Classes préparatoires aux grandes écoles

Filière scientifique • Mathématiques

Variables aléatoires discrètes

Mouad El Assadi

Professeur agrégé de mathématiques

Ingénieur d'État en génie civil — École Hassania des Travaux Publics (EHTP)

elassadimouad1@gmail.com

Table des matières

I Loi d'une variable aléatoire discrète **2**
I.1 Généralités sur les variables aléatoires discrètes . . . 2
I.2 Loi d'une variable aléatoire discrète . . . 3
I.3 Loi de $f(X)$ . . . 4
I.4 Exemples 15.1.19 - Loi d'une somme, loi du minimum . . . 4
I.5 Loi d'un vecteur aléatoire réel discret . . . 5
I.6 Variables indépendantes . . . 6

II Espérance et variance **8**
II.1 Espérance d'une variable positive . . . 8
II.2 Espérance d'une variable complexe . . . 9
II.3 Formule de transfert et conséquences . . . 9
II.4 Moment d'ordre 2 et variance d'une variable réelle . . . 11
II.5 Covariance . . . 12
II.6 Variance d'une somme . . . 13

III Variables aléatoires à valeurs dans $\mathbb{N}$, lois usuelles **14**
III.1 Fonction génératrice d'une variable à valeurs dans $\mathbb{N}$ . . . 14
III.2 Rappels sur les lois usuelles finies . . . 16
III.3 Loi géométrique . . . 17
III.4 Loi de Poisson . . . 18
III.5 Stabilités (HP classique) . . . 18

IV Inégalités probabilistes **19**
IV.1 Inégalités de Markov et de Bienaymé-Tchebychev . . . 19
IV.2 Loi faible des grands nombres . . . 19

Dans ce chapitre, sauf mention contraire, $(\Omega, \mathcal{T}, \mathbb{P})$ désigne un espace probabilisé quelconque.
La situation à laquelle vous serez le plus fréquemment confrontés est celle d'une variable aléatoire discrète 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 discrète à valeurs dans n'importe quel ensemble E.
La définition d'une variable aléatoire X générale (non discrète) dépasse en revanche le cadre du programme, et nécessite un cadre un peu plus précis : l'ensemble E dans lequel X prend ses valeurs doit lui-même être muni d'une tribu, et X doit vérifier une propriété de « mesurabilité », qui est en quelque sorte un respect des tribus, comme la continuité est le respect des topologies.

I. Loi d'une variable aléatoire discrète

I.1 Généralités sur les variables aléatoires discrètes

Définition I.1 (Variable aléatoire discrète)**.** *Soit $(\Omega, \mathcal{T})$ un espace probabilisable. Une variable aléatoire discrète X à valeurs dans E est une application $X : \Omega \to E$ telle que :*
(i) $X(\Omega)$ (l'ensemble des valeurs possibles de X) est au plus dénombrable; (ii) pour tout $x \in X(\Omega)$, $X^{-1}(\{x\}) \in \mathcal{T}$.

Définition I.2 (Variables et vecteurs aléatoires réels discrets)**.** *- si $E = \mathbb{R}$ et X vérifie les conditions de la définition 15.1.1, on dit que X est une variable aléatoire réelle discrète (abrégé en v.a.r.d.) - si $E = \mathbb{R}^n$ et X vérifie les conditions de la définition 15.1.1, on dit que X est un vecteur aléatoire réel discret.*

Proposition I.3 (Image réciproque des parties de E)**.** *Soit $(\Omega, \mathcal{T})$ un espace probabilisable, et $X : \Omega \to E$ une variable aléatoire discrète. Alors,*

$$\forall A \in \mathcal{P}(E), X^{-1}(A) \in \mathcal{T}.$$

Éléments de preuve. C'est la stabilité par union finie ou dénombrable, du fait que $X(\Omega) \cap A$ est fini ou dénombrable. □

Remarque I.4 (Notation 15.1.4 - Événements liés à une variable aléatoire)**.** *Soit (Ω, T) un espace probabilisable, et X une variable aléatoire à valeurs dans E.*
Soit $A \in \mathcal{P}(E)$. L'événement $X^{-1}(A) \in \mathcal{T}$ est noté $(X \in A)$ ou $\{X \in A\}$ ou encore $[X \in A]$.
Soit $x \in A$. L'événement $X^{-1}(\{x\}) = (X \in \{x\})$ sera simplement noté $(X = x)$, ou $\{X = x\}$ ou $[X = x]$.

Remarque I.5 (15.1.5)**.** *Les notations $(X \in A)$ et $\{X \in A\}$ figurent toutes deux dans le programme, contrairement à $[X \in A]$ qui est néanmoins largement présente dans beaucoup d'ouvrages (en particulier pour les événements $[X = x]$). Nous laisserons désormais de côté cette notation.*

Remarque I.6 (Notation 15.1.6 - Événements liés à une v.a.r.d.)**.** *Dans le cas d'une variable aléatoire réelle discrète,*
l'événement $\{\omega \mid X(\omega) < a\}$ est noté $(X < a)$ ou $\{X < a\}$
l'événement $\{\omega \mid X(\omega) \leqslant a\}$ est noté $(X \leqslant a)$ ou $\{X \leqslant a\}$
l'événement $\{\omega \mid a < X(\omega) < b\}$ est noté $(a < X < b)$ ou $\{a < X < b\}$
l'événement $\{\omega \mid a < X(\omega) \leqslant b\}$ est noté $(a < X \leqslant b)$ ou $\{a < X \leqslant b\}$ etc.

Proposition I.7 (Stabilité par produit cartésien)**.** *Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires discrètes. Alors le couple (X, Y) est une variable aléatoire discrète.*

Proposition I.8 (Composition d'une variable aléatoire discrète)**.** *Soit X une variable aléatoire discrète sur (Ω, T), 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. $f(X)^{-1}(\{y\})$ est une union au plus dénombrable de $X^{-1}(\{x\})$. $f : X(\Omega) \to f(X)(\Omega)$ est une surjection, et $X(\Omega)$ est au plus dénombrable. □

On en déduit :

Proposition I.9 (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 Ω, λ un réel quelconque :*
1. 1 est une v.a.r.d. 2. $X+Y$ est une v.a.r.d. 3. λX est une v.a.r.d. 4. XY est une v.a.r.d.
De plus, si X est une v.a.r.d., il en est de même de $|X|$.

Éléments de preuve. 2 et 4 s'obtiennent en composant (X, Y) par la somme ou le produit. □

I.2 Loi d'une variable aléatoire discrète

Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans E. Par définition, $\{X = x\}$ est un événement, donc $\mathbb{P}(X = x)$ est donc bien défini.

Proposition I.10. *Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{T}, \mathbb{P})$ à valeurs dans E. Alors - la famille $(\{X = x\})_{x\in X(\Omega)}$ est un SCE d'événements; - la famille $(\mathbb{P}(X = x))_{x\in X(\Omega)}$ est une distribution de probabilités.*

Éléments de preuve. Le premier point est de la vérification facile (c'est en fait la partition associée à une surjection). Le deuxième point est la σ-additivité utilisée sur le SCE précédent. □

Définition I.11 (Loi d'une variable aléatoire discrète). *Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \to E$ une variable aléatoire. La loi de X est la mesure de probabilité définie sur $(X(\Omega), \mathcal{P}(X(\Omega)))$ par la distribution de probabilités discrète $(P(X = x))_{x\in X(\Omega)}$.*

Remarque I.12 (15.1.12). *Au besoin, il est possible de définir $\mathbb{P}$ sur l'espace probabilisable $(F, \mathcal{P}(F))$, pour tout ensemble F tel que $X(\Omega) \subset F$, en prolongeant la mesure de probabilité $(P(X = x))_{x\in X(\Omega)}$, en posant*

$$\forall A \in \mathcal{P}(F), \mathbb{P}(A) = \sum_{x\in A\cap X(\Omega)} \mathbb{P}(X = x).$$

Cela permet par exemple de définir $\mathbb{P}_X$ sur $(E, \mathcal{P}(E))$ tout entier.

Proposition I.13 (Expression de la loi d'une v.a. discrète). *Soit $A \in \mathcal{P}(X(\Omega))$*

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

cette somme étant bien définie.

Éléments de preuve. Écrire A comme union de singletons. □

Remarque I.14 (15.1.14). *En définissant $\mathbb{P}_X$ sur $(E, \mathcal{P}(E))$, on peut plus généralement écrire*

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

Remarque I.15 (Notation 15.1.15 - Variables de même loi)**.** *On note $X \sim Y$ pour désigner le fait que X et Y suivent la même loi.*
Plus précisément, étant données deux variables aléatoires $X : \Omega_1 \to E$ et $Y : \Omega_2 \to E$, définies sur les espaces probabilisés $(\Omega_1, \mathcal{T}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{T}_2, \mathbb{P}_2)$, $X \sim Y$ si et seulement si les mesures $\mathbb{P}_X$ et $\mathbb{P}_Y$ définies sur E vérifient

$$\mathbb{P}_X = \mathbb{P}_Y.$$

1. Contrairement à l'exemple précédent, la relation $X \sim Y$ ne nécessite pas que X et Y soient définis sur le même espace probabilisé. 2. La relation $\sim$ est clairement réflexive, symétrique et transitive. Mais dans le cadre du programme, il est un peu abusif d'en déduire que c'est une relation d'équivalence, car son « domaine » de définition n'est pas très bien défini. Il n'est pas clair que l'ensemble des variables aléatoires discrètes soit bien un ensemble, car, même en fixant E, la relation nécessite de faire varier Ω. Or, la collection des ensembles n'est pas un ensemble.

On peut bien évidemment aussi définir des lois conditionnelles. En effet, la définition d'une variable aléatoire $X : \Omega \to E$ n'implique que l'espace probabilisable (Ω, T), et non l'espace probabilisé $(\Omega, T, \mathbb{P})$. Ainsi, étant donné un événement $B \in T$, on peut munir (Ω, T) de la mesure de probabilité conditionnelle $\mathbb{P}_B$ et définir la loi de la variable X en référence à l'espace probabilisé $(\Omega, T, \mathbb{P}_B)$. Concrètement, la loi conditionnelle de X est donc définie par la distribution de probabilités

$$\forall x \in X(\Omega), \mathbb{P}_{B,X}(x) = \mathbb{P}_B(X = x).$$

I.3 Loi de $f(X)$

Proposition I.16 (Loi de $f(X)$)**.** *1. Soit X une variable aléatoire discrète sur $(\Omega, T, \mathbb{P})$, à valeurs dans E, et $f : E \to F$. La loi de $f(X)$, définie sur $(F, \mathcal{P}(F))$ (ou seulement sur $(f(X(\Omega)), \mathcal{P}(f(X(\Omega))))$, est donnée par*

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

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

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

2. En particulier,

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

Éléments de preuve. 1. $\{X \in f^{-1}(A)\} = \{f(X) \in A\}$
2. $\{f(X) = x\} = \{X \in f^{-1}(x)\}$ □

I.4 Exemples 15.1.19 - 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.

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

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

Corollaire I.17 (Stabilité de $\sim$ par image directe)**.** *Soit $X : \Omega_1 \to E$ et $Y : \Omega_2 \to E$ telles que $X \sim Y$, et $f : E \to F$. Alors $f(X) \sim f(Y)$.*

I.5 Loi d'un vecteur aléatoire réel discret

Proposition I.18 (Caractérisation d'un vecteur aléatoire par ses coordonnées)**.** *Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \to \mathbb{R}^n$ une application. On note $X = (X_1, \ldots, X_n)$ les coordonnées de X. Les propositions suivantes sont équivalentes:*
(i) X est une variable aléatoire discrète (donc un vecteur aléatoire discret) (ii) pour tout $i \in [1, n]$, X_i est une variable aléatoire réelle discrète.

Remarque I.19 (Notation 15.1.22 - $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n)$)**.** *Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire discret, à valeurs dans $\mathbb{R}^n$. Pour alléger les notations, on note $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n)$ au lieu de $\mathbb{P}([[X_1 \in A_1] \cap \cdots \cap \{X_n \in A\}])$. En particulier, on utilisera la notation $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$.*

Pour éviter l'abus de notations techniques stériles, nous donnons les définitions dans le cas de couples de variables aléatoires, elles s'étendent facilement au cas de vecteurs aléatoires à davantage de coordonnées. Pour commencer, on peut remarquer que si (X, Y) est un couple de variables aléatoires réelles discrètes, et $V = (X, Y)$, alors

$$V(\Omega) \subset X(\Omega) \times Y(\Omega).$$

Définition I.20 (Loi conjointe, lois marginales)**.** *Soit (X, Y) un couple de variables aléatoires réelles discrètes.*
La loi conjointe du couple (X, Y) est la loi du vecteur aléatoire $V = (X, Y)$. Il s'agit donc d'une mesure de probabilité $\mathbb{P}(X, Y)$ définie par la distribution de probabilités

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

La première loi marginale du couple (X, Y) est la loi de X.
La seconde loi marginale du couple (X, Y) est la loi de Y.
De la même façon, on parle de la loi conjointe, et de la k-ième loi marginale d'un n-uplet $(X_1, \ldots, X_n)$.

De façon évidente :

Proposition I.21 (Expression d'une loi marginale)**.** *Soit (X, Y) un couple de variables aléatoires réelles discrètes. La loi conjointe détermine les lois marginales par les formules :*

$$\begin{cases} \forall x \in X(\Omega), & \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y). \\ \forall y \in Y(\Omega), & \mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y). \end{cases}$$

De façon similaire, la k-ième loi marginale d'un n-uplet $(X_1, \ldots, X_n)$ s'obtient en sommant sur toutes les valeurs possibles prises par les $n - 1$ autres variables aléatoires.

Remarque I.22 (Avertissement 15.1.25)**.** *Les lois marginales sont déterminées par la loi conjointe, mais la réciproque est fausse : les lois marginales ne déterminent pas la loi conjointe!*

Exemple I.23 (15.1.26)**.** *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.*

I.6 Variables indépendantes

Proposition I.24 (Variables aléatoires discrètes indépendantes)**.** *Soit X et Y deux variables aléatoires discrètes définies sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans E_1 et E_2 respectivement. Les propositions suivantes sont équivalentes:*
(i) Pour tout $(A, B) \in \mathcal{P}(E_1) \times \mathcal{P}(E_2)$ $\mathbb{P}_{(X,Y)}(A \times B) = \mathbb{P}_X(A)\mathbb{P}_Y(B)$
(ii) Pour tout $(A, B) \in \mathcal{P}(E_1) \times \mathcal{P}(E_2)$, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants, i.e.

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B);$$

(iii) Pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants, i.e.

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

Dans ce cas, on dit que X et Y sont indépendantes, et on note $X \perp Y$.

Remarque I.25 (15.1.28)**.** *Conformément à la remarque faite lors de la définition de la loi d'une variable aléatoire, on peut se contenter de considérer $(A, B) \in \mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega))$.*

Proposition I.26 (Famille de variables mutuellement indépendantes)**.** *Soit $(X_k)_{k\in K}$ une famille quelconque de variables aléatoires discrètes sur $(\Omega, \mathcal{T}, \mathbb{P})$, X_k étant à valeurs dans E_k. Les propositions suivantes sont équivalentes:*
(i) pour toute famille $(A_k)_{k\in K}$ telle que pour tout $k \in K$, $A_k \subset E_k$, les événements $[X_k \in A_k]$, $k \in K$, sont indépendants;
(ii) pour tout $J \in \mathcal{P}_f(K)$ et toute famille $(A_j)_{j\in J}$ telle que pour tout $j \in J$, $A_j \subset E_j$,

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

(iii) pour toute famille $(x_k)_{k\in K}$ telle que pour tout $k \in K$, $x_k \in X_k(\Omega)$, les événements $\{X_k = x_k\}$, $k \in K$, sont indépendants ;
(iv) pour tout $J \in \mathcal{P}_f(K)$ et toute famille $(x_j)_{j\in J}$ telle que pour tout $j \in J$, $x_j \in X_j(\Omega)$,

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

Dans ce cas, on dit que les variables aléatoires X_k, $k \in K$, sont indépendantes (ou mutuellement indépendantes).

Proposition I.27 (Sous-famille d'une famille de variables indépendantes)**.** *Si la famille $(X_k)_{k\in K}$ est une famille de variables aléatoires discrètes indépendantes, alors, pour tout $J \subset K$, $(X_j)_{j\in J}$ est une famille de variables aléatoires discrètes indépendantes.*

Corollaire I.28 (Indépendance 2 à 2)**.** *Si $(X_k)_{k\in K}$ est une famille de variables aléatoires discrètes indépendantes, alors les X_k sont 2 à 2 indépendants. La réciproque est fausse.*

En revanche, on peut caractériser l'indépendance par l'indépendance des sous-familles finies.

Proposition I.29 (Caractérisation de l'indépendance par les familles finies)**.** *Soit $(X_k)_{k\in K}$ une famille de variables aléatoires discrètes. Les propositions suivantes sont équivalentes:*
(i) $(X_k)_{k\in K}$ est une famille de variables indépendantes ; (ii) pour tout $J \in \mathcal{P}_f(K)$, $(X_j)_{j\in J}$ est une famille de variables indépendantes.

Proposition I.30 (Fonctions de variables indépendantes)**.** *Soit $X \perp Y$ deux v.a.r.d. indépendantes, et f et g deux fonctions définies respectivement sur $X(\Omega)$ et $Y(\Omega)$. Alors $f(X) \perp g(Y)$.*

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

Théorème I.31 (Lemme des coalitions, pour 2 coalitions)**.** *Soit $1 \leq m < n$, et $(X_1, \dots, X_m, \dots X_n)$ une famille de variables aléatoires discrètes indépendantes. Soit f et g deux fonctions définies sur $X_1(\Omega) \times \dots \times X_m(\Omega)$ et $X_{m+1}(\Omega) \times \cdots \times X_n(\Omega)$ 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 exprimant tout en fonction des répartitions de probabilités des X_i. □

Par récurrence, on en déduit la version générale, qu'on exprime volontairement sous une forme un peu lâche.

Théorème I.32 (Lemme des coalitions)**.** *Soit $(X_k)_{k \in K}$ une famille de variables aléatoires discrètes indépendantes, et $(K_i)_{i \in I}$ une famille de sous-ensembles finis deux à deux disjoints de K. Soit $(Y_i)_{i \in I}$ une famille de variables aléatoires, telle que pour tout $i \in I$, Y_i s'écrive comme fonction des X_k, $k \in K_i$ (et pas des autres). Alors $(Y_i)_{i \in I}$ est une famille de variables aléatoires discrètes indépendantes.*

Éléments de preuve. Le cas K infini et I fini s'obtient par récurrence à partir du lemme des 2 coalitions (seul un nombre fini de X_k intervenant, puisque les parties K_i sont finies).
Le cas I infini provient alors du fait que l'indépendance d'une famille équivaut à l'indépendance de toutes ses sous-familles finies. □

On admet le résultat suivant, qui permet d'assurer l'existence d'espaces probabilisés modélisant des situations un peu complexes, notamment infinies. Par exemple, la modélisation précise d'une succession infinie de tirages à Pile ou Face indépendants n'est pas chose facile. L'univers Ω est naturellement $\{0,1\}^{\mathbb{N}}$, mais cet ensemble est en bijection avec $\mathbb{R}$, et le choix de la tribu $\mathcal{P}(\Omega)$ ne permet pas ici de modéliser correctement le tirage à pile ou face.
En effet, on peut montrer que toute mesure de probabilité sur $(\mathbb{R}, \mathcal{P}(\mathbb{R}), \mathbb{P})$ est déterminée par une répartition de probabilités discrète. Or, pour tout $\omega \in \Omega$, on obtient alors dans cet espace probabilisé $\mathbb{P}(\omega) = 0$ (par le même calcul que la probabilité de n'obtenir que des Piles). Ceci contredit le fait que $\mathbb{P}$ soit déterminée par une répartition de probabilités discrète.
La construction d'un modèle convenable est difficile, et n'est pas un objectif du programme. Nous utiliserons donc le théorème suivant, admis, pour justifier l'existence d'un modèle convenable.

Théorème I.33 (Espace probabilisé modélisant une suite de lois, admis)**.** *Soit E un ensemble et $(P_n)_{n \in \mathbb{N}}$ une suite de lois de probabilités discrètes sur E (ce qui équivaut à la donnée d'une suite de distributions de probabilités discrètes).*
Alors il existe un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et une suite $(X_n)_{n \in \mathbb{N}}$ de variables indépendantes, définies sur $(\Omega, \mathcal{T}, \mathbb{P})$, à valeurs dans E, et telle que

$$\forall n \in \mathbb{N}, \mathbb{P}_{X_n} = P_n.$$

Éléments de preuve. Démonstration hors-programme. □

Il est fréquent, pour étudier les caractéristiques d'une expérience, de la répéter une infinité de fois dans les mêmes conditions, de façon indépendante, et d'observer le résultat sous forme de valeur prise par une variable aléatoire.

Définition I.34 (variables i.i.d.)**.** *Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. On dit que la suite (X_n) est indépendante identiquement distribuée (en abrégé officiel i.i.d.), si les X_n sont indépendants, et suivent la même loi (i.e. pour tout $(i,j) \in \mathbb{N}^2$ $X_i \sim X_j$).*
Si les X_i suivent une loi classique, on dira, par exemple, que $(X_n)_{n \in \mathbb{N}}$ est une suite i.i.d. de variables suivant une loi binomiale $B(n,p)$.

Corollaire I.35 (Modélisation d'une répétition infinie d'expériences indépendantes)**.** *Soit P une loi de probabilités discrète sur E. Alors il existe un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et $(X_n)_{n\in\mathbb{N}}$ une suite i.i.d. de variables de loi P.*

En appliquant cela à une loi de Bernoulli $B(p)$, on obtient l'existence d'un espace probabilisé modélisant une succession infinie de tirages (éventuellement déséquilibrés) à Pile ou Face.

Corollaire I.36 (Espace probabilisé modélisant des tirages à P/F)**.** *Il existe un espace probabilisé modélisant une succession infinie de tirages à Pile ou Face indépendants, avec une probabilité p d'obtenir Pile à chaque tirage.*

II. Espérance et variance

On étend les définitions vues en première année au cas des variables discrètes. Dans toute cette section, $(\Omega, \mathcal{T}, \mathbb{P})$ désigne un espace probabilisé qu'on ne réintroduira pas à chaque énoncé ; sauf mention explicite du contraire, les variables sont définies sur $(\Omega, \mathcal{T}, \mathbb{P})$ et à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$.

II.1 Espérance d'une variable positive

Définition II.1 (Espérance d'une variable positive)**.** *Soit X une variable aléatoire à valeurs dans $\overline{\mathbb{R}}_+$. L'espérance de X est définie par*

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

où, par convention, $+\infty \times 0 = 0$. L'espérance est toujours bien définie dans $\overline{\mathbb{R}}_+$.

Remarque II.2 (15.2.2)**.** *Si $+\infty \in X(\Omega)$: si $\mathbb{P}(X = +\infty) > 0$, alors $\mathbb{E}(X) = +\infty$; si $\mathbb{P}(X = +\infty) = 0$, alors X a même espérance que sa restriction à $\Omega' = \Omega\backslash\{X = +\infty\}$.*

Plus généralement,

Proposition II.3 (Espérance de variables de même loi)**.** *Si X et Y sont des variables à valeurs dans $\overline{\mathbb{R}}_+$, et $X \sim Y$, alors $\mathbb{E}(X) = \mathbb{E}(Y)$.*

Remarque II.4 (15.2.4)**.** *Une variable aléatoire X positive peut admettre une espérance infinie même si X ne prend que des valeurs finies.*

Exemple II.5 (15.2.5)**.**

$$X(\Omega) = \mathbb{N}^* \text{ et } \forall n \in \mathbb{N}^*, \mathbb{P}(X = n) = \frac{6}{\pi^2 n^2}.$$

Proposition II.6 (Positivité et stricte positivité)**.** *Soit X une variable aléatoire à valeurs dans $\overline{\mathbb{R}}_+$.*
1. $\mathbb{E}(X) \geqslant 0$ 2. $\mathbb{E}(X) = 0 \iff X \sim 0$ (i.e. X est presque sûrement nulle).

Dans le cas de variables à valeurs dans $\mathbb{N}$, on dispose d'une autre description possible de $\mathbb{E}(X)$.

Proposition II.7 (Réexpression de l'espérance lorsque $X(\Omega) \subset \mathbb{N} \cup \{+\infty\}$)**.** *Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N} \cup \{+\infty\}$. Alors*

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

II.2 Espérance d'une variable complexe

Définition II.8 (Espérance). *Soit X une variable aléatoire discrète à valeurs dans $\mathbb{C}$. On dit que X admet une espérance finie 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 II.9 (15.2.9). *Il arrive souvent que l'ensemble dénombrable $X(\Omega)$ soit énuméré, c'est-à-dire décrit sous forme d'une suite $(x_n)_{n\in\mathbb{N}}$. La condition d'existence d'une espérance finie équivaut dans ce cas à la convergence absolue de la série $\sum x_n\mathbb{P}(X=x_n)$.*
On prendra garde au fait que la semi-convergence n'est pas suffisante pour assurer la sommabilité. En effet, comme on l'a vu, la valeur de la somme est dans ce cas fortement dépendante de l'ordre de sommation. Or l'ordre de sommation défini par l'énumération $(x_n)_{n\in\mathbb{N}}$ est purement arbitraire, et n'a pas de raison d'être privilégié par rapport à un autre.

Exemple II.10 (15.2.10).

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

Remarque II.11 (15.2.11). *Lorsque $X\geqslant 0$, $X\in L^1 \iff \mathbb{E}(X)<+\infty$.*
En revanche, lorsque X n'est pas positive, cela n'a pas de sens de donner une valeur infinie à $\mathbb{E}(X)$. Comme dans le cas positif, on peut comparer l'espérance de deux variables de même loi.

Proposition II.12 (Espérance de variables de même loi). *Soit $X\sim Y$ deux variables discrètes à valeurs complexes de même loi. Alors :*
- X admet une espérance finie si Y admet une espérance finie; - dans ce cas, $\mathbb{E}(X)=\mathbb{E}(Y)$.

II.3 Formule de transfert et conséquences

Théorème II.13 (Formule de transfert, version positive). *Soit X une variable aléatoire à valeurs dans E et $f : X(\Omega)\to\overline{\mathbb{R}}_+$. Alors, dans $\overline{\mathbb{R}}_+$, l'égalité suivante est toujours valide :*

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

Éléments de preuve. Partir de l'espérance de $f(X)$, et, par propriété d'associativité positive, regrouper les valeurs de x selon les parties $\{f(X)=y\}$. □

Théorème II.14 (Formule de transfert, version complexe). *Soit X une variable aléatoire à valeurs dans E et $f : X(\Omega)\to\mathbb{C}$. Les propositions suivantes sont équivalentes:*
(i) $f(X)\in L^1$; (ii) $(f(x)\mathbb{P}(X=x))_{x\in X(\Omega)}$ est sommable.
Dans ce cas,

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

Exemple II.15 (15.2.15). *Si X est une variable discrète réelle. En cas de sommabilité,*

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

On dit dans ce cas que $X\in L^k$, et $\mathbb{E}(X^k)$ est appelé moment d'ordre k de X.

Remarque II.16 (15.2.16). *Si (X,Y) est un couple de variables aléatoires définies sur E et F respectivement, et $f : E \times F \to \mathbb{C}$, on peut considérer (X,Y) comme une variable aléatoire à valeurs dans $E \times F$, ce qui permet d'utiliser la formule de transfert : $f(X,Y)$ admet une espérance finie si $(f(x,y)\mathbb{P}(X=x,Y=y))_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est sommable. Dans ce cas,*

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

La formule de transfert permet d'obtenir des propriétés importantes de l'espérance.

Proposition II.17 (Inégalité triangulaire). *Soit X une variable aléatoire à valeurs complexes. Alors*

$$X \in L^1 \iff \mathbb{E}(|X|) < +\infty,$$

et dans ce cas,

$$|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|).$$

Proposition II.18 (Linéarité de $\mathbb{E}$, cas positif). *Soit X et Y deux variables aléatoires discrètes à valeurs dans $\overline{\mathbb{R}}_+$, et $\lambda \in \mathbb{R}_+^*$. Alors, dans $\overline{\mathbb{R}}_+$,*

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

Proposition II.19 (Linéarité de $\mathbb{E}$, cas complexe). *Soit X et Y deux variables aléatoires discrètes à valeurs dans $\mathbb{C}$, et $\lambda \in \mathbb{C}$. Si $X \in L^1$ et $Y \in L^1$, alors $\lambda X + Y \in L^1$ et dans ce cas,*

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

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. □

La propriété de positivité de la proposition suivante a déjà été démontrée.

Proposition II.20 (Positivité et croissance de l'espérance). *1. Soit (X,Y) deux variables à valeurs dans $\mathbb{R}_+$. Alors, dans $\mathbb{R}_+$,*

$$X \leqslant Y \Rightarrow \mathbb{E}(X) \leqslant \mathbb{E}(Y).$$

2. Soit $X, Y \in L^1$ deux variables aléatoires réelles discrètes admettant une espérance.
Si $X \geqslant 0$ ps, alors $\mathbb{E}(X) \geqslant 0$, avec égalité ssi $X = 0$ ps.
Si $X \geqslant Y$ ps, alors $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$.

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

Corollaire II.21 (L^1 par majoration). *Soit X et Y deux variables à valeurs dans $\mathbb{C}$. Si $|X| \leqslant |Y|$ et $Y \in L^1$, alors $X \in L^1$.*

Proposition II.22 (Espérance d'un produit de variables indépendantes). *Soit X et Y deux variables à valeurs dans $\mathbb{C}$. Si X et Y sont L^1, alors $XY \in L^1$ et*

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

Par récurrence immédiate, si $X_1, \dots, X_n$ sont des variables indépendantes et L^1, alors $X_1 \cdots X_n$ est aussi L^1, et

$$\mathbb{E}(X_1 \dots X_n) = \prod_{k=1}^{n} \mathbb{E}(X_k).$$

II.4 Moment d'ordre 2 et variance d'une variable réelle

Dans ce paragraphe et le suivant, les variables aléatoires sont discrètes et réelles.
Comme dans le cas des variables finies, la valeur moyenne n'est pas une donnée suffisante pour avoir une bonne compréhension d'une variable aléatoire. 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.
Pour des raisons techniques, il est plus facile de travailler avec la moyenne quadratique $\mathbb{E}((X - \mathbb{E}(X))^2)$ de la distance à $\mathbb{E}(X)$, plutôt que la moyenne arithmétique $\mathbb{E}(|X - \mathbb{E}(X)|)$. En effet, ce choix nous ramène à des méthodes bilinéaires.
On a déjà défini les moments d'ordre k. Les seuls au programme sont les moments d'ordre 1 (l'espérance) et 2.

Définition II.23 (Moment d'ordre 2)**.** *Soit X une variable aléatoire réelle discrète.*
- Le moment d'ordre 2 de X est l'espérance de X^2. Ce moment d'ordre 2 est toujours bien défini dans $\mathbb{R}_+$.
- Ainsi, d'après la formule de transfert,

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

- On dit que X admet un moment d'ordre 2 fini (et on note $X \in L^2$) si $\mathbb{E}(X^2) < +\infty$.

Proposition II.24 (Espace des variables L^2 sur Ω)**.** *1. Soit X et Y deux variables aléatoires réelles discrètes sur Ω. Si X et Y sont dans L^2, alors $XY \in L^1$.*
2. L'ensemble $L^2(\Omega, \mathbb{R})$ des variables aléatoires définies sur Ω et admettant un moment d'ordre 2 fini est un espace vectoriel sur $\mathbb{R}$.

Corollaire II.25 (Espérance d'une variable L^2)**.** *Soit X une variable aléatoire réelle discrète. Si $X \in L^2$, alors $X \in L^1$.*

Éléments de preuve. Prendre $Y = 1$. □

Théorème II.26 (Inégalité de Cauchy-Schwarz pour l'espérance)**.** *- Pour toutes variables aléatoires réelles discrètes X et Y dans L^2,*

$$\mathbb{E}(XY)^2 \leqslant \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

- Dans l'inégalité précédente, le cas d'égalité est obtenu si et seulement si X et Y sont colinéaires presque sûrement, donc s'il existe des constantes λ et μ réelles telles que $\lambda X + \mu Y \sim 0$.

Éléments de preuve. C'est l'inégalité de Cauchy-Schwarz pour la forme bilinéaire symétrique positive $(X, Y) \mapsto \mathbb{E}(XY)$ définie sur $L^2(\Omega, \mathbb{R})$.
On rappelle que l'inégalité de Cauchy-Schwarz pour une forme bilinéaire symétrique positive ϕ peut être obtenue sans l'hypothèse de séparation, et que le cas d'égalité est alors obtenu pour le couple (x, y) si et seulement s'il existe λ et μ tels que $\phi(\lambda x + \mu y, \lambda x + \mu y) = 0$. □

Afin d'alléger les énoncés, nous ne rappelons plus par la suite à chaque fois que les variables considérées sont réelles discrètes (ce que nous sous-entendrons par l'écriture $X \in L^2$ que nous n'avons définie que dans ce cas).

Définition II.27 (Variance et écart-type d'une variable aléatoire)**.** *Soit $X \in L^2$. Alors, en vertu de ce qui précède, $X - \mathbb{E}(X)$ est bien défini et L^2. On définit alors :*
- la variance de X :

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

- l'écart type de X :

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

Proposition II.28 (Positivité et stricte positivité de $\mathbb{V}$)**.** *Soit $X \in L^2$. Alors $\mathbb{V}(X) \geqslant 0$, l'égalité étant obtenue si et seulement si X est constante presque sûrement.*

Dans la pratique, on calcule souvent la variance avec le résultat suivant.

Théorème II.29 (Formule de König-Huygens)**.** *Soit $X \in L^2$. Alors*

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

Proposition II.30 (Quadraticité et invariance par translation de la variance)**.** *Soit a et b deux réels et $X \in L^2$. Alors $aX + b \in L^2$, et*

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

Il arrive souvent de normaliser les variables aléatoires, afin de pouvoir les comparer. Il faut dans ce cas à la fois normaliser la valeur moyenne et la dispersion.

Définition II.31 (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 II.32 (Variable centrée réduite associée à X)**.** *Soit $X \in L^2$ telle que $\sigma(X) > 0$. Il existe d'uniques 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 $\mathbb{V}$. □

II.5 Covariance

Pour l'instant, contrairement au cas de l'espérance, nous n'avons pas donné de moyen de calculer la variance d'une somme. Nous avons en effet besoin pour cela d'un outil supplémentaire, qui nous permet de mesurer la dépendance entre deux variables.
En effet on comprend bien que la variance de $X + Y$ va fortement dépendre de la corrélation entre X et Y. Pour prendre deux cas extrêmes, supposons que X et Y suivent une même loi symétrique par rapport à 0 :
- si $X = Y$, on aura alors $\mathbb{V}(X+Y) = \mathbb{V}(2X) = 4\mathbb{V}(X)$; - si $X = -Y$, on aura alors $\mathbb{V}(X+Y) = 0$.
Ainsi, la seule donnée des lois de X et Y ne peut pas suffire à déterminer $\mathbb{V}(X+Y)$. La covariance $\mathrm{cov}(X, Y)$ va nous permettre de mesurer la propension de deux variables à évoluer de façon parallèle ou opposée.

Définition II.33 (Covariance)**.** *Soit X et Y deux variables L^2. Alors $(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \in L^1$ et on définit la covariance de X et Y par :*

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

Remarque II.34 (Interprétation de la covariance). *Si* $\mathrm{cov}(X,Y) > 0$, $X - \mathbb{E}(X)$ *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 :*
$\mathrm{cov}(X,Y) > 0$ *signifie que* X *et* Y *ont tendance à évoluer parallèlement.*
$\mathrm{cov}(X,Y) < 0$ *signifie que* X *et* Y *ont tendance à avoir une évolution opposée.*

Définition II.35 (Variables décorrélées). *Les variables* X *et* Y *sont dites décorrélées lorsque* $\mathrm{cov}(X,Y) = 0$.

Proposition II.36 (Propriétés de la covariance). *Soit* $X, Y \in L^2$.
1. $\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
2. $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ *(symétrie)*
3. cov *est bilinéaire sur l'espace des v.a.r.d. sur* $(\Omega, \mathcal{T}, \mathbb{P})$ *admettant un moment d'ordre 2.*
4. $\mathrm{cov}(X,X) = \mathbb{V}(X)$
5. $\mathrm{cov}(1,X) = 0$
6. Si X *et* Y *sont indépendantes,* $\mathrm{cov}(X,Y) = 0$.

Éléments de preuve. Assez immédiat par linéarité de l'espérance. □

Remarque II.37 (Avertissement 15.2.38). *La réciproque de la dernière propriété est fausse : il existe des variables décorrélées, mais non indépendantes. Voir les exercices pour des exemples.*

Les propriétés 2, 3 et 4 permettent d'affirmer que cov est une forme bilinéaire symétrique positive, de forme quadratique associée égale à la variance. En particulier, on dispose d'une deuxième inégalité de Cauchy-Schwarz :

Théorème II.38 (Inégalité de Cauchy-Schwarz pour la covariance, HP). *Soit* $X, Y \in L^2$. *Alors*

$$|\mathrm{cov}(X,Y)| \leqslant \sigma(X)\sigma(Y),$$

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* $aX + bY + c = 0$*).*

Remarque II.39 (15.2.40). *Sans être explicitement au programme, cette inégalité peut être considérée comme étant au programme, puisqu'il ne s'agit de rien d'autre que l'inégalité de Cauchy-Schwarz pour l'espérance appliquée aux variables centrées* $X - \mathbb{E}(X)$ *et* $Y - \mathbb{E}(Y)$.

II.6 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 II.40 (Formule de polarisation). *Soit* $X, Y \in L^2$. *Alors :*

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

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

Proposition II.41 (Variance d'une somme). *1. Si* $X, Y \in L^2$, *alors* $X + Y \in L^2$, *et :*

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

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

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

Proposition II.42 (Variance de $X_1+\cdots+X_n$)**.** *1. Soit $X_1,\dots,X_n \in L^2$; alors $X_1+\cdots+X_n \in L^2$ et :*

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

2. Si de plus les X_i sont 2 à 2 indépendantes, ou plus généralement 2 à 2 décorrélées,

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

On peut réexprimer cette formule matriciellement.

Proposition II.43 (Réexpression matricielle de la variance d'une somme)**.** *Soit $C = \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$. Alors : $\mathbb{V}(X_1+\cdots+X_n) = C^\top\cdot V(X_1,\dots,X_n)\cdot C$,*
où $V(X_1,\dots,X_n)$ est la matrice ci-dessous, appelée matrice des variances-covariances :

$$V(X_1,\dots,X_n) = (\operatorname{cov}(X_i,X_j))_{(i,j)\in[\![1,n]\!]^2}.$$

Ainsi, la variance est la somme des coefficients de la matrice des variances-covariances. Ce point de vue simplifie souvent les manipulations sommatoires, notamment lorsque l'expression des covariances suit un motif en diagonales (constante sur chaque diagonale par exemple). Les considérations combinatoires deviennent plus évidentes sur la représentation matricielle.

III. Variables aléatoires à valeurs dans $\mathbb{N}$, lois usuelles

Nous voyons maintenant les lois classiques au programme. Avant de faire leur étude, on remarque que toutes les lois classiques sont à valeurs dans $\mathbb{N}$. Dans ce contexte on dispose d'un outil calculatoire efficace, apprécié également dans un contexte combinatoire.

III.1 Fonction génératrice d'une variable à valeurs dans $\mathbb{N}$

Définition III.1 (Série et fonction génératrice d'une variable à valeurs dans $\mathbb{N}$)**.** *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La série génératrice de X est la série entière $\sum_{n=0}^{+\infty}\mathbb{P}(X=n)t^n$. La fonction génératrice de X est la somme de la série génératrice et est notée G_X :*

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

pour tout t dans le domaine de convergence de la série génératrice.

Proposition III.2 (Rayon de convergence de la série génératrice)**.** *La série génératrice $\sum\mathbb{P}(X=n)t^n$ converge normalement sur $\overline{B}(0,1)\subset\mathbb{C}$. En particulier, son rayon R vérifie $R\geq 1$.*

En utilisant la formule de transfert, on obtient alors :

Corollaire III.3 (Réexpression de G_X comme une espérance)**.** *Pour tout $t\in\dot{B}(0,R)\cup\overline{B}(0,1)$,*

$$G_X(t) = \mathbb{E}(t^X).$$

Corollaire III.4 (Régularité de G_X)**.** *1. La fonction génératrice G_X est de classe $\mathcal{C}^\infty$ sur $\dot{B}(0,R)$ (donc en particulier sur $\dot{B}(0,1)$).*
2. La fonction génératrice est continue sur $\overline{B}(0,1)$.

Remarque III.5 (15.3.5)**.** *Si $R>1$, le point 2 est redondant. En revanche, il apporte une information nouvelle si $R=1$.*

Les fonctions génératrices donnent un outil calculatoire efficace d'étude des variables aléatoires. En particulier, elles permettent de reconnaître et comparer des variables aléatoires. En effet :

Théorème III.6 (Détermination de la loi de X par sa fonction génératrice)**.** *1. Soit X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. Si $G_X = G_Y$, alors $X \sim Y$.*
2. Plus précisément,

$$\forall n \in \mathbb{N}, \mathbb{P}(X = n) = \frac{G_X^{(n)}(0)}{n!}.$$

La fonction génératrice donne également un accès direct à l'espérance, et presque direct à la variance.

Théorème III.7 (Expression de l'espérance à l'aide de la fonction génératrice)**.** *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, et G_X sa fonction génératrice, considérée comme fonction d'une variable réelle. Les propositions suivantes sont équivalentes:*
(i) $X \in L^1$ (ii) G_X est dérivable en 1.
Dans ce cas, on a $\mathbb{E}(X) = G'_X(1)$.

Éléments de preuve. Si le rayon R de la série génératrice vérifie $R > 1$, R est aussi le rayon de la série dérivée, et cela résulte des propriétés de dérivation des séries entières. Sinon, le résultat est un corollaire du théorème 13.2.15 (la dérivabilité correspond dans ce cas à la dérivabilité à gauche).
Comme le théorème 13.2.15 n'est pas explicitement au programme, on rappelle que le sens direct $(i) \implies (ii)$ est une application simple du théorème d'Abel radial. La démonstration du sens réciproque $(ii) \implies (i)$ n'est pas exigible. □

Théorème III.8 (Expression de la variance à l'aide de la fonction génératrice)**.** *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Les propositions suivantes sont équivalentes:*
(i) $X \in L^2$ (ii) G_X est deux fois dérivable en 1.
Dans ce cas, on a $\mathbb{V}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.

Éléments de preuve. $X \in L^2$ si $X(X-1) \in L^1$. Cela correspond, par formule de transfert, à la dérivée seconde de la série génératrice.
Ensuite utiliser la formule de König-Huygens. □

Une dernière propriété importante est l'accès facile à la loi d'une somme de variables indépendantes.

Théorème III.9 (Fonction génératrice d'une somme de variables indépendantes)**.** *1. Soit X et Y deux variables aléatoires indépendantes, à valeurs dans $\mathbb{N}$, et R_X et R_Y les rayons de convergence de leurs séries génératrices. Alors*

$$\forall t \in \dot{B}(0, \min(R_X, R_Y)) \cup \overline{B}(0,1), G_{X+Y}(t) = G_X(t)G_Y(t).$$

2. Soit $X_1, \ldots, X_n$ des variables indépendantes à valeurs dans $\mathbb{N}$, et R_i le rayon de la série génératrice de X_i. Alors

$$\forall t \in \dot{B}(0, \min(R_i)) \cup \overline{B}(0,1), G_{X_1+\cdots+X_n}(t) = G_{X_1}(t) \cdots G_{X_n}(t).$$

Éléments de preuve. Le point 1 par produit de Cauchy. Le point 2 par récurrence. □

III.2 Rappels sur les lois usuelles finies

On rappelle rapidement les lois usuelles finies classiques, en complétant leur étude par la description de leur fonction génératrice. On présente une alternative au calcul direct de leur espérance et de leur variance, en se servant des séries génératrices, et on se sert des fonctions génératrices également pour justifier une propriété de stabilité.

Définition III.10 (Loi uniforme)**.** *Soit $n \in \mathbb{N}^*$. Une v.a.r. X suit la loi uniforme de paramètre n si $X(\Omega) = [1, n]$ et si :*

$$\forall k \in [1, n], \mathbb{P}(X = k) = \frac{1}{n}.$$

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

Proposition III.11 (Fonction génératrice, espérance et variance d'une loi uniforme)**.** *Soit $X \sim \mathcal{U}(n)$. Alors*

$$\forall t \in \mathbb{R}, G_X(t) = \frac{t(1-t^n)}{n(1-t)} \quad (\text{pour } t \neq 1), \quad G_X(1) = 1,$$

et :

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

Éléments de preuve. C'est juste pour refaire le calcul avec la fonction génératrice, qui donne une version plutôt plus compliquée ici que la méthode directe. □

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 :

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

Proposition III.12 (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 $[a, b]$, elle vérifie :

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

Définition III.13 (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 :*

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

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

Proposition III.14 (Espérance et variance d'une loi de Bernoulli)**.** *Soit $X \sim \mathcal{B}(p)$. Alors*

$$\forall t \in \mathbb{R}, G_X(t) = (1-p) + tp, \qquad \mathbb{E}(X) = p \qquad \text{et} \qquad \mathbb{V}(X) = pq.$$

Éléments de preuve. Un peu trivial... Mais le vérifier aussi avec la série génératrice. □

Exemple III.15 (Fonction indicatrice d'un événement)**.** *Étude de $X(\omega) = \mathbf{1}_A(\omega)$, $A \in \mathcal{P}(\Omega)$.*

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

Proposition III.16 (Expérience-type associée à une loi de Bernoulli)**.** *Soit une expérience à 2 issues (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 III.17 (15.3.17)**.** *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 $p=1$. De façon équivalente, si une variable aléatoire à valeurs dans $\mathbb{N}$ a une fonction génératrice affine, alors il s'agit d'une variable de Bernoulli (avec cette définition étendue).*

Définition III.18 (Loi binomiale)**.** *Soit X une v.a.r. On dit que X suit la loi binomiale de paramètres (n,p) (et on note $X \sim B(n,p)$) si :*

$$X(\Omega) = [0,n], \qquad \text{et} \qquad \forall k \in [0,n], \mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k}, \text{ avec } q = 1-p.$$

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

Un calcul élémentaire (fait en MPSI) amène :

Proposition III.19 (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 $X \sim B(n,p)$.*

Proposition III.20 (Espérance et variance d'une loi binomiale)**.** *Soit $X \sim B(n,p)$. Alors*

$$\forall t \in \mathbb{R}, G_X(t) = ((1-p)+tp)^n, \qquad \mathbb{E}(X) = np, \qquad \text{et} \qquad \mathbb{V}(X) = npq.$$

Éléments de preuve. La série génératrice se calcule par la formule du binôme. Dériver pour obtenir l'espérance et la variance. □

III.3 Loi géométrique

Nous ajoutons à ce catalogue deux nouvelles lois, discrètes, mais non finies.
La première correspond à la loi du temps d'attente dans une succession infinie de tirages à Pile ou Face. On sait, d'après le théorème 15.1.36, définir un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ modélisant cette expérience. On se place dans cet espace probabilisé.

Proposition III.21 (Temps d'attente du premier succès)**.** *Soit T la variable égale au rang du premier succès lors d'une succession infinie d'expériences de Bernoulli indépendantes de paramètre p, éventuellement égale à $+\infty$ si le succès n'est jamais obtenu. Alors $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$, et*

$$\forall n \in \mathbb{N}^*, \mathbb{P}(X=n) = pq^{n-1}, \qquad \text{et} \qquad P(X=+\infty) = 0.$$

Ainsi, la valeur $+\infty$ n'est pas pertinente dans la description de la loi, et on peut définir la loi géométrique de X sans la valeur infinie.

Définition III.22 (Loi géométrique)**.** *On dit que X suit une loi géométrique de paramètre $p \in]0,1[$, si $(X \in \mathbb{N}^*)$ est presque sûr, et*

$$\forall n \in \mathbb{N}^*, \mathbb{P}(X=n) = pq^{n-1}, \qquad \text{avec } q = 1-p.$$

On note $X \sim \mathcal{G}(p)$.

Et par notre motivation même :

Proposition III.23 (Expérience-type associée à une loi géométrique)**.** *Soit X le temps d'attente du premier succès lors d'une répétition infinie d'expériences de Bernoulli indépendantes de même paramètre p. Alors $X \sim \mathcal{G}(p)$.*

Proposition III.24 (Fonction génératrice, espérance et variance d'une loi géométrique)**.** *Soit $X \sim \mathcal{G}(p)$. Alors*

$$\forall t \in]-\frac{1}{q}, \frac{1}{q}[, G_X(t) = \frac{pt}{1-qt}, \qquad \mathbb{E}(X) = \frac{1}{p} \qquad \textit{et} \qquad \mathbb{V}(X) = \frac{q}{p^2}.$$

Éléments de preuve. On présente le calcul de l'espérance de X par fonction génératrice et par calcul direct. □

III.4 Loi de Poisson

Définition III.25 (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 λ, et on note $X \sim \mathcal{P}(\lambda)$, si :*

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

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

De plus :

Proposition III.26 (Espérance et variance d'une loi de Poisson)**.** *Soit $X \sim \mathcal{P}(\lambda)$. Alors*

$$\forall t \in \mathbb{R}, G_X(t) = \mathrm{e}^{\lambda(t-1)}, \qquad \mathbb{E}(X) = \lambda \qquad \text{et} \qquad \mathbb{V}(X) = \lambda.$$

Proposition III.27 (Limite en loi de variables binomiales)**.** *Soit (X_n) une suite de variables binomiales de paramètres (n, p_n), telles que $p_n \sim \frac{\lambda}{n}$ lorsque $n \to +\infty$. Alors*

$$\forall k \in \mathbb{N}, \mathbb{P}(X_n = k) \longrightarrow \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}.$$

On dit que la suite (X_n) converge en loi vers une loi de Poisson de paramètre λ.

Remarque III.28 (15.3.28)**.** *Pour cette raison, la loi de Poisson modélise le nombre de succès lorsque le succès est un événement rare (i.e. de petite probabilité) et que la population est très grande.*
Une situation typique est l'affluence (à un magasin, à un parc d'attraction etc.), lors d'un laps de temps donné (par exemple une journée), dans une population très grande. Chaque individu de la population a une probabilité p très faible d'aller à ce lieu lors d'une journée donnée, et la population N est très grande. Ainsi, la loi $B(N, p)$ représentant l'affluence au lieu donné pendant la période donnée est assez voisine de la loi $\mathcal{P}(Np)$. Ainsi, en posant $\lambda = Np$, la loi de Poisson $\mathcal{P}(\lambda)$ est une bonne modélisation de l'affluence.

III.5 Stabilités (HP classique)

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

Proposition III.29 (Stabilité des lois binomiales)**.** *Soit $(m, n) \in (\mathbb{N}^*)^2$ et $p \in]0, 1[$. Si $X \sim B(n, p)$ et $Y \sim B(m, p)$ sont indépendantes, alors $X + Y \sim B(n + m, p)$.*

Éléments de preuve. Se comprend bien par l'expérience. Assez immédiat par fonctions génératrices. □

Corollaire III.30 (Somme de variables de Bernoulli indépendantes). *Soit $X_1, \dots, X_n$ des variables aléatoires mutuellement indépendantes suivant la loi $\mathcal{B}(p)$. Alors $X_1 + \cdots + X_n \sim B(n,p)$.*

Proposition III.31 (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. □

IV. Inégalités probabilistes

Pour terminer, nous généralisons au cas de variables discrètes les inégalités de Markov et de Bienaymé-Tchebychev vues en MPSI pour les variables finies.

IV.1 Inégalités de Markov et de Bienaymé-Tchebychev

Théorème IV.1 (Inégalité de Markov). *Soit X une variable aléatoire réelle discrète dans L^1. Alors, pour tout $a > 0$,*

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}(X)}{a}.$$

Éléments de preuve. Par croissance de $\mathbb{E}$, en remarquant que $X \geqslant a\mathbb{1}_{\{X \geqslant a\}}$. □

Corollaire IV.2 (Réexpression quadratique de l'inégalité de Markov dans le cas L^2). *Soit X une variable aléatoire réelle discrète dans L^2, et $a \in \mathbb{R}_+^*$. Alors*

$$\mathbb{P}(|X| \geqslant a) \leqslant \frac{\mathbb{E}(X^2)}{a^2}.$$

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

Théorème IV.3 (Inégalité de Bienaymé-Tchebychev). *Soit Y une variable aléatoire réelle discrète dans L^2, d'espérance m et de variance σ^2. Alors :*

$$\forall \epsilon > 0, \mathbb{P}\big(|Y - m| \geqslant \epsilon\big) \leqslant \frac{\sigma^2}{\epsilon^2}.$$

Éléments de preuve. Inégalité de Markov quadratique appliquée à $Y - m$. □

IV.2 Loi faible des grands nombres

On a déjà vu la notion de convergence en loi de X_n vers X. Cette notion de convergence ne dit rien en terme de réalisation des variables, mais juste sur les probabilités.
Une notion plus probabiliste de convergence mesure à quel point deux variables ont tendance à prendre des valeurs proches l'une de l'autre.

Définition IV.4 (Convergence en probabilités). *Soit (X_n) une suite de variables aléatoires L^1 et X une variable L^1. On dit que (X_n) converge en probabilités vers X si*

$$\forall \epsilon > 0, \lim_{n \to +\infty} \mathbb{P}(|X_n - X| \geqslant \epsilon) = 0.$$

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 IV.5 (Loi faible des grands nombres, HP)**.** *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles discrètes L^2, i.i.d, d'espérance m et de variance σ^2. Soit $(Z_n)_{n\in\mathbb{N}^*}$ définie par :*

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

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

$$\forall \epsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}\big(|Z_n - m| \geqslant \epsilon\big) \leqslant \frac{\sigma^2}{n\epsilon^2}.$$

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

Exemple IV.6 (Théorème d'or de Bernoulli)**.** *1. La loi faible des grands nombres utilisée avec une suite de variables de Bernoulli i.i.d. de loi $\mathcal{B}(p)$, avec p non connu, on obtient*

$$\forall \epsilon > 0, \forall n \in \mathbb{N}^*, \mathbb{P}\big(|Z_n - p| \geqslant \epsilon\big) \leqslant \frac{1}{4n\epsilon^2}.$$

Le fait qu'on puisse faire cette majoration indépendamment de p permet d'estimer la valeur de p.
2. 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 $0,9$.