Chapitre 3: Variables aléatoires discrètes, espérance, variance et loi des grands nombres.

*

1 Introduction

Le nombre de piles obtenus au cours d'une série de n lancers de pile ou face ou plus généralement dans un jeu de hasard (roulette,dés,...), le gain d'un parieur est une grandeur variable qui dépend du déroulement aléatoire du jeu. Une telle grandeur numérique qui est fonction des éventualités d'une expérience aléatoire s'appelle une *variable aléatoire.*

Pour rendre compte mathématiquement d'une variable aléatoire, on doit la considérer comme une application $X : \omega \mapsto X(\omega)$ définie sur un univers des possibles Ω est à valeurs réelles. Dans ce chapitre, et pour des raisons de simplicité, toutes les variables aléatoires considérées seront **discrètes** (i.e. l'ensemble $X(\Omega)$ des valeurs prises par X est fini ou dénombrable).

2 Généralités

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé modélisant une certaine expérience aléatoire. Dans ce chapitre l'espace Ω sera le plus souvent discret et dans ce cas on supposera que la tribu $\mathcal{F}$ des événements est égale à l'ensemble $\mathcal{P}(\Omega)$ de tous les sous-ensembles de Ω.

Définition 2.1 *: 1) On appelle* ***variable aléatoire (v.a.*** *en abrégé)* ***discrète*** *définie sur Ω toute application $X : \Omega \to \mathbb{R}$ telle que*
a) L'ensemble $X(\Omega) = \{x_i;\ i \in D\}$ des valeurs prises[1] *par X est fini ou dénombrable.*
b) (condition de mesurabilité) Pour tout $x_i \in X(\Omega)$, on a

$$[X = x_i] := \{\omega \in \Omega;\ X(\omega) = x_i\} \in \mathcal{F} \tag{1}$$

(i.e. l'ensemble $[X = x_i]$ est un événement)[2]*. On dit que c'est l'événement «X* ***prend la valeur*** *x_i».*

*Notes du cours de Probabilités de M1 de M. L. Gallardo, Université de Tours, année 2009-2010. Les démonstrations sont détaillées dans le cours oral.

[1] On suppose que $D = \{1, 2, \dots N\}$ (resp.$\mathbb{N}^*$) si l'ensemble $X(\Omega)$ est fini (resp. dénombrable)

[2] cette condition est toujours satisfaite si Ω est discret et $\mathcal{F} = \mathcal{P}(\Omega)$.

4) La famille $(p_i)_{i\in D}$ des nombres $p_i = \mathbb{P}(X = x_i)$ $(i \in D)$ s'appelle la **distribution de probabilité** *(ou loi de probabilité) de la v.a. X.*

Remarque 2.2 *:* Il est important de noter que les événements $[X = x_i]$ $(i \in D)$ forment un système complet d'événements et par conséquent $\sum_{i\in D} p_i = 1$.

Exemple 2.3 *:* On lance trois fois une pièce régulière. Lorsqu'il sort pile, on gagne 1E, s'il sort face 0E. Soit X le gain obtenu par le joueur. L'expérience aléatoire considérée a 8 éventualités équi-probables :

$$\Omega = \{000, 001, 010, 100, 011, 101, 110, 111\}$$

et à chaque éventualité $\omega \in \Omega$, correspond une valeur $X(\omega)$ du gain X. On peut résumer ceci par un tableau *:*

ω	*000*	*001*	*010*	*100*	*011*	*101*	*110*	*111*
$X(\omega)$	*0*	*1*	*1*	*1*	*2*	*2*	*2*	*3*
$\mathbb{P}(\omega)$	*1/8*	*1/8*	*1/8*	*1/8*	*1/8*	*1/8*	*1/8*	*1/8*

On voit que $X(\Omega) = \{0, 1, 2, 3\}$ et on détermine aussitôt les événements $[X = x_i]$ *:*

$[X = 0] = \{000\}$
$[X = 1] = \{001,\ 010,\ 100\}$
$[X = 2] = \{011,\ 101,\ 110\}$
$[X = 3] = \{111\}$

On en déduit la loi de probabilité de X :

x_i	*0*	*1*	*2*	*3*
$p_i = \mathbb{P}(X = x_i)$	*1/8*	*3/8*	*3/8*	*1/8*

Définition 2.4 *: Supposons que les valeurs prises par la v.a. X puissent s'écrire dans l'ordre croissant $x_1 < x_2 < x_3 < \cdots < x_n < \cdots$ (ce qui est le cas par exemple si $X(\Omega) = \mathbb{N}$). On appelle alors **probabilités cumulées** la suite (finie ou infinie) (a_k) des nombres*

$$a_k = \mathbb{P}(X \leq x_k) = \sum_{i=1}^{k} p_i \ .$$

Remarque 2.5 *:* Pour les v.a. dont les lois sont tabulées, les probabilités cumulées sont intéressantes du point de vue pratique car pour des valeurs $x_l < x_m$, on a aussitôt

$$\mathbb{P}(x_l \leq X \leq x_m) = a_m - a_{l-1}.$$

2.1 Opérations sur les variables aléatoires

La notion de variable aléatoire discrète se prête bien aux tranformations déterministes ainsi qu'aux opérations algébriques usuelles qu'on fait sur les applications à valeurs réelles. Ainsi par exemple si les gains de plusieurs joueurs sont des variables aléatoires relatives à un certain jeu, toute fonction déterministe de leur gain (comme le carré ou la puissance troisième) mais aussi la somme totale (ou le produit) de leurs gains sont aussi des variables aléatoires relatives au jeu considéré.

Toutes les variables aléatoires qu'on considère dans ce paragraphe sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Soient X et Y des variables aléatoires discrètes, $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R}^2 \to \mathbb{R}$ des applications.

Théorème 2.6 *: Les applications*

$$f(X) : \omega \mapsto f(X)(\omega) := f(X(\omega)) \tag{2}$$

et

$$g(X,Y) : \omega \mapsto g(X,Y)(\omega) := g(X(\omega)), Y(\omega)) \tag{3}$$

définies sur Ω sont des variables aléatoires discrètes.

Par exemple X^n (n entier ≥ 0), λX ($\lambda \in R$), $X+Y$, $X-Y$, XY, etc...sont aussi des variables aléatoires discrètes. Mais plus généralement si $X_1, \ldots, X_n$ sont desvariables aléatoires discrètes et si $g : \mathbb{R}^n \to \mathbb{R}$ est une application, $g(X_1, \ldots, X_n)$ est aussi une variable aléatoire discrète.

démonstration du théorème : Il est clair que les ensembles $f(X)(\Omega)$ et $g(X,Y)(\Omega)$ sont discrets. Il reste à voir que la condition de mesurabilité (1) est satisfaite. Montrons le pour $f(X)$. Soit $y \in f(X)(\Omega)$ et $A_y = \{x \in X(\Omega);\ f(x) = y\}$. L'ensemble A_y est fini ou dénombrable et on voit facilement que

$$[f(X) = y] = \cup_{x \in A_y}[X = x] \in \mathcal{F},$$

car $\mathcal{F}$ est stable par réunions finies ou dénombrables. D'où le résultat. De la même façon on montre aussi que $g(X,Y)$ est une v.a. □

3 Les lois classiques

3.1 Loi de Bernoulli

Soit A un événement relatif à une expérience aléatoire modélisée par un espace $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 3.1 *: On appelle* ***v.a. indicatrice*** *de l'événement A la v.a. définie sur Ω par*

$$\mathbf{1}_A(\omega) = 1 \ \text{si}\ \omega \in A$$
$$\mathbf{1}_A(\omega) = 0 \ \text{si}\ \omega \in \bar{A}.$$

Si on note $p = \mathbb{P}(A)$ la probabilité de l'événement A. La loi de probabilité de la v.a. $\mathbf{1}_A$ (appellée loi de Bernoulli $\mathcal{B}(1,p)$) est donnée par

$\mathbf{1}_A$	0	1
p_i	1-p	p

3.2 Loi binomiale

On considère une expérience aléatoire à deux issues S(succès) et E(échec) avec $\mathbb{P}(S) = p$ et $\mathbb{P}(E) = 1 - p$ $(0 \leq p \leq 1)$. On fait n répétitions indépendantes de cette expérience qu'on modélise par l'espace produit $\Omega = \{S, E\}^n$ muni de la probabilité produit comme expliqué au chapitre 2.

Définition 3.2 *: La variable aléatoire X=«nombre total de succès» (au cours des n répétitions) est appelée **v.a. binomiale de paramètres** (n, p). Pour abréger la loi d'une telle v.a. sera désignée par $\mathcal{B}(n, p)$* [3]*.*

Proposition 3.3 *: L'expression de la loi $\mathcal{B}(n, p)$ est*

$$p_k = \mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k} \quad (k = 0, 1, \ldots, n).$$

démonstration : L'événement $[X = k]$ est l'ensemble des n-uplets composés de k lettres S et $n - k$ lettres E qui sont au nombre de C_n^k. Tous ces n uplets ont la même probabilité $p^k(1 - p)^{n-k}$ (par définition de la probabilité produit). D'où le résultat. □

3.3 Loi géomètrique (ou loi de Pascal)

On considère une infinité de répétitions indépendantes d'une expérience à deux issues S et E modélisée par l'espace produit infini $\Omega = \{S, E\}^{\mathbb{N}^*}$ des suites infinies $\omega = (x_k)_{k \geq 1}$ $(x_k \in \{S, E\})$, muni de la probabilité produit comme expliqué au chapitre 1. On considère la v.a. X=«instant du premier succès»[4] qui est telle que

$$(4) \qquad X(\omega) = n \quad \text{si} \quad \forall k \in \{1, \ldots n-1\}\,,\ x_k = E \quad \text{et} \quad x_n = S.$$

Si on considère les événements E_i=«échec au i-ème coup» et S_i=«succès au i-ème coup», on peut écrire l'événement $[X = n]$ sous la forme

$$(5) \qquad [X = n] = E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap S_n$$

(intersection d'événements indépendants) d'où il résulte aussitôt que

$$(6) \qquad \mathbb{P}(X = n) = (1 - p)^{n-1} p \quad (n \in \mathbb{N}^*).$$

Définition 3.4 *: La loi de probabilité (6) de la v.a. «instant du premier succès» considérée en (4) s'appelle **loi géométrique** (ou loi de Pascal).*

---

[3]Si $n = 1$, c'est la loi de Bernoulli.

[4]attention X est bien discrète mais ici l'espace Ω où X est définie n'est pas discret ! il convient de vérifier la condition de mesurabilité de la définition 2.1 b) . Mais c'est évident car pour tout $n \in \mathbb{N}^*$, l'ensemble $[X = n]$ est cylindrique d'après la formule (5) donc il appartient à la tribu produit.

3.4 Loi de Poisson

Définition 3.5 *: Soit $\lambda > 0$ un paramètre fixé. On dit qu'une v.a. X à valeurs dans l'ensemble des entiers $\mathbb{N}$ suit la* ***loi de Poisson de paramètre*** *λ si*

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

Remarque 3.6 : On notera que les nombres $p_n = e^{-\lambda} \frac{\lambda^n}{n!}$ $(n \in \mathbb{N})$ constituent bien une loi de probabilité sur $\mathbb{N}$ puisqu'ils sont positifs et de somme $\sum_{n=0}^{+\infty} p_n = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1$.

Proposition 3.7 *(les succès rares et la loi de Poisson) : Soit $n \in \mathbb{N}$ et $0 < p < 1$. Supposons que n tende vers $+\infty$ et que p tende vers 0 de telle sorte que np tende vers une constante $\lambda > 0$. Alors pour tout $k \in \mathbb{N}$ fixé,*

$$C_n^k p^k (1-p)^{n-k} \to e^{-\lambda} \frac{\lambda^k}{k!} \cdot$$

démonstration :

$$\begin{aligned} C_n^k p^k (1-p)^{n-k} &= \frac{1}{k!} \left[(n-k+1)(n-k+2) \cdots n \right] p^k (1-p)^{n-k} \\ &= \frac{1}{k!} (np)^k (1-p)^n \left(1 - \frac{k-1}{n}\right) \cdots \left(1 - \frac{2}{n}\right) (1-p)^{-k}. \end{aligned} \tag{7}$$

Si on pose $np = \lambda + \epsilon$ où $\lim \epsilon = 0$, alors

$$(1-p)^n = \left(1 - \frac{\lambda}{n} + \frac{\epsilon}{n}\right)^n \to e^{-\lambda} \quad (\epsilon \to 0)$$

et comme les facteurs à la droite de $(1-p)^n$ dans la formule (7) tendent tous vers 1, le résultat en découle. □

Remarque 3.8 : La loi de Poisson apparaît donc comme une approximation de la loi binomiale quand n est "grand" et p est "petit" (succès rare). Par exemple si $n = 100$ et $p = 0,1$ la loi binomiale $\mathcal{B}(100;\ 0,1)$ est très bien approchée par la loi de Poisson de paramètre $\lambda = 10$. Ce fait est exploité dans la construction des tables de la loi binomiale.

4 Espérance, variance et moments d'une v.a.

4.1 Introduction

Soit X une v.a. prenant un nombre fini de valeurs $x_1, x_2, \ldots, x_n$ avec les probabilités respectives $p_1, p_2, \ldots, p_n$. La somme

$$p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$$

peut être interprétée de deux manières :

a) **mathématiquement** c'est une moyenne pondérée. Plus précisément, c'est le barycentre du système (x_i, p_i) $(i = 1, \ldots, n)$, le «point» x_i étant affecté de la «masse» p_i.

b) **heuristiquement** supposons qu'on répète N fois l'expérience aléatoire attachée à X et soit N_i le nombre de fois où X prend la valeur x_i $(i = 1, \ldots, n)$. La moyenne arithmétique des valeurs de X observées au cours des N essais est :

$$\frac{N_1 x_1 + N_2 x_2 + \cdots + N_n x_n}{N} = \frac{N_1}{N} x_1 + \frac{N_2}{N} x_2 + \cdots + \frac{N_n}{N} x_n,$$

qui puisque $\frac{N_i}{N}$ est proche de p_i si N est grand (loi empirique des grands nombres[5]), est pratiquement égale à $p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$ si N est grand.

4.2 Généralités

4.2.1 Définitions de l'espérance et des moments

Considérons une v.a. X discrète prenant les valeurs x_i avec la probabilité p_i $(i \in D$, où D est fini ou infini dénombrable$)$[6].

Définition 4.1 *: 1) Si $\sum_{i \in D} p_i |x_i| < +\infty$ [7], on dit que X a un* ***moment d'ordre un****. Dans ce cas, le nombre réel*

$$\mathbb{E}(X) = \sum_{i \in D} p_i x_i$$

(somme finie si D est fini et somme d'une série convergente si D est dénombrable) est appelé ***espérance*** *(ou moyenne) de la v.a. X.*

2) Si $\mathbb{E}(X) = 0$ on dit que la v.a. est ***centrée****.*

3) Pour $r > 0$, le nombre $\mathbb{E}(X^r)$, lorsqu'il existe, est appelé ***moment d'ordre*** *r de la v.a. X.*

4) Le nombre $Var(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$, lorsqu'il existe, est appelé ***variance*** *de X et le nombre $\sigma_X = \sqrt{Var(X)}$ est* ***l'écart type*** *de X.*

5) Une v.a. X telle que $\mathbb{E}(X) = 0$ et $\sigma_X = 1$ est dite ***centrée et réduite****.*

Remarque 4.2 *:* Si D est dénombrable, la condition $\sum_{i \in D} p_i |x_i| < +\infty$ (i.e. la convergence absolue de la série $\sum_{i \in D} p_i x_i$) est fondamentale pour la définition de l'espérance. En effet on sait (cours d'analyse du L2) que lorsqu'une série est absolument convergente, elle est évidemment convergente mais elle est surtout commutativement convergente, c'est à dire que la somme $\sum_{i \in D} p_i x_i$ est la même quel que soit l'ordre dans lequel on écrit les termes. Cette propriété est indispensable, comme on va le voir ci-dessous.

Remarque 4.3 *:* Comme on l'a vu dans l'introduction, l'espérance d'une v.a. X peut-être vue comme une position moyenne des valeurs de X, c'est donc un indicateur de position. La variance (resp. l'écart type) est un indicateur de dispersion qui mesure l'espérance des écarts

[5] voir le chapitre 1
[6] on peut supposer $D = \{1, \ldots, N\}$ si D est fini et $D = \mathbb{N}$ si D est dénombrable.
[7] ce qui est toujours le cas si D est fini.

quadratiques entre X et sa valeur moyenne $\mathbb{E}(X)$, le nombre $Var(X)$ (resp. σ_X) étant d'autant plus petit que les valeurs de X sont plus concentrées autour de $\mathbb{E}(X)$. On pourrait aussi utiliser l'espérance des écarts absolus, c'est à dire le nombre $\mathbb{E}(|X - \mathbb{E}(X)|)$, pour mesurer la dispersion mais ce nombre ne se prête pas bien au calcul contrairement à la variance qui a de très bonnes propriétés mathématiques comme nous aurons l'occasion de le voir.

Exercice 4.4 *: Supposons que l'univers Ω où la v.a. X est définie, est discret. L'espérance de X existe si et seulement si la série $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$ est absolument convergente*[8] *et dans ce cas on peut la calculer par la formule*[9]

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega). \tag{8}$$

solution : 1)($\Rightarrow$) Supposons que l'espérance de X existe. Comme les événements $[X = x_i]$ ($x_i \in X(\Omega)$) forment une partition de Ω, la somme S de la série à termes positifs $\sum_{\omega\in\Omega} |X(\omega)|\mathbb{P}(\omega)$ peut se décomposer en blocs[10] :

$$\begin{aligned} S &= \sum_{\omega\in[X=x_1]} |X(\omega)|\mathbb{P}(\omega) + \cdots + \sum_{\omega\in[X=x_i]} |X(\omega)|\mathbb{P}(\omega) + \cdots \\ &= |x_1| \sum_{\omega\in[X=x_1]} \mathbb{P}(\omega) + \cdots + |x_i| \sum_{\omega\in[X=x_i]} \mathbb{P}(\omega) + \cdots \\ &= |x_1|\mathbb{P}(X = x_1) + \cdots + |x_i|\mathbb{P}(X = x_i) + \cdots \qquad (9) \end{aligned}$$

Mais comme $\sum_{i\in D} p_i|x_i| < +\infty$ par hypothèse, on a $S < +\infty$ et donc la série $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$ est absolument convergente. Mais la série $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$ est aussi commutativement convergente[11] et on peut reprendre mot à mot le calcul précédent en remplaçant $|X(\omega)|$ par $X(\omega)$ ce qui montre que $\mathbb{E}(X)$ s'exprime par la formule (8).
2) ($\Leftarrow$) Réciproquement, supposons que la série $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$ soit absolument convergente. En reprenant le calcul ci-dessus (formule (9)) on montre que $\sum_{i\in D} p_i|x_i| < +\infty$ donc l'espérance existe. De plus la série $\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega)$ étant commutativement convergente, le calcul (9) avec $|X(\omega)|$ remplacé par $X(\omega)$ montre que $\mathbb{E}(X)$ est donné par la formule (8). D'où le résultat. $\square$

4.2.2 Espérance d'une somme de variables aléatoires

Dans ce paragraphe toutes les v.a. considérées sont discrètes et sont supposées définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

---

[8] i.e. si $\sum_{\omega\in\Omega} |X(\omega)|\mathbb{P}(\omega) < +\infty$, ce qui, par exemple est toujours vérifié si Ω est fini.

[9] attention il faut que l'univers Ω soit discret sinon cette formule n'a pas de sens ! Dans le cas général d'un univers quelconque nous verrons dans un prochain chapitre que la sommation doit être remplacée par une intégrale : $\int_\Omega X(\omega)d\mathbb{P}(\omega)$.

[10] ceci est valable que S soit finie ou égale à $+\infty$

[11] car elle est absolument convergente

Théorème 4.5 : *Si X et Y sont deux v.a. ayant une espérance, la v.a. $X+Y$ a aussi une espérance et*

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) \tag{10}$$

démonstration : Pour $x \in X(\Omega)$ et $y \in Y(\Omega)$, posons

$$A_{xy} = [X = x] \cap [Y = y].$$

Les événements A_{xy} ($(x,y) \in X(\Omega) \times Y(\Omega)$) forment un système complet et satisfont les deux relations suivantes

$$\left\{ \begin{array}{lcl} \sum_{x \in X(\Omega)} \mathbb{P}(A_{xy}) & = & \mathbb{P}(Y = y) \\ \sum_{y \in Y(\Omega)} \mathbb{P}(A_{xy}) & = & \mathbb{P}(X = x) \end{array} \right. \tag{11}$$

Posons $Z = X + Y$, admettons pour l'instant qu'on ait démontré la condition

$$\sum_{z \in Z(\Omega)} |z| \mathbb{P}(Z = z) < +\infty \tag{12}$$

et calculons $\mathbb{E}(Z) = \sum_{z \in Z(\Omega)} z \mathbb{P}(Z = z)$. Pour tout $z \in Z(\Omega)$, considérons l'ensemble

$$B_z = \{(x,y) \in X(\Omega) \times Y(\Omega);\ x + y = z\}.$$

Il est clair que les ensembles B_z ($z \in Z(\Omega)$) forment une partition de $X(\Omega) \times Y(\Omega)$ et que $\mathbb{P}(Z = z) = \sum_{(x,y) \in B_z} \mathbb{P}(A_{xy})$. On a donc

$$\begin{aligned}
\mathbb{E}(Z) &= \sum_{z \in Z(\Omega)} z \sum_{(x,y) \in B_z} \mathbb{P}(A_{xy}) \\
&= \sum_{z \in Z(\Omega)} \sum_{(x,y) \in B_z} (x+y) \mathbb{P}(A_{xy}) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (x+y) \mathbb{P}(A_{xy}) \\
&= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} x \mathbb{P}(A_{xy}) + \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} y \mathbb{P}(A_{xy}) \\
&= \sum_{x \in X(\Omega)} x \sum_{y \in Y(\Omega)} \mathbb{P}(A_{xy}) + \sum_{y \in Y(\Omega)} y \sum_{x \in X(\Omega)} \mathbb{P}(A_{xy}) \qquad (13) \\
&= \mathbb{E}(X) + \mathbb{E}(Y).
\end{aligned}$$

(où pour (13) on a utilisé la propriété de Fubini et les relations (11) pour obtenir la dernière égalité. Enfin, pour justifier la condition (12), on fait un calcul analogue en remplaçant z par $|z|$ et (avec les mêmes notations), on obtient une majoration :

$$\sum_{z \in Z(\Omega)} |z| \mathbb{P}(Z = z) \leq \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} (|x| + |y|) \mathbb{P}(A_{xy})$$

et le membre de droite est fini puisqu'il est égal à $\mathbb{E}(|X|) + \mathbb{E}(|Y|)$. $\square$

Remarque 4.6 *:* Dans le calcul précédent, on notera que la légitimité de la sommation par paquets : $\sum_{z\in Z(\Omega)}\sum_{(x,y)\in B_z} = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)}$ et l'utilisation de la propriété de Fubini, est assurée par la condition de convergence absolue (12).

Corollaire 4.7 *(linéarité de l'espérance) : Soient $X_1,\dots,X_n$ des v.a. ayant une espérance et $a_1,\dots,a_n$ des réels. Alors la v.a. $a_1X_1+\cdots+a_nX_n$ a une espérance donnée par*

$$\mathbb{E}(a_1X_1+\cdots+a_nX_n) = a_1\mathbb{E}(X_1)+\cdots+a_n\mathbb{E}(X_n).$$

Exercice 4.8 *:* On suppose que l'espace Ω est discret. Démontrer le résultat précédent en utilisant la formule de l'espérance donnée à l'exercice 4.4.

solution : Il suffit de prouver le résultat pour $n=2$. Il est facile de voir que $\mathbb{E}(|a_1X_1+a_2X_2|)\le |a_1|\mathbb{E}(|X_1|)+|a_2|\mathbb{E}(|X_2|) < +\infty$, donc $a_1X_1+a_2X_2$ a une espérance et d'après 4.4 on a

$$\begin{aligned}\mathbb{E}(a_1X_1+a_2X_2) &= \sum_{\omega\in\Omega}(a_1X_1(\omega)+a_2X_2(\omega))\,\mathbb{P}(\omega)\\ &= a_1\sum_{\omega\in\Omega}X_1(\omega)\mathbb{P}(\omega)+a_2\sum_{\omega\in\Omega}X_2(\omega)\mathbb{P}(\omega)\\ &= a_1\mathbb{E}(X_1)+a_2\mathbb{E}(X_2).\end{aligned}$$

4.2.3 Fonction d'une variable aléatoire

Étant donné une v.a. X on considère souvent des v.a. de la forme $Y=g(X)$ où $g:\mathbb{R}\to\mathbb{R}$ est une fonction usuelle. Pour trouver si Y a une espérance, il faudrait en principe déterminer la loi de probabilité de Y ce qui peut s'avérer difficile. Voici un résultat qui permet le calcul de $\mathbb{E}(Y)$ à partir de la loi de X uniquement.

Théorème 4.9 *: Soit X une v.a. discrète , g une application de $\mathbb{R}$ dans $\mathbb{R}$ et $Y=g(X)$. Alors l'espérance de la v.a. Y existe si et seulement si $\sum_{x\in X(\Omega)}|g(x)|\mathbb{P}(X=x)<+\infty$ et on peut dans ce cas la calculer par la formule*

$$\mathbb{E}(Y)=\sum_{x\in X(\Omega)} g(x)\mathbb{P}(X=x) \tag{14}$$

démonstration : Pour tout $y\in Y(\Omega)$, soit $B_y=\{x\in X(\Omega); g(x)=y\}$. Il est clair que les ensembles B_y $(y\in Y(\Omega))$ forment une partition de l'ensemble $X(\Omega)$ des valeurs prises par X et que $\sum_{x\in B_y}\mathbb{P}(X=x)=\mathbb{P}(Y=y)$

a) Supposons la condition $\sum_{x\in X(\Omega)}|g(x)|\mathbb{P}(X=x)<+\infty$ vérifiée. Alors l'espérance de Y existe puisque

$$\begin{aligned}\sum_{x\in X(\Omega)}|g(x)|\mathbb{P}(X=x) &= \sum_{y\in Y(\Omega)}\sum_{x\in B_y}|g(x)|\mathbb{P}(X=x)\\ &\sum_{y\in Y(\Omega)}|y|\sum_{x\in B_y}\mathbb{P}(X=x)\\ &= \sum_{y\in Y(\Omega)}|y|\mathbb{P}(Y=y).\end{aligned}$$

Dans ces conditions en reprenant le calcul précédant avec y au lieu de $|y|$ et $g(x)$ au lieu de $|g(x)|$, on obtient $\mathbb{E}(Y) = \sum_{x\in X(\Omega)} g(x)\mathbb{P}(X = x)$.
b) Réciproquement, si $\mathbb{E}(Y)$ existe, on montre immédiatement en «remontant» le calcul précédent que $\sum_{x\in X(\Omega)} |g(x)|\mathbb{P}(X = x) < +\infty$ et que $\mathbb{E}(Y)$ est donnée par la formule (14). $\square$

Exemple 4.10 *:* Pour toute v.a. discrète X prenant les valeurs x_i avec les probabilités p_i $(i \in D)$, X admet un moment d'ordre r(entier > 0)[12] si et seulement si $\sum_{i\in D} p_i|x_i|^r < +\infty$ et dans ce cas

$$\mathbb{E}(X^r) = \sum_{i\in D} p_i x_i^r.$$

Il suffit en effet d'appliquer le théorème 4.9 avec la fonction $g(x) = x^r$.

Corollaire 4.11 *(Formule pour la variance) : Soit X une v.a. discrète prenant les valeurs x_i avec les probabilités p_i $(i \in D)$ et ayant un moment d'ordre 2. Alors elle a aussi un moment d'ordre 1 et*

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \tag{15}$$

$$= \left(\sum_{i\in D} p_i(x_i)^2\right) - \left(\sum_{i\in D} p_i x_i\right)^2 \tag{16}$$

démonstration : X a un moment d'ordre 1 car d'après l'inégalité de Cauchy-Schwarz $\sum_{i\in D} p_i|x_i| \leq \left(\sum_{i\in D} p_i\right)^{1/2} \left(\sum_{i\in D} p_i(x_i)^2\right)^{1/2} < +\infty$. Par définition $Var(X) = \mathbb{E}\left((X - \mathbb{E}(X)^2\right) = \mathbb{E}\left(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2\right)$ et la formule (15) découle alors de la linéarité de l'espérance (corollaire 4.7). La deuxième formule (16) résulte alors du théorème 4.9 (appliqué avec $g(x) = x^2$). $\square$

Remarque 4.12 *:* Pour une v.a. discrète, il est évident que $Var(X) = 0$ si et seulement si X est constante. Avec le résultat précédent, on montre alors immédiatement que si X a un moment d'ordre 2, il en est de même pour $aX + b$ (a et b réels fixés) et on a

$$Var(aX + b) = a^2 Var(X). \tag{17}$$

4.3 Exemples

4.3.1 Loi binomiale et loi de Poisson

Proposition 4.13 *: 1) Si X suit la loi binomiale $\mathcal{B}(n, p)$ on a $\mathbb{E}(X) = np$ et $Var(X) = np(1 - p)$.*
2) Si X suit la loi de Poisson de paramètre $\lambda(> 0)$, X a des moments de tous les ordres et $\mathbb{E}(X) = Var(X) = \lambda$.

[12] on notera que si X est à valeurs positives, on peut aussi considérer la notion de moment d'ordre $r > 0$ (pas forcément entier).

démonstration : 1) X a des moments de tous ordres[13]. Calculons $\mathbb{E}(X)$:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} kC_n^k p^k(1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k(1-p)^{n-k}\\ &= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1}(1-p)^{n-1-(k-1)}\\ &= np\sum_{k=0}^{n-1} C_{n-1}^k p^k(1-p)^{n-1-k} = np\end{aligned}$$

Le calcul de la variance de la loi $\mathcal{B}(n,p)$ est laissé en exercice.

2) La v.a. de Poisson X prend toutes les valeurs entières. Soit $r>0$ fixé. D'après le théorème 4.9 [14], la v.a. X^r a une espérance si et seulement si la série $\sum_{k=0}^{+\infty} k^r e^{-\lambda}\frac{\lambda^k}{k!}$ converge ce qui est vrai (exercice de niveau L2). On a alors

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Le calcul de la variance de X est laissé en exercice (indication : commencer par calculer $\mathbb{E}(X(X-1))$ et en déduire $\mathbb{E}(X^2)$ puis $Var(X)$).

4.3.2 Jeux équitables

Deux joueurs A et B jouent à un jeu d'argent où la probabilité de gagner est égale à p pour A et à $1-p$ pour B $(0<p<1)$. Les mises de A et B sont respectivement s et s' euros et le vainqueur empoche le total des enjeux. Soient X et X' les gains des joueurs A et B. Le jeu est dit **équitable** si $\mathbb{E}(X)=\mathbb{E}(Y)$: On a

$$\mathbb{E}(X) = s'p - s(1-p) \quad \text{et} \quad \mathbb{E}(Y) = (1-p)s - s'p.$$

Le jeu est donc équitable si $\frac{s}{p} = \frac{s'}{1-p}$, autrement dit si les enjeux des joueurs sont proportionnels à leur probabilité de succès.

4.3.3 Le jeu de St Petersbourg

Imaginons le jeu de casino suivant : On lance une pièce[15] jusqu'à l'apparition du premier pile. Si cela se produit au n-ième lancer, la banque verse au joueur la somme $X=2^n$ euros. Quel doit être l'enjeu que la banque devrait exiger du joueur pour ne pas être perdante ?

Pour que le jeu soit équitable, la mise doit être égale à l'espérance de gain du joueur. Mais l'espérance de X n'est pas finie car X prend les valeurs 2^n $(n\geq 1)$ et $\mathbb{P}(X=2^n)=\frac{1}{2^n}$ donc $\sum_{n=1}^{+\infty} 2^n\mathbb{P}(X=2^n)=+\infty$.

[13] car elle prend un nombre fini de valeurs.
[14] appliqué avec la fonction $g(x)=x^r$.
[15] non truquée

5 La loi des grands nombres

5.1 Variables aléatoires indépendantes

Définition 5.1 *: 1) On dit que les v.a. X et Y sont **indépendantes** si pour tout $x \in X(\Omega)$ et tout $y \in Y(\Omega)$, les événements $[X = x]$ et $[Y = y]$ sont indépendants i.e.*

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

2) La définition se généralise au cas de n v.a. $X_1, \ldots, X_n$ qui sont dites indépendantes (dans leur ensemble) si pour tout choix de $a^{(i)} \in X_i(\Omega)$ $(i = 1, \ldots, n)$, les événements $\left[X_i = a^{(i)}\right]$ $(i = 1, \ldots, n)$ sont indépendants.

3) Une suite $(X_n)_{n\in\mathbb{N}}$ de v.a. indépendantes est une suite telle que pour tout entier N, les v.a. $X_1, \ldots, X_N$ sont indépendantes

Exemple 5.2 *:* Des v.a. X et Y sont indépendantes si elles se rapportent à deux expériences aléatoires indépendantes. Ainsi lorsqu'on fait plusieurs répétitions indépendantes d'une même expérience : Soit X_1 le résultat du premier essai,..., X_n le résultat du n-ième essai. Les v.a. $X_1, \ldots, X_n$ sont indépendantes.

Théorème 5.3 *(espérance d'un produit de v.a. indépendantes) : Si les v.a. X et Y ont un moment d'ordre un et sont indépendantes, alors la v.a. XY a un moment d'ordre un et on a*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{18}$$

démonstration : 1) Supposons que X et Y soient à valeurs positives. Les valeurs de Z sont de la forme $z = xy$, pour $x \in X(\Omega)$ et $y \in Y(\Omega)$. Pour $z \in Z(\Omega)$, considérons l'ensemble

$$A_z = \{(x, y) \in X(\Omega) \times Y(\Omega);\ xy = z\}$$

Il est clair que les A_z forment une partition de $X(\Omega) \times Y(\Omega)$ et que

$$[Z = z] = \cup_{(x,y)\in A_z} [X = x] \cap [Y = y]$$

(réunion d'événements deux à deux incompatibles). Alors

$$\begin{aligned}
\sum_{z\in Z(\Omega)} z\mathbb{P}(Z = z) &= \sum_{z\in Z(\Omega)} \left(z \sum_{(x,y)\in A_z} \mathbb{P}\left([X = x] \cap [Y = y]\right) \right) \\
&= \sum_{z\in Z(\Omega)} \left(z \sum_{(x,y)\in A_z} \mathbb{P}(X = x)\mathbb{P}(Y = y) \right) \\
&= \sum_{z\in Z(\Omega)} \left(\sum_{(x,y)\in A_z} xy\mathbb{P}(X = x)\mathbb{P}(Y = y) \right) \\
&= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \sum_{x\in X(\Omega)} x\mathbb{P}(X = x) \sum_{y\in y(\Omega)} y\mathbb{P}(Y = y) < +\infty.
\end{aligned}$$

(la deuxième égalité d'après l'indépendance des v.a. X et Y et la quatrième par le fait qu'une série à termes positifs est sommable par paquets).

2) Si X et Y sont de signe quelconque, le 1) montre que Z a un moment d'ordre un. On peut alors reprendre le calcul précédent pour prouver la formule (18) car la sommation par paquets est valable pour des séries absolument convergentes.

Corollaire 5.4 *(Variance d'une somme de v.a. indépendantes) : Si X et Y ont un moment d'ordre deux et sont indépendantes, $X+Y$ a un moment d'ordre deux et*

$$Var(X+Y) = Var(X) + Var(Y). \tag{19}$$

démonstration : En développant $(X+Y)^2$ et en utilisant le corollaire 4.7, on voit tout de suite que $X+Y$ a un moment d'ordre deux. On a alors

$$\begin{aligned} Var(X+Y) &= \mathbb{E}\left((X-\mathbb{E}(X)+Y-\mathbb{E}(Y))^2\right) \\ &= Var(X) + Var(Y) + 2\left(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)\right) \end{aligned}$$

et le terme résiduel[16] est nul puisque $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. □

5.2 La loi des grands nombres

Théorème 5.5 *(inégalité de Bienaymé-Tchebychev) : Soit X une v.a. ayant un moment d'ordre deux. On note $m = \mathbb{E}(X)$ et $\sigma = \sqrt{Var(X)}$. Alors* **pour tout** *$a > 0$, on a l'inégalité*

$$\mathbb{P}\left(\left|\frac{X-m}{\sigma}\right| \geq a\right) \leq \frac{1}{a^2} \tag{20}$$

démonstration : La v.a. centrée réduite $Y = \frac{X-m}{\sigma}$ est telle que $\mathbb{E}(Y^2) = 1$. Or si on désigne par y_i les valeurs prises par Y,

$$\mathbb{E}(Y^2) = \sum_i y_i^2 \mathbb{P}(Y = y_i) \geq \sum_{|y_i| \geq a} y_i^2 \mathbb{P}(Y = y_i)$$

(on somme seulement sur les indices i tels que $|y_i| \geq a$). Ainsi on a

$$1 = \mathbb{E}(Y^2) \geq a^2 \sum_{|y_i| \geq a} \mathbb{P}(Y = y_i) = a^2 \mathbb{P}(|Y| \geq a)$$

et l'inégalité (20) en découle. □

Remarque 5.6 : L'inégalité de Bienaymé-Tchebychev est valable pour tout $a > 0$. Pour a petit, $\frac{1}{a^2}$ est grand et l'inégalité ne nous apprend rien. Par contre si a est grand, l'inégalité nous dit qu'il est très improbable que X s'écarte de m de plus de $a\sigma$.

[16] La quantité $\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ est appelée la covariance des v.a. X et Y. Si X et Y sont indépendantes la covariance est nulle mais la réciproque est fausse.

Théorème 5.7 *(loi des grands nombres) : Soit $X_1, \ldots, X_n, \ldots$ une suite de v.a. indépendantes et de même loi, d' espérance m et de variance σ^2. Alors quel que soit $\epsilon > 0$, on a*

$$\lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - m\right| \geq \epsilon\right) = 0.$$

démonstration : Posons $Y_n = \frac{X_1+\cdots+X_n}{n}$. Alors $\mathbb{E}(Y_n) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = m$ et $Var(Y_n) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}$. L'inégalité de Bienaymé-Tchebychev implique

$$\forall a > 0,\ \mathbb{P}\left(|Y_n - m| \geq \frac{a\sigma}{\sqrt{n}}\right) \leq \frac{1}{a^2}\ .$$

En particulier si $a = \frac{\sqrt{n}\epsilon}{\sigma}$, on obtient

$$\mathbb{P}(|Y_n - m| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{si} \quad n \to +\infty.$$

C'est le résultat annoncé. □

Corollaire 5.8 *: Soit $\mathcal{E}$ une expérience aléatoire et A un événement pouvant se réaliser avec la probabilité p. Soit $\frac{N_A}{N}$ la fréquence de A au cours de N répétitions indépendantes de l'expérience $\mathcal{E}$. Alors pour tout $\epsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{N_A}{N} - p\right| \geq \epsilon\right) \to 0 \quad \textit{quand} \quad N \to +\infty.$$

démonstration : Considérons les v.a.

$$X_i = \begin{cases} 1 & \text{si } A \text{ se réalise au } i\text{-ème coup} \\ 0 & \text{sinon} \end{cases}$$

Les v.a. X_i $(i = 1, \ldots, N)$ sont indépendantes et de même loi (de Bernoulli), avec $\mathbb{E}(X_i) = p$ et $Var(X_i) = p(1-p)$. De plus $X_1 + \cdots + X_N = N_A$. Le résultat découle donc du théorème 5.7. □

Remarque 5.9 *:* La notion de convergence qui apparaît dans le théorème et son est une notion probabiliste qu'on appelle **convergence en probabilité** : On dit qu'une suite (X_n) de v.a. converge en probabilité vers une v.a. X si

(21) $$\forall \epsilon > 0,\ \mathbb{P}(|X_n - X| \geq \epsilon) \to 0 \ \text{ si } n \to +\infty.$$

La moyenne arithmétique $\frac{X_1+\cdots X_n}{n}$ converge donc vers $m = \mathbb{E}(X_i)$ en probabilité quand $n \to +\infty$. Concrètement cela signifie que pour tout $\epsilon > 0$, l'événement

$$\left[\left|\frac{X_1 + \cdots X_n}{n} - m\right| \geq \epsilon\right]$$

que la moyenne arithmétique s'écarte de m de plus de ϵ, est pratiquement nulle si n est assez grand.
Le corollaire peut donc être considéré comme la justification mathématique de la loi empirique des grands nombres dont on a parlé au chapitre 1.