Variables aléatoires finies

L'idée : plutôt que de modéliser une expérience aléatoire par un espace probabilisé trop grossier, on *postule* qu'un espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ le modélise finement et on considère une application $X : \Omega \to E$ où E correspond aux évéments observables ou qui nous intéressent.

Exemple 1 On lance un dé de 6 non pipé. Modéliser cette expérience aléatoire en posant $\Omega = \{1, 2, \ldots, 6\}$ oblige à identifier des situations très différentes, et est abusif car un résultat de l'expérience n'est pas un nombre. On peut aussi postuler l'existence d'un espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ modélisant finement cette expérience et considérer l'application

$$X : \left\{ \begin{array}{lll} \Omega & \to & \{1, 2, \ldots, 6\} \\ \omega & \mapsto & \text{le nombre lu sur la face supérieure du dé lorsque } \omega \text{ est réalisée.} \end{array} \right.$$

Exemple 2 On lance deux dés de 6 non pipés. Quelle que soit la modélisation choisie pour cette expérience aléatoire, on peut souhaiter ne s'intéresser qu'à la somme des deux valeurs obtenues (par exemple parce qu'on joue à un jeu dont le gagnant est celui qui réalise la plus grande somme). On s'intéresse alors naturellement à l'application

$$X : \left\{ \begin{array}{lll} \Omega & \to & \{2, 3, \ldots, 12\} \\ \omega & \mapsto & \text{la somme des résultats des dés lorsque } \omega \text{ est réalisée.} \end{array} \right.$$

Les valeurs prises par X **correspondent aux évéments qui nous intéressent**.
En particulier les événements $\{\omega \in \Omega, X(\omega) = i\}$, où i parcourt les valeurs prises par X :

- auraient pu former l'ensemble Ω si on avait voulu modéliser très grossièrement (mais beurk !) ;
- forment un système complet d'événements (est-ce qu'on sait encore ce que c'est ?).

Dans tout le chapitre, $(\Omega, \mathcal{P}(\Omega), P)$ désigne un espace probabilisé, et E désigne un ensemble.

I Variables aléatoires

I.1 Définition et autres exemples

Définition 1 : Variable aléatoire.

1. On appelle variable aléatoire (si grosse paresse : v.a.) une application $X : \Omega \to E$.
2. Si $E \subset \mathbb{R}$, on dit que X est une variable aléatoire réelle, dans toute la suite v.a.r.
3. Si $E \subset \mathbb{R}^n$ pour un entier n, on dit que X est un vecteur aléatoire, dans la suite $\overrightarrow{\text{v.a.}}$.

On a déjà vu deux exemples dans l'introduction. Deux autres :

Exemples 3

1. On lance une pièce équilibrée.
   On note X la variable aléatoire qui prend la valeur 0 si la pièce donne face et 1 si la pièce donne pile.
   $$\text{Autrement dit } X : \left\{ \begin{array}{lll} \Omega & \to & \{0, 1\} \\ \omega & \mapsto & \left\{ \begin{array}{l} 1 \text{ si } \omega \text{ réalise "face"} \\ 0 \text{ si } \omega \text{ réalise "pile".} \end{array} \right. \end{array} \right.$$
2. On considère le jeu suivant. On lance un dé non pipé. Si la résultat est 5 ou 6, on gagne 10. Sinon, on perd 6. On note Y la variable aléatoire qui donne le gain à ce jeu.
   $$\text{Ainsi } Y : \left\{ \begin{array}{lll} \Omega & \to & \{-6, 10\} \\ \omega & \mapsto & \left\{ \begin{array}{l} 10 \text{ si } \omega \text{ réalise 5 ou 6} \\ -6 \text{ sinon.} \end{array} \right. \end{array} \right.$$

I.2 Terminologie et notations

Reprenons l'exemple 2. L'espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ modélise le lancé de deux dés et

$$X : \begin{cases} \Omega & \to & \{2, 3, \dots, 12\} \\ \omega & \mapsto & \text{la somme des résultats des dés lorsque } \omega \text{ est réalisée.} \end{cases}$$

À quoi correspond l'événement « j'obtiens une somme égale à 4 » ?

$$\begin{aligned} A &= \{\omega \in \Omega, X(\omega) = 4\} \\ &= X^{-1}(\{4\}) \\ &=: (X = 4) \end{aligned}$$

Notation 1 Si $X : \Omega \to E$ est une variable aléatoire, on notera :

- $(X = x)$ l'événement $X^{-1}\big(\{x\}\big)$, pour $x \in E$;
- $(X \in A)$ l'événement $X^{-1}\big(A\big)$, pour $A \subset E$.

Si de plus X est une v.a.r. (*i. e.* si on a $E \subset \mathbb{R}$), on notera :

- $(X \leqslant x)$ l'événement $X^{-1}\big(]-\infty, x]\big)$;
- $(X < x)$ l'événement $X^{-1}\big(]-\infty, x[\big)$;
- $(X \geqslant x)$ l'événement $X^{-1}\big([x, \infty[\big)$;
- $(X > x)$ l'événement $X^{-1}\big(]x, +\infty[\big)$.

On a maintenant le vocabulaire pour écrire noir sur blanc la remarque faite dans l'introduction :

Remarque 1

La famille des $(X = x)_{x \in X(\Omega)}$ forme un système complet d'événements.

Rappel "Tout ce pase bien par image réciproque"

i. e. l'image réciproque par X préserve $\begin{cases} \cup \\ \cap \end{cases}$

$$\begin{aligned} & \coprod_{x \in X(\Omega)} \{x\} = X(\Omega) \\ \text{donc } & X^{-1}\left(\coprod_{x \in X(\Omega)} \{x\}\right) = X^{-1} \circ X(\Omega) && \text{en appliquant } X^{\leftarrow} \\ \text{ie } & \coprod_{x \in X(\Omega)} X^{\leftarrow}(\{x\}) = \{\omega \in \Omega, X(\omega) = X(\Omega)\} \\ \text{ie } & \coprod_{x \in X(\Omega)} (X = x) = \Omega \end{aligned}$$

Notation 2 Si F est un ensemble et $f : E \to F$ est une application, on pourra noter $f(X)$ la variable aléatoire $f \circ X$. En effet, ces gens (les probabilistes) sont des barbares.

Exemple 4 Reprenons la variable aléatoire de l'exemple 3-2. On a :

I.3 Loi d'une variable aléatoire réelle

Définition 2 : Loi.

Soit $X : \Omega \to E$ une variable aléatoire. On appelle <u>loi de X</u> l'application $P_X : \left\{ \begin{array}{ccc} \mathcal{P}\big(X(\Omega)\big) & \to & [0,1] \\ A & \mapsto & P(X \in A) \end{array} \right.$

Remarque 2

On identifiera fréquemment P_X avec son prolongement naturel à $\mathcal{P}(E)$, parfois plus pratique à décrire :

$$P_X : \left\{ \begin{array}{ccc} \mathcal{P}(E) & \to & [0,1] \\ A & \mapsto & P(X \in A) \end{array} \right. .$$

Exemple 5 Reprenons l'exemple 1. La loi de X est $\left\{ \begin{array}{ccc} \mathcal{P}\big(\{1,\dots,6\}\big) & \to & [0,1] \\ A & \mapsto & \dots\dots\dots \end{array} \right.$

Dans cet exemple on remarque que P_X est ....

En général :

Théorème 1 .

Soit $X : \Omega \to E$ une variable aléatoire.

Alors $P_X : \mathcal{P}(X(\Omega)) \to [0,1]$ est une probabilité.

Bien sûr, de même, $P_X : \mathcal{P}(E) \to [0,1]$ est une probabilité.

DÉMONSTRATION. 1.

$$\begin{aligned} P(X(\Omega)|X) &= P(X \in X(\Omega)) \\ &= P(X^{\leftarrow}(X(\Omega))) \\ &= P(\Omega) && \text{Il est certain que } X \in X(\Omega) \\ &= 1 \end{aligned}$$

2. Soient $A \amalg B \in \mathcal{P}(X(\Omega))$

$$\begin{aligned} P(A \amalg B|X) &= P(X \in A \amalg B) \\ &= P(X^{\leftarrow}(A \amalg B)) \\ &= P\left(X^{\leftarrow}(A) \amalg X^{\leftarrow}(B)\right) \\ &= P(X^{\leftarrow}(A)) + P(X^{\leftarrow}(B)) \\ &= P(A|X) + P(B|X) \end{aligned}$$

□

/!\ En raison du théorème précédent, on décrira systématiquement une loi à l'aide du théorème de caractérisation, c'est-à-dire en explicitant uniquement les $P_X\big(\{x\}\big)$ pour $x \in X(\Omega)$.

Exemple 6 On décrit les lois des deux variables aléatoires de l'exemple 3 comme suit :

$$\begin{cases} P(X(\{1\})) = P(X=1) = \frac{1}{2} \\ P(X(\{0\})) = P(X=0) = \frac{1}{2} \end{cases}$$

$$\begin{cases} P(Y(\{-6\})) = P(X=-6) = \frac{8}{10} = 0.8 \\ P(Y(\{10\})) = P(X=10) = \frac{2}{10} = 0.2 \end{cases}$$

II Lois usuelles

Dans toute cette section, on considère une var X (rappel : une variable aléatoire réelle est une variable aléatoire qui ne prend que des valeurs réelles).

Rappel : les événements de la forme $(X = x)$ sont les événements qui nous intéressent lors de l'étude de X. Je les appelle plus bas les événements "sélectionnés par X".

II.1 Loi uniforme finie

Elle correspond au cas où les événements sélectionnés par X sont équiprobables.

Définition 3 .

On dit que X suit la loi uniforme sur E et on note $X \sim \mathcal{U}(E)$ lorsqu'on a $P_X = P_u$ (sur E), i. e. lorsqu'on a $\forall x \in E,\ P(X = x) = \dfrac{1}{|E|}$. Si $E = \{a, a+1, \dots, b\}$ on note $X \sim \mathcal{U}(a; b)$.

Exemples 7

1. `random.randint(1, n)` $\sim \mathcal{U}(1; n)$
2. $X : \begin{cases} \Omega & \to \{1, 0\} \\ \omega & \mapsto \begin{cases} 1 & \text{si la pièce réalise face} \\ 2 & \text{si la pièce réalise pile} \end{cases} \end{cases}$
   Si la pièce est équilibrée alors $X \sim \mathcal{U}(0; 1)$
3. $X : \begin{cases} \Omega & \to \{1, \dots, 6\} \\ \omega & \mapsto \text{la valeur du dé lancé} \end{cases}$
   Si le dé n'est pas pipé alors $X \sim \mathcal{U}(1; 6)$

II.2 Loi de Bernoulli

Elle correspond au cas où il n'y a que deux événements sélectionnés par X : on parle de "succès" pour $(X = 1)$ et d'"échec" pour $(X = 0)$.

Définition 4 .

Soit $p \in [0, 1]$. On dit que X suit la loi de Bernoulli de paramètre p et on note $X \sim \mathcal{B}(p)$ lorsqu'on a $X(\Omega) \subset \{0, 1\}$ et $P(X = 1) = p$, donc $P(X = 0) = 1 - p = q$.

Exemples 8

1. $\mathcal{B}(\frac{1}{2}) \sim \mathcal{U}(0, 1)$
2. Une pièce truquée tombe sur face 3 fois sur 4. $X : \Omega \to \{0, 1\} = \omega \mapsto \begin{cases} 1 & \text{si la pièce donne face} \\ 0 & \text{si la pièce donne pile} \end{cases} \sim \mathcal{B}(\frac{3}{4})$.

3. $Y : \begin{cases} \Omega & \to \{-6, 10\} \\ \omega & \mapsto \begin{cases} 10 & \text{si le dé donne 5 ou 6} \\ -6 & \text{sinon} \end{cases} \end{cases}$ ne suit pas une $\mathcal{B}$: $\{-6, 10\} \neq \{0, 1\}$.

Mais $Y = f(X)$ avec $X \sim \mathcal{B}(\frac{1}{3})$ où $X : \begin{cases} \Omega & \to \{0, 1\} \\ \omega & \mapsto \begin{cases} 1 & \text{si le dé donne 5 ou 6} \\ 0 & \text{sinon} \end{cases} \end{cases}$ et $f : \begin{cases} \{-6, 10\} & \to \{0, 1\} \\ -6 & \mapsto 0 \\ 10 & \mapsto 1 \end{cases}$.

Théorème 2 .
Une v.a.r. suit une loi de Bernoulli si et seulement si c'est une fonction indicatrice.
Autrement dit : $\Big(\exists p \in [0, 1],\ X \sim \mathcal{B}(p)\Big) \Leftrightarrow \Big(\exists A \in \mathcal{P}(\Omega),\ X = \mathbb{1}_A\Big)$.

DÉMONSTRATION.

$$\begin{aligned} & \exists p \in [0, 1], X \sim \mathcal{B}(p) \\ \Leftrightarrow\ & \exists p \begin{cases} X(\Omega) & = \{0, 1\} \\ P(X = 1) & = p \\ P(X = 0) & = 1 - p \end{cases} \\ \Leftrightarrow\ & \exists p \begin{cases} X(\Omega) & = \{0, 1\} \\ P(X = 1) & = p \end{cases} \end{aligned}$$

Procédons par double implication

$\boxed{\Longrightarrow}$ On suppose qu'il existe $p \in [0, 1]$ tel que
$\begin{cases} X(\Omega) & = \{0, 1\} \\ P(X = 1) & = p \end{cases}$
Posons $A = (X = 1) = X^{\leftarrow}(\{1\})$

$$\begin{aligned} \mathbb{1}_A &= \begin{cases} \Omega & \to \{0, 1\} \\ \omega & \mapsto \begin{cases} 0 & \text{si } \omega \in A \\ 1 & \text{sinon} \end{cases} \end{cases} \\ &= \begin{cases} \Omega & \to \{0, 1\} \\ \omega & \mapsto \begin{cases} 1 & \text{si } X(\omega) = 1 \\ 0 & \text{si } X(\omega) = 0 \end{cases} \end{cases} = X \end{aligned}$$

$\boxed{\Longleftarrow}$ Posons $p = P(A)$

$$\begin{aligned} P(X = 1) &= P(\mathbb{1}_A = 1) \\ &= P(A) \\ &= p \end{aligned}$$

□

II.3 Loi binomiale

Elle correspond au cas où les événements sélectionnés par X correspondent à l'obtention de 0, 1, ..., ou n succès lors de la réalisation de n expériences de Bernoulli indépendantes. (Le théorème 13 que l'on verra plus bas permet de donner une traduction mathématique à cette phrase.)

Définition 5 .
Soit $n \in \mathbb{N}$ et $p \in [0,1]$. On dit que X suit la loi binomiale de paramètres n et p et on note $X \sim \mathcal{B}(n,p)$ lorsqu'on a $X(\Omega) \subset \{0,1,\ldots,n\}$ et $\forall k \in \{0,\ldots,n\},\ P(X=k) = \binom{n}{k} p^k q^{n-k}$ où bien sûr $q = 1-p$.

Exemples 9

1. On lance n fois une pièce équilibrée et on compte le nombre de fois qu'on a obtenu face. $X \sim \mathcal{B}(n, \frac{1}{2})$
2. On lance un dé n fois et on compte le nombre de 6. On a $X \sim \mathcal{B}(n, \frac{1}{6})$

II.4 Exercice classique : décrire la loi d'une v.a.r. donnée

Point méthode :

1. Déterminer $X(\Omega)$ (les valeurs prises par X) ;
2. Déterminer les $P(X=x)$ pour $x \in X(\Omega)$;
3. Conclure à l'aide du théorème de caractérisation.

On peut se rassurer en vérifiant que $\sum\limits_{x \in X(\Omega)} P(X=x) = 1$.

Exemple 10 On lance au plus n fois un dé en s'arrêtant dès qu'on obtient un 6. On note X la variable aléatoire égale à k si on a obtenu 6 au bout de k lancés, et à 0 si on n'a pas obtenu 6. Décrivons la loi de X.

1. $X(\Omega) = \{\underbrace{0}_{\text{on a pas obtenu 6}}, 1, 2, \ldots, n\}$
2. $$\begin{aligned} P(X=0) &= \left(\frac{5}{6}\right)^n \\ P(X=1) &= \frac{1}{6} \\ &\vdots \\ P(X=k) &= \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6} && \text{pour } k \neq 0 \\ &\vdots \\ P(X=n) &= \left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6} \end{aligned}$$

   D'après le théorème de caractérisation, ceci détermine la loi de X $P(X)$.

   Remarque 3

   $$\begin{aligned} & \sum_{k=0}^{n} P(X=k) = 1 \\ \Leftrightarrow\ & P(X=0) + \sum_{k=1}^{n} P(X=k) = 1 && \text{d'après Chasles} \\ \Leftrightarrow\ & \left(\frac{5}{6}\right)^n + \frac{1}{6}\frac{1-\left(\frac{5}{6}\right)^n}{1-\frac{5}{6}} = 1 \\ \Leftrightarrow\ & 1 = 1 \end{aligned}$$

   Donc ok.

III Espérance. Variance.

Dans toute cette section, X est une v.a.r., et donc $X(\Omega) \subset \mathbb{R}$.

III.1 Espérance

C'est la moyenne des valeurs prises par X, pondérées par les probabilités qu'elles soient prises.

Définition 6 : Espérance.

On appelle espérance de X le nombre $E(X) = \displaystyle\sum_{x \in X(\Omega)} P(X = x)\, x$.

Terminologie : lorsqu'on a $E(X) = 0$, on dit que X est centrée.

Exemples 11

1. X compte le nombre de "pile" après $2n$ lancés d'une pièce équillibrée.

   On a $E(X) = \sum_{k=0}^{2n} P(X = k)k$ car $X(\Omega) = \{0, \ldots, 2n\}$

   Mieux : $X \sim \mathcal{B}(2n, \frac{1}{2})$ donc $P(X = k) = \binom{2n}{k} \times \left(\frac{1}{2}\right)^{2n}$

   On a

$$\begin{aligned}
E(X) &= \frac{1}{2^{2^n}} \sum_{k=0}^{2n} k \binom{2n}{k} \\
&= \frac{1}{2^{2^n}} \sum_{k=1}^{2n} 2n \binom{2n-1}{k-1} \\
&= \frac{2n}{2^{2^n}} \sum_{k=0}^{2n-1} \binom{2n-1}{k} \\
&= \frac{2n}{2^{2^n}} 2^{2n-1} \\
&= n
\end{aligned}$$

2. X compte le nombre de "pile" après $2n + 1$ lancés d'une pièce équillibrée.

   On a $E(X) = n + \frac{1}{2}$

/!\ L'espérance peut très bien ne pas être une valeur prise par Ω

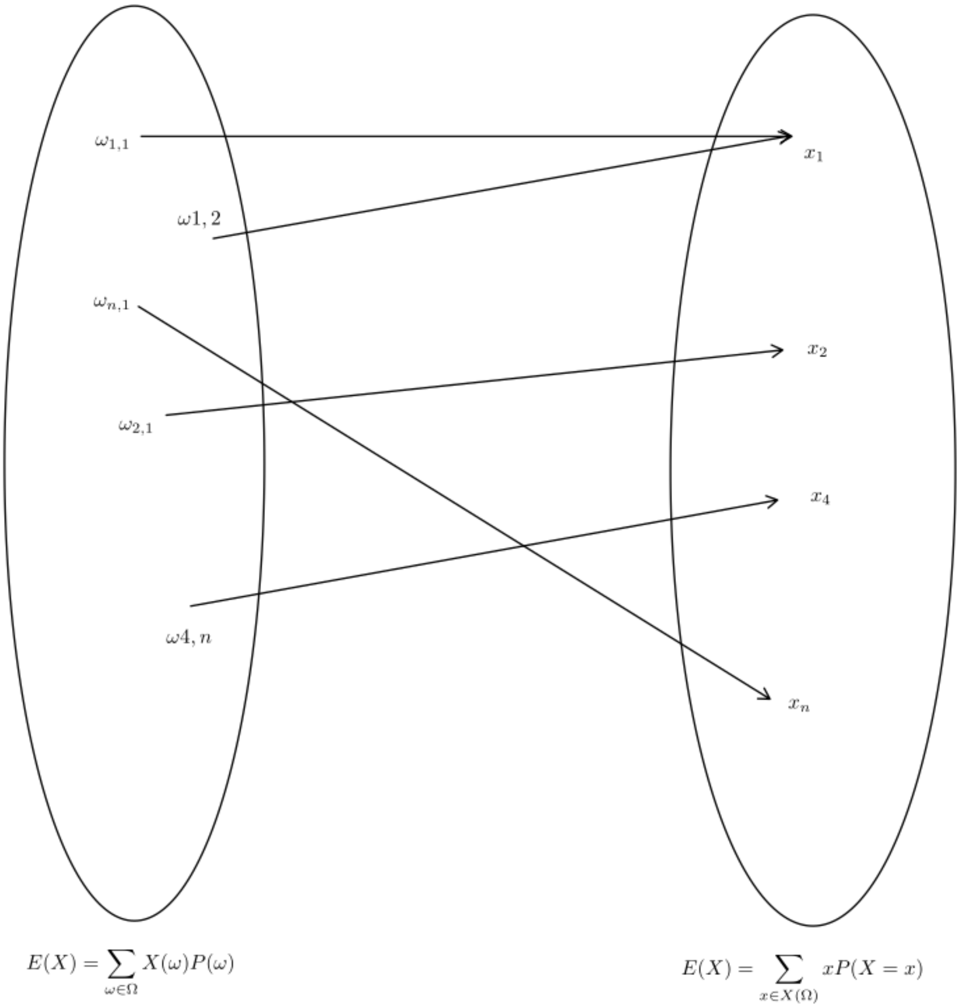
Exemple 12 Reprenons l'exemple 10.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} P(X=k)k \\
&= \sum_{k=1}^{n} P(X=k)k && \text{car pour } k=0,\ P(X=k)k=0 \\
&= \frac{1}{6}\sum_{k=1}^{n} k\left(\frac{5}{6}\right)^{k-1} \\
&= \frac{1}{6}\sum_{k=0}^{n} k\left(\frac{5}{6}\right)^{k-1} && \text{d'après Chasles} \\
&= \frac{1}{6}\sum_{k=0}^{n} kX^{k-1} && \text{en posant } X=\frac{5}{6} \\
&= \frac{1}{6}\sum_{k=0}^{n} \frac{\mathrm{d}}{\mathrm{d}x}X^{k} \\
&= \frac{1}{6}\frac{\mathrm{d}}{\mathrm{d}x}\sum_{k=0}^{n} X^{k} \\
&= \frac{1}{6}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1-X^{n+1}}{1-X}\right) \\
&= \frac{1}{6}\left(-\frac{1}{(1-x)^2}-\frac{(n+1)x^n(1-x)+x^{n+1}}{(1-x)^2}\right) \\
&= \frac{1/6}{(1-x)^2}\left(1-(n+1)x^n+-(n+1)x^{n+1}-x^{n+1}\right) \\
&= \frac{1/6}{(1-x)^2}\left(1-(n+1)x^n+nx^{n+1}\right) \\
&= 6\left(1-\underbrace{(n+1)\left(\frac{5}{6}\right)^{n}+n\left(\frac{5}{6}\right)^{n+1}}_{\sim 0}\right) && \text{car } x=\frac{5}{6}
\end{aligned}$$

Théorème 3 : Propriétés de l'espérance.

1. On a : $E(X)=\displaystyle\sum_{\omega\in\Omega} X(\omega)P(\{\omega\})$.
2. L'espérance est :
   - linéaire : $\forall X,Y:\Omega\to\mathbb{R},\ \forall\lambda,\mu\in\mathbb{R}^2,\ E(\lambda X+\mu Y)=\lambda E(X)+\mu E(Y)$;
   - positive : $\forall X:\Omega\to\mathbb{R},\ X\geqslant 0\Rightarrow E(X)\geqslant 0$;
   - croissante : $\forall X,Y:\Omega\to\mathbb{R},\ X\leqslant Y\Rightarrow E(X)\leqslant E(Y)$.
3. Si X est constante, i. e. il existe m tel que $X(\Omega)=\{m\}$, alors $E(X)=m$.

DÉMONSTRATION. 1. Pour rappel $(X=x)_{x\in X(\Omega)}$ forme un sce *i. e.* $\coprod_{x\in X(\Omega)}(X=x)=\Omega$



Note : $P(\omega) := P(\{\omega\})$ par abus

Donc

$$\begin{aligned}
\sum_{\omega \in \Omega} X(\omega)P(\{\omega\}) &= \sum_{x \in X(\Omega)} \sum_{\omega \in (X=x)} X(\omega)P(\{\omega\}) \\
&= \sum_{x \in X(\Omega)} \sum_{\omega \in (X=x)} xP(\{\omega\}) \\
&= \sum_{x \in X(\Omega)} x \sum_{\omega \in (X=x)} P(\{\omega\}) \\
&= \sum_{x \in X(\Omega)} xP\left(\coprod_{\omega \in (X=x)} \omega \right) \\
&= \sum_{x \in X(\Omega)} xP(X = x)
\end{aligned}$$

2.

$$\begin{aligned}E(\lambda X+\mu Y) &= \sum_{\omega\in\Omega}(\lambda X+\mu Y)(\omega)P(\{\omega\})\\ &= \sum_{\omega\in\Omega}(\mu X(\omega)+\mu X(\omega))P(\{\omega\})\\ &= \sum_{\omega\in\Omega}\lambda X(\omega)P(\{\omega\})+\mu Y(\omega)P(\{\omega\})\\ &= \lambda\sum_{\omega\in\Omega}X(\omega)P(\{\omega\})+\mu\sum_{\omega\in\Omega}Y(\omega)P(\{\omega\})\end{aligned}$$

3.

$$E(X)=\sum_{x\in X(r)}\underbrace{x}_{\geq 0}\underbrace{P(X=x)}_{\in[0,1]}\geq 0 \qquad \text{comme somme de positifs}$$

4.

$$\begin{aligned}Z &:= Y-X & \geq 0\\ \text{donc } E(Y-X) &\geq 0 & \text{par positivité}\\ \text{ie } E(Y)-E(X) &\geq 0 & \text{par linéarité}\\ \text{ie } E(X) &\leq E(Y)\end{aligned}$$

5.

$$\begin{aligned}X(\Omega) &= \{\mu\}\\ \text{donc } E(X) &= \sum_{x\in X(r)}xP(X=x)\\ &= \mu P(X=\mu)\\ &= \mu\cdot 1\\ &= \mu\end{aligned}$$

□

Application 1 Si $E(X)=\mu$ alors $X-E(X)$ est une variable aléatoire .
$X-E(X)$ est centrée (*i. e.* d'espérance nulle)

III.2 Variance

La variance permet de mesurer à quel point X est (ou pas) concentrée autour de son espérance.
On dit que $V(X)$ est un indicateur de dispersion, alors que $E(X)$ est un indicateur de position.

Définition 7 : Variance.
On appelle variance de X le nombre $V(X)=E\big((X-E(X))^2\big)$.

Remarque 4

On met un carré car :

— ça permet d'augmenter les gros écarts et diminuer les petits écarts (erreurs de mesure, e.g.)

— ça devient presque un produit skyler

Remarque 5

Par positivité de l'espérance, on a toujours $V(X) \geqslant 0$. Ceci permet de définir l'écart-type de X $\sigma(X) = \sqrt{V(X)}$.

Terminologie : lorsqu'on a $V(X) = 1$ (*i. e.* $\sigma(X) = 1$), on dit que X est réduite.

Exemple 13 Déterminons la variance de la variable aléatoire de l'exemple 1. On reprend $X : \begin{cases} \Omega & \to \{1, \dots, 6\} \\ \omega & \mapsto \text{la valeur du dé} \end{cases}$

Remarque 6

C'est un cas particulier de la loi uniforme

$$\begin{aligned} E(X) &= \sum_{k=1}^{6} k \underbrace{P(X = k)}_{X} \\ &= \frac{1}{6} \sum_{k=1}^{6} k \\ &= \frac{1}{6} \frac{6 \cdot 7}{2} \\ &= 3.5 \end{aligned}$$

Notons $Y = (X - E(X))^2$

$$\begin{aligned} V(X) &= E(Y) \\ &= \sum_{y \in Y(\Omega)} y P(Y = y) \\ &= \frac{1}{3} \left(\frac{1}{2}\right)^2 + \frac{1}{3} \left(\frac{3}{2}\right)^2 + \frac{1}{3} \left(\frac{5}{2}\right)^2 \\ &= \frac{1 + 9 + 25}{12} = \frac{35}{12} \end{aligned}$$

$$Y(\Omega) = \{ \underbrace{\left(\frac{1}{2}\right)^2}_{X \in \{3,4\}}, \underbrace{\left(\frac{3}{2}\right)^2}_{X \in \{2,5\}}, \underbrace{\left(\frac{5}{2}\right)^2}_{X \in \{1,6\}} \}$$

$$Y(Y = \left(\frac{1}{2}\right)^2) = \frac{1}{3}$$

$$Y(Y = \left(\frac{3}{2}\right)^2) = \frac{1}{3}$$

$$Y(Y = \left(\frac{5}{2}\right)^2) = \frac{1}{3}$$

Théorème 4 : Propriétés de la variance.

1. Formule de Kœnig : $V(X) = E(X^2) - E(X)^2$.
2. Si $a, b \in \mathbb{R}$ alors $V(aX + b) = a^2 V(X)$.
3. Si X est constante, alors $V(X) = 0$.

Démonstration. 1.

$$
\begin{aligned}
V(X) &= E((X - E(X))^2) \\
&= E(X^2 - 2E(X)X + E(X)^2) && \text{en développant} \\
&= E(X^2) - 2E(X)E(X) + E(E(X)^2) && \text{linéarité de } E \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

2.

$$
\begin{aligned}
V(aX + b) &= E((aX + b - E(aX + b))^2) \\
&= E((aX + b - (aE(X) + b)^2) \\
&= E((a(X - E(X)))^2) \\
&= E(a^2(X - E(X))^2) \\
&= a^2 V(X)
\end{aligned}
$$

3. Soit μ une constante.

$$
\begin{aligned}
V(\mu) &= E((\mu - E(\mu))^2) \\
&= E((\mu - \mu)^2) \\
&= E(0^2) \\
&= 0
\end{aligned}
$$

□

On peut retrouver la variance de l'exemple 13 avec Kœnig : $V(X) = E(X^2) - E(X)^2$

$$
\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
X^2(\Omega) &= \{1^2, 2^2, \dots, 6^2\} \\
P(X = x) &= \frac{1}{6} && \text{pour } x \in \{1^2, \dots, 6^2\}
\end{aligned}
$$

$$
\begin{aligned}
X(\Omega) &= \text{les valeurs prises par } X \\
X^2(\Omega) &= \text{les valeurs prises par } X^2
\end{aligned}
$$

$$
\begin{aligned}
E(X^2) &= \frac{1}{6}\sum_{k=1}^{6} k^2 = \frac{1}{6}\frac{6 \cdot 7 \cdot 13}{6} = \frac{7 \cdot 13}{6} \\
E(X)^2 &= \left(\frac{7}{2}\right)^2 = \frac{7 \cdot 7}{4}
\end{aligned}
$$

$$\begin{aligned} V(X) &= \frac{7\cdot 13}{6} - \frac{7\cdot 7}{4} \\ &= \frac{7\cdot 26 - 7\cdot 21}{12} \\ &= \frac{35}{12} \end{aligned}$$

Application 2 La v.a.r. $\dfrac{X-E(X)}{\sigma(X)}$ est toujours une variable aléatoire centrée réduite.

III.3 Théorème de transfert

On remarque en calculant $V(X)$ avec Kœnig qu'on a toujours $E(X^2) = \displaystyle\sum_{x\in X(\Omega)} x^2 P(X=x)$.

La formule de transfert généralise ceci.

Théorème 5 : Formule de transfert.
Soit $Z : \Omega \to E$ une variable aléatoire (non nécessairement réelle) et $f : E \to \mathbb{R}$. Alors :

$$E(f(Z)) = \sum_{z\in Z(\Omega)} f(z)P(Z=z)$$

J'ai noté Z la variable aléatoire de l'énoncé car dans cette section X est supposée être une v.a.r., mais on a en particulier pour $f : \mathbb{R} \to \mathbb{R}$: $E(f(X)) = \displaystyle\sum_{x\in X(\Omega)} f(x)P(X=x)$.

DÉMONSTRATION. $X = f \circ Z$
$(Z=z)_{z\in Z(\Omega)}$ forme un sce

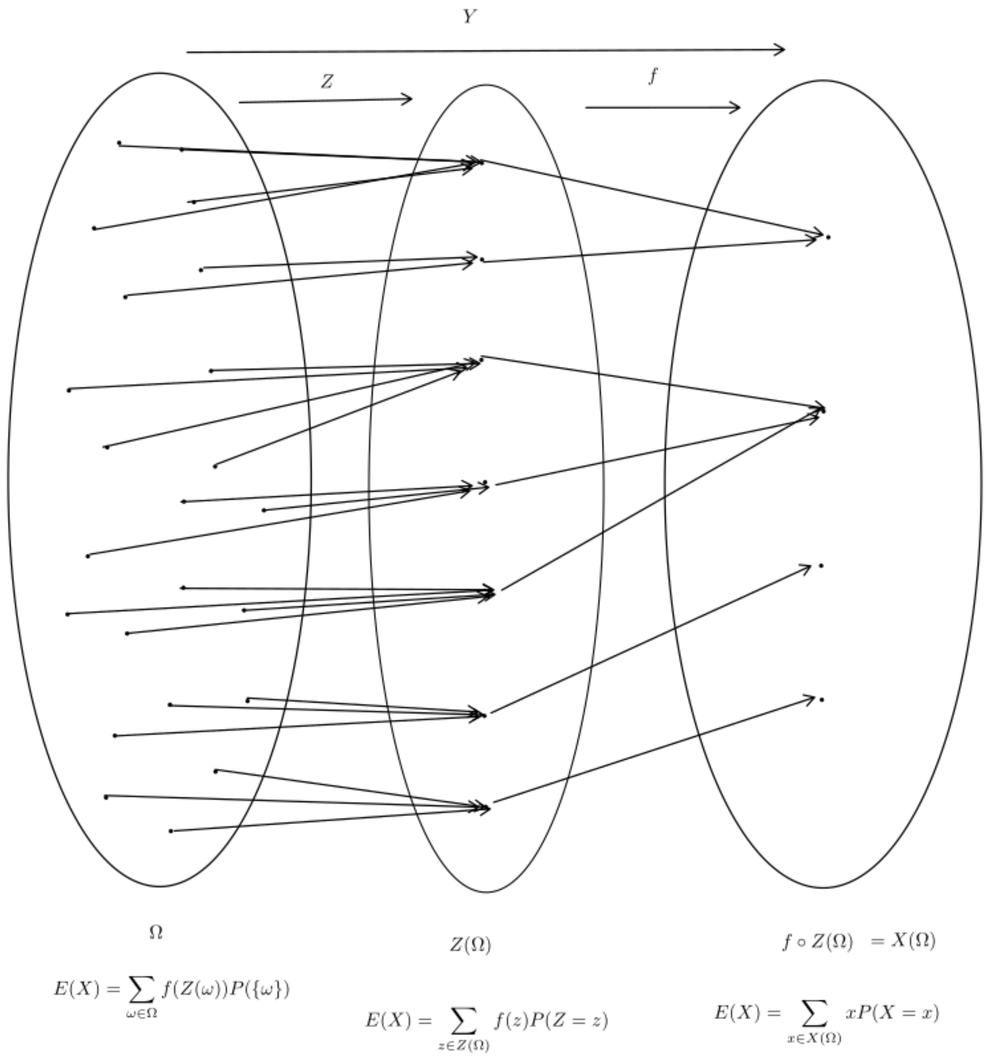
$$\coprod_{z\in Z(\Omega)} (Z=z) = \Omega$$

$$\begin{aligned} \text{donc } E(f(Z)) &= \sum_{\omega\in\Omega} f(Z(\omega))P(\{\omega\}) \\ &= \sum_{z\in Z}\sum_{\omega\in(Z=z)} f(Z(\omega))P(\{\omega\}) && \text{Chasles itéré} \\ &= \sum_{z\in Z(\Omega)}\sum_{\omega\in(Z=z)} f(z)P(\{\omega\}) && \text{par définition} \\ &= \sum_{z\in Z(\Omega)} f(z) \sum_{\omega\in(Z=z)} P(\{\omega\}) \\ &= \sum_{z\in Z(\Omega)} f(z)P(Z=z) && \text{propriété des probabilités} \end{aligned}$$

□

Application 3 On peut en particulier calculer ainsi tous les <u>moments</u> de f, c'est-à-dire les $E(X^k)$ avec $k \in \mathbb{N}$:

- $E(X^2)$ est le moment d'ordre 2 ;
- $E(X^3)$ est le moment d'ordre 3 $= \sum_{x\in X(\Omega)} x^3 P(X=x)$;
- etc.



III.4 Deux inégalités fondamentales

Les deux inégalités que l'on verra dans cette sous-section répondent à deux interrogations qui m'aident à les retenir.

Première interrogation : on tire au hasard, uniformément, un individu dans une population d'actifs.
Quelle est la probabilité qu'il gagne au moins 7 fois plus que le salaire moyen de cette population ?

$$X : \begin{cases} \Omega & \to \mathbb{R} \\ \omega & \mapsto \text{le salaire de } \omega \end{cases}$$

$$P(X \geq 7E(X)) \leq \frac{1}{7}$$

Théorème 6 : Inégalité de Markov.

Supposons X **positive**. Soit $\alpha > 0$. Alors $P(X \geqslant \alpha) \leqslant \dfrac{E(X)}{\alpha}$.

Démonstration. L'inégalité équivaut à

$$E(X) \geq \alpha P(X \geq \alpha)$$

Or

$$\begin{aligned}
E(X) &= \sum_{x\in X(\Omega)} xP(X=x) \\
&= \sum_{x\in X(\Omega)\wedge x<\alpha} xP(X=x) + \sum_{x\in X(\Omega)\wedge x\geq\alpha} xP(X=x) \\
&\geq \sum_{x\in X(\Omega)\wedge x\geq\alpha} xP(X=x) && \text{car la somme de gauche a seulement des termes positifs} \\
&\geq \alpha \sum_{x\geq\alpha} P(X=x) && = \alpha P(X\geq\alpha)
\end{aligned}$$

□

Seconde interrogation : on a vu que $P(X = E(X))$ peut être nulle, mais quelle est la probabilité que X soit **proche** de son espérance ? C'est-à-dire qu'on ait $E(X) - \varepsilon < X < E(X) + \varepsilon$?

Théorème 7 : Inégalité de Bienaymé-Tchebychev.

Soit $\varepsilon > 0$. Alors $P\big(|X - E(X)| \geqslant \varepsilon\big) \leqslant \dfrac{V(X)}{\varepsilon^2}$.

En passant par l'événement contraire on a donc $P\big(E(X) - \varepsilon < X < E(X) + \varepsilon\big) \geqslant 1 - \dfrac{V(X)}{\varepsilon^2}$: cette probabilité est donc d'autant plus grande que ε est grand (ce qui n'est pas une surprise) et aussi d'autant plus grande que $V(X)$ est petite (ce qui n'est pas une surprise non plus).

Démonstration. Notons $Y = (X - E(X))^2 \geq 0$

$$\begin{aligned}
P(\underbrace{|X - E(X)| \geq \epsilon} &= P((X-E(X))^2 \geq \epsilon^2) \\
&\leq \frac{E(Y)}{\epsilon^2} && \text{Markov avec } Y \geq 0 \\
&= \frac{E((X-E(X))^2)}{\epsilon^2} \\
&= \frac{V(X)}{\epsilon^2}
\end{aligned}$$

□

III.5 Espérance et variance des lois usuelles

Remarque 7

Si $X \sim \mathcal{U}(a;b)$, l'espérance et la variance de X ne sont pas explicitement au programme ; mais on les verra en TD.

$$X(\Omega) \subset \{0,1\} \qquad \begin{cases} P(X=0) & = 1-p = q \\ P(X=1) & = p \end{cases}$$

- $E(X) = \sum_{x\in X(\Omega)} P(X=x)x = 0q + 1p = p$
- $V(X) = E(X^2) - E(X)^2$ or $X^2 = X$ donc $V(X) = E(X) - E(X)^2 = p - p^2 = p(1-p) = pq$

Théorème 8 : Cas d'une loi de Bernoulli.

On suppose $X \sim \mathcal{B}(p)$ pour un certain $p \in [0,1]$ (on note $q = 1-p$). Alors :

$$X(\Omega) \subset \{0, \dots, n\}$$
$$P(X = k) = \binom{n}{k} p^k q^{n-k}$$

- $E(X) = np$
- $V(X) = npq$

Démonstration. □

Théorème 9 : Cas d'une loi Binomiale.

On suppose $X \sim \mathcal{B}(n,p)$ pour un certain $(n,p) \in \mathbb{N} \times [0,1]$ (on note $q = 1-p$). Alors :

- $E(X) = np$
- $V(X) = npq$

Démonstration.

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} kP(X=k) \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k} \\ &= n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k q^{(n-1)-(k-1)} \\ &= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i q^{(n-1)-i} \\ &= np \underbrace{(p+q)^{n-1}}_{1} \\ &= np \end{aligned}$$

$$\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= \sum_{k=0}^{n} k^2 P(X=k) - \left(\sum_{k=0}^{n} kP(X=k)\right)^2 && \text{transfert} \\
&= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} - n^2p^2 \\
&= \sum_{k=0}^{n} nk \binom{n-1}{k-1} p^k q^{n-k} - n^2p^2 \\
&= n\sum_{k=0}^{n} (k-1) \binom{n-1}{k-1} p^k q^{n-k} + n\sum_{k=0}^{n} \binom{n-1}{k-1} p^k q^{n-k} - n^2p^2 \\
&= n\sum_{k=0}^{n} (n-1) \binom{n-2}{k-2} p^k q^{n-k} + np - n^2p^2 \\
&= n(n-1)p^2 \underbrace{\sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{(n-2)-(k-2)}}_{(p+q)^{n-2}=1} + np - n^2p^2 \\
&= \cancel{n^2p^2} - np^2 + np \cancel{- n^2p^2} \\
&= npq
\end{aligned}$$

□

IV Indépendance

IV.1 Indépendance de deux variables aléatoires

Idée : Deux variables aléatoires sont indépendantes lorsque les événements qu'elles sélectionnent sont tous indépendants entre eux.

Proposition-Définition 8 .
Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires.
Les propositions suivantes sont équivalentes ; lorsqu'elles sont réalisées on dit que X et Y sont indépendantes :

1. $\forall x \in E,\ \forall y \in F,\ P(X = x \text{ et } Y = y) = P(X = x)P(Y = y)$;
2. $\forall A \subset E,\ \forall B \subset F,\ P(X \in A \text{ et } Y \in B) = P(X \in A)P(Y \in B)$.

La première proposition est plus facile à obtenir ; on la privilégiera pour **montrer** que deux variables aléatoires sont indépendantes. La seconde proposition exprime un peu plus (en apparence) ; on la privilégiera si on veut **utiliser** que deux variables aléatoires sont indépendantes.

Démonstration. $\boxed{2 \implies 1}$ On prend le cas particulier

$$\begin{cases} A &= \{x\} \\ B &= \{y\} \end{cases}$$

$\boxed{1 \implies 2}$
Soit $(A, B) \in \mathcal{P}(E) \times \mathcal{P}(F)$

$$\begin{aligned}
A &= \coprod_{a\in A}\{a\} \\
B &= \coprod_{b\in B}\{b\} \\
(X\in A)\cap(Y\in B) &= X^{-1}(A)\cap Y^{-1}(B) \\
&= X^{-1}\left(\coprod_{a\in A}\{a\}\right)\cap Y^{-1}\left(\coprod_{b\in B}\{b\}\right) \\
&= \left(\coprod_{a\in A}X^{-1}(\{a\})\right)\cap\left(\coprod_{b\in B}Y^{-1}(\{b\})\right) && X^{-1}(\amalg)=\amalg X^{-1} \\
&= \coprod_{(a,b)\in A\times B} X^{-1}(\{a\})\cap Y^{-1}(\{b\}) && \text{par distributivité de } \cap \text{ sur } \amalg \\
&= \coprod_{(a,b)\in A\times B} ((X=a)\cap(Y=a))
\end{aligned}$$

D'où

$$\begin{aligned}
P((X\in A)\cap(Y\in B)) &= \sum_{(a,b)\in A\times B} P((X=a)\cap(Y=b)) && \text{prop 2 des probas} \\
&= \sum_{(a,b)\in A\times B} P(X=a)P(Y=b) && \text{d'après 1.} \\
&= \left(\sum_{a\in A}P(X=a)\right)\left(\sum_{b\in B}P(Y=b)\right) && \text{fubini séparé} \\
&= P(X\in A)P(Y\in B)
\end{aligned}$$

□

Exemple 14 *Un exemple déjà vu :* on lance deux dés, on note X le numéro donné par le premier et Y le numéro donné par le second. **On suppose seulement qu'on a $(X,Y)\sim\mathcal{U}(\{1,\dots,6\}^2)$.** Montrer qu'on a :

- X et Y sont indépendantes ;
- $X\sim\mathcal{U}(\{1,\dots,6\})$;
- $Y\sim\mathcal{U}(\{1,\dots,6\})$.

Et donc finalement l'équiprobabilité des tirages de couples implique non seulement l'équiprobabilité des tirages de chaque coordonnée **mais aussi** l'indépendance des coordonnées entre elles.
Soient $i,j\in[\![1,6]\!]$
On veut montrer que $P(X,Y=i,j)=P(X=i)P(Y=j); P(X=i)=P(X=j)=\frac{1}{6}$
par hypothèse

$$P(X,Y=i,j)=\frac{1}{36}$$

$$\begin{aligned}
P(X=i) &= P(X,Y \in \{i\} \times [\![1,6]\!]) \\
&= P\left(\coprod_{k=1}^{6} (X,Y=i,k)\right) \\
&= \sum_{k=1}^{6} \underbrace{P(X,Y=i,k)}_{\frac{1}{36}} \\
&= \frac{6}{36} = \frac{1}{6}
\end{aligned}$$

De même,

$$\begin{aligned}
P(Y=j) &= P\left(\coprod_{k=1}^{6} X,Y=k,j\right) \\
&= \sum_{k=1}^{6} \underbrace{P(X,Y=k,j)}_{\frac{1}{36}} \\
&= \frac{6}{36} = \frac{1}{6}
\end{aligned}$$

Contre-exemple 15 On dispose de r boules rouges et b boules bleues, chacune rangée dans sa boite individuelle (il y a donc $r+b$ boites différentes, chacune contenant une boule). Dimitri Matias-Pujante sort les $r+b$ boules de leurs boites et joue avec, puis remet chaque boule dans une boite aléatoirement, sans se préoccuper de les remettre dans la bonne boite. On note alors R le nombre de boules rouges rangées à leur place et B le nombre de boules bleues rangées à leur place. Alors R et B ne sont pas indépendantes. En effet :

$$\begin{cases} R(\Omega) & = [\![0,r]\!] \\ B(\Omega) & = [\![0,b]\!] \end{cases}$$

$$\underbrace{P(R=r \cap B=b-1)}_{=0} \neq \underbrace{\underbrace{P(R=r)}_{\neq 0}\underbrace{P(B=b-1)}_{\neq 0}}_{\neq 0}$$

Application 4 Soient $n_1, n_2 \in \mathbb{N}$ et $p \in [0,1]$. Soient X_1 et X_2 deux variables aléatoires telles que $X_1 \sim \mathcal{B}(n_1,p)$, $X_2 \sim \mathcal{B}(n_2,p)$, **et $\boldsymbol{X_1}$ et $\boldsymbol{X_2}$ sont indépendantes**. Déterminons la loi de X_1+X_2.

Montrons $X_1 + X_2 \sim \mathcal{B}(n_1+n_2, p)$

$$\begin{aligned}(X_1+X_2)(\Omega) &\subset X_1(\Omega)+X_2(\Omega)\\ &\subset [\![0, n_1+n_2]\!]\end{aligned}$$

Soit $k \in [\![0, n_1+n_2]\!]$

$$\begin{aligned}P(X_1+X_2=k) &= P\left(\coprod_{i=0}^{k}((X_1=i)\cap(X_2=k-i))\right)\\ &= \sum_{i=0}^{k} P((X_1=i)\cap(X_2=k-i))\\ &= \sum_{i=0}^{k} P(X_1=i)P(X_2=k-i) && \text{par indépendance}\\ &= \sum_{i=0}^{k}\binom{n_1}{i}p^i q^{n_1-i}\binom{n_2}{k-i}p^{k-i}q^{n_2-k+i}\\ &= p^k q^{n_1+n_2-k}\sum_{i=0}^{k}\binom{n_1}{i}\binom{n_2}{k-i}\\ &= p^k q^{n_1+n_2-k}\binom{n_1+n_2}{k} && \text{d'après Vandermonde}\end{aligned}$$

Théorème 10 .

Soient $X:\Omega\to E$ et $Y:\Omega\to F$ deux variables aléatoires indépendantes, et $f:E\to E'$ et $g:F\to F'$ deux applications. Alors $f(X)$ et $g(Y)$ sont indépendantes.

On verra une généralisation en TD, et un énoncé encore plus fort en seconde année.

IV.2 Indépendance, espérance et variance

Théorème 11 .

Soient X et Y deux v.a.r. Si X et Y sont indépendantes, alors $E(XY)=E(X)E(Y)$.

/!\ Ce n'est pas vrai si X et Y ne sont pas indépendantes. Contre-exemple : e.g. $Y = X \sim \mathcal{B}(\Omega)$, $p \notin \{0,1\}$

$$\begin{aligned}XY &= X^2\\ &= X\\ \implies E(XY) &= E(X) = p\\ E(X)E(Y) &= E(X)^2 = r^2\end{aligned}$$

Démonstration. Notons $Z = XY = f(X,Y)$ où $f:\mathbb{R}^2\to\mathbb{R} = (x,y)\mapsto x\cdot y$

$$
\begin{aligned}
E(Z) &= E \circ f(X,Y) \\
&= \sum_{(x,y)\in(X,Y)(\Omega)} f(x)yP(X,Y=x,y) && \text{transfert} \\
&= \sum_{(x,y)\in(X,Y)(\Omega)} xyP(X=x\cap Y=y) \\
(X,Y)(\Omega) &= \{(X,Y)(\omega), \omega\in\Omega\} \\
&= \{X(\omega), Y(\omega), \omega\in\Omega\} \\
&\subset \{X(\omega_1), Y(\omega_2), \binom{\omega_1}{\omega_2}\in\Omega^2\} && = X(\Omega)\times Y(\Omega)
\end{aligned}
$$

Même si l'inclusion est stricte,

$$
\begin{aligned}
\text{pour } (x,y)\in X(\Omega)\times Y(\Omega)\setminus(X,Y)(\Omega), P(X=x\cap Y=y) = 0 \\
\text{donc } E(Z) &= \sum_{x,y\in(X,Y)(\Omega)} xyP(X=x\cap Y=y) && \text{quitte à rajouter des 0} \\
&= \sum_{x,y\in(X,Y)(\Omega)} xyP(X=x)P(Y=y) && \text{par indépendance} \\
&= \sum_{x\in X(\Omega)} xP(X=x)\sum_{y\in Y(\Omega)} yP(Y=y) \\
&= E(X)E(Y)
\end{aligned}
$$

□

/!\ La réciproque de ce théorème n'est pas vraie non plus, on le verra dans la dernière section.

Corollaire 1 .

Soient X et Y deux v.a.r. Si X et Y sont indépendantes, alors $V(X+Y) = V(X)+V(Y)$.

DÉMONSTRATION. Si X et Y sont indépendantes :

$$
\begin{aligned}
V(X+Y) &= E((X+Y-E(X+Y))^2) \\
&= E((X-E(X)+Y-E(Y))^2) \\
&= E((X-E(X))^2+(Y-E(Y))^2+2(X-E(X))(Y-E(Y))) \\
&= E((X-E(X))^2)+E((Y-E(Y))^2)+2E(XY-XE(Y)-YE(X)+E(X)E(Y)) \\
&= V(X)+V(Y)+2E(XY-XE(Y)-YE(X)+E(X)E(Y)) \\
&= V(X)+V(Y)+2\left(\underbrace{E(XY)-E(X)E(Y)}_{\text{0 par indépendance}}\right) \\
&= V(X)+V(Y)
\end{aligned}
$$

□

De même :

Théorème 12 .

Soient $X_1,\dots,X_n$ des v.a.r. indépendantes deux à deux, alors $V(X_1+\dots+X_n) = V(X_1)+\dots+V(X_n)$.

DÉMONSTRATION. Idem. Exercice. □

IV.3 Indépendance mutuelle

Proposition-Définition 9 .

Soient $X_1 : \Omega \to E_1$, ..., et $X_n : \Omega \to E_n$ des variables aléatoires. Les propositions suivantes sont équivalentes ; lorsqu'elles sont réalisées on dit que $X_1, \dots, X_n$ sont mutuellement indépendantes :

1. $\forall (x_1, \dots, x_n) \in E_1 \times \dots \times E_n,\ P(X_1 = x_1 \text{ et ... et } X_n = x_n) = P(X_1 = x_1) \times \dots \times P(X_n = x_n)$;
2. $\forall (A_1, \dots, A_n) \in \mathcal{P}(E_1) \times \dots \times \mathcal{P}(E_n),\ P(X_1 \in A_1 \text{ et ... et } X_n \in A_n) = P(X_1 \in A_1) \times \dots \times P(X_n \in A_n)$.

Démonstration. La même que pour l'indépendance 2 à 2. Exercice. □

L'indépendance mutuelle implique l'indépendance 2 à 2, mais la réciproque est fausse.

Théorème 13 .

Soient $p \in [0, 1]$ et $X_1, X_2, \dots, X_n$ des variables aléatoires **mutuellement indépendantes** telles que $\forall i,\ X_i \sim \mathcal{B}(p)$. Alors $X_1 + X_2 + \dots + X_n \sim \mathcal{B}(n, p)$.

Ce théorème est donc un énoncé mathématique précis donnant un sens au blabla introductif sur la loi binomiale.

Démonstration.

$$\forall i, X_i(\Omega) \subset \{0, 1\} \text{ donc } (X_1, X_2, \dots, X_n)(\Omega) \subset \{0, 1, \dots, n\}$$

Soit $k \in \{0, \dots, n\}$

$$\begin{aligned}
P(X_1 + \dots + X_n = k) &= P\left(\coprod_{(\epsilon_1, \dots, \epsilon_n) \in \{1, \dots, n\}^n \wedge \sum_{i=1}^n \epsilon_i = k} (X_1 = \epsilon_1) \cap \dots \cap (X_n = \epsilon_n) \right) \\
&= \sum_{(\epsilon_1, \dots, \epsilon_n) \in \{1, \dots, n\}^n \wedge \sum_{i=1}^n \epsilon_i = k} P(X_1 = \epsilon_1 \cap \dots \cap X_n = \epsilon_n) \\
&= \sum_{\text{la même}} \underbrace{P(X_1 = \epsilon_1) \cdots P(X_n = \epsilon_n)}_{p^k q^{n-k}} && \text{par indépendance} \\
&= p^k q^{n-k} \sum_{(\epsilon_1, \dots, \epsilon_n) \in \{0,1\}^n \wedge \epsilon_1 + \dots + \epsilon_n = k} 1 \\
&= p^k q^{n-k} \# \{(\epsilon_1, \dots, \epsilon_n) \in \{0, 1\}^n, \epsilon_1 + \dots + \epsilon_n = k\} \\
&= \binom{n}{k} p^k q^{n-k}
\end{aligned}$$

□

V Vecteurs aléatoires

V.1 Exemples

On rappelle qu'un vecteur aléatoire est une application $f : \Omega \to E$ avec $E \subset \mathbb{R}^n$.

Exemple 16 On considère l'expérience aléatoire suivante : on tire un candidat au concours commun des Mines en 2019. On souhaite savoir si le candidat tiré au sort est admissible ; on se dit qu'il suffit de comparer la moyenne de ses résultats aux épreuves écrites pondérée par les coefficients de ces épreuves, et de la comparer à la barre d'admissibilité. **Mais ce n'est pas si simple !**

> Sont déclarés admissibles les candidats qui sont titulaires d'un premier minimum de points obtenu pour l'ensemble des épreuves d'admissibilité et, en outre, d'un second minimum de points obtenu pour l'ensemble des quatre compositions de mathématiques et de physique. [...] Tout candidat ayant obtenu une note inférieure à 2 à l'épreuve de français est déclaré non admissible.

Extrait de la notice du Concours Commun Mines Ponts.

Bref, ce qu'on veut mesurer dans cette expérience aléatoire c'est la donnée de toutes les notes, et on ne peut pas vraiment faire mieux. On s'intéresse donc au vecteur aléatoire

$$Z : \begin{cases} \Omega & \to & \mathbb{R}^9 \\ \omega & \mapsto & \begin{pmatrix} M_1(\omega) \\ M_2(\omega) \\ P_1(\omega) \\ \vdots \\ F(\omega) \end{pmatrix} \end{cases}$$

où $M_1(\omega)$ est la note obtenue à l'épreuve Maths 1 par le candidat lors de la réalisation de ω, etc.

Dans toute la suite on s'intéresse surtout aux couples aléatoires, c'est-à-dire au cas $n = 2$. La généralisation au cas général ne pose pas de difficulté particulière.

Se donner un couple aléatoire, c'est donc se donner deux variables aléatoires, mais attention, elles doivent être définies sur le même espace probabilisé. Est-ce si grave ?

Remarque 8

"Théorème de Nil Venet" : quitte à considérer que Ω modélise la réalisation de toutes les expériences aléatoires possibles et imaginables, on peut toujours considérer que deux variables aléatoires quelconques sont définies sur le même univers Ω.

Considérons encore deux exemples très simplistes que l'on utilisera dans la suite pour illustrer les définitions à venir.

Exemples 17
Soit X telle que $X \sim \mathcal{U}(-1,0,1)$, $Y_1 = X^2$ et $Y_2 = \min(X,0)+1$: $Z_1 = \binom{X}{Y_1}$ et $Z_2 = \binom{X}{Y_2}$ sont deux couples aléatoires.

V.2 Lois associées à un vecteur aléatoire

Définition 10 .

Soit $Z = \binom{X}{Y}$ un couple aléatoire.

- On appelle loi conjointe de X et Y la loi de Z.
- On appelle lois marginales de Z les lois de X et Y.

Exemple 18 Reprenons l'exemple 17. On peut représenter toutes ces lois à l'aide de tableaux.

Table 1 – $Z_1 = \begin{pmatrix} X \\ Y_1 \end{pmatrix}$

$\frac{X}{Y_1}$	-1	0	1	Σ
0	0	$\frac{1}{3}$	0	$1/3$
1	$\frac{1}{3}$	0	$\frac{1}{3}$	$2/3$
Σ	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

Table 2 – $Z_2 = \begin{pmatrix} X \\ Y_2 \end{pmatrix}$

$\frac{X}{Y_2}$	-1	0	1	Σ
0	$\frac{1}{3}$	0	0	$1/3$
1	0	$\frac{1}{3}$	$\frac{1}{3}$	$2/3$
Σ	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

1. Le tableau donne seulement les $P(Z = (x, y))$ mais d'après .... cela suffit.
2. Le tableau de la loi conjointe illustre qu'en général on a seulement $(X, Y)(\Omega)$ ..... $X(\Omega) \times Y(\Omega)$.
3. Une fois obtenu le tableau de la loi conjointe, on en déduit les lois marginales en ....
4. Cet exemple en particulier illustre que ....
5. **Cependant**, ...

Exercice 1. Donner la loi conjointe de X et Y_3 telles que
$$\begin{cases} X & \sim \mathcal{U}(-1, 1) \\ Y_3 & \sim \mathcal{B}(\frac{2}{3}) \\ X & \text{indép } Y_3 \end{cases}$$

$\frac{X}{Y_3}$	-1	0	1	
0	$1/9$	$1/9$	$1/9$	$1/3$
1	$2/9$	$2/9$	$2/9$	$2/3$
	$1/3$	$1/3$	$1/3$	

Définition 11 .

Soit $Z = \binom{X}{Y}$ et $A \in X(\Omega)$. On appelle <u>loi de Y conditionnellement à $X \in A$</u> (ou loi de Y sachant $X \in A$) l'application $\left\{ \begin{array}{ccc} \mathcal{P}(Y(\Omega)) & \to & [0, 1] \\ B & \mapsto & P_{X \in A}(Y \in B) \end{array} \right.$.

Exemple 19 On reprend l'exemple 17. La loi de Y_1 est .... Mais la loi de Y_1 sachant $X \neq -1$ est ....

V.3 Covariance

Définition 12 .

Soit $\binom{X}{Y}$ un couple aléatoire. On appelle <u>covariance de X et Y</u> le réel $\mathrm{Cov}(X,Y) = E((X-E(X))(Y-E(Y)))$.

On a ainsi $V(X) =$ ....

Théorème 14 : Kœnig-Huygens.

Soit $\binom{X}{Y}$ un couple aléatoire. Alors $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.

Démonstration.

□

Corollaire :

Théorème 15 .

Soit $\binom{X}{Y}$ un couple aléatoire. Si X et Y sont indépendantes alors ....

Remarque 9

La réciproque est fausse **et on l'a déjà vu**!

En effet :

Théorème 16 .

La covariance est une forme bilinéaire, symétrique, et positive (seulement).

DÉMONSTRATION. On veut montrer :

□

Corollaire 2 : Cauchy-Schwarz.

Soit $\binom{X}{Y}$ un couple aléatoire. Alors $|\mathrm{Cov}(X, Y)| \leqslant \sigma(X)\sigma(Y)$.

DÉMONSTRATION. On copie la preuve dans les préhilbertiens. Celle-ci ne nécessite que la positivité. Par contre, pour le cas d'égalité de Cauchy-Schwarz dans les préhilbertiens, on a besoin de la définie-positivité, du coup le cas d'égalité n'est plus le même ici, on le verra dans le corollaire 3. □

On a maintenant une formule générale pour la variance d'une somme :

Théorème 17 .

Soient X_1, ..., X_n des var sur Ω. $V(X_1 + \cdots + X_n) =$ ....

V.4 Indépendance vs décorrélation

Définition 13 .

Soit $\binom{X}{Y}$ un couple aléatoire. On appelle coefficient de corrélation de X et Y le réel $\rho_{X,Y} = \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$.

C'est le truc qu'on utilise en physique pour valider un modèle (et que Regressi ou Excel calculent tout seul). On va essayer de comprendre pourquoi.

Cauchy-Schwarz se reformule : $-1 \leqslant \rho_{X,Y} \leqslant 1$.

Définition 14 .

Soit $\binom{X}{Y}$ un couple aléatoire. On dit que X et Y sont décorrélées lorsqu'on a $\rho_{X,Y} = 0$, i. e. $\mathrm{Cov}(X,Y) = 0$.

Le corollaire 15 et la remarque 10 se reformulent :

- **l'indépendance implique la décorrélation ;**
- **mais deux variables décorrélées peuvent ne pas être indépendantes.**

Lemme 1 .

Soit X une v.a.r. finie. On a $V(X) = 0 \Leftrightarrow X$ constante.

Attention, lorsqu'on sort du cadre fini, ce n'est plus vrai.

DÉMONSTRATION.

□

Corollaire 3 .

Soit $\binom{X}{Y}$ un couple aléatoire avec X non constante. On a $|\rho_{X,Y}| = 1 \Leftrightarrow \exists (a,b) \in \mathbb{R}^2,\ Y = aX + b$.

Ainsi un coefficient de corrélation maximal en valeur absolue correspond au fait qu'il existe une relation de dépendance linéaire entre X et Y. Mais le coefficient de corrélation peut être nul même pour des couples (X,Y) tels que Y est de la forme $f(X)$ avec f une fonction non linéaire...

DÉMONSTRATION.

□