Variables aléatoires sur un espace probabilisé fini

Lycée Berthollet, MPSI1 2023-24

Dans tout ce chapitre, on se place sur un espace probabilisé **fini** (Ω, P).

Remarque 1 Il est important de comprendre que les définitions de certains objets (variables aléatoires, *etc*.) données cette année sont **spécifiques** au cas fini et ne sont pas valables dans le cas général. Par exemple toute application partant de Ω est une variable aléatoire, mais ce ne sera plus le cas lorsque Ω sera infini.

I Définitions

Définition 2

Une *variable aléatoire (v.a.)* sur Ω est une application $X : \Omega \to E$.

Si $E = \mathbb{R}$, on dit que la v.a. est *réelle (v.a.r.)*.

Plus généralement, si $E \subset \mathbb{R}$, on convient de changer l'ensemble d'arrivée de X en $\mathbb{R}$, ce qui fait de X une v.a.r.

Exemples 3

1. On modélise un jeu de pile ou face à un coup, où le joueur gagne 1 (euro) si la pièce (éventuellement truquée) tombe sur pile et rien si elle tombe sur face, par l'espace (Ω, P). On a alors $\Omega = \mathcal{P} \sqcup \mathcal{F}$, avec $\mathcal{P}$ l'évènement “la pièce tombe sur pile” et $\mathcal{F}$ l'évènement “la pièce tombe sur face”. L'application $X : \Omega \to \mathbb{R}$ définie par $X(\omega) = 1$ si $\omega \in \mathcal{P}$ et $X(\omega) = 0$ si $\omega \in \mathcal{F}$, est la v.a.r. qui représente le *gain* du joueur. Son image est incluse dans $\{0, 1\}$.
2. Si on lance un dé pipé, dont la probabilité d'obtenir 6 est $\frac{1}{2}$ et celle d'obtenir un $i \neq 6$ donné est $\frac{1}{10}$, le résultat du dé sera modélisé par une v.a.r. $Y : \Omega \to \mathbb{R}$. Elle vérifie $Y(\Omega) = [\![1, 6]\!]$.
3. Si on lance deux dés non pipés, alors l'application S, qui fait correspondre à chaque $\omega \in \Omega$ la somme des résultats des deux dés, est une v.a.r. $S : \Omega \to \mathbb{R}$. On a $S(\Omega) = [\![2, 12]\!]$.
4. Si on lance n fois la pièce précédente, avec la même règle de gain pour chaque coup, le gain total est représenté par une v.a.r. S_n. On a ici $S_n(\Omega) \subset [\![0, n]\!]$.

Notation 4

Pour une v.a. $X : \Omega \to E$, $A \subset E$ et $x \in E$, on note $(X \in A)$ l'événement $X^{-1}(A)$ et $(X = x)$ l'événement $X^{-1}(\{x\})$.

Si de plus $E \subset \mathbb{R}$ (*i.e.* X est une v.a.r.), on note, en “étendant” l’ensemble d’arrivée de X à $\mathbb{R}$, $(X \geq x)$ l’événement $X^{-1}([x,+\infty[)$, *etc.*

On note alors $\mathrm{P}(X \in A)$, $\mathrm{P}(X = x)$, $\mathrm{P}(X \geq x)$... les probabilités de ces évènements.

Exemple 5 Dans l’exemple précédent de n lancers d’une pièce, pour $k \in [\![0,n]\!]$ fixé, la probabilité de gagner moins de k euros est $\mathrm{P}(S_n < k)$. Savez-vous calculer cette probabilité ?

Définition 6

Pour X un v.a. sur Ω, on définit la *loi de X* comme l’application

$$\mathrm{P}_X : \left\{ \begin{array}{ccc} \mathcal{P}(X(\Omega)) & \longrightarrow & [0,1] \\ A & \longmapsto & \mathrm{P}_X(A) = \mathrm{P}(X \in A) \end{array} \right. .$$

Remarque 7 On dit aussi que X *suit la loi* P_X.

Notation 8

Pour deux v.a. $X, Y \in E^{\Omega}$ on note $X \sim Y$ ssi $\mathrm{P}_X = \mathrm{P}_Y$, *i.e.* X et Y ont même loi.

Cela définit clairement une relation d’équivalence sur E^{Ω}.

La notation P_X est justifiée par le résultat suivant :

Proposition 9

Pour toute v.a. X sur Ω, sa loi P_X est une probabilité sur $X(\Omega)$.

Elle est uniquement déterminée par la donnée des $\mathrm{P}(X = x) = \mathrm{P}_X(\{x\})$, pour $x \in X(\Omega)$:

$$\forall A \subset X(\Omega), \quad \mathrm{P}_X(A) = \sum_{x \in A} \mathrm{P}_X(\{x\}) \quad \left(\textit{i.e. } \mathrm{P}(X \in A) = \sum_{x \in A} \mathrm{P}(X = x) \right).$$

Démonstration: Soit X une v.a. Comme Ω est fini, alors $X(\Omega)$ aussi. Par définition de P_X, elle va bien de $\mathcal{P}(X(\Omega))$ vers $[0,1]$. Par ailleurs $\mathrm{P}_X(X(\Omega)) = \mathrm{P}\left(X^{-1}(X(\Omega))\right)$. D’après les résultats ensemblistes, $\Omega \subset X^{-1}(X(\Omega))$. Mais par ailleurs, toute image réciproque de partie par X est dans son ensemble de départ Ω, donc $X^{-1}(X(\Omega)) \subset \Omega$ et donc $X^{-1}(X(\Omega)) = \Omega$. Ainsi $\mathrm{P}_X(X(\Omega)) = \mathrm{P}(\Omega) = 1$, puisque P est une probabilité sur Ω.

Enfin, si $A, B \subset X(\Omega)$ sont disjoints, alors d’une part, $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$ et d’autre part, $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B) = \varnothing$, donc, comme P est une probabilité sur Ω,

$$\mathrm{P}_X(A \cup B) = \mathrm{P}\left(X^{-1}(A \cup B)\right) = \mathrm{P}\left(X^{-1}(A) \cup X^{-1}(B)\right) = \mathrm{P}\left(X^{-1}(A)\right) + \mathrm{P}\left(X^{-1}(B)\right) = \mathrm{P}_X(A) + \mathrm{P}_X(B).$$

Ainsi P_X est bien une probabilité sur $X(\Omega)$.

Soit $A \subset \Omega$. Comme $A = \bigsqcup_{x \in A} \{x\}$ et P_X est une probabilité, alors $\mathrm{P}_X(A) = \sum_{x \in A} \mathrm{P}_X(\{x\})$. □

En pratique, lorsqu’on donne ou demande la loi d’une v.a. X, c’est donc de $X(\Omega)$ et des probabilités $\mathrm{P}(X = x)$, pour $x \in X(\Omega)$, dont on parle.

Parfois, on donne seulement un ensemble $F \supset X(\Omega)$ et les $\mathrm{P}(X = x)$, pour $x \in F$. Cela suffit pour travailler avec X. C’est par exemple le cas pour une variable de Bernoulli (*i.e.* à valeurs dans $\{0,1\}$) dont on ne connaît pas le paramètre $p \in [0,1]$ (*i.e.* $p = \mathrm{P}(X = 1)$). Si $p = 0$, alors il se pourrait que la valeur 1 ne soit pas atteinte et si $p = 1$, c’est le cas de la valeur 0. On dit alors que la loi est donnée par

$$X(\Omega) \subset \{0,1\}, \; \mathrm{P}(X = 0) = 1 - p \; \text{ et } \; \mathrm{P}(X = 1) = p.$$

Exemple 10 Reprenons l'exemple du dé pipé précédent et de la v.a. Y retournant le résultat du dé. La loi P_Y est la probabilité sur $[\![1,6]\!]$ déterminée de manière unique par

$$\mathrm{P}_Y(\{6\}) = \frac{1}{2} \quad \text{et} \quad \forall i \in [\![1,5]\!],\ \mathrm{P}_Y(\{i\}) = \frac{1}{10}.$$

Souvent, les v.a. considérées sont des "fonctions" de v.a. plus simples :

Définition 11
Pour une v.a. $X : \Omega \to E$ et $f : E \to F$ une application, on note $f(X)$ l'application $f \circ X$ et on l'appelle *v.a. image de X par f*. C'est une v.a. sur Ω à valeurs dans F.

Remarque 12 D'après la proposition précédente, sa loi $\mathrm{P}_{f(X)}$ est uniquement déterminée par les nombres $\mathrm{P}(f(X) = y) = \mathrm{P}\left(X \in f^{-1}(\{y\})\right)$, pour $y \in f(X(\Omega))$, donc uniquement déterminée par f et la loi de X :

$$\forall X, Y \in E^{\Omega},\ \forall f \in F^E,\quad X \sim Y \Longrightarrow f(X) \sim f(Y).$$

Exercice 13 Toujours avec l'exemple du dé pipé, on imagine qu'un joueur lance le dé et gagne le carré de la différence entre le résultat du dé et 3 (en euros). La v.a. représentant le gain est alors $Z = (Y-3)^2$.
Vérifier que sa loi est une probabilité sur l'ensemble $Z(\Omega) = \{0,1,4,9\}$ et que

$$\mathrm{P}_Z(\{0\}) = \frac{1}{10},\quad \mathrm{P}_Z(\{1\}) = \frac{1}{5},\quad \mathrm{P}_Z(\{4\}) = \frac{1}{5} \quad \text{et} \quad \mathrm{P}_Z(\{9\}) = \frac{1}{2}. \quad \oplus$$

Enfin, il est possible de définir des lois conditionnelles :

Définition 14
Pour $X \in E^{\Omega}$ et $A \subset \Omega$ de probabilité non nulle, on définit l'application

$$\mathrm{P}_A(X \in \cdot) : \left\{ \begin{array}{ccc} \mathcal{P}(X(\Omega)) & \longrightarrow & [0,1] \\ B & \longmapsto & \mathrm{P}_A(X \in B) = \mathrm{P}(X \in B | A), \end{array} \right.$$

qu'on appelle loi de X *sachant* A.

On a alors la

Proposition 15
L'application $\mathrm{P}_A(X \in \cdot)$ *est une probabilité sur* $X(\Omega)$.

Démonstration: Il suffit de remarquer que c'est la loi $(\mathrm{P}_A)_X$ de la v.a. X lorsqu'on la considère sur l'espace probabilisé (Ω, P_A). □

II Lois usuelles

1 Variables aléatoires constantes

Comme leur nom l'indique, ce sont les v.a. $X : \Omega \to E$ telles qu'il existe $x_0 \in E$ vérifiant

$$\forall \omega \in \Omega,\quad X(\omega) = x_0.$$

2 Variables aléatoires uniformes

Définition 16

Soient $X : \Omega \to E$ et $F \subset E$.

La v.a. X est *uniforme d'image F* ssi $X(\Omega) = F$ et P_X est une probabilité uniforme sur F, *i.e.*

$$\forall x \in F, \quad \mathrm{P}(X = x) = \frac{1}{\operatorname{Card} F}.$$

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

Exemples 17 Une telle loi intervient lorsque les résultats d'une expérience aléatoire sont équiprobables, suite à une hypothèse explicite (dé non pipé, pièce non truquée ou équilibrée, boules d'un urne indiscernables au toucher, *etc.*) ou implicite (on dit par exemple qu'on choisit "au hasard" un nombre entre 1 et 100, ou autre cas où la loi d'un tirage aléatoire n'est pas spécifiée, mais où le problème n'est pas raisonnable si on ne suppose pas la loi uniforme). Par exemple, si X est (la v.a. modélisant) le résultat du lancer d'un dé non pipé, alors $X \sim \mathcal{U}([\![1,6]\!])$. Si Y modélise le résultat d'un lancer de pièce équilibrée, $Y \sim \mathcal{U}(\{\text{pile}, \text{face}\})$.

3 Variables aléatoires de Bernoulli

Définition 18

Soit $p \in [0,1]$.

Une v.a.r. X est une v.a. *de Bernoulli de paramètre p* ssi

$$X(\Omega) \subset \{0,1\} \quad \text{et} \quad \mathrm{P}(X = 1) = p.$$

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

On a alors $\mathrm{P}(X = 0) = 1 - p$ et on pose classiquement $q = 1 - p$.

Remarque 19 Il se peut que $X(\Omega) = \{0\}$ (lorsque $p = 0$) ou $X(\Omega) = \{1\}$ (lorsque $p = 1$).

Remarque 20 Toute v.a.r. X telle que $X(\Omega) \subset \{0,1\}$ vérifie $X \sim \mathcal{B}(\mathrm{P}(X = 1))$.

Exemples 21

1. Rappelons que la fonction "caractéristique" $\mathbf{1}_A$ (qu'on appelle plutôt dans ce contexte fonction "indicatrice") d'un événement $A \subset \Omega$, est définie ainsi : pour $\omega \in \Omega$, $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ si $\omega \notin A$. Cette application est une v.a. de Bernoulli de paramètre $\mathrm{P}(A)$.
2. Lorsqu'une pièce à une probabilité p de tomber sur pile et qu'un joueur gagne 1 euro si elle tombe sur pile et 0 euro sinon, la v.a. X de gain vérifie $X \sim \mathcal{B}(p)$.
3. Lorsqu'une urne contient N boules dont R boules roses, la v.a. Y définie par $Y = 1$ si la boule tirée est rose, et 0 sinon, vérifie $Y \sim \mathcal{B}\left(\frac{R}{N}\right)$.

Plus généralement, lorsqu'on effectue une expérience aléatoire de Bernoulli, c'est-à-dire dont le résultat est soit un succès, soit un échec, et qu'on note X la v.a. valant 1 en cas de succès et 0 en cas d'échec, alors $X \sim \mathcal{B}(p)$ où p est la probabilité du succès.

4 Variables aléatoires binomiales

Définition 22
Soient $n \in \mathbb{N}^\star$ et $p \in [0,1]$. On note $q = 1-p$.
Une v.a.r. X est une v.a. *binomiale de paramètres n* et p ssi

$$X(\Omega) \subset [\![0,n]\!] \quad \text{et} \quad \forall k \in [\![0,n]\!],\ \mathrm{P}(X=k) = \binom{n}{k} p^k q^{n-k}.$$

On note cela $X \sim \mathcal{B}(n,p)$.

Remarque 23 $X \sim \mathcal{B}(1,p) \iff X \sim \mathcal{B}(p)$.

Exercice 24 Montrer que si $X \sim \mathcal{B}(n,p)$, alors $Y = n-X \sim \mathcal{B}(n,1-p)$.

Proposition 25 *Soient $n \in \mathbb{N}^\star$ et $p \in [0,1]$. Lors d'une succession de n expériences aléatoires "indépendantes" de Bernoulli, de même probabilité de succès p, le nombre total X de succès vérifie $X \sim \mathcal{B}(n,p)$.*

Démonstration: Remarquons que lorsqu'on parle d'"indépendance" pour des expériences aléatoires, cela signifie une indépendance "physique", *i.e.* les résultats des différentes expériences n'influent pas sur les résultats des autres pour des raisons physiques, comme les lancers de plusieurs pièces, ou même plusieurs lancers d'une même pièce. On modélise alors cela par de l'indépendance "mathématique".

Soit $p \in [0,1]$ fixé. On démontre alors le résultat par récurrence sur $n \geq 1$, en posant

> $\mathcal{A}_n$: *Lors d'une succession de n expériences aléatoires "indépendantes" de Bernoulli, de probabilité de succès p, le nombre total X de succès vérifie $X \sim \mathcal{B}(n,p)$.*

Initialisation : Pour $n=1$, le nombre de succès X vérifie $X \sim \mathcal{B}(p)$, d'après ce qu'on a vu sur les lois de Bernoulli, donc $X \sim \mathcal{B}(1,p)$.
Hérédité : Supposons $\mathcal{A}_n$ vérifiée pour un certain $n \geq 1$. On considère alors une succession de $n+1$ expériences aléatoires "indépendantes" de Bernoulli, de probabilité de succès p et on note X_n le nombre de succès obtenus au cours des n premières épreuves et X_{n+1} le nombre total de succès. Par hypothèse de récurrence, $X_n \sim \mathcal{B}(n,p)$. En notant S_{n+1} l'évènement "la $(n+1)$-ième expérience est un succès", on a alors, pour $k \in [\![0,n+1]\!]$,

$$(X_{n+1}=k) = (X_n = k) \cap \overline{S_{n+1}} \sqcup (X_n = k-1) \cap S_{n+1}.$$

En utilisant l'additivité disjointe et l'indépendance des évènements $(X_n = i)$ et S_{n+1} découlant de l'"indépendance" des expériences,

$$\begin{aligned} \mathrm{P}(X_{n+1}=k) &= \mathrm{P}\big((X_n=k)\cap\overline{S_{n+1}}\big) + \mathrm{P}\left((X_n=k-1)\cap S_{n+1}\right) \\ &= \mathrm{P}(X_n=k)\,\mathrm{P}\left(\overline{S_{n+1}}\right) + \mathrm{P}(X_n=k-1)\,\mathrm{P}(S_{n+1}). \end{aligned}$$

En utilisant le fait que $X_n \sim \mathcal{B}(n,p)$, et la convention habituelle de nullité des binomiaux $\binom{n}{j}$ lorsque $j \notin [\![0,n]\!]$, on obtient

$$\mathrm{P}(X_{n+1}=k) = \left(\binom{n}{k} + \binom{n}{k-1}\right) p^k q^{(n+1)-k} = \binom{n+1}{k} p^k q^{(n+1)-k},$$

donc $\mathcal{A}_{n+1}$ est vérifiée. □

Exemples 26

1. On lance n fois une pièce qui a une probabilité p de tomber sur face. Le nombre X de “faces” obtenus vérifie $X \sim \mathcal{B}(n,p)$.
2. On tire n fois avec remise une boule dans une urne contenant N boules dont R boules roses. Le nombre X de tirages ayant donné une boule rose vérifie $X \sim \mathcal{B}\left(n, \frac{R}{N}\right)$.

III Couples de v.a.

Définition 27

Pour $X : \Omega \to E$ et $Y : \Omega \to F$ deux v.a. sur Ω, on définit la v.a.

$$(X,Y) : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & E \times F \\ \omega & \longmapsto & (X,Y)(\omega) = (X(\omega), Y(\omega)). \end{array} \right.$$

Une telle v.a. est appelée un *couple de variables aléatoires*.
La loi $\mathrm{P}_{(X,Y)}$ de (X,Y) est alors appelée la *loi conjointe* de X et Y.
Les lois P_X et P_Y de X et Y sont appelées les *lois marginales* du couple (X,Y).

En notant $X(\Omega) = \{x_1, \ldots, x_k\}$ et $Y(\Omega) = \{y_1, \ldots, y_l\}$, on peut représenter toutes ces lois dans un tableau du type suivant

X \ Y	y_1	$\cdots$	y_j	$\cdots$	y_l	P_X
x_1						
$\vdots$						
x_i			$\mathrm{P}\left(X = x_i, Y = y_j\right)$			$\mathrm{P}(X = x_i)$
$\vdots$						
x_k						
P_Y			$\mathrm{P}\left(Y = y_j\right)$			1

Exemple 28 Voici un exemple de loi d’un de couple (X,Y) de v.a. tel que $X(\Omega) = \{A, B\}$ et $Y(\Omega) = \{0, 1\}$:

X \ Y	0	1	P_X
A	$\frac{1}{3}$	0	$\frac{1}{3}$
B	$\frac{5}{12}$	$\frac{1}{4}$	$\frac{2}{3}$
P_Y	$\frac{3}{4}$	$\frac{1}{4}$	1

On remarque que la loi de X est simplement obtenue en faisant la somme de chaque ligne, pour les colonnes y_1 à y_l, et celle de Y en faisant la somme de chaque colonne, pour les lignes x_1 à x_k. C’est un fait général, comme le montre la proposition suivante.

Proposition 29 *La loi d'un couple (X,Y) détermine les lois marginales par les formules*

$$\forall x \in X(\Omega), \quad \mathrm{P}(X=x) = \sum_{y \in Y(\Omega)} \mathrm{P}(X=x, Y=y),$$

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

Démonstration: Évidente par additivité disjointe, puisque, pour $x \in X(\Omega)$, $(X=x) = \bigsqcup_{y \in Y(\Omega)} (X=x) \cap (Y=y)$, et de même pour la loi de Y. □

Remarque 30 En revanche, les lois marginales **ne déterminent pas** la loi conjointe, comme le montrent les trois couples suivants, dont les loi marginales sont identiques, mais pas les lois conjointes :

X \ Y	1	2	3	P_X
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
P_Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

X \ Y	1	2	3	P_X
0	$\frac{1}{3}$	0	0	$\frac{1}{3}$
1	0	$\frac{1}{3}$	0	$\frac{1}{3}$
2	0	0	$\frac{1}{3}$	$\frac{1}{3}$
P_Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

X \ Y	1	2	3	P_X
0	$\frac{1}{9}$	$\frac{2}{9}$	0	$\frac{1}{3}$
1	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$
2	$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{1}{3}$
P_Y	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Remarque 31 Dans le premier des tableaux précédents, il ne semble pas y avoir de "dépendance" entre X et Y, alors que dans le deuxième, la donnée de X détermine entièrement celle de Y. Dans le troisième cas, la donnée de X amène des informations sur Y, mais pas sa valeur exacte. Ces remarques nous mènent à la définition de l'indépendance de variables aléatoires, que nous verrons dans la prochaine section.

Enfin, pour $j \in [\![1,l]\!]$, la table permet de retrouver facilement la loi de X sachant $Y = y_j$, en divisant chaque case de la colonne "y_j" par la case marginale correspondante. De la même façon, on trouve la loi de Y sachant $X = x_i$ en utilisant la ligne "x_i".

Exercice 32 On considère un couple dont la loi conjointe est donnée par le tableau suivant :

X \ Y	A	B	C
S	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{3}$
E	$\frac{1}{8}$	0	$\frac{1}{6}$

Déterminer les lois marginales et les lois de X sachant $Y=A$, de X sachant $Y=B$, de X sachant $Y=C$, de Y sachant $X=S$ et de Y sachant $X=N$.

IV V.a. indépendantes

1 Indépendance de deux v.a.

Définition 33 Deux v.a. X et Y sur un même univers **fini** sont dites *indépendantes* ssi pour tous $A \subset X(\Omega)$ et $B \subset Y(\Omega)$, les évènements $(X \in A)$ et $(Y \in B)$ sont indépendants, ce qui se réécrit

$$\forall A \subset X(\Omega),\ \forall B \subset Y(\Omega), \quad \mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\,\mathrm{P}(Y \in B)$$

ou encore

$$\forall A \subset X(\Omega),\ \forall B \subset Y(\Omega), \quad \mathrm{P}_{(X,Y)}(A \times B) = \mathrm{P}_X(A)\mathrm{P}_Y(B).$$

Cette définition est très pratique lorsque cette indépendance provient de la modélisation du problème ou de l'énoncé. Cependant, si l'on doit démontrer mathématiquement cette indépendance, il est plus pratique d'utiliser la caractérisation suivante :

Proposition 34 *L'indépendance de X et Y équivaut à*

$$\forall (x,y) \in X(\Omega) \times Y(\Omega), \quad \mathrm{P}(X = x, Y = y) = \mathrm{P}(X = x)\,\mathrm{P}(Y = y)\,.$$

Démonstration: Le sens direct est immédiat, en appliquant la définition de l'indépendance, pour chaque couple $(x,y) \in X(\Omega) \times Y(\Omega)$, aux ensembles $A = \{x\}$ et $B = \{y\}$, ce qui donne bien

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

Pour montrer la réciproque, on fait un calcul. Supposons que la propriété ci-dessus soit vérifiée. Soient $A \subset X(\Omega)$ et $B \subset Y(\Omega)$. On a, par réunion disjointe

$$\begin{aligned}
\mathrm{P}(X \in A, Y \in B) &= \mathrm{P}((X,Y) \in A \times B) = \sum_{(x,y) \in A \times B} \mathrm{P}((X,Y) = (x,y)) \\
&= \sum_{(x,y) \in A \times B} \mathrm{P}(X = x, Y = y) = \sum_{(x,y) \in A \times B} \mathrm{P}(X = x)\,\mathrm{P}(Y = y) \\
&= \sum_{x \in A} \sum_{y \in B} \mathrm{P}(X = x)\,\mathrm{P}(Y = y) = \left(\sum_{x \in A} \mathrm{P}(X = x)\right)\left(\sum_{y \in B} \mathrm{P}(Y = y)\right) \\
&= \mathrm{P}(X \in A)\,\mathrm{P}(Y \in B)\,.
\end{aligned}$$

□

On voit immédiatement que

Proposition 35 *Toute v.a. constante est indépendante de toute v.a. définie sur le même espace probabilisé.*

En pratique, à l'aide d'un tableau du type précédent représentant la loi conjointe et les lois marginales, on vérifie si la valeur dans la case correspondant à $X = x$ et $Y = y$ est bien le produit des nombres en bout de ligne et en bout de colonne.

Exemple 36 Parmi les trois couples de la remarque 30, seul le premier donne des v.a. X et Y indépendantes.

On voit ainsi immédiatement que

Proposition 37 *Pour deux v.a.* ***indépendantes****, les lois marginales déterminent la loi conjointe.*

On en déduit facilement que

Proposition 38 *Un couple (X,Y) de v.a. uniformes est uniforme si et seulement si X et Y sont indépendantes.*

On a aussi

Proposition 39 *La somme Z de deux v.a. indépendantes $X \sim \mathcal{B}(m,p)$ et $Y \sim \mathcal{B}(n,p)$ vérifie $Z \sim \mathcal{B}(m+n,p)$.*

Démonstration: On note $Z = X+Y$ et $q = 1-p$. Il est clair que $Z(\Omega) \subset [\![0,m+n]\!]$. Pour $k \in [\![0,m+n]\!]$,

$$(Z=k) = \bigsqcup_{i=0}^{k} (X=i) \cap (Y=k-i)$$

Remarquons que certains de ces évènements peuvent être impossibles, par exemple dans le cas où $k > m$, mais l'égalité est toujours vérifiée. En utilisant le prolongement usuel des coefficients binomiaux $\binom{a}{b}$ (coefficient nul pour $b \in \mathbb{Z} \setminus [\![0,a]\!]$), on a alors, par additivité disjointe et indépendance de X et Y, lorsque p et q sont non nuls :

$$\begin{aligned}
\mathrm{P}(Z=k) &= \sum_{i=0}^{k} \mathrm{P}(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} \mathrm{P}(X=i)\,\mathrm{P}(Y=k-i) \\
&= \sum_{i=0}^{k} \binom{m}{i} p^i q^{m-i} \binom{n}{k-i} p^{k-i} q^{n-(k-i)} \\
&= \left(\sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i} \right) p^k q^{(m+n)-k}.
\end{aligned}$$

et cette formule fonctionne encore lorsque p ou q est nul (le vérifier !).

On a vu précédemment une démonstration combinatoire du fait que $\sum_{i=0}^{k} \binom{m}{i}\binom{n}{k-i} = \binom{m+n}{k}$, qu'on rappelle rapidement. On prend deux ensembles disjoints E et F de cardinaux respectifs m et n. On a la partition

$$\mathcal{P}_k(E \sqcup F) = \bigsqcup_{i=0}^{k} \{A \in \mathcal{P}_k(E \sqcup F) \mid \mathrm{Card}\,(A \cap E) = i\}$$

et choisir un ensemble $A \in \mathcal{P}_k(E \sqcup F)$ tel que $\mathrm{Card}\,(A \cap E) = i$, c'est choisir un élément de $\mathcal{P}_i(E)$ et un élément de $\mathcal{P}_{k-i}(F)$, soit $\binom{m}{i}\binom{n}{k-i}$ choix. En prenant les cardinaux de l'égalité ensembliste ci-dessus, on obtient par réunion disjointe l'égalité cherchée.

On a ainsi montré que $\mathrm{P}(Z=k) = \binom{m+n}{k} p^k q^{(m+n)-k}$. Cela étant valable pour tout $k \in [\![0,m+n]\!]$, $Z \sim \mathcal{B}(m+n,p)$. □

Exemple 40 On lance deux fois un dé pipé de telle manière que le résultat 6 apparaisse avec une probabilité $\frac{1}{2}$ et les autres avec une probabilité $\frac{1}{10}$. Quelle est la probabilité d'obtenir le résultat du premier supérieur ou égal à 3 et celui du second inférieur ou égal à 2 ?

En notant X la v.a. modélisant le résultat du premier lancer et Y celui du second, on a, par indépendance physique des lancers qu'on modélise par l'indépendance mathématique de X et Y,

$$\begin{aligned} \mathrm{P}(X \geq 3, Y \leq 2) &= \mathrm{P}(X \geq 3)\mathrm{P}(Y \leq 2) \\ &= (\mathrm{P}(X=3)+\mathrm{P}(X=4)+\mathrm{P}(X=5)+\mathrm{P}(X=6))(\mathrm{P}(Y=1)+\mathrm{P}(Y=2)) \\ &= \frac{8}{10}\cdot\frac{2}{10} = \frac{4}{25}. \end{aligned}$$

Enfin on a le résultat très utile suivant concernant les v.a. composées :

Proposition 41

Soient deux v.a. indépendantes $X : \Omega \to E$, $Y : \Omega \to F$ et deux applications $f : E \to E'$, $g : F \to F'$.

Alors $f(X)$ et $g(Y)$ sont indépendantes.

Démonstration: Soient $x' \in f(X(\Omega))$ et $y' \in g(Y(\Omega))$. On a alors $(f(X) = x') = (X \in f^{-1}(\{x'\}))$ et $(g(Y) = y') = (Y \in g^{-1}(\{y'\}))$, donc

$$\begin{aligned} \mathrm{P}\left(f(X) = x', g(Y) = y'\right) &= \mathrm{P}\left(X \in f^{-1}(\{x'\}), Y \in g^{-1}(\{y'\})\right) \\ &= \mathrm{P}\left(X \in f^{-1}(\{x'\})\right)\mathrm{P}\left(Y \in g^{-1}(\{y'\})\right) \\ &= \mathrm{P}\left(f(X) = x'\right)\mathrm{P}\left(g(Y) = y'\right). \end{aligned}$$

Ainsi $f(X)$ et $g(Y)$ sont indépendantes. □

Exemple 42 Pour deux v.a.r. indépendantes X et Y, les v.a. X^2 et Y^3 sont indépendantes.

2 Indépendance mutuelle d'un nombre fini de v.a.

On généralise la définition de l'indépendance à un nombre fini de v.a. :

Définition 43 Les v.a. $X_1, X_2 \ldots X_n$ (définies sur un même espace probabilisé) sont dites *mutuellement indépendantes* si et seulement si pour tout $(A_1, A_2, \ldots, A_n) \in \prod_{i=1}^{n} \mathcal{P}(X_i(\Omega))$, les événements $(X_1 \in A_1)$, $(X_2 \in A_2) \ldots (X_n \in A_n)$ sont mutuellement indépendants.

Remarque 44 La mutuelle indépendance implique évidemment l'indépendance deux-à-deux, mais la réciproque est fausse, comme pour les évènements.

On veut maintenant, suivant le plan du cas $n = 2$, caractériser plus simplement cette mutuelle indépendance. Rappelons que, pour $n \geq 3$, la simple égalité

$$\mathrm{P}(X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n) = \mathrm{P}(X_1 \in A_1)\mathrm{P}(X_2 \in A_2)\cdots\mathrm{P}(X_n \in A_n)$$

n'implique pas du tout l'indépendance mutuelle des évènements $(X_i \in A_i)$, comme on pourra le voir en relisant avec profit le premier chapitre de probabilités.

Cependant, un résultat analogue à celui donné pour l'indépendance de deux variables aléatoires est vrai, avec une démonstration plus subtile :

Théorème 45 *Cette indépendance mutuelle équivaut à*

$$\forall(A_1,A_2,...,A_n)\in\prod_{i=1}^{n}\mathcal{P}(X_i(\Omega)),\quad \mathrm{P}(X_1\in A_1,X_2\in A_2,\ldots,X_n\in A_n)=\mathrm{P}(X_1\in A_1)\mathrm{P}(X_2\in A_2)\cdots\mathrm{P}(X_n\in A_n).$$

Démonstration: Le sens direct est évident puisque la mutuelle indépendance des évènements $(X_1\in A_1)$, $(X_2\in A_2)$... $(X_n\in A_n)$ implique l'égalité

$$\mathrm{P}(X_1\in A_1,X_2\in A_2,\ldots,X_n\in A_n)=\prod_i \mathrm{P}(X_i\in A_i).$$

Montrons la réciproque. On suppose que

$$\forall(B_1,B_2,...,B_n)\in\prod_{i=1}^{n}\mathcal{P}(X_i(\Omega)),\quad \mathrm{P}(X_1\in B_1,X_2\in B_2,\ldots,X_n\in B_n)=\mathrm{P}(X_1\in B_1)\mathrm{P}(X_2\in B_2)\cdots\mathrm{P}(X_n\in B_n)$$

et on prend $(A_1,A_2,\ldots,A_n)\in\prod_{i=1}^{n}\mathcal{P}(X_i(\Omega))$. Pour montrer l'indépendance mutuelle des évènements $(X_i\in A_i)$, on doit montrer que, pour tout choix d'indices $1\le i_1<\ldots<i_k\le n$, on a $\mathrm{P}(X_{i_1}\in A_{i_1},X_{i_2}\in A_{i_2},\ldots,X_{i_k}\in A_{i_k})=\prod_{j=1}^{k}\mathrm{P}\left(X_{i_j}\in A_{i_j}\right)$.

Soient $1\le i_1<\ldots<i_k\le n$. On définit alors les ensembles $A'_1,A'_2,\ldots,A'_n$ ainsi : pour tout $j\in[\![1,k]\!]$, $A'_{i_j}=A_{i_j}$ et pour tout $i\in[\![1,n]\!]\setminus\{i_1,\ldots,i_k\}$, $A'_i=X_i(\Omega)$. Remarquons alors que pour $i\in[\![1,n]\!]\setminus\{i_1,\ldots,i_k\}$, $(X_i\in A'_i)$ est certain et donc $\mathrm{P}(X_i\in A'_i)=1$. Ainsi,

$$\begin{aligned}\mathrm{P}(X_{i_1}\in A_{i_1},X_{i_2}\in A_{i_2},\ldots,X_{i_k}\in A_{i_k}) &= \mathrm{P}\left(X_1\in A'_1,X_2\in A'_2,\ldots,X_n\in A'_n\right)\\ &= \mathrm{P}\left(X_1\in A'_1\right)\mathrm{P}\left(X_2\in A'_2\right)\cdots\mathrm{P}\left(X_n\in A'_n\right)\\ &= \mathrm{P}(X_{i_1}\in A_{i_1})\mathrm{P}(X_{i_2}\in A_{i_2})\cdots\mathrm{P}(X_{i_k}\in A_{i_k})\end{aligned}$$

□

On en déduit alors la caractérisation très pratique

Corollaire 46 *Les v.a. X_1, X_2... X_n sont mutuellement indépendantes ssi*

$$\forall(x_1,\ldots,x_n)\in X_1(\Omega)\times\ldots\times X_n(\Omega),\quad \mathrm{P}(X_1=x_1,\ldots,X_n=x_n)=\prod_{i=1}^{n}\mathrm{P}(X_i=x_i).$$

Démonstration: Le passage de la caractérisation du théorème à celle-ci se fait exactement de la même manière que pour deux v.a., avec des notations cependant plus lourdes. □

On peut aussi généraliser la proposition 41 :

Théorème 47 ***(Lemme des coalitions)***

Soient $X_1,\ldots,X_n$ des v.a. mutuellement indépendantes et $I_1,\ldots,I_k$ des parties deux-à-deux disjointes de $[\![1,n]\!]$.

Si, pour tout $j\in[\![1,k]\!]$, Z_j est une v.a. "fonction" des X_i, $i\in I_j$, alors les v.a. $Z_1,\ldots,Z_k$ sont mutuellement indépendantes.

Démonstration: Pour écrire une démonstration, il faudrait aller plus avant dans la formalisation (nommer en particulier le nombre d'éléments et surtout les éléments des ensembles I_j), ce qui n'est pas l'optique du cours. Cependant le fond de la démonstration est le même que celui de la démonstration de la proposition 41. □

Exemple 48 Si $X_1, \ldots, X_9$ sont des v.a. réelles mutuellement indépendantes, alors le v.a. $Z_1 = X_5 + X_7$, $Z_2 = \cos(X_8)$, $Z_3 = \sin(X_1 + X_9)$ et $Z_4 = X_3$ sont mutuellement indépendantes.

On a en particulier le corollaire trivial :

Corollaire 49 *Si $X_1, \ldots, X_n$ sont mutuellement indépendantes et $1 \leq i_1 < \cdots < i_k \leq n$, alors les v.a. $X_{i_1}, \ldots, X_{i_k}$ sont mutuellement indépendantes.*

Enfin, concernant les v.a. classiques on a le résultat suivant :

Proposition 50
Soient $n \in \mathbb{N}^\star$ et $p \in [0,1]$.
La somme de n v.a. de Bernoulli mutuellement indépendantes de même paramètre p est binomiale de paramètres n et p.

Démonstration: On peut démontrer cela par récurrence exactement comme dans la preuve de la proposition 25, en se servant du théorème précédent pour justifier l'indépendance.

Mais on peut aller plus vite en utilisant le résultat 39 sur les sommes de binomiales indépendantes, ce qu'on fait ici. On fixe p et on raisonne par récurrence rapide sur n.

Pour $n = 2$, on a une somme de deux v.a. indépendantes de type $\mathcal{B}(1,p)$, qui est donc de type $\mathcal{B}(2,p)$.

Supposons qu'on connaisse le résultat pour un $n \geq 2$. On prend alors des v.a. $X_1, \ldots, X_{n+1}$ mutuellement indépendantes. En appliquant le théorème précédent, les v.a. $X_1, \ldots, X_n$ sont mutuellement indépendantes et on peut leur appliquer l'hypothèse de récurrence. Ainsi $Y = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n,p)$. On a aussi $X_{n+1} \sim \mathcal{B}(1,p)$ et toujours d'après le théorème précédent, Y et X_{n+1} sont indépendantes. En appliquant encore une fois la proposition 39,

$$\sum_{i=1}^{n+1} X_i = Y + X_{n+1} \sim \mathcal{B}(n+1,p).$$

□

V Espérance d'une v.a.r.

1 Introduction heuristique

On imagine qu'une v.a.r. X représente le gain qu'on obtient à un jeu. Elle peut prendre les valeurs $x_1, \ldots, x_N$ avec des probabilités $p_1, \ldots, p_N$. On veut calculer le gain moyen (par coup) qu'on peut en attendre si l'on joue un "grand" nombre de fois, c'est-à-dire si l'on fait tendre le nombre n de réalisations vers $+\infty$. C'est ce gain moyen qu'on appellera l'"espérance" $\mathrm{E}(X)$ de X. Par ailleurs, comme nous l'avons déjà évoqué dans le premier chapitre de probabilités, la probabilité p_j est la limite théorique de la proportion du nombre d'occurences de x_j parmi les n résultats, lorsque n tend vers $+\infty$. Pour un nombre n de réalisations fixé, notons y_i le résultat de la i-ème réalisation. La moyenne des gains est, par regroupement des résultats identiques,

$$\frac{y_1 + \cdots + y_n}{n} = \sum_{j=1}^{N} \frac{\operatorname{Card}\left\{i \in [\![1,n]\!] \mid y_i = x_j\right\}}{n} x_j,$$

ce qui donne en passant à la limite

$$\mathrm{E}(X) = \sum_{j=1}^{N} p_j x_j,$$

qui est une moyenne des valeurs prises par X, pondérées par leurs probabilités d'occurence, puisque $\sum_j p_j = 1$.

Nous allons prendre cela comme définition de l'espérance et un résultat théorique viendra l'an prochain justifier la démonstration "heuristique" ci-dessus : il s'agit de la loi des grands nombres, qui dit que, pour une suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ de même loi (admettant une espérance notée m) et mutuellement indépendantes (en un sens qui sera défini l'an prochain), on a

$$\frac{X_1 + \cdots + X_n}{n} \underset{n\to+\infty}{\longrightarrow} m$$

en un sens qu'on définira. La convergence de cette suite d'applications a en fait lieu en beaucoup de sens différents, dont la plupart ne seront pas définis en CPGE, mais le seront dans les écoles.

2 Définition

Définition 51

Soit X une v.a.r. sur une espace probabilisé **fini**.

On note $\mathrm{E}(X)$ (ou $\mathrm{E}[X]$) et on appelle *espérance de* X le nombre

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

On définit ainsi une application $\mathrm{E} : \mathbb{R}^{\Omega} \to \mathbb{R}$.

Remarque 52 Le nombre $\mathrm{E}(X)$ ne dépend en fait que de la loi de X.

Remarque 53 Pour toute partie finie F de $\mathbb{R}$ contenant $X(\Omega)$, $\mathrm{E}(X) = \sum_{x\in F} \mathrm{P}(X = x)\ x$, puisque les termes rajoutés sont nuls.

On peut même étendre cela au cas où F est infinie avec la convention sur les sommes de familles presque nulles. Certain auteurs écrivent ainsi $\mathrm{E}(X) = \sum_{x\in\mathbb{R}} \mathrm{P}(X = x)\ x$, voire tout simplement $\mathrm{E}(X) = \sum_{x} \mathrm{P}(X = x)\ x$.

Exemple 54 On joue avec le dé pipé déjà utilisé précédemment (le résultat 6 apparaît avec une probabilité $\frac{1}{2}$ et les autres avec une probabilité $\frac{1}{10}$). Si le résultat est pair, le joueur gagne 1 euro. S'il vaut 1, le joueur perd 2 euros. S'il vaut 3, le joueur ne gagne ni ne perd rien. S'il vaut 5, le joueur perd 3 euros. On se demande si le jeu est favorable au joueur, équitable ou défavorable.

Pour cela, on peut par exemple modéliser l'expérience par l'univers $\Omega = [\![1, 6]\!]$ muni de la probabilité uniquement déterminée par $\mathrm{P}(\{i\}) = \frac{1}{10}$ pour $i \neq 6$ et $\mathrm{P}(\{6\}) = \frac{1}{2}$. On appelle X la v.a. qui représente le gain et on calcule son espérance. On a $X(\Omega) = \{-3, -2, 0, 1\}$ et

$$\begin{aligned} \mathrm{E}(X) &= \mathrm{P}(X = -3)\cdot(-3) + \mathrm{P}(X = -2)\cdot(-2) + \mathrm{P}(X = 0)\cdot 0 + \mathrm{P}(X = 1)\cdot 1 \\ &= \mathrm{P}(\{5\})\cdot(-3) + \mathrm{P}(\{1\})\cdot(-2) + \mathrm{P}(\{3\})\cdot 0 + (\mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}))\cdot 1 \end{aligned}$$

$$= \frac{1}{10}\cdot(-3)+\frac{1}{10}\cdot(-2)+\left(\frac{1}{10}+\frac{1}{10}+\frac{1}{2}\right)$$
$$= \frac{1}{5}.$$

Ainsi le jeu est favorable au joueur, et il gagne en moyenne 20 centimes d'euros par coup.

Remarquons que si l'on distribue le produit avec la parenthèse dans le calcul ci dessus, on obtient

$$\mathrm{E}(X) = \mathrm{P}(\{1\})\cdot X(1)+\mathrm{P}(\{2\})\cdot X(2)+\mathrm{P}(\{3\})\cdot X(3)+\mathrm{P}(\{4\})\cdot X(4)+\mathrm{P}(\{5\})\cdot X(5)+\mathrm{P}(\{6\})\cdot X(6).$$

Cela est une formule générale, énoncée ci-après.

Proposition 55 $\mathrm{E}(X) = \sum_{\omega\in\Omega}\mathrm{P}(\{\omega\})\cdot X(\omega)$.

Démonstration: On écrit que

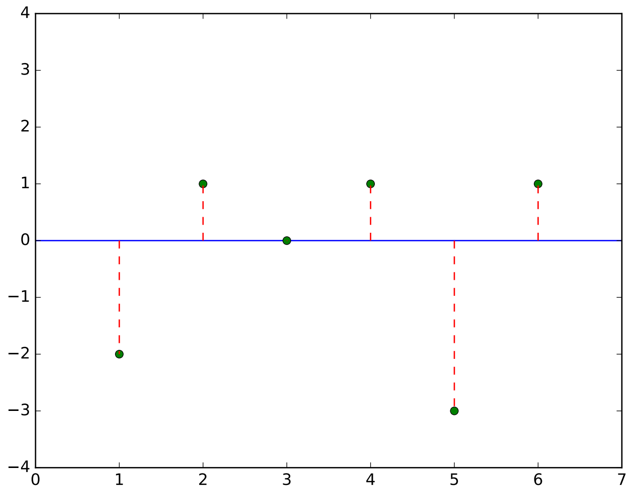
$$\Omega = \bigsqcup_{x\in X(\Omega)} (X = x)$$

et on en déduit que

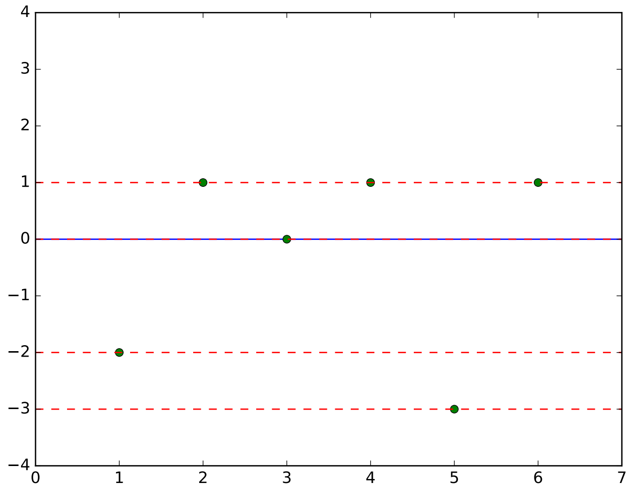
$$\sum_{\omega\in\Omega}\mathrm{P}(\{\omega\})\cdot X(\omega) = \sum_{x\in X(\Omega)}\left(\sum_{\omega\in(X=x)}\mathrm{P}(\{\omega\})\right)\cdot x = \sum_{x\in X(\Omega)}\mathrm{P}(X=x)\ x = \mathrm{E}(X).$$

□

Remarque 56 On peut considérer l'espérance comme une "intégrale finie". La formule de la proposition précédente s'apparente alors à la formule des rectangles, où la "hauteur" de chaque point est pondérée par sa "probabilité", comme on peut le voir sur la figure suivante.



A contrario, la formule de la définition de l'espérance, qui regroupe tous les évènements correspondant à un même gain, correspond à une stratification horizontale, comme on peut le voir sur la figure suivante.



Ces deux tendances se retrouvent dans l'intégration des fonctions, la première dans l'intégrale de Riemann, dont une version simplifiée est utilisée en CPGE, et la seconde dans l'intégrale de Lebesgue, qui sera enseignée dans les écoles.

Définition 57 Un v.a.r. X est dite centrée ssi $\mathrm{E}(X) = 0$.

Si X est une v.a.r. quelconque, la variable $X - \mathrm{E}(X)$ est centrée.

(On peut le voir en utilisant la linéarité de l'espérance et l'espérance d'une v.a.r. constante, qu'on voit ci-dessous.)

Exemple 58 Un jeu est équitable ssi l'espérance de gain est nulle, *i.e.* la v.a. de gain est centrée.

Comme l'espérance est une "intégrale finie", il est naturel qu'elle vérifie les propriétés suivantes.

Proposition 59

1. ***Linéarité.*** *Si X,Y sont deux v.a.r. et $\lambda,\mu \in \mathbb{R}$, alors $\mathrm{E}(\lambda X + \mu Y) = \lambda \mathrm{E}(X) + \mu \mathrm{E}(Y)$.* *En termes algébriques, $\mathrm{E} \in \mathcal{L}(\mathbb{R}^\Omega, \mathbb{R})$. On dit que E est une forme linéaire sur l'espace vectoriel réel $\mathbb{R}^\Omega$.*
2. ***Positivité.*** *Si X est une v.a.r. telle que $X(\Omega) \subset \mathbb{R}_+$, alors $\mathrm{E}(X) \geq 0$.*
3. ***Croissance.*** *Si X,Y sont deux v.a.r. telles que, pour tout $\omega \in \Omega$, $X(\omega) \leq Y(\omega)$, alors $\mathrm{E}(X) \leq \mathrm{E}(Y)$.*

Démonstration: En utilisant la formule de la proposition 55, les démonstrations sont de simples manipulations de sommes finies. □

On peut calculer les espérances des v.a. classiques vues précédemment.

Proposition 60

1. ***Constante.*** *Si $\lambda \in \mathbb{R}$, $\mathrm{E}(\lambda) = \lambda$.*
2. ***Indicatrice/Bernoulli.*** *Si $A \subset \Omega$, $\mathrm{E}(\mathbf{1}_A) = \mathrm{P}(A)$. Si $X \sim \mathcal{B}(p)$, $\mathrm{E}(X) = p$.*
3. ***Binomiale.*** *Si $X \sim \mathcal{B}(n,p)$, $\mathrm{E}(X) = np$.*

Démonstration: Les trois premiers calculs sont triviaux. Pour la binomiale, on va donner trois démonstrations, chacune étant à connaître. On rappelle qu'ici, on a $n \geq 1$.

La première est un calcul direct :

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \cdot k \\
&= \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} \cdot k \\
&= \sum_{k=1}^{n} \frac{n!k}{k!(n-k)!} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1} q^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!((n-1)-j)!} p^j q^{(n-1)-j} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{(n-1)-j} \\
&= np(p+q)^{n-1} \\
&= np \cdot 1^{n-1} \\
&= np.
\end{aligned}$$

La deuxième utilise la fonction polynomiale $f : x \longmapsto (x+q)^n = \sum_{k=0}^{n} \binom{n}{k} x^k q^{n-k}$, dérivable sur $\mathbb{R}$. En dérivant les deux expressions, on obtient, pour tout $x \in \mathbb{R}$,

$$n(x+q)^{n-1} = \sum_{k=1}^{n} \binom{n}{k} k x^{k-1} q^{n-k}.$$

En appliquant cela en $x = p$ et en multipliant par p, on obtient

$$np = np(p+q)^{n-1} = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} \cdot k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \cdot k = \mathrm{E}(X).$$

La troisième est la moins calculatoire. On considère n v.a. mutuellement indépendantes $X_1, \ldots, X_n$ telles que, pour tout $i \in [\![1, n]\!]$, $X_i \sim \mathcal{B}(p)$ (on admet que cela existe). Alors $Y = X_1 + \cdots + X_n \sim \mathcal{B}(n, p)$ d'après la section précédente. Comme l'espérance est linéaire,

$$\mathrm{E}(Y) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum_{i=1}^{n} p = np.$$

Par ailleurs, l'espérance d'une v.a. ne dépendant que de sa loi, et X et Y ayant même loi, $\mathrm{E}(X) = \mathrm{E}(Y) = np$. □

Exercice 61 On joue à pile ou face avec une pièce ayant une probabilité $\frac{3}{4}$ de tomber sur pile et on gagne deux euro à chaque fois qu'on tombe sur pile et rien sinon. Montrer que l'espérance de gain total lorsqu'on joue 4 coups est 6 euros. ⊕

Exercice 62 Pour $a,b \in \mathbb{Z}$ tels que $a \leq b$, quelle est l'espérance d'une v.a. uniforme d'image $[\![a,b]\!]$? (Réponse : $\frac{a+b}{2}$)

Lors du calcul de l'espérance d'une v.a. composée, une formule fondamentale est la suivante :

Théorème 63 ***(Formule de transfert)***
Pour $X : \Omega \to E$ et $f : X(\Omega) \to \mathbb{R}$, on a

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

Démonstration: On écrit que

$$\Omega = \bigsqcup_{x \in X(\Omega)} (X = x)$$

et on en déduit que

$$\mathrm{E}(f(X)) = \sum_{\omega \in \Omega} \mathrm{P}(\{\omega\}) \cdot f(X(\omega)) = \sum_{x \in X(\Omega)} \left(\sum_{\omega \in (X=x)} \mathrm{P}(\{\omega\}) \right) \cdot f(x) = \sum_{x \in X(\Omega)} \mathrm{P}(X = x)\ f(x).$$

□

Remarque 64 Comme précédemment, $\mathrm{E}(f(X))$ ne dépend que de f et de **la loi** de X.

Remarque 65 Si $F \supset X(\Omega)$ et $f : F \to \mathbb{R}$, alors $\mathrm{E}(f(X)) = \sum_{x \in F} \mathrm{P}(X = x)\ f(x)$.

Exercice 66 Quel est l'espérance de X^2, si $X \sim \mathcal{B}(p)$? (expliquer pourquoi la formule de transfert est ici inutile)

Exercice 67 Utiliser la formule de transfert pour calculer l'espérance de X^2, pour $X \sim \mathcal{B}(n,p)$. On pourra utiliser des méthodes analogues à celles des démonstrations de la proposition 60 pour le calcul de la somme obtenue. $\bigoplus$

On verra une autre méthode pour le calcul de cette espérance, à l'aide de la variance, dans la prochaine section.

En cas d'indépendance, on a un résultat pour l'espérance d'un produit :

Théorème 68 *Si X,Y sont deux v.a.r. indépendantes, alors* $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$.

Démonstration: En appliquant la formule de transfert à la v.a. (couple) $Z = (X,Y)$ et la fonction

$$f : \left\{ \begin{array}{ccc} X(\Omega) \times Y(\Omega) & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & xy \end{array} \right. ,$$

on obtient, en remarquant que les éléments de $(X(\Omega) \times Y(\Omega)) \setminus (X,Y)(\Omega)$ ne sont jamais atteints par Z et en utilisant l'indépendance de X et Y,

$$\mathrm{E}(XY) \quad = \quad \mathrm{E}(f(Z))$$

$$
\begin{aligned}
&= \sum_{z\in Z(\Omega)} \mathrm{P}(Z=z)\, f(z)\\
&= \sum_{(x,y)\in(X,Y)(\Omega)} \mathrm{P}((X,Y)=(x,y))\, xy\\
&= \sum_{(x,y)\in(X,Y)(\Omega)} \mathrm{P}((X,Y)=(x,y))\, xy + \sum_{(x,y)\in(X(\Omega)\times Y(\Omega))\setminus(X,Y)(\Omega)} \mathrm{P}((X,Y)=(x,y))\, xy\\
&= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} \mathrm{P}((X,Y)=(x,y))\, xy\\
&= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} \mathrm{P}(X=x, Y=y)\, xy\\
&= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} \mathrm{P}(X=x)\,\mathrm{P}(Y=y)\, xy\\
&= \left(\sum_{x\in X(\Omega)} \mathrm{P}(X=x)\, x\right)\left(\sum_{y\in Y(\Omega)} \mathrm{P}(Y=y)\, y\right)\\
&= \mathrm{E}(X)\,\mathrm{E}(Y).
\end{aligned}
$$

□

Remarque 69 La réciproque est fausse. Considérer les trois exemples de la remarque 30. ⊕

Exercice 70 Calculer l'espérance du produit des résultats de deux dés non pipés (réponse $\frac{49}{4}$). ⊕

Remarque 71 En combinant ce résultat avec un précédent, on obtient : si $X \in E^\Omega$ et $Y \in F^\Omega$ sont deux v.a. indépendantes et $f \in \mathbb{R}^{X(\Omega)}$, $g \in \mathbb{R}^{Y(\Omega)}$, alors $\mathrm{E}(f(X)g(Y)) = \mathrm{E}(f(X))\,\mathrm{E}(g(Y))$.

VI Variance, écart-type, covariance

1 Variance et écart-type d'une v.a.r.

L'espérance permet de "localiser" une v.a.r. Par exemple, dans le cas d'un jeu avec gain, la position de cette espérance par rapport à 0 a permis de déterminer si un jeu est favorable, équitable ou défavorable au joueur.

La variance (ou l'écart-type) permet de mesurer la "dispersion moyenne" d'une v.a.r. X autour de son espérance. Cette mesure est faite en utilisant les carrés des distances des valeurs à l'espérance :

Définition 72

Soit X une v.a.r.

On appelle *variance* de X, la quantité $\mathrm{V}(X) = \mathrm{E}\left[(X-\mathrm{E}(X))^2\right]$.

On appelle *écart-type* de X la racine carrée de la variance, *i.e.* la quantité $\sigma(X) = \sqrt{\mathrm{V}(X)}$.

Remarque 73 Comme l'espérance, la variance de X ne dépend que de la loi de X, par la formule de transfert.

Exemple 74 Soit X la v.a. résultat d'un lancer de dé non pipé, X_1, X_2 deux v.a. mutuellement indépendantes de même loi que X correspondant à 2 lancers successifs du dé et $Y = \frac{1}{2}(X_1 + X_2)$ la moyenne des résultats de ces deux lancers. On voit facilement que $\mathrm{E}(Y) = \mathrm{E}(X) = \frac{7}{2}$. Calculer les variances et écarts-types de X et Y. $\oplus$

On préfère souvent manipuler des v.a. "standardisées".

Définition 75
Une v.a.r. X est dite *réduite* ssi $\sigma(X) = 1$, *i.e.* $\mathrm{V}(X) = 1$.
Elle est dite *centrée réduite* ssi $\mathrm{E}(X) = 0$ et $\sigma(X) = 1$.

On calcule le plus souvent la variance par la formule suivante :

Proposition 76 $\mathrm{V}(X) = \mathrm{E}\left(X^2\right) - (\mathrm{E}(X))^2$.

Démonstration: En exercice, très instructif. $\oplus$ □

Exercice 77 Calculer la variance d'une v.a. de Bernoulli. $\oplus$

À cause du carré dans sa formule, la variance **n'est pas linéaire**, cependant, on a la formule suivante :

Proposition 78 *Pour une v.a.r. X et $a, b \in \mathbb{R}$,*

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

Démonstration: En exercice. □

Remarque 79 En particulier, $V(-X) = V(X)$.

Cette formule a pour conséquences :

Proposition 80

1. *Pour une v.a.r. X,*
$$\mathrm{V}(X - \mathrm{E}(X)) = \mathrm{V}(X).$$
2. *Si de plus* $\mathrm{V}(X) > 0$*, alors la v.a.r.*
$$\frac{X - \mathrm{E}(X)}{\sigma(X)}$$
est centrée réduite.

Exercice 81 Centrer-réduire une v.a. binomiale par un calcul direct. C'est très instructif, mais bien noter aussi qu'il existe une méthode beaucoup plus rapide vue plus bas.

2 Covariance

Commençons par un avertissement :

Remarque 82 **La variance d'une somme n'est en général pas la somme des variances**, Cependant, on va voir que cela est vrai sous une hypothèse de “décorrélation”, qui utilise la notion suivante de covariance.

Définition 83 Pour deux v.a.r., X et Y, la *covariance* de X et Y est

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))\right].$$

Lorsque $\mathrm{Cov}(X,Y)=0$, les v.a. X et Y sont dites *décorrélées*.

Remarque 84 On a $\mathrm{Cov}(X,X) = \mathrm{V}(X)$.

Remarque 85 Il est évident que $\mathrm{Cov}(X-\mathrm{E}(X), Y-\mathrm{E}(Y)) = \mathrm{Cov}(X,Y)$.

Comme pour la variance, il existe une formule très pratique :

Proposition 86 *Pour toutes v.a.r., X et Y,*

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y).$$

Démonstration: En exercice. □

On en déduit immédiatement la

Proposition 87 *Deux v.a.r. indépendantes sont décorrélées.*

Remarque 88 Attention, la réciproque est fausse, puisque la relation $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ **ne suffit pas** pour avoir l'indépendance de X et Y, comme vu précédemment.

On peut alors exprimer la variance d'une somme de variables aléatoires réelles :

Proposition 89 *Pour toutes v.a.r. $X_1, X_2, \ldots, X_n$,*

$$\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{V}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cov}\left(X_i, X_j\right).$$

En particulier, pour deux v.a.r. X et Y,

$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) + 2\mathrm{Cov}(X,Y).$$

Démonstration: Soient $X_1, \ldots, X_n$ des v.a.r.

Par la proposition 80 et la remarque 85, centrer les v.a.r. X_i ne change ni leur variance, ni la variance de leur somme, ni leurs covariances. On peut donc supposer sans perte de généralité que les v.a.r. X_i sont centrées. On a alors, puisque la somme est aussi centrée :

$$\mathrm{V}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{E}\left(\left(\sum_{i=1}^{n} X_i\right)^2\right)$$

$$\begin{aligned}
&= \mathrm{E}\left(\sum_{i=1}^{n} X_i^2 + 2 \sum_{1\le i<j\le n} X_iX_j\right) \\
&= \sum_{i=1}^{n} \mathrm{E}\left(X_i^2\right) + 2 \sum_{1\le i<j\le n} \mathrm{E}\left(X_iX_j\right) \\
&= \sum_{i=1}^{n} \mathrm{V}\left(X_i\right) + 2 \sum_{1\le i<j\le n} \mathrm{Cov}\left(X_i,X_j\right).
\end{aligned}$$

□

Cette proposition a comme conséquence immédiate un résultat très utile :

Corollaire 90 *La variance d'une somme de variables aléatoires* ***deux à deux décorrélées*** *est la somme de leurs variances.*

C'est en particulier le cas lorsque les v.a. sont deux-à-deux indépendantes et a fortiori lorsqu'elles sont mutuellement indépendantes.

Cela permet d'obtenir **rapidement** la variance d'une v.a. binomiale.

Proposition 91

On note comme d'habitude $q = 1 - p$.

Si $X \sim \mathcal{B}(n,p)$, alors $\mathrm{V}(X) = npq$.

On retrouve que la variance d'une v.a. de Bernoulli de paramètre p est pq.

Démonstration: Soit Y une somme de n v.a. Y_i de Bernoulli de paramètre p mutuellement indépendantes. Alors, par le corollaire précédent, $\mathrm{V}(Y) = \sum_{i=1}^{n} \mathrm{V}(Y_i) = npq$.

De plus, on sait que $Y \sim \mathcal{B}(n,p)$, donc X et Y ont même loi, donc même variance, donc $\mathrm{V}(X) = npq$. □

Exercice 92 Pour $X \sim \mathcal{B}(n,p)$, montrer que $\mathrm{E}\left(X^2\right) = np(q+np)$.

VII Inégalités

1 Inégalité de Markov

Proposition 93 *Pour une v.a.r. X telle que $X(\Omega) \subset \mathbb{R}_+$,*

$$\forall c > 0, \quad \mathrm{P}(X \ge c) \le \frac{\mathrm{E}(X)}{c}.$$

Démonstration: On a

$$\mathrm{E}(X) = \sum_{x\in X(\Omega)} \mathrm{P}(X=x)x \quad = \sum_{x\in X(\Omega)\cap]-\infty,c[} \mathrm{P}(X=x)x \quad + \sum_{x\in X(\Omega)\cap[c,+\infty[} \mathrm{P}(X=x)x.$$

Comme $X(\Omega) \subset \mathbb{R}_+$,

$$\sum_{x\in X(\Omega)\cap]-\infty,c[} \mathrm{P}(X=x)x \quad = \sum_{x\in X(\Omega)\cap[0,c[} \mathrm{P}(X=x)\ x \quad \ge 0$$

par positivité des probabilités $\mathrm{P}(X=x)$, donc

$$\mathrm{E}(X) \quad \geq \sum_{x\in X(\Omega)\cap[c,+\infty[} \mathrm{P}(X=x)x \quad \geq \left(\sum_{x\in X(\Omega)\cap[c,+\infty[} \mathrm{P}(X=x)\right)c \quad = \mathrm{P}(X\geq c)c,$$

une nouvelle fois par positivité des probabilités $\mathrm{P}(X=x)$. Il suffit alors de diviser l'inégalité par $c>0$. □

Remarque 94 Cette inégalité nous permettra de démontrer l'inégalité de Bienaymé-Tchebychev. Elle servira aussi, dans le cas des univers infinis, à contrôler la manière dont $\mathrm{P}(X\geq c)$ tend vers 0 lorsque c tend vers $+\infty$. En pratique, si on veut être sûr que $\mathrm{P}(X\geq c)$ est inférieure à un certain seuil et si on connaît $\mathrm{E}(X)$, on choisit alors c suffisamment grand pour avoir cette majoration.

Remarque 95 Une autre formulation est la suivante : si $X:\Omega\to E$ est une v.a. et $f:X(\Omega)\to\mathbb{R}_+$, alors

$$\forall c>0,\quad \mathrm{P}(f(X)\geq c)\leq\frac{\mathrm{E}(f(X))}{c}.$$

Remarque 96 Comme on ne fait aucune hypothèse sur la loi de X, ou presque, l'inégalité de Markov est un outil universel mais peu précis, surtout dans le cas d'un univers fini, comme le montre l'exemple suivant.

Exemple 97 On tire 100 fois à pile ou face avec une pièce pipée dont le probabilité de tomber sur pile est $\frac{2}{5}$. On gagne alors le carré du nombre de fois où la pièce est tombée sur pile, en euros. On cherche à l'aide de l'inégalité de Markov un N "raisonnable" tel qu'on soit "sûr à 95%" que le gain soit inférieur ou égal à N.

Soit X la v.a. représentant le nombre de lancers donnant pile. Le gain est alors $Y=X^2$. Comme on l'a déjà vu précédemment, $X\sim\mathcal{B}\left(100,\frac{2}{5}\right)$. L'espérance de Y, calculée dans un exercice précédent, est $100\cdot\frac{2}{5}\cdot(\frac{3}{5}+100\cdot\frac{2}{5})=1624$. En prenant $c=\frac{1624}{0{,}05}=32480$, on obtient, par l'inégalité de Markov, $\mathrm{P}(Y\geq c)\leq\frac{\mathrm{E}(Y)}{c}=0,05$. Le nombre $N=c-1=32479$ convient, puisqu'en considérant l'évènement contraire,

$$\mathrm{P}(Y\leq 32479)\geq 0,95.$$

Cependant, il est immédiat que le nombre $N=100^2=10000$ fonctionne aussi, donc l'inégalité de Markov ne nous a été ici d'aucune utilité !

2 Inégalité de Bienaymé-Tchebychev

Théorème 98 *Pour une v.a.r. X et $\varepsilon>0$,*

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

Démonstration: Il suffit d'appliquer l'inégalité de Markov à la v.a. $Y=(X-E(X))^2$, avec $c=\varepsilon^2$, et de remarquer que, puisqu'on a des quantités positives, $|X-E(X)|\geq\varepsilon\iff(X-E(X))^2\geq\varepsilon^2$. □

Remarque 99 Cette inégalité est classiquement énoncée avec le symbole ε, ce qui est trompeur car elle n'est pas utilisée en pratique pour de “petites” valeurs de ε.

Exercice 100 On dispose de n boîtes de conserves, dont exactement deux sont avariées. On les ouvre successivement toutes. Soit X la v.a égale au rang d'ouverture de la première boîte avariée.

1. Pour $k \in [\![0, n]\!]$, montrer que

$$\mathrm{P}(X > k) = \frac{(n-k)(n-k-1)}{n(n-1)}$$

   et en déduire la loi de X.

2. Pour $\varepsilon > 0$, montrer que

$$\mathrm{P}\left(\left|X - \frac{n+1}{3}\right| \geq \varepsilon\right) \leq \frac{(n+1)(n-2)}{18\varepsilon^2}.$$