

**Cours de mathématiques**  
**Partie V – Probabilités**  
MP2I

**Alain TROESCH**

---

Version du:

19 mai 2023



# Table des matières

<b>28</b>	<b>Espaces probabilisés</b>	<b>3</b>
I	Espaces probabilisables . . . . .	3
I.1	Notion d'expérience aléatoire . . . . .	3
I.2	$\sigma$ -algèbres d'événements (ou tribus) . . . . .	4
II	Espaces probabilisés . . . . .	7
II.1	Mesures de probabilité . . . . .	7
II.2	Probabilités uniformes sur un univers fini . . . . .	10
II.3	Ensembles négligeables . . . . .	10
III	Conditionnement et indépendance . . . . .	11
III.1	Probabilités conditionnelles . . . . .	11
III.2	Indépendance . . . . .	11
IV	Les trois principaux théorèmes du calcul des probabilités . . . . .	14
IV.1	Formule des probabilités totales . . . . .	14
IV.2	Formule des probabilités composées . . . . .	15
IV.3	Formules de Bayes . . . . .	16
V	Principes généraux du calcul des probabilités . . . . .	17
<b>29</b>	<b>Variables aléatoires</b>	<b>21</b>
I	Variables aléatoires . . . . .	21
I.1	Applications mesurables . . . . .	21
I.2	Variables aléatoires . . . . .	23
I.3	Loi d'une variable aléatoire . . . . .	25
I.4	Variables aléatoires discrètes . . . . .	26
I.5	Loi de probabilité d'une variable aléatoire discrète . . . . .	27
I.6	Loi de $f(X)$ . . . . .	27
I.7	Loi d'un vecteur aléatoire . . . . .	28
II	Espérance mathématique . . . . .	30
II.1	Espérance d'une variable aléatoire réelle discrète . . . . .	30
II.2	Théorème de transfert et propriétés de l'espérance . . . . .	31
II.3	Théorème de l'espérance totale (HP) . . . . .	33
III	Variance, dispersion, moments . . . . .	33
III.1	Moments d'ordre $k$ . . . . .	34
III.2	Variance . . . . .	35
III.3	Variables centrées réduites . . . . .	36
IV	Indépendance de variables aléatoires . . . . .	36

---

IV.1	Couples de variables aléatoires indépendantes . . . . .	36
IV.2	Familles de v.a.r. indépendantes . . . . .	37
IV.3	Coalitions . . . . .	38
V	Covariance . . . . .	39
V.1	Espérance d'un produit . . . . .	39
V.2	Covariance . . . . .	40
V.3	Variance d'une somme . . . . .	41
VI	Lois discrètes classiques . . . . .	42
VI.1	Loi quasi-certaine . . . . .	42
VI.2	Loi uniforme . . . . .	43
VI.3	Loi de Bernoulli . . . . .	43
VI.4	Loi binomiale – nombre de succès . . . . .	44
VI.5	Loi géométrique – temps d'attente du premier succès . . . . .	45
VI.6	Loi de Poisson . . . . .	46
VI.7	Stabilité des lois classiques . . . . .	47
VI.8	Tableau récapitulatif . . . . .	47
VII	Inégalités et convergences . . . . .	47
VII.1	Convergence en probabilités . . . . .	48
VII.2	Inégalités en probabilités . . . . .	48
VII.3	Loi faible des grands nombres . . . . .	49

# Espaces probabilisés

« Ce n'est qu'en essayant continuellement que l'on finit par réussir, donc plus ça rate, plus on a de chances que ça marche »

« La fusée interplanétaire des Shadoks n'était pas très au point, mais ils avaient calculé qu'elle avait quand même une chance sur un million de marcher. Et ils se dépêchaient de bien rater les 999999 premiers essais pour être sûrs que le millionième marche. »

(Jacques Rouxel)

Il s'agit dans ce chapitre de donner un cadre rigoureux au calcul des probabilités, et à établir un certain nombre de formules calculatoires essentielles, permettant de mener à bien et de façon rigoureuse la plupart des calculs de probabilités.

## I Espaces probabilisables

### I.1 Notion d'expérience aléatoire

Une *expérience aléatoire* est une donnée intuitive : c'est une expérience dont le résultat ne dépend que du hasard. Plus explicitement, cela signifie que s'il est possible de répéter l'expérience plusieurs fois dans les mêmes conditions et de façon identique, les différents résultats ne seront pas nécessairement les mêmes. Ainsi, le résultat d'une expérience aléatoire n'est en général pas prévisible. L'exemple typique est le lancer d'une pièce, qui va retomber sur Pile ou Face, ou bien le lancer d'un dé : les lois physiques régissant la mécanique d'une pièce lancée en l'air ou d'un dé roulant sur une table sont trop complexes et trop sensibles à des variations des conditions initiales, et aux éventuelles perturbations extérieures pour espérer obtenir un résultat déterministe, même avec un lancer très précis.

Une expérience, malgré son nom, peut être purement théorique, dans le sens où elle ne peut pas être réalisée de façon effective. Par exemple l'expérience consistant à lancer une infinité de fois une pièce de monnaie. Il peut aussi s'agir d'une simple observation d'un phénomène naturel ne nécessitant pas d'intervention de l'observateur. C'est le cas en physique (notamment en physique des particules).

#### Définition 28.1.1 (Univers)

- Un *résultat*, ou une *issue* de l'expérience est une donnée issue de l'expérience aléatoire ; une même expérience peut fournir différents résultats, suivant ce qu'on veut étudier de l'expérience.
- L'*univers des possibles*  $\Omega$  (ou plus simplement l'univers) est l'ensemble des issues possibles d'une expérience.

Une même expérience peut fournir plusieurs univers différents suivant ce qu'on veut en tirer.

Intuitivement, un événement correspond à un groupement d'issues possibles vérifiant une certaine propriété. Ainsi, il s'agit d'un ensemble d'issues, donc d'un sous-ensemble de l'univers.

**Définition 28.1.2 (Événement, définition intuitive)**

Un événement est un sous-ensemble de  $\Omega$ . Pour certaines raisons techniques, lorsque  $\Omega$  n'est pas fini, on est parfois amené à se restreindre et à ne pas considérer tous les sous-ensembles comme des événements, ce que nous formaliserons plus loin avec la notion de tribu.

La plupart des notions ensemblistes ont une traduction dans le langage des probabilités, afin de mieux traduire leur interprétation intuitive.

**Terminologie 28.1.3 (Vocabulaire probabiliste)**

- *Événement élémentaire*, ou *épreuve* : un singleton  $\{\omega\} \subset \Omega$ .
- *Événement certain* : l'événement  $\Omega$ .
- *Événement impossible* : l'événement  $\emptyset$ .
- *Événement contraire de l'événement  $A$*  : l'événement  $\bar{A} = \mathbb{C}_\Omega A$ .
- *$A$  entraîne  $B$  si  $A \subset B$*
- *Événement «  $A$  et  $B$  »* :  $A \cap B$  (réalisation simultanée)
- *Événement «  $A$  ou  $B$  »* :  $A \cup B$
- *$A$  et  $B$  sont dits incompatibles si  $A \cap B = \emptyset$ .*
- une famille  $\{A_i, i \in I\}$  est dite *constituée d'événements deux à deux incompatibles* si pour tout  $(i, j) \in I^2$  tel que  $i \neq j$ ,  $A_i$  et  $A_j$  sont incompatibles.

**Définition 28.1.4 (Système complet d'événements)**

Un système complet d'événements est une famille  $\{A_i, i \in I\}$  formant une partition de  $\Omega$ . Autrement dit :

- Les événements  $A_i$  sont non vides ;
- La famille est constituée d'événements deux à deux incompatibles ;
- $\biguplus_{i \in I} A_i = \Omega$ .

**I.2  $\sigma$ -algèbres d'événements (ou tribus)**

Souvent, la définition de la mesure de probabilité sur tous les sous-ensembles de  $\Omega$ , n'est pas pertinente, ou peut poser des problèmes techniques. Pour cette raison, il peut être intéressant de pouvoir ne définir la mesure de probabilité que sur une classe particulière de sous-ensembles de  $\Omega$ . L'objet de ce paragraphe et du suivant est de définir un type satisfaisant de classe de sous-ensembles à laquelle se restreindre.

Voici nos exigences pour la construction d'une telle classe :

- L'événement certain  $\Omega$  et l'événement impossible ont une probabilité, donc appartiennent à la classe ;
- Pour tout événement  $A$  dont on sait calculer la probabilité, on voudrait pouvoir calculer la probabilité de l'événement contraire  $\bar{A}$  ;
- Si on dispose d'une probabilité sur  $A$  et  $B$ , on voudrait disposer d'une probabilité de  $A \cup B$ .
- Plus généralement, on souhaite pouvoir calculer la probabilité d'une union infinie dénombrable d'événements.

**Remarque 28.1.5**

Le dernier point est motivé par l'étude d'un tirage infini à Pile ou Face : quelle est la probabilité qu'un tirage donné (par exemple n'obtenir que des Piles) ait lieu ? Pour ce calcul, on peut encore se dispenser des unions (ou intersections par passage au complémentaire) infinies, en effectuant des majorations, si on définit correctement la probabilité ; mais en revanche, si on se demande maintenant quelle est

la probabilité qu'un résultat soit stationnaire, on a besoin de regrouper une infinité de cas possibles (stationnaires à partir d'un certain rang).

Cela nous amène à la définition suivante :

**Définition 28.1.6 ( $\sigma$ -algèbre, ou tribu, Spé)**

Soit  $\Omega$  un univers (fini ou non). Une  $\sigma$ -algèbre  $\mathcal{A}$  d'événements sur  $\Omega$  (ou tribu) est un sous-ensemble  $\mathcal{T}$  de  $\mathcal{P}(\Omega)$  telle que :

1.  $\Omega \in \mathcal{A}$ ;
2.  $\forall A \in \mathcal{P}(\Omega), A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$  (stabilité par complémentation);
3. Pour toute famille dénombrable  $(A_i)_{i \in I}$  d'éléments de  $\mathcal{T}$ ,  $\bigcup_{i \in I} A_i$  est dans  $\mathcal{T}$  (stabilité par union dénombrable)

Nous complétons les propriétés imposées par la définition par les suivantes :

**Proposition 28.1.7 (Propriétés des tribus, Spé)**

Soit  $\mathcal{T}$  une tribu d'événements sur  $\Omega$ .

1.  $\emptyset \in \mathcal{T}$ ;
2.  $\forall (A, B) \in \mathcal{P}(\Omega), (A, B) \in \mathcal{T}^2 \implies A \cup B \in \mathcal{T}$  (stabilité par union finie);
3. Pour toute famille dénombrable  $(A_n)_{n \in \mathbb{N}}$  d'éléments de  $\mathcal{T}$ ,  $\bigcap_{n=0}^{+\infty} A_n \in \mathcal{T}$ .
4.  $\forall (A, B) \in \mathcal{P}(\Omega), (A, B) \in \mathcal{T}^2 \implies A \cap B \in \mathcal{T}$  (stabilité par intersection finie);

◁ **Éléments de preuve.**

Le point 2 se montre en complétant la famille  $(A, B)$  en une famille dénombrable, en ajoutant l'ensemble vide, ce qui ne change pas l'union. Les autres s'obtiennent par passage au complémentaire.

▷

Par une récurrence immédiate, les points 2 et 4 se généralisent facilement aux unions ou intersections d'un nombre fini quelconque d'événements.

**Définition 28.1.8 (Espace probabilisable et événements, Spé)**

- Un *espace probabilisable* est un couple  $(\Omega, \mathcal{T})$ , où  $\Omega$  est un ensemble quelconque, appelé univers, et  $\mathcal{T}$  une  $\sigma$ -algèbre d'événements sur  $\Omega$ .
- Les éléments de  $\mathcal{T}$  sont appelés *événements* (de l'espace probabilisable  $(\Omega, \mathcal{T})$ ).

**Lemme 28.1.9 (Intersection de tribus, Spé)**

Soit  $(\mathcal{T}_i)_{i \in I}$  une famille de  $\sigma$ -algèbres sur  $\Omega$ . Alors  $\bigcap_{i \in I} \mathcal{T}_i$  est une  $\sigma$ -algèbre d'événements.

◁ **Éléments de preuve.**

Vérification immédiate de la préservation des trois points de la définition par intersection. ▷

**Définition 28.1.10 (Tribu engendrée par une famille, Spé)**

Soit  $\mathcal{C} = \{A_i, i \in I\}$  une famille quelconque d'événements dans  $\mathcal{P}(\Omega)$ . La  $\sigma$ -algèbre  $\mathcal{T}$  (notée  $\sigma(\mathcal{C})$  ou  $\mathcal{T}(\mathcal{C})$ ) engendrée par la famille  $\mathcal{C}$  est la plus petite  $\sigma$ -algèbre contenant  $\mathcal{C}$ . Elle vérifie donc :

1.  $\mathcal{C} \subset \mathcal{T}$ ;
2. pour toute  $\sigma$ -algèbre  $\mathcal{T}'$ ,  $\mathcal{C} \subset \mathcal{T}' \implies \mathcal{T} \subset \mathcal{T}'$ .

**Proposition 28.1.11 (Existence de la tribu engendrée)**

Cette  $\sigma$ -algèbre existe.

◁ **Éléments de preuve.**

C'est l'intersection de toutes les tribus  $\mathcal{T}'$  vérifiant  $\mathcal{C} \subset \mathcal{T}'$ . ▷

**Remarque 28.1.12**

La  $\sigma$ -algèbre engendrée par une famille contient au moins tous les complémentaires des événements de cette famille, ainsi que toutes les unions et intersections finies ou dénombrables des événements de la famille et de leurs complémentaires.

**Proposition 28.1.13 (Tribu engendrée par les singletons)**

Soit  $\Omega$  un univers dénombrable. La  $\sigma$ -algèbre engendrée par l'ensemble des événements élémentaires  $\{\{\omega\}, \omega \in \Omega\}$  est  $\mathcal{P}(\Omega)$ .

◁ **Éléments de preuve.**

Tout sous-ensemble de  $\Omega$  est intersection au plus dénombrable de ses singletons. ▷

**Remarque 28.1.14**

Dans le cas où  $\Omega$  est dénombrable, si on veut pouvoir considérer les événements élémentaires (réalisations ponctuelles) comme des événements, la seule tribu possible sur  $\Omega$  est  $\mathcal{P}(\Omega)$ . Ainsi, lorsque  $\Omega$  est fini ou dénombrable, l'espace probabilisable que l'on considère généralement est  $(\Omega, \mathcal{P}(\Omega))$ .

Enfin, voici une tribu d'une grande importance dans l'étude des variables aléatoires réelles :

**Définition 28.1.15 (tribu des boréliens, HP)**

La *tribu borélienne* sur  $\mathbb{R}$  est la tribu  $\mathcal{B}^1$ , souvent notée plus simplement  $\mathcal{B}$ , engendrée par tous les intervalles  $] - \infty, a]$ . Les ensembles de cette tribu sont appelés *ensembles boréliens*, ou tout simplement *boréliens*.

**Proposition 28.1.16 (D'autres systèmes de générateurs de  $\mathcal{B}^1$ )**

La tribu des boréliens contient tous les intervalles. Elle est aussi engendrée par les intervalles  $] - \infty, a[$ , ou encore par les  $[a, b]$ , ou encore par les  $[a, b[$  etc.

◁ **Éléments de preuve.**

Par exemple, pour  $] - \infty, a[$ , montrer que ces ensembles sont dans  $\mathcal{B}^1$  (les écrire comme union de  $] - \infty, a_n]$ ). Réciproquement, les  $] - \infty, a]$  sont dans la tribu engendrée par les  $] - \infty, b[$  (écrire cette fois  $] - \infty, a]$  comme intersection de  $] - \infty, a_n[$ ). Les autres se démontrent sur le même principe. ▷

**Définition 28.1.17 (Tribu des boréliens sur  $\mathbb{R}^n$ , HP)**

La tribu  $\mathcal{B}^n$  des boréliens de  $\mathbb{R}^n$  est la tribu sur  $\mathbb{R}^n$  engendrée par les pavés  $I_1 \times \cdots \times I_n$ , les  $I_k$  étant des intervalles (dont on peut imposer le type, à condition qu'ils soient de longueur non nulle)

**II Espaces probabilisés**

Les notions de ce paragraphe ne sont au programme de Sup que dans le cas où  $\Omega$  est fini et où l'espace probabilisable considéré est  $(\Omega, \mathcal{P}(\Omega))$ . Vous pouvez vous limiter à ce cas, mais nous généralisons à un espace probabilisable  $(\Omega, \mathcal{T})$  quelconque (programme de Spé).

**II.1 Mesures de probabilité**

Intuitivement, une probabilité mesure la possibilité qu'un événement donné se produise. Il s'agit donc d'un réel  $p$  indiquant la fréquence probable de réalisation de l'événement lorsqu'on répète un grand nombre de fois l'expérience dans des situations similaires, c'est-à-dire le rapport moyen entre le nombre de réalisations et le nombre d'expériences.

Ainsi, une probabilité est un réel positif et inférieur à 1. De plus, toutes les expériences amenant une issue de  $\Omega$ , la probabilité de l'événement  $\Omega$  doit être 1. Enfin, si  $A$  et  $B$  sont deux événements ne pouvant pas être réalisés simultanément, le nombre d'expériences amenant  $A$  ou  $B$  est la somme du nombre d'expériences amenant  $A$  et du nombre d'expériences amenant  $B$  (puisque'une expérience ne peut amener à la fois  $A$  et  $B$ ). De l'interprétation ci-dessus, il doit alors venir que la probabilité de l'union disjointe  $A \sqcup B$  doit être la somme des probabilités de  $A$  et de  $B$ . Cela motive la définition suivante :

**Définition 28.2.1 (Mesure de probabilité)**

Soit  $(\Omega, \mathcal{T})$  un espace probabilisable. Une mesure de probabilité sur  $(\Omega, \mathcal{T})$  est une application  $\mathbb{P} : \mathcal{T} \rightarrow \mathbb{R}$  telle que :

1.  $\forall A \in \mathcal{T}, 0 \leq \mathbb{P}(A) \leq 1$  ;
2.  $\mathbb{P}(\Omega) = 1$
3. (propriété de  $\sigma$ -additivité) Pour toute famille dénombrable  $(A_n)_{n \in \mathbb{N}}$  d'événements deux à deux incompatibles,  $\sum \mathbb{P}(A_n)$  converge, et on a :  $\mathbb{P}\left(\biguplus_{n \in \mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$ .

**Remarque 28.2.2**

Les probabilités étant positives, la convergence de  $\sum \mathbb{P}(A_n)$  est absolue, et  $(\mathbb{P}(A_n))_{n \in \mathbb{N}}$  est donc une famille sommable. Les propriétés de convergence, ainsi que la valeur de la somme sont donc indépendantes de l'ordre de sommation. On peut donc plus généralement considérer une famille  $(A_i)_{i \in I}$  dénombrable (mais indexée sur un autre ensemble dénombrable  $I$ , ne possédant pas d'ordre naturel). Ainsi, si  $(A_i)_{i \in I}$  est une famille dénombrable d'événements 2 à 2 incompatibles,

$$\mathbb{P}\left(\biguplus_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i).$$

**Proposition 28.2.3 (Propriétés d'une mesure de probabilité)**

Soit  $\mathbb{P}$  une mesure de probabilités sur  $(\Omega, \mathcal{T})$ . Alors :

1.  $\mathbb{P}(\emptyset) = 0$  ;
2. (Additivité) Pour tout couple  $(A, B)$  d'événements incompatibles,  $\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B)$  ;

3.  $\forall A \in \mathcal{T}, \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$  ;
4.  $\forall (A, B) \in \mathcal{T}, A \subset B \implies \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$  ;
5.  $\forall (A, B) \in \mathcal{T}, A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$  ;
6.  $\forall (A, B) \in \mathcal{T}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$  ;
7. (Formule du crible de Poincaré, HP) *Plus généralement, soit  $(A_1, \dots, A_n) \in \mathcal{T}^n$ . Alors :*

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{\substack{I \subset [1, n] \\ I \neq \emptyset}} (-1)^{|I|+1} \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

◁ **Éléments de preuve.**

$\mathbb{P}(\emptyset)$  s'obtient en étudiant la convergence de  $\sum \mathbb{P}(\emptyset)$ , convergence assurée par la  $\sigma$ -additivité. L'additivité en découle en complétant la famille finie  $(A, B)$  en une famille dénombrable par ajout de  $\emptyset$ . Les points 3,4,5,6 sont évident, en décomposant le plus grand ensemble en union disjointe d'ensembles plus petits. La formule du crible s'obtient par récurrence, exactement comme dans le cas des cardinaux. ▷

**Remarque 28.2.4**

Dans le cas où  $\Omega$  est fini (cadre du programme de Sup), la  $\sigma$ -additivité n'est pas pertinente, car une famille infinie  $(A_n)_{n \in \mathbb{N}}$  ne peut être constituée d'éléments 2 à 2 disjoints que si seul un nombre fini d'événements  $A_i$  ne sont pas vides. Ainsi, dans le cas où  $\Omega$  est fini, on peut se contenter de définir une mesure de probabilité en remplaçant la propriété de  $\sigma$ -additivité par la propriété d'additivité.

Le théorème suivant découle directement de la  $\sigma$ -additivité. Il n'a d'intérêt que dans le cas où  $\Omega$  n'est pas fini (sinon, les suites considérées sont stationnaires). Pour cette raison, il sort du cadre du programme de Sup (mais pas de Spé).

**Théorème 28.2.5 (Propriété de limite monotone, ou continuité monotone de la mesure  $\mathbb{P}$ )**

1. Soit  $(A_n)_{n \in \mathbb{N}}$  une suite croissante d'événements (pour l'inclusion). Alors :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n).$$

2. Soit  $(A_n)_{n \in \mathbb{N}}$  une suite décroissante d'événements. Alors :

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

◁ **Éléments de preuve.**

Écrire  $\bigcup_{n \in \mathbb{N}} A_n = \bigsqcup_{n \in \mathbb{N}} (A_n \setminus A_{n-1})$ , avec  $A_{-1} = \emptyset$ , puis  $\sigma$  additivité. Passer au complémentaire pour l'intersection. ▷

**Corollaire 28.2.6**

Soit  $(A_n)_{n \in \mathbb{N}}$  une famille quelconque d'événements. Alors

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

En définissant pour une suite croissante  $(A_n)$ ,  $\bigcup_{n \in \mathbb{N}} A_n = \lim \uparrow A_n$  et de même pour une suite décroissante,  $\bigcap_{n \in \mathbb{N}} A_n = \lim \downarrow A_n$ , la propriété ci-dessus se réécrit :

$$\mathbb{P}(\lim \uparrow A_n) = \lim \mathbb{P}(A_n) \quad \text{et} \quad \mathbb{P}(\lim \downarrow A_n) = \lim \mathbb{P}(A_n).$$

Il s'agit donc d'une propriété de continuité séquentielle, d'où le nom donné au théorème.

Les unions et intersections ci-dessus dans le cas de suites monotones sont un cas particulier d'une construction plus générale :

**Définition 28.2.7 (Limite supérieure et limite inférieure, HP)**

Soit  $(A_n)$  une suite d'événements. On définit

- la limite supérieure  $\limsup A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k=n}^{\infty} A_k$  : réalisation d'une infinité de  $A_k$  ;
- la limite inférieure  $\liminf A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{k=n}^{\infty} A_k$  : réalisation de presque tous les  $A_k$  (tous sauf un nombre fini, c'est-à-dire tous à partir d'un certain rang).

**Corollaire 28.2.8**

Soit  $(A_n)$  une suite d'événements.

- La probabilité qu'un nombre infini de  $A_k$  se réalisent est

$$\mathbb{P}(\limsup A_n) = \lim \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right).$$

- La probabilité que presque tous les  $A_k$  se réalisent est

$$\mathbb{P}(\liminf A_n) = \lim \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k\right).$$

**Remarque 28.2.9**

Si  $(A_n)$  est décroissante,  $\bigcup_{k=n}^{+\infty} A_k = A_n$  et  $\limsup A_n = \lim \downarrow A_n$  ; de même si  $(A_n)$  est croissante,  $\liminf A_n = \lim \uparrow A_n$ . Le corollaire précédent englobe donc le théorème de continuité monotone.

Nous avons maintenant tout ce qu'il faut pour nous définir un cadre rigoureux pour nos probabilités.

**Définition 28.2.10 (Espace probabilisé, ou modèle probabiliste de Kolmogorov)**

Un *espace probabilisé* est un triplet  $(\Omega, \mathcal{T}, \mathbb{P})$  où  $(\Omega, \mathcal{T})$  est un espace probabilisable, et  $\mathbb{P}$  une mesure de probabilité sur  $(\Omega, \mathcal{T})$ .

On termine cette section en donnant un résultat permettant de décrire de façon rapide et commode une mesure de probabilité sur un univers fini ou dénombrable.

**Définition 28.2.11 (Distribution de probabilités)**

Soit  $I$  un ensemble. Une distribution de probabilités sur  $I$  est une famille sommable  $(p_i)_{i \in I}$  telle que  $\sum_{i \in I} p_i = 1$ .

**Proposition 28.2.12 (Détermination d'une probabilité par l'image des singletons)**

Soit  $\Omega$  un ensemble fini ou dénombrable. Une mesure de probabilité est entièrement déterminée par la distribution de probabilités  $(\mathbb{P}(\{\omega\}))_{\omega \in \Omega}$

◁ **Éléments de preuve.**

Unicité car les singletons engendrent  $\mathcal{P}(\Omega)$ . Expliciter  $p$  ainsi obtenue en fonction des probabilités des singletons, et vérifier qu'il s'agit bien d'une mesure de probabilité. ▷

Ainsi, pour définir une mesure de probabilité sur  $\Omega$  fini ou dénombrable, il suffit de se donner une distribution de probabilités  $(p_\omega)_{\omega \in \Omega}$ .

**II.2 Probabilités uniformes sur un univers fini**

Soit  $\Omega$  un univers fini. On considère l'espace probablisable  $(\Omega, \mathcal{P}(\Omega))$ .

D'après la proposition 28.2.12, pour définir une mesure de probabilité  $\mathbb{P}$  sur  $(\Omega, \mathcal{P}(\Omega))$ , il suffit de définir  $\mathbb{P}$  sur les singletons de sorte que la somme soit égale à 1. Voici un cas particulier très important :

**Définition 28.2.13 (Probabilité uniforme sur un univers fini)**

La *probabilité uniforme* (ou *équipartition*) sur  $(\Omega, \mathcal{P}(\Omega))$  est la probabilité définie par :

$$\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}.$$

On retrouve alors, à partir du cadre formel qu'on s'est fixé, le résultat que vous utilisez intuitivement depuis longtemps déjà :

**Théorème 28.2.14 (Formule de Laplace)**

Soit  $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$  l'espace probablisé uniforme sur l'univers fini  $\Omega$ . Alors, pour tout  $A \in \mathcal{P}(\Omega)$  :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

◁ **Éléments de preuve.**

Écrire  $A$  comme union de ses singletons, puis utiliser l'additivité. ▷

**II.3 Ensembles négligeables****Définition 28.2.15 (Ensembles négligeables, HP)**

Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace probablisé, et soit  $A \in \mathcal{T}$ .

1. On dit que  $A$  est *négligeable*, ou *presque-impossible*, ou *quasi-impossible* si  $\mathbb{P}(A) = 0$ .
2. On dit que  $A$  est *presque-certain* ou *quasi-certain* si  $\mathbb{P}(A) = 1$ .
3. Une propriété est vraie *presque sûrement* si l'événement « la propriété est satisfaite » est presque-certain.

**Exemple 28.2.16**

L'obtention d'une suite infinie de Pile dans une succession infinie de tirages d'une pièce équilibrée est un événement presque-impossible, mais pas impossible.

## III Conditionnement et indépendance

### III.1 Probabilités conditionnelles

On cherche maintenant à voir comment la connaissance de certaines informations sur le résultat d'une expérience modifie les probabilités. Pour cela, on définit la probabilité conditionnelle d'un événement  $A$  sachant que l'événement  $B$  est réalisé : la connaissance de la réalisation de  $B$  peut influencer sur les chances que  $A$  se réalise aussi. Par exemple si on tire un dé à 6 faces équilibré, la probabilité d'obtenir 1 est  $\frac{1}{6}$ . Mais si on sait qu'on a obtenu une valeur impaire (événement  $B$ ), on a une chance sur 3 pour que ce soit 1.

**Définition 28.3.1 (Probabilités conditionnelles)**

Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace probabilisé,  $A$  et  $B$  deux événements de  $\mathcal{T}$  tels que  $B$  ne soit pas presque-impossible. Alors, la *probabilité conditionnelle de  $A$  en  $B$*  (ou la *probabilité de  $A$  sachant  $B$* ) est :

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

On trouve également souvent la notation  $\mathbb{P}(A | B)$ .

Cette définition se comprend bien pour la mesure uniforme : il s'agit de la proportion des éléments de  $B$  qui satisfont  $A$  (c'est-à-dire la proportion des éléments de  $B$  qui sont dans  $A \cap B$ ).

**Proposition 28.3.2**

Soit  $B$  un événement non presque-impossible. Alors  $\mathbb{P}_B$  définit une mesure de probabilité sur l'espace probabilisable  $(\Omega, \mathcal{T})$ .

◁ **Éléments de preuve.**

Vérifications sans difficulté.

▷

**Remarque 28.3.3**

On utilise souvent la relation définissant les probabilités conditionnelles sous la forme suivante :  $\mathbb{P}(A \cap B) = \mathbb{P}(A | B)\mathbb{P}(B)$ .

En effet, on se sert souvent des probabilités conditionnelles pour calculer les probabilités d'intersections, et non l'inverse, car les probabilités conditionnelles sont souvent plus simples à calculer. On verra un peu plus loin une généralisation de cette formule à une intersection de  $n$  événements (formule des probabilités composées, théorème 28.4.7)

### III.2 Indépendance

Les probabilités conditionnelles mesurent l'impact de la réalisation d'un événement sur un autre. Si cet impact est nul (c'est-à-dire si la connaissance de la réalisation d'un événement n'influe pas sur la probabilité de l'autre), on dira que ces deux événements sont indépendants.

Ainsi, intuitivement, deux événements  $A$  et  $B$  sont indépendants si la connaissance de l'un ne modifie pas la probabilité de l'autre, donc si  $\mathbb{P}(A | B) = \mathbb{P}(A)$  et  $\mathbb{P}(B | A) = \mathbb{P}(B)$ . L'inconvénient de cette définition réside dans le fait qu'elle n'est valable que sous une hypothèse de non quasi-impossibilité, afin de pouvoir considérer les probabilités conditionnelles. On constate cependant que dans le cas de non quasi-impossibilité, les deux égalités sont équivalentes à l'égalité :  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Cette égalité pouvant être considérée sans hypothèse de quasi-impossibilité, c'est elle que nous prenons comme définition de l'indépendance. Elle a en outre l'avantage d'être symétrique.

**Définition 28.3.4 (Événements indépendants)**

Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace probabilisé, et  $(A, B) \in \mathcal{T}^2$ . Les événements  $A$  et  $B$  sont *indépendants* si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .

En faisant le cheminement inverse de celui précédant la définition, il vient :

**Proposition 28.3.5 (Probabilités conditionnelles et indépendance)**

Soit  $A$  et  $B$  deux événements indépendants tel que  $B$  ne soit pas quasi-impossible. Alors

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

On a évidemment aussi  $\mathbb{P}(B) = \mathbb{P}(B | A)$  si  $A$  n'est pas quasi-impossible.

On note dans ce qui suit  $\mathcal{P}_f(I)$  l'ensemble des parties **finies** de  $I$ .

**Définition 28.3.6 (Mutuelle indépendance)**

Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace probabilisé, et  $(A_i)_{i \in I}$  une famille d'événements. On dit que les  $A_i, i \in I$ , sont mutuellement indépendants si et seulement si pour tout  $J \in \mathcal{P}_f(I)$ ,

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j).$$

En particulier si  $I$  lui-même est fini en prenant  $J = I$  :

**Proposition 28.3.7 (Probabilité d'une intersection d'événements mut. indépendants)**

Soit  $(A_i)_{i \in I}$  une famille finie d'événements mutuellement indépendants. Alors :

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

**Avertissement 28.3.8**

L'égalité ne suffit pas à avoir l'indépendance mutuelle. Il faut bien contrôler TOUS les produits des sous-familles (finies)

On verra dans le paragraphe suivant comment calculer la probabilité d'une intersection lorsque les événements ne sont pas mutuellement indépendants (formule des probabilités composées, théorème 28.4.7).

**Exemple 28.3.9**

On fait des tirages à pile ou face

- $A$  est réalisé si et seulement si le premier tirage est Pile.
- $B$  est réalisé si et seulement si lors des 3 premiers tirages, il y a au plus un Pile.
- $C = B$

On a  $\mathbb{P}(A \cap B \cap C) = \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ , mais les événements ne sont clairement pas mutuellement indépendants !

De la définition même découle :

**Proposition 28.3.10 (Stabilité de la mutuelle indépendance par restriction)**

Soit  $(A_i)_{i \in I}$  une famille d'événements mutuellement indépendants et  $J \subset I$ . Alors  $(A_j)_{j \in J}$  est encore composée d'événements mutuellement indépendants.

En particulier, en prenant toutes les sous-familles de 2 événements, la mutuelle indépendance implique l'indépendance 2 à 2 des événements.

**Avertissement 28.3.11**

La réciproque est fautive : l'indépendance 2 à 2 d'une famille n'implique pas l'indépendance mutuelle.

**Exemple 28.3.12**

On procède à trois tirages à Pile ou Face. Soit  $A$  l'événement consistant à obtenir exactement 1 Pile lors des tirages 2 et 3,  $B$  de même lors des tirages 1 et 3,  $C$  de même lors des tirages 1 et 2. Les événements  $A$ ,  $B$  et  $C$  sont 2 à 2 indépendants, mais pas mutuellement.

**Proposition 28.3.13 (Indépendance de sous-familles disjointes)**

Soit  $(A_i)_{i \in I}$  une famille d'événements mutuellement indépendants, et  $(I_k)_{k \in K}$  une famille de sous-ensembles finis de  $I$ , 2 à 2 disjoints. Alors la famille  $\left( \bigcap_{i \in I_k} A_i \right)_{k \in K}$  est constitué d'événements mutuellement indépendants.

◁ **Éléments de preuve.**

Prendre une intersection finie de tels ensembles, c'est aussi une intersection finie de  $A_i$ . Écrire par mutuelle indépendance la probabilité comme produit de  $\mathbb{P}(A_i)$ , puis regrouper par paquets de  $I_k$  en réutilisant la mutuelle indépendance sur ces paquets. ▷

**Proposition 28.3.14 (Indépendance et complémentation)**

Si  $A$  et  $B$  sont indépendants, alors  $A$  et  $\overline{B}$  sont indépendants.

◁ **Éléments de preuve.**

Calculer  $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A \setminus (A \cap B))$ . ▷

**Proposition 28.3.15 (Mutuelle indépendance et complémentation)**

Si  $A_1, \dots, A_n$  sont mutuellement indépendants, il en est de même de  $\overline{A_1}, A_2, \dots, A_n$

◁ **Éléments de preuve.**

Considérer l'intersection d'une sous-famille. Seul cas à voir :  $\overline{A_1}$  est dans la sous-famille. Se ramener au cas précédent en justifiant que  $A_1$  est indépendant de l'intersection des autres termes de la sous-famille. ▷

En appliquant plusieurs fois de suite cette propriété, on peut compléter un nombre quelconque des événements  $A_i$ . Par ailleurs, il est facile de voir que cela reste valable pour des familles quelconques. Cette propriété pourrait se généraliser en terme de tribu, mais cela dépasse largement le cadre du programme.

## IV Les trois principaux théorèmes du calcul des probabilités

### IV.1 Formule des probabilités totales

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

**Définition 28.4.1 (Système quasi-complet d'événements, HP)**

Soit  $\mathcal{C} \subset \mathcal{T}$ . On dit que  $\mathcal{C}$  est un système quasi-complet d'événements si les événements de  $\mathcal{C}$  sont non impossibles, deux à deux incompatibles, et si  $\sum_{A \in \mathcal{C}} \mathbb{P}(A) = 1$ .

Ainsi, la différence avec un système complet réside uniquement dans le fait que l'union des événements n'est égale à  $\Omega$  qu'à un ensemble négligeable près (l'union est presque-certaine et non certaine).

Un système complet est bien entendu aussi quasi-complet, la réciproque étant fausse.

**Théorème 28.4.2 (Formule des probabilités totales)**

Soit  $(A_i)_{i \in I}$  un système quasi-complet fini ou dénombrable, tels que les événements  $A_i$  ne sont pas presque-impossibles, et soit  $B \in \mathcal{T}$ . Alors

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

cette somme étant (absolument) convergente.

◁ **Éléments de preuve.**

Écrire  $B = \left( \bigsqcup_{i \in I} B \cap A_i \right) \sqcup (B \cap C)$ , où  $\mathbb{P}(C) = 0$ . ▷

**Remarque 28.4.3**

La formule des probabilités totales permet de retrouver une probabilité  $\mathbb{P}(B)$  connaissant les probabilités conditionnelles  $\mathbb{P}(B | A_i)$ . Ainsi, on peut calculer  $\mathbb{P}(B)$  en « distinguant suivant un certain nombre de cas » (suivant que  $A_i$  est réalisé), si ces cas partagent l'ensemble (ou presque) des cas possibles. Ainsi, cette formule est adapté au cas où le résultat d'une expérience dépend de résultats antérieurs, donc si on est amené à faire une discussion pour décrire le résultat de l'expérience.

**Remarque 28.4.4**

La somme intervenant dans la formule des probabilités totales étant absolument convergente (la convergence absolue découlant ici de la convergence, puisque la somme est à termes positifs), elle est correctement définie, de façon indépendante de l'ordre de sommation. Ce point est crucial ici, puisque nous ne nous sommes pas donné d'ordre sur l'ensemble  $I$  des indices.

Voici un cas particulier très important.

**Corollaire 28.4.5 (Formule des probabilités totales pour le système complet  $(A, \bar{A})$ )**

Soit  $A$  un événement non quasi-impossible et non quasi-certain. Alors, pour tout événement  $B$ ,

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

En anticipant un peu sur les variables aléatoires, nous donnons un autre cas important. Vous pouvez passer ce cas à première lecture, et y revenir un peu plus tard.

Étant donné une variable aléatoire réelle discrète (donc une fonction de  $\Omega$  dans  $\mathbb{R}$  telle que son image  $X(\Omega)$  soit au plus dénombrable) définissons les événements

$$[X = x] = X^{-1}(x) = \{\omega \in \Omega \mid X(\omega) = x\}$$

Les événements  $[X = x]$ ,  $x \in X(\Omega)$  forment un système complet d'événements.

**Corollaire 28.4.6 (Formule des probabilités totales associée à une v.a.r.d.)**

Soit  $X$  une variable aléatoire réelle discrète. Quitte à enlever quelques parts négligeables de sorte à se ramener à un système quasi-complet, on peut supposer que pour tout  $x \in X(\Omega)$ ,  $\mathbb{P}([X = x]) > 0$ . On a alors, pour tout événement  $B$  :

$$\mathbb{P}(B) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \mathbb{P}(B \mid X = x).$$

## IV.2 Formule des probabilités composées

La définition des probabilités conditionnelles permet d'exprimer la probabilité d'une intersection à l'aide de probabilités conditionnelles :

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

Cette formule est à comprendre de la façon suivante : en considérant les événements  $A$  et  $B$  comme successifs, pour obtenir  $A \cap B$ , il faut d'abord obtenir  $A$ , puis,  $A$  étant obtenu (ce qui donne la condition de la deuxième probabilité), il faut obtenir  $B$ . On pourrait bien sûr continuer ainsi : si  $C$  est un troisième événement, consécutif à  $A$  et  $B$ , une fois réalisé  $A \cap B$  (ce qui donne la condition de la troisième probabilité), il faut réaliser  $C$ . Ainsi :

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

De façon plus générale, on obtient :

**Théorème 28.4.7 (Formule des probabilités composées)**

Soit  $n \geq 2$ , et  $(A_1, \dots, A_n) \in \mathcal{T}^n$  tel que  $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) \neq 0$  (donc  $A_1 \cap \dots \cap A_{n-1}$  n'est pas presque-impossible). Alors :

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right) &= \mathbb{P}(A_1) \mathbb{P}(A_2 \mid A_1) \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \prod_{i=2}^n \mathbb{P}\left(A_i \mid \bigcap_{j=1}^{i-1} A_j\right). \end{aligned}$$

◁ Éléments de preuve.

Récurrence sur  $n$ .

▷

**Remarque 28.4.8**

L'intérêt de cette formule est de donner une formule pour le calcul des probabilités d'intersections d'événements, notamment dans le cas d'une succession d'expérience. Cette formule dévoile toute son utilité lorsque les événements considérés ne sont pas mutuellement indépendants. En cas d'indépendance mutuelle, elle ne dit rien de plus que la proposition 28.3.7.

### IV.3 Formules de Bayes

La troisième formule permet d'inverser des conditions. Sa version simple découle de façon directe de la définition d'une probabilité conditionnelle.

**Théorème 28.4.9 (Formule de Bayes simple)**

Soit  $A$  et  $B$  deux événements tels que  $\mathbb{P}(A) > 0$  et  $\mathbb{P}(B) > 0$ . Alors

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

◁ **Éléments de preuve.**

Calculer  $\mathbb{P}(A \cap B)$  en conditionnant de deux manières. ▷

De façon plus générale, la connaissance des probabilités d'un système complet, et des probabilités d'un événement  $B$  conditionnées au système complet permet de retourner une à une les probabilités conditionnelles :

**Théorème 28.4.10 (Formule de Bayes sur un système complet)**

Soit  $(A_i)_{i \in I}$  un système quasi-complet fini ou dénombrable, et  $B$  un événement non quasi-impossible, tel que pour tout  $i \in I$ ,  $\mathbb{P}(A_i) \neq 0$ . Soit  $j \in I$ . Alors :

$$\mathbb{P}(A_j | B) = \frac{\mathbb{P}(B | A_j)\mathbb{P}(A_j)}{\sum_{i \in I} \mathbb{P}(B | A_i)\mathbb{P}(A_i)}.$$

◁ **Éléments de preuve.**

Exprimer  $\mathbb{P}(B)$  avec la formule des probabilités totales dans la formule précédente. ▷

**Remarque 28.4.11 (Intérêt de la formule de Bayes)**

Le système  $(A_i)$  représente souvent une liste de causes pouvant amener l'événement  $B$  lors d'une étape suivante de l'expérience. Il est alors généralement facile de déterminer la probabilité qu'une certaine conséquence  $B$  ait lieu, sachant que la cause  $A_i$  a eu lieu, c'est-à-dire la probabilité conditionnelle  $\mathbb{P}(B | A_i)$  (on respecte dans ce cas l'ordre temporel). Ces données permettent, grâce à la formule de Bayes, de remonter le temps, en déterminant la probabilité qu'une certaine cause  $A_i$  ait eu lieu sachant la conséquence  $B$ . Pour cette raison, cette formule est aussi souvent appelée *formule de probabilité des causes*.

**Exemple 28.4.12**

Dans une population composée d'autant d'hommes que de femmes, 5 % des hommes et 0,25 % des femmes sont daltoniens. Quelle est la probabilité pour qu'une personne daltonienne choisie au hasard soit une femme ?

Voici d'autres exemples de situations de la vie courante dans lesquelles la formule de Bayes peut trouver des utilisations pratiques

**Exemple 28.4.13**

1. Diagnostics médicaux (retrouver la cause des symptômes)
2. Anti-SPAM

## V Principes généraux du calcul des probabilités

Nous exposons ici le principe général du calcul (et de la rédaction de ce calcul) d'une probabilité. Évidemment, dans la pratique, les situations sont très variées, et on peut être amené à s'écarter légèrement de la voie tracée ci-dessous.

### Méthode 28.5.1 (Comment aborder un calcul de probabilité)

1. Si on est dans une situation d'équiprobabilité, on peut se diriger vers un argument combinatoire (et utiliser la formule de Laplace). Dans les autres cas, la démarche générale est indiquée ci-dessous.
2. Définir des événements simples issus de l'expérience décrite, en n'oubliant pas la numérotation de ces événements en cas d'expérience répétée. Ces événements simples auront pour vocation de décrire ceux qui nous intéressent.  
Il ne faut pas hésiter à introduire des événements par vous-même, à donner des notations. Ce travail n'est pas nécessairement fait pour vous dans le sujet.
3. Décrire à l'aide d'une phrase l'événement dont on recherche la probabilité, sous forme une condition nécessaire et suffisante de réalisation :

« L'événement  $A$  est réalisé si et seulement si ... »

Cette condition nécessaire et suffisante de réalisation de l'événement  $A$  doit s'exprimer en fonction de la réalisation ou non des événements simples définis dans la première étape. Il s'agit donc de « décomposer » l'événement  $A$  en événements simples.

4. Traduire cette description dans le langage ensembliste (c'est-à-dire sous forme d'union, intersection ou complémentaires d'événements élémentaires). Les deux étapes sont nécessaires : la première (description verbale) fournit une explication facilement lisible, et éclaire la formule, la seconde (description ensembliste) donne de la rigueur à l'argument, et permet ensuite de s'orienter vers la bonne méthode de calcul.
5. Calcul de la probabilité  $\mathbb{P}(A)$ , en se servant de la décomposition ensembliste précédente, et des règles de calcul des probabilités d'unions, d'intersections ou de complémentaires. Nous rappelons l'essentiel de ces règles ci-dessous.

Ainsi, il est important de savoir s'orienter rapidement vers la bonne technique de calcul (probabilité d'une union, d'une intersection), suivant la situation rencontrée. Voici les différents cas rencontrés les plus fréquemment :

### Méthode 28.5.2 (Calcul de la probabilité d'un complémentaire)

Une seule formule à connaître, issue de la définition d'une mesure de probabilité :

$$\forall A \in \mathcal{T}, \quad \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A).$$

### Méthode 28.5.3 (Calcul de la probabilité d'une intersection)

Suivant les cas :

- Intersection dénombrable d'événements décroissants pour l'inclusion :  
On utilise le théorème de la limite monotone.
- Intersection d'un nombre fini d'événements mutuellement indépendants :  
On utilise la relation issue de la définition de l'indépendance mutuelle :

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2) \dots \mathbb{P}(A_n).$$

- Intersection d'un nombre fini d'événements non nécessairement indépendants :

On utilise la formule des probabilités composées.

Cette formule est surtout utile lorsqu'il y a un enchaînement temporel des événements ( $A_1$  étant le premier à advenir,  $A_n$  le dernier), chaque résultat influant sur les suivants. C'est le cas par exemple dans le cadre de tirages successifs dans une urne évolutive, l'évolution se faisant différemment suivant la boule tirée à un tirage donné.

Cette formule permet de formaliser le calcul de la probabilité associée à une branche de l'arbre des possibilités.

#### Méthode 28.5.4 (Calcul de la probabilité d'une union)

- Union dénombrable d'événements croissants pour l'inclusion :

On utilise ici aussi le théorème de la limite monotone.

- Union d'un nombre fini ou dénombrable d'événements 2 à 2 incompatibles :

On utilise l'additivité (cas fini) ou la  $\sigma$ -additivité (cas dénombrable) de la mesure de probabilité.

- Union d'un nombre fini d'événements mutuellement indépendants :

On a intérêt dans ce cas à considérer l'événement complémentaire. On est alors ramené au calcul de la probabilité d'une intersection d'événements mutuellement indépendants.

- Union d'un nombre fini d'événements lorsqu'on a des informations sur les intersections :

On utilise la formule du crible de Poincaré. Les cas  $n = 2$  et  $n = 3$  sont les plus utiles :

$$* n = 2 : \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$$

$$* n = 3 : \mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_2 \cap A_3) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$$

#### Méthode 28.5.5 (Cas d'une expérience dont l'issue dépend de résultats antérieurs)

Dans certaines situations, par exemple lorsque le mode opératoire d'une expérience (par exemple la composition d'une urne) dépend du résultat d'une expérience précédente, il peut être nécessaire de discuter suivant le résultat obtenu lors de l'expérience antérieure : autrement dit, dans ces situations, il est aisé de calculer des probabilités conditionnellement à chacun des résultats possibles de la première expérience. Il faut ensuite récupérer la probabilité globale à partir de toutes ces probabilités conditionnelles. On utilise pour cela la formule des probabilités totales.

A retenir : discussion à faire  $\longrightarrow$  formule des probabilités totales

Enfin, vu qu'elles interviennent dans de nombreuses formules, nous faisons quelques remarques concernant le calcul des probabilités conditionnelles.

#### Méthode 28.5.6 (Calcul de probabilités conditionnelles)

Puisque la probabilité conditionnelle  $\mathbb{P}_B$  sachant un événement  $B$  donné est une mesure de probabilité, tous les résultats, et toutes les règles vues ci-dessus pour le calcul des probabilités s'appliquent au calcul des probabilités conditionnelles. Lorsque ces formules font elles-même intervenir des probabilités conditionnelles, les conditions s'ajoutent à celle déjà présente (sous forme d'une intersection) :

$$(\mathbb{P}_B)_C = \mathbb{P}_{B \cap C}.$$

À titre d'exemple, voici la formule des probabilités totales pour une probabilité conditionnelle :  $(A_i)_{i \in I}$  étant un système complet au plus dénombrable tel que pour tout  $i \in I$   $\mathbb{P}_C(A_i) \neq 0$ , on a :

$$\mathbb{P}(B | C) = \sum_{i \in I} \mathbb{P}(A_i | C) \mathbb{P}(B | C \cap A_i).$$

**Avertissement 28.5.7**

Attention, on ne peut pas définir de notion d'événement conditionnel. Ce n'est pas l'événement qui est conditionnel, mais sa réalisation. Ainsi, si vous adoptez la démarche générale de rédaction :

- ne décrivez pas l'événement conditionnel (cela n'a pas de sens), mais donnez bien une condition nécessaire et suffisante de réalisation, conditionnellement à un certain événement :  
« Sachant que  $B$  est réalisé,  $A$  est réalisé si et seulement si ... »
- Gardez-vous bien de transcrire cette phrase de manière ensembliste ; cela vous amènerait inévitablement à considérer des « événements conditionnels ». Sautez cette étape et passez directement aux probabilités. Vous pouvez transcrire votre phrase de façon ensembliste à l'intérieur des probabilités, la condition étant alors précisée non pas au niveau des ensembles, mais au niveau de la probabilité.

**Avertissement 28.5.8**

L'indépendance est une notion dépendant de la mesure de probabilité. Ce n'est pas parce que  $A$  et  $B$  sont indépendants (lorsqu'on dit cela, on sous-entend l'indépendance pour la mesure de probabilité totale  $\mathbb{P}$ ) qu'ils sont indépendants pour une mesure de probabilité conditionnelle  $\mathbb{P}_C$ .

**Exemple 28.5.9**

On effectue 3 tirages à Pile ou Face indépendants. L'événement  $C$  est réalisé si et seulement si on obtient exactement 1 Pile lors des tirages 1 et 2,  $A$  de même avec les tirages 2 et 3. et  $B$  de même avec les tirages 1 et 3.

Les événements  $A$  et  $B$  sont indépendants pour  $\mathbb{P}$ , mais pas pour  $\mathbb{P}_C$ .



# Variables aléatoires

*Alea jacta est.*

(Jules, un vieux pote d'Asterix)

Dans ce chapitre, sauf mention contraire,  $(\Omega, \mathcal{T}, \mathbb{P})$  désigne un espace probabilisé quelconque.

Le programme se limite au cas où  $\Omega$  est fini. Nous aborderons le cas un peu plus général d'une variable aléatoire discrète. La difficulté supplémentaire dans ce cadre réside souvent dans des problèmes de convergence. Ces problèmes de convergence n'ont pas lieu dans le cadre d'un univers fini, puisque toutes les sommes considérées sont alors finies. Nous rappelons pour cela le résultat fondamental suivant :

*Si  $\sum a_n$  est absolument convergente alors ni la convergence ni la valeur de la somme ne dépendent de l'ordre de sommation, ce qui permet de considérer dans ce cas une somme sur des ensembles  $I$  dénombrables non ordonnés (il suffit d'avoir convergence absolue pour une énumération possible des éléments de  $I$  pour que la somme sur  $I$  ait un sens bien défini). On dira dans ce cas que la famille  $(a_i)_{i \in I}$  est sommable.*

Nous énonçons les résultats dans le contexte général, mais nous nous contenterons de donner les preuves dans le cas fini, conformément au programme, afin d'éviter les justifications un peu douloureuses de convergence et de commutation de sommes. Les justifications précises de ces points seront vues en Spé.

La situation la plus fréquente est celle d'une variable aléatoire à valeurs réelles (variable aléatoire réelle), ou à valeurs dans  $\mathbb{R}^n$  (on parle de vecteur aléatoire, c'est équivalent à la donnée de  $n$  variables aléatoires réelles) mais on peut définir une variable aléatoire à valeurs dans n'importe quel ensemble muni d'une tribu.

## I Variables aléatoires

### I.1 Applications mesurables

#### Définition 29.1.1 (application mesurable)

Soit  $(E, \mathcal{S})$  et  $(F, \mathcal{T})$  deux espaces munis d'une tribu (on parlera d'espaces mesurables). Soit  $f : E \rightarrow F$ . On dit que  $f$  est mesurable si pour tout  $B \in \mathcal{T}$ ,  $f^{-1}(B) \in \mathcal{S}$ .

#### Proposition 29.1.2 (Composition)

Soit  $(E, \mathcal{S})$ ,  $(F, \mathcal{T})$  et  $(G, \mathcal{U})$  trois espaces mesurables, et  $f : E \rightarrow F$  et  $g : F \rightarrow G$  deux applications mesurables. La composée  $g \circ f$  est encore mesurable.

**Proposition 29.1.3 (Caractérisation des applications mesurables)**

Avec les notations précédentes, soit  $f : E \rightarrow F$ , et  $\mathcal{C}$  tel que  $\mathcal{T} = \sigma(\mathcal{C})$ . Alors  $f$  est mesurable si et seulement si pour tout  $B \in \mathcal{C}$ ,  $f^{-1}(B) \in \mathcal{S}$ .

◁ **Éléments de preuve.**

Montrer que l'ensemble des  $B$  de  $\mathcal{T}$  tels que  $f^{-1}(B) \in \mathcal{S}$  est une tribu contenant  $\mathcal{C}$ . ▷

On rappelle que la tribu  $\mathcal{B}^n$  des boréliens sur  $\mathbb{R}^n$  est engendrée par les ensembles  $I_1 \times I_2 \times \dots \times I_n$ , où les  $I_k$  sont des intervalles ouverts  $]a_k, b_k[$  (pavés ouverts, ou ouverts élémentaires). Un autre système de générateurs est par exemple l'ensemble des pavés fermés (produit cartésien d'intervalles fermés bornés).

**Lemme 29.1.4 (des éléments de  $\mathcal{B}^n$ )**

Les produits  $I_1 \times \dots \times I_n$ , où les  $I_k$  sont des intervalles quelconques, sont des éléments de  $\mathcal{B}^n$ .

◁ **Éléments de preuve.**

On l'avait déjà évoqué : récupérer des bornes ouvertes par intersection dénombrable, des bornes infinies par union dénombrable. ▷

**Corollaire 29.1.5**

Soit, pour  $k \in \llbracket 1, n \rrbracket$ ,  $\pi_k : \mathbb{R}^n \rightarrow \mathbb{R}$  la projection sur le  $k$ -ième facteur. Alors  $\pi_k$  est mesurable.

**Lemme 29.1.6 (D'autres systèmes générateurs de  $\mathcal{B}^1$ )**

La tribu  $\mathcal{B}^1$  sur  $\mathbb{R}$  est aussi engendrée par les intervalles  $] -\infty, a ]$ .

◁ **Éléments de preuve.**

Retrouver les  $[a, b]$  par intersection, union. ▷

On montrerait sur le même principe de  $\mathcal{B}^1$  est engendrée par les intervalles  $] -\infty, a [$ , ou par les intervalles  $[a, +\infty [$ , ou par les intervalles  $] a, +\infty [$ , ou encore par les intervalles  $] a, b [$ , ou encore  $[a, b [$ , ou  $] a, b ]$ . Ces différents points sont laissés en exercice.

**Corollaire 29.1.7 (Caractérisation par les coordonnées)**

Soit  $(E, \mathcal{T})$  un espace mesurable, et  $f : E \rightarrow \mathbb{R}^n$  une application. On note  $f_1, \dots, f_n$  les coordonnées de  $E$ . Ainsi, pour tout  $x \in E$ ,  $f(x) = (f_1(x), \dots, f_n(x))$ . L'ensemble  $\mathbb{R}^n$  étant muni de la tribu  $\mathcal{B}^n$  et  $\mathbb{R}$  de la tribu  $\mathcal{B}^1$ , l'application  $f$  est mesurable si et seulement si  $f_1, \dots, f_n$  le sont.

◁ **Éléments de preuve.**

- Sens direct : composer par la projection qui est mesurable.
- Sens réciproque : L'image réciproque d'un pavé par  $f$  est une intersection d'images réciproques par les  $f_i$ .

▷

Une autre propriété importante, issue de la mesurabilité de la projection est :

**Théorème 29.1.8 (Produit cartésien de boréliens)**

Si  $A_1, \dots, A_n$  sont des éléments de  $\mathcal{B}_1$ , alors  $A_1 \times \dots \times A_n$  est un élément de  $\mathcal{B}_n$ . Autrement dit, le produit cartésien de boréliens est un borélien.

◁ **Éléments de preuve.**

Remarquer que  $A_1 \times \cdots \times A_n = \bigcap_{k=1}^n p_k^{-1}(A_k)$ .

▷

## I.2 Variables aléatoires

Soit  $(\Omega, \mathcal{T}, \mathbb{P})$  un espace probabilisé,  $E$  un ensemble et  $\mathcal{T}'$  une tribu sur  $E$ .

### Définition 29.1.9 (Variable aléatoire)

- Une variable aléatoire à valeurs dans  $E$  est une application mesurable  $X : \omega \mapsto X(\omega)$  de  $\Omega$  dans  $(E, \mathcal{T}')$ .
- Une variable aléatoire réelle (v.a.r.) est une variable aléatoire à valeurs dans  $(\mathbb{R}, \mathcal{B}^1)$
- Un vecteur aléatoire réel  $X$  est une variable aléatoire à valeurs dans  $(\mathbb{R}^n, \mathcal{B}^n)$ . D'après ce qui précède, la donnée d'un vecteur aléatoire équivaut à la donnée de  $n$  variables aléatoires réelles  $X_1, \dots, X_n$ .

Comme nous l'avons vu dans le paragraphe précédent, la mesurabilité peut se justifier en se restreignant à l'étude des images réciproques d'un système générateur de  $\mathcal{T}'$ . En appliquant ce résultat à la tribu  $\mathcal{B}^1$  des boréliens, on obtient :

### Proposition 29.1.10 (Caractérisation des variables aléatoires réelles)

Soit  $X : (\Omega, \mathcal{T}) \rightarrow \mathbb{R}$  une application.  $X$  est une variable aléatoire si et seulement si pour tout  $a \in \mathbb{R}$ ,  $X^{-1}(]-\infty, a]) \in \mathcal{T}$ .

On admet parfois la définition suivante, un peu plus large, et utile dans certaines expériences infinies :

### Définition 29.1.11 (Variable aléatoire, notion étendue)

Une *variable aléatoire*  $X$  sur  $\Omega$  à valeurs dans  $E$  est une application mesurable définie sur  $\Omega' \in \mathcal{T}$  telle que  $\mathbb{P}(\Omega') = 1$ , donc définie presque partout.

### Exemple 29.1.12

On lance une infinité de fois une pièce, et  $X$  correspond au rang de lancer du premier Pile obtenu. En adoptant la convention précédente,  $X$  est une v.a.r., prenant les valeurs  $X(\Omega) = \mathbb{N}^*$ . Remarquez que l'événement « n'obtenir que des faces », qui est quasi-impossible, mais pas impossible, n'admet pas d'image par  $X$ .

### Remarque 29.1.13

Cette définition étendue se ramène à la définition générale :

- soit en considérant  $X$  comme une variable aléatoire sur  $\Omega'$ , muni de la tribu  $\{A \cap \Omega', A \in \mathcal{T}\} = \{A \in \mathcal{T} \mid A \subset \Omega'\}$ ,
- soit en considérant  $X$  comme une variable aléatoire à valeurs dans  $(\overline{\mathbb{R}}, \overline{\mathcal{B}^1})$ , où  $\overline{\mathcal{B}^1}$  est la tribu  $\sigma(\mathcal{B}^1, \{-\infty\}, \{+\infty\})$ , et où les éléments de  $\overline{\Omega'}$  sont envoyés par  $X$  sur  $+\infty$ .

### Notation 29.1.14 (Événements liés à une variable aléatoire)

Soit  $X$  une variable aléatoire à valeurs dans  $(E, \mathcal{T}')$ .

- Soit  $A \in \mathcal{T}'$  (en particulier  $A \subset E$ ). L'événement  $X^{-1}(A) \in \mathcal{T}$  est noté  $[X \in A]$  ou  $\{X \in A\}$  ou encore  $(X \in A)$

- Soit  $x \in A$ . Si  $\{x\} \in \mathcal{T}'$ , l'événement  $X^{-1}(\{x\}) = [X \in \{x\}]$  sera simplement noté  $[X = x]$ , ou  $(X = x)$ , ou  $\{X = x\}$ .

**Remarque 29.1.15**

Si  $\mathcal{T} = \mathcal{P}(\Omega)$  (cadre du programme), on peut toujours considérer sur  $E$  la tribu  $\mathcal{P}(E)$ . Dans ce cas, l'événement  $[X \in A]$  est bien défini pour tout  $A \subset E$ .

Ainsi, l'ensemble des événements qu'on peut décrire à l'aide de  $X$  sont tous les événements  $[X \in A]$ , c'est-à-dire  $X^{-1}(A)$ , pour  $A \in \mathcal{T}'$ . Cet ensemble admet naturellement une structure de tribu :

**Proposition/Définition 29.1.16 (Tribu associée à une v.a.)**

Soit  $X : (\Omega, \mathcal{T}) \rightarrow (E, \mathcal{T}')$  une variable aléatoire. L'ensemble  $\{X^{-1}(A), A \in \mathcal{T}'\} \subset \mathcal{T}$  est une tribu, notée  $\mathcal{T}_X$ , appelée tribu associée à la variable  $X$ .

Intuitivement,  $\mathcal{T}_X$  est la plus petite sous-tribu de  $\mathcal{T}$  rendant  $X$  mesurable.

Dans le cas d'une variable à valeurs réels, on note de la façon suivante les événements associés aux images réciproques des intervalles :

**Notation 29.1.17 (Événements liés à une v.a.r.)**

Dans le cas d'une variable aléatoire réelle,  $E = \mathbb{R}$ , muni de la tribu des boréliens. On utilise les notations suivantes pour les images réciproques des différents intervalles possibles (les intervalles appartiennent tous à la tribu des boréliens) :

- L'événement  $\{\omega \mid X(\omega) < a\}$  est noté  $[X < a]$  ou  $(X < a)$  ou... ;
- L'événement  $\{\omega \mid X(\omega) \leq a\}$  est noté  $[X \leq a]$  ou... ;
- L'événement  $\{\omega \mid a < X(\omega) < b\}$  est noté  $[a < X < b]$ ,
- L'événement  $\{\omega \mid a < X(\omega) \leq b\}$  est noté  $[a < X \leq b]$ ,
- etc.

Par ailleurs, les singletons étant dans la tribu des boréliens, on peut définir pour tout  $x \in \mathbb{R}$  l'événement  $[X = x] = X^{-1}(\{x\})$ , qui peut éventuellement être vide.

**Proposition 29.1.18 (Système complet associé à une variable aléatoire)**

Soit  $X$  une variable (ou un vecteur) aléatoire réelle discrète sur  $(\Omega, \mathcal{T})$ . Alors  $\{[X = x], x \in X(\Omega)\}$  est un système complet d'événements.

◁ **Éléments de preuve.**

C'est la partition associée à une surjection. Le seul point à justifier est que ce sont bien des éléments de la tribu. Cela provient du fait que les singletons sont des boréliens. ▷

De façon plus générale, ceci reste valable dès lors que  $X$  est à valeurs dans  $(E, \mathcal{T}')$  et que pour tout  $x \in X(\Omega)$ ,  $\{x\} \in \mathcal{T}'$ .

**Remarque 29.1.19**

En utilisant la définition étendue d'une variable aléatoire, il ne s'agit que d'un système quasi-complet, mais cela permet tout de même d'utiliser les formules classiques.

### I.3 Loi d'une variable aléatoire

**Définition 29.1.20 (Loi d'une variable aléatoire)**

Soit  $X$  une variable aléatoire (quelconque) sur  $(\Omega, \mathcal{T}, \mathbb{P})$ , à valeurs dans  $(E, \mathcal{T}')$ . La loi de probabilité de  $X$  est la fonction  $\mathbb{P}_X$  définie de  $\mathcal{T}'$  dans  $[0, 1]$  par :

$$\forall A \in \mathcal{T}', \quad \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)).$$

**Proposition 29.1.21 (La mesure  $\mathbb{P}_X$ )**

La fonction  $\mathbb{P}_X$  est une mesure de probabilité sur l'espace probablisable  $(E, \mathcal{T}')$ .

## ◁ Éléments de preuve.

Vérification simple, en se servant des propriétés de l'image réciproque. ▷

Dans le cas d'une variable réelle (donc à valeurs dans  $\mathbb{R}$  muni de la tribu borélienne), il n'est pas nécessaire de déterminer  $\mathbb{P}_X$  sur tous les éléments de  $\mathcal{T}$ . Des techniques fines sur les espaces mesurables permettent en effet de montrer qu'une mesure de probabilité est entièrement déterminée par ses valeurs sur un  $\pi$ -système  $\mathcal{C}$  engendrant la tribu  $\mathcal{T}'$ , un  $\pi$ -système étant un sous-ensemble de  $\mathcal{T}'$  stable par intersections finies. Ce résultat est une conséquence du « lemme de classe monotone ». On en déduit :

**Proposition 29.1.22 (Détermination de la loi d'une v.a.r., admis)**

Soit  $X$  une variable aléatoire réelle. Alors la loi  $\mathbb{P}_X$  de  $E$  est entièrement déterminée par les probabilités  $\mathbb{P}(X \leq x)$ ,  $x \in \mathbb{R}$ .

**Définition 29.1.23 (Fonction de répartition)**

Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a.r.. La fonction de répartition de  $X$  est la fonction, notée généralement  $F_X$ , définie pour tout  $x \in \mathbb{R}$  par  $F_X(x) = \mathbb{P}(X \leq x)$ .

Ainsi, d'après ce qui précède, la fonction de répartition détermine entièrement la loi  $\mathbb{P}_X$ .

**Théorème 29.1.24 (Propriétés caractéristiques des fonctions de répartition)**

Soit  $X$  une v.a.r., et  $F_X$  sa fonction de répartition. Alors :

1.  $F_X$  est une fonction croissante sur  $\mathbb{R}$  ;
2.  $F_X$  est continue à droite en tout point de  $\mathbb{R}$  ;
3.  $F_X$  admet des limites en  $\pm\infty$ , et  $\lim_{x \rightarrow -\infty} F_X(x) = 0$ ,  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

## ◁ Éléments de preuve.

1. C'est la croissance de la probabilité (pour l'inclusion)
2. La croissance donne déjà l'existence de la limite à droite. Le théorème de limite monotone montre que  $F_X(a) = \lim_{n \rightarrow +\infty} F_X\left(a + \frac{1}{n}\right)$ .
3. Considérer une intersection d'ensembles de la forme  $] -\infty, a]$  donnant  $\emptyset$ , une union donnant  $\mathbb{R}$ .

▷

Réciproquement, on peut montrer (on l'admettra) que ces propriétés caractérisent les fonctions de répartition, c'est-à-dire que toute fonction vérifiant cela peut être vue comme la fonction de répartition d'une certaine variable aléatoire réelle.

**Proposition 29.1.25 (Expression de  $\mathbb{P}(X \in I)$  à l'aide de la fonction de répartition)**

Soit  $X$  une v.a.r., et  $a \leq b$  dans  $\overline{\mathbb{R}}$ . On a :

- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
- $\mathbb{P}(a < X < b) = \lim_{y \rightarrow b^-} F_X(y) - F_X(a)$ ,  $a < b$
- $\mathbb{P}(a \leq X \leq b) = F_X(b) - \lim_{x \rightarrow a^-} F_X(x)$
- $\mathbb{P}(a \leq X < b) = \lim_{y \rightarrow b^-} F_X(y) - \lim_{x \rightarrow a^-} F_X(x)$ .

◁ **Éléments de preuve.**

Même principe en décrivant tous ces ensembles comme unions, intersections, complémentations d'ensembles de la forme  $] - \infty, c]$ . ▷

En particulier, on a, pour tout  $x \in \mathbb{R}$  :

$$\mathbb{P}(X = x) = F(x) - \lim_{y \rightarrow x^-} F(y).$$

Les probabilités ponctuelles correspondent donc aux sauts de discontinuité de  $F_X$ .

**I.4 Variables aléatoires discrètes**

Les définitions diffèrent un peu suivant les contextes et les besoins. La définition suivante est la moins contraignante, et est amplement suffisante pour toute la théorie.

**Définition 29.1.26 (Variable aléatoire discrète)**

- Une *variable aléatoire discrète*  $X$  est une variable aléatoire telle que  $X(\Omega)$  soit un ensemble fini ou dénombrable.
- Si  $X$  est une variable aléatoire réelle, on abrègera « variable aléatoire réelle discrète » en v.a.r.d.

**Remarque 29.1.27**

Soit  $X$  une variable aléatoire réelle discrète à valeurs dans  $E$ . Quitte à corestreindre  $X$  à son image, on peut donc supposer que  $E$  est au plus dénombrable. Il peut alors être muni de la tribu  $\mathcal{P}(E)$ , qui raffine la tribu initiale sur  $E$ . **On supposera désormais cette condition toujours vérifiée.** Ainsi, une v.a.r.d. sera toujours à valeurs dans un espace mesurable du type  $(E, \mathcal{P}(E))$ . On ne précisera donc plus la tribu sur  $E$ .

**Proposition 29.1.28 (Composition d'une variable aléatoire discrète)**

Soit  $X$  une variable aléatoire discrète sur  $(\Omega, \mathcal{T}, \mathbb{P})$ , à valeurs dans  $E$ , et  $f$  une application de  $X(\Omega)$  dans un ensemble  $F$ . Alors  $f \circ X$  est une variable aléatoire discrète, notée  $f(X)$ .

◁ **Éléments de preuve.**

C'est une variable aléatoire car  $f$  est mesurable, puisque  $E$  est muni de la tribu  $\mathcal{P}(E)$  en vertu de la remarque précédente.

De plus  $f : X(\Omega) \rightarrow f(X)(\Omega)$  est une surjection, et  $X(\Omega)$  est au plus dénombrable. ▷

On en déduit :

**Proposition 29.1.29 (Algèbre des v.a.r.d.)**

Le sous-ensemble de  $\mathbb{R}^\Omega$  constitué des variables aléatoires réelles discrètes est une sous-algèbre de  $\mathbb{R}^\Omega$ . Ainsi, si  $X$  et  $Y$  deux v.a.r.d. sur  $\Omega$ ,  $\lambda$  un réel quelconque :

1.  $X + Y$  est une v.a.r.d.
2.  $\lambda X$  est une v.a.r.d.
3.  $XY$  est une v.a.r.d.

**Remarque 29.1.30**

Plus généralement, la somme, le produit, mais aussi l'infimum ou le supremum d'une famille finie de v.a.r.d. est encore une v.a.r.d.

**I.5 Loi de probabilité d'une variable aléatoire discrète**

Soit  $X$  une variable aléatoire discrète sur  $(\Omega, \mathcal{T}, \mathbb{P})$ , à valeurs dans  $E$ . Conformément à la remarque 29.1.27, on considère que la tribu sur  $E$  est  $\mathcal{P}(E)$ , donc contient tous les singletons. Ainsi, les ensembles  $\{X = x\}$  sont des éléments de la tribu  $\mathcal{T}$ , et  $\mathbb{P}(X = x)$  est donc bien défini.

**Proposition 29.1.31**

Soit  $X$  une variable aléatoire discrète sur  $(\Omega, \mathcal{T}, \mathbb{P})$  à valeurs dans  $E$ . Alors la famille  $(\mathbb{P}(X = x))_{x \in X(\Omega)}$  est une distribution de probabilités.

◁ **Éléments de preuve.**

C'est la  $\sigma$ -additivité utilisée sur le SCE précédent. ▷

La sommabilité (issuée de la convergence absolue, ici automatique en cas de convergence, la série étant à termes positifs) assure que cette somme est bien définie, indépendamment de l'ordre de sommation.

**Proposition 29.1.32 (Détermination de la loi d'une v.a. discrète)**

La loi d'une v.a. discrète est entièrement déterminée par la distribution de probabilités  $(\mathbb{P}(X = x))_{x \in X(\Omega)}$  pour  $x \in X(\Omega)$ . Plus précisément, pour tout  $A \subset E$ ,

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

cette somme étant bien définie.

◁ **Éléments de preuve.**

Écrire  $A$  comme union de singletons. ▷

**Notation 29.1.33**

On note  $X \sim Y$  pour désigner le fait que  $X$  et  $Y$  suivent la même loi.

**Exemple 29.1.34**

Si  $X = Y$  presque sûrement (i.e. il existe  $\Omega' \subset \Omega$  tel que  $\mathbb{P}(\Omega') = 1$  et pour tout  $\omega \in \Omega'$ ,  $X(\omega) = Y(\omega)$ ), alors  $X \sim Y$ .

**I.6 Loi de  $f(X)$**

**Proposition 29.1.35 (Loi de  $f(X)$ )**

1. Soit  $X$  une variable aléatoire discrète sur  $(\Omega, \mathcal{T}, \mathbb{P})$ , à valeurs dans  $E$ . La loi de  $f(X)$  est donnée par

$$\mathbb{P}_{f(X)} = \mathbb{P}_X \circ \widehat{f^{-1}}.$$

Autrement dit, pour tout  $A \in \mathcal{P}(F)$ ,

$$\mathbb{P}(f(X) \in A) = \mathbb{P}_{f(X)}(A) = \mathbb{P}_X(f^{-1}(A)) = \mathbb{P}(X \in f^{-1}(A)).$$

2.  $f(X)$  étant discrète, sa loi est déterminée par les probabilités ponctuelles :

$$\forall x \in f(X(\Omega)), \quad \mathbb{P}(f(X) = x) = \sum_{\substack{y|f(y)=x \\ y \in X(\Omega)}} \mathbb{P}(X = y).$$

◁ **Éléments de preuve.**

1.  $[X \in f^{-1}(A)] = [f(X) \in A]$ .
2.  $[f(X) = x] = [X \in f^{-1}(x)]$ .

▷

**Exemples 29.1.36 (Loi d'une somme, loi du minimum)**

Voici deux exemples particulièrement importants, à bien connaître, dans le cas où  $X$  et  $Y$  sont des variables aléatoires réelles discrètes.

1.  $\forall z \in \mathbb{R}, \quad \mathbb{P}(X + Y = z) = \sum_{\substack{(x,y) \in X(\Omega) \times Y(\Omega) \\ x+y=z}} \mathbb{P}(X = x, Y = y).$
2.  $\forall z \in X(\Omega) \cup Y(\Omega), \quad \mathbb{P}(\min(X, Y) = z) = \sum_{\substack{y \in Y(\Omega) \\ y \geq z}} \mathbb{P}(X = z, Y = y) + \sum_{\substack{x \in X(\Omega) \\ x > z}} \mathbb{P}(X = x, Y = z).$

Exemple : tirage de deux dés.

**I.7 Loi d'un vecteur aléatoire**

Soit  $(X_1, \dots, X_n)$  un vecteur aléatoire, à valeurs dans  $(\mathbb{R}^n, \mathcal{B}^n)$ , ou plus généralement, dans un produit cartésien d'espaces mesurés. Pour commencer, introduisons une notation simplifiée pour désigner les événements associés au vecteur  $(X_1, \dots, X_n)$  (c'est-à-dire les éléments de  $\mathcal{T}_{(X_1, \dots, X_n)}$ ).

**Notation 29.1.37 ( $\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n)$ )**

Pour alléger les notations, on note  $\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n)$  au lieu de  $\mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n])$ . En particulier, on utilisera la notation  $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$ .

**Définition 29.1.38 (Loi conjointe, lois marginales)**

Soit  $V = (X_1, \dots, X_n)$  un vecteur aléatoire, à valeurs dans  $\mathbb{R}^n$ .

- On appelle loi conjointe de  $V$  la loi de  $(X_1, \dots, X_n)$ . Il s'agit donc d'une mesure de probabilité  $\mathbb{P}_{(X_1, \dots, X_n)}$  définie sur l'espace probabilisable  $(\mathbb{R}^n, \mathcal{B}^n)$
- On appelle  $k$ -ième loi marginale la loi de  $X_k$ .

De façon évidente :

**Proposition 29.1.39 (Expression d'une loi marginale)**

Soit  $A$  un borélien de  $\mathbb{R}$

1. L'événement  $(X_k \in A)$  est égal à  $[X_1 \in \mathbb{R}, \dots, X_k \in A, \dots, X_n \in \mathbb{R}]$ .
2. La  $k$ -ième loi marginale s'obtient de la loi conjointe par l'expression :

$$\mathbb{P}(X_k \in A) = \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_k \in A, \dots, X_n \in \mathbb{R}) = \mathbb{P}_V(\mathbb{R} \times \dots \times A \times \dots \times \mathbb{R}).$$

**Remarque 29.1.40**

Ce résultat reste vrai dans un cadre plus général où  $V$  est à valeurs dans un produit cartésien  $E_1 \times \dots \times E_n$ , en remplaçant  $X_i \in \mathbb{R}$  par  $X_i \in E_i$ , et  $A \in \mathbb{R}$  par  $A \in \mathcal{T}_i$ , où  $\mathcal{T}_i$  est la tribu sur  $E_i$ .

La propriété 30.1.41 reste valide dans ce contexte : la donnée de probabilités  $\mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$  définit une loi d'un vecteur aléatoire si et seulement si ces probabilités sont toutes positives, de somme égale à 1. On trouve la probabilité d'un événement associé à  $(X_1, \dots, X_n)$  en sommant les probabilités sur tous les  $n$ -uplets  $(x_1, \dots, x_n)$  constituant l'événement. En particulier :

**Proposition 29.1.41 (Expression des lois marginales par la loi conjointe)**

Les lois marginales se déduisent de la loi conjointe : pour tout  $x_k \in X_k(\Omega)$  :

$$\mathbb{P}(X_k = x_k) = \sum_{(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) \in X_1(\Omega) \times \dots \times X_{k-1}(\Omega) \times X_{k+1}(\Omega) \times \dots \times X_n(\Omega)} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$$

◁ **Éléments de preuve.**

Cette somme est d'après 30.1.46 égale à  $\mathbb{P}(V \in \mathbb{R} \times \dots \times \mathbb{R} \times \{x_k\} \times \mathbb{R} \times \dots \times \mathbb{R})$ . ▷

Par exemple, pour un couple  $(X, Y)$ ,

$$\mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y).$$

**Avertissement 29.1.42**

Les lois marginales sont déterminées par la loi conjointe, mais la réciproque est fautive : les lois marginales ne déterminent pas la loi conjointe !

**Exemple 29.1.43**

Soit une urne contenant 1 boule blanche, 1 boule noire. On effectue un tirage, avec équiprobabilité. On note  $X_1, Y_1$  et  $X_2$  les trois variables (égales l'une à l'autre) égales à 1 si on tire la boule noire, et 0 sinon. On note  $Y_2$  la variable égale à 1 si on tire la boule blanche, et 0 sinon. Alors, les lois marginales de  $(X_1, Y_1)$  et de  $(X_2, Y_2)$  sont les mêmes, pourtant les lois conjointes sont distinctes.

La loi conjointe de  $(X, Y)$  est souvent plus facile à déterminer que les lois de  $X$  et  $Y$ . On se sert donc du calcul de cette loi conjointe pour déterminer ensuite les lois de  $X$  et  $Y$ .

**Exemple 29.1.44**

On lance deux dés équilibrés, on note  $X$  le maximum, et  $Y$  le minimum. Déterminer la loi conjointe de  $(X, Y)$ , puis les lois marginales.

## II Espérance mathématique

Dans le cas de variables aléatoires réelles, ou plus généralement de variables à valeurs dans une  $\mathbb{R}$ -algèbre, on peut considérer des sommes, produits et moyennes de valeurs de  $X$ . Ainsi, on peut envisager de définir une valeur moyenne prise par  $X$ . On peut par exemple calculer une moyenne empirique, en répétant un grand nombre de fois l'expérience associée à  $X$ , et en faisant la moyenne sur ces expériences des valeurs obtenues pour  $X$ . On peut aussi définir une moyenne théorique (c'est l'espérance). Des propriétés de convergence (qui ne sont pas au programme de cette année) permettent de justifier que la moyenne empirique tend vers la moyenne théorique (l'espérance) lorsque le nombre d'expériences tend vers  $+\infty$ . C'est la loi faible des grands nombres.

Nous étudions dans ce paragraphe ces notions de moyennes, et les mesures d'écart associés.

### II.1 Espérance d'une variable aléatoire réelle discrète

L'espérance mathématique correspond à la moyenne théorique.

Par exemple, si on dispose d'un dé à 6 faces, dont 5 faces sont notées 1 et une face est numérotée 2, le 2 sortira en moyenne 1 fois sur 6. et le 1 apparaîtra 5 fois sur 6. Ainsi, sur 6 tirage, on peut « espérer » avoir en moyenne 5 fois 1 et 1 fois 2. La moyenne de ces 6 tirages est alors

$$\frac{1}{6}(5 \times 1 + 1 \times 2) = 1 \times \mathbb{P}(X = 1) + 2 \times \mathbb{P}(X = 2),$$

où  $X$  désigne le résultat d'un tirage.

Cela motive la définition suivante de la moyenne théorique, appelée plus généralement espérance mathématique.

#### Définition 29.2.1 (espérance mathématique)

Soit  $X$  une v.a.r.d.. On dit que  $X$  admet une espérance si la famille  $(x\mathbb{P}(X = x))_{x \in X(\Omega)}$  est sommable, et dans ce cas :

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x).$$

La notation  $X \in L^1$  signifie que  $X$  admet une espérance finie.

#### Remarque 29.2.2

L'espérance peut ne pas exister :

- série divergente :  $X(\Omega) = \mathbb{N}^*$  et  $\forall n \in \mathbb{N}^*, \mathbb{P}(X = n) = \frac{6}{\pi^2 n^2}$ .
- série semi-convergente :  $X(\Omega) = \{(-1)^n n, n \in \mathbb{N}^*\}$ , et :  $\forall n \in \mathbb{N}^*, \mathbb{P}(X = (-1)^n n) = \frac{6}{\pi^2 n^2}$ .

En effet, la semi-convergence n'est pas suffisante pour définir correctement l'espérance, car,  $X(\Omega)$  n'étant pas muni d'un ordre naturel, il faut s'assurer de l'indépendance du résultat par rapport à l'ordre de sommation.

Dans le premier cas cependant, on peut donner un sens à l'espérance dans  $\overline{\mathbb{R}}$ , la série définissant l'espérance étant alors de somme infinie, quel que soit l'ordre de sommation choisi.

L'espérance de  $X$  ne dépend que de la loi de  $X$ . Ainsi, si  $X \sim Y$ , alors  $\mathbb{E}(X) = \mathbb{E}(Y)$ . En particulier, c'est le cas si  $X = Y$  presque sûrement.

#### Définition 29.2.3 (Espérance infinie)

Soit  $X$  une variable aléatoire **positive**, telle que  $\sum_{x \in X(\omega)} x\mathbb{P}(X = x)$  diverge. On pourra alors dire que  $\mathbb{E}(X) = +\infty$

Ainsi, lorsque  $X$  est positive, l'espérance existe toujours dans  $\overline{\mathbb{R}}$ , et dire qu'elle existe dans  $\mathbb{R}$  équivaut à dire que  $\mathbb{E}(X) < +\infty$ .

**Remarque 29.2.4**  
 Dans le cadre du programme de Sup, seul le cas de variables aléatoires finies (i.e.  $X(\Omega)$  fini) est à considérer. Dans ce cas, la somme définissant l'espérance est finie, et il n'y a pas de convergence à justifier. Nous résumons cela dans la proposition suivante.

**Proposition 29.2.5 (Espérance d'une variable finie)**  
*Toute variable finie admet une espérance.*

◁ **Éléments de preuve.**

La somme définissant  $\mathbb{E}(X)$  est alors finie! ▷

Plus généralement :

**Proposition 29.2.6 (Espérance d'une variable bornée ps)**  
*Soit  $X$  une v.a.r.d. bornée ps. Alors  $X \in L^1$ .*

◁ **Éléments de preuve.**

Par majoration,  $|x\mathbb{P}(X = x)| \leq M\mathbb{P}(X = x)$ , et cette famille est sommable. ▷

Du fait même de l'interprétation en terme de moyenne pondérée, on obtient :

**Proposition 29.2.7 (Réexpression de l'espérance)**  
*Soit  $X$  une variable aléatoire réelle discrète sur  $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ,  $\Omega$  étant dénombrable. Alors  $X \in L^1$  si et seulement si la famille  $(\mathbb{P}(\{\omega\})X(\omega))_{\omega \in \Omega}$  est sommable, et dans ce cas :*

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\})X(\omega).$$

◁ **Éléments de preuve.**

Groupements de terme (propriété d'associativité pour les familles sommables) sur les  $[X = x]$  pour passer de l'une à l'autre. ▷

## II.2 Théorème de transfert et propriétés de l'espérance

**Théorème 29.2.8 (Théorème de transfert)**  
*Soit  $E$  un borélien de  $\mathbb{R}^n$  et  $X$  un vecteur aléatoire réel discret à valeurs dans  $E \subset \mathbb{R}^n$ , et  $f : E \rightarrow \mathbb{R}$ . Alors  $f(X) \in L^1$  si et seulement si la famille  $(f(x)\mathbb{P}(X = x))_{x \in X(\Omega)}$  est sommable et dans ce cas :*

$$\mathbb{E}(f(X)) = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x).$$

◁ **Éléments de preuve.**

La tribu induite est la tribu  $\mathcal{B}^n \cap \mathcal{P}(E)$  de  $E$ .  
 Groupement de terme (propriété d'associativité) sur les  $f^{-1}(\{y\})$  pour passer d'une expression à l'autre. ▷

Ainsi, l'espérance de  $f(X)$  est entièrement déterminée par la loi de  $X$  et la fonction  $f$ .

**Remarque 29.2.9**

1. Ce théorème est fréquemment utilisé pour  $n = 1$  ( $X$  est une variable aléatoire réelle discrète). Par exemple, en cas de sommabilité, on peut calculer les moments :

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

2. Mais on peut aussi utiliser le théorème de transfert pour des variables vectorielles. Par exemple, si  $(X, Y)$  est un couple de vecteurs aléatoires, et  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  une fonction mesurable, la formule se traduit par :

$$\mathbb{E}(f(X, Y)) = \sum_{(x, y) \in V(\Omega)} f(x, y) \mathbb{P}(V = (x, y)) = \sum_{(x, y) \in X(\Omega) \times Y(\Omega)} f(x, y) \mathbb{P}(X = x, Y = y),$$

où  $V = (X, Y)$ .

3. Le théorème d'associativité restant vrai dans  $\overline{\mathbb{R}}_+$  pour des familles à valeurs dans  $\overline{\mathbb{R}}_+$ , le théorème de transfert reste aussi valide dans ce contexte, lorsque  $f$  est elle-même à valeurs dans  $\overline{\mathbb{R}}_+$ .

On peut aussi remarquer que la condition de convergence absolue imposée dans la définition de  $\mathbb{E}(X)$  et le théorème de transfert nous donnent immédiatement la propriété suivante :

**Proposition 29.2.10 (Condition d'existence de  $\mathbb{E}(X)$ )**

$X \in L^1$  si et seulement si  $|X| \in L^1$ , donc si et seulement si  $\mathbb{E}(|X|) < +\infty$

**Théorème 29.2.11 (Linéarité de l'espérance)**

1. Soit  $X$  et  $Y$  deux variables aléatoires réelles discrètes sur  $(\Omega, \mathcal{T}, \mathbb{P})$  et  $\lambda \in \mathbb{R}$ . Si  $X \in L^1$  et  $Y \in L^1$ , alors  $\lambda X + Y \in L^1$ , et

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

Autrement dit, l'ensemble des v.a.r.d. admettant une espérance est un espace vectoriel, et  $\mathbb{E}$  est une forme linéaire sur cet espace.

2. Si  $X$  et  $Y$  sont positives et  $\lambda \geq 0$ , alors cette égalité reste vraie dans  $\overline{\mathbb{R}}_+$  dans tous les cas.

◁ **Éléments de preuve.**

Théorème de transfert, appliqué à  $f : (x, y) \mapsto \lambda x + y$ , puis manipulations simples de somme et expression des lois marginales en fonction de la loi conjointe.

Le second point provient de la validité du théorème de transfert dans  $\overline{\mathbb{R}}_+$ . ▷

**Proposition 29.2.12 (Positivité et croissance de l'espérance)**

Soit  $X, Y \in L^1$  deux variables aléatoires réelles discrètes admettant une espérance

- Si  $X \geq 0$  ps, alors  $\mathbb{E}(X) \geq 0$ , avec égalité ssi  $X = 0$  ps.
- Si  $X \geq Y$  ps, alors  $\mathbb{E}(X) \geq \mathbb{E}(Y)$ .

◁ **Éléments de preuve.**

Dire que  $X \geq 0$  ps signifie qu'il existe  $\Omega' \subset \Omega$  tel que  $\mathbb{P}(\Omega') = 1$  et  $\forall \omega \in \Omega', X(\omega) \geq Y(\omega)$ .

Le premier point est évident, le cas d'égalité nécessitant  $\mathbb{P}(X = x) = 0$  dès que  $x > 0$ . Le deuxième s'en déduit par linéarité de l'espérance. ▷

**Corollaire 29.2.13 (existence de  $\mathbb{E}$  par majoration)**

Soit  $X$  et  $Y$  deux v.a.r.d. Si  $0 \leq |X| \leq |Y|$  et si  $Y \in L^1$ , alors  $X \in L^1$ .

◁ **Éléments de preuve.**

Utiliser l'égalité précédente dans  $\overline{\mathbb{R}}$ , avec l'égalité  $|X| + Z = |Y|$ , où  $Z \geq 0$ . ▷

**II.3 Théorème de l'espérance totale (HP)**

Nous donnons dans ce paragraphe un théorème très utile, formalisant un certain nombre de raisonnements intuitifs de calculs sur les moyennes. Ce résultat, malgré sa simplicité, n'est malheureusement pas au programme, mais peut être mis en place assez rapidement si on maîtrise bien ses techniques de sommabilité.

**Définition 29.2.14 (Espérance conditionnelle, HP)**

Soit  $A$  un événement non presque-impossible et  $X$  une variable aléatoire réelle discrète. Alors, on dit que  $X$  admet une espérance conditionnelle sachant  $A$  (ou espérance conditionnée à  $A$ ) si  $X$  admet une espérance conditionnelle pour la mesure de probabilité  $\mathbb{P}_A$ , donc si la famille  $(x\mathbb{P}_A([X = x]))_{x \in X(\Omega)}$  est sommable. Dans ce cas, on définit :

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

**Théorème 29.2.15 (Formule de l'espérance totale, HP)**

Soit  $(A_i)_{i \in I}$  un système quasi-complet fini ou dénombrable, constitué d'événements non quasi-impossibles, et  $X$  une v.a.r.d..

1. Les propriétés suivantes sont équivalentes :

- (i) La famille  $(x\mathbb{P}(X = x | A_i)\mathbb{P}(A_i))_{(x,i) \in X(\omega) \times I}$  est sommable
- (ii)  $X \in L^1$
- (iii) La famille  $(\mathbb{E}(|X| | A_i)\mathbb{P}(A_i))_{i \in I}$  est bien définie et sommable.

2. Dans ce cas, on a alors

$$\mathbb{E}(X) = \sum_{i \in I} \mathbb{E}(X | A_i)\mathbb{P}(A_i).$$

◁ **Éléments de preuve.**

C'est juste Fubini, en utilisant d'un côté la formule des probabilités totales, et de l'autre la définition de l'espérance conditionnelle. ▷

Ce théorème est une formalisation de raisonnements en moyennes que l'on est amené à faire de façon assez intuitive. Par exemple, si, lorsqu'on choisit un restaurant, on choisit le restaurant « À la bonne fourchette » avec une probabilité de  $\frac{3}{4}$  et le restaurant « Ah ! la chère fourchette ! » avec une probabilité de  $\frac{1}{4}$ , et qu'en moyenne, on dépense 20 euros en allant dans le premier et 80 euros en allant dans le second, une soirée au restaurant nous coûte en moyenne  $\frac{3}{4} \times 20 + \frac{1}{4} \times 80 = 35$  euros.

**III Variance, dispersion, moments**

Connaître la valeur moyenne prise par une variable est une chose, mais n'a pas forcément beaucoup de pertinence pour avoir un ordre de grandeur du résultat d'une expérience. En effet, l'espérance ne mesure

pas la dispersion de la variable aléatoire : les valeurs prises peuvent tout aussi bien rester toujours très proches de  $\mathbb{E}(X)$ , ou alors s'en éloigner beaucoup, de part et d'autre, de façon à ce qu'en moyenne cela se compense. Il est donc intéressant de définir une quantité mesurant cette dispersion. La dispersion correspond à la moyenne de l'écart à  $\mathbb{E}(X)$ , c'est-à-dire de  $|X - \mathbb{E}(X)|$ . La moyenne arithmétique correspondrait donc à l'espérance de  $|X - \mathbb{E}(X)|$ . Pour des raisons techniques, ce choix n'est pas bon (les valeurs absolues rendent le calcul de cette quantité difficile). D'ailleurs, s'agissant d'une distance, le choix euclidien usuel est plutôt quadratique. Pour cette raison, on a plutôt adopté la moyenne quadratique, c'est-à-dire la racine carrée de la moyenne de  $(X - \mathbb{E}(X))^2$ . Ce choix est surtout motivé par la facilité de calcul qui en découle, notamment grâce à la formule de König-Huygens.

### III.1 Moments d'ordre $k$

#### Définition 29.3.1 (Moments d'ordre $k$ )

1. Le moment d'ordre  $k$  de  $X$  est  $\mathbb{E}(X^k)$  (si ce moment existe). La notation  $X \in L^k$  signifie que  $X$  admet un moment d'ordre  $k$  fini.
2. Si  $X$  admet une espérance, le moment centré d'ordre  $k$  de  $X$  est  $\mathbb{E}((X - \mathbb{E}(X))^k)$ , donc le moment d'ordre  $k$  de la variable centrée  $X - \mathbb{E}(X)$ .

Le théorème de transfert nous donne alors

#### Proposition 29.3.2 (Expression des moments)

Soit  $X$  une variable aléatoire réelle discrète.

1.  $X \in L^k$  si et seulement si  $(x^k \mathbb{P}(X = x))_{x \in X(\Omega)}$  est absolument sommable, et dans ce cas

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

2.  $X$  admet (dans  $\mathbb{R}$ ) un moment centré d'ordre  $k$  si et seulement si  $((x - \mathbb{E}(X))^k \mathbb{P}(X = x))_{x \in X(\Omega)}$  est absolument sommable, et dans ce cas

$$\mathbb{E}((X - \mathbb{E}(X))^k) = \sum_{x \in X(\omega)} (x - \mathbb{E}(X))^k \mathbb{P}(X = x).$$

Par exemple, le moment d'ordre 1 est égal à l'espérance, tandis que le moment centré d'ordre 1 est nul.

#### Théorème 29.3.3

Soit  $X$  une variable aléatoire réelle discrète. Si  $X \in L^k$ , alors  $X \in L^j$  pour tout  $j \in \llbracket 1, k \rrbracket$ .

#### ◁ Éléments de preuve.

Majorer en discutant suivant que  $|X| \leq 1$  ou  $|X| \geq 1$ . On peut formaliser cette discussion par le théorème de l'espérance totale, ou encore plus simplement, avec des indicatrices :

$$|X|^j = \mathbb{1}_{|X| \leq 1} |X|^j + \mathbb{1}_{|X| > 1} |X|^j \leq 1 + \mathbb{1}_{|X| > 1} |X|^k.$$

▷

#### Proposition 29.3.4 (Positivité des moments d'ordre pair)

Soit  $X$  une variable aléatoire admettant un moment d'ordre pair  $2k$ . Alors :

- $\mathbb{E}(X^{2k}) \geq 0$ , avec égalité si et seulement si  $X$  est nulle presque sûrement.
- $\mathbb{E}((X - \mathbb{E}(X))^{2k}) \geq 0$ , avec égalité si et seulement si  $X$  est constante presque sûrement.

## ◁ Éléments de preuve.

Le cas d'égalité découle du cas d'égalité dans 29.2.12.

▷

**III.2 Variance**

Un cas particulier important est le suivant :

**Définition 29.3.5 (Variance d'une variable aléatoire)**

Soit  $X$  une v.a.r.d. admettant une espérance  $\mathbb{E}(X)$ . Alors, sous réserve d'existence (c'est-à-dire sous réserve de sommabilité), la variance de  $X$  est :

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \sum_{x \in X(\Omega)} (x - \mathbb{E}(X))^2 \mathbb{P}(X = x).$$

Ainsi,  $\mathbb{V}$  est le moment centré d'ordre 2 de  $X$ .

Comme cas particulier de 29.3.4, on a toujours  $\mathbb{V}(X) \geq 0$  avec égalité si et seulement si  $X$  est constante ps.

On peut alors définir l'écart quadratique moyen à la moyenne :

**Définition 29.3.6 (Écart-type)**

L'écart-type est l'écart quadratique moyen à la moyenne, à savoir :

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

**Remarque 29.3.7**

La variance peut ne pas exister même si l'espérance existe. Pouvez-vous trouver un exemple ?

**Théorème 29.3.8 (variance d'une variable bornée)**

Soit  $X$  une variable bornée (par exemple finie) sur  $(\Omega, \mathcal{T}, \mathbb{P})$ . Alors  $X$  admet une variance et un écart-type.

## ◁ Éléments de preuve.

$(X - \mathbb{E}(X))^2$  est aussi bornée.

▷

Dans la pratique, la variance se calcule le plus souvent avec le :

**Théorème 29.3.9 (Koenig-Huyghens)**

Soit  $X$  une v.a.r.d. Alors  $X$  admet une variance si et seulement si  $X \in L^2$  et dans ce cas :

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

## ◁ Éléments de preuve.

Développer  $(X - \mathbb{E}(X))^2$ . La relation obtenue montre que si 2 des 3 variables  $X$ ,  $X^2$  et  $(X - \mathbb{E}(X))^2$  admettent une espérance, la troisième aussi. Or, si  $X$  admet une variance, par définition, elle admet une espérance, et si  $X$  admet un moment d'ordre 2, par 29.3.3, elle admet une espérance. La formule de K-H est alors immédiate par linéarité de l'espérance.

▷

La définition-même de la variance amène :

**Proposition 29.3.10 (propriétés de la variance)**

Soit  $a$  et  $b$  deux réels et  $X \in L^2$ . Alors  $aX + b \in L^2$ , et

$$\mathbb{V}(aX + b) = a^2\mathbb{V}(X).$$

**Avertissement 29.3.11**

En général, on n'a pas  $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ . On verra dans la suite du chapitre qu'une condition suffisante pour que ce soit vrai est que  $X$  et  $Y$  soient indépendantes, dans un sens qu'on définira dans ce chapitre. Dans le cas général, la variance d'une somme s'exprime à l'aide de la covariance, ce que nous verrons plus loin.

### III.3 Variables centrées réduites

**Définition 29.3.12 (variable centrée, réduite)**

Une variable  $X$  est dite :

- centrée si  $X \in L^1$ , et  $\mathbb{E}(X) = 0$
- réduite si  $X \in L^2$ , et  $\mathbb{V}(X) = 1$
- centrée réduite si elle est à la fois centrée et réduite.

**Proposition 29.3.13 (Variable centrée réduite associée à  $X$ )**

Soit  $X \in L^2$  telle que  $\mathbb{V}(X) \neq 0$ . Il existe d'unique réels  $a > 0$  et  $b$  tels que la variable  $X^*$  définie par  $X^* = aX + b$  soit centrée réduite. Plus précisément,

$$X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}.$$

On dit que  $X^*$  est la variable aléatoire centrée réduite associée à  $X$ .

◁ Éléments de preuve.

Analyse synthèse par linéarité de  $\mathbb{E}$  et quadraticité de  $\mathbb{V}$ .

▷

## IV Indépendance de variables aléatoires

### IV.1 Couples de variables aléatoires indépendantes

**Définition 29.4.1 (Variables indépendantes)**

Soit  $X$  et  $Y$  deux variables aléatoires définies sur  $(\Omega, \mathcal{T}, \mathbb{P})$ , à valeurs dans  $(E_1, \mathcal{T}_1)$  et  $(E_2, \mathcal{T}_2)$  respectivement. Les variables  $X$  et  $Y$  sont indépendantes (noté  $X \perp\!\!\!\perp Y$ ) si et seulement si pour tout  $(A, B) \in \mathcal{T}_1 \times \mathcal{T}_2$ , les événements  $[X \in A]$  et  $[Y \in B]$  sont indépendants, c'est-à-dire :

$$\forall (A, B) \in \mathcal{T}_1 \times \mathcal{T}_2, \quad \mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Autrement dit,  $X \perp\!\!\!\perp Y$  si et seulement si les tribus  $\mathcal{T}_X$  et  $\mathcal{T}_Y$  sont indépendantes (dans le sens où un prenant un événement quelconque dans chacune de ces tribus, ces deux événements sont indépendants).

En particulier, le lemme de classe monotone, qu'on a déjà évoqué sans l'énoncer, amène, pour les variables aléatoires réelles :

**Proposition 29.4.2 (v.a.r. indépendantes, admis)**

Soit  $X$  et  $Y$  deux variables aléatoires réelles. Alors  $X$  et  $Y$  sont indépendantes si et seulement si pour tout  $(x, y) \in \mathbb{R}^2$ , les événements  $[X \leq x]$  et  $[Y \leq y]$  sont indépendants, c'est-à-dire :

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y).$$

Pour les variables aléatoires réelles discrètes, on obtient une description ponctuelle :

**Proposition 29.4.3 (v.a.r.d. indépendantes)**

Soit  $X$  et  $Y$  deux variables aléatoires réelles discrètes. Alors  $X \perp\!\!\!\perp Y$  si et seulement si pour tout  $x \in X(\Omega)$  et pour tout  $y \in Y(\Omega)$ , les événements  $[X = x]$  et  $[Y = y]$  sont indépendants, c'est-à-dire :

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

◁ **Éléments de preuve.**

Exprimer  $A$  et  $B$  comme union de singletons.

▷

**Remarque 29.4.4**

Ce dernier résultat affirme qu'en cas d'indépendance, les lois marginales déterminent la loi conjointe d'un couple de variables aléatoires réelles discrètes. Ce fait reste vrai dans une situation plus générale (variables non discrètes), mais repose sur des résultats un peu plus délicat sur les tribus (notamment le lemme de classe monotone).

En effet, dans le cas où  $X$  et  $Y$  sont deux variables aléatoires réelles indépendantes, la loi du couple  $(X, Y)$  coïncide avec le produit des lois de  $X$  et de  $Y$  sur les boréliens de  $\mathbb{R}^2$  de la forme  $A \times B$ , où  $A$  et  $B$  sont des boréliens de  $\mathbb{R}$ . Or, les boréliens du type  $A \times B$  engendrent (par définition)  $\mathcal{B}_2$  et forment une classe stable par intersection finie. Or, d'après le lemme  $\lambda$ - $\pi$  de Dynkin vu en DM, il existe une unique mesure de probabilité sur  $\mathbb{R}^2$  prenant des valeurs déterminées sur les  $A \times B$ . Ainsi,  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  déterminent  $\mathbb{P}_{(X, Y)}$ .

Ce raisonnement reste valable plus généralement pour des variables quelconques à valeurs respectivement dans  $(E_1, \mathcal{T}_1)$  et  $(E_2, \mathcal{T}_2)$ , en remarquant que le couple  $(X, Y)$  définit dans ce cas une variable aléatoire à valeur dans le produit  $E_1 \times E_2$ , muni de la tribu produit, engendrée par les produits  $A \times B$ , pour  $A \in \mathcal{T}_1$  et  $B \in \mathcal{T}_2$ .

## IV.2 Familles de v.a.r. indépendantes

On généralise ici la notion d'indépendance au cas d'un nombre plus grand de variables aléatoires. Comme dans le cas des événements, la notion d'indépendance deux à deux est souvent insuffisante pour traduire complètement les situations d'indépendance rencontrées. Nous devons ici aussi distinguer l'indépendance deux à deux de l'indépendance mutuelle définie ci-dessous.

Soit  $(X_k)_{k \in K}$  une famille quelconque de variables aléatoires,  $X_k$  étant à valeurs dans  $(E_k, \mathcal{T}_k)$

**Définition 29.4.5 (Famille de variables mutuellement indépendantes)**

On dit que les variables aléatoires  $X_k$ , pour  $k \in K$ , sont mutuellement indépendantes si pour toute famille  $(A_k)_{k \in K}$  telle que pour tout  $k \in K$ ,  $A_k \in \mathcal{T}_k$ , les événements  $[X_k \in A_k]$ , pour  $k \in K$ , sont

mutuellement indépendants, donc si et seulement si pour tout sous-ensemble fini  $J$  de  $K$ ,

$$\mathbb{P}\left(\bigcap_{j \in J} [X_j \in A_j]\right) = \prod_{j \in J} \mathbb{P}(X_j \in A_j).$$

De même que pour 2 variables, cela se traduit par le fait que les tribus  $\mathcal{T}_{X_j}$  forment une famille de tribus mutuellement indépendantes (dans le sens où en prenant un événement quelconque de chaque tribu, on obtient toujours une famille mutuellement indépendante).

Dans le cas de variables aléatoires réelles discrètes, on a la définition équivalente ci-dessous.

**Proposition 29.4.6 (Caractérisation ponctuelle de la mutuelle indépendance)**

Si les  $X_k$  sont réelles discrètes, elles sont mutuellement indépendantes si et seulement si pour toute famille  $(x_k)_{k \in K}$  telle que pour tout  $k \in K$ ,  $x_k \in X_k(\Omega)$ , les événements  $[X_k = x_k]$ , pour  $k \in K$ , sont mutuellement indépendants, donc si et seulement si pour tout sous-ensemble fini  $J$  de  $K$ ,

$$\mathbb{P}\left(\bigcap_{j \in J} [X_j = x_j]\right) = \prod_{j \in J} \mathbb{P}(X_j = x_j).$$

◁ **Éléments de preuve.**

On peut se contenter de montrer l'équivalence dans le cas  $|J| = 2$  (récurrence). Décomposer les  $[X_k \in A_k]$  en union finie ou dénombrable de  $[X_k = x_{k,i}]$ . ▷

### IV.3 Coalitions

**Proposition 29.4.7 (Fonctions de variables indépendantes)**

Soit  $X \perp\!\!\!\perp Y$  deux v.a.r.d. indépendantes, et  $f$  et  $g$  deux fonctions définies respectivement sur  $X(\Omega)$  et  $Y(\Omega)$ . Alors  $f(X) \perp\!\!\!\perp g(Y)$ .

Ceci est un cas particulier du résultat plus général ci-dessous :

**Proposition 29.4.8 (Lemme des coalitions)**

Soit  $1 \leq m < n$ , et  $(X_1, \dots, X_m, \dots, X_n)$  une famille de v.a.r.d. mutuellement indépendantes. Soit  $f$  et  $g$  deux fonctions définies sur  $\mathbb{R}^m$  et  $\mathbb{R}^{m-n}$  respectivement. Alors  $f(X_1, \dots, X_m)$  et  $g(X_{m+1}, \dots, X_n)$  sont indépendantes.

◁ **Éléments de preuve.**

Démonstration un peu technique, en faisant la vérification ponctuellement. ▷

**Remarque 29.4.9**

1. La démonstration s'adapte bien pour une famille  $(X_i)_{i \in I}$  de variables mutuellement indépendantes, une partition finie  $(K_j)_{j \in J}$  de  $I$ , et des fonctions  $f_j$  de  $|K_j|$  variables, évaluées sur le  $|K_j|$ -uplet  $(X_i, i \in K_j)$  (on obtient alors la mutuelle indépendance). Mais à la fois l'énoncé et la démonstration deviennent beaucoup plus techniques à écrire, alors que les cas particuliers généralement utilisés sont limpides.

2. Ainsi, on obtient la mutuelle indépendance de variable  $Y_j$  qui s'écrivent comme fonctions (mesurables) de variables aléatoires  $X_i$  mutuellement indépendantes, chacune des  $X_i$  n'intervenant que dans la définition d'au plus un  $Y_j$ .
3. On peut aussi sans difficulté supplémentaire adapter au cas de variables aléatoires discrètes non réelles.

## V Covariance

### V.1 Espérance d'un produit

On a vu que l'espérance est linéaire. On peut se demander si elle est multiplicative. La réponse est non en général. On peut cependant obtenir :

**Théorème 29.5.1 (Espérance d'un produit de variables indépendantes)**

*Si  $X$  et  $Y$  sont des variables aléatoires réelles discrètes indépendantes et admettent une espérance, alors  $XY$  admet une espérance, et*

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

◁ **Éléments de preuve.**

Théorème de transfert avec la fonction  $f : (x, y) \mapsto xy$ . ▷

Il est important de noter que ce résultat entre en défaut lorsque  $X$  et  $Y$  ne sont pas indépendants :

**Exemple 29.5.2**

Soit  $X_1, X_2$ , et  $X_3$ , indépendantes, suivant des lois de Bernoulli de paramètre  $p \in ]0, 1[$ . Soit  $Y_1 = X_1X_2$  et  $Y_2 = X_2X_3$ . Alors  $\mathbb{E}(Y_1Y_2) \neq \mathbb{E}(Y_1)\mathbb{E}(Y_2)$ .

On donne tout de même un résultat d'existence :

**Lemme 29.5.3 (CS d'existence de  $\mathbb{E}(XY)$ )**

*Si  $X$  et  $Y$  sont des v.a.r.d. admettant des moments d'ordre 2, alors  $XY$  admet une espérance.*

◁ **Éléments de preuve.**

Majoration  $|XY| \leq \frac{1}{2}(X^2 + Y^2)$  dans la somme déterminant  $XY$  par le théorème de transfert. ▷

On en déduit en particulier :

**Proposition 29.5.4 (L'espace des v.a.r.d. admettant un moment d'ordre 2)**

*L'ensemble des variables aléatoires réelles discrètes sur  $(\Omega, \mathcal{T}, \mathbb{P})$  admettant un moment d'ordre 2 est un sous-espace vectoriel de l'espace de toutes les v.a.r.d..*

◁ **Éléments de preuve.**

La linéarité nécessite d'étudier  $(\lambda X + Y)^2$ , et fait donc intervenir  $XY$ . ▷

## V.2 Covariance

Nous introduisons maintenant une quantité permettant de comparer les variations de  $X$  et de  $Y$ . Si  $X$  et  $Y$  sont indépendantes, ces variations sont indépendantes ce qui se traduira par le fait que cette quantité est nulle. Cette propriété sera le reflet de l'égalité que l'on vient d'obtenir  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ . Ainsi, notre mesure du défaut d'indépendance se définira comme la différence  $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ . Afin de mieux comprendre la signification de cette différence, nous l'introduisons sous une forme légèrement différente.

### Définition 29.5.5 (Covariance)

Soit  $X$  et  $Y$  deux variables aléatoires réelles discrètes. Sous réserve d'existence de ces espérances, on définit la *covariance* de  $X$  et  $Y$  par :

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

### Remarque 29.5.6 (Interprétation de la covariance)

Si  $\text{cov}(X, Y) > 0$ ,  $X - \mathbb{E}(X)$  et  $Y - \mathbb{E}(Y)$  ont tendance à être de même signe, donc  $X$  et  $Y$  ont tendance à se situer du même côté de leur espérance. Ainsi :

- $\text{cov}(X, Y) > 0$  signifie que  $X$  et  $Y$  ont tendance à évoluer parallèlement
- $\text{cov}(X, Y) < 0$  signifie que  $X$  et  $Y$  ont tendance à avoir une évolution opposée.

### Définition 29.5.7 (Variables décorréllées)

Les variables  $X$  et  $Y$  sont dites décorréllées lorsque  $\text{cov}(X, Y) = 0$ .

### Proposition 29.5.8 (CS d'existence de $\text{cov}(X, Y)$ )

Si  $X \in L^2$  et  $Y \in L^2$ , alors  $\text{cov}(X, Y)$  existe.

◁ Éléments de preuve.

Elles admettent alors aussi des moments d'ordre 1. Développer. ▷

### Proposition 29.5.9 (Propriétés de la covariance)

Soit  $X, Y \in L^2$ .

1.  $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$  ;
2.  $\text{cov}(X, Y) = \text{cov}(Y, X)$  (*symétrie*)
3.  $\text{cov}$  est bilinéaire sur l'espace des v.a.r.d. sur  $(\Omega, \mathcal{T}, \mathbb{P})$  admettant un moment d'ordre 2.
4.  $\text{cov}(X, X) = \mathbb{V}(X)$ .
5.  $\text{cov}(1, X) = 0$ .
6. Si  $X$  et  $Y$  sont indépendantes,  $\text{cov}(X, Y) = 0$ .

◁ Éléments de preuve.

Assez immédiat par linéarité de l'espérance. ▷

### Avertissement 29.5.10

La réciproque de la dernière propriété est fautive : il existe des variables décorréllées, mais non indépendantes. Voir les exercices pour des exemples.

Les propriétés 2, 3 et 4 permettent d'affirmer que  $\text{cov}$  est une forme bilinéaire symétrique positive, de forme quadratique associée égale à la variance. En particulier, on dispose dans cette situation de l'inégalité de Cauchy-Schwarz :

**Théorème 29.5.11 (Inégalité de Cauchy-Schwarz)**

Soit  $X, Y \in L^2$ . Alors

$$|\text{cov}(X, Y)| \leq \sigma(X)\sigma(Y),$$

avec égalité si et seulement s'il existe une relation affine presque sûrement entre  $X$  et  $Y$  (c'est-à-dire une relation non triviale  $aX + bY + c = 0$ )

◁ **Éléments de preuve.**

Si cela n'a pas encore été fait à ce moment, on renvoie au cours d'algèbre bilinéaire, pour la preuve de cette inégalité, plus généralement pour des formes bilinéaires symétriques positives. ▷

**Définition 29.5.12 (Coefficient de corrélation)**

Le coefficient de corrélation  $\rho(X, Y)$  est défini, pour des variables non quasi-certaines, par :

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

D'après le résultat précédent,  $\rho(X, Y) \in [-1, 1]$ ;  $\rho(X, Y) = 0$  indique une décorrélation, alors que  $|\rho(X, Y)| = 1$  indique une corrélation très forte (les variables aléatoires  $X$  et  $Y$  sont liées par la relation de dépendance la plus forte qu'on puisse imaginer).

**Remarque 29.5.13 (Une autre inégalité de Cauchy-Schwartz)**

Pour  $(X, Y) \in (L^2)^2$ ,  $(X, Y) \mapsto \mathbb{E}(XY)$  est également bilinéaire symétrique positive, donc on a aussi :

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

### V.3 Variance d'une somme

Une autre formule issue directement de l'algèbre bilinéaire (et du lien entre une forme bilinéaire symétrique et la forme quadratique associée) est la formule de polarisation :

**Proposition 29.5.14 (Formule de polarisation)**

Soit  $X, Y \in L^2$ . Alors :

$$\text{cov}(X, Y) = \frac{1}{2}(\mathbb{V}(X + Y) - \mathbb{V}(X) - \mathbb{V}(Y)).$$

Cette formule est dans notre contexte surtout utilisée pour calculer la variance d'une somme à l'aide des covariances. Nous la reexprimons donc de la façon suivante :

**Proposition 29.5.15 (Variance d'une somme)**

1. Si  $X, Y \in L^2$ , alors  $X + Y \in L^2$ , et :

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2 \cdot \text{cov}(X, Y).$$

2. Si de plus  $X \perp\!\!\!\perp Y$ , alors  $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$ .

**Remarque 29.5.16**

La multiplication des réels est aussi une forme bilinéaire symétrique. A quoi correspond la formule de polarisation dans ce contexte ?

En utilisant la propriété de bilinéarité généralisée, on obtient de même l'expression de la variance d'une somme de  $n$  terme :

**Proposition 29.5.17 (Variance de  $X_1 + \dots + X_n$ )**

1. Soit  $X_1, \dots, X_n \in L^2$  ; alors  $X_1 + \dots + X_n \in L^2$  et :

$$\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n) + 2 \cdot \sum_{1 \leq i < j \leq n} \text{cov}(X_i, X_j).$$

2. Si de plus les  $X_i$  sont 2 à 2 indépendantes,

$$\mathbb{V}(X_1 + \dots + X_n) = \mathbb{V}(X_1) + \dots + \mathbb{V}(X_n).$$

**Remarque 29.5.18**

Il est remarquable qu'il n'y ait besoin que d'une hypothèse d'indépendance 2 à 2 ici, et non d'une indépendance mutuelle. Une hypothèse de non corrélation 2 à 2 serait même suffisante.

**Proposition 29.5.19 (Reexpression matricielle de la variance d'une somme)**

Soit  $C = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$ . Alors :

$$\mathbb{V}(X_1 + \dots + X_n) = C^T \cdot \mathbb{V}(X_1, \dots, X_n) \cdot C,$$

où  $\mathbb{V}(X_1 + \dots + X_n)$  est la matrice ci-dessous, appelée matrice des variances-covariances :

$$V(X_1, \dots, X_n) = (\text{cov}(X_i, X_j))_{(i,j) \in \llbracket 1, n \rrbracket^2}.$$

## VI Lois discrètes classiques

Un certain nombre de lois interviennent fréquemment. De nombreuses situations peuvent s'y ramener. Il est donc intéressant de les connaître afin de s'économiser un certain nombre de calculs d'espérance ou de variance. Il est crucial de bien connaître l'expérience-type associée à chacune de ces lois, car c'est *via* la description de cette expérience que ces lois interviennent dans des situations diverses.

### VI.1 Loi quasi-certaine

**Définition 29.6.1 (Loi quasi-certaine)**

Une variable suit la loi quasi-certaine (ou est quasi-certaine) de valeur  $c$  si  $X = c$  ps, donc  $\mathbb{P}(X = c) = 1$ , et pour tout  $x \neq c$ ,  $\mathbb{P}(X = x) = 0$ .

**Proposition 29.6.2 (Espérance et variance d'un loi quasi-certaine)**

Soit  $X$  une variable quasi-certaine de valeur  $c$ . Alors

$$\mathbb{E}(X) = c \quad \text{et} \quad \mathbb{V}(X) = 0.$$

**Proposition 29.6.3 (Caractérisation des variables quasi-certaines)**

Réciproquement, si  $X \in L^2$  et  $\mathbb{V}(X) = 0$ , alors  $X$  est une variable quasi-certaine.

◁ **Éléments de preuve.**

Tous les termes de la somme  $\sum_{x \in X(\Omega)} (x - \mathbb{E}(X))^2 \mathbb{P}(X = x)$  doivent être nuls. ▷

**VI.2 Loi uniforme**

L'expérience consiste à tirer une boule dans une urne contenant  $n$  boules numérotées de 1 à  $n$ . La variable  $X$  est le numéro de la boule tirée. Ainsi, on définit :

**Définition 29.6.4 (Loi uniforme)**

Soit  $n \in \mathbb{N}^*$ . Une v.a.r.  $X$  suit la loi uniforme de paramètre  $n$  si  $X(\Omega) = \llbracket 1, n \rrbracket$  et si :

$$\forall k \in \llbracket 1, n \rrbracket, \quad \mathbb{P}(X = k) = \frac{1}{n}.$$

On note  $X \sim \mathcal{U}(n)$ .

**Proposition 29.6.5 (Espérance et variance d'une loi uniforme)**

Soit  $X \sim \mathcal{U}(n)$ . Alors  $\mathbb{E}(X) = \frac{n+1}{2}$  et  $\mathbb{V}(X) = \frac{n^2-1}{12}$ .

◁ **Éléments de preuve.**

Utiliser KH. On est alors ramené à des sommes d'entiers et de carrés d'entiers. ▷

Pour tout ensemble fini  $E$ , on note plus généralement  $X \sim \mathcal{U}(E)$  pour une v.a.r. suivant la loi uniforme sur  $X$ , c'est-à-dire une v.a.r. telle que :

$$\forall x \in E, \quad \mathbb{P}(X = x) = \frac{1}{|E|}.$$

**Proposition 29.6.6 (Expérience-type associée à une loi uniforme)**

Soit  $U$  une urne contenant  $n$  boules numérotées de 1 à  $n$ . On effectue un tirage avec équiprobabilité, et on note  $X$  le numéro de la boule obtenu. Alors  $X \sim \mathcal{U}(n)$ .

Plus généralement, si  $X$  est une variable suivant une loi uniforme sur l'ensemble  $\llbracket a, b \rrbracket$ , elle vérifie :

$$\mathbb{E}(X) = \frac{a+b}{2} \quad \text{et} \quad \mathbb{V}(X) = \frac{(b-a+1)^2-1}{12}.$$

**VI.3 Loi de Bernoulli**

L'expérience consiste à tirer dans une urne contenant une proportion  $p$  de boules blanches. On note  $X$  la variable aléatoire égale à 1 si on tire une boule blanche, et 0 sinon.

On peut aussi définir  $X$  à l'aide de tirages à pile ou face avec une pièce déséquilibrée, dont la probabilité de pile est  $p$ ;  $X$  prend alors la valeur 1 si on tire pile, et 0 si on tire face.

Ainsi,  $[X = 1]$  représente le succès dans une expérience ayant une probabilité  $p$  de succès.

De façon plus générale, une situation similaire se produit dès lors qu'on a une expérience à deux issues : succès et échec. L'événement  $[X = 1]$  représente alors le succès, et  $[X = 0]$  représente l'échec.

On définit donc :

**Définition 29.6.7 (Loi de Bernoulli)**

Soit  $X$  une v.a.r. et  $p \in ]0, 1[$ . On dit que  $X$  suit une loi de Bernoulli de paramètre  $p$  (et on note  $X \sim \mathcal{B}(1, p)$  ou  $X \sim \mathcal{B}(p)$ ) si :

$$X(\Omega) = \{0, 1\} \quad \text{et} \quad \begin{cases} \mathbb{P}(X = 0) = 1 - p \\ \mathbb{P}(X = 1) = p \end{cases}$$

Désormais, on se donne un réel  $p \in ]0, 1[$ , et on note  $q = 1 - p$ .

**Proposition 29.6.8 (Espérance et variance d'une loi de Bernoulli)**

Soit  $X \sim \mathcal{B}(p)$ . Alors  $\mathbb{E}(X) = p$  et  $\mathbb{V}(X) = pq$ .

◁ **Éléments de preuve.**

Un peu trivial...

▷

**Exemple 29.6.9 (Fonction indicatrice d'un événement)**

Étude de  $X(\omega) = \mathbb{1}_A(\omega)$ ,  $A \in \mathcal{P}(\Omega)$ .

La motivation même de la définition de la loi de Bernoulli amène :

**Proposition 29.6.10 (Expérience-type associée à une loi de Bernoulli)**

Soit une expérience à 2 issue (par exemple P/F, ou tirage dans une urne contenant 2 types de boules), l'une représentant le succès et l'autre l'échec, la probabilité d'obtenir un succès étant  $p$ . Soit  $X$  prenant la valeur 1 en cas de succès et la valeur 0 en cas d'échec. Alors  $X \sim \mathcal{B}(p)$ .

Une expérience telle que décrite dans la proposition précédente est usuellement appelée « expérience de Bernoulli ».

**Remarque 29.6.11**

Toute variable aléatoire prenant ses valeurs dans  $\{0, 1\}$  est une variable de Bernoulli, à condition d'élargir la définition en acceptant de considérer les cas dégénérés  $p = 0$  et  $p = 1$ .

## VI.4 Loi binomiale – nombre de succès

Soit  $n \in \mathbb{N}^*$ . On répète  $n$  fois une expérience de Bernoulli, et on note  $X$  le nombre de succès obtenus. Par exemple on compte le nombre de Pile obtenus dans une succession de  $n$  lancers (indépendants) avec une pièce déséquilibrée donnant Pile avec une probabilité de  $p$ .

Déterminons la loi de  $X$ . Tout d'abord, on peut avoir de 0 à  $n$  succès, ainsi  $X(\Omega) = \llbracket 0, n \rrbracket$ . Soit donc  $k \in \llbracket 0, n \rrbracket$ . L'événement  $[X = k]$  est obtenu si on a eu  $k$  succès. Soit, pour tout  $i \in \llbracket 1, n \rrbracket$ ,  $S_i$  l'événement :

obtenir un succès à la  $i$ -ième expérience. Alors :

$$[X = k] = \bigcup_{\substack{I \subset \llbracket 1, n \rrbracket \\ |I|=k}} \left( \bigcap_{i \in I} S_i \cap \bigcap_{i \in \llbracket 1, n \rrbracket \setminus I} \overline{S}_i \right)$$

Les événements  $S_i$ ,  $i \in \llbracket 1, n \rrbracket$ , sont mutuellement indépendants, et les événements de l'union sont deux à deux incompatibles. Ainsi,

$$\mathbb{P}(X = k) = \sum_{\substack{I \subset \llbracket 1, n \rrbracket \\ |I|=k}} \left( \prod_{i \in I} p \cdot \prod_{i \in \llbracket 1, n \rrbracket \setminus I} (1-p) \right) = \sum_{\substack{I \subset \llbracket 1, n \rrbracket \\ |I|=k}} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Le coefficient binomial correspond à la position des  $k$  succès,  $p^r$  est la probabilité d'obtention de ces  $r$  succès et  $(1-p)^r$  est la probabilité d'obtention des échecs aux autres tirages.

On définit donc :

**Définition 29.6.12 (Loi binomiale)**

Soit  $X$  une v.a.r.. On dit que  $X$  suit la loi binomiale de paramètres  $(n, p)$  (et on note  $X \sim \mathcal{B}(n, p)$ ) si :

$$X(\Omega) = \llbracket 0, n \rrbracket, \quad \text{et} \quad \forall k \in \llbracket 0, n \rrbracket, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}, \quad \text{où } q = 1 - p.$$

Ceci définit bien une loi de probabilités. En effet :  $\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$ .

Le calcul inductif nous ayant servi de motivation à la définition de la loi binomiale amène directement :

**Proposition 29.6.13 (Expérience-type associée à la loi binomiale)**

Soit  $X$  le nombre de succès obtenus lors d'une répétition de  $n$  épreuves de Bernoulli indépendantes de même paramètre  $p$ . Alors  $X \sim \mathcal{B}(n, p)$ .

**Proposition 29.6.14 (Espérance et variance d'une loi binomiale)**

Soit  $X \sim \mathcal{B}(n, p)$ . Alors  $\mathbb{E}(X)$  et  $\mathbb{V}(X)$  existent, et :

$$\mathbb{E}(X) = np, \quad \text{et} \quad \mathbb{V}(X) = npq.$$

◁ **Éléments de preuve.**

On est ramené à des sommes du binôme, éventuellement après un ou deux comités-président. Il peut être intéressant de remplacer le calcul de  $\mathbb{E}(X^2)$  par le calcul de l'espérance d'un autre polynôme en  $X$  de degré 2, permettant directement l'utilisation répétée de la formule comité-président. ▷

On peut obtenir ce résultat soit à l'aide d'un calcul direct, soit en remarquant que  $X$  est une somme de  $n$  variables de Bernoulli indépendantes.

## VI.5 Loi géométrique – temps d'attente du premier succès

On considère maintenant une succession infinie d'expériences de Bernoulli indépendantes de paramètre  $p$ . On considère  $X$  la variable aléatoire correspondant au rang du premier succès.

Par exemple  $X$  est le rang du premier pile obtenu dans une succession de lancers avec une probabilité d'obtention de pile égale à  $p$  pour chaque tirage.

Déterminons la loi de  $X$ . Les valeurs possibles de  $X$  sont toutes les valeurs entières strictement positives :  $X(\Omega) = \mathbb{N}^*$ . Soit  $k \in \mathbb{N}^*$ . Alors, en utilisant les mêmes notations que précédemment,

$$\mathbb{P}(X = k) = \mathbb{P}(\overline{S}_1 \cap \dots \cap \overline{S}_{k-1} \cap S_k) = \mathbb{P}(\overline{S}_1) \dots \mathbb{P}(\overline{S}_{k-1}) \mathbb{P}(S_k) = pq^{k-1}$$

On définit donc :

**Définition 29.6.15 (Loi géométrique)**

Soit  $X$  une v.a.r.. On dit que  $X$  suit la loi géométrique de paramètre  $p$ , et on note  $X \sim \mathcal{G}(p)$ , si :

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

Cela définit bien une loi de probabilité :  $\sum_{k=1}^{+\infty} \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} pq^{k-1} = \frac{p}{1-q} = 1$ .

Et par notre motivation-même :

**Proposition 29.6.16 (Expérience-type associée à une loi géométrique)**

Soit  $X$  le temps d'attente du premier succès (c'est-à-dire le nombre d'expériences nécessaires pour obtenir le premier succès) lors d'une répétition infinie d'expériences de Bernoulli indépendantes de même paramètre  $p$ . Alors  $X \sim \mathcal{G}(p)$ .

**Proposition 29.6.17 (Espérance et variance d'une loi géométrique)**

Soit  $X \sim \mathcal{G}(p)$ . Alors  $X$  admet une espérance et une variance, et :

$$\mathbb{E}(X) = \frac{1}{p} \quad \text{et} \quad \mathbb{V}(X) = \frac{q}{p^2}.$$

◁ **Éléments de preuve.**

On est ramené à des sommes géométriques dérivées. Comme plus haut, on peut avoir intérêt à ne pas considérer directement  $\mathbb{E}(X^2)$ . ▷

**Remarque 29.6.18**

Il est parfois commode de définir  $X$  à valeurs dans  $\mathbb{N}^* \cup \{+\infty\}$ , et en définissant la probabilité  $\mathbb{P}(X = +\infty) = 0$ . Cela permet d'englober le cas du malchanceux éternel.

## VI.6 Loi de Poisson

On définit une dernière loi, un peu à part puisqu'elle ne correspond pas à une expérience précise, mais qui apparaît souvent comme modélisation, ou comme loi limite (il s'agit d'une limite dans un certain sens de lois binomiales).

**Définition 29.6.19 (Loi de Poisson)**

Soit  $\lambda \in \mathbb{R}_+$  et  $X$  une v.a.r.. On dit que  $X$  suit la loi de Poisson de paramètre  $\lambda$ , et on note  $X \sim \mathcal{P}(\lambda)$ , si :

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

D'après les résultats sur les séries exponentielles, cela définit bien une loi de probabilité. De plus :

**Proposition 29.6.20 (Espérance et variance d'une loi de Poisson)**

Soit  $X \sim \mathcal{P}(\lambda)$ . Alors  $X$  admet une espérance et une variance, et :

$$\mathbb{E}(X) = \lambda \quad \text{et} \quad \mathbb{V}(X) = \lambda.$$

## ◁ Éléments de preuve.

Se ramener à des séries exponentielles, et toujours la même remarque. ▷

## VI.7 Stabilité des lois classiques

Par stabilité d'une loi classique, on entend la chose suivante : si  $X$  et  $Y$  suit un certain type de loi et sont indépendante, alors  $X + Y$  suit une loi de même type (avec des paramètres éventuellement différents). Nous énonçons ici deux propriétés de stabilité.

**Proposition 29.6.21 (Stabilité des lois binomiales)**

Soit  $(m, n) \in (\mathbb{N}^*)^2$  et  $p \in ]0, 1[$ . Si  $X \sim \mathcal{B}(n, p)$  et  $Y \sim \mathcal{B}(m, p)$  sont indépendantes, alors  $X + Y \sim \mathcal{B}(n + m, p)$ .

En particulier si  $X \sim \mathcal{B}(n, p)$ ,  $X$  est la somme de  $n$  variables de Bernoulli de paramètre  $p$  mutuellement indépendantes.

## ◁ Éléments de preuve.

Se comprend bien par l'expérience. Le démontrer par le calcul en décomposant  $[X + Y = k]$  en union de  $[X = i] \cap [Y = k - i]$ , et en calculant ces probabilités par indépendance. ▷

**Corollaire 29.6.22 (Somme de variables de Bernoulli indépendantes)**

Soit  $X_1, \dots, X_n$  des variables aléatoires mutuellement indépendantes suivant la loi  $\mathcal{B}(p)$ . Alors  $X_1 + \dots + X_n \sim \mathcal{B}(n, p)$ .

**Proposition 29.6.23 (Stabilité des lois de Poisson)**

Soit  $(\lambda, \mu) \in (\mathbb{R}_+)^2$ . Si  $X \sim \mathcal{P}(\lambda)$  et  $Y \sim \mathcal{P}(\mu)$  sont indépendantes, alors  $X + Y \sim \mathcal{P}(\lambda + \mu)$ .

## ◁ Éléments de preuve.

Même principe de calcul. ▷

## VI.8 Tableau récapitulatif

La table 29.1 récapitule les résultats obtenus pour les différentes lois classiques étudiées. Nous indiquons en gras les lois qui sont au programme de Sup et de Spé. Nous avons indiqué dans ce tableau quelques autres lois classiques, que vous pouvez étudier en exercice :

- Loi de Pascal  $\mathcal{P}(r, p)$  : temps d'attente du  $r$ -ième succès dans une répétition d'épreuves de Bernoulli indépendantes de même paramètre. Lorsque  $r = 1$ , on retrouve une loi géométrique. On a une propriété de stabilité au sens du paragraphe précédent, qui se comprend bien de façon intuitive.
- Loi binomiale négative  $\mathcal{J}(r, p)$  : nombre d'échecs avant le  $r$ -ième succès. On peut remarquer que  $X \sim \mathcal{J}(r, p)$  ssi  $X + r \sim \mathcal{P}(r, p)$ .
- Loi hypergéométrique  $\mathcal{H}(N, n, p)$  : nombre de boules blanches tirées lors de  $n$  tirages sans remise dans une urne contenant initialement  $Np$  boules blanches et  $N(1 - p)$  boules noires
- Les temps d'attente de la première boule blanche, ou de la  $r$ -ième, dans le même contexte.

## VII Inégalités et convergences

Il est fréquent d'aboutir à l'étude d'une suite de variables aléatoires et de s'intéresser à la variable obtenue « par passage à la limite ». Il faut pour cela définir ce qu'on entend par passage à la limite sur

des variables aléatoires et se donner quelques outils d'étude. Il existe plusieurs définitions de différents types de convergence. Les deux types les plus fréquemment utilisés sont la convergence en probabilités et la convergence en loi. Les deux étant hors-programme, nous nous contenterons d'étudier la convergence en probabilités. Attention cependant au fait que si les convergences sont hors-programme, les deux inégalités (Markov et Bienaymé-Tchebychev) sont elles bien au programme.

## VII.1 Convergence en probabilités

### Définition 29.7.1 (Convergence en probabilités, HP)

Soit  $(X_n)_{n \in \mathbb{N}}$  une suite de v.a.r.d. sur  $(\Omega, \mathcal{T}, \mathbb{P})$ , et  $X$  une autre v.a.r.d.. On dit que  $X_n$  converge en probabilité (ou converge stochastiquement) vers  $X$  si :

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

## VII.2 Inégalités en probabilités

### Théorème 29.7.2 (Inégalité de Markov)

Soit  $X$  une v.a.r.d. sur  $(\Omega, \mathcal{T}, \mathbb{P})$  telle que  $X(\Omega) \subset \mathbb{R}_+$ , admettant une espérance non nulle  $\mathbb{E}(X)$ . Alors, pour tout  $\lambda \in \mathbb{R}_+^*$  :

$$\mathbb{P}(X \geq \lambda \mathbb{E}(X)) \leq \frac{1}{\lambda}.$$

◁ Éléments de preuve.

Minorer  $X$  par  $\lambda \mathbb{E}(X) \mathbb{1}_{X \leq \lambda \mathbb{E}(X)}$ . ▷

Voici deux variantes de l'inégalité de Markov, parfois plus commodes à utiliser :

### Corollaire 29.7.3 (Réexpression de l'inégalité de Markov)

Sous les mêmes hypothèses, pour tout  $\varepsilon > 0$ ,

$$\mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon} \quad \text{et} \quad \mathbb{P}(X > \varepsilon) \leq \frac{\mathbb{E}(X)}{\varepsilon}$$

◁ Éléments de preuve.

Avec  $\lambda = \frac{\varepsilon}{\mathbb{E}(X)}$ . ▷

### Corollaire 29.7.4 (Inégalité de Markov quadratique)

Soit  $X$  une variable aléatoire (discrète ou à densité) admettant un moment d'ordre 2. Alors :

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

◁ Éléments de preuve.

Si  $\mathbb{E}(X^2) \neq 0$ , appliquer ce qui précède à  $X^2$ . Et sinon ? ▷

**Théorème 29.7.5 (Inégalité de Bienaymé-Tchebychev)**

Soit  $Y$  une v.a.r.d. admettant une espérance  $m$  et une variance  $\sigma^2$ . Alors :

$$\forall \varepsilon > 0, \mathbb{P}(|Y - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

◁ **Éléments de preuve.**

Inégalité de Markov quadratique appliquée à...

▷

**VII.3 Loi faible des grands nombres**

Un des exemples les plus importants de convergence en probabilité est celle qui résulte du calcul de la moyenne obtenue lors d'une répétition de variables aléatoires.

**Théorème 29.7.6 (Loi faible des grands nombres, HP)**

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de v.a.r.d. mutuellement indépendantes suivant une même loi (on dit « indépendantes identiquement distribuées », ou, en abrégé : *i.i.d.*), ayant une espérance  $m$  et une variance  $\sigma^2$ . Soit  $(Z_n)_{n \in \mathbb{N}^*}$  définie par :

$$\forall n \in \mathbb{N}, Z_n = \frac{X_1 + \dots + X_n}{n}.$$

Alors  $(Z_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers la variable certaine égale à  $m$ . Plus précisément :

$$\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}(|Z_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

◁ **Éléments de preuve.**

Inégalité de B-T appliquée à  $Z_n$  dont on sait facilement déterminer l'espérance et la variance.

▷

**Théorème 29.7.7 (Théorème d'or de Bernoulli, HP)**

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables mutuellement indépendantes, suivant toutes une loi de Bernoulli de paramètre  $p$ . Soit :

$$\forall n \in \mathbb{N}^*, Z_n = \frac{X_1 + \dots + X_n}{n}.$$

Alors  $(Z_n)_{n \in \mathbb{N}^*}$  converge en probabilité vers la v.a.r. certaine égale à  $p$ . Plus précisément :

$$\forall \varepsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}(|Z_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}.$$

◁ **Éléments de preuve.**

Conséquence immédiate de la loi faible des grands nombres, en utilisant une certaine inégalité archi-classique.

▷

Ainsi, la fréquence statistique d'un événement tend vers la probabilité de réalisation de l'événement.

**Exemple 29.7.8**

On tire 1000 fois à pile ou face avec une pièce déséquilibrée dont la probabilité d'obtention de Pile est  $p$ . On obtient 570 fois Pile. Donner un intervalle  $I$  tel que la probabilité que  $p \in I$  soit supérieure à 0.9.

Nom	Paramètres	Notation	Valeurs	Loi	$\mathbb{E}(X)$	$\mathbb{V}(X)$
<b>Uniforme</b>	$n \in \mathbb{N}^*$	$\mathcal{U}(n)$	$\llbracket 1, n \rrbracket$	$\mathbb{P}(X = k) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
<b>Bernoulli</b>	$p \in ]0, 1[$	$\mathcal{B}(1, p)$	$\{0, 1\}$	$\mathbb{P}(X = 1) = p$	$p$	$pq$
<b>Binomiale</b>	$(n, p)$	$\mathcal{B}(n, p)$	$\llbracket 0, n \rrbracket$	$\mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}$	$np$	$npq$
<b>Géométrique</b>	$p$	$\mathcal{G}(p)$	$\mathbb{N}^*$	$\mathbb{P}(X = k) = pq^{k-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$
Pascal	$(r, p)$	$\mathcal{P}(r, p)$	$\llbracket r, +\infty \llbracket$	$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$
Binomiale négative	$(r, p)$	$\mathcal{J}(r, p)$	$\mathbb{N}$	$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r q^k$	$\frac{rq}{p}$	$\frac{rq}{p^2}$
Hypergéométrique	$(N, n, p)$	$\mathcal{H}(N, n, p)$	$\subset \llbracket 0, n \rrbracket$	$\mathbb{P}(X = k) = \frac{\binom{Np}{k} \binom{Nq}{n-k}}{\binom{N}{n}}$	$np$	$npq \frac{N-n}{N-1}$
Attente du 1 <sup>er</sup> succès (tirage sans remise)	$(N, p)$		$\llbracket 1, Nq + 1 \rrbracket$	$\mathbb{P}(X = k) = \frac{Np}{k} \cdot \frac{\binom{Nq}{k-1}}{\binom{N}{k}}$		
Attente du $r^e$ succès (tirage sans remise)	$(r, N, p)$		$\llbracket r, Nq + r \rrbracket$	$\mathbb{P}(X = k) = \frac{r}{k} \frac{\binom{Np}{r} \binom{Nq}{k-r}}{\binom{N}{k}}$		
<b>Poisson</b>	$\lambda \in \mathbb{R}_+$	$\mathcal{P}(\lambda)$	$\mathbb{N}$	$\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$	$\lambda$	$\lambda$

FIGURE 29.1 – Table des lois discrètes classiques