Chapitre 1

Combinatoire

Chapitre 2

Modèle probabiliste

Chapitre 3

Conditionnement et indépendance

Chapitre 4

Variables aléatoires

4.1 Variables aléatoires

4.1.1 Définition

Définition (variable aléatoire).([1]) Soit (Ω, P) un espace probabilisé. Une variable aléatoire est une application de Ω dans $\mathbb{R}$.

Convention. On désigne usuellement les variables aléatoires par des lettres majuscules.

4.1.2 Exemples

1. Vladimir vous propose le jeu suivant. L'un de vous deux lance un dé. Si le résultat vaut 1, vous perdez 8 euros ; si le résultat vaut 2, 3 ou 4 vous gagnez 1 euros ; si le résultat est 5 ou 6 vous gagnez 2 euros. On peux modéliser le problème ainsi. On choisit pour univers l'ensemble $\Omega = \{1, ..., 6\}$. On le munit de la mesure de probabilité uniforme que l'on note P. On définit une variable aléatoire G par :

$$G(1) = -8, \quad G(2) = G(3) = G(4) = 1, \quad G(5) = G(6) = 2.$$

   Cette variable aléatoire modélise votre gain lors d'une partie.

2. On lance deux dés et on s'intéresse aux résultats des deux dés ainsi qu'à leur somme. On peux modéliser le problème ainsi. On choisit pour univers l'ensemble $\Omega = \{1, ..., 6\}^2$. On le munit de la mesure de probabilité uniforme que l'on note P. On définit une variable aléatoire X par $X(a, b) = a$, pour tout (a, b) dans Ω. On définit similairement une variable aléatoire Y par $Y(a, b) = b$, pour tout (a, b) dans Ω. On définit enfin une variable aléatoire S par $S = X + Y$. Formellement, il s'agit de la somme de deux applications. Alternativement, on peut décrire S ainsi : $S(a, b) = a + b$ pour tout (a, b) dans Ω.

   La variable aléatoire X modélise le résultat du premier dé ; Y modélise le résultat du deuxième dé ; S modélise la somme des résultats des deux dés.

[1]Nous passons sous silence les questions de mesurabilité, qui sont sans importance pour nous

4.1.3 Notations

Reprenons l'exemple 2 de la section 4.1.2. Intéressons-nous tout d'abord à l'évènement A : "le résultat du premier dé est 3" ([2]). Voici plusieurs manières équivalentes de décrire cet évènement :

$$\begin{aligned} A &= \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\} \\ &= \{(a,b)\in\Omega : a=3\} \\ &= \{(a,b)\in\Omega : X(a,b)=3\} \\ &= \{\omega\in\Omega : X(\omega)=3\}. \end{aligned}$$

Toutes les écritures précédentes sont équivalentes et sont compréhensibles par tout mathématicien. Les probabilistes, par abus d'écriture, préfère décrire cet ensemble ainsi :

$$A = \{X=3\}.$$

De la même manière, les probabilistes écrivent abusivement $B=\{X\le 3\}$ pour définir l'évènement

$$\begin{aligned} B &= \{\omega\in\Omega : X(\omega)\le 3\} \\ &= \{(a,b)\in\Omega : X(a,b)\le 3\} \\ &= \{(a,b)\in\Omega : a\le 3\} \\ &= \{(1,1),\dots,(1,6),(2,1),\dots,(2,6),(3,1),\dots,(3,6)\}. \end{aligned}$$

Similairement, $C=\{2\le X\le 4\}$ est un abus d'écriture correspondant à $C=\{\omega\in\Omega : 2\le X(\omega)\le 4\}$. Toujours via le même abus, les probabilistes écrivent $P(X\le 5)$ plutôt que $P(\{\omega\in\Omega : X(\omega)\le 5\}$ ou encore $P(X=5 \text{ et } Y\le 4)$ plutôt que $P(\{\omega\in\Omega : X(\omega)=5 \text{ et } Y(\omega)\le 4\})$ ou $P(\{(5,1),(5,2),(5,3),(5,4)\})$.

Un premier intérêt de l'introduction des variables aléatoires et des abus d'écritures associés est ainsi de permettre des écritures légères et intuitives dans lesquelles l'univers est sous-entendu ([3]). Formellement, une variable aléatoire est une application de l'univers dans $\mathbb{R}$. Intuitivement, on essaye d'y penser comme à un "nombre aléatoire". Dans notre exemple, on pense à X comme au résultat du premier dé. La modélisation de l'évènement réel "le résultat du premier dé est 3" par $\{X=3\}$ se fait alors assez naturellement.

Remarque. Je vous invite à ne négliger aucun des deux aspects (rigueur et intuition). Expliciter régulièrement ce dont les abus sont des abus est un exercice pas inintéressant.

4.2 Loi d'une variable aléatoire

Théorème 4.1 *Soit (Ω,P) un espace probabilisé. Soit $X:\Omega\to\mathbb{R}$ une variable aléatoire. On définit une application P_X des sous-ensembles de $\mathbb{R}$ dans $\mathbb{R}$ en posant, pour tout $A\subset\mathbb{R}$:*

$$P_X(A) = P(X\in A) = P(\{\omega\in\Omega : X(\omega)\in A\}).$$

Alors P_X est une mesure de probabilité sur R.

[2]Plus exactement, A est l'évènement de notre modèle qui correspond, via notre modélisation, à l'évènement réel "le résultat du premier dé est 3"

[3]Il y a toujours un univers, mais il est souvent compliqué ; on est donc content de pouvoir le cacher

Définition.L'application P_X ainsi associée à une variable aléatoire X est appelée la loi de X.

Exemple. Je reprend les exemples de la section 4.1.2.

1. On a par exemple $P_G(\{1\}) = P(G = 1) = P(\{2, 3, 4\}) = 1/2$.
2. On a par exemple $P_X(\{2\}) = P(X = 2) = P(\{(2, 1), \dots, (2, 6)\}) = 1/6$. On a aussi $P_X([-3, 2]) = P_X(\{1, 2\}) = 1/3$. On peut également vérifier que X et Y ont la même loi.

4.3 Fonction de répartition

Cette partie est la moins fondamentale de ce chapitre.

Définition (fonction de répartition). Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire. On appelle fonction de répartition de X la fonction $F_X : \mathbb{R} \to [0, 1]$ qui à tout réel x associe :

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

Remarque. La fonction de répartition d'une variable aléatoire ne dépend d'elle qu'à travers sa loi. Autrement dit, si deux variables aléatoires X et Y ont la même loi, alors elles ont la même fonction de répartition. En effet, on a alors, pour tout réel x : $F_X(x) = P_X(]-\infty, x]) = P_Y(]-\infty, x]) = F_Y(x)$. La réciproque est également vraie (si deux variables aléatoires ont la même fonction de répartition, elles ont la même loi) mais c'est plus délicat à établir.

Proposition 4.2 *Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire. On note F_X sa fonction de répartition. Alors :*

1. *La fonction F_X est croissante.*
2. *Si a et b sont deux réels vérfiant $a \leq b$, alors on a $P(a < X \leq b) = F_X(b) - F_X(a)$.*
3. *La fonction F_X converge vers 0 en $-\infty$ et vers 1 en $+\infty$.*

Preuve. Les deux premiers points sont des exercices raisonnables. Le dernier est un exercice plus délicat (mais abordable). □

Remarque. À défaut de prouver cette proposition, le minimum est de chercher à comprendre pourquoi elle est raisonnable.

4.4 Indépendance

Définition (indépendance de variables aléatoires). Soit (Ω, P) un espace probabilisé. Soient $X_1, ..., X_n$ des variables aléatoires. On dit qu'elles sont indépendantes si, pour tout A_1,...,A_n sous-ensembles de $\mathbb{R}$, les ensembles $\{X_1 \in A_1\}, ..., \{X_n \in A_n\}$ sont indépendants.

Remarques.

1. On peut paraphraser (de manière plutôt vague) la définition ainsi : des variables aléatoires sont indépendantes si les évènements que l'on peut construire avec ces différentes variables aléatoires sont indépendants.
2. De manière encore plus vague, dire que des variables aléatoires sont indépendantes signifie que savoir ce que valent certaines d'entres elles ne change pas la probabilité que les autres prennent certaines valeurs.

Théorème 4.3 *Soit (Ω, P) un espace probabilisé. Soient $X_1, ..., X_n$ des variables aléatoires. Elles sont indépendantes si et seulement si, pour tout A_1,...,A_n sous-ensembles de $\mathbb{R}$, on a :*

$$P(\{X \in A_1\} \textit{ et } \ldots \textit{ et } \{X \in \ A_n\}) = \prod_i P(X \in A_i).$$

Preuve. La preuve n'est ni tout à fait immédiate, ni difficile, ni très importante. □

Remarque. Explicitons le cas le plus usuel. Deux variables aléatoires X et Y sont indépendantes si et seulement si, pour tout ensemble A et B on a $P(X \in A \text{ et } Y \in B) = P(X \in A)P(Y \in B)$.

Exemple. Les variables aléatoires X et Y du deuxième exemple de la section 4.1.2 sont indépendantes (c'est un exercice pénible avec les outils actuels ; ce sera un exercice facile avec les résultats du chapitre sur les variables aléatoires discrètes).

Chapitre 5

Variables aléatoires discrètes

Dans ce chapitre, nous étudions des variables aléatoires particulières : les variables aléatoires discrètes. Typiquement, ce sont des variables aléatoires qui ne prennent qu'un nombre fini de valeurs ou qui ne prennent que des valeurs entières.

5.1 Définition

Définition (dénombrable). On dit qu'un ensemble est dénombrable s'il existe une bijection entre cet ensemble et $\mathbb{N}$.

Exemples :
- Ensembles dénombrables : l'ensemble des entiers positifs ou nuls $\mathbb{N}$, l'ensemble des entiers relatifs $\mathbb{Z}$, l'ensemble des rationnels $\mathbb{Q}$, l'ensemble des entiers pairs $2\mathbb{Z}$, tout sous-ensemble infini de l'un des ensembles précédents, ...
- Ensembles non dénombrables : l'intervalle $[0, 1]$, l'ensemble des réels $\mathbb{R}$, ...

Définition (variable aléatoire discrète). Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire. On dit que X est une variable aléatoire discrète si $X(\Omega)$ est fini ou dénombrable

Rappel. L'ensemble $X(\Omega)$ est l'image de Ω par X, autrement dit c'est l'ensemble des valeurs que peux prendre l'application X.

Remarque. Typiquement, une variable aléatoire est discrète si elle prend un nombre fini de valeurs ou si elle ne prend que des valeurs entières. On peut adopter cette définition en première lecture.

Exemple de variables aléatoires discrètes : Celles de la section 4.1.2.

5.2 Loi d'une variable aléatoire discrète

La loi d'une variable aléatoire a été définie dans le chapitre 4. C'est un objet complexe (c'est une mesure de probabilité sur $\mathbb{R}$). Dans le cas de variables aléatoires discrètes, elle peut être décrite de manière relativement simple grâce à la proposition suivante.

Proposition 5.1 *Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire discrète. On note P_X sa loi. Alors, pour tout sous-ensemble A de $\mathbb{R}$ on a :*

$$P_X(A) = \sum_{x \in X(\Omega) \cap A} P_X(\{x\}).$$

Remarque. Cette proposition est une manière atroce de dire quelque chose de relativement simple. Expliquons cela sur un exemple.

Exemple. Reprenons la variable aléatoire X des exemples de la section 4.1.2 (X modélise le résultat du lancer d'un dé). On a alors $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$. Appliquons la proposition précédente pour $A = [-1, 2.3]$. On obtient $P_X([-1, 2.3]) = P_X(\{1\}) + P_X(\{2\})$. Autrement dit :

$$P(-1 \leq X \leq 2.3) = P(X = 1) + P(X = 2).$$

On peut énoncer cette égalité ainsi : la probabilité que le résultat d'un dé soit compris entre -1 et 2.3 est égale à la somme de la probabilité que la résultat vaille 1 et de la probabilité que le résultat vaille 2. L'idée est simplement que dire que le résultat d'un dé est compris entre -1 et 2.3 est équivalent à dire que ce résultat vaut 1 ou 2.

Preuve. La preuve est un exercice de difficulté raisonnable et d'intérêt moyen. L'important est de comprendre l'idée. $\square$

Conséquence. Pour spécifier la loi d'une variable aléatoire discrète, il suffit de donner l'ensemble des valeurs qu'elle peut prendre (le $X(\Omega)$) ainsi que la probabilité qu'elle prenne chacune de ces valeurs (les $P_X(\{x\})$, autrement dit les $P(X = x)$). En effet, à partir de ces seules données, on peut récupérer $P_X(A)$ pour tout A : on peut récupérer la loi de X.

Résumé pragmatique. En pratique, quand on demande la loi d'une variable aléatoire discrète, il suffit de donner l'ensemble des valeurs qu'elle peut prendre ainsi que la probabilité qu'elle prenne chacune de ces valeurs (sauf éventuellement si la loi porte un nom).

Exemples.

1. Continuous avec la variable aléatoire X de la section 4.1.2 et explicitons sa loi. On a $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$ et $P(X = 1) = P(X = 2) = \cdots = P(X = 6) = 1/6$. On appelle cette loi la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$ (voir le chapitre 6).
2. Reprenons maintenant la variable aléatoire G de la section 4.1.2 (celle qui modélise le gain dans le jeu de Vladimir) et explicitons sa loi. On a $G(\Omega) = \{-8, 1, 2\}$ et $P(G = -8) = 1/6$, $P(G = 1) = 1/2$, $P(G = 2) = 1/3$.

5.3 Fonction de répartition d'une variable aléatoire discrète

On considère X une variable aléatoire discrète définie sur un univers Ω. On note F_X sa fonction de répartition et P_X sa loi. Grâce à la proposition 5.1 on obtient, pour tout réel a :

$$F_X(a) = \sum_{x \in X(\Omega), x \leq a} P_X(x).$$

Exercice. Construire le graphe de F_G où G est la variable aléatoire de la section 4.1.2 (le gain dans le jeu de Vladimir).

5.4 Espérance

5.4.1 Introduction, motivation de la définition

Statut de la section. L'objectif de cette section est de faire apparaître la notion d'espérance. Cette section me semble intéressante pour développer son intuition, elle est par contre inutile sur le plan formel.

Vladimir vous propose de jouer à son jeu toute la soirée (voir la section 4.1.2). La question est de déterminer s'il est ou non dans votre intérêt de jouer.

Pour cela, on suppose que vous ayez accepté et on essaye d'évaluer vos gains ou vos pertes à la fin de la soirée. On note n le nombre de vois que vous avez joué. On suppose que n est grand. On note n_{-8} le nombre de fois que avez perdu 8 euros, n_1 le nombre de fois que vous avez gagné 1 euros et n_2 le nombre de fois que vous avez gagné 2 euros. On s'attend à avoir :

$$\frac{n_{-8}}{n} \approx P(G=-8) = 1/6, \quad \frac{n_1}{n} \approx P(G=1) = 1/2, \text{ et } \frac{n_2}{n} \approx P(G=2) = 1/3.$$

et donc

$$n_{-8} \approx nP(G=-8) = n/6, \quad n_1 \approx nP(G=1) = n/2, \text{ et } n_2 \approx nP(G=2) = n/3.$$

On s'attend donc à ce que votre gain soit :

$$nP(G=-8)*(-8) + nP(G=1)*1 + nP(G=2)*2$$

ce qui se réécrit

$$\big(P(G=-8)*(-8) + P(G=1)*1 + P(G=2)*2\big)n$$

et qui vaut ici

$$(-1/6)n.$$

Le gain espéré étant négatif, il est préférable pour vous de ne pas jouer. Il faut noter que, pour prendre cette décision, il suffit de connaître le signe de la quantité :

$$P(G=-8)*(-8) + P(G=1)*1 + P(G=2)*2.$$

Il faut également noter que cette quantité s'interprète naturellement comme le gain moyen que l'on peut espérer par partie. Tout cela justifie l'intérêt que l'on porte à cette quantitée et justifie donc l'introduction de la définition de la section suivante.

Remarque. On peut également noter que l'on a ici $P(G>0) = 5/6$ et $P(G<0) = 1/6$. Ces deux quantités (et en particulier le fait que l'on ait $P(G>0) > 1/2$) ne sont pas pertinentes pour notre problème (i.e. pour déterminer le signe probable de notre gain après un grand nombre de partie).

5.4.2 Définition

Définition (espérance).([1]) Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire discrète sur Ω. On définit l'espérance de X, notée $E(X)$, par la formule :

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x). \tag{5.1}$$

Exemple. Reprenons la variable aléatoire G du jeu de Vladimir (voir la section 4.1.2). Rappelons que l'on a $G(\Omega) = \{-8, 1, 2\}$ et $P(X = -8) = 1/6$, $P(X = 1) = 1/2$, $P(X = 2) = 1/3$. On a donc

$$E(G) = \frac{1}{6}(-8) + \frac{1}{2}1 + \frac{1}{3}2 = -\frac{1}{6}.$$

On retrouve la quantité que nous avons mis en évidence dans la section précédente.

Compréhension de la formule. L'espérance d'une variable aléatoire est la *moyenne* des valeurs que peut prendre la variable aléatoire pondérée par la probabilité qu'elle prenne chacune de ces valeurs.

Remarque. L'espérance est ainsi une moyenne avec des pondérations, des coefficients, comme les moyennes de notes. Vous pouvez ainsi récupérer l'intuition que vous avez au sujet de ces dernières moyennes (comme par exemple le fait que la moyenne des notes soit comprise entre la plus mauvaise note et la meilleure note - ce qui se correspond ici par le fait que l'espérance d'une variable aléatoire est comprise entre la plus petite des valeurs qu'elle peut prendre et la plus grande des valeurs qu'elle peut prendre).

5.4.3 Une autre expression pour l'espérance

Cette section peut être omise en première lecture.

Lemme 5.2 *Soit (Ω, P) un espace probabilisé. Soit X un v.a. discrète sur Ω. Soient $A_1, \ldots, A_n$ des évènements disjoints de réunion Ω. Soient $a_1, \ldots, a_n$ des réels. On suppose que sur chaque A_i, X est constante et égale à a_i. Alors on a*

$$E(X) = \sum a_i P(A_i).$$

Idée de la preuve sur un exemple. Supposons par exemple $n = 5$, $a_1 = a_2 = 2$, $a_3 = 3$, $a_4 = a_5 = 6$. On a $X(\Omega) = \{2, 3, 6\}$. On a donc :

$$E(X) = 2P(X = 2) + 3P(X = 3) + 6P(X = 6). \tag{5.2}$$

Remarquons maintenant que l'évènement $\{X = 2\}$ est la réunion disjointe des évènements A_1 et A_2. Par conséquent, $P(X = 2) = P(A_1) + P(A_2)$. Similairement, on obtient $P(X = 3) = P(A_3)$ et $P(X = 6) = P(A_4) + P(A_5)$. En injectant ces égalités dans (5.2) on obtient :

$$E(X) = 2P(A_1) + 2P(A_2) + 3P(A_3) + 6P(A_4) + 6P(A_5),$$

i.e. l'égalité demandé dans le cas de cet exemple. □

[1]On cache les problèmes d'intégrabilités, qui ne sont pas importants pour nous. Ces problèmes sont liés au fait que la formule que nous utilisons pour définir l'espérance n'a pas toujours de sens quand X prend une infinité de valeurs. Il y a des problèmes de convergence.

5.4.4 Quelques propriétés de l'espérance

Proposition 5.3 *Soit* (Ω, P) *un espace probabilisé. Soit* X *un v.a. discrète sur* Ω*. Soit* $f : \mathbb{R} \to \mathbb{R}$ *une application. Alors :*

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

Remarques.

- Dans cette proposition, $f(X)$ est un abus d'écriture qui désigne la variable aléatoire $f \circ X$. Autrement dit, $f(X)$ est l'application de Ω dans $\mathbb{R}$ qui vaut $f(X(\omega))$ pour tout ω dans Ω.
- La formule peut se comprendre (et donc se retenir) ainsi : on fait la moyenne sur toutes les valeurs x possibles pour X de la valeur de f en x pondérée par la probabilité que X vaille x.

Preuve dans le cas où X ne prend qu'un nombre fini de valeurs. La lecture de cette preuve peut être omise en première lecture. Notons $x_1, \ldots, x_n$ les différentes valeurs que peut prendre X. Notons $A_1, \ldots, A_n$ les évènements $\{X = x_1\}, \ldots, \{X = x_n\}$. Ces évènements sont disjoints et de réunion Ω. Remarquons que, sur chacun des A_i, la variable aléatoire $f(X)$ est constante égale à $f(x_i)$. On peut donc appliquer le lemme 5.2. On obtient alors :

$$E(f(X)) = f(x_1)P(A_1) + \cdots + f(x_n)P(A_n).$$

En reprenant la définition des évènements A_i on obtient :

$$E(f(X)) = f(x_1)P(X = x_1) + \cdots + f(x_n)P(X = x_n).$$

La proposition est prouvée dans ce cas particulier. □

Proposition 5.4 *Soit* (Ω, P) *un espace probabilisé. Soient* X *et* Y *deux variables aléatoires.*

1. *Pour tout réel* λ *on a* $E(\lambda X) = \lambda E(X)$*.*
2. *On a* $E(X + Y) = E(X) + E(Y)$*.*
3. *Pour tout réel* c*, on a* $E(c) = c$*.*

Remarques.

- Les deux premières propriétés indiquent que l'espérance est linéaire.
- Dans l'égalité $E(c) = c$ énoncée dans la dernière propriété, il y a un abus d'écriture : dans le membre de gauche, c désigne la variable aléatoire constante égale à c. (Dans le membre de droite il n'y a pas d'abus d'écriture, c désigne bien le réel c.)
- Il faut s'efforcer de réfléchir à ces propriétés jusqu'à les trouver naturelles. On peut les paraphraser ainsi : si on multiple une quantité aléatoire par deux, on multiplie également sa moyenne par deux ; si on additionne deux quantités aléatoires, la moyenne de la somme est la somme des moyennes ; si une quantité aléatoire est en fait constante, sa moyenne est précisément cette constante.

Preuve. La preuve de la première propriété est un exercice raisonnable. La preuve de la troisième propriété est un exercice facile. Dans la suite, qui peut être omise en première lecture, nous prouvons la deuxième propriété dans le cas où X et Y prennent un nombre fini de valeurs. L'idée est de découper l'univers en évènements sur lesquels X et Y sont constants, d'utiliser le lemme 5.2 pour en déduire une expression de $E(X+Y)$, puis de recoller les morceaux.

Pour tout x dans $X(\Omega)$ et tout y dans $Y(\Omega)$, notons $A_{x,y}$ l'évènement :

$$A_{x,y} = \{X = x \text{ et } Y = y\}.$$

Remarquons que, sur cette évènement, la variable aléatoire $X+Y$ est constante égale à $x+y$. Remarquons également que Ω est la réunion disjointe de ces évènements. En appliquant le lemme 5.2 à la variable aléatoire $X+Y$ et à cette famille d'évènement on obtient :

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

Écrivons :

$$E(X+Y) = \sum_{x\in X(\Omega), y\in Y(\Omega)} xP(A_{x,y}) + \sum_{x\in X(\Omega), y\in Y(\Omega)} yP(A_{x,y}) \tag{5.3}$$

et considérons tout d'abord le premier terme du membre de droite. On a :

$$\begin{aligned} \sum_{x\in X(\Omega), y\in Y(\Omega)} xP(A_{x,y}) &= \sum_{x\in X(\Omega)} \left(\sum_{y\in Y(\Omega)} xP(A_{x,y}) \right) \\ &= \sum_{x\in X(\Omega)} x \left(\sum_{y\in Y(\Omega)} P(A_{x,y}) \right). \end{aligned} \tag{5.4}$$

Fixons x et considérons :

$$\sum_{y\in Y(\Omega)} P(A_{x,y})$$

Remarquons que les évènements $A_{x,y} = \{X = x \text{ et } Y = y\}$ (nous ne faisons varier que y) sont disjoints et de réunion $\{X = x\}$. Par conséquent, la somme précédente vaut $P(X = x)$ En injectant ce résultat dans (5.4) on obtient :

$$\sum_{x\in X(\Omega), y\in Y(\Omega)} xP(A_{x,y}) = \sum_{x\in X(\Omega)} xP(X = x) = E(X).$$

Le premier terme du membre de droite de (5.3) vaut donc $E(X)$. On montre de même que le deuxième terme vaut $E(Y)$. L'égalité (5.3) donne alors $E(X+Y) = E(X) + E(Y)$. □

5.5 Variance et écart-type

Définition (variance). Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire discrète sur Ω. La variance $\mathrm{var}(X)$ de X est l'espérance de la variable aléatoire $(X - E(X))^2$:

$$\mathrm{var}(X) = E\big((X - E(X))^2\big).$$

Remarques.

- Formulons cette définition légèrement différement. Notons m l'espérance de X : $m = E(X)$. Posons $Z = (X - m)^2$. C'est une variable aléatoire. (C'est l'application de Ω dans $\mathbb{R}$ qui à ω associe $(X(\omega) - m)^2$.) La variance de X est l'espérance de cette variable aléatoire : $\mathrm{var}(X) = E(Z)$.
- La variance de X est une mesure de la dispersion de X autour de sa moyenne (plus X a tendance à prendre des valeurs éloignées de son espérance $E(X)$, plus sa variance $\mathrm{var}(X)$ est grande).
- La définition peut sembler obscure, elle se justifie par son interprétation géométrique et par le fait qu'elle intervient dans différents théorèmes (notamment dans le théorème centrale limite).
- On appelle écart-type de X la quantité $\sigma_X = \sqrt{\mathrm{var}(X)}$.

On a :

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

En effet, si on note f l'application de $\mathbb{R}$ dans $\mathbb{R}$ définie par $f(x) = (x - E(X))^2$, on a : $\mathrm{var}(X) = E(f(X))$. Il suffit alors d'appliquer la proposition 5.3.

Proposition 5.5 *Soit (Ω, P) un espace probabilisé. Soit X une variable aléatoire discrète sur Ω.*

1. *On a* $\mathrm{var}(X) = E(X^2) - (E(X))^2$.
2. *Si a et b sont deux réels, on a* $\mathrm{var}(aX + b) = a^2 \mathrm{var}(X)$ *et* $\sigma_X = |a|\sigma_X$.

Preuve. Exercices. □

5.6 Indépendance

Proposition 5.6 *Soit (Ω, P) un espace probabilisé. Des variables aléatoires dicrètes $X_1, \dots, X_n$ sur Ω sont indépendantes si et seulement si pour tout x_1 dans $X_1(\Omega)$, ..., pour tout x_n dans $X_n(\Omega)$ on a :*

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n).$$

Exercice. Utiliser cette proposition pour montrer que les variables X et Y du deuxième exemples de la section 4.1.2 sont indépendantes.

Proposition 5.7 *Soit (Ω, P) un espace probabilisé. Soient $X_1, \dots, X_n$ des variables aléatoires discrètes indépendantes sur Ω. Alors :*

$$E(X_1 \dots X_n) = E(X_1) \dots E(X_n).$$

Chapitre 6

Quelques lois discrètes classiques

L'objectif de ce chapitre est de présenter sommairement quelques familles de loi discrètes fréquemment recontrées.

6.1 Loi de Bernoulli

6.1.1 Définition et premières propriétés

Nous introduisons dans cette section une loi qui a peu d'intérêt pour elle même mais qui est la brique élémentaire dans la définition des deux lois que nous étudierons ensuite.

Définition (loi de Bernoulli). Soit X une variable aléatoire définie sur un espace probabilisé (Ω, P). Soit $p \in [0,1]$. On dit que la loi de X est une loi de Bernoulli de paramètre p si :

1. X est à valeurs dans $\{0,1\}$ [1].
2. $P(X=1)=p$ et $P(X=0)=1-p$.

Usages. Il est usuel de parler de "succès" pour désigner l'évènement $\{X=1\}$ et d'"échec" pour désigner l'évènement $\{X=0\}$.

6.1.2 Exemple

On tire une boule au hasard dans une urne contenant 5 boules rouges et 3 boules noires. On s'intéresse à l'évènement "la boule tirée est rouge". Voilà une manière de modéliser cette expérience :

- On numérote de 1 à 5 les boules rouges, de 6 à 8 les boules noires. On adopte pour univers l'ensemble $\Omega = \{1, 2, \ldots, 8\}$. Par exemple, dans notre modélisation, l'élément 3 de notre univers correspond à l'évènement réel "on tire la boule numéro 3 (et donc une boule rouge)".
- On adopte la mesure de probabilité uniforme P sur Ω.
- On s'intéresse à l'évènement $A = \{1, 2, \ldots, 5\}$.

[1] L'important est en fait que $P(X \in \{0,1\}) = 1$. La même remarque s'applique à d'autres endroits du cours

Dans ce modèle, on peut calculer $P(A)$. On obtient, sans surprises, $P(A) = 5/8$. Nous n'avons pas utilisé de variable aléatoire dans cette modélisation. Nous en introduisons une de manière un peu artificielle (l'intérêt apparaitra dans la suite du chapitre) :

– Soit X la variable aléatoire définie par $X(1) = X(2) = \cdots = X(5) = 1$ et $X(6) = X(7) = X(8) = 0$.

Cette variable aléatoire prend ses valeurs dans $\{0, 1\}$. C'est donc une variable aléatoire de Bernoulli. Son paramètre est $P(X = 1) = P(\{1, 2, \ldots, 5\}) = 5/8$. L'évènement qui nous intéresse est $A = \{X = 1\}$. Sa probabilité est donc le paramètre de notre variable aléatoire de Bernoulli : $5/8$.

6.1.3 Espérance et variance d'une loi de Bernoulli

Proposition 6.1 *Soit X une variable aléatoire de loi de Bernoulli de paramètre $p \in [0, 1]$. On a alors $E(X) = p$ et $\mathrm{var}(X) = p(1-p)$.*

Preuve. C'est un exercice facile. □

6.2 Loi binomiale

6.2.1 Définition et premières propriétés

Définition (Loi binomiale). Soit X une variable aléatoire définie sur un espace probabilisé (Ω, P). Soient $n \geq 1$ un entier et $p \in [0, 1]$. On dit que la loi de X est une loi binomiale de paramètres n et p si :

1. X est à valeurs dans $\{0, 1, \ldots, n\}$.
2. Pour tout $k \in \{0, 1, \ldots, n\}$, on a $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Le résultat suivant justifie l'intérêt porté à cette loi :

Théorème 6.2 *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre p. On pose $X = X_1 + \cdots + X_n$. Alors la loi de X est une loi binomiale de paramètres n et p.*

Preuve. Il s'agit d'établir les deux points suivants :

1. X prend ses valeurs dans $\{0, 1, \ldots, n\}$.
2. Pour tout $k \in \{0, 1, \ldots, n\}$, on a $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Le premier point résulte du fait que X est la somme de n variables aléatoires prenant leurs valeurs dans $\{0, 1\}$. Prouvons maintenant le second point. Pour cela, fixons $k \in \{0, 1, \ldots, n\}$. Il s'agit d'établir l'égalité suivante :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Pour simplifier la présentation, supposons $n = 4$ et $k = 3$. Il s'agit, dans ce cas, d'établir l'égalité suivante :

$$P(X = 3) = 4p^3(1-p). \tag{6.1}$$

Remarquons que l'évènement $\{X = 3\}$ est la réunion disjointes des quatre évènements suivants :

$$\begin{aligned} A_1 &= \{X_1 = 0 \text{ et } X_2 = 1 \text{ et } X_3 = 1 \text{ et } X_4 = 1\}, \\ A_2 &= \{X_1 = 1 \text{ et } X_2 = 0 \text{ et } X_3 = 1 \text{ et } X_4 = 1\}, \\ A_3 &= \{X_1 = 1 \text{ et } X_2 = 1 \text{ et } X_3 = 0 \text{ et } X_4 = 1\}, \\ A_4 &= \{X_1 = 1 \text{ et } X_2 = 1 \text{ et } X_3 = 1 \text{ et } X_4 = 0\}. \end{aligned}$$

Par conséquent, on a

$$P(X = 3) = P(A_1) + P(A_2) + P(A_3) + P(A_4). \tag{6.2}$$

Par indépendance, on obtient :

$$P(A_1) = P(X_1 = 0)P(X_2 = 1)P(X_3 = 1)P(X_4 = 1).$$

Comme les X_i sont des variables aléatoires de Bernoulli de paramètre p, on obtient alors :

$$P(A_1) = (1 - p)p^3.$$

De même, on a $P(A_2) = P(A_3) = P(A_4) = p^3(1 - p)$. Grâce à (6.2) on en déduit alors (6.1).

Les différents facteurs de $4p^3(1 - p)$ se comprennent ainsi :

1. 4 est le nombre de manières de choisir les 3 X_i qui valent 1 et le X_i qui vaut 0. Dans le cas général, le nombre 4 est remplacé par le nombre de manières de choisir les k X_i qui valent 1 et les $n - k$ X_i qui valent 0, i.e. $\binom{n}{k}$.
2. p^3 est la probabilité que 3 X_i fixés vaillent 1. Dans le cas général, p^3 est remplacé par p^k.
3. $(1 - p)$ est la probabilité que 1 X_i fixé vaille 0. Dans le cas général, $(1 - p)$ est remplacé par $(1 - p)^{n-k}$.

Dans le cas général ($0 \leq k \leq n$ quelconques), en suivant les mêmes idées, on obtient ainsi $P(X = k) = \binom{n}{k}p^k(1 - p)^{n-k}$. Cela conclut la preuve. □

Remarques.

1. Comme les X_i prennent leurs valeurs dans $\{0, 1\}$, $X = X_1 + X_2 + \cdots + X_n$ est le nombre de X_i valant 1.
2. Reprenons la terminologie usuelle pour décrire les évènements faisant intervenir des variables de loi de Bernoulli. Utilisons l'expression "succès à l'expérience i" pour décrire l'évènement $\{X_i = 1\}$. Utilisons l'expression "échec à l'expérience i" pour décrire l'évènement $\{X_i = 0\}$. Le théorème précédant peut alors s'énoncer ainsi : si on effectue n expériences aléatoires indépendantes et qu'à chaque expérience, la probabilité de succès est p, alors le nombre de succès suit une loi binomiale de paramètres n et p.

6.2.2 Exemple

On tire 5 fois, avec remise, une boule dans une urne contenant 5 boules rouges et 3 boules noires. On s'intéresse au nombre de fois que l'on a tiré une boule rouge.

Modélisation. Soient $X_1, X_2, \ldots, X_5$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre 5/8. Soit $S = X_1 + X_2 + \cdots + X_5$.

Rapport entre le modèle et l'expérience réelle. L'évènement $\{X_i = 1\}$ correspond à l'évènement réel "la i-ème boule tirée est rouge". L'évènement $\{X_i = 0\}$ correspond à l'évènement réel "la i-ème boule tirée est noire". La variable aléatoire S, qui compte le nombre de X_i valant 1, correspondant donc, dans la réalité, au nombre de fois que l'on a tiré une boule rouge.

Question vague. L'indépendance entre les X_i vous parait-il être une hypothèse raisonnable ? Le fait que le paramètre des X_i soit 5/8 vous parait-il raisonnable ?

Conclusion. Comme S est la somme de 5 variables aléatoires indépendantes de loi de Bernoulli de paramètre 5/8, le théorème 6.2 nous indique que S suit une loi binomiale de paramètres 5 et 5/8.

Remarque. D'un point de vue pratique (dans une copie), on peut s'épargner l'explicitation de la modélisation en paraphrasant ce qui précède ainsi. On effectue 5 tirages indépendants, à chaque tirage la probalilité de tirer une boule rouge est de 5/8. Par conséquent, le nombre de fois que l'on tire une boule rouge suit une loi binomiale de paramètres 5 et 5/8.

6.2.3 Espérance et variance d'une loi binomiale

Proposition 6.3 *Soit X une variable aléatoire suivant une loi binomiale de paramètres n et p. Alors :*

1. $E(X) = np$.
2. $\mathrm{var}(X) = np(1-p)$.

Preuve du premier point. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre p. Posons $S = X_1 + \cdots + X_n$. Par le théorème 6.2, S et X ont la même loi. Ils ont donc la même espérance. On a donc :

$$E(X) = E(S) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$$

où nous avons utilisé la linéarité de l'espérance à la dernière étape. Par ailleurs, comme chacune des variables X_i suit une loi de Bernoulli de paramètre p, chacune d'elle est d'espérance p. On en déduit $E(X) = np$. □

Remarque. On peut alternativement essayer de prouver ce résultat à partir de la formule définissant l'espérance, qui donne ici :

$$E(X) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k.$$

La suite de cette section est dédiée à la preuve du deuxième point de la proposition ci-dessus. Elle peut être omise en première lecture. Nous aurons besoin du lemme suivant :

Lemme 6.4 *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes. Alors :*

$$\mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n).$$

Remarque. Ce lemme n'est pas vrai sans l'hypothèse d'indépendance. Ainsi, $\mathrm{var}(X + X) = \mathrm{var}(2X) = 4\mathrm{var}(X)$.

Preuve du lemme. Pour simplifier les notations, on se place dans le cas $n = 2$. Il est alors plus agréable d'appeler les variables X et Y. Nous avons donc à prouver (sous l'hypothèse X et Y indépendantes) :

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

Introduisons les variables aléatoires recentrées $\widetilde{X} = X - E(X)$ et $\widetilde{Y} = Y - E(Y)$. Remarquons :

- $\mathrm{var}(X) = E(\widetilde{X}^2)$;
- $\mathrm{var}(Y) = E(\widetilde{Y}^2)$;
- $\mathrm{var}(X + Y) = E((\widetilde{X} + \widetilde{Y})^2)$.

Les deux premiers points sont immédiats. Le dernier se vérifie ainsi. La variance de $X+Y$ est l'espérance de la variable aléatoire $Z = (X + Y - E(X + Y))^2$. Mais, par linéarité de l'espérance, on a $Z = (\widetilde{X} + \widetilde{Y})^2$. Cela donne le troisième point.

Le problème est donc ramené à la preuve de l'égalité suivante :

$$E((\widetilde{X} + \widetilde{Y})^2) = E(\widetilde{X}^2) + E(\widetilde{Y}^2).$$

Or, en développant le carré et par linéarité de l'espérance on obtient :

$$E((\widetilde{X} + \widetilde{Y})^2) = E(\widetilde{X}^2) + 2E(\widetilde{X}\widetilde{Y}) + E(\widetilde{Y}^2).$$

Le problème est donc ramené à montrer que le terme $E(\widetilde{X}\widetilde{Y})$ est nul. En utilisant le fait que X et Y sont indépendantes, on peut vérifier que $\widetilde{X}$ et $\widetilde{Y}$ sont indépendantes. On a donc :

$$E(\widetilde{X}\widetilde{Y}) = E(\widetilde{X})E(\widetilde{Y}).$$

Mais $E(\widetilde{X}) = E(X - E(X)) = E(X) - E(E(X))$ (par linéarité), d'où $E(\widetilde{X}) = E(X) - E(X) = 0$ (car l'espérance d'une constante est égale à cette constante). Ainsi on a bien

$$E(\widetilde{X}\widetilde{Y}) = 0,$$

ce qui conclut. □

Remarque obscure. Ce qui précède est une application du théorème de Pythagore.

Preuve du deuxième point de la proposition. Comme pour la preuve du premier point, on se ramène à étudier la variance de $S = X_1 + \cdots + X_n$ (je reprend les mêmes notations). Comme les X_i sont indépendantes, le lemme précédant nous donne :

$$\mathrm{var}(X) = \mathrm{var}(S) = \mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n).$$

Comme les X_i suivent une loi de Bernoulli de paramètre p, elles admettent toutes $p(1-p)$ pour variance. Cela conclut la preuve. □

6.3 Loi géométrique

6.3.1 Définition et premières propriétés

Définition (Loi géométrique). Soit X une variable aléatoire définie sur un espace probabilisé (Ω, P). Soit $p \in]0,1]$. On dit que la loi de X est une loi géométrique de paramètre p si :

1. X prend ses valeurs dans $\mathbb{N} \setminus \{0\}$ (l'ensemble des entiers supérieurs ou égaux à 1).
2. Pour tout $k \in \mathbb{N} \setminus \{0\}$, on a $P(X = k) = (1-p)^{k-1}p$.

Le résultat suivant justifie l'intérêt porté à cette loi :

Théorème 6.5 *Soient* [2] $X_1, X_2, X_3, \ldots$ *des variables aléatoires indépendantes* [3] *de loi de Bernoulli de paramètre p. On pose*

$$X = \min\{k : X_k = 1\}.$$

Alors X est bien définie [4] *et X suit une loi géométrique de paramètre p.*

Preuve. Il s'agit d'établir les trois points suivants :

1. X est bien définie [4].
2. X prend ses valeurs dans $\mathbb{N} \setminus \{0\}$ [5].
3. Pour tout $k \in \mathbb{N} \setminus \{0\}$, on a $P(X = k) = (1-p)^{k-1}p$.

Commençons par prouver les deux premiers points (cette partie de la preuve peut être omise en première lecture). Soit $n \geq 1$ un entier. On a :

$$P(\text{Les } X_k \text{ sont tous nuls}) \leq P(X_1 = 0, \ldots, X_n = 0).$$

Par indépendance, on en déduit :

$$P(\text{Les } X_k \text{ sont tous nuls}) \leq P(X_1 = 0) \ldots P(X_n = 0).$$

Les X_k suivant une loi de Bernoulli de paramètre p, on obtient :

$$P(\text{Les } X_k \text{ sont tous nuls}) \leq (1-p)^n.$$

En faisant tendre n vers l'infini et en utilisant le fait que p appartient à $]0,1]$, on en déduit :

$$P(\text{Les } X_k \text{ sont tous nuls}) = 0.$$

---

[2] L'existence d'une telle suite soulève des problèmes théoriques ; d'un point de vue pratique ce n'est pas important.

[3] L'indépendance signifie que, pour tout $n \geq 1$, les variables $X_1, \ldots, X_n$ sont indépendantes.

[4] Voilà un énoncé plus précis. Avec probabilité 1, il existe un entier k tel que $X_k = 1$. Par conséquent, avec probabilité 1, il existe un plus petit entier k tel que X_k=1 et X_k et donc bien défini.

[5] Nous allons en réalité montrer que cette propriété est vraie avec probabilité 1.

Par conséquent, avec probabilité 1, il existe un entier k tel que $X_k = 1$. Avec probabilité 1, l'expression $X = \min\{k : X_k = 1\}$ a donc un sens et définit un entier supérieur ou égal à 1. Les deux premiers points sont prouvés.

Montrons maintenant le dernier point. Soit $k \geq 1$ un entier. On a :

$$P(X = k) = P(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1).$$

Par indépendance des X_i on en déduit :

$$P(X = k) = P(X_1 = 0) \ldots P(X_{k-1} = 0)P(X_k = 1).$$

Comme les X_i suivent une loi de Bernoulli de paramètre p, on en déduit finalement :

$$P(X = k) = (1 - p)^{k-1}p.$$

Cela conclut. □

Remarques.

1. Reprenons la terminologie usuelle pour décrire les évènements faisant intervenir des variables de loi de Bernoulli. Utilisons l'expression "succès à l'étape i" pour décrire l'évènement $\{X_i = 1\}$. Utilisons l'expression "échec à l'étape i" pour décrire l'évènement $\{X_i = 0\}$. Le théorème précédant peut alors s'énoncer ainsi. Si on effectue des expériences aléatoires indépendantes et qu'à chaque expérience la probabilité de succès est p, alors :
   - On finit par réussir.
   - Le numéro de l'expérience lors de laquelle on réussit pour la première fois suit une loi géométrique de paramètre p.
2. La formule $P(X = k) = (1 - p)^{k-1}p$ se comprend ainsi : pour que X vaille k, il faut et il suffit que les $k - 1$ premiers essais soient des échecs (d'où le facteur $(1 - p)^{k-1}$) et que l'essai suivant soit un succès (d'où le facteur p).

6.3.2 Exemple

On considère la même urne que précédemment. On tire avec remise une boule jusqu'à obtenir une boule rouge. On s'intéresse au nombre de tirages effectués.

Modélisation. Soient $X_1, X_2, \ldots$ des variables aléatoires indépendantes de loi de Bernoulli de paramètre $5/8$. Soit S le plus petit entier k tel que $X_k = 1$.

Rapport entre le modèle et l'expérience réelle. Nous nous sommes donné une infinité de variables aléatoires X_i. L'évènement $\{X_i = 0\}$ correspond, dans le cas où on effectue au moins i tirages, à l'évènement réel "la i-ème boule tirée est noire". Si on effectue moins de tirages, l'évènement $\{X_i = 0\}$ ne correspond à rien dans la réalité.

Conclusion. Par le théorème 6.5, N suit une loi géométrique de paramètre p.

Remarque. D'un point de vue pratique (dans un examen), on peut s'épargner l'explicitation de la modélisation en paraphrasant ce qui précède ainsi. On effectue des tirages indépendants, à chaque tirage la probalilité de tirer une boule rouge est de $5/8$. Par conséquent, le nombre de tirages nécessaires pour obtenir une boule rouge suit une loi géométrique de paramètre $5/8$.

6.3.3 Espérance d'une loi géométrique

Proposition 6.6 *Soit X une variable aléatoire suivant une loi géométrique de paramètre p. Alors $E(X) = 1/p$.*

Preuve. Cette preuve peut être réservée à une seconde lecture. On a :

$$E(X) = \sum_{k=1}^{\infty} p(1-p)^{k-1}k.$$

Soit $\phi :]-1, 1[\to \mathbb{R}$ définie par

$$\phi(x) = \sum_{k=0}^{\infty} x^k.$$

La série entière définissant cette fonction a pour rayon de convergeance 1. La fonction ϕ est donc bien définie, elle est dérivable et on a, pour tout x :

$$\phi'(x) = \sum_{k=1}^{\infty} kx^{k-1}.$$

On constate :

$$E(X) = p\phi'(1-p).$$

Or, pour tout x, on a $\phi(x) = 1/(1-x)$ et donc $\phi'(x) = 1/(1-x)^2$. D'où :

$$E(X) = p/(1-(1-p))^2 = 1/p.$$

$\square$

Exemple. Par exemple, si on lance un dé jusqu'à obtenir un 1, alors le nombre moyens d'essais nécessaires est $1/(1/6)$ c'est-à-dire 6.

Chapitre 7

Variables aléatoires à densité

Chapitre 8

Quelques lois à densité classiques

Chapitre 9

Théorèmes limites