orthogonale

On va maintenant revenir sur la notion d'espérance conditionnelle et montrer qu'elle se caractérise très facilement dans le cas gaussien. Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé, et  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  une v.a. réelle. On rappelle que l'on note par  $L^2(\Omega, \mathcal{A}, P)$  l'espace des v.a. réelles  $X$  de carré intégrable i.e. telles que  $\mathbb{E}(X^2) < +\infty$ . De plus, l'espace  $L^2(\Omega, \mathcal{A}, P)$  est

un espace de Hilbert lorsqu'il est muni du produit scalaire  $\langle X, Y \rangle = \mathbb{E}(XY)$ , et de la norme quadratique associée  $\|X\|_2 = \sqrt{\mathbb{E}(X^2)}$ .

Une remarque très importante est alors la suivante. Si  $\begin{pmatrix} X \\ Y \end{pmatrix} \in \mathbb{R}^2$  est un vecteur gaussien d'espérance nulle, alors l'**orthogonalité de  $X$  et  $Y$**  dans l'espace  $L^2(\Omega, \mathcal{A}, P)$  i.e.

$$\langle X, Y \rangle = \mathbb{E}(XY) = 0$$

et l'**indépendance de  $X$  et  $Y$**  sont des propriétés équivalentes car

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

Utilisons donc de nouveau le cadre de la géométrie hilbertienne pour caractériser l'espérance conditionnelle. Soit  $Z = (Z_1, \dots, Z_n)' \in \mathbb{R}^n$  un vecteur aléatoire gaussien d'espérance nulle, i.e.  $\mathbb{E}(Z_i) = 0$  pour tout  $i = 1, \dots, n$ . Introduisons la définition suivante :

**Définition 9.7.** Soient  $\{i_1, \dots, i_p\} \subset \{1, \dots, n\}$  avec  $p \leq n$ . Le sous-espace vectoriel (fermé) engendré par les v.a.  $Z_{i_1}, \dots, Z_{i_p}$  est

$$\text{Vect}(Z_{i_1}, \dots, Z_{i_p}) = \overline{\left\{ \sum_{k=1}^p \alpha_k Z_{i_k}, \alpha_k \in \mathbb{R} \right\}}$$

On a alors les propriétés suivantes :

**Propriété 9.14.** Soit  $Y \in L^2(\Omega, \mathcal{A}, P)$  une v.a.r. telle que  $(Y, Z_{i_1}, \dots, Z_{i_p}) \in \mathbb{R}^{p+1}$  est un vecteur aléatoire gaussien d'espérance nulle. Alors, l'espérance conditionnelle de  $Y$  sachant le vecteur  $\tilde{Z} = (Z_{i_1}, \dots, Z_{i_p})$ , notée  $\mathbb{E}(Y|Z_{i_1}, \dots, Z_{i_p})$  ou  $\mathbb{E}(Y|\tilde{Z})$ , est la projection orthogonale de  $Y$  sur l'espace

$$\text{Vect}(Z_{i_1}, \dots, Z_{i_p}) = \overline{\left\{ \sum_{k=1}^p \alpha_k Z_{i_k}, \alpha_k \in \mathbb{R} \right\}},$$

et dans le cas  $p = 1$

$$\mathbb{E}(Y|Z_{i_1}) = \frac{\text{Cov}(Z_{i_1}, Y)}{\text{Var}(Z_{i_1})} Z_{i_1}.$$

On peut alors remarquer que  $\mathbb{E}(Y|Z_{i_1}, \dots, Z_{i_p})$  est donc une v.a. gaussienne ce qui montre la **stabilité de la loi gaussienne par conditionnement**. De plus on a également la décomposition suivante lors du conditionnement par rapport à deux vecteurs gaussiens indépendants :

**Propriété 9.15.** Soit  $Z \in \mathbb{R}^n$  et  $W \in \mathbb{R}^m$  deux vecteurs aléatoires gaussien indépendants et d'espérance nulle. Soit  $Y \in L^2(\Omega, \mathcal{A}, P)$  une v.a.r. telle que  $(Y, Z', W') \in \mathbb{R}^{n+m+1}$  est un vecteur aléatoire gaussien d'espérance nulle. Alors, l'espérance conditionnelle de  $Y$  sachant le vecteur  $\begin{pmatrix} Z \\ W \end{pmatrix}$ , notée  $\mathbb{E}(Y|Z, W)$ , est la projection orthogonale de  $Y$  sur l'espace

$$\text{Vect}(Z, W) = \overline{\left\{ \sum_{k=1}^n \alpha_k Z_k + \sum_{j=1}^m \beta_j W_j, \alpha_k, \beta_j \in \mathbb{R} \right\}} = \text{Vect}(Z) \oplus \text{Vect}(W)$$

et donc

$$\mathbb{E}(Y|Z, W) = \mathbb{E}(Y|Z) + \mathbb{E}(Y|W).$$

### 9.2.4 Lois conditionnelles et prédiction

Etudions maintenant les lois conditionnelles de certaines coordonnées d'un vecteur gaussien sachant la valeur des autres coordonnées. Soit  $Z$  un vecteur aléatoire gaussien de dimension  $n$  divisé en deux blocs de coordonnées  $X$  et  $Y$  de dimensions  $n_X \geq 1$  et  $n_Y \geq 1$  telles que  $n = n_X + n_Y$ , i.e.  $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$ . L'espérance de  $Z$  s'écrit alors sous la forme

$$\mathbb{E}(Z) = \begin{pmatrix} \mathbb{E}(X) \\ \mathbb{E}(Y) \end{pmatrix},$$

et la matrice de covariance  $\Sigma_Z$  de  $Z$  peut se décomposer de la façon suivante

$$\Sigma_Z = \begin{pmatrix} \Sigma_X & \Sigma_0 \\ \Sigma_0' & \Sigma_Y \end{pmatrix},$$

où

$$\Sigma_0 = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))').$$

La notion de densité conditionnelle vue précédemment pour les couples de variables aléatoires réelles peut être étendue au cas des vecteurs aléatoires. La densité conditionnelle de  $Y$  sachant  $X = x$  est alors la fonction

$$f_Y^{X=x}(y) = \frac{f_Z(x, y)}{f_X(x)},$$

et on a le résultat suivant.

**Propriété 9.16** (admis). *Supposons que  $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$  soit un vecteur gaussien, et que la matrice de covariance  $\Sigma_X$  de  $X$  est définie positive. Soit  $x \in \mathbb{R}^{n_X}$ . La loi conditionnelle de  $Y$  sachant  $X = x$  est alors une loi gaussienne d'espérance*

$$\mathbb{E}(Y|X = x) = \mathbb{E}(Y) + \Sigma_0' \Sigma_X^{-1} (x - \mathbb{E}(X))$$

et de matrice de covariance

$$\Sigma_Y^{X=x} = \Sigma_Y - \Sigma_0' \Sigma_X^{-1} \Sigma_0,$$

et donc

$$\forall y \in \mathbb{R}^{n_Y}, f_Y^{X=x}(y) = \frac{1}{(2\pi)^{n_Y/2}} \frac{1}{\sqrt{\det(\Sigma_Y^{X=x})}} \exp\left(-\frac{1}{2} (y - m_Y^{X=x})' [\Sigma_Y^{X=x}]^{-1} (y - m_Y^{X=x})\right).$$

avec  $m_Y^{X=x} = \mathbb{E}(Y) + \Sigma_0' \Sigma_X^{-1} (x - \mathbb{E}(X))$ .



# Chapitre 10

## PROCESSUS DE MARKOV

### 10.1 Introduction aux Processus Aléatoires

La théorie des processus aléatoires concerne l'étude mathématique de phénomènes physiques, biologiques ou économiques évoluant dans le temps, et dont l'évolution est de caractère aléatoire, c'est-à-dire non prévisible avec certitude. Pour définir un processus aléatoire, il faut :

#### 1- Un espace des temps $T$ ( $T \subset \mathbb{R}_+$ )

Les deux espaces des temps les plus utilisés sont :

- $T = \mathbb{N}$  : le processus est dit *discret*; on regarde ce qu'il se passe à chaque unité de temps, ou bien on fait une suite d'opérations et on regarde ce qu'il se passe à chaque opération (ex : lancer d'une pièce).
- $T = \mathbb{R}_+$  : le processus est dit *continu* : on garde les yeux fixés sur un système qui évolue dans le temps à partir d'un instant  $t_0$  que l'on prend pour origine des temps ( $t = 0$ ).

#### 2- Un espace des états $E$

L'ensemble  $E$  peut être :

- discret : c'est-à-dire fini ou dénombrable. Il sera, dans ce cas, souvent pratique d'identifier  $E$  avec une partie de  $\mathbb{N}$  ou de  $\mathbb{Z}$ .
- non discret : par exemple  $E = \mathbb{R}$  ou  $E \subset \mathbb{R}^2$  (partie du plan) ou  $E \subset \mathbb{R}^3$  (partie de l'espace)

#### 3- Une famille de variables aléatoires $(X_t)_{t \in T}$ .

Ces variables aléatoires sont toutes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans l'espace des états  $E$ . Ainsi, à chaque instant  $t \in T$ , on associe, non pas une valeur déterministe (comme dans le calcul d'une trajectoire mécanique) mais une valeur aléatoire décrite par une variable aléatoire  $X_t$  à valeurs dans  $E$ . La variable aléatoire  $X_t$  peut représenter les résultats d'essais successifs comme par exemple, le jet d'une pièce à pile ou face, ou des observations successives sur une caractéristique d'une population.

**Définition 10.1.** *Un processus aléatoire est une famille de variables aléatoires  $(X_t)_{t \in T}$ .*

Un processus aléatoire est une généralisation d'un vecteur aléatoire. Comme dans le cas du vecteur aléatoire, la connaissance de la loi de  $X_t$  pour tout  $t \in T$  est loin de caractériser le processus. Les éléments principaux qui différencient les processus aléatoires généraux sont l'espace des états  $E$ , l'espace des temps  $T$  et les relations de dépendance entre les  $X_t$ .

#### 4- Quelques relations de dépendance.

- Processus de comptage :

Un processus de comptage  $(N_t)_{t \in T}$ , où  $T \subset \mathbb{R}$ , est un processus croissant (si  $s \leq t$ , alors  $N_s \leq N_t$ ), à valeurs dans  $E = \mathbb{N}$ .

- Processus à accroissements indépendants :

Un processus croissant  $(X_t)_{t \geq 0}$  est dit à accroissements indépendants si pour tout  $n \in \mathbb{N}^*$ , et pour tous  $t_1, \dots, t_n$  tels que  $t_1 < t_2 < \dots < t_n$ , les accroissements  $X_{t_1} - X_0, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$  sont des variables aléatoires indépendantes.

- Processus homogène dans le temps :

Le processus  $(X_t)_{t \geq 0}$  est dit homogène, si pour tout  $t$  et pour tout  $s$ , la loi de  $X_{t+s} - X_s$  ne dépend pas de  $s$ .

- Processus de Markov :

Le processus  $(X_t)_{t \geq 0}$  est dit de Markov, si pour tout  $n \in \mathbb{N}$ , pour tous  $t_1, \dots, t_n, t_{n+1}$  tels que  $t_1 < t_2 < \dots < t_n < t_{n+1}$  :

$$P([X_{t_{n+1}} = e_{n+1}] | [X_{t_1} = e_1] \cap \dots \cap [X_{t_n} = e_n]) = P([X_{t_{n+1}} = e_{n+1}] | [X_{t_n} = e_n])$$

c'est-à-dire que seul le passé le plus proche est pris en compte pour déterminer les probabilités d'occupation de chaque état.

### Exemples de processus aléatoires

#### Processus de ramification

Ce processus est utilisé pour suivre l'évolution de certaines populations animales ou cellulaires, ou bien d'un nom chez les humains. Chaque individu génère, indépendamment des autres, un nombre aléatoire de descendants.  $X_n$  est le nombre d'individus total à la  $n^{\text{ième}}$  génération.

$$E = \mathbb{N} ; T = \mathbb{N}.$$

#### Files d'attente

Des clients se présentent à des guichets à des temps aléatoires, pour y recevoir des services de durée aléatoire (ex : banque, magasin, péage d'autoroute...).  $X_t$  est le nombre de clients en attente à l'instant  $t$ .

$$E = \mathbb{N} ; T = \mathbb{R}_+.$$

## 10.2 Généralités sur les Processus de Markov

Le processus de Markov fournit un outil simple de modélisation d'une classe particulière de systèmes à espace d'états discret. L'analyse des processus de Markov est un préliminaire nécessaire à l'étude des systèmes de files d'attente.

**Définition 10.2.** Le processus  $(X_t)_{t \geq 0}$  est dit de Markov, si

a) axiome de Markov : pour tous  $t_1 < t_2 < \dots < t_n < t_{n+1}$ , pour tous  $x_1, \dots, x_{n+1}$  :

$$P([X_{t_{n+1}} = x_{n+1}] | [X_{t_1} = x_1] \cap \dots \cap [X_{t_n} = x_n]) = P([X_{t_{n+1}} = x_{n+1}] | [X_{t_n} = x_n])$$

b) axiome d'homogénéité : pour tous  $s$  et  $t$ , pour tous  $x, y \in E$ ,  $P([X_{t+s} = y] | [X_s = x])$  ne dépend que de  $t$  (et non des instants  $s$  et  $t + s$ ).

L'axiome de Markov traduit que la probabilité de n'importe quel comportement futur, le présent étant connu, n'est pas modifié par toute connaissance supplémentaire du passé.

**Notations :** On pose  $p_{x,y}(t) = P([X_{t+s} = y] | [X_s = x]) = P([X_t = y] | [X_0 = x])$  et  $P(t) = (p_{x,y}(t))_{x,y \in E}$ . On pose aussi  $\vec{\pi}(t) = P_{X_t}$  (vecteur ligne de composantes  $\pi_x(t) = P([X_t = x])$ ).

**Propriété 10.1.**

a)  $P(t)$  est une matrice stochastique, i.e.  $p_{x,y}(t) \geq 0$  et  $\sum_y p_{x,y}(t) = 1$  pour tout  $x$ .

b) Pour tout  $s$  et pour tout  $t$ ,  $P(s+t) = P(s)P(t)$ .

c) Pour tout  $s$  et pour tout  $t$ ,  $\vec{\pi}(s+t) = \vec{\pi}(s)P(t)$ .

*Démonstration.* a)  $p_{x,y}(t) \in [0, 1]$  car c'est une probabilité. De plus, la ligne  $x$  correspond à la loi  $P_{X_t}^{[X_0=x]}$  et on a bien

$$\sum_y p_{x,y}(t) = \sum_y P^{[X_0=x]}([X_t = y]) = 1.$$

b) Pour calculer  $p_{x,y}(t+s) = P^{[X_0=x]}([X_{t+s} = y])$ , on va faire intervenir les différents états pouvant être occupés à l'instant  $t$  :

$$\begin{aligned} P^{[X_0=x]}([X_{t+s} = y]) &= \frac{P([X_0 = x] \cap [X_{t+s} = y])}{P([X_0 = x])} \\ &= \sum_{z \in E} \frac{P([X_0 = x] \cap [X_t = z] \cap [X_{t+s} = y])}{P([X_0 = x])} \\ &= \sum_{z \in E} \frac{P([X_{t+s} = y] | [X_0 = x] \cap [X_t = z]) P([X_0 = x] \cap [X_t = z])}{P([X_0 = x])} \\ &= \sum_{z \in E} P([X_{t+s} = y] | [X_t = z]) P([X_t = z] | [X_0 = x]) \end{aligned}$$

d'après l'axiome de Markov ( $X_0$  n'apporte rien de plus que  $X_t$  pour déterminer la loi de  $X_{t+s}$ ). Ainsi, on a

$$p_{x,y}(t+s) = \sum_{z \in E} p_{x,z}(t) p_{z,y}(s),$$

qui est le coefficient  $(x, y)$  de la matrice produit  $P(t)P(s)$ .

c)  $\pi_y(t+s) = P([X_{t+s} = y]) = \sum_{x \in E} P([X_{t+s} = y] | [X_t = x])P([X_t = x])$  c'est-à-dire

$$\pi_y(t+s) = \sum_{x \in E} \pi_x(t)p_{x,y}(s).$$

□

### 10.3 Chaînes de Markov à temps discret

Pour cette classe particulière de processus, on ne s'intéresse à l'état du système qu'en des instants particuliers  $t_n$  de leur évolution. Pour simplifier la présentation, on prendra ici  $T = \mathbb{N}$ .

#### 10.3.1 Matrice de transition et graphe d'une chaîne de Markov

On notera  $P$  la matrice  $P(1)$  définie précédemment (et  $p_{x,y} = p_{x,y}(1)$ ). Comme, pour tous entiers  $m$  et  $n$ , on a  $P(n+m) = P(n)P(m)$ , on a en particulier, pour tout  $n$ ,  $P(n+1) = P(n)P(1) = P(n)P$  et donc, comme de plus,  $P(0) = I = P^0$

$$P(n) = P^n \text{ et } \vec{\pi}(n) = \vec{\pi}(0)P^n \text{ pour tout } n \in \mathbb{N}.$$

Ainsi, la seule donnée de la matrice de transition  $P$  et de  $\vec{\pi}(0)$  suffit à déterminer la loi de  $X_n$  pour tout  $n \in \mathbb{N}$ . La matrice de transition  $P$  caractérise la chaîne et se prête bien aux calculs mais, pour mieux visualiser les transitions entre états, il est souvent utile de faire un graphe de la chaîne, équivalent à la donnée de  $P$  où :

- les états sont représentés par des points ;
- une probabilité de transition  $p_{x,y} > 0$  est représentée par un arc orienté de  $x$  à  $y$  au dessus duquel est notée la valeur de  $p_{x,y}$ .

#### 10.3.2 Classification des états

**Définition 10.3.** Soit  $x$  et  $y$  deux états. On dit que  $x$  mène à  $y$  s'il existe  $n \in \mathbb{N}$  tel que  $p_{x,y}(n) > 0$ . On dit que  $x$  et  $y$  communiquent si  $x$  mène à  $y$  et si  $y$  mène à  $x$  i.e. s'il existe  $n_1 \in \mathbb{N}$  tel que  $p_{x,y}(n_1) > 0$  et  $n_2 \in \mathbb{N}$  tel que  $p_{y,x}(n_2) > 0$ .

*Remarque :* Pour  $n \geq 1$ ,  $p_{x,y}(n) = \sum_{x_1, \dots, x_{n-1}} p_{x,x_1} p_{x_1,x_2} \cdots p_{x_{n-1},y}$  donc, si  $p_{x,y}(n) > 0$ , il existe  $x_1, \dots, x_{n-1}$  tels que  $p_{x,x_1} > 0$ ,  $p_{x_1,x_2} > 0, \dots, p_{x_{n-1},y} > 0$ , c'est-à-dire que l'on peut trouver sur le graphe un "chemin" de  $x$  à  $y$  si  $x \neq y$  (il peut même y en avoir plusieurs!). Attention toutefois,  $p_{x,x}(0) = 1$ , donc  $x$  mène toujours à  $x$  même s'il n'y a pas de chemin de  $x$  vers lui-même.

**Définition 10.4.** Soit  $x$  et  $y$  deux états. La relation de communication est une relation d'équivalence i.e.  $x \leftrightarrow y$  s'il existe  $(n_1, n_2) \in \mathbb{N}^2$  tels que  $p_{x,y}^{(n_1)} > 0$  et  $p_{y,x}^{(n_2)} > 0$ .

**Définition 10.5.** Une chaîne à une seule classe (pour la relation d'équivalence  $\leftrightarrow$ ) est dite irréductible.

*Remarque :* Dans une chaîne irréductible, on a donc que tous les états communiquent.

**Définition 10.6.** Un état  $x$  est dit non essentiel s'il existe un état  $y$  et  $n \in \mathbb{N}$  tels que  $p_{x,y}(n) > 0$  et  $p_{y,x}(m) = 0$  pour tout  $m \in \mathbb{N}$ . Dans le cas contraire  $x$  est dit essentiel.

Les états d'une même classe de communication sont de même nature : on parlera de classe non essentielle (classe que l'on peut quitter) et classe essentielle (ou absorbante) (classe que l'on ne peut pas quitter).

**Définition 10.7.** La période  $d(x)$  d'un état  $x$  est définie par :

$$d(x) = P.G.C.D.\{n \in \mathbb{N}^* ; p_{x,x}(n) > 0\}$$

(avec  $d(x) = 0$  si pour tout  $n \in \mathbb{N}^*$ ,  $p_{x,x}(n) = 0$ ). Si  $d(x) = 1$ ,  $x$  est dit apériodique.

**Propriété 10.2.** Tous les états d'une même classe de communication ont même période.

*Démonstration.* Si  $x$  et  $x'$  communiquent, il existe  $k$  et  $l$  tels que  $p_{x,x'}(k) > 0$  et  $p_{x',x}(l) > 0$ . Alors  $p_{x,x}(k+l) > 0$ .

Soit  $m$  tel que  $p_{x',x'}(m) > 0$ . Alors  $p_{x,x}(k+m+l) > 0$ . Donc  $d(x)$  divise  $k+l$  et  $d(x)$  divise  $k+m+l$ ; donc  $d(x)$  divise  $m$  et ceci, pour tout  $m$  tel que  $p_{x',x'}(m) > 0$ , donc  $d(x)$  divise  $d(x')$ . Comme  $x$  et  $x'$  jouent le même rôle,  $d(x) = d(x')$ .  $\square$

*Remarque :* Une classe dont un élément admet une boucle, (c'est-à-dire  $p_{x,x} > 0$ ), est obligatoirement apériodique (c'est-à-dire de période 1), mais ce n'est pas une condition nécessaire.

Introduisons, maintenant le temps d'atteinte d'un état  $x$

**Définition 10.8.**  $T_x = \begin{cases} \min\{n \geq 1 ; X_n = x\} & \text{si il existe } m \geq 1 \text{ tel que } X_m = x \\ +\infty & \text{si } X_m \neq x \text{ pour tout } m \geq 1 \end{cases}$ .

**Définition 10.9.** Un état  $x$  est dit récurrent si  $P^{[X_0=x]}([T_x < +\infty]) = 1$ , c'est-à-dire, si partant de  $x$ , on repasse presque sûrement en  $x$ . Dans le cas contraire,  $x$  est dit transitoire.

**Définition 10.10.** La probabilité de premier passage de  $x$  à  $y$  en  $n$  pas est donnée par  $f_{x,y}^{(n)} = P^{[X_0=x]}([T_y = n])$  pour  $n \geq 1$  (et par convention  $f_{x,y}^{(0)} = 0$ ).

On introduit également les quantités suivantes.

**Définition 10.11.** La probabilité d'atteindre  $y$  en partant de  $x$  est donnée par

$$f_{x,y} = \sum_{n \geq 1} f_{x,y}^{(n)}.$$

**Définition 10.12.** Le temps moyen de premier passage par  $y$  en partant de  $x$  est donné par

$$\mu_{x,y} = \begin{cases} \sum_{n \geq 1} n f_{x,y}^{(n)} & \text{si } f_{x,y} = 1 \\ +\infty & \text{si } f_{x,y} < 1 \end{cases}$$

**Propriété 10.3.**

- 1) L'état  $x$  est récurrent si et seulement si  $f_{x,x} = \sum_{n \geq 1} f_{x,x}^{(n)} = 1$ .
- 2) L'état  $x$  est transitoire si et seulement si  $f_{x,x} = \sum_{n \geq 1} f_{x,x}^{(n)} < 1$ .

*Démonstration.* Celle-ci découle immédiatement des définitions précédentes d'un état récurrent ou transitoire.  $\square$

**Définition 10.13.** Le temps moyen de retour en  $x$  est donné par

$$\mu_x = \mathbb{E}^{[X_0=x]}(T_x) = \mu_{x,x} = \sum_{n \geq 1} n f_{x,x}^{(n)}.$$

Si  $x$  est un état transitoire, alors  $\mu_x = +\infty$ . Toutefois,  $\mu_x$  peut être infini, même si  $x$  est récurrent. On est donc conduit à une classification plus fine des états récurrents.

**Définition 10.14.** Un état récurrent est dit récurrent positif si  $\mu_x < +\infty$ . Dans le cas contraire, il est dit récurrent nul.

### 10.3.3 Formules de récurrence

**Propriété 10.4.** Pour  $n \geq 1$  et  $m \geq 1$  tels que  $n - m \geq 0$ , et pour tout  $x, y$  dans  $E$  on a que

$$p_{x,y}(n) = \sum_{z \in E} p_{x,z}(m) p_{z,y}(n - m). \quad (10.1)$$

*Démonstration.* Le résultat découle immédiatement de la relation  $P^n = P^m P^{n-m}$ .  $\square$

La formule (10.1) s'appelle l'équation de Chapman-Kolmogorov. Elle s'interprète comme le fait que la probabilité d'aller de  $x$  à  $y$  en  $n$  pas, c'est aller de  $x$  à  $z$  en  $m$  pas, puis aller de  $z$  à  $y$  en  $n - m$  pas en passant par les états  $z$  qui le permettent.

**Propriété 10.5.** Pour  $n \geq 1$ , on a  $p_{x,y}(n) = \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n - k)$ .

*Démonstration.* Le processus passe de  $x$  à  $y$  en  $n$  étapes si et seulement s'il passe de  $x$  à  $y$  pour la première fois en  $k$  étapes ( $0 \leq k \leq n$ ) et s'il passe ensuite de  $y$  à  $y$  en les  $n - k$  étapes suivantes. Ces chemins, pour des  $k$  distincts, sont disjoints, et la probabilité d'un chemin pour  $k$  fixé est  $f_{x,y}^{(k)} p_{y,y}(n - k)$ .

Ce raisonnement intuitif peut être rendu rigoureux de la façon suivante.

Comme, pour  $n \geq 1$ ,

$$[X_n = y] = \bigcup_{k=1}^{n-1} ([T_y = k] \cap [X_n = y]) \cup [T_y = n],$$

on en déduit, les  $[T_y = k]$  étant disjoints,

$$\begin{aligned} P([X_n = y]/[X_0 = x]) &= \sum_{k=1}^{n-1} P([T_y = k] \cap [X_n = y]/[X_0 = x]) + f_{x,y}^{(n)} \\ &= \sum_{k=1}^{n-1} P([T_y = k]/[X_0 = x]) P([X_n = y]/[T_y = k] \cap [X_0 = x]) + f_{x,y}^{(n)} \end{aligned}$$

Or, pour  $1 \leq k \leq n-1$ ,  $[T_y = k] \cap [X_0 = x]$  est de la forme  $A \cap [X_k = y]$  où  $A$  ne dépend que de  $X_0, \dots, X_{k-1}$ . Par conséquent,

$$P([X_n = y]/[T_y = k] \cap [X_0 = x]) = P([X_n = y]/A \cap [X_k = y]) = P([X_n = y]/[X_k = y]) = p_{y,y}(n-k).$$

Comme  $P([X_n = y]/[X_0 = x]) = p_{x,y}(n)$  et  $P([T_y = k]/[X_0 = x]) = f_{x,y}^{(k)}$ , il en résulte

$$p_{x,y}(n) = \sum_{k=1}^{n-1} f_{x,y}^{(k)} p_{y,y}(n-k) + f_{x,y}^{(n)} = \sum_{k=1}^n f_{x,y}^{(k)} p_{y,y}(n-k) = \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k)$$

avec la convention  $f_{x,y}^{(0)} = 0$ . □

### 10.3.4 Caractérisation de la nature d'un état

#### Théorème 10.1.

- 1) Un état  $y$  est récurrent positif si et seulement si  $\sum_{n=0}^{+\infty} p_{y,y}(n) = +\infty$  et  $\lim_{n \rightarrow +\infty} p_{y,y}(n) > 0$ .
- 2) Un état  $y$  est récurrent nul si et seulement si  $\sum_{n=0}^{+\infty} p_{y,y}(n) = +\infty$  et  $\lim_{n \rightarrow +\infty} p_{y,y}(n) = 0$ .
- 3) Si un état  $y$  est récurrent, on a alors que  $\sum_{n=0}^{+\infty} p_{x,y}(n) = +\infty$  pour tout  $x$  qui mène à  $y$ .
- 4) Un état  $y$  est transitoire si et seulement si  $\sum_{n=0}^{+\infty} p_{y,y}(n) < +\infty$ .

On a alors  $\sum_{n=0}^{+\infty} p_{x,y}(n) < +\infty$  pour tout  $x \in E$ .

*Démonstration.* On considère les séries entières  $F_{x,y}(s) = \sum_{n=0}^{+\infty} s^n f_{x,y}^{(n)}$  et  $P_{x,y}(s) = \sum_{n=0}^{+\infty} s^n p_{x,y}(n)$ .

On établit que  $\boxed{P_{x,y}(s) = \delta_{x,y} + F_{x,y}(s)P_{y,y}(s)}$ .

En effet, par la propriété 10.5,  $P_{x,y}(s) = \sum_{n=0}^{+\infty} s^n p_{x,y}(n) = p_{x,y}(0) + \sum_{n=1}^{+\infty} s^n \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k)$  avec  $p_{x,y}(0) = \delta_{x,y}$ . De plus  $f_{x,y}^{(0)} p_{y,y}(0) = 0$ , donc on a bien

$$\sum_{n=1}^{+\infty} s^n \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k) = \sum_{n=0}^{+\infty} s^n \sum_{k=0}^n f_{x,y}^{(k)} p_{y,y}(n-k) = F_{x,y}(s)P_{y,y}(s)$$

d'après la définition du produit de deux séries.

En appliquant ceci à  $x = y$ , on obtient  $P_{y,y}(s) = 1 + F_{y,y}(s)P_{y,y}(s)$ , soit  $P_{y,y}(s) = \frac{1}{1 - F_{y,y}(s)}$ .

• Si  $y$  est récurrent,  $f_{y,y} = 1$ , soit, par le théorème d'Abel,  $\lim_{s \rightarrow 1^-} F_{y,y}(s) = 1$  et donc  $\lim_{s \rightarrow 1^-} P_{y,y}(s) = +\infty$ . La distinction entre un état  $y$  récurrent positif ou récurrent nul via la condition  $\lim_{n \rightarrow +\infty} p_{y,y}(n) > 0$  ou  $\lim_{n \rightarrow +\infty} p_{y,y}(n) = 0$  est plus délicate à établir. C'est une conséquence du théorème 10.3 donné un peu plus loin dans ce chapitre.

• Si  $x \neq y$  et si  $x$  mène à  $y$ , alors  $F_{x,y}(1) \neq 0$ , et, comme  $P_{x,y}(s) = F_{x,y}(s)P_{y,y}(s)$ , si  $\lim_{s \rightarrow 1^-} P_{y,y}(s) = +\infty$ , alors  $\lim_{s \rightarrow 1^-} P_{x,y}(s) = +\infty$ .

• Si  $y$  est transitoire, alors  $\lim_{s \rightarrow 1^-} P_{y,y}(s) < +\infty$  et, comme on a  $F_{x,y}(1) \leq 1$ , on a  $\lim_{s \rightarrow 1^-} P_{x,y}(s) < +\infty$ , c'est-à-dire  $\sum_{n=0}^{+\infty} p_{x,y}(n) < +\infty$ . □

*Conséquence :* Si  $y$  est transitoire, alors  $\sum_{n=0}^{+\infty} p_{x,y}(n) < +\infty$  pour tout  $x$ , et, en particulier,  $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$ , mais la réciproque est fautive en général.

**Propriété 10.6.** *Les états d'une même classe de communication sont, soit tous récurrents, soit tous transitoires.*

*Démonstration.* Si  $x$  et  $x'$  communiquent, il existe  $k$  et  $l$  tels que  $p_{x,x'}(k) > 0$  et  $p_{x',x}(l) > 0$ . On a donc  $p_{x,x}(k+n+l) \geq p_{x,x'}(k)p_{x',x'}(n)p_{x',x}(l)$  pour tout  $n \geq 1$ . Donc, par le théorème 10.1, si  $x'$  est récurrent, alors  $x$  l'est aussi et, comme  $x$  et  $x'$  jouent le même rôle, si  $x$  est récurrent, alors  $x'$  l'est aussi. □

### Cas particulier des chaînes finies.

**Propriété 10.7.** *Si  $E$  est fini alors il existe au moins un état récurrent.*

*Démonstration.* Si tous les états étaient transitoires, on aurait  $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$  pour tout  $y$  et alors, comme la somme ne comporte qu'un nombre fini de termes,  $\lim_{n \rightarrow +\infty} \sum_{y \in E} p_{x,y}(n) = \sum_{y \in E} \lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$ , ce qui contredit  $\sum_{y \in E} p_{x,y}(n) = 1$ . □

### 10.3.5 Distribution stationnaire

**Définition 10.15.** *Une famille  $\vec{\pi} = (\pi_x)_{x \in E}$  est dite distribution stationnaire d'une chaîne, de matrice de transition  $P$ , si c'est une probabilité qui vérifie  $\vec{\pi} = \vec{\pi}P$ .*

*Remarques :*

• Si  $\vec{\pi}(0) = \vec{\pi}$ , alors,  $\vec{\pi}(n) = \vec{\pi}$  pour tout  $n$  (preuve par récurrence sur  $n$  : étant donné que  $\vec{\pi}(n+1) = \vec{\pi}(n)P$ , si  $\vec{\pi}(n) = \vec{\pi}$ , alors  $\vec{\pi}(n+1) = \vec{\pi}P = \vec{\pi}$ .)

• Si  $(X_n)$  converge en loi,  $\lim_{n \rightarrow +\infty} \vec{\pi}(n)$  est une probabilité notée  $\vec{\pi}^{(\infty)}$  qui est distribution stationnaire de la chaîne. En effet, on a  $\vec{\pi}(n+1) = \vec{\pi}(n)P$  pour tout  $n \in \mathbb{N}$  et, en prenant la limite quand  $n \rightarrow +\infty$ , on obtient bien  $\vec{\pi}^{(\infty)} = \vec{\pi}^{(\infty)}P$ .

•  $\vec{\pi}$  peut ne pas exister : si  $E = \mathbb{N}$  et  $p_{k,k+1} = 1$  pour tout  $k \in \mathbb{N}$ ,  $\vec{\pi} = \vec{\pi}P$  conduit à  $\pi_0 = 0$  et  $\pi_k = \pi_{k-1}$  pour tout  $k \in \mathbb{N}^*$  : on aurait alors  $\pi_k = 0$  pour tout  $k \in \mathbb{N}$  et  $\vec{\pi}$  ne peut pas être une probabilité.

- Une chaîne peut admettre une infinité de distributions stationnaires.

*Interprétations :*

•  $\pi_y = \sum_{x \in E} \pi_x p_{x,y}$  s'écrit aussi  $\sum_{x \in E} \pi_y p_{y,x} = \sum_{x \in E} \pi_x p_{x,y}$  ( car  $\sum_{x \in E} p_{y,x} = 1$ ). On peut interpréter  $\pi_y p_{y,x}$  comme le nombre moyen de transitions de  $y$  vers  $x$  par unité de temps et  $\sum_{x \in E} \pi_y p_{y,x}$  s'interprète alors comme le flux moyen de sortie de l'état  $y$ . De même,  $\sum_{x \in E} \pi_x p_{x,y}$  est le flux moyen d'entrée dans l'état  $y$ . Ainsi, en régime permanent :

pour tout état  $y$ , il y a égalité entre le flux sortant de  $y$  et le flux entrant dans  $y$ .

- $\pi_y$  peut s'interpréter comme la proportion de temps passé dans l'état  $y$ .

**Propriété 10.8.** Si  $\vec{\pi}$  est distribution stationnaire d'une chaîne irréductible, alors  $\pi_x > 0$  pour tout  $x \in E$ .

*Démonstration.* On a  $\vec{\pi} = \vec{\pi}P^n$  pour tout  $n \in \mathbb{N}$ , donc  $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$  pour tout état  $y$ .

Supposons qu'il existe  $y$  tel que  $\pi_y = 0$ . Alors, comme  $\pi_x p_{x,y}(n) \geq 0$ , on a, pour tout  $x \in E$  et pour tout  $n \in \mathbb{N}$ ,  $\pi_x p_{x,y}(n) = 0$ .

Or, la chaîne étant irréductible, il existe  $n$  tel que  $p_{x,y}(n) > 0$ ; d'où  $\pi_x = 0$ , et ceci, pour tout  $x$ . On a alors  $\vec{\pi} = \vec{0}$ , ce qui est impossible, puisque  $\vec{\pi}$  est une probabilité. □

Si  $x$  est récurrent positif, on pose, pour tout  $y \in E$ ,  $\rho_x(y) = \mathbb{E}^{[X_0=x]} \left( \sum_{n \geq 0} \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right)$  : c'est le nombre moyen de visites dans l'état  $y$  entre deux visites dans l'état  $x$ .

**Propriété 10.9.** Si  $x$  est un état récurrent positif, alors  $\vec{\pi} = \left( \frac{\rho_x(y)}{\mu_x} \right)_{y \in E}$  est une distribution stationnaire.

*Démonstration.* On a

$$\begin{aligned} \rho_x(y) &= \mathbb{E}^{[X_0=x]} \left( \sum_{n \geq 0} \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right) = \sum_{n \geq 0} \mathbb{E}^{[X_0=x]} (\mathbb{I}_{[X_n=y] \cap [T_x > n]}) \\ &= \sum_{n \geq 0} P^{[X_0=x]} ([X_n = y] \cap [T_x > n]) \end{aligned}$$

Si  $y \neq x$ , on a  $P^{[X_0=x]} ([X_0 = y] \cap [T_x > 0]) = 0$  et, pour  $n \geq 1$ , on a aussi  $P^{[X_0=x]} ([X_n = y] \cap [T_x = n]) = 0$ , de sorte que l'on a deux expressions pour  $\rho_x(y)$  :

$$\rho_x(y) = \sum_{n \geq 1} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) ; \quad (*)$$

$$\rho_x(y) = \sum_{n \geq 1} P^{[X_0=x]} ([X_n = y] \cap [T_x > n]).$$

D'autre part,  $P^{[X_0=x]} ([X_0 = x] \cap [T_x > 0]) = 1$  et, pour  $n \geq 1$ , les probabilités  $P^{[X_0=x]} ([X_n = x] \cap [T_x > n])$  sont nulles, de sorte que  $\rho_x(x) = 1$ . Comme  $x$  est récurrent, la relation  $1 = f_{x,x} = \sum_{n \geq 1} P^{[X_0=x]} ([T_x = n])$

est vérifiée.

Or, elle peut encore s'écrire  $1 = \rho_x(x) = \sum_{n \geq 1} P^{[X_0=x]} ([X_n = x] \cap [T_x > n - 1])$ . Ainsi, (\*) est vérifiée pour tout  $y \in E$ . On la réécrit :

$$\begin{aligned} \rho_x(y) &= P^{[X_0=x]} ([X_1 = y] \cap [T_x > 0]) + \sum_{n \geq 2} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{n \geq 2} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]). \end{aligned}$$

Maintenant, pour  $n \geq 2$ ,

$$P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) = \sum_{z \neq x} P^{[X_0=x]} ([X_{n-1} = z] \cap [X_n = y] \cap [T_x > n - 1]) = K$$

avec  $K = \sum_{z \neq x} P^{[X_0=x]} ([X_n = y] / [X_{n-1} = z] \cap [T_x > n - 1]) P^{[X_0=x]} ([X_{n-1} = z] \cap [T_x > n - 1])$ .

Or,

$$[T_x > n - 1] \cap [X_{n-1} = z] = [X_1 \neq x] \cap \dots \cap [X_{n-1} \neq x] \cap [X_{n-1} = z] = [X_1 \neq x] \cap \dots \cap [X_{n-2} \neq x] \cap [X_{n-1} = z],$$

et, par l'axiome de Markov

$$P^{[X_0=x]} ([X_n = y] / [X_{n-1} = z] \cap [T_x > n - 1]) = P^{[X_0=x]} ([X_n = y] / [X_{n-1} = z]).$$

Ainsi,

$$\begin{aligned} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) &= \sum_{z \neq x} P^{[X_0=x]} ([X_n = y] / [X_{n-1} = z]) P^{[X_0=x]} P([X_{n-1} = z] \cap [T_x > n - 1]) \\ &= \sum_{z \neq x} p_{z,y} P^{[X_0=x]} P([X_{n-1} = z] \cap [T_x > n - 1]). \end{aligned}$$

Par suite,

$$\begin{aligned} \rho_x(y) &= p_{x,y} + \sum_{n \geq 2} P^{[X_0=x]} ([X_n = y] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{n \geq 2} \sum_{z \neq x} p_{z,y} P^{[X_0=x]} ([X_{n-1} = z] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{z \neq x} p_{z,y} \sum_{n \geq 2} P^{[X_0=x]} ([X_{n-1} = z] \cap [T_x > n - 1]) \\ &= p_{x,y} + \sum_{z \neq x} p_{z,y} \sum_{n \geq 1} P^{[X_0=x]} ([X_n = z] \cap [T_x > n]) \\ &= p_{x,y} + \sum_{z \neq x} p_{z,y} \rho_x(z) = \sum_z \rho_x(z) p_{z,y}. \end{aligned} \tag{10.2}$$

D'autre part, on a

$$\begin{aligned} \sum_y \rho_x(y) &= \sum_y \mathbb{E}^{[X_0=x]} \left( \sum_{n \geq 0} \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right) = \mathbb{E}^{[X_0=x]} \left( \sum_{n \geq 0} \sum_y \mathbb{I}_{[X_n=y] \cap [T_x > n]} \right) \\ &= \mathbb{E}^{[X_0=x]} \left( \sum_{n \geq 0} \mathbb{I}_{[T_x > n]} \right) = \sum_{n \geq 0} P^{[X_0=x]} ([T_x > n]) = \mathbb{E}^{[X_0=x]} (T_x) = \mu_x. \end{aligned}$$

Si  $x$  est récurrent positif, alors  $\mu_x < +\infty$  et on peut poser  $\pi_y = \frac{\rho_x(y)}{\mu_x}$ . Par l'égalité (10.2), il en résulte bien que  $\vec{\pi}$  est une distribution stationnaire. □

**Théorème 10.2.**

1) Une chaîne irréductible admet une distribution stationnaire  $\vec{\pi}$  si et seulement si tous ses états sont récurrents positifs. Dans ce cas,  $\vec{\pi}$  est unique et  $\pi_x = 1/\mu_x$  pour tout  $x \in E$ .

2) Une chaîne quelconque admet une distribution stationnaire si et seulement si elle possède une classe récurrente positive. Dans ce cas, si  $\vec{\pi}$  est une distribution stationnaire et si  $(R_i)_i$  sont les classes récurrentes positives, il existe des réels positifs  $\lambda_i$  de somme 1 tels que :

$$\pi_x = 0 \text{ si } x \text{ n'est pas récurrent positif et } \pi_x = \lambda_i/\mu_x \text{ si } x \in R_i.$$

*Démonstration.* 1) • On montre facilement que l'existence d'une distribution stationnaire implique la récurrence de la chaîne. En effet, on a alors  $\vec{\pi} = \vec{\pi}P^n$  pour tout  $n \in \mathbb{N}$ , d'où  $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$  pour tout  $y \in E$ . Si tous les états étaient transitoires, on aurait, pour tout  $x \in E$ ,  $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$ .

Comme  $p_{x,y}(n) \leq 1$  et que  $\sum_x \pi_x$  converge, on peut appliquer le théorème de convergence dominée de Lebesgue, c'est-à-dire intervertir limite et somme.

On a alors  $\pi_y = \lim_{n \rightarrow +\infty} \sum_{x \in E} \pi_x p_{x,y}(n) = \sum_{x \in E} \pi_x \lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$ , ce qui est contradictoire.

• On montre ensuite que, si  $\vec{\pi}$  est une distribution stationnaire, que l'on prend comme distribution initiale (i.e.  $\pi_y = P^{[X_0=y]}$ ), alors  $\pi_y \mu_y = 1$  pour tout état  $y$ .

En effet,  $\mu_y = \mathbb{E}^{[X_0=y]}(T_y) = \sum_{n \geq 0} P^{[X_0=y]}([T_y > n]) = \sum_{n \geq 1} P^{[X_0=y]}([T_y \geq n])$ ; d'où

$$\pi_y \mu_y = \sum_{n \geq 1} P^{[X_0=y]}([T_y \geq n]) P([X_0 = y]) = \sum_{n \geq 1} P([X_0 = y] \cap [T_y \geq n]).$$

On pose  $a_n = P([X_m \neq y \text{ pour } 0 \leq m \leq n])$ . On a  $P([X_0 = y] \cap [T_y \geq 1]) = P([X_0 = y]) = 1 - a_0$  et, pour  $n \geq 2$ ,

$$\begin{aligned} P([X_0 = y] \cap [T_y > n - 1]) &= P([X_0 = y] \cap [X_m \neq y \text{ pour } 1 \leq m \leq n - 1]) \\ &= P([X_m \neq y \text{ pour } 1 \leq m \leq n - 1]) - P([X_m \neq y \text{ pour } 0 \leq m \leq n - 1]) \\ &= P([X_m \neq y \text{ pour } 0 \leq m \leq n - 2]) - P([X_m \neq y \text{ pour } 0 \leq m \leq n - 1]) \\ &= a_{n-2} - a_{n-1} \end{aligned}$$

d'après l'axiome d'homogénéité. Donc  $\mu_y \pi_y = \sum_{n \geq 1} P([X_0 = y] \cap [T_y \geq n]) = 1 - \lim_{n \rightarrow +\infty} a_n$  et  $\lim_{n \rightarrow +\infty} a_n = P([X_m \neq y \text{ pour tout } m \geq 0]) = 0$  car  $y$  est récurrent. Donc  $\mu_y \pi_y = 1$  et, comme  $\pi_y > 0$  par la proposition 10.8, on a  $\mu_y < +\infty$  et  $\pi_y = \frac{1}{\mu_y}$ .

On a ainsi montré que tout état  $y$  est récurrent positif et que la distribution  $\vec{\pi}$  est unique.

• Réciproquement, si les états sont récurrents positifs, on a une distribution stationnaire donnée par la propriété 10.9, et, d'après ce que l'on vient de voir, cette distribution est donc unique et vérifie  $\pi_y = \frac{1}{\mu_y}$  pour tout état  $y$ .

2) Si  $\vec{\pi}$  est distribution stationnaire, alors  $\vec{\pi} = \vec{\pi}P^n$  pour tout  $n \in \mathbb{N}$ . D'où  $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$

pour tout  $y \in E$ .

• Si  $y$  est transitoire, on vérifie exactement comme dans 1) que  $\pi_y = 0$ .

• Si  $y$  est récurrent, alors  $\pi_y = \sum_{x \in E} \pi_x p_{x,y} = \sum_{x \in \text{Cl}(y)} \pi_x p_{x,y}$  car si  $x$  est transitoire,  $\pi_x = 0$  et si  $x$  est

récurrent, dans une autre classe,  $p_{x,y} = 0$ .

En posant  $C = \text{Cl}(y)$  et  $\vec{\pi}(C) = \sum_{x \in \text{Cl}(y)} \pi_x$ , on a :

→ soit  $\vec{\pi}(C) = 0$  et alors  $\pi_x = 0$  pour tout  $x \in C$  ;

→ soit  $\vec{\pi}(C) \neq 0$  et alors  $\vec{\pi}_C = \left( \frac{\pi_x}{\pi(C)} \right)_{x \in C}$  est distribution stationnaire de la chaîne irréductible

réduite à  $C$ . Par conséquent, d'après l'unicité vue au 1), on a  $\pi_x = \frac{\vec{\pi}(C)}{\mu_x}$  et  $x$  est récurrent positif. De

plus, on a  $\sum_C \vec{\pi}(C) = 1$ .

□

*Conséquences :*

1)  $\vec{\pi}$  n'est pas unique s'il existe plusieurs classes récurrentes positives.

2) Les états d'une même classe de communication sont, soit tous transitoires, soit tous récurrents positifs, soit tous récurrents nuls.

En effet, soit  $x$  récurrent positif. Alors la chaîne réduite à la classe de communication de  $x$  est irréductible et admet une distribution stationnaire puisqu'elle possède un état récurrent positif. Donc, tous les états de la classe de communication de  $x$  sont récurrents positifs.

3) Si la chaîne est irréductible, le nombre moyen de visites à  $y$  entre deux passages par  $x$  vérifie  $\rho_x(y) = \frac{\mu_x}{\mu_y}$ .

En effet,  $\vec{\pi}$  définie par  $\pi_y = \frac{\rho_x(y)}{\mu_x}$  est distribution stationnaire et d'après l'unicité,  $\pi_y = \frac{1}{\mu_y}$ .

### 10.3.6 Comportement asymptotique

La loi conditionnelle de  $X_n$  sachant  $[X_0 = x]$  est représentée par la famille  $(p_{x,y}(n))_{y \in E}$ . Pour avoir la convergence en loi, il est nécessaire que  $\lim_{n \rightarrow +\infty} p_{x,y}(n)$  existe pour tout  $y$  mais, en général, ceci n'est pas suffisant car il n'est pas évident que la famille limite soit une probabilité. C'est pourquoi, on introduit la notion plus restrictive suivante :

**Définition 10.16.** On dit que la chaîne  $(X_n)_{n \geq 0}$  de matrice de transition  $P$  admet une distribution limite si, pour tout  $(x, y) \in E^2$ ,  $p_{x,y}(n)$  admet une limite  $l_y$  indépendante de  $x$ , quand  $n$  tend vers  $+\infty$ , la famille  $\vec{l} = (l_y)_{y \in E}$  étant non identiquement nulle. La famille  $\vec{l} = (l_y)_{y \in E}$  est alors la distribution limite.

*Remarque :* Il n'est pas évident, *a priori*, que si  $\vec{l}$  existe, elle mérite le nom de distribution. Il est clair que  $l_y \geq 0$ , mais de  $\sum_{y \in E} p_{x,y}(n) = 1$ , on déduit *a priori* seulement que  $\sum_{y \in E} l_y \leq 1$ .

Pour pouvoir aller plus loin, il va falloir admettre le théorème suivant, dit théorème ergodique :

**Théorème 10.3** (ergodique). *Si  $y$  est apériodique ou non récurrent positif, alors*

$$\lim_{n \rightarrow +\infty} p_{x,y}(n) = \frac{f_{x,y}}{\mu_y} \text{ pour tout } x \in E, \text{ où } f_{x,y} = \sum_{n \geq 1} f_{x,y}^{(n)} = \sum_{n \geq 1} P^{[X_0=x]}([T_y = n]).$$

*Conséquences :*

- En particulier, si  $y$  est récurrent nul, alors  $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$  car  $f_{y,y} = 1$  et  $\mu_y = +\infty$ .
- Si  $x = y$  apériodique ou non récurrent positif, alors  $\lim_{n \rightarrow +\infty} p_{y,y}(n) = \frac{1}{\mu_y}$ . En effet, si  $y$  est récurrent, alors  $f_{y,y} = 1$  et si  $y$  est non récurrent, alors  $\mu_y = +\infty$  et comme  $f_{y,y} \leq 1$ , on a  $\frac{f_{y,y}}{\mu_y} = \frac{1}{\mu_y} = 0$ .

**Propriété 10.10.** *Si  $\vec{l}$  est une distribution limite, alors  $\vec{l}$  est l'unique distribution stationnaire de la chaîne.*

*Démonstration.* Soit  $y$  tel que  $\lim_{n \rightarrow +\infty} p_{x,y}(n) = l_y \neq 0$ . Alors, nécessairement,  $y$  est récurrent positif, d'après le théorème ergodique. Il existe donc au moins une classe récurrente positive. Mais il ne peut pas en exister plusieurs car sinon, en prenant  $x$  dans une autre classe récurrente positive, on aurait  $\lim_{n \rightarrow +\infty} p_{x,y}(n) = 0$ . D'après le théorème 10.2, il existe donc une unique distribution stationnaire  $\vec{\pi}$  avec  $\pi_x = \frac{1}{\mu_x}$  pour tout état récurrent positif, et  $\pi_x = 0$  sinon.

On a donc, pour tout état  $y$ ,  $\pi_y = \sum_{x \in E} \pi_x p_{x,y}(n)$  et, comme  $p_{x,y}(n) \leq 1$  et que  $\sum_{x \in E} \pi_x$  converge, on peut appliquer le théorème de convergence dominée de Lebesgue, c'est-à-dire intervertir limite et somme.

On a  $\pi_y = \lim_{n \rightarrow +\infty} \sum_{x \in E} \pi_x p_{x,y}(n) = \sum_{x \in E} \pi_x \lim_{n \rightarrow +\infty} p_{x,y}(n) = l_y \sum_{x \in E} \pi_x = l_y$ , donc  $\vec{l} = \vec{\pi}$ .

En particulier, on en déduit que  $\vec{l}$  est une loi de probabilité. □

**Théorème 10.4.** *La distribution limite existe si et seulement si la suite  $(X_n)_{n \geq 0}$  converge en loi, la loi limite n'étant pas fonction de la loi initiale. La limite en loi est alors la distribution limite, qui est aussi l'unique distribution stationnaire de la chaîne.*

*Démonstration.* Celle-ci résulte de la définition de la distribution limite et de la propriété précédente. □

## 10.4 Processus de Markov continus

### 10.4.1 Régime transitoire

Contrairement à ce qui se passe pour les chaînes de Markov à temps discret, on ne dispose pas ici d'un historique complet du processus : on observe celui-ci à certains instants dans le temps, choisis aussi nombreux que l'on veut et répartis comme on veut mais la notion d'"unité de temps" n'a plus de sens ici et la matrice  $P = P(1)$  ne permet pas de déterminer  $P(t)$  pour tout  $t$ .

L'idée est alors de considérer  $P(h)$  lorsque  $h \rightarrow 0$ .

Grâce à la relation  $P(t+h) = P(t)P(h) = P(h)P(t)$  et à  $P(0) = I$ , on a :

$$\frac{P(t+h) - P(t)}{h} = P(t) \frac{P(h) - I}{h} = \frac{P(h) - I}{h} P(t).$$

Sous réserve d'existence des limites, si on pose  $A = \lim_{h \rightarrow 0} \frac{P(h) - I}{h} = P'(0)$ , on a alors :

$$P'(t) = P(t)A = AP(t) \text{ et } P(0) = I.$$

Cette équation différentielle matricielle admet l'unique solution :

$$P(t) = e^{tA} = \sum_{n=0}^{+\infty} \frac{t^n}{n!} A^n.$$

Ainsi, on détermine  $A$  à partir de la famille  $(P(t))_{t \geq 0}$ , ( $A = P'(0)$ ), mais réciproquement, la seule connaissance de  $A$  permet de retrouver tous les  $P(t)$  ( $P(t) = e^{At}$ ).

La matrice  $A$  est appelée le générateur infinitesimal du processus.

On pose  $A = (a_{x,y})_{x,y \in E}$ . Ainsi,  $a_{x,y} = \lim_{h \rightarrow 0} \frac{p_{x,y}(h) - \delta_{x,y}}{h}$ .

- Si  $x \neq y$ ,  $a_{x,y} \geq 0$  et  $p_{x,y}(h) = a_{x,y}h + o(h)$
- Si  $x = y$ ,  $a_{x,x} \leq 0$  et  $p_{x,x}(h) = 1 + a_{x,x}h + o(h)$ .

On a ici,  $\boxed{\text{pour tout } x \in E, \sum_{y \in E} a_{x,y} = 0}$  (la somme sur chaque ligne de  $A$  vaut 0).

**Propriété 10.11.** Pour tout état  $x \in E$ , le temps passé dans  $x$  avant de le quitter suit la loi exponentielle  $\mathcal{E}(\lambda_x)$  avec  $\lambda_x = -a_{x,x} = \sum_{y \neq x} a_{x,y}$ .

*Démonstration.* Si  $T_x$  désigne ici le temps passé en  $x$  avant de le quitter, on a :

$$P([T_x > t+h]) = P([T_x > t+h]/[T_x > t])P([T_x > t])$$

Mais, d'après l'homogénéité dans le temps, on a  $P([T_x > t+h]/[T_x > t]) = P([T_x > h])$ . Ainsi, si on pose  $G_x(t) = P([T_x > t])$ , on obtient :

$$G_x(t+h) = G_x(t)G_x(h) \text{ avec } G_x(h) = p_{x,x}(h) + o(h) = 1 + a_{x,x}h + o(h),$$

donc  $\frac{G_x(t+h) - G_x(t)}{h} = G_x(t) \left( a_{x,x} + \frac{o(h)}{h} \right)$  et, en faisant  $h \rightarrow 0$ , il vient :

$$G'_x(t) = a_{x,x}G_x(t) \text{ avec } G_x(0) = 1,$$

soit  $G_x(t) = e^{a_{x,x}t}$  et  $\boxed{P([T_x \leq t]) = 1 - e^{a_{x,x}t}}$ .

□

L'évolution d'un processus de Markov à temps continu peut se voir comme une répétition de deux phases :

→ on reste un certain temps (de loi exponentielle) dans un état ;

→ lorsqu'on quitte cet état, on choisit l'état vers lequel on sort, cette destination ne dépendant ni du temps passé dans l'état, ni du chemin par lequel on est arrivé à l'état. On notera  $p_{x,y}$  la probabilité de se rendre dans l'état  $y$  en quittant l'état  $x$ .

**Propriété 10.12.** Pour tout  $(x, y) \in E^2$  tel que  $x \neq y$ ,  $p_{x,y} = \frac{a_{x,y}}{\sum_{z \neq x} a_{x,z}}$  (et  $p_{x,x} = 0$ ).

*Démonstration.* On a, si  $\lambda_x = -a_{x,x} = \sum_{z \neq x} a_{x,z}$ ,

$$P([X_{t+h} \neq x] | [X_t = x]) = \lambda_x h + o(h)$$

et, pour  $y \neq x$ ,

$$P([X_{t+h} = y] | [X_t = x]) = a_{x,y} h + o(h) = p_{x,y} \lambda_x h + o(h)$$

d'après la définition de  $p_{x,y}$ . C'est donc bien que  $p_{x,y} = \frac{a_{x,y}}{\lambda_x} = \frac{a_{x,y}}{\sum_{z \neq x} a_{x,z}}$ .

□

**Définition 10.17.** La chaîne de Markov discrète de matrice de transition  $P = (p_{x,y})_{x,y \in E}$  est appelée *chaîne de Markov induite du processus*.

### 10.4.2 Régime permanent

On rappelle que, si  $\vec{\pi}(t)$  est la loi de  $X_t$ , i.e.  $\vec{\pi}(t) = (\pi_x(t))_{x \in E}$  où  $\pi_x(t) = P([X_t = x])$ , on a :

$$\vec{\pi}(t) = \vec{\pi}(0)P(t).$$

**Définition 10.18.** On dit que le processus  $(X_t)_{t \geq 0}$  est *stationnaire* si la loi de  $X_t$  est indépendante de  $t$ .

On a alors  $\vec{\pi}(t) = \vec{\pi}(0)$  pour tout  $t \geq 0$  et en dérivant l'équation précédente, on obtient :

$$0 = \vec{\pi}(0)P'(t) = \vec{\pi}(0)P(t)A = \vec{\pi}(t)A = \vec{\pi}(0)A.$$

**Définition 10.19.** On appelle *distribution stationnaire* toute probabilité  $\vec{\pi}$  qui vérifie

$$\vec{\pi}A = \vec{0}.$$

Si  $(X_t)$  converge en loi et si  $\vec{\pi}^{(\infty)} = \lim_{t \rightarrow +\infty} \vec{\pi}(t)$ , alors  $\vec{\pi}^{(\infty)}$  est distribution stationnaire du processus.

On a  $(\vec{\pi}A)_y = \sum_{x \in E} \pi_x a_{x,y} = \sum_{x \neq y} \pi_x a_{x,y} + \pi_y a_{y,y}$  avec  $a_{y,y} = -\sum_{x \neq y} a_{y,x}$  ainsi :

$$\vec{\pi}A = \vec{0} \text{ équivaut à } \sum_{x \neq y} \pi_x a_{x,y} = \sum_{x \neq y} \pi_y a_{y,x}$$

Cette relation traduit l'équilibre, en régime stationnaire du flux rentrant en  $y$  ( $\sum_{x \neq y} \pi_x a_{x,y}$ ) et du flux sortant de  $y$  ( $\sum_{x \neq y} \pi_y a_{y,x}$ ).

# Chapitre 11

## PROCESSUS DE POISSON

### 11.1 Introduction

De nombreux phénomènes aléatoires se manifestent par des “arrivées” survenant une par une à des instants aléatoires successifs.

*Exemples :*

- arrivées d’appels à un central téléphonique
- passage de véhicules à un péage d’autoroute
- arrivées de clients à un guichet
- occurrence d’accidents dans une ville
- pannes de machines dans une usine

De tels phénomènes peuvent se définir par la famille  $(A_n)_{n \in \mathbb{N}^*}$  des temps d’arrivées qui sont des variables aléatoires. Mais on peut aussi le faire à partir du processus de comptage  $(N_t)_{t \in \mathbb{R}_+}$ , ou par la famille  $(T_n)_{n \in \mathbb{N}^*}$  des intervalles de temps entre deux arrivées.

$N_t$  est le nombre d’événements apparus jusqu’à l’instant  $t$ .

$N_{t+u} - N_u$  est le nombre d’événements apparus entre  $u$  et  $u + t$ .

L’espace des états du processus  $(N_t)_{t \in \mathbb{R}_+}$  est  $E = \mathbb{N}$  et l’espace des temps est  $T = \mathbb{R}_+$ . Le processus qui modélise convenablement les exemples cités est le processus de Poisson. On conviendra que  $N_0 = 0$ .

On note :

- $A_n$  l’instant de réalisation du  $n^{\text{ième}}$  événement ;
- $T_n$  la durée séparant le  $(n-1)^{\text{ième}}$  événement du  $n^{\text{ième}}$  événement pour  $n \geq 2$  et  $T_1 = A_1$ .

On a :

→  $A_n = T_1 + T_2 + \dots + T_n$  pour tout  $n \in \mathbb{N}^*$  :  
 →  $T_1 = A_1$  et  $T_n = A_n - A_{n-1}$  pour tout  $n \geq 2$ .

Ainsi, la connaissance de la famille  $(A_n)_{n \in \mathbb{N}^*}$  équivaut à celle de la famille  $(T_n)_{n \in \mathbb{N}^*}$ .

D'autre part,  $A_n \leq t$  signifie que le  $n^{\text{ième}}$  événement a eu lieu à l'instant  $t$  ou avant, c'est-à-dire qu'à l'instant  $t$ , au moins  $n$  événements ont eu lieu, c'est-à-dire que  $N_t \geq n$ . Ainsi

$$F_{A_n}(t) = P([A_n \leq t]) = P([N_t \geq n]) = 1 - \sum_{k=0}^{n-1} P([N_t = k])$$

et

$$P([N_t = n]) = P([N_t \geq n]) - P([N_t \geq n + 1]) = P([A_n \leq t]) - P([A_{n+1} \leq t]).$$

Par conséquent, la connaissance de  $(N_t)_{t \in \mathbb{R}_+}$  équivaut à celle de la famille  $(A_n)_{n \in \mathbb{N}^*}$ .

## 11.2 Définition et description du processus

Soit  $(N_t)_{t \in \mathbb{R}_+}$  un processus aléatoire à valeurs dans  $\mathbb{N}$  tel que  $N_0 = 0$ .

**Définition 11.1.** Le processus  $(N_t)_{t \in \mathbb{R}_+}$  est appelé processus de comptage si c'est un processus croissant, c'est-à-dire si pour tout  $s \leq t$ ,  $N_s \leq N_t$ . La variable aléatoire  $N_t - N_s$  est alors appelée accroissement du processus sur ]s, t].

**Définition 11.2.** Un processus de comptage  $(N_t)_{t \in \mathbb{R}_+}$  est appelé processus à accroissements indépendants si pour tout  $n \in \mathbb{N}^*$  et pour tous  $t_1, \dots, t_n$  tels que  $t_1 < t_2 < \dots < t_n$ , les accroissements  $N_{t_1} - N_0, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$  sont des variables aléatoires indépendantes.

Les processus vérifiant cette hypothèse sont assez nombreux : il semble en effet assez naturel que les arrivées se produisant dans des intervalles disjoints ne soient pas liées entre elles.

**Définition 11.3.** Le processus est dit stationnaire (ou homogène dans le temps), si pour tout  $s$  et pour tout  $t$ , l'accroissement  $N_{t+s} - N_s$  a même loi que  $N_t$ .

*Remarque :* Cette propriété est semblable à l'axiome d'homogénéité pour les chaînes de Markov : seule la durée écoulée entre deux instants (et non pas les deux instants eux-mêmes) compte pour déterminer la loi de l'accroissement.

**Définition 11.4.** Un processus à accroissements indépendants stationnaire  $(N_t)_{t \in \mathbb{R}_+}$  est dit à événements rares si  $\lim_{h \rightarrow 0_+} P([N_h > 0]) = 0$  et si  $\lim_{h \rightarrow 0_+} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$ .

*Remarque :* La deuxième propriété traduit l'improbabilité d'arrivées simultanées. Si on pose  $f(t) = P([N_t = 0])$  ( et donc  $f(0) = 1$ ), la première hypothèse traduit la continuité de  $f$  en 0, car  $P([N_h > 0]) = 1 - f(h)$  et donc  $\lim_{h \rightarrow 0_+} f(h) = f(0)$ .

**Définition 11.5.** Un processus de comptage  $(N_t)_{t \in \mathbb{R}_+}$  tel que  $N_0 = 0$  est un processus de Poisson si

C1 :  $(N_t)_{t \in \mathbb{R}_+}$  est stationnaire

C2 :  $(N_t)_{t \in \mathbb{R}_+}$  est un processus à accroissements indépendants ;

C3 :  $(N_t)_{t \in \mathbb{R}_+}$  est un processus à événements rares .

Le nom donné au processus de Poisson s'explique par ce qui suit :

**Propriété 11.1.** Un processus de comptage  $(N_t)_{t \in \mathbb{R}_+}$  tel que  $N_0 = 0$  est un processus de Poisson si et seulement si :

C1 :  $(N_t)_{t \in \mathbb{R}_+}$  est stationnaire ;

C2 :  $(N_t)_{t \in \mathbb{R}_+}$  est un processus à accroissements indépendants ;

C3' : il existe  $\lambda > 0$  tel que, pour tout  $t \geq 0$ , la variable aléatoire  $N_t$  suit la loi de Poisson de paramètre  $\lambda t$ .

*Démonstration.* Un processus qui vérifie C3' vérifie également C3.

En effet, si on a C3', alors

$$P([N_h = 0]) = e^{-\lambda h} = 1 - \lambda h + o(h)$$

d'où  $P([N_h > 0]) = \lambda h + o(h)$  et  $\lim_{h \rightarrow 0} P([N_h > 0]) = 0$ . De plus,  $P([N_h = 1]) = e^{-\lambda h} \lambda h = \lambda h + o(h) \sim \lambda h$  et

$$P([N_h > 1]) = 1 - P([N_h = 0]) - P([N_h = 1]) = o(h)$$

et donc  $\lim_{h \rightarrow 0+} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$ .

On va maintenant montrer que, pour un processus de Poisson, l'axiome C3' est vérifié.

$$P([N_{t+s} = k]) = \sum_{i=0}^k P([N_{t+s} = k] \cap [N_t = i]) \quad (11.1)$$

$$= \sum_{i=0}^k P([N_t = i] \cap [N_{t+s} - N_t = k - i]) \quad (11.2)$$

$$= \sum_{i=0}^k P([N_t = i]) P([N_{t+s} - N_t = k - i]) \quad (11.3)$$

$$= \sum_{i=0}^k P([N_t = i]) P([N_s = k - i]) \quad (11.4)$$

(On passe de (11.2) à (11.3) en utilisant C2 et de (11.3) à (11.4) en utilisant C1.)

On pose alors  $P([N_t = k]) = \pi_k(t)$  et on va établir que  $\pi_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$  par récurrence sur  $k$ .

Étude pour  $k = 0$  :  $\pi_0(t) = f(t)$  avec  $f$  qui vérifie

$$f(t + s) = f(t)f(s) \quad (*)$$

$f$  est continue sur  $\mathbb{R}_+$  car  $\lim_{h \rightarrow 0} (f(t+h) - f(t)) = \lim_{h \rightarrow 0} f(t)(f(h) - 1) = 0$  par C3.

Or, les seules solutions continues de (\*), à valeurs dans  $\mathbb{R}_+$ , sont les fonctions de la forme  $f(t) = e^{at}$ .

(En effet,  $\psi = \ln f$  vérifie  $\psi(s + t) = \psi(t) + \psi(s)$ . En particulier,  $\psi(nt) = n\psi(t)$ , puis, avec  $t = \frac{1}{n}$ ,  $\psi\left(\frac{1}{n}\right) = \frac{1}{n}\psi(1)$ , puis  $\psi\left(\frac{m}{n}\right) = \frac{m}{n}\psi(1)$ . On a alors  $\psi(r) = r\psi(1)$  pour tout  $r \in \mathbb{Q}$  et on conclut que

$\psi(x) = x\psi(1)$  par densité de  $\mathbb{Q}$  dans  $\mathbb{R}$  ( $x = \lim r_n$ ) et grâce à la continuité de  $f$  ( $f(x) = \lim f(r_n)$ ). Ainsi  $f(x) = e^{\psi(x)} = e^{x\psi(1)}$ .

Ici  $f$  est à valeurs dans  $[0, 1]$ , non identiquement nulle, donc  $a \leq 0$ . Le cas  $a = 0$  n'a pas d'intérêt car alors on aurait  $P([N_t = 0]) = 1$  pour tout  $t$ . Donc,  $a < 0$  et il existe  $\lambda = -a$  tel que

$$f(t) = \pi_0(t) = P([N_t = 0]) = e^{-\lambda t}.$$

Récurrence sur  $k$  : On suppose maintenant que  $\pi_i(t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!}$  pour tout  $t \geq 0$  et pour tout  $i \leq k$  et on va montrer que  $\pi_{k+1}(t) = e^{-\lambda t} \frac{(\lambda t)^{k+1}}{(k+1)!}$ .

Pour cela on va établir une équation différentielle du premier ordre vérifiée par  $\pi_{k+1}(t)$  : on regarde combien d'arrivées se sont produites jusqu'à l'instant  $t+h$  en faisant intervenir le nombre d'arrivées qui se sont produites jusqu'à l'instant  $t$ .

$$[N_{t+h} = k+1] = \bigcup_{i=0}^{k+1} [N_t = i] \cap [N_{t+h} - N_t = k+1-i]$$

et donc, toujours en utilisant C2 puis C1, il vient

$$P([N_{t+h} = k+1]) = \sum_{i=0}^{k+1} P([N_t = i])P([N_{t+h} - N_t = k+1-i]) = \sum_{i=0}^{k+1} P([N_t = i])P([N_h = k+1-i])$$

soit, avec la notation introduite ici

$$\pi_{k+1}(t+h) = \sum_{i=0}^{k+1} \pi_i(t)\pi_{k+1-i}(h)$$

Or, pour  $i \leq k-1$ , on a  $k+1-i \geq 2$ ; d'où  $P([N_h = k+1-i]) = \pi_{k+1-i}(h) = o(\pi_1(h))$  car  $\lim_{h \rightarrow 0} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$  et

$$\pi_{k+1}(t+h) = \sum_{i=0}^{k-1} \pi_i(t)\pi_{k+1-i}(h) + \pi_1(h)\pi_k(t) + \pi_0(t)\pi_{k+1}(h) = \pi_1(h)\pi_k(t) + \pi_0(t)\pi_{k+1}(h) + o(\pi_1(h)).$$

Avec  $\pi_0(h) = 1 - \pi_1(h) + o(\pi_1(h))$ , on obtient alors :

$$\pi_{k+1}(t+h) - \pi_{k+1}(t) = \pi_1(h) [\pi_k(t) - \pi_{k+1}(t)] + o(h).$$

Donc  $\frac{\pi_{k+1}(t+h) - \pi_{k+1}(t)}{h} = \frac{\pi_1(h)}{h} [\pi_k(t) - \pi_{k+1}(t)] + \frac{o(h)}{h}$  et

$$\pi'_{k+1}(t) = \lim_{h \rightarrow 0} \frac{\pi_1(h)}{h} [\pi_k(t) - \pi_{k+1}(t)].$$

Or  $\pi_1(h) = 1 - \pi_0(h) + o(\pi_1(h))$  donc  $\lim_{h \rightarrow 0} \frac{1 - \pi_0(h)}{\pi_1(h)} = 1 = \lim_{h \rightarrow 0} \frac{1 - \pi_0(h)}{h} \times \frac{h}{\pi_1(h)}$  et comme

$$\lim_{h \rightarrow 0} \frac{1 - \pi_0(h)}{h} = \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h}}{h} = \lambda,$$

c'est donc que  $\lim_{h \rightarrow 0} \frac{\pi_1(h)}{h} = \lambda$ , d'où  $\pi'_{k+1}(t) = \lambda [\pi_k(t) - \pi_{k+1}(t)]$ , soit :

$$\boxed{\pi'_{k+1}(t) + \lambda \pi_{k+1}(t) = \lambda \pi_k(t).}$$

Pour résoudre cette équation différentielle linéaire du premier ordre, on applique la méthode dite “de variation de la constante”.

On a donc  $\pi_{k+1}(t) = C(t)e^{-\lambda t}$  avec

$$C'(t)e^{-\lambda t} = \lambda\pi_k(t) = e^{-\lambda t} \frac{\lambda^{k+1}t^k}{k!}.$$

D'où  $C'(t) = \frac{\lambda^{k+1}t^k}{k!}$  et

$$C(t) - C(0) = \int_0^t C'(u)du = \frac{\lambda^{k+1}}{k!} \int_0^t u^k du = \frac{\lambda^{k+1}}{k!} \left[ \frac{u^{k+1}}{k+1} \right]_0^t = \frac{(\lambda t)^{k+1}}{(k+1)!}$$

avec  $\pi_{k+1}(0) = C(0) = P([N_0 = k+1]) = 0$  car  $N_0 = 0$ . On a donc bien  $\pi_{k+1}(t) = e^{-\lambda t} \frac{(\lambda t)^{k+1}}{(k+1)!}$ . □

*Remarque :*

Pour tous  $s$  et  $t$  et pour tous  $i \leq k$  :

$$\begin{aligned} P([N_{t+s} = k] \cap [N_s = i]) &= P([N_s = i] \cap [N_{t+s} - N_s = k - i]) \\ &= P([N_s = i])P([N_{t+s} - N_s = k - i]) \\ &= P([N_s = i])P([N_t = k - i]). \end{aligned}$$

Ainsi,

$$\begin{aligned} P([N_{t+s} = k] | [N_s = i]) &= \frac{P([N_s = i] \cap [N_{t+s} = k])}{P([N_s = i])} \\ &= P([N_t = k - i]) = e^{-\lambda t} \frac{(\lambda t)^{k-i}}{(k-i)!}. \end{aligned}$$

Ceci détermine les probabilités de transition de  $i$  à  $k$  entre les instants  $s$  et  $t + s$ .

### 11.3 Caractérisation d'un processus de Poisson par ses temps d'arrivée

Soit  $A_n$  l'instant de la  $n^{\text{ième}}$  arrivée :  $A_n = \inf\{t \geq 0 ; N_t = n\}$  et  $T_n$  le  $n^{\text{ième}}$  temps d'attente pour  $n \in \mathbb{N}^*$  :  $T_n = A_n - A_{n-1}$  (en convenant  $A_0 = 0$ ).

On a  $A_n = \sum_{i=1}^n T_i$  et  $N_t = \max\{n \geq 0 ; A_n \leq t\}$ .

**Théorème 11.1.**  $(N_t)_{t \in \mathbb{R}_+}$  est un processus de Poisson de paramètre  $\lambda$  si et seulement si les variables aléatoires  $T_n$  sont indépendantes de même loi exponentielle  $\mathcal{E}(\lambda)$  (de densité

$$f_{T_n}(t) = \lambda e^{-\lambda t} \mathbb{I}_{]0, +\infty[}(t).$$

*Démonstration.*

$$P([T_1 > t]) = P([N_t = 0]) = e^{-\lambda t} = 1 - F_1(t)$$

où  $F_1$  est la fonction de répartition de  $T_1$ . On a donc bien  $T_1$  qui suit la loi exponentielle  $\mathcal{E}(\lambda)$ .

$$\begin{aligned}
P^{[T_1=t_1]}([T_2 > t]) &= P([N_{t_1+t} = 1]/[N_{t_1} = 1] \cap [N_s = 0 \text{ pour tout } s < t_1]) \\
&= P([N_{t_1+t} - N_{t_1} = 0]/[N_{t_1} = 1] \cap [N_s = 0 \text{ pour tout } s < t_1]) \\
&= P([N_{t_1+t} - N_{t_1} = 0])
\end{aligned}$$

d'après l'indépendance des accroissements.

Or  $P([N_{t_1+t} - N_{t_1} = 0]) = P([N_{t_1+t-t_1} = 0]) = P([N_t = 0])$  d'après la stationnarité; et c'est aussi  $e^{-\lambda t}$  car  $N_t$  suit la loi de Poisson  $\mathcal{P}(\lambda t)$ .

Donc  $T_2$  est bien indépendante de  $T_1$  et de même loi exponentielle  $\mathcal{E}(\lambda)$ .

De façon plus générale,

$$\begin{aligned}
P([T_k > t]/[T_1 = t_1] \cap \dots \cap [T_{k-1} = t_{k-1}]) &= P([N_{t_{k-1}+t} - N_{t_{k-1}} = 0]) \\
&= P([N_t = 0]) = e^{-\lambda t}.
\end{aligned}$$

Donc  $T_k$  est indépendante de  $T_1, \dots, T_{k-1}$  et de même loi exponentielle  $\mathcal{E}(\lambda)$ .

La réciproque sera admise. □

*Conséquence :* Les variables aléatoires  $A_n$  suivent la loi Gamma  $\gamma(\lambda, n)$  (ou loi d'Erlang), de densité définie par

$$f_{A_n}(t) = \frac{\lambda^n}{(n-1)!} e^{-\lambda t} t^{n-1} \mathbb{I}_{]0, +\infty[}(t).$$

**Propriété 11.2.** Si  $(N_t)_{t \in \mathbb{R}_+}$  est un processus de Poisson de paramètre  $\lambda$ , le temps aléatoire  $U$  qui sépare un instant  $\theta$  du prochain événement et le temps aléatoire  $V$  qui sépare  $\theta$  du dernier événement suivent la loi exponentielle  $\mathcal{E}(\lambda)$ .

*Démonstration.*

$$P([U > x]) = P([N_{\theta+x} - N_\theta = 0]) = P([N_x = 0]) = e^{-\lambda x}$$

car  $[U > x]$  signifie que pendant la durée  $x$  qui suit  $\theta$ , il n'y a aucune arrivée. De même,

$$P([V > x]) = P([N_\theta - N_{\theta-x} = 0]) = P([N_x = 0]) = e^{-\lambda x}$$

car  $[V > x]$  signifie que pendant la durée  $x$  qui précède  $\theta$ , il n'y a eu aucune arrivée. □

*Remarque :* On a alors  $\mathbb{E}(U + V) = \mathbb{E}(U) + \mathbb{E}(V) = \frac{2}{\lambda}$  alors que  $\mathbb{E}(T_n) = \frac{1}{\lambda}$  pour tout  $n \in \mathbb{N}^*$ . C'est donc que  $U + V = T_{N_\theta}$  n'a pas même loi que les  $T_n$  alors que sur  $[N_\theta = n]$ , on a  $T_{N_\theta} = T_n$ .

On peut terminer ce paragraphe en remarquant que  $\mathbb{E}(N_1) = \lambda$  et  $\mathbb{E}(T_n) = \frac{1}{\lambda}$ . Ainsi, plus  $\lambda$  est grand, plus le nombre moyen d'arrivées par unité de temps est important, et plus l'intervalle entre 2 arrivées est court, ce qui semblait *a priori* évident. Pour cette raison, on appelle également le paramètre  $\lambda$  l'*intensité* du processus.

## 11.4 Propriétés supplémentaires

### 11.4.1 Processus de Poisson et loi binomiale

**Propriété 11.3.** Pour  $s \leq t$ , la loi conditionnelle de  $N_s$  sachant  $[N_t = n]$  est la loi binomiale  $\mathcal{B}\left(n, \frac{s}{t}\right)$ .

*Démonstration.*

$$\begin{aligned} P^{[N_t=n]}([N_s = k]) &= \frac{P([N_s = k] \cap [N_t = n])}{P([N_t = n])} = \frac{P([N_s = k] \cap [N_t - N_s = n - k])}{P([N_t = n])} \\ &= \frac{P([N_s = k])P([N_t - N_s = n - k])}{P([N_t = n])} = \frac{P([N_s = k])P([N_{t-s} = n - k])}{P([N_t = n])} \\ &= \frac{e^{-\lambda s} \frac{(\lambda s)^k}{k!} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{n-k}}{(n-k)!}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} = C_n^k \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}. \end{aligned}$$

□

### 11.4.2 Processus non homogènes

Soit  $(X_t)_{t \geq 0}$  un processus de comptage à accroissements indépendants vérifiant  $X_0 = 0$ ,

$P([X_{t+h} = i + 1] / [X_t = i]) = \lambda(t)h + o(h)$  et  $P([X_{t+h} = i] / [X_t = i]) = 1 - \lambda(t)h + o(h)$ .

On note  $A_n$  le  $n^{\text{ième}}$  temps d'arrivée du processus.

**Propriété 11.4.**  $X_t$  suit la loi de Poisson de paramètre  $\Lambda(t) = \int_0^t \lambda(u)du$  et  $T_1 = A_1$  suit la loi de densité  $t \mapsto \lambda(t) \exp\left[-\int_0^t \lambda(u)du\right] \mathbb{I}_{]0, +\infty[}(t)$ .

*Remarques :*

[1] Pour un processus de Poisson  $(X_t)$  d'intensité  $\lambda$ , on a

$$P([X_{t+h} = i + 1] / [X_t = i]) = \lambda h + o(h) \text{ et } P([X_{t+h} = i] / [X_t = i]) = 1 - \lambda h + o(h).$$

c'est-à-dire que  $\lambda$  ne dépend pas de  $t$ .

[2] Le processus étudié ici est une généralisation du processus de Poisson. Pour  $\lambda(t) = \lambda$  constant, on retrouve que  $X_t$  suit la loi de Poisson de paramètre  $\Lambda(t) = \lambda \times t$  et que  $T_1$  suit la loi exponentielle de paramètre  $\lambda$ .

*Démonstration.* On reprend la méthode utilisée pour montrer que, si  $(X_t)$  est un processus à accroissements indépendants, stationnaire et à événements rares, alors  $X_t$  suit la loi de Poisson  $\mathcal{P}(\lambda t)$  : on pose  $\pi_n(t) = P([X_t = n])$  puis, en considérant  $\pi_n(t+h)$  et en passant à la limite quand  $h \rightarrow 0$ , on va établir une équation différentielle.

$$P([X_{t+h} = 0]) = P([X_t = 0] \cap [X_{t+h} - X_t = 0]) = P([X_t = 0])(1 - \lambda(t)h + o(h)), \text{ soit}$$

$$\pi_0(t+h) - \pi_0(t) = -\lambda(t)h\pi_0(t) + o(h).$$

D'autre part, pour  $n \geq 1$ ,

$$\begin{aligned} P([X_{t+h} = n]) &= \sum_k P([X_t = k] \cap [X_{t+h} - X_t = n - k]) \\ &= P([X_t = n])(1 - \lambda(t)h) + P([X_t = n - 1])\lambda(t)h + o(h) \end{aligned}$$

soit  $\pi_n(t+h) - \pi_n(t) = (-\lambda(t)\pi_n(t) + \lambda(t)\pi_{n-1}(t))h + o(h)$ . On a donc, en divisant par  $h$  et en faisant  $h \rightarrow 0$  :

$$\boxed{\begin{cases} \pi'_0(t) + \lambda(t)\pi_0(t) = 0 \\ \pi'_n(t) + \lambda(t)\pi_n(t) = \lambda(t)\pi_{n-1}(t) \end{cases} \text{ avec } \pi_0(0) = 1 \text{ et } \pi_n(0) = 0 \text{ pour } n \geq 1.}$$

On résout alors, soit en utilisant les équations différentielles du premier ordre (et la méthode de variation de la constante), soit à l'aide de la fonction génératrice de  $X_t$  :

$$G(z, t) = \sum_{n=0}^{+\infty} z^n \pi_n(t).$$

Première méthode :

$$\frac{\pi'_0(t)}{\pi_0(t)} = [\ln(\pi_0(t))]' = -\lambda(t) \text{ et } \ln(\pi_0(t)) - \ln(\pi_0(0)) = -\int_0^t \lambda(u)du.$$

Ainsi,  $\pi_0(t) = P([X_t = 0]) = \exp(-\Lambda(t))$ . D'autre part, par la méthode de variation de la constante,

$$\pi_n(t) = C(t) \exp(-\Lambda(t)) \text{ avec } C'(t) \exp(-\Lambda(t)) = \lambda(t)\pi_{n-1}(t)$$

et par récurrence, si on suppose que  $\pi_{n-1}(t) = \exp(-\Lambda(t)) \frac{(\Lambda(t))^{n-1}}{(n-1)!}$ , on a alors  $C'(t) = \lambda(t) \frac{(\Lambda(t))^{n-1}}{(n-1)!}$

et, comme  $C(0) = 0$  ( $\pi_n(0) = 0$  pour  $n \geq 1$ ), on a alors  $C(t) = \frac{(\Lambda(t))^n}{n!}$  et on a bien :

$$\boxed{P([X_t = n]) = \pi_n(t) = \exp(-\Lambda(t)) \frac{(\Lambda(t))^n}{n!}.}$$

Deuxième méthode : On a  $\frac{\partial G}{\partial t}(z, t) + \lambda(t)G(z, t) = \lambda(t) \sum_{n=1}^{+\infty} z^n \pi_{n-1}(t) = \lambda(t)zG(z, t)$ , soit  $\frac{G'}{G} = (z-1)\lambda$

et, comme  $G(0) = \pi_0(0) = 1$ ,  $G(z, t) = \exp\left[(z-1) \int_0^t \lambda(u)du\right]$ . On reconnaît bien là la fonction génératrice de la loi de Poisson de paramètre  $\Lambda(t) = \int_0^t \lambda(u)du$ .

Pour la loi de  $T_1$ ,  $P([T_1 > t]) = P([X_t = 0]) = \exp(-\Lambda(t)) = 1 - F_{T_1}(t)$  pour  $t > 0$ . On a donc bien

$$\boxed{f_{T_1}(t) = \lambda(t) \exp(-\Lambda(t)) \text{ pour } t > 0.}$$

□