Chapitre 3:
Variables aléatoires discrètes
Espérance-Variance
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 $x_i \in X(\Omega)$, l'ensemble

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

*Notes du cours de Probabilités de M1 de M. L. Gallardo, Université de Tours, année 2008-2009. 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)

est un événement (i.e. appartient à la tribu $\mathcal{F}$)[2]*. On l'appelle événement «X* ***prend la valeur*** *x_i».*
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]cette condition est toujours satisfaite si Ω est discret et $\mathcal{F} = \mathcal{P}(\Omega)$.

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*

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

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

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

probabilité produit comme expliqué au chapitre 1. On considère la v.a. X=«instant du premier succès»[4] qui est telle que

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

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

$$(2) \qquad [X = n] = E_1 \cap E_2 \cap \cdots \cap E_{n-1} \cap S_n$$

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

$$(3) \qquad \mathbb{P}(X = n) = (1-p)^{n-1}p \quad (n \in \mathbb{N}^*).$$

Définition 3.4 *: La loi de probabilité (3) de la v.a. «instant du premier succès» considérée en (1) s'appelle* ***loi géométrique*** *(ou loi de Pascal).*

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 :

$$\begin{aligned} C_n^k p^k (1-p)^{n-k} &= \frac{1}{k!}\left[(n-k+1)(n-k+2)\cdots n\right] p^k (1-p)^{n-k} \\ (4) \qquad &= \frac{1}{k!}(np)^k (1-p)^n \left(1 - \frac{k-1}{n}\right)\cdots\left(1 - \frac{2}{n}\right)(1-p)^{-k}. \end{aligned}$$

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 (4) tendent tous vers 1, le résultat en découle. $\square$

[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 (2) donc il appartient à la tribu produit.

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

$$p_1x_1 + p_2x_2 + \cdots + p_nx_n$$

peut être interprêtée 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 :

$$\frac{N_1x_1 + N_2x_2 + \cdots + N_nx_n}{N} = \frac{N_1}{N}x_1 + \frac{N_2}{N}x_2 + \cdots + \frac{N_n}{N}x_n,$$

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 $p_1x_1 + p_2x_2 + \cdots + p_nx_n$ 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*

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

*(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**.*

[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.

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.

Proposition 4.3 *(Autre formule pour exprimer $\mathbb{E}(X)$) : Supposons que l'univers Ω où la v.a. X est définie, est un espace* ***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]

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega). \tag{5}$$

démonstration : 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] :

$$\begin{aligned} S &= \sum_{\omega\in[X=x_1]} |X(\omega)|\mathbb{P}(\omega) + \cdots + \sum_{\omega\in[X=x_i]} |X(\omega)|\mathbb{P}(\omega) + \cdots \\ &= |x_1| \sum_{\omega\in[X=x_1]} \mathbb{P}(\omega) + \cdots + |x_i| \sum_{\omega\in[X=x_i]} \mathbb{P}(\omega) + \cdots \\ &= |x_1|\mathbb{P}(X = x_1) + \cdots + |x_i|\mathbb{P}(X = x_i) + \cdots \end{aligned} \tag{6}$$

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 (5).
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 (6))

[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

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 (6) avec $|X(\omega)|$ remplacé par $X(\omega)$ montre que $\mathbb{E}(X)$ est donné par la formule (5). D'où le résultat. □

Corollaire 4.4 *(linéarité de l'espérance) : Soient $X_1, \ldots, X_n$ des v.a. ayant une espérance et $a_1, \ldots, a_n$ des réels. Alors la v.a. $a_1X_1 + \cdots + a_nX_n$*[12] *a une espérance donnée par*

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

démonstration : Supposons pour simplifier que Ω est discret[13]. Il suffit de prouver le résultat pour $n = 2$. Il est facile de voir que $\mathbb{E}(|a_1X_1 + a_2X_2|) \leq |a_1|\mathbb{E}(|X_1|) + |a_2|\mathbb{E}(|X_2|) < +\infty$, donc $a_1X_1 + a_2X_2$ a une espérance d'après la proposition 4.3 et on a

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

4.2.2 Fonction d'une variable aléatoire

Étant donné une v.a. X on considère souvent des v.a. de la forme $Y = g(X)$ qui sont fonction déterministe de X i.e. $Y(\omega) = g(X(\omega))$ où $g : \mathbb{R} \to \mathbb{R}$ est une fonction usuelle. Par exemple $Y = X^2$, $Y = |X|$, $Y = \exp(X)$, etc... Pour trouver si Y a une espérance, il faudrait 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 4.5 *: Soit X une v.a. discrète prenant les valeurs x_i avec les probabilités p_i ($i \in D$), g une application de $\mathbb{R}$ dans $\mathbb{R}$ et $Y = g(X)$. Alors l'espérance de Y existe si et seulement si $\sum_{i\in D} |g(x_i)|p_i < +\infty$ et on peut dans ce cas la calculer par la formule*

$$\mathbb{E}(Y) = \sum_{i\in D} g(x_i)p_i \tag{7}$$

démonstration : On se place pour simplifier dans le cas particulier où l'espace Ω est discret. D'après la proposition 4.3, $\mathbb{E}(Y)$ existe (et vaut alors $\sum_{\omega\in\Omega} g(X(\omega))\mathbb{P}(\omega)$) si et seulement si $\sum_{\omega\in\Omega} |g(X(\omega))|\mathbb{P}(\omega) < +\infty$. À l'aide de la partition $[X = x_i]$ ($x_i \in X(\Omega)$) de Ω, on montre comme en (6) que

$$\sum_{\omega\in\Omega} |g(X(\omega))|\mathbb{P}(\omega) = \sum_{i\in D} |g(x_i)|p_i$$

puis dans le cas où cette série converge, en reprenant le même calcul sans les valeurs absolues, on obtient la formule (7). □

[12] définie par $(a_1X_1 + \cdots + a_nX_n)(\omega) = a_1X_1(\omega) + \cdots + a_nX_n(\omega)$

[13] le théorème est vrai même si Ω n'est pas discret.

Exemple 4.6 *:* 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)[14] 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 4.5 avec la fonction $g(x) = x^r$.

Corollaire 4.7 *(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*

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \tag{8}$$

$$= \left(\sum_{i\in D} p_i (x_i)^2\right) - \left(\sum_{i\in D} p_i x_i\right)^2 \tag{9}$$

démonstration : D'après l'inégalité de Cauchy-Schwarz $\sum_{i\in D} p_i|x_i| \leq \left(\sum_{i\in D} p_i\right)^{1/2} \left(\sum_{i\in D} p_i (x_i)^2\right)^{1/2} < +\infty$ donc X a un moment d'ordre 1. 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 (8) découle alors de la linéarité de l'espérance (corollaire 4.4). La deuxième formule (9) résulte alors du théorème 4.5 (appliqué avec $g(x) = x^2$). □

Remarque 4.8 *:* Pour une v.a. discrète, il est évident que $Var(X) = 0$ si et seulement si X est constante. Avec la 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

$$Var(aX + b) = a^2 Var(X). \tag{10}$$

4.3 Exemples

4.3.1 Loi binomiale et loi de Poisson

Proposition 4.9 *: 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 $\mathbb{E}(X) = Var(X) = \lambda$.

démonstration : 1) X a des moments de tous ordres[15]. Calculons $\mathbb{E}(X)$:

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

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

[15]car elle prend un nombre fini de valeurs.

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 4.5 [16], 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

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

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[17] 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 $\sum_{n=1}^{+\infty} 2^n \mathbb{P}(X = 2^n) = +\infty$.

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 $a^{(i)} \in$

[16] appliqué avec la fonction $g(x) = x^r$.
[17] non truquée

$X_i(\Omega)$ $(i = 1, \ldots, n)$, 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 rapportant 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*

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \tag{11}$$

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 (11) 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*

$$Var(X+Y) = Var(X) + Var(Y). \tag{12}$$

démonstration : En développant $(X+Y)^2$ et en utilisant le corollaire 4.4, 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[18] 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*

$$\mathbb{P}\left(\left|\frac{X-m}{\sigma}\right| \geq a\right) \leq \frac{1}{a^2} \tag{13}$$

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 (13) 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 $a\sigma$.

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*

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

[18] 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.

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

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

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 corollaire 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

$$\forall \epsilon > 0,\ \mathbb{P}(|X_n - X| \geq \epsilon) \to 0 \ \text{ si } n \to +\infty. \tag{14}$$

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.

6 Simulation d'une v.a. discrète (notions)

6.1 Nombres au hasard entre 0 et 9

Considérons une suite $(X_i)_{i\geq 1}$ de variables aléatoires indépendantes et de loi uniforme sur l'ensemble des entiers 0, 1, 2,...,9. Si on procède effectivement à une suite de tirages avec remise dans une urne contenant 10 boules numérotées de 0 à 9, on obtient un échantillon (observé) de valeurs :

(15) 0222 8519 4874 5524 8969 1553 0020 8848 9508 0047 etc...

on dit que c'est une liste de nombres au hasard. Si on fait un très grand nombre de tirages, on obtient ce qu'on appelle une **table de nombres au hasard**. On a groupé ces nombres par paquets de 4 pour des raisons pratiques de lisibilité. Naturellement, on peut produire de tels nombres plus rapidement avec des logiciels informatique, on obtient des pseudo-nombres au hasard, en général de bonne qualité mais ceci est un autre problème.
Il est important de noter qu'aucune table de nombres au hasard n'est universelle (contrairement à une table de logarithmes par exemple) et qu'on peut soi même fabriquer sa propre table qui ne coïncidera pas avec celle donnée ci-dessus.

6.2 Nombres au hasard entre 0 et 99

Proposition 6.1 *: Si on lit une table de nombres au hasard par tranches de 2 chiffres, on obtient une table de nombres au hasard entre* 0 *et* 99

démonstration : Soit $(X_i)_{i\geq 1}$ la suite de variables aléatoires indépendantes et de loi uniforme sur l'ensemble des entiers 0, 1, 2,...,9 dont la réalisation a donné la table de nombres au hasard précédente. Considérons maintenant les variables aléatoires

$$(16)\quad Z_1 = 10X_1 + X_2,\ Z_2 = 10X_3 + X_4,\ Z_3 = 10X_5 + X_6, \\ \ldots, Z_k = 10X_{2k-1} + X_{2k}, \ldots$$

dont la réalisation est la table lue par tranches de 2 chiffres. Les Z_k sont des v.a. indépendantes et de même loi uniforme sur $\{0, 1, 2, \ldots, 99\}$. En effet

$$(17)\quad \mathbb{P}(Z_k = \overline{ij}) = \mathbb{P}(X_{2k-1} = i, X_{2k} = j) \\ = \mathbb{P}(X_{2k-1} = i)\mathbb{P}(X_{2k} = j) = \frac{1}{10} \times \frac{1}{10} = \frac{1}{100}.$$

On démontre de la même façon qu'on obtient des nombres au hasard entre 0 et 999 en lisant une table de nombres au hasard par tranches de 3 chiffres etc...

6.3 Modification d'une table de nombres au hasard

On peut supprimer un certain nombre **d'emplacements** dans une table de nombres au hasard par exemple la 1-ière ligne, la 1-ière colonne, etc... ce qui reste est encore une table de nombres au hasard.
Attention si on supprime systématiquement un nombre par exemple le 3, on n'obtient plus une table de nombres au hasard :

Exemple : Si on supprime tous les chiffres 0, 7, 8, 9 d'une table de nombres au hasard, il reste une table de nombres au hasard entre 1 et 6. Le résultat n'est pas étonnant mais la démonstration rigoureuse nécessite de prouver que **la table obtenue en barrant les** 0**,** 7**,** 8**,** 9 **est une réalisation d'une suite** $(Y_i)_{i\geq 1}$ **de variables aléatoires indépendantes et de loi uniforme sur l'ensemble** $A = \{1, 2, 3, 4, 5, 6\}$: On doit considérer la suite $(X_i)_{i\geq 1}$ qui a produit la table initiale comme un processus aléatoire. Soient

$$T_1 = \inf\{k \geq 1;\ X_k \in A\},\ T_2 = \inf\{k > T_1;\ X_k \in A\}, ...,$$
$$T_n = \inf\{k > T_{n-1};\ X_k \in A\},\ \text{etc...}$$

les instants successifs où le processus passe dans A. Posons

$$Y_1 = X_{T_1},\ Y_2 = X_{T_2}, \ldots, Y_n = X_{T_n}, \ldots$$

Il est clair que la table barrée est la réalisation de la suite des Y_n qui sont des variables aléatoires indépendantes (car les X_k le sont). Il reste à prouver que Y_n est de loi uniforme sur A. Pour $a \in A$, par la formule de la probabilité totale, on a

$$\begin{aligned}\mathbb{P}(Y_n = a) &= \mathbb{P}(X_{T_n} = a)\\ &= \sum_{i=1}^{+\infty} \mathbb{P}(X_i = a | T_n = i)\mathbb{P}(T_n = i).\end{aligned}$$

Mais

$$\begin{aligned}\mathbb{P}(X_i = a | T_n = i) &= \mathbb{P}(X_i = a | 1 \leq X_i \leq 6)\\ &= \frac{\mathbb{P}(X_i = a)}{\mathbb{P}(1 \leq X_i \leq 6)} = \frac{1}{6}\end{aligned}$$

D'où $\mathbb{P}(Y_n = a) = \frac{1}{6}\sum_{i=1}^{+\infty} \mathbb{P}(T_n = i) = \frac{1}{6}$. Q.E.D.

6.4 Simulation d'une variable aléatoire discrète

Le vocabulaire qui suit est utilisé en statistiques :

Définition 6.2 *: 1) On appelle* n***-échantillon*** *d'une loi de probabilité* $\mathcal{L}$ *toute suite finie* $X_1, \ldots, X_n$ *de* n *v.a. indépendantes et de même loi* $\mathcal{L}$*.*

2) On appelle n***-échantillon observé*** *d'une loi* $\mathcal{L}$ *toute suite* $X_1(\omega)$, $\ldots, X_n(\omega)$ *de* n *valeurs prises par un* n*-échantillon* $X_1, \ldots, X_n$ *de loi* $\mathcal{L}$*.*

3) Simuler un n-échantillon de la loi $\mathcal{L}$ (ou de la v.a. X) c'est trouver un procédé explicite[19] *pour obtenir un n-échantillon observé de la loi $\mathcal{L}$.*

Nous allons présenter le principe de la simulation d'une v.a. sur un exemple

Soit X une variable aléatoire de loi donnée par

(18)

i	0	1	2	3
P(X=i)	1/8	3/8	3/8	1/8

Ecrivons les fractions sous forme décimale : $1/8 = 0,125$ et $3/8 = 0,375$.

Considérons une table de nombres au hasard qu'on va lire par tranches de 3 chiffres. On remplace alors par :

0 tout nombre de 000 à 124 inclus
1 tout nombre de 125 à 499 inclus
2 tout nombre de 500 à 874 inclus
3 tout nombre de 875 à 999 inclus.

Avec la table (15), on obtient le 13-échantillon suivant

0 1 1 2 2 1 2 2 0 0 1 2 0

On démontrera en exercice qu'on a bien un 13-échantillon observé de $(Y_i)_{1\leq i\leq 13}$ où les Y_i sont des variables aléatoires i.i.d. de loi donnée par le tableau (18).

Autre méthode : On barre les 8 et les 9 de la table (15) et on remplace par

0 tout 0 rencontré dans la table,
1 tout nombre 1, 2 ou 3,
2 tout nombre 4, 5 ou 6,
3 tout nombre 7.

Ceci donne le 15-échantillon observé suivant

0 1 1 1 2 1 2 3 2 2 2 1 2 2 1

On justifiera en exercice qu'il s'agit bien d'un 15-échantillon observé de la loi (18).

Note à l'attention des lecteurs : Merci de me signaler les coquilles ou erreurs que vous auriez pu remarquer dans ce fichier. Votre attention permettra d'améliorer la prochaine version de ces notes de cours.

[19] utilisant une table de nombres au hasard ou d'un logiciel informatique produisant des nombres au hasard.