Chapitre 27

Variables aléatoires sur un univers fini

On se donne dans tout le chapitre un espace probabilisé fini $(\Omega, \mathscr{P}(\Omega), P)$, ainsi que des ensembles E et F non vides.

I Variable aléatoire

I.1 Introduction et notation

Définition. Une variable aléatoire est une application définie sur Ω à valeurs dans E. L'ensemble $X(\Omega) = \{X(\omega) \,|\, \omega \in \Omega\}$ est l'ensemble des valeurs prises par la variable aléatoire X. On l'appelle *univers image* de X.

En abrégé : v.a.

L'univers image (qui est donc par définition une partie de E) est aussi parfois appelé support de X.

Remarques :

- Comme Ω est fini, $X(\Omega)$ l'est aussi. Si $n = \text{card}(X(\Omega))$, alors on peut écrire $X(\Omega) = \{x_1; \dots; x_n\}$, où les x_i, $1 \leqslant i \leqslant n$ sont des réels distincts.
- Si $\omega \in \Omega$, on dit que $X(\omega)$ est une réalisation de X.
- Le terme « variable aléatoire » peut être trompeur : une variable aléatoire est une **fonction**.
- Pourquoi cette définition ? Intuitivement, une v.a. X est un objet qui varie de manière aléatoire. Autrement dit, la valeur de X dépend du hasard, et comme on a représenté le hasard par le choix d'un élément ω dans l'univers des éventualités Ω, il est naturel de considérer X comme un objet qui varie avec ω c'est-à-dire comme une application qui à chaque ω associe un nombre $X(\omega) \in \mathbb{R}$.
- Ce point est fondamental car la définition intuitive de X comme nombre qui varie aléatoirement fait de X un objet mouvant sur lequel on n'a pas de prise et qui ne se prête pas à l'analyse mathématique, tandis que si on considère les v.a. comme des fonctions définies sur Ω, on peut définir des opérations sur ces fonctions et ainsi développer un calcul sur ces variables aléatoires.

On se donne dans la suite $X : \Omega \to E$ une variable aléatoire.

Notations : on va en quelque sorte « transférer » sur E les opérations faites jusque là sur Ω ce qui sera plus pratique car E est en général plus maniable que Ω (la plupart du temps, E sera égal à $\mathbb{R}$).

Définition. Soient $(a, b) \in \mathbb{R}^2$ tel que $a < b$ et X une variable aléatoire réelle finie sur Ω. On définit alors les événements suivants :

$$[X = a] = \{\omega \in \Omega \,|\, X(\omega) = a\}, \quad [X < a] = \{\omega \in \Omega \,|\, X(\omega) < a\},$$
$$[X \leqslant a] = \{\omega \in \Omega \,|\, X(\omega) \leqslant a\}, \quad [X > a] = \{\omega \in \Omega \,|\, X(\omega) > a\},$$
$$[X \geqslant a] = \{\omega \in \Omega \,|\, X(\omega) \geqslant a\}, \quad [a \leqslant X \leqslant b] = \{\omega \in \Omega \,|\, a \leqslant X(\omega) \leqslant b\}.$$

Plus généralement, si A est une partie de $\mathbb{R}$, alors on note

$$[X \in A] = \{\omega \in \Omega \,|\, X(\omega) \in A\}.$$

On trouve aussi les notations $(X = a)$ (i.e. avec des parenthèses) ou $\{X = a\}$ (i.e. avec des accolades). De plus, pour ne pas surcharger la notation, on écrira la plupart du temps $P(X = a)$ plutôt que $P([X = a])$ etc.

Remarque : Les ensembles ci-dessus sont des parties de Ω, pas des parties de E ! Plus précisément, l'ensemble $[X \dots]$ est l'ensemble des $\omega \in \Omega$ dont l'image par X, notée $X(\omega)$, vérifie la condition ...

Exemples :

- On s'intéresse à l'expérience qui consiste à lancer deux fois un dé équilibré, et on définit X la variable aléatoire qui donne la somme des deux résultats obtenus. Formellement, on se place sur $\Omega = [\![1\,;\,6]\!]^2$ muni de $\mathscr{P}(\Omega)$ muni de l'équiprobabilité (le dé est équilibré) et on définit :

$$X : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ (i,j) & \longmapsto & i+j \end{array} \right.$$

Alors on a par exemple :

  - ⋆ $[X \leqslant 4] = \{(1,1);(1,2);(1,3);(2,1);(2,2);(3,1)\}$.
  - ⋆ $[X = 6] = \{(1,5);(2,4);(3,3);(4,2);(5,1)\}$.

Ce sont des parties de $\Omega = [\![1\,;\,6]\!]^2$. En particulier, ce sont des événements : on peut donc calculer leur probabilité (puisqu'on a muni Ω de l'équiprobabilité) : on trouve

$$\begin{aligned} P(X \leqslant 4) &= \frac{\mathrm{card}(X \leqslant 4)}{\mathrm{card}(\Omega)} \\ &= \frac{6}{36} \\ &= \frac{1}{6} \end{aligned}$$

et on trouve de même que $P(X = 6) = 5/36$.

- On dispose d'une urne contenant n boules numérotées de 1 à n. On tire successivement et sans remise p boules (avec $p \leqslant n$). Un univers associé à cette expérience est Ω l'ensemble des p-uplets d'éléments distincts de $[\![1\,;\,n]\!]$, muni de $\mathscr{P}(\Omega)$ et de l'équiprobabilité. L'application

$$X : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ (\omega_1, \dots, \omega_p) & \longmapsto & \displaystyle\max_{1 \leqslant i \leqslant n} \omega_i \end{array} \right.$$

est une variable aléatoire réelle finie qui représente le numéro maximal parmi les p boules tirées. On a alors $X(\Omega) = [\![p\,;\,n]\!]$.

- On choisit une carte au hasard dans un jeu de 52 cartes, et on gagne 5 euros si on tire une figure (y compris l'as) et on perd 5 euros sinon. Un univers associé à cet expérience est Ω, l'ensemble des 52 cartes, muni de $\mathscr{P}(\Omega)$ et de l'équiprobabilité. L'application

$$X : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & \left\{ \begin{array}{ll} 5 & \text{si } \omega \text{ est une figure.} \\ -5 & \text{sinon} \end{array} \right. \end{array} \right.$$

est une variable aléatoire réelle qui représente le gain (algébrique) de l'expérience aléatoire. Par exemple, $[X = 5]$ est l'ensemble à 16 éléments formé par les 4 valets, les 4 dames, les 4 rois et les 4 as, si bien que $P(X = 5) = 16/52 = 4/13$ et $P(X = -5) = 9/13$.

Remarques :

- Il faut bien comprendre que les ensembles du type $[X = x]$ etc. sont des **ensembles** et non pas des éléments de E (la plupart du temps, des réels). Par exemple, écrire $[X = x] + [X < x]$ n'a aucun sens !
- De plus, ces ensembles sont des parties de Ω (c'est-à-dire des événements) et non pas des parties de $\mathbb{R}$. Cependant, en pratique, on ne les explicitera jamais (car on n'explicitera que très rarement Ω, comme dans le chapitre précédent), mais ce ne sera pas très grave, cette notation permettant de les manipuler beaucoup plus facilement, en ne travaillant que sur E.

Exemples :

- Si X est une v.a. à valeurs réelles (i.e. si $E=\mathbb{R}$) et si $x\in\mathbb{R}$, alors $[X\leqslant x]=[X<x]\cup[X=x]$. De plus, comme ces événements sont incompatibles, alors $P(X\leqslant x)=P(X<x)+P(X=x)$.

  On peut l'affirmer directement, mais si on a un doute, on le démontre de la façon suivante : soit $\omega\in\Omega$

$$\begin{aligned}\omega\in[X\leqslant x] &\iff X(\omega)\leqslant x\\ &\iff X(\omega)<x \quad\text{ou}\quad X(\omega)=x\\ &\iff \omega\in[X<x]\quad\text{ou}\quad \omega\in[X=x]\\ &\iff \omega\in[X<x]\cup[X=x]\end{aligned}$$

  Les deux ensembles $[X\leqslant x]$ et $[X<x]\cup[X=x]$ ont les mêmes éléments donc sont égaux. Donnons d'autres exemples.
- $[X\leqslant a]\subset[X\leqslant b]$ si $a\leqslant b$.
- $[a<X\leqslant b]=[X\leqslant b]\backslash[X\leqslant a]$ (toujours si $a\leqslant b$). En particulier, $P(a<X\leqslant b)=P(X\leqslant b)-P(X\leqslant a)$.
- $[X\leqslant b]\cap[X\geqslant a]=[a\leqslant X\leqslant b]$.
- ⚠ Si X est à valeurs dans $\mathbb{Z}$ et si $k\in\mathbb{Z}$, alors $[X=k]=[X\geqslant k]\backslash[X\geqslant k+1]$. En particulier, $P(X=k)=P(X\geqslant k)-P(X\geqslant k+1)$.

Rappelons (après tout, l'enseignement est l'art de la répétition) que ce sont des parties de Ω.

Ce résultat est très important car il apparaît dans de nombreux exercices, par exemple l'exercice 16.

Proposition. Soit X une variable aléatoire réelle finie sur Ω. La famille $([X=x])_{x\in X(\Omega)}$ est un système complet fini d'événements. On l'appelle le système complet associé à X.

Démonstration. Pour tout $x\in X(\Omega)$, $[X=x]\in\mathscr{P}(\Omega)$ par définition. Si x et y sont distincts dans $X(\Omega)$, alors $[X=x]\cap[X=y]=\varnothing$ si bien que les événements $[X=x]$ et $[X=y]$ sont incompatibles. Enfin, si $\omega\in\Omega$, alors $\omega\in[X=X(\omega)]\subset\displaystyle\bigcup_{x\in X(\Omega)}[X=x]\subset\Omega$. Ainsi $\displaystyle\bigcup_{x\in X(\Omega)}[X=x]=\Omega$. $\square$

On appliquera souvent la formule des probabilités totales à ce système complet d'événements. Rappelons que la formule des probabilités totales est le théorème le plus important de tout le cours de probabilités. De plus, on en déduit que

$$\sum_{x\in E}P(X=x)=P(X\in\Omega)=1$$

On retrouve le fameux : « la somme des probas vaut 1 ». On s'en servira pour caractériser la loi de X, cf. paragraphe suivant.

Exemple : On lance trois fois une pièce de monnaie. On peut considérer l'univers $\Omega=\{P;F\}^3$. Soit X la variable aléatoire qui donne le nombre de Pile obtenus. On a $X(\Omega)=\{0;1;2;3\}$ et $[X=0]=\{(F,F,F)\}$, $[X=3]=\{(P,P,P)\}$,

$$[X=1]=\{(P,F,F);(F,P,F);(F,F,P)\},$$

$$[X=2]=\{(P,P,F);(F,P,P);(P,F,P)\},$$

On voit que $\big([X=0],[X=1],[X=2],[X=3]\big)$ est bien un système complet d'événements.

I.2 Loi d'une variable aléatoire

I.2.a Exemples

Rappelons que, l'ensemble Ω étant fini, $X(\Omega)$, l'image de Ω par X, est aussi un ensemble fini.

Exemples :

- Dans le cas où X est la somme des deux numéros obtenus (voir ci-dessus), $X(\Omega)=[\![2\,;12]\!]$.
- Dans le cas où on tire une carte et où on gagne ou on perd 5 euros, $X(\Omega)=\{\pm5\}$.

$X(\Omega)$ étant un ensemble fini, on cherche à le munir d'une probabilité. Dans le dernier exemple, il semble absurde de le munir de l'équiprobabilité. On va plutôt le munir de la probabilité P_X définie par $P_X(\{5\}) = \dfrac{4}{13}$ et $P_X(\{-5\}) = \dfrac{9}{13}$. Rappelons en effet (cf. chapitre précédent) qu'une probabilité P_X sur $X(\Omega)$ est entièrement déterminée par la distribution de probabilités $(P_X(\{y\}))_{y\in X(\Omega)}$.

I.2.b Définition

Proposition/Définition. L'application

$$P_X : \left\{ \begin{array}{ccc} \mathscr{P}(X(\Omega)) & \to & [\,0\,;1\,] \\ A & \mapsto & P(X\in A) = P\left(\{\omega\in\Omega \mid X(\omega)\in A\}\right) \end{array} \right.$$

est une probabilité sur $(X(\Omega), \mathscr{P}(\Omega))$, appelée la loi de X.

Démonstration.

- P étant une probabilité, elle est à valeurs dans $[\,0\,;1\,]$ donc P_X aussi.
- $P_X(X(\Omega)) = P(X\in X(\Omega)) = P(\Omega) = 1$.
- Soient A et B deux éléments incompatibles de $\mathscr{P}(X(\Omega))$. Alors $P_X(A\cup B) = P(X \in A\cup B)$. Montrons que $[X\in A\cup B] = [X\in A]\cup[X\in B]$.

  Soit $\omega\in\Omega$ (on rappelle que ces ensembles sont des parties de Ω).

$$\begin{aligned} \omega\in[X\in A\cup B] &\iff X(\omega)\in A\cup B \\ &\iff X(\omega)\in A \text{ ou } X(\omega)\in B \\ &\iff \omega\in[X\in A] \text{ ou } \omega\in[X\in B] \\ &\iff \omega\in[X\in A]\cup[X\in B] \end{aligned}$$

  Ainsi, $[X\in A\cup B] = [X\in A]\cup[X\in B]$ donc $P(X\in A\cup B) = P([X\in A]\cup[X\in B])$. Or, A et B sont incompatibles donc $[X\in A]$ et $[X\in B]$ également (exo). P étant une probabilité,

$$\begin{aligned} P(X\in A\cup B) &= P(X\in A) + P(X\in B) \\ &= P_X(A) + P_X(B) \end{aligned}$$

□

Remarque : On rappelle que $[X\in A]$ est l'ensemble des antécédents des éléments de A par X. Ainsi, cette définition est complètement intuitive : la probabilité de A est en fait la probabilité de l'ensemble de ses antécédents !

I.3 Caractérisation de la loi d'une variable aléatoire

Remarque : Comme dit ci-dessous (cf. chapitre précédent), une probabilité est entièrement caractérisée par son image en les singletons. On peut adapter ce résultat au cas de la loi d'une variable aléatoire.

Corollaire.

- La famille $(P(X=x))$ est une distribution de probabilités sur $X(\Omega)$.

- La loi de X est entièrement déterminée par la distribution de probabilités $(P(X = x))_{x\in X(\Omega)}$.

DÉMONSTRATION. Les événement $[X = x]$ formant un système complet d'événements, la somme de leurs probabilités vaut 1 : la famille $(P(X = x))_{x\in X(\Omega)}$ est donc une famille de réels positifs indexée par E et de somme 1, c'est donc par définition une distribution de probabilités sur $X(\Omega)$. Le deuxième point en découle, puisqu'une probabilités est entièrement déterminée par la donnée d'une distribution de probabilités (cf. chapitre précédent).

Conclusion : La loi d'une v.a. X est entièrement déterminée par $X(\Omega)$ et les $P(X = x)$ pour $x \in X(\Omega)$, c'est-à-dire que quand on demande la loi d'une v.a., il suffit de connaître $X(\Omega)$ et $P(X = x)$ pour tout $x \in X(\Omega)$. On peut alors calculer $P(X \in A)$ pour tout $A \subset X(\Omega)$, cf. chapitre précédent et ci-dessous. Plus précisément :

On peut aussi répondre en disant que X suit une loi de référence (cf. paragraphe IV).

Corollaire. Si A est une partie de E, alors

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

Remarque : On peut se poser la question inverse : étant donnée une probabilité sur un ensemble fini E, existe-t-il un espace probabilisé $(\Omega, \mathscr{P}(\Omega), P)$ et une variable aléatoire $X : \Omega \to E$ telle que cette probabilité soit la loi de X ?

Théorème. Soit $(p_e)_{e\in E}$ une distribution de probabilités sur E. Il existe un un espace probabilisé $(\Omega, \mathscr{P}(\Omega), P)$ et une variable aléatoire $X : \Omega \to E$ telle que pour tout $e \in E, P(X = e)$. et

Il n'y a unicité ni de $(\Omega, \mathscr{P}(\Omega), P)$, ni de la variable aléatoire X.

DÉMONSTRATION. Il suffit de prendre $\Omega = E$ et $X = \mathrm{Id}_E$. Pour que X ait la bonne loi, il suffit de munir Ω de la probabilité P définie par la distribution de probabilités $(p_e)_{e\in E}$ c'est-à-dire l'unique probabilité telle que $P(\{e\}) = p_e$ pour tout $e \in E$. On sait que P est une probabilité, que X est une variable aléatoire de Ω dans E, et que pour tout $e \in E$, $P(X = e) = P(\{e\}) = p_e$.

Exemple : Soit $n \in \mathbb{N}$. La formule du binôme de Newton entraîne que $\displaystyle\sum_{k=0}^{n} \binom{n}{k} = 2^n$ donc $\displaystyle\sum_{k=0}^{n} \frac{1}{2^n}\binom{n}{k} = 1$. Il s'agit d'une somme de réels positifs donc il existe une variable aléatoire réelle finie X telle que $X(\Omega) = [\![0\,;\,n]\!]$ et telle que :

$$\forall k \in [\![0\,;\,n]\!], \qquad P(X = k) = \frac{1}{2^n}\binom{n}{k}.$$

Dans le paragraphe IV.4, nous verrons que X suit alors une loi binomiale de paramètres n et $1/2$.

I.3.a Un exemple classique

Pour donner la loi d'une variable aléatoire réelle finie X, on peut donner une formule générale pour $P(X = x)$, $x \in X(\Omega)$. On peut aussi (lorsque $X(\Omega)$ contient peu de valeurs), dresser un tableau de valeurs ou construire un diagramme en bâtons (pour chaque $x \in X(\Omega)$, on trace un bâton de longueur $P(X = x)$ situé à l'abscisse x).

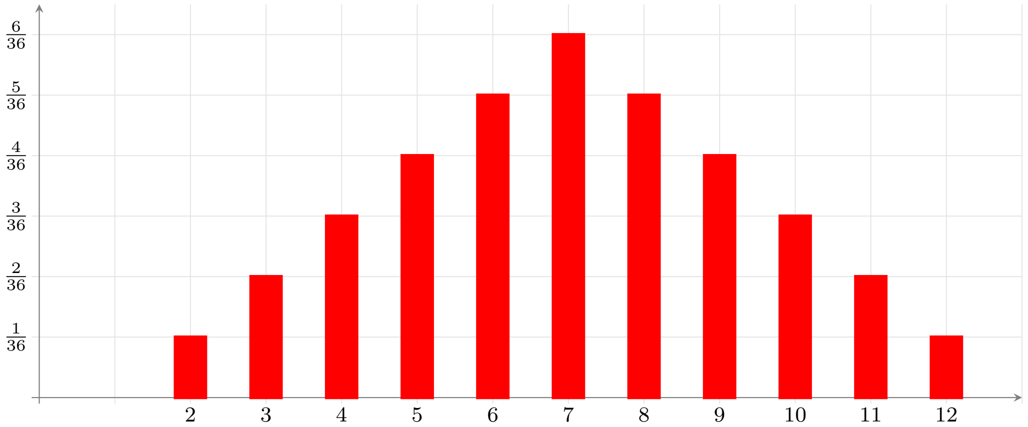
Exemple : Reprenant l'exemple de la variable aléatoire X donnant la somme des chiffres des faces obtenues par le lancer de deux dés à 6 faces (on munit l'univers $\Omega = [\![1\,;\,6]\!]^2$ de la probabilité uniforme). Déterminons sa loi : on a $X(\Omega) = [\![2\,;\,12]\!]$ et, si $k \in [\![2\,;\,12]\!]$, $P(X = k) = \dfrac{\mathrm{card}([X = k])}{\mathrm{card}(\Omega)} = \dfrac{\mathrm{card}([X = k])}{36}$. On a

$$[X=2]=\{(1,1)\},\quad [X=3]=\{(1,2);(2,1)\},$$
$$[X=4]=\{(2,2);(1,3);(3,1)\},\quad [X=5]=\{(1,4);(2,3);(3,2);(4,1)\},\quad \cdots$$

donc on obtient le tableau :

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

On peut aussi tracer son diagramme en bâtons :



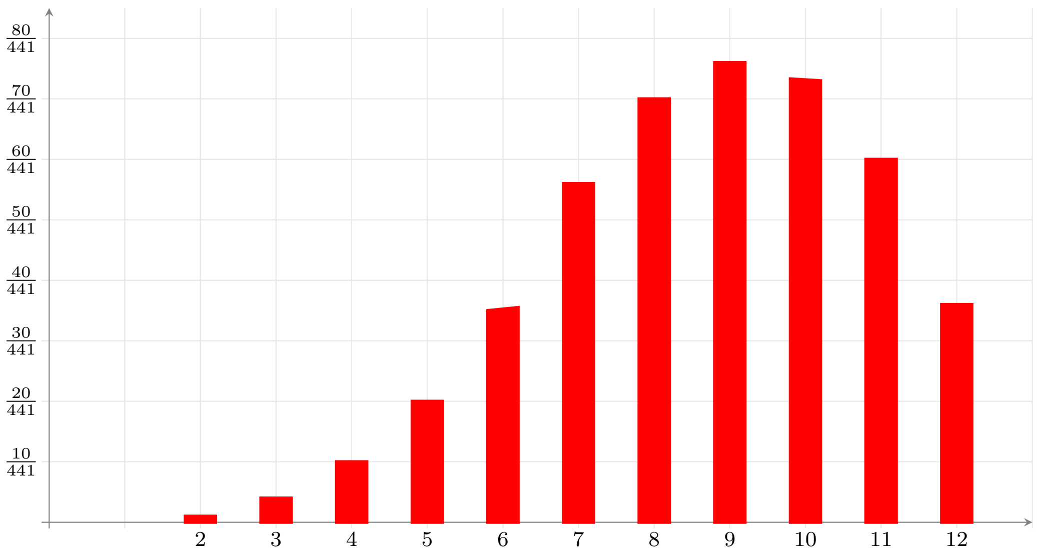
On remarque que le diagramme est symétrique par rapport à 7 : c'est un cas particulier de loi triangulaire (cf. exercice 60).

Remarque : Attention, la loi de X dépend de la probabilité P ! Par exemple, le tableau ci-dessus a été construit en supposant que le dé est équilibré, c'est-à-dire en supposant que P est l'équiprobabilité. Mais si ce n'est plus le cas ?
Exemple : On refait cette expérience avec le dé du chapitre précédent, avec lequel la probabilité d'obtenir l'entier $k \in [\![1\,;6]\!]$ est proportionnelle à k. On rappelle que pour tout $k \in [\![1\,;6]\!]$, $P(X=k)=k/21$. Les ensembles $[X=k]$ ci-dessus sont inchangés, c'est leur probabilité, et donc la loi de X, qui change :

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{441}$	$\frac{4}{441}$	$\frac{10}{441}$	$\frac{20}{441}$	$\frac{35}{441}$	$\frac{56}{441}$	$\frac{70}{441}$	$\frac{76}{441}$	$\frac{73}{441}$	$\frac{60}{441}$	$\frac{36}{441}$

L'histogramme devient :

On remarque qu'il est « décalé vers la droite ». C'est normal : les plus gros numéros ont une plus grosse probabilité de sortir !



La loi de X a changé alors que X est toujours la même fonction de Ω dans $\mathbb{R}$: c'est la proba P sur $X(\Omega)$ qui a changé.

I.3.b Égalité en loi

Remarques :

- On a vu dans le chapitre précédent que l'univers Ω joue très souvent un rôle inerte. C'est encore pire avec les variables aléatoires : les notations $[X=a]$ etc. permettent de transférer sur E le travail fait auparavant sur Ω. Par exemple, l'égalité $[X=k] = [X \geqslant k] \backslash [X \geqslant k+1]$ se comprend bien, même sans passer sur Ω (même s'il faut savoir la prouver au cas où).
- Dans certains cas, Ω est vraiment un ensemble compliqué mais heureusement le théorème précédent nous permet d'affirmer son existence sans nous poser de questions : nous n'expliciterons plus Ω dans la suite. Nous savons qu'il existe, et ça nous suffit. Nous pourrons par exemple dire : soit X une variable aléatoire définie par $P(X=0)=p$ et $P(X=1)=1-p$ avec $p \in [\,0\,;1\,]$.
- L'idée sous-jacente est que seule la loi compte. Illustrons cette idée par un exemple.

Exemple : Réalisons deux expériences aléatoires. D'une part on lance deux fois un dé équilibré : on prend $\Omega_1 = [\![\,1\,;\,6\,]\!]^2$ muni de $\mathscr{P}(\Omega)$ et de l'équiprobabilité. Soit :

$$X : \left\{ \begin{array}{ccc} \Omega_1 & \to & \mathbb{R} \\ & & \\ (a,b) & \mapsto & \left\{ \begin{array}{ll} 1 & \text{si } a=1 \text{ ou } (a=2 \text{ et } b \leqslant 3) \\ 2 & \text{si } (a=2 \text{ et } b \geqslant 4) \text{ ou } a=3 \\ 3 & \text{si } a=4 \text{ ou } (a=5 \text{ et } b \leqslant 3) \\ 4 & \text{si } (a=5 \text{ et } b \geqslant 4) \text{ ou } a=6 \end{array} \right. \end{array} \right.$$

Alors $X(\Omega) = [\![\,1\,;\,4\,]\!]$ et $P(X=1) = P(X=2) = P(X=3) = P(X=4) = \dfrac{1}{4}$. D'autre part, on prend une carte au hasard dans un jeu de 52 cartes. On prend $\Omega_2 = \{52 \text{ cartes}\}$ muni de $\mathscr{P}(\Omega)$ et de l'équiprobabilité. Soit

$$Y : \left\{ \begin{array}{ccc} \Omega_2 & \to & \mathbb{R} \\ & & \\ \omega & \mapsto & \left\{ \begin{array}{ll} 1 & \text{si } \omega \text{ est un coeur} \\ 2 & \text{si } \omega \text{ est un carreau} \\ 3 & \text{si } \omega \text{ est un trèfle} \\ 4 & \text{si } \omega \text{ est un pique} \end{array} \right. \end{array} \right.$$

Alors $X(\Omega) = [\![\,1\,;\,4\,]\!]$ et $P(Y=1) = P(Y=2) = P(Y=3) = P(Y=4) = \dfrac{1}{4}$. X et Y ont donc la même loi.

En gros : la loi de X est la loi du gain si on joue au jeu X. Si X et Y ont même loi, on a les mêmes probas de gagner une certaine somme, la loi est tout ce qui nous intéresse, on se fiche du fait que les jeux soient différents ! C'est pour cela qu'on ne s'intéressera qu'à la loi des variables aléatoire, les espaces probabilités sur lesquels elles sont définies n'ont pas un grand intérêt, seule compte leur existence.

Notation : Soient X et Y deux variables aléatoires. Si X et Y ont la même loi, on note $X \sim Y$. C'est évidemment une relation d'équivalence.

Les variables ne sont pas forcément définies sur le même Ω. Comme on l'a déjà dit : Ω ne joue plus qu'un rôle inerte !

Remarque : ⚠ « Avoir la même loi » ne signifie pas « être égales » ! On voit bien avec l'exemple ci-dessus que cela n'a rien à voir : X et Y ont la même loi mais n'ont aucune raison d'être égales. Plus fort : ces fonctions ne sont même pas définies sur le même Ω donc ne peuvent pas être la même fonction ! De plus, ce n'est pas parce qu'elles ont la même loi qu'elles prennent les mêmes valeurs en même temps ! Plus fort : elles peuvent avoir la même loi et ne jamais être égales !

Exemple : Soit X une v.a. dont la loi est donnée par $P(X=0)=1/2=P(X=1)$ et posons $Y=1-X$. Alors $Y(\Omega)=\{0;1\}$,

$$\begin{array}{rcl} P(Y=0) & = & P(1-X=0) \\ & = & P(X=1) \\ & = & 1/2 \end{array}$$

et on a de même (ou parce que la somme fait 1) $P(Y=1)=1/2$. Par conséquent, $X(\Omega)=Y(\Omega), P(X=0)=P(Y=0)$ et $P(X=1)=P(Y=1)$: $X\sim Y$, X et Y ont la même loi. Cependant, quand $X=0$, alors $Y=1$, et quand $X=1$, alors $Y=0$: X et Y ont la même loi mais sont toujours différentes !

Nous dirons plus tard que X suit une loi de Bernoulli de paramètre $1/2$.

Morale de l'histoire : La loi peut être interprétée (quand on est sur $\mathbb{R}$) comme la probabilité du gain (algébrique). Deux variables aléatoires ont la même loi quand « on peut gagner les mêmes sommes avec les mêmes probabilités », ce qui ne veut pas du tout dire qu'on les gagne au même moment !

Utilité : lorsque des variables aléatoires ont la même loi, les probabilités de l'une sont égales aux probabilités correspondantes de l'autre, les espérances et les variances (voir ci-dessous) sont égales, cela évite de faire plusieurs fois les mêmes calculs, et cela permet de ne s'intéresser qu'à une des variables aléatoires. Par exemple :

- Si la loi de X est symétrique par rapport à 0 (i.e. si $P(X=x)=P(X=-x)$ pour tout x), alors X et $-X$ ont la même loi. Cela implique par exemple que $P(X>0)=P(X<0)$ et donc que

$$\begin{array}{rcll} P(X\neq 0) & = & P([X>0]\cup[X<0]) & \text{(union de deux événements incompatibles)} \\ & = & P(X>0)+P(X<0) & \\ & = & 2P(X>0) & \end{array}$$

- Si $X_1,\dots,X_n$ ont la même loi, alors $P(X_1\geqslant 0)+\dots+P(X_n\geqslant 0)=nP(X_1\geqslant 0)$.
- Si $X_1,\dots,X_n$ sont de même loi, et si on pose $S_n=X_1+\dots,X_n$, alors $S_n-S_{n-1}=X_n$ est de même loi que X_1.
- etc.

Si, en plus, on ajoute l'indépendance (cf. paragraphe VI), on peut dire (par exemple) que $S_n-S_{n-2}=X_n+X_{n-1}$ est de même loi que $S_2=X_1+X_2$, etc. et on pourra l'affirmer directement.

I.4 Transfert d'une variable aléatoire

Proposition/Définition. Soit $X:\Omega\to E$ une variable aléatoire et soit $g:E\to F$. Alors :

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

est une variable aléatoire finie notée $g(X)$.

On a aussi $Y=g\circ X$. On dit que Y est un transfert de X par g. Il n'y a rien à prouver : $g(X)$ est une fonction de Ω dans F par définition.

Exemple : Soit X une v.a. de loi donnée par $P(X=-1)=1/4$, $P(X=1)=1/2$ et $P(X=2)=1/4$. Donner les lois de $Y=2X+1, Z=X^2$ et $T=e^X$.

- $Y(\Omega)=\{-1;3;5\}$ et :

$$\begin{array}{rcl} P(Y=-1) & = & P(2X+1=-1) \\ & = & P(X=-1) \\ & = & 1/4 \end{array}$$

On trouve de même $P(Y=3)=1/2$ $P(Y=5)=1/4$.

- $Z(\Omega) = \{-1;4\}$ et :

$$\begin{array}{rcl} P(Z=1) & = & P(X^2=1) \\ & = & P([X=1]\cup[X=-1]) \\ & = & P(X=1)+P(X=-1) \\ & = & 3/4 \end{array}$$

Union de deux événements incompatibles.

  De plus, X étant à valeurs dans $\{-1;1;2\}$ (plus précisément puisque X ne peut pas prendre la valeur -2), $P(Z=4) = P(X^2=4) = P(X=2) = 1/4$ (on aurait aussi pu faire $1 - P(Z=1)$).
- Enfin, $T(\Omega) = \{e^{-1}; e; e^2\}$ si bien que

$$\begin{array}{rcl} P(T=e^{-1}) & = & P(e^X = e^{-1}) \\ & = & P(X=-1) \\ & = & 1/4 \end{array}$$

  etc.

Remarque : Ces égalités sont toutes assez intuitives, mais pourquoi sont-elles vraies ? Par exemple, pourquoi a-t-on $P\left(e^X = e^{-1}\right) = P(X=-1)$?

Montrons que $\left[e^X = e^{-1}\right] = [X=-1]$ (n'oublions pas que ces deux **ensembles** sont des parties de Ω). Soit $\omega \in \Omega$.

$$\begin{array}{rcl} \omega \in \left[e^X = e^{-1}\right] & \iff & e^{X(\omega)} = e^{-1} \\ & \iff & X(\omega) = -1 \\ & \iff & \omega \in [X=-1] \end{array}$$

Par conséquent, $\left[e^X = e^{-1}\right] = [X=-1]$. En particulier, ces deux ensembles ont la même probabilité.

Montrons de même que $\left[e^{-X} \leqslant 2\right] = [X \geqslant -\ln(2)]$. Soit $\omega \in \Omega$.

$$\begin{array}{rcll} \omega \in \left[e^{-X} \leqslant 2\right] & \iff & e^{-X(\omega)} \leqslant 2 & \\ & \iff & -X(\omega) \leqslant \ln(2) & \text{(car la fonction ln est strictement croissante)} \\ & \iff & X(\omega) \geqslant -\ln(2) & \\ & \iff & \omega \in [X \geqslant -\ln(2)] & \end{array}$$

et on conclut de la même façon (l'argument clef est ici la stricte croissance de la fonction ln). En pratique, on peut affirmer tout cela directement, mais s'il y a un doute, on le montre de la même façon que ci-dessus. Donnons un dernier exemple pour la route, un peu plus difficile.

Montrons que si $f : \mathbb{R} \to \mathbb{R}$ est croissante, alors $P(X \geqslant 1) \leqslant P(f(X) \geqslant f(1))$. On ne cherche plus à montrer une égalité ici, donc on ne va pas travailler par équivalences (et si on essaye tout de même, ça coince car f n'est pas strictement croissante). On va non plus montrer que les ensembles $[X \geqslant 1]$ et $[f(X) \geqslant f(1)]$ sont égaux, mais que le premier est inclus dans le deuxième, ce qui permettra de conclure. Soit $\omega \in \Omega$. Supposons que $\omega \in [X \geqslant 1]$. Alors $X(\omega) \geqslant 1$ donc, f étant croissante (et ici, la réciproque serait fausse, c'est pour cela qu'on ne peut pas raisonner par équivalences), $f(X(\omega)) \geqslant f(1)$ donc $\omega \in [f(X) \geqslant f(1)]$. En d'autres termes, $[X \geqslant 1] \subset [f(X) \geqslant f(1)]$ ce qui permet de conclure.

Nous utiliserons ce résultat dans le paragraphe VII.

Proposition. Soient X et Y deux variables aléatoires à valeurs dans E. Soit f une fonction définie sur E. Si $X \sim Y$ alors $f(X) \sim f(Y)$, c'est-à-dire que si X et Y ont la même loi, alors $f(X)$ et $f(Y)$ ont toujours la même loi.

Démonstration. Notons $X(\Omega) = Y(\Omega) = \{e_1; \dots; e_n\}$ si bien que $f(X(\Omega)) = f(Y(\Omega)) = \{f(e_1); \dots; f(e_n)\}$ (certains de ces éléments pouvant être égaux si f n'est pas injective). Quitte à supprimer les éléments égaux, on suppose que tous les $f(e_i)$ sont distincts. Soit $i \in [\![1 ; n]\!]$ et notons $y_i = f(e_i)$ et $A_i = f^{-1}(\{y_i\})$ l'ensemble des antécédents de y_i par f (attention, f n'est pas forcément injective). Alors $f(X) = y_i$ si et seulement si $X \in A_i$, ce qui donne :

$$\begin{aligned}
P(f(X) = y_i) &= P(X \in A_i) \\
&= P\left(X \in \bigcup_{x_k \in X(\Omega)\cap A_i} \{x_k\}\right) \\
&= \sum_{x_k \in X(\Omega)\cap A_i} P(X = x_k) \\
&= \sum_{x_k \in X(\Omega)\cap A_i} P(Y = x_k) \\
&= P(f(Y) = y_i)
\end{aligned}$$

□

Union disjointe.

$X \sim Y$

I.5 Loi conditionnelle

Proposition/Définition. Soit B un événement avec $P(B) \neq 0$. On appelle loi conditionnelle de X sachant que l'événement B est réalisé la loi de X pour la probabilité P_B, c'est-à-dire l'application

$$\begin{cases} \mathscr{P}(X(\Omega)) & \to & [0;1] \\ A & \mapsto & P_B(X \in A) = P_B(\{\omega \in \Omega \mid X(\omega) \in A\}) \end{cases}$$

Rappelons que pour conditionner par un événement, il est indispensable que celui-ci ait une probabilité non nulle. De plus, d'après le chapitre 26, P_B est bien une probabilité sur $(\Omega, \mathscr{P}(\Omega))$.

Remarques :

- Comme dans le chapitre précédent, apporter une information (B est réalisé) change parfois les probabilités, et donc, ici, la loi de la variable aléatoire X. La nouvelle loi connaissant cette information est appelée loi conditionnelle sachant que l'événement B est réalisé.
- En pratique, pour donner une loi conditionnelle, on donne comme pour une loi normale, l'univers image et les probabilites conditionnelles de chaque élément de l'univers image.

Ou, comme pour une loi normale, on donne la réponse sous la forme d'une loi de référence. cf. paragraphe IV.2.b pour un exemple.

II Espérance d'une variable aléatoire.

II.1 Introduction et définition

Prenons l'exemple du dé non équilibré ci-dessus, où la probabilité d'obtenir un numéro est proportionnelle à ce numéro : si on note X le résultat obtenu, on se dit qu'après un grand nombre N de répétitions, la proportion f_1 de coups où le résultat vaut 1 sera environ égale à $1/21$, la proportion f_2 de coups où le résultat vaut 2 sera environ égale à $2/21$, etc. La somme de toutes les faces obtenues égale à

$$S = N \times f_1 \times 1 + N \times f_2 \times 2 + N \times f_3 \times 3 + N \times f_4 \times 4 + N \times f_5 \times 5 + N \times f_6 \times 6$$

sera donc estimée par :

$$S \approx N \times \frac{1}{21} \times 1 + N \times \frac{2}{21} \times 2 + N \times \frac{3}{21} \times 3 + N \times \frac{4}{21} \times 4 + N \times \frac{5}{21} \times 5 + N \times \frac{6}{21} \times f_6$$

et donc la valeur moyenne, obtenue en divisant par N, est estimée par :

$$\frac{1}{21}\times 1+\frac{2}{21}\times 2+\frac{3}{21}\times 3+\frac{4}{21}\times 4+\frac{5}{21}\times 5+\frac{6}{21}\times f_6=\sum_{k=1}^{6} kP(X=k)$$

Une fois généralisé, ce raisonnement heuristique conduit naturellement à la définition suivante.

> On peut définir de la même façon l'espérance d'une v.a. complexe, et les propriétés se généralisent facilement (sauf celles concernant la positivité, naturellement). Mais attention, une v.a. quelconque à valeurs dans un ensemble E n'admet pas forcément d'espérance car la somme ci-contre n'est pas forcément définie.

Définition. Soit X une variable aléatoire réelle. On appelle espérance de X le réel :

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

On dit que X est centrée si $E(X)=0$.

Remarques :

- En posant $X(\Omega)=\{x_1;\dots;x_n\}$, on a :

  $$E(X)=\sum_{i=1}^{n} x_i\,P(X=x_i)$$

  En particulier, si $X(\Omega)=[\![1\,;\,n]\!]$:

  $$E(X)=\sum_{k=1}^{n} kP(X=k)$$

> C'est même valable si $X(\Omega)=[\![0\,;\,n]\!]$ puisque le terme $kP(X=k)$ est alors nul si $k=0$.

- $E(X)$ n'est rien d'autre que la moyenne pondérée. On somme chaque valeur possible pour X pondérée par la proba que X prenne cette valeur.
- Si deux v.a. X et Y ont la même loi, alors elles prennent les mêmes valeurs avec les mêmes probas donc ont la même espérance. Ce sera la même chose avec la variance (cf. paragraphe III).

> Évidemment, réciproque fausse : par exemple, une loi certaine égale à 1/2 et une loi de Bernoulli de paramètre 1/2 (cf. IV) ont la même espérance mais n'ont pas la même loi.

On se donne dans la suite de cette partie et dans la suivante une variable aléatoire X à valeurs réelles.

II.2 Premiers exemples

Exemple : Si X est la v.a. donnant le résultat obtenu lors du lancer d'un dé équilibré. Alors $X(\Omega)=[\![1\,;\,6]\!]$, $P(X=k)=1/6$ pour tout $k\in[\![1\,;\,6]\!]$.

$$\begin{aligned} E(X) &= 1\times P(X=1)+2\times P(X=2)+3\times P(X=3)+4\times P(X=4)+5\times P(X=5)+6\times P(X=6)\\ &= \frac{1}{6}(1+2+3+4+5+6)\\ &= \frac{7}{2}\end{aligned}$$

> Nous dirons dans le paragraphe IV.2 que X suit une loi uniforme sur $[\![1\,;\,6]\!]$.

Exemple : Même chose mais cette fois avec le dé qui donne k avec proba $k/21$.

$$\begin{aligned} E(X) &= \sum_{k=1}^{6} kP(X=k)\\ &= \sum_{k=1}^{6}\frac{k^2}{21}\\ &= \frac{6\times 7\times 13}{6\times 21}\\ &= \frac{13}{3}\end{aligned}$$

Exemple : On lance une pièce équilibrée, on gagne 1 si on tombe sur Pile, on perd 1 si on tombe sur Face. On note X le gain algébrique. Alors $X(\Omega) = \{\pm 1\}$ et $P(X=1) = P(X=-1) = 1/2$ si bien que

$$E(X) = 1 \times P(X=1) - 1 \times P(X=-1) = 0$$

Cette variable est donc centrée.

Exemple : Si on note X la somme de deux dés équilibrés :

$$E(X) = \sum_{k=1}^{12} kP(X=k)$$

Après calcul (pas de mystère : on calcule une somme de 12 termes, pas de formule pour aller plus vite), on trouve que $E(X) = 7$.

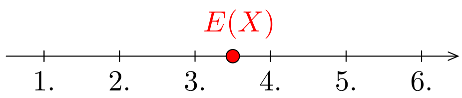
II.3 Interprétation

L'espérance représente « la valeur moyenne » de X, c'est-à-dire que (en gros) si on joue n fois, on gagne en moyenne n fois l'espérance de X (quand n est **très grand**).

Cela permet de mieux « situer », de mieux se faire une première idée (insuffisante !) de X : on dit parfois que l'espérance est un indicateur de position.

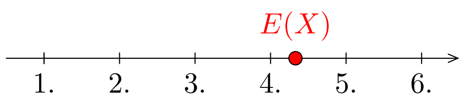
Remarque : Attention, $E(X)$ n'est pas la valeur que X prend avec la plus grande probabilité. $E(X)$ peut même être une valeur que X ne prend jamais : cf exemples précédents.

Reprenons l'exemple d'un lancer de dé équilibré :



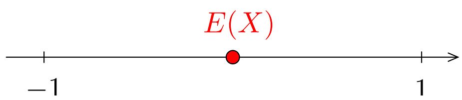
L'espérance est au milieu, ce qui est cohérent puisque le dé est équilibré.

Pour l'exemple du dé non équilibré :



L'espérance est décalée vers la droite, ce qui est cohérent puisque les gros numéros sortent de façon plus probable que les petits.

Pour l'exemple de la pièce équilibrée :



Remarque : Historiquement, la notion d'espérance a permis de définir un jeu équitable, c'est-à-dire un jeu pour lequel l'espérance du gain est nulle, par exemple le jeu de l'exemple de la pièce équilibrée. Si l'espérance est strictement positive, on est « avantagé », et on est « désavantagé » si elle est strictement négative.

Remarque : Attention, cet avantage ou ce désavantage ne sont visible que si on fait un très grand nombre de parties ! Ce n'est pas parce que l'espérance est positive que l'on doit être incité à jouer !

Exemple : On considère un jeu où on perd 10000 euros avec probabilité $\dfrac{9999}{10000}$ et où on gagne 10^8 euros avec probabilité $\dfrac{1}{10000}$ (ce qui représente environ la probabilité de faire 13 « Pile » de suite). Soit X le gain (algébrique). Sa loi est donnée par

$$P(X=-10000)=\frac{9999}{10000} \qquad \text{et} \qquad P\left(X=10^8\right)=\frac{1}{10000}$$

On a donc :

$$\begin{aligned} E(X) &= -10000\times P(X=-10000)+10^8\times P(X=10^8) \\ &= -10000\times\frac{9999}{10000}+10^8\times\frac{1}{10000} \\ &= -9999+10^4 \\ &= 1 \end{aligned}$$

L'espérance est donc strictement positive. Cependant, si on fait un tirage, on a « toutes les chances de perdre ». Ce n'est qu'avec un grand nombre de tirages que les pertes successives seront compensées par les rares gains. Mais on sera sûrement ruiné avant.

II.4 Propriétés de l'espérance

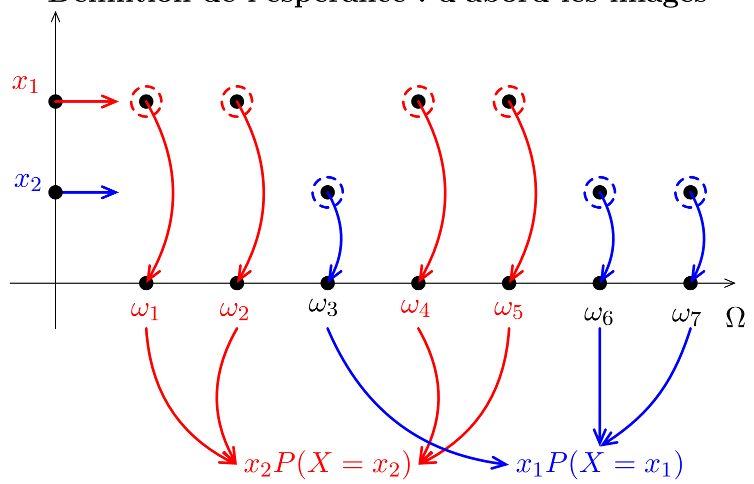
Lemme. $E(X)=\displaystyle\sum_{\omega\in\Omega} X(\omega)P(\{\omega\})$.

Remarque : Ce lemme est totalement évident : l'espérance est la somme des images (i.e. des éléments x de $X(\Omega)$) multipliées par la probabilité de leur ensemble d'antécédents (i.e. $P(X=x)$), tandis que, pour calculer la somme de droite, on part plutôt de chaque antécédent (au lieu de partir des images) $\omega\in\Omega$, dont on multiplie la probabilité par la valeur de son image (i.e. $X(\omega)$).

Dans les deux cas, on fait la même chose : on multiplie les images par la proba des éventualités correspondantes (i.e. les probas des antécédents par la valeur des images) et ensuite on somme, mais dans la définition de l'espérance, on prend plutôt le parti des images (premier dessin ci-dessous) tandis que dans la somme de droite, on prend plutôt le parti des antécédents (deuxième dessin ci-dessous). La seule différence est que, pour l'espérance, on part des images et on prend l'ensemble de ses antécédents : on fait un regroupement par paquets (au lieu de prendre chaque éventualité séparément, comme dans la somme de droite). C'est d'ailleurs l'idée de la preuve ci-dessous.

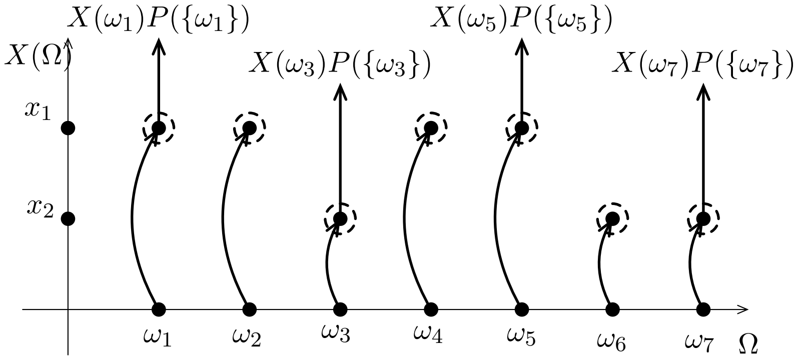
On peut voir cela comme une somme double : sommer d'abord selon les lignes ou les colonnes (voir les dessins ci-dessous).

Définition de l'espérance : d'abord les images



Deuxième écriture de l'espérance : d'abord les antécédents

On ne trace pas toutes les flèches pour des raisons de lisibilité



Démonstration. L'idée est de partir de la somme de droite et de faire un regroupement par paquets en regroupant les ω qui ont la même image. Notons $X(\Omega) = \{x_1; \dots; x_n\}$ si bien que :

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Puisque les $[X = x_i]$ forment un système complet d'événements, Ω est l'union disjointe des $[X = x_i]$. Par conséquent (on utilise le théorème de sommation par paquets vu au chapitre 3) :

$$\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{i=1}^{n} \left(\sum_{\omega \in [X = x_i]} X(\omega) P(\{\omega\}) \right)$$

Or, par hypothèse, tous les termes $X(\omega)$ de la deuxième somme sont égaux à x_i, terme qui ne dépend pas de « l'indice de sommation » ω et donc on peut le sortir de la somme :

$$\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{i=1}^{n} x_i \left(\sum_{\omega \in [X = x_i]} P(\{\omega\}) \right)$$

Par conséquent (union disjointe)

$$\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{i=1}^{n} x_i P \left(\bigcup_{\omega \in [X = x_i]} \{\omega\} \right)$$

Or, la probabilité de l'union des éléments d'un ensemble étant la probabilité de cet ensemble, on a :

$$\sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{i=1}^{n} x_i P(X = x_i)$$

□

ce qui est le résultat voulu.

Proposition. Soient X et Y deux variables aléatoires définies sur le même univers Ω et à valeurs réelles.

- **Linéarité de l'espérance :** Pour tous réels λ et μ, $E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$.
- **Positivité de l'espérance :** Si X est à valeurs positives, alors $E(X) \geqslant 0$, avec égalité si et seulement si $P(X = 0) = 1$ (c'est-à-dire si $X = 0$ presque sûrement, cf. chapitre 26).
- **Croissance de l'espérance :** Si $X \geqslant Y$ alors $E(X) \geqslant E(Y)$, avec égalité si et seulement si $P(X = Y) = 1$.

La linéarité de l'espérance et l'inégalité triangulaire sont encore valables pour des variables aléatoires à valeurs complexes, en prenant λ et μ complexes pour la linéarité, et le module au lieu de la valeur absolue pour l'inégalité triangulaire.

- **Inégalité triangulaire :** $|E(X)| \leqslant E(|X|)$.

DÉMONSTRATION. • Soit $(\lambda, \mu) \in \mathbb{R}^2$. D'après le lemme précédent :

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \qquad \text{et} \qquad E(Y) = \sum_{\omega \in \Omega} Y(\omega) P(\{\omega\})$$

Par conséquent :

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

- Par hypothèse sur X, $X(\Omega) \subset \mathbb{R}^+$, et comme les probabilités sont des réels positifs, $E(X)$ est une somme de termes positifs donc est positive. De plus, celle-ci est nulle si et seulement si tous les termes sont nuls (encore une fois car c'est une somme de termes positifs). Ainsi :

$$\begin{aligned} E(X) = 0 &\iff \forall x \in X(\Omega), x P(X = x) = 0 \\ &\iff \forall x \neq 0, P(X = x) = 0 \\ &\iff P(X \neq 0) = 0 \\ &\iff P(X = 0) = 1 \end{aligned}$$

Si X n'est pas positive, on n'a pas forcément $E(X) \geqslant 0$, et le cas d'égalité n'est plus valable (prendre l'exemple ci-dessus, où $P(X = 1) = P(X = -1) = 0$.)

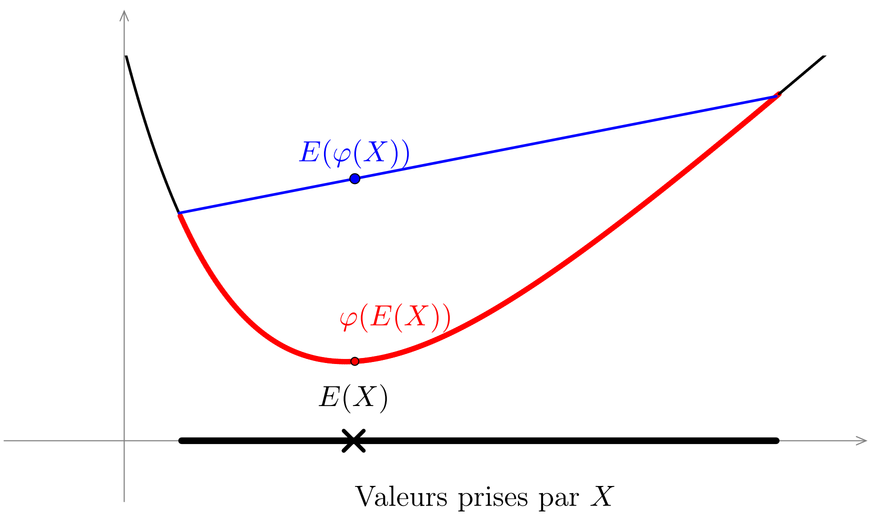
- Il suffit d'appliquer le résultat de positivité à $Z = X - Y$ ainsi que la linéarité de l'espérance.
- Toujours d'après le lemme, par inégalité triangulaire (sur $\mathbb{R}$) et puisqu'une probabilité est positive :

$$|E(X)| = \left| \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) \right| \leqslant \sum_{\omega \in \Omega} |X(\omega)| P(\{\omega\}) = E(|X|)$$

$\square$

Il n'y a pas forcément égalité. Par exemple, si X est la variable aléatoire définie par $P(X = 1) = P(X = -1) = 1/2$ (cf. paragraphe I.3.b), alors $E(X) = 0$ et $E(|X|) = 1$.

Remarque : L'inégalité triangulaire est un cas particulier de ce qu'on appelle l'inégalité de Jensen (pour l'espérance) : si φ est convexe, alors $\varphi(E(X)) \leqslant E(\varphi(X))$ (cf. exercice 14). Cela se voit bien sur un dessin : si φ est convexe, l'image de la moyenne (avec tous les dangers que ce raccourci implique, cf. paragraphe II.3) est inférieure à la moyenne des images.



Corollaire (Linéarité de l'espérance). Soient $(X_1, \dots, X_n)$ des v.a. définies sur le même Ω à valeurs réelles, et $\lambda_1, \dots, \lambda_n$ des réels. Alors :

$$E\left(\sum_{k=1}^{n} \lambda_k X_k\right) = \sum_{k=1}^{n} \lambda_k E(X_k)$$

Ce n'est pas vrai avec un produit ! On n'a pas forcément $E(XY) = E(X) \times E(Y)$! Il faut une condition supplémentaire, cf. paragraphe VI.4.

DÉMONSTRATION. Par récurrence : $\rightsquigarrow$ EXERCICE.

II.5 Théorème de transfert

Lemme. Soit $(a, b) \in \mathbb{R}^2$. Alors : $E(aX + b) = aE(X) + b$.

DÉMONSTRATION. Il suffit de prouver que, si on note Y la variable aléatoire constante égale à b, alors $E(Y) = b$. Le résultat découlera alors immédiatement de la linéarité de l'espérance. Or, si Y est constante égale à b, alors $Y(\Omega) = \{b\}$ et $P(Y = b) = 1$ donc $E(Y) = bP(Y = b) = b$ ce qui permet de conclure.

Nous dirons dans le paragraphe IV.1 que Y suit une loi certaine.

Remarques :

- Comme ci-dessus, on assimilera souvent une constante b à la variable aléatoire constante égale à b. On écrira donc souvent des égalités du type $E(b) = b$.
- Morale de l'histoire : la linéarité de l'espérance est valable pour les constantes additives aussi bien que pour les constantes multiplicatives. Par exemple : $E(X + 1) = E(X) + 1$. Attention cependant : comme on l'a déjà dit, autant on peut affirmer que $E(X + Y) = E(X) + E(Y)$, même quand Y n'est pas constante, par linéarité de l'espérance, autant l'égalité $E(X \times Y) = E(X) \times E(Y)$ est fausse en général lorsque Y n'est pas constante ! cf. paragraphe VI.4.

Corollaire (Variable centrée associée à une v.a. X). La variable aléatoire $X - E(X)$ est centrée et est appelée variable centrée associée à X.

DÉMONSTRATION. Par linéarité de l'espérance, $E(X - \underbrace{E(X)}_{=b}) = E(X) - E(X) = 0$.

Même s'il y a du X dans $E(X)$, il ne faut pas perdre de vue que $E(X)$ est un **réel** qui n'a rien d'aléatoire, contrairement à X ! Par conséquent, on peut le « sortir de l'espérance » par linéarité, comme n'importe quel réel. Ce sera la même chose avec la formule de König-Huygens dans le paragraphe suivant.

Remarque : Le lemme précédent pourrait laisser croire que, si on a une fonction f définie sur $X(\Omega)$, alors $E(f(X)) = f(E(X))$. Ce résultat est faux en général : par exemple, pour la valeur absolue, on dispose de l'inégalité triangulaire (cf. paragraphe II.4) mais on n'a pas une égalité en général. Cependant, il est possible de calculer $E(f(X))$ « en restant sur E sans passer par Ω » grâce au théorème fondamental suivant.

Théorème (Théorème de transfert). Soient $X : \Omega \to E$ une v.a. et $g : X(\Omega) \to \mathbb{R}$. Alors :

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

DÉMONSTRATION. La démonstration est la même que celle du lemme de la partie précédente. Notons $X(\Omega) = \{x_1; \dots; x_n\}$. Ω est donc l'union disjointe des $[X = x_i]$, si bien que :

$$\begin{aligned} E(g(X)) &= \sum_{\omega\in\Omega} g(X(\omega))P(\{\omega\}) \\ &= \sum_{i=1}^{n}\left(\sum_{\omega\in[X=x_i]} g(X(\omega))P(\{\omega\})\right) \\ &= \sum_{i=1}^{n} g(x_i)\left(\sum_{\omega\in[X=x_i]} P(\{\omega\})\right) \\ &= \sum_{i=1}^{n} g(x_i)P\left(\bigcup_{\omega\in[X=x_i]}\{\omega\}\right) \\ &= \sum_{i=1}^{n} g(x_i)P(X=x_i) \end{aligned}$$

$\square$

Le théorème de transfert (aussi appelé formule de transfert) permet de calculer l'espérance de $g(X)$ uniquement à l'aide de la loi de X (et pas de la loi de $g(X)$, qui peut être difficile à calculer). En clair : on applique g aux valeurs prises par x mais pas aux probabilités ! Précisons (cela sera utile pour calculer l'espérance d'un produit de deux variables aléatoires indépendantes, cf. paragraphe VI.4) que X n'a pas besoin d'être à valeurs réelles (mais g, oui).

Exemples : Soit X une v.a. à valeurs dans $\mathbb{R}_+^*$. Notons $X(\Omega)=\{x_1;\dots;x_n\}$. Alors :

- $E\left(\dfrac{1}{X}\right)=\displaystyle\sum_{k=1}^{n}\frac{1}{x_k}\times P(X=x_k)$.
- $E\left(X^2\right)=\displaystyle\sum_{k=1}^{n}{x_k}^2\times P(X=x_k)$.
- $E\left(e^X\right)=\displaystyle\sum_{k=1}^{n}e^{x_k}\times P(X=x_k)$.
- etc.

Comme on vient de le dire : on applique g aux valeurs prises par X, on ne touche pas aux probabilités.

Remarque : Attention, comme on l'a déjà dit, en général, $E(g(X))\neq g(E(X))$! En particulier, $E\left(X^2\right)\neq E(X)^2$.

III Variance d'une v.a

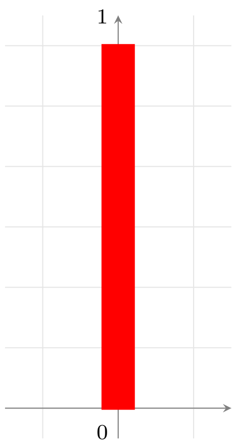
Dans cette partie également, X désigne une variable aléatoire à valeurs réelles.

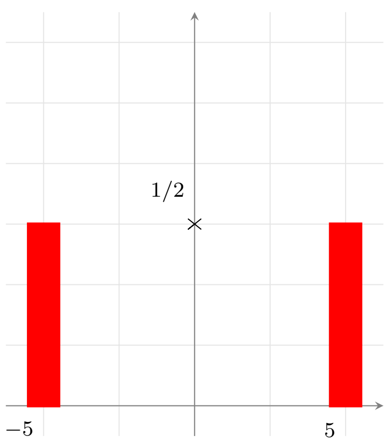
Cette fois, la définition change pour une v.a. complexe, mais c'est HP.

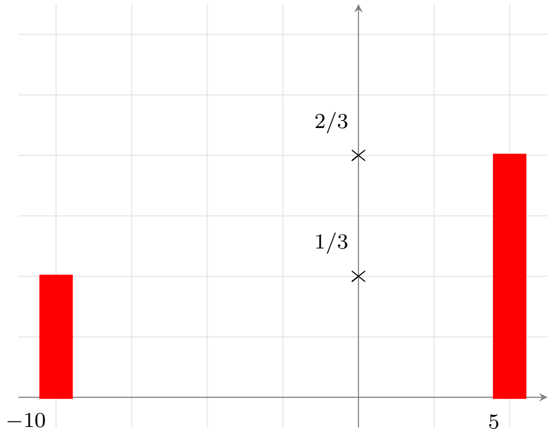
III.1 Introduction

La donnée de la valeur moyenne d'une v.a. X ne permet pas en général de se faire une idée de la loi de X.

Une v.a. centrée (c'est-à-dire d'espérance nulle) peut être très concentrée au voisinage de 0 ou tout aussi bien être concentrée au voisinage des points 5 et -10 ou des points -10 et 10 :







Il est donc nécessaire de définir un nombre qui mesure l'écart entre une v.a. X et son espérance. On peut dans un premier temps songer à prendre la moyenne de $X - E(X)$, c'est-à-dire $E(X - E(X))$. Cependant, par linéarité de l'espérance (cf. paragraphe II.4), $E(X - E(X)) = E(X) - E(X) = 0$. Ceci est dû au fait que, dans l'espérance ci-dessus, les valeurs positives et négatives se compensent.

On peut ensuite penser à $E(|X - E(X)|)$ pour éviter les compensations. D'après le théorème de transfert,

$$E(|X - E(X)|) = \sum_{x \in X(\Omega)} |x - E(X)| \times P(X = x)$$

et cette quantité est difficile à calculer : les valeurs absolues et les sommes ne font pas bon ménage ! On va plutôt prendre la moyenne de $(X - E(X))^2$.

Cela nous permettra de plus d'interpréter l'écart-type en terme de distance.

III.2 Définition

Définition. On appelle variance de X le réel :

$$V(X) = E\big((X - E(X))^2\big).$$

Remarque : La variance mesure la dispersion d'une variable aléatoire par rapport à son espérance, c'est-à-dire la manière dont la variable aléatoire se répartit autour de son espérance. De manière intuitive, quand la variance est petite, la v.a. est faiblement dispersée : elle a peu de chances de prendre une valeur éloignée de sa valeur moyenne, donnée par $E(X)$. En particulier, une seule observation de X va sans doute donner un résultat proche de $E(X)$ et permet d'estimer économiquement cette espérance, tandis qu'une variable avec une grande variance prend souvent des valeurs éloignées de son espérance, et une unique observation a peu de chance de fournir une bonne estimation de l'espérance.

On dit pour cela que c'est un indicateur de dispersion (alors qu'on dit plutôt, cf. paragraphe II.3, que l'espérance est un indicateur de position).

Proposition. $V(X) \geqslant 0$ avec égalité si et seulement si $X = E(X)$ presque sûrement, i.e. $P(X = E(X)) = 1$.

Démonstration. $(X - E(X))^2$ est une v.a. positive donc, par positivité de l'espérance, son espérance est positive et est nulle si et seulement si $P\left((X - E(X))^2\right) = 1$ ce qui permet de conclure.

Remarque : X est dite alors constante presque sûrement, ou certaine, cf IV.1. Le résultat précédent est totalement intuitif : la variance servant à mesurer la dispersion d'une variable aléatoire par rapport à son espérance, il est intuitif que cette dispersion est nulle si et seulement si la variable aléatoire et son espérance sont égales (avec probabilité 1).

Définition. On appelle écart-type de X le réel $\sigma(X) = \sqrt{V(X)}$.

Remarque : L'écart-type est bien défini puisque la variance est positive. Il peut être interprété comme une distance entre X et $E(X)$: cf TD. Dans ce cas, $V(X)$ est la distance au carré : ne pas dire que $V(X)$ est la distance même si la variance sert à mesurer la dispersion à l'espérance. Votre professeur de physique préféré dirait que la variance et l'espérance ne sont pas homogènes (il y a un carré) mais que l'écart-type et l'espérance le sont.

III.3 Calcul pratique

Théorème ((formule de König-Huygens)). Alors $V(X) = E\left(X^2\right) - E(X)^2$.

DÉMONSTRATION. $V(X) = E\left[X^2 - 2E(X)X + E(X)^2\right]$. Par linéarité de l'espérance

$$\begin{aligned} V(X) &= E(X^2) - 2E(X) \times E(X) + E(X)^2 \\ &= E(X)^2 - E(X)^2 \end{aligned}$$

□

$E(X)$ et $E(X)^2$ sont des **réels** (i.e. ne sont pas aléatoires) ! De plus, on sait que si b est un réel, alors $E(b) = b$.

Remarque : Pour faire simple : il faut systématiquement utiliser cette formule pour calculer la variance. On calcule dans un premier temps l'espérance, puis $E(X^2)$ (à l'aide du théorème de transfert) puis on applique la formule de König-Huygens.

Exemple : Notons X la variable aléatoire égale au résultat obtenu lors du lancer d'un dé équilibré. On sait déjà que $E(X) = 7/2$. D'après la formule de transfert :

$$\begin{aligned} E(X^2) &= \sum_{k=1}^{6} k^2 P(X = k) \\ &= \frac{1}{6}\sum_{k=1}^{6} k^2 \\ &= \frac{6 \times 7 \times 13}{6^2} \\ &= \frac{91}{6} \end{aligned}$$

D'après la formule de König-Huygens :

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 \\ &= \frac{91}{6} - \frac{49}{4} \\ &= \frac{35}{12} \end{aligned}$$

Exemple : Si X est la v.a. définie par $P(X = 1) = P(X = -1) = 1/2$, X^2 est constante égale à 1 donc $E(X^2) = 1$ si bien que $V(X) = E(X^2) - E(X)^2 = 1 - 0 = 1$.

Proposition. Soit $(a, b) \in \mathbb{R}^2$. Alors $V(aX + b) = a^2 V(X)$ et $\sigma(aX + b) = |a|\sigma(X)$.

Remarque : ⚠ La variance n'est pas linéaire ! De plus, en général, $V(X+Y) \neq V(X) + V(Y)$!

DÉMONSTRATION. D'après la formule de König-Huygens :

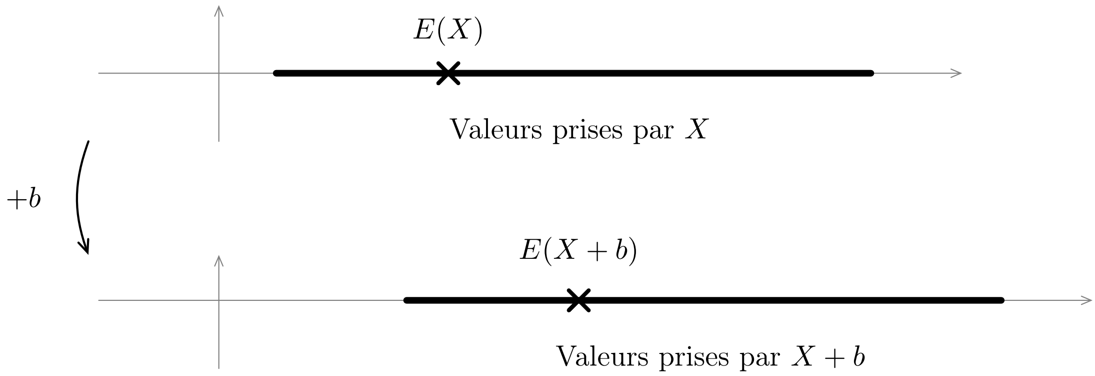
$$
\begin{array}{rcl}
V(aX+b) & = & E\left[(aX+b)^2\right] - (E(aX+b))^2 \\
& = & E\left[a^2X^2 + 2abX + b^2\right] - (E(aX+b))^2 \\
& = & a^2E(X^2) + 2abE(X) + b^2 - (aE(X)+b)^2 \\
& = & a^2E(X^2) + 2abE(X) + b^2 - a^2E(X)^2 - 2abE(X) - b^2 \\
& = & a^2(E(X^2) - E(X)^2) \\
& = & a^2V(X)
\end{array}
$$
□

Par linéarité de l'espérance.

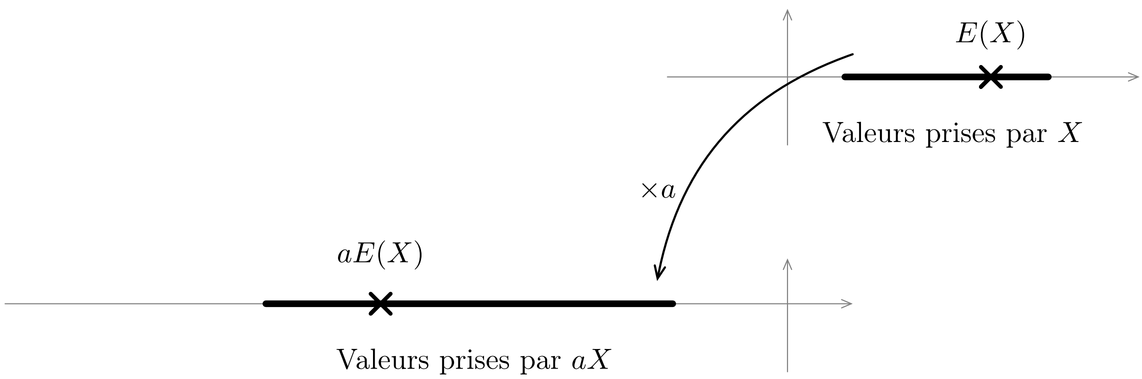
Il suffit ensuite de prendre la racine carrée pour obtenir l'écart-type.

Question : Où passe le b ?

- Quand on translate, l'écart à l'espérance reste le même, donc b n'a aucune influence sur la variance.



- Par contre, quand on multiplie par a, on « dilate » et l'écart-type est multiplié par $|a|$.



Définition. X est dite réduite si $V(X) = 1$.

Proposition/Définition. Si $\sigma(X) \neq 0$, la variable aléatoire $X^* = \dfrac{X - E(X)}{\sigma(X)}$ est centrée réduite et est appelée variable centrée réduite associée à X.

Il faut diviser par l'écart-type et pas par la variance !

DÉMONSTRATION. $\rightsquigarrow$ EXERCICE.

IV Lois usuelles

IV.1 Loi certaine

Définition. On dit qu'une variable aléatoire X suit une loi certaine s'il existe $m \in E$ tel que :

- $X(\Omega) = \{m\}$.
- $P(X = m) = 1$.

Rappelons que la loi d'une variable aléatoire est entièrement déterminée par l'univers image et la distribution de probabilités $(P(X = e))_{e \in X(\Omega)}$.

Exemple : Si on a un dé pipé qui donne toujours 4, alors X, le nombre donné par le dé, suit une loi certaine égale à 4.

Remarques :

- On peut aussi avoir $X(\Omega) = \{x_1; \dots; x_n\}$ et $P(X = x_1) = 1, P(X = x_2) = \cdots = P(X = x_n) = 0$. Tout dépend des conventions, en général on ne garde dans $X(\Omega)$ que les éléments dont les probas sont non nulles. En tout cas, la propriété fondamentale est : « $\exists m, P(X = m) = 1$ ».
- On parle parfois de v.a. presque certaine ou de v.a. presque sûrement égale à m.

Proposition. Si X est certaine égale à m, alors $E(X) = m$ et $V(X) = 0$.

DÉMONSTRATION. $\rightsquigarrow$ EXERCICE.

Remarques :

- Ainsi (mais on l'a déjà vu plus haut), $E(m) = m$.
- On a déjà vu la réciproque : si $V(X) = 0$, alors X suit une loi certaine.

IV.2 Loi uniforme

IV.2.a Cadre

La loi uniforme sur un ensemble E est l'équiprobabilité sur E.

En d'autres termes, la probabilité de n'importe quelle partie A de E est égale à $\operatorname{card}(A)/\operatorname{card}(E)$.

IV.2.b Définition et notation

Définition. Soit E un ensemble fini non vide. On dit qu'une variable aléatoire X suit une loi uniforme sur E si :

- $X(\Omega) = E$.
- $\forall e \in E, P(X = e) = \dfrac{1}{\operatorname{card}(E)}$.

On note alors $X \sim U(E)$.

Remarque : En d'autres termes, une variable aléatoire suit une loi uniforme sur un ensemble **fini** lorsque tous les éléments de l'ensembles sont équiprobables i.e. ont la même probabilité. Cette probabilité commune est alors égale à $1/\operatorname{card}(E)$ puisque la somme des probabilités est égale à 1.

Exemple : Soit $X \sim U(\llbracket 1\,;\,2n\,\rrbracket)$. Donner la loi conditionnelle de X sachant que l'événement $[X \leqslant n]$ est réalisé.

Supposons donc l'événement $[X \leqslant n]$ réalisé. Soit $k \in \llbracket 1\,;\,2n\,\rrbracket$. Si $k \geqslant n+1$, $P_{[X \leqslant n]}(X = k) = 0$. Supposons à présent $k \leqslant n$. Alors :

Rédaction ! cf. chapitre précédent.

$$\begin{aligned} P_{[X \leqslant n]}(X = k) &= \frac{P([X \leqslant n] \cap [X = k])}{P(X \leqslant n)} \\ &= \frac{P(X = k)}{P(X \leqslant n)} \end{aligned}$$

puisque $[X = k] \subset [X \leqslant n]$. Or, $P(X = k) = 1/2n$ et

$$\begin{aligned} P(X \leqslant n) &= \frac{\mathrm{card}(\llbracket 1\,;\,n\,\rrbracket)}{\mathrm{card}(\llbracket 1\,;\,2n\,\rrbracket)} \\ &= \frac{n}{2n} \\ &= \frac{1}{2} \end{aligned}$$

si bien que la probabilité recherchée vaut $1/n$: la loi de X conditionnellement à la réalisation de l'événement $[X \leqslant n]$ est la loi uniforme sur $\llbracket 1\,;\,n\,\rrbracket$. C'est intuitif ! Si X suit une loi uniforme sur $\llbracket 1\,;\,2n\,\rrbracket$, X peut prendre n'importe quelle valeur entre 1 et $2n$ avec la même probabilité, et si on ajoute l'information que $X \leqslant n$, alors X ne prendra que des valeurs entre 1 et n mais aucune n'est plus probable qu'une autre : on garde une loi uniforme.

Comme pour la loi certaine, on peut « retirer de l'univers image » les éléments ayant une probabilité nulle : on peut donc « faire comme si » l'univers image était égal à $\llbracket 1\,;\,n\,\rrbracket$.

Remarque : Attention, ce genre de raisonnement peut parfois être dangereux : il n'est par exemple pas vrai que les marginales d'une loi uniforme suivent encore une loi uniforme, cf. paragraphe V.2.

IV.3 Loi de Bernoulli

IV.3.a Cadre

On a une expérience avec deux issues possibles : succès (avec proba $p \in \,]\,0\,;1\,[\,$) et échec (avec proba $q = 1 - p$).

On suppose que X vaut 1 en cas de succès, et 0 en cas d'échec. On dit alors que X suit une loi de Bernoulli de paramètre p.

Il est fréquent de poser $q = 1 - p$ quand on manipule une loi de Bernoulli ou une loi binomiale (ou, l'année prochaine, une loi géométrique).

IV.3.b Définition et notation

Définition. Soit $p \in \,]\,0\,;1\,[$. On dit qu'une variable aléatoire réelle finie X suit une loi de Bernoulli de paramètre p si :

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

On note alors $X \sim B(p)$.

Le paramètre p (resp. $q = 1 - p$) est appelé probabilité de succès (resp. d'échec).

Remarques :

- Les cas $p = 0$ et $p = 1$ sont écartés le plus souvent car ils correspondent à une loi certaine (égale à 0 et à 1 respectivement), même si on peut les trouver dans certains énoncés.
- Pour pouvoir affirmer qu'une variable aléatoire suit une loi de Bernoulli, il suffit que $X(\Omega) = \{0; 1\}$. Si on note $p = P(X = 1)$, on aura alors automatiquement $P(X = 0) = 1 - p$.

Bernoulli, avec un seul i !

- Dès lors, pour déterminer totalement une loi de Bernoulli, il suffit de connaître $P(X = 1)$. En effet, on a alors automatiquement $P(X = 0) = 1 - P(X = 1)$ (ce n'est bien sûr valable que si on sait déjà que X suit une loi de Bernoulli et qu'on cherche le paramètre). Par conséquent, dans la suite, si on sait déjà que $X(\Omega) = \{0; 1\}$, alors on pourra affirmer directement que X suit une loi de Bernoulli et le paramètre sera $P(X = 1)$. Par exemple, si $X(\Omega) = \{0; 1\}$ et si $P(X = 1) = 1/3$, on peut dire directement que $X \sim B(1/3)$, il n'est pas nécessaire de calculer $P(X = 0)$!

Exemples :

- On lance une pièce de monnaie truquée de telle sorte que la probabilité qu'elle tombe sur Pile soit $p \in]\,0\,;1\,[$. Soit X la variable aléatoire qui vaut 1 si la pièce tombe sur Pile et 0 si elle tombe sur face. Alors $X \sim B(p)$.
- Soit X la variable aléatoire qui vaut 1 si on tire un roi dans un jeu de 52 cartes et qui vaut 0 sinon. Alors $X \sim B(1/13)$.
- Si $X \sim B(p)$, alors $X^2 \sim B(p)$. En effet, comme X ne prend que 0 et 1 pour valeurs, on a $X^2 = X$.
- Si $X \sim B(p)$, alors $1 - X \sim B(1 - p)$. En effet $1 - X$ ne prend que $1 = 1 - 0$ et $0 = 1 - 1$ pour valeurs et $P(1 - X = 1) = P(X = 0) = p$.

En particulier, si $X \sim B(1/2)$, alors $1 - X \sim B(1/2)$: on a alors $X \sim 1 - X$ ce qui ne veut pas dire que $X = 1 - X$. Plus fort : X et $1 - X$ sont toujours différentes !

Remarque : La loi de Bernoulli est la loi que l'on rencontre lorsqu'il n'y a que deux issues possibles à une expérience : succès et échec. On code le succès par la valeur 1 et l'échec par la valeur 0. Plus précisément : voir les variables indicatrices ci-dessous. Cependant, attention : ce n'est pas parce qu'il n'y a que deux valeurs que la variable aléatoire suit forcément une loi de Bernoulli, il faut que ces valeurs soient 0 et 1. Cependant, on peut dans tous les cas s'y ramener facilement : si a et b sont deux réels tels que $a < b$ et si X est une variable aléatoire réelle finie qui ne prend que les valeurs a et b (avec probabilités non nulles), alors :

$$Y = \frac{X - a}{b - a} \sim B(p)$$

avec $p = P(X = b)$ (exo).

Par exemple, si les deux valeurs sont ± 1, on parle plutôt de loi de Rademacher.

IV.3.c Espérance et variance

Proposition. Soit $p \in]\,0\,;1\,[$. Si $X \sim B(p)$, alors $E(X) = p$ et $V(X) = p(1 - p)$.

Démonstration. $\rightsquigarrow$ Exercice.

Répétons ce que nous avons dit plus haut : si une variable aléatoire ne prend que les valeurs 0 et 1, on peut affirmer directement qu'elle suit une loi de Bernoulli, il ne manque que le paramètre p. Or,

$$p = P(X = 1) = E(X).$$

Ainsi, selon les cas, il suffit de calculer $P(X = 1)$ ou $E(X)$ pour donner le paramètre : il est inutile de donner $P(X = 0)$!

IV.3.d Lien avec les indicatrices

Définition. Soit $A \in \mathscr{P}(\Omega)$ un événement. Soit :

$$X : \left\{ \begin{array}{ccl} \Omega & \to & \mathbb{R} \\ \omega & \mapsto & : \left\{ \begin{array}{ll} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{array} \right. \end{array} \right.$$

X est appelée variable indicatrice de l'événement A et est noté $\mathbb{1}_A$.

La définition est évidemment la même que pour l'indicatrice d'un ensemble quelconque, cf. chapitre 4.

Proposition. Soit $A \in \mathscr{P}(\Omega)$ un événement. Alors $\mathbb{1}_A \sim B(P(A))$ et en particulier : $E(\mathbb{1}_A) = P(A)$

Démonstration. $\mathbb{1}_A(\Omega) = \{0; 1\}$ donc X suit une loi de Bernoulli et $P(\mathbb{1}_A = 1) = P(A)$.

Remarque : ⚠ Le résultat $E(\mathbb{1}_A) = P(A)$ (c'est-à-dire que la probabilité est égale à l'espérance de l'indicatrice) est très important et doit être retenu. Ce résultat permet de montrer certains résultats très facilement en utilisant des propriétés très simples vérifiées par l'espérance (croissance, linéarité etc.) que ne vérifie pas la probabilité. Voir par exemple l'inégalité de Markov dans la partie VII.1.

IV.4 Loi binomiale

IV.4.a Cadre

On procède à n répétitions indépendantes d'une expérience de Bernoulli, c'est-à-dire d'une expérience comportant deux issues possibles : succès (avec proba p) et échec (avec proba $1-p$) et on note X le nombre de succès obtenus. On modélise cette situation en prenant l'espace probabilisé suivant :

$$\Omega = \{0;1\}^n = \{(a_1, \dots, a_n) \mid \forall i \in [\![1 ; n]\!], a_i \in \{0;1\}\}$$

où les 0 correspondent aux échecs et les 1 aux succès (et où a_i correspond au résultat de la i-ème répétition de l'expérience). On munit Ω de $\mathscr{P}(\Omega)$ et on se demande quelle probabilité mettre sur $(\Omega, \mathscr{P}(\Omega))$.
La probabilité uniforme n'est pas la probabilité naturelle : par exemple, puisque la proba de succès est p et la proba d'un échec $1-p$, il est intuitif que la proba d'avoir p succès est p^n et non pas $1/2^n$. Il est plus naturel de définir une probabilité P de la façon suivante : si $\omega = (a_1, \dots, a_n) \in \Omega$, on pose $P(\{\omega\}) = p^k(1-p)^{n-k}$ (où k est le nombre de 1 et $n-k$ le nombre de 0). Par exemple,

$$P(\{(0;\dots;0)\}) = (1-p)^n \qquad \text{et} \qquad P(\{(1;0;\dots;0)\}) = p(1-p)^{n-1}$$

On définit la variable aléatoire suivante :

$$X : \left\{ \begin{array}{ccc} \Omega & \to & \mathbb{R} \\ (a_1, \dots, a_n) & \mapsto & a_1 + \dots + a_n \end{array} \right.$$

Puisque les éléments de Ω ont des coordonnées égales à 0 ou à 1, la somme des coordonnées est égale au nombre de coordonnées égales à 1 (ici, le nombre de succès). X est donc bien la v.a. qui donne le nombre de succès. X prend les valeurs $0, \dots, n$ et si $k \in [\![0 ; n]\!]$:

$$\begin{aligned} P(X=k) &= P(\{\omega \in \Omega \,|\, X(\omega) = k\}) \\ &= P\left(\bigcup_{\omega \in \Omega, X(\omega)=k} \{\omega\} \right) \\ &= \sum_{\omega \in \Omega, X(\omega)=k} P(\{\omega\}) \\ &= \sum_{\omega \in \Omega, X(\omega)=k} p^k (1-p)^{n-k} \end{aligned}$$

Or, on peut représenter un tel ω par un chemin dans un arbre binaire, chemin avec k succès et $n-k$ échecs lors de n répétitions d'une expérience aléatoire. Il y a donc $\binom{n}{k}$ éventualités ω vérifiant $X(\omega) = k$ donc

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

IV.4.b Définition et notation

Définition. Soient $n \in \mathbb{N}^*$ et $p \in]\,0;1\,[$. On dit qu'une variable aléatoire réelle finie X suit une loi binomiale de paramètres n et p si :

- $X(\Omega) = [\![0 ; n]\!]$.

Le paramètre p (resp. $q = 1-p$) est appelé probabilité de succès (resp. d'échec).

- $\forall k \in [\![0\,;\, n]\!], P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$.

On note alors $X \sim B(n,p)$.

Remarques :

- On écarte en général $p=0$ et $p=1$ car on a alors une loi certaine (égale à 0 et à n respectivement).
- La loi $B(1,p)$ est la loi $B(p)$.
- Comme dit dans le paragraphe précédent, une loi binomiale de paramètres n et p compte le nombre de succès obtenus lors de n répétitions indépendantes d'une expérience aléatoire où la probabilité de succès est p. Nous en reparlerons dans le paragraphe VI.3.

Exercice type : On lance 10000 dés équilibrés. Quelle est la probabilité d'obtenir 2023 fois 6 ?

Rédaction attendue : Soit X la variable aléatoire qui compte le nombre de 6 obtenus. X compte le nombre de succès (obtenir 6, proba 1/6) obtenus lors de 2023 répétitions indépendantes d'une expérience aléatoire (lancer un dé). Ainsi, $X \sim B(10000, 1/6)$ et en particulier :

Ne pas oublier de préciser qu'elles sont indépendantes !

$$P(X=2023) = \binom{10000}{2023} \left(\frac{1}{6}\right)^{2023} \left(\frac{5}{6}\right)^{7077}$$

IV.4.c Espérance et variance

Proposition. Soient $n \in \mathbb{N}^*$ et $p \in]\,0\,;1\,[$. Si $X \sim B(n,p)$, alors $E(X) = np$ et $V(X) = np(1-p)$.

Remarque : On peut remarquer qu'on a n fois l'espérance et la variance d'une loi $B(p)$: nous en reparlerons au paragraphe VI.4.

Démonstration. On a

On enlève le terme d'indice 0 puisqu'il est nul.

$j = k-1$, $k = j+1$

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\, P(X=k) \\
&= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \times \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n \times (n-1)!}{(k-1)!(n-1-(k-1))!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= n \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j+1} (1-p)^{n-(j+1)} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} (1-p)^{n-1-j}
\end{aligned}$$

D'après le binôme de Newton, il vient : $E(X) = np(p+1-p)^{n-1} = np$. D'après la formule de transfert :

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2\, P(X=k) \\
&= \sum_{k=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \times \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \times \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} (k-1) \times \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}
\end{aligned}$$

On enlève le terme d'indice 0 puisqu'il est nul.

Comme en intégration : méthode du $+1-1$.

La seconde somme est égale à $E(X) = np$. La première somme commence en fait en $k=2$ puisque le terme d'indice $k=1$ est nul. On déroule ensuite les calculs :

$$\begin{aligned}
E(X^2) &= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} + np \\
&= \sum_{k=2}^{n} \frac{n(n-1) \times (n-2)!}{(k-2)!(n-2-(k-2))!} p^k (1-p)^{n-k} + np \\
&= n(n-1) \sum_{k=2}^{n} \binom{n-2}{k-2} p^k (1-p)^{n-k} + np \\
&= n(n-1) \sum_{j=0}^{n-2} \binom{n-2}{j} p^{j+2} (1-p)^{n-(j+2)} + np \\
&= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j} + np \\
&= n(n-1)p^2 (p+1-p)^{n-2} + np \\
&= n(n-1)p^2 + np
\end{aligned}$$

Enfin, d'après la formule de König-Huygens :

$$\begin{aligned}
V(X) &= E(X^2) - E(X)^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}$$

□

V Loi d'un couple de variables aléatoires

On revient au cas général : les variables aléatoires sont à nouveau à valeurs dans un ensemble quelconque (sauf quand nous parlerons d'espérance ou de variance).

V.1 Un couple de v.a. = une « grosse » v.a.

Proposition. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires. Alors le couple de variables aléatoires (X, Y) est une variable aléatoire de Ω dans $E \times F$. En d'autres termes, un couple de variables aléatoires est une variable aléatoire dans l'ensemble produit.

Démonstration. Il suffit de voir que $(X, Y) : \omega \to (X(\omega), Y(\omega))$ est une fonction de Ω dans $E \times F$.

Remarques :

- On peut généraliser au cas d'un n-uplet de variables aléatoires : si, pour tout $i \in [\![1\,;\,n]\!]$, X_i est une variable aléatoire de Ω dans un ensemble E_i non vide, alors le n-uplet $(X_1, \dots, X_n)$ est une variable aléatoire de Ω dans l'ensemble produit $E_1 \times \cdots \times E_n$.
- Puisqu'un couple (ou, plus généralement, un n-uplet) de variables aléatoires n'est rien d'autre qu'**une** variable aléatoire dans un ensemble plus gros, tous les résultats précédents concernant une variable aléatoire à valeurs dans un ensemble E (penser à « truc ») s'appliquent. En particulier, on peut parler de la loi d'un couple de variables aléatoires, et celle-ci est entièrement déterminée par l'univers image et les probabilités de chacun des éléments de l'univers image. Plus précisément :

Proposition/Définition.

- La loi d'un couple de variables aléatoires (X, Y) sur Ω est entièrement déterminée par la distribution de probabilité $(P([X = x] \cap [Y = y]))_{(x,y) \times (X,Y)(\Omega)}$.
- Cette loi est appelée loi conjointe du couple (X, Y).
- Les lois de X et Y sont appelées lois marginales du couple (X, Y). Plus précisément, la loi de X est appelée première loi marginale du couple (X, Y), et la loi de Y sa deuxième loi marginale.

On généralise sans mal à un n-uplet de variables aléatoires.

Démonstration. C'est tout simplement le dernier résultat du paragraphe I.3 avec (X, Y) à la place de X (penser à « variable aléatoire machin ») en remarquant que $P((X, Y) = (x, y)) = P([X = x] \cap [Y = y])$.

Remarques :

- D'où sort cette intersection ? C'est très simple : un élément de l'univers image $(x, y) \in X(\Omega) \times Y(\Omega)$ est atteint par la variable aléatoire (X, Y) lorsqu'on a à la fois $X = x$ et $Y = y$, et on sait que l'intersection est la traduction ensembliste du « et » logique. L'événement $[X = x] \cap [Y = y]$ est parfois noté $[X = x]$ et $[Y = y]$ ou parfois plus simplement avec une virgule, c'est-à-dire qu'on trouve parfois la notation $P(X = x, Y = y)$ à la place de $P([X = x] \cap [Y = y])$: il ne faut alors pas perdre de vue que cette virgule représente une intersection, surtout quand on voudra utiliser de l'indépendance.
- L'ensemble $(X, Y)(\Omega)$ est inclus dans $X(\Omega) \times Y(\Omega)$ mais il n'y a pas égalité en général. En effet, si $(x, y) \in X(\Omega) \times Y(\Omega)$, alors x peut être atteint par X et y peut être atteint par y mais il n'existe peut-être pas d'éventualité ω telle qu'on ait à la fois $X(\omega) = x$ **et** $Y(\omega) = y$. Il peut être difficile de donner explicitement $(X, Y)(\Omega)$ alors que $X(\Omega) \times Y(\Omega)$ est simple à donner. Il suffit de faire « comme si » les deux ensembles étaient égaux, i.e. de donner comme ensemble image $X(\Omega) \times Y(\Omega)$ et de poser $P([X = x] \cap [Y = y]) = 0$ lorsque $(x, y) \notin (X, Y)(\Omega)$: voir l'exemple ci-dessous.

Exemple : Une urne contient trois boules blanches et une boule noire. On tire successivement et sans remise deux boules dans l'urne : on note C_1 la couleur de la première boule tirée, et C_2 la couleur de la deuxième boule tirée. Donner la loi conjointe du couple (C_1, C_2).

$C_1(\Omega) = C_2(\Omega) = \{B; N\}$ donc $(C_1, C_2)(\Omega) = \{B; N\}^2$.

- Tout d'abord, on ne peut pas tirer une boule noire aux deux tirages puisqu'il n'y a qu'une boule noire dans l'urne. Par conséquent, $P([C_1 = N] \cap [C_2 = n]) = 0$.
- Cherchons $P([C_1 = N] \cap [C_2 = B])$. Cependant, les deux événements ne sont pas indépendants (la probabilité de tirer une boule lors du second tirage dépend de ce qu'on a tiré au premier tirage). D'après la formule des probabilités composées :

$$\begin{aligned} P([C_1 = N] \cap [C_2 = B]) &= P(C_1 = N) \times P_{[C_1 = N]}(C_2 = B) \\ &= \frac{1}{4} \times 1 \\ &= \frac{1}{4} \end{aligned}$$

En effet, supposons $[C_1 = N]$ réalisé. Alors on a tiré une boule noire au premier tirage, si bien que l'urne, avant le second tirage, ne contient plus que des boules blanches, d'où le résultat. On trouve de même que $P([C_1 = B] \cap [C_2 = N]) = 1/4$ et $P([C_1 = B] \cap [C_2 = B]) = 1/2$. Là aussi, représentons les résultats sous forme de tableau (mais on ne peut pas les représenter sous forme d'histogramme) :

C_1 \ C_2	B	N
B	$1/2$	$1/4$
N	$1/4$	0

Pas exactement en fait : (N, N) n'est pas atteint par (C_1, C_2), mais ce n'est pas très grave, il suffit de lui attribuer une probabilité nulle.

V.2 Lien entre loi conjointe et lois marginales

Connaissant la loi conjointe d'un couple (X, Y), peut-on en déduire les lois marginales ? La réponse est oui, grâce à la formule des probabilités totales. En effet, les $[Y = y]_{y \in Y(\Omega)}$ forment un système complet d'événements donc, d'après la formule des probabilités totales, pour tout $x \in X(\Omega)$:

$$P(X = x) = \sum_{y \in Y(\Omega)} P([X = x] \cap [Y = y])$$

et donc, si les $P([X = x] \cap [Y = y])$ sont connues, c'est-à-dire si la loi conjointe du couple (X, Y) est connue, on connaît la loi de X. De même, on connaît la loi de Y. Morale de l'histoire :

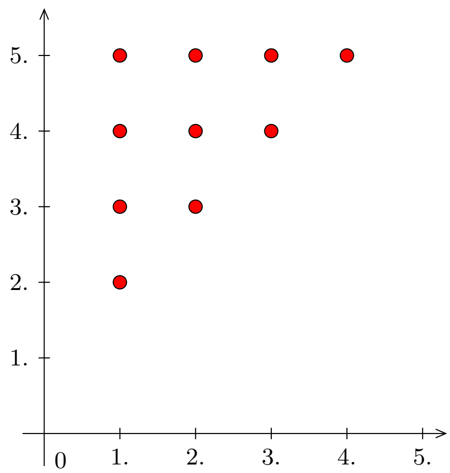
Les résultats de ce paragraphe s'adaptent facilement à un n-uplet de variables aléatoires.

Comme dit au chapitre 26, c'est le seul cadre où on utilise la formule des probas totales avec des intersections. Méthode à connaître !

Théorème. Si on connaît la loi conjointe d'un couple (X, Y), on peut toujours en déduire les lois marginales du couple.

Exemple : Supposons que le couple (X, Y) suive une loi uniforme sur le « triangle » $\{(1, 2); (1, 3); (1, 4); (1, 5); (2, 3); \dots; (4, 5)\}$ représenté par les points ci-dessous :

Si on écrit les diverses probabilités de la loi de (X, Y) sous forme de tableau comme dans l'exemple ci-dessus, alors les marginales sont obtenues en sommant les termes des lignes, pour la première coordonnée, ou des colonnes, pour la seconde. On retrouve le fait qu'on peut retrouver les marginales à partir de la loi conjointe.



En d'autres termes, la probabilité que (X, Y) soit égale à l'un des couples appartenant à l'ensemble vaut $1/10$ ($1/\operatorname{card}(E)$). Donner la loi marginale de X. Tout d'abord, $X(\Omega) = [\![1 \,;\, 4]\!]$. Puisque $Y(\Omega) = [\![2 \,;\, 5]\!]$, les événements $[Y = 2], [Y = 3], [Y = 4]$ et $[Y = 5]$ forment un système complet d'événements donc, d'après la formule des probabilités totales :

$$
\begin{aligned}
P(X=1) &= P_{[Y=2]}(X=1)P(Y=2) + P_{[Y=3]}(X=1)P(Y=3) \\
&\qquad + P_{[Y=4]}(X=1)P(Y=4) + P_{[Y=5]}(X=1)P(Y=5) \\
&= P([X=1]\cap[Y=2]) + P([X=1]\cap[Y=3]) \\
&\qquad + P([X=1]\cap[Y=4]) + P([X=1]\cap[Y=25]) \\
&= \frac{4}{10}
\end{aligned}
$$

On trouve de même que $P(X = 2) = 3/10$, $P(X = 3) = 2/10$ et $P(X = 4) = 1/10$: X ne suit pas une loi uniforme ! Cela se voit très bien sur le dessin : X a plus de chances d'être égal à 1 qu'à 4 ! Attention à ne pas conclure trop vite que les marginales d'une loi uniforme suivent une loi uniforme...

On peut retrouver les lois marginales à partir de la loi conjointe, oui, mais il y a un peu de travail et les résultats sont parfois contre-intuitifs !

On peut se poser la question inverse : connaissant les lois marginales, peut-on en déduire la loi conjointe du couple ? La réponse est non en général. Par exemple, si $X \sim B(1/2)$ et si $Y = X$ alors la loi du couple (X, Y) est donnée par le tableau suivant :

X \ Y	0	1
0	1/2	0
1	0	1/2

tandis que si $X \sim B(1/2)$ et $Y = 1 - X$ alors la loi du couple est donnée par le tableau suivant :

X \ Y	0	1
0	0	1/2
1	1/2	0

On a dans les deux cas $X \sim B(1/2)$ et $Y \sim B(1/2)$ mais, dans les deux cas, la loi du couple n'est pas la même : la loi du couple, en général, ne peut pas être déduite des lois marginales. La raison est que, même en connaissant X et Y séparément, on ne connaît pas la façon dont X et Y se comportent l'une par rapport à l'autre.

Connaître la loi d'un couple est beaucoup plus général que connaître la loi de chaque coordonnée : les deux coordonnées sont-elles égales ? sont-elles toujours différentes ? etc.

VI Variables aléatoires indépendantes

VI.1 Définition et caractérisation

VI.1.a Indépendance de deux variables aléatoires

Définition. Deux variables aléatoires X et Y définies sur le même Ω sont indépendantes si pour tous $A \in \mathscr{P}(X(\Omega))$ et $B \in \mathscr{P}(Y(\Omega))$, les événements $[X \in A]$ et $[Y \in B]$ sont indépendants, c'est-à-dire si $P([X \in A] \cap [Y \in B]) = P(X \in A) \times P(Y \in B)$. On note alors : $X \perp\!\!\!\perp Y$.

Remarque : En d'autres termes, deux v.a. sont indépendantes lorsque tout événement relatif à X est indépendant de tout événement relatif à Y : tout ce qui se passe du côté de X est indépendant de tout ce qui se passe du côté de Y !

Proposition. Deux variables aléatoires X et Y sur Ω sont indépendantes si et seulement si :

$$\forall (x,y) \in X(\Omega) \times Y(\Omega), P([X = x] \cap [Y = y]) = P(X = x) \times P(Y = x)$$

En d'autres termes, pour savoir si deux variables aléatoires sont indépendantes, il suffit de regarder « comment elles se comportent vis-à-vis des singletons.

Remarque : En particulier, lorsque deux variables aléatoires X et Y sont indépendantes, on peut retrouver la loi conjointe du couple (X, Y) à l'aide des lois marginales.

Démonstration. Si X et Y sont indépendantes, alors pour tous $x \in X(\Omega)$ et $y \in Y(\Omega)$, on posant $A = [X = x]$ et $B = [Y = y]$, on obtient bien que $P([X = x] \cap [Y = y]) = P(X = x) \times P(Y = y)$.

Réciproquement, supposons que pour tous x et y, on ait $P([X = x] \cap [Y = y]) = P(X = x) \times P(Y = x)$. Soient $A \in \mathscr{P}(\Omega)$ et $B \in \mathscr{P}(Y(\Omega))$. Notons $A = \{x_1; \dots; x_n\}$ et $B = \{y_1; \dots; y_p\}$ si bien que :

$$\begin{aligned}
P([X \in A] \cap [Y \in B]) &= P\left(\left(\bigcup_{i=1}^{n} [X = x_i]\right) \cap \left(\bigcup_{j=1}^{p} [Y = y_j]\right)\right) \\
&= P\left(\bigcup_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p} [X = x_i] \cap [Y = y_j]\right) \\
&= \sum_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p} P([X = x_i] \cap [Y = y_j]) \\
&= \sum_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p} P(X = x_i) \times P(Y = y_j) \\
&= \left(\sum_{i=1}^{n} P(X = x_i)\right) \times \left(\sum_{j=1}^{p} P(Y = y_j)\right) \\
&= P(A) \times P(B) \qquad \square
\end{aligned}$$

Distributivité de l'intersection sur l'union.

Événements deux à deux incompatibles.

Par hypothèse sur X et Y.

Distributivité du produit sur la somme.

et donc $X \perp\!\!\!\perp Y$.

Exemple : Soient $X \sim B(1/2)$ et $Y = 1 - X$. Alors $P(X = 1) = P(Y = 1) = 0$ mais $P([X = 1] \cap [Y = 1]) = 1/2 \neq P(X = 1) \times P(Y = 1)$: les deux variables aléatoires ne sont pas indépendantes, ce qui est complètement intuitif : si on connaît l'une, alors on connaît l'autre ! L'interprétation vue au chapitre précédent sur les événements indépendants est en effet encore valable pour les variables aléatoires, à savoir : deux variables aléatoires sont indépendantes lorsque la valeur de l'une ne donne aucune information sur la valeur de l'autre.

VI.1.b Indépendance mutuelle de n variables aléatoires

Ici, la situation est plus simple que pour l'indépendance mutuelle de n événements.

Proposition/Définition. n variables aléatoires $X_1, \dots, X_n$ définies sur le même Ω sont mutuellement indépendantes si, pour tout choix de $A_1 \in \mathscr{P}(X_1(\Omega)), \dots, A_n \in \mathscr{P}(X_n(\Omega))$, on a :

$$P([X_1 \in A_1] \cap \cdots \cap [X_n \in A_n]) = P(X_1 \in A_1) \times \cdots \times P(X_n \in A_n)$$

De plus, cette condition est équivalente à : $\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)$,

$$P([X_1 = x_1] \cap \cdots \cap [X_n = x_n]) = P(X_1 = x_1) \times \cdots \times P(X_n = x_n)$$

On note alors : $X_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp X_n$.

Là aussi, il suffit de s'intéresser aux singletons.

Démonstration. Analogue au cas de deux variables aléatoires.

Lorsque des variables sont à la fois indépendantes et de même loi, on dit qu'elles sont i.i.d. : indépendantes identiquement distribuées.

Remarque : Attention, on rappelle que pour montrer que n événements sont indépendants, s'intéresser à l'intersection des n événements en eux-mêmes est insuffisant, il faut s'intéresser à l'intersection sur toute partie I de $\llbracket 1\,;\, n \rrbracket$, cf. chapitre 26. Cependant, pour des variables aléatoires, ça marche. La raison est que, quand on regarde les intersections des événements $[X_i \in A_i]$, si on prend $A_i = X_i(\Omega)$ tout entier, on intersecte en fait avec Ω ce qui fait que l'événement « disparaît de l'intersection » (intersecter avec tout l'ensemble ne change rien).

Remarque : En pratique, des variables aléatoires indépendantes servent à modéliser n expériences aléatoires indépendantes ou n répétitions indépendantes d'une expérience aléatoire : on note X_1 le résultat de la première expérience, et ainsi de suite jusqu'à X_n. Cependant, est-ce toujours possible ? Comme dans le paragraphe I.3, la réponse est oui :

Théorème. Soient $(p_e)_{e \in E}$ une distribution de probabilités de E et $(q_f)_{f \in F}$ une distribution de probabilités sur F. Il existe un espace probabilisé $(\Omega, \mathscr{P}(\Omega), P)$ et des variables aléatoires $X : \Omega \to E$ et $Y : \Omega \to F$ indépendantes telles que pour tout $(e, f) \in E \times F, P(X = e) = p_e$ et $P(Y = f) = q_f$.

Cet énoncé se généralise sans mal à un nombre quelconque fini de variables aléatoires.

Démonstration. Tout d'abord, par distributivité du produit sur la somme :

$$\sum_{e \in E, f \in F} p_e q_f = \left(\sum_{e \in E} p_e\right) \times \left(\sum_{f \in F} q_f\right) = 1$$

□

La famille $(p_e q_f)$ est donc une distribution de probabilités sur l'ensemble $E \times F$: il existe donc un espace probabilisé $(\Omega, \mathscr{P}(\Omega), P)$ et une variable aléatoire $(X, Y) : \Omega \to E \times F$ telle que, pour tous e et f : $P([X = e] \cap [Y = f]) = p_e \times q_f$.

On trouve de même qu'au paragraphe I.3 que les lois marginales de X et de Y vérifient alors : $P(X = e) = p_e$ pour tout $e \in E$, et idem pour Y. En particulier, pour tous e et f, $P([X = e] \cap [Y = f]) = P(X = e) \times P(Y = f)$: X et Y sont indépendantes.

Remarque : Par conséquent, comme dans le paragraphe I.3, on pourra toujours se donner des variables indépendantes suivant des lois quelconques sans avoir à justifier leur existence.

Exemple : Soient $(p_1, \dots, p_n) \in \,]\,0\,;1\,[^n$ et $(X_1, \dots, X_n)$ des v.a. indépendantes de loi respective $B(p_1), \dots, B(p_n)$. Donnons la loi de $X_1 \times \cdots \times X_n$.

De telles v.a. existent d'après le théorème précédent.

$X_1 \times \cdots \times X_n$ est à valeurs dans $\{0; 1\}$ donc suit une loi de Bernoulli. Trouvons son paramètre.

$$\begin{aligned} P(X_1 \times \cdots \times X_n = 1) &= P([X_1 = 1] \cap \cdots \cap [X_n = 1]) \\ &= P(X_1 = 1) \times \cdots \times P(X_n = 1) \\ &= p_1 \times \cdots \times p_n \end{aligned}$$

Par indépendance.

On en déduit que $X_1 \times \cdots \times X_n \sim B(p_1 \times \cdots \times p_n)$.

Remarque : Quand on aura des variables aléatoires mutuellement indépendantes et de même loi (ce qu'on appelle indépendantes et identiquement distribuées, ou i.i.d, donc), on pourra aller plus loin que dans le paragraphe I.3.b en ce qui concerne l'égalité des lois, puisqu'on pourra l'affirmer pour des sommes, des produits, des fonctions de plusieurs variables aléatoires (et non pas pour une seule). Un exemple vaut mieux qu'un long discours.

Exemple : Soient $X_1, \dots, X_{2n}$ des variables aléatoires indépendantes et de même loi. On pose $S_n = X_1 + \cdots + X_n$ et $S_{2n} = X_1 + \cdots + X_{2n}$. Alors

$$S_{2n} - S_n = X_{n+1} + \cdots + X_{2n}$$

qui est égale en loi à S_n. En pratique, il suffit de dire que les variables aléatoires sont indépendantes et de même loi, mais soyons plus explicites ici. Tout d'abord, $X_{n+1} \sim X_1, \dots, X_{2n} \sim X_n$. En d'autres termes, les deux n-uplets $Y_1 = (X_1, \dots, X_n)$ et $Y_2 = (X_{n+1}, \dots, X_{2n})$ ont les mêmes marginales. Les variables aléatoires étant mutuellement indépendantes, on peut retrouver les lois des n-uplets à l'aide des marginales, donc les n-uplets Y_1 et Y_2 ont la même loi, si bien que, pour toute fonction f, $f(Y_1)$ et $f(Y_2)$ ont la même loi (cf. paragraphe I.4), et en particulier pour la fonction qui à un n-uplet associe la somme de ses coordonnées.

Il découlera également du lemme des coalitions (voir plus bas) que $S_{2n} - S_n$ est indépendante de S_n.

Ce cas de figure se produit souvent, mais inutile de justifier autant en pratique, il suffit de dire que les X_i sont indépendantes et ont la même loi pour conclure. On pourra donc affirmer par exemple, le cas échéant, que $X_1 \times \cdots \times X_n$ et $X_{n+1} \times \cdots \times X_{2n}$ ont la même loi.

VI.2 Exemple important : loi d'un maximum ou d'un minimum

Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{Z}$. Si on demande de déterminer la loi de $M = \max(X, Y)$ (respectivement de $W = \min(X, Y)$), on pense immédiatement à calculer, pour tout $n \in \mathbb{Z}$, les probabilités $P(M \leqslant n)$ (respectivement $P(W \geqslant n)$). En effet :

Si elles sont discrètes mais pas à valeurs dans $\mathbb{Z}$, on calcule $P(M \leqslant t)$ et $P(W \geqslant t)$ pour tout $t \in \mathbb{R}$ et on applique la même méthode.

$$P(M \leqslant n) = P([X \leqslant n] \cap [Y \leqslant n]) \quad \text{et} \quad P(W \geqslant n) = P([X \geqslant n] \cap [Y \geqslant n]).$$

Si on connaît la loi du couple (X, Y), comme lorsque X et Y sont indépendantes, alors on aboutit à une formule explicite et, Pour retrouver la loi, on utilise le fait que

$$P(M = n) = P(M \leqslant n) - P(M \leqslant n - 1).$$

et $P(W = n) = P(W \geqslant n) - P(W \geqslant n + 1)$.

L'idée est que le maximum (respectivement minimum) de deux réels est majoré (respectivement minoré) par n si et seulement si ces deux réels le sont : le plus grand élève d'une classe fait moins de deux mètres ssi tous les élèves font moins de deux mètres et le plus petit fait plus d'un mètre ssi tous les élèves font plus d'un mètre.

Exemple : Soit $N \in \mathbb{N}^*$. On dispose d'un dé équilibré à N faces que l'on lance deux fois. On note X et Y les résultats respectifs des deux lancers. Il s'agit de deux variables aléatoires indépendantes de loi $U(\llbracket 1\,;\,N \rrbracket)$. Déterminons les lois de $M = \max(X, Y)$ et $W = \min(X, Y)$.

- On a $M(\Omega) = W(\Omega) = \llbracket 1\,;\,N \rrbracket$.
- Pour tout $n \in \llbracket 1\,;\,N \rrbracket$, $P(X \leqslant n) = \displaystyle\sum_{k=1}^{n} P(X = k) = \frac{n}{N}$.
  De même $P(Y \leqslant n) = \dfrac{n}{N}$.

- Pour tout $n \in [\![1\,;\,N]\!]$, on a donc

$$P(M \leqslant n) = P([X \leqslant n] \cap [Y \leqslant n]) = P(X \leqslant n)P(Y \leqslant n) = \frac{n^2}{N^2},$$

par indépendance de X et Y. Cette formule reste vraie si $n = 0$. Ainsi, pour tout $n \in [\![1\,;\,N]\!]$,

$$P(M = n) = P(M \leqslant n) - P(M \leqslant n-1) = \frac{n^2}{N^2} - \frac{(n-1)^2}{N^2} = \frac{2n-1}{N^2}.$$

- Pour tout $n \in [\![1\,;\,N]\!]$, on a

$$P(W \geqslant n) = P([X \geqslant n] \cap [Y \geqslant n]) = P(X \geqslant n)P(Y \geqslant n).$$

par indépendance de X et Y. Ainsi

$$P(W \geqslant n) = (1 - P(X \leqslant n-1))(1 - P(Y \leqslant n-1)) = \left(1 - \frac{n-1}{N}\right)^2.$$

Par conséquent, pour tout $n \in [\![1\,;\,N]\!]$,

$$\begin{aligned} P(W = n) &= P(W \geqslant n) - P(W \geqslant n+1) \\ &= \left(1 - \frac{n-1}{N}\right)^2 - \left(1 - \frac{n}{N}\right)^2 = \frac{2N+1-2n}{N^2} \end{aligned}$$

On généralise sans mal à un plus grand nombre de variables aléatoires.

VI.3 Transfert de familles de variables aléatoires et lemme des coalitions

Proposition. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires, et soient f et g deux fonctions définies respectivement sur $X(\Omega)$ et $Y(\Omega)$. Si X et Y sont indépendantes, alors $f(X)$ et $g(Y)$ sont indépendantes. .

On généralise aisément à un nombre quelconque de variables aléatoires indépendantes

Remarque : C'est intuitif ! Si les valeurs prises par X n'apportent aucune information sur les valeurs prises par Y, les valeurs prises par $f(X)$ n'apportent aucune information sur les valeurs prises par $g(Y)$.

Démonstration. Soient $a \in f(X(\Omega))$ et $b \in g(Y(\Omega))$. Notons $C = f^{-1}(\{a\})$ et $D = g^{-1}(B)$ les ensembles des antécédents de a par f et de b par g. Par conséquent, $X \in C \iff f(X) = a$ et idem pour Y et D. Dès lors :

$$\begin{aligned} P([f(X) = a] \cap [g(Y) = b]) &= P([X \in C] \cap [Y \in D]) \\ &= P(X \in C) \times P(Y \in D) \\ &= P(f(X) = a) \times P(g(Y) = b) \end{aligned}$$

Par indépendance de X et Y.

□

Avant de généraliser à des « coalitions » de plus d'une variable aléatoire, définissons (de même que dans le paragraphe I.4) des fonctions de variables aléatoires.

Là non plus, comme dans le paragraphe I.4, il n'y a rien à prouver, et on dit que $g(X, Y)$ est un transfert de (X, Y). On généralise aisément à un nombre quelconque de variables aléatoires.

Proposition/Définition. Soient X et Y deux variables aléatoires sur Ω. Soit g une fonction définie sur $X(\Omega) \times Y(\Omega)$. Alors $Z : \omega \mapsto g(X(\omega), Y(\omega))$ est une variable aléatoire finie notée $g(X, Y)$.

Lemme (Lemme des coalitions). Soient $X_1, \dots, X_n$ des v.a. indépendantes sur Ω. Soit $m \geqslant 1$ et soient f et g définies respectivement sur $X_1(\Omega) \times \cdots \times X_m(\Omega)$ et $X_{m+1}(\Omega) \times \cdots \times X_n(\Omega)$. Alors $f(X_1, \dots, X_m)$ et $g(X_{m+1}, \dots, X_n)$ sont indépendantes.

Démonstration. Notons Z la variable aléatoire produit $(X_1, \dots, X_m)$ et T la variable aléatoire $(X_{m+1}, \dots, X_n)$. Prouvons que les deux variables aléatoires Z et T sont indépendantes. Soient $z = (z_1, \dots, z_m) \in X_1(\Omega) \times \cdots \times X_m(\Omega)$ et $t = (t_{m+1}, \dots, t_n) \in X_{m+1}(\Omega) \times \cdots \times X_n(\Omega)$. Par indépendance de $X_1, \dots, X_n$:

$$\begin{aligned} P([Z=z] \cap [T=t]) &= P([X_1 = z_1] \cap \cdots \cap [X_m = z_m] \cap [X_{m+1} = t_{m+1}] \cap \cdots \cap [X_n = t_n]) \\ &= P(X_1 = z_1) \times \cdots \times P(X_m = z_m) \times P(X_{m+1} = t_{m+1}) \times \cdots \times P(X_n = t_n) \\ &= P([X_1 = z_1] \cap \cdots \cap [X_m = z_m]) \times P([X_{m+1} = t_{m+1}] \cap \cdots \cap [X_n = t_n]) \\ &= P(Z = z) \times P(T = t) \end{aligned}$$

□

c'est-à-dire que Z et T sont indépendantes : il suffit ensuite d'appliquer le résultat précédent, c'est-à-dire que $f(Z)$ et $g(T)$ sont indépendantes.

Remarque : Ce résultat a l'air compliqué, mais il dit juste que si on dispose de variables indépendantes, on peut les regrouper de la façon qu'on veut (des coalitions, sans élément en commun évidemment) et leur appliquer les fonctions qu'on veut, on obtiendra encore des variables indépendantes. Par exemple, si X, Y, Z sont indépendantes, alors $X^2 + \sin(Y)$ et e^Z sont indépendantes : les informations éventuelles sur une coalition n'apportent aucune information sur l'autre coalition. On généralise aisément à un nombre quelconque de coalitions.

Autre exemple : si $X_1, \dots, X_{2n}$ sont i.i.d, alors $X_1 + \cdots + X_n$ est indépendante de $X_{n+1} + \cdots + X_{2n}$.

Proposition. Soient $p \in]\,0\,;1\,[$ et $X_1, \dots X_n$ des v.a. indépendantes suivant chacune une loi de Bernoulli de paramètre p. Alors $X_1 + \cdots + X_n \sim B(n,p)$.

Remarque : Ce résultat est intuitif : une loi de Bernoulli représente le résultat d'une expérience aléatoire (1 pour succès, 0 pour échec). Par conséquent, la somme de n variables aléatoires indépendantes suivant chacune une loi de Bernoulli de paramètres p représente le nombre de succès lors de n répétitions indépendantes d'une expérience aléatoire. Or, c'est précisément ce que représente une variable aléatoire suivant une loi binomiale !

Rappelons que, quand on dispose de réels égaux à 0 ou 1, la somme de ces réels est égale au nombre de ces réels qui sont égaux à 1.

Démonstration. Par récurrence sur n. Commençons par un lemme.

Lemme. Si $X \sim B(n,p)$ et si $Y \sim B(m,p)$ et si X et Y sont indépendantes, alors $X + Y \sim B(n+m,p)$.

En d'autres termes, compter les succès lors de $n+m$ répétitions revient à compter les succès lors des n premières répétitions puis à compter les succès lors des m expériences suivantes, et enfin à sommer les nombres de succès obtenus. Attention, le paramètre p doit être le même pour les deux lois !

Démontrons ce lemme. Notons $Z = X + Y$. Tout d'abord, $Z(\Omega) = [\![0\,; n+m]\!]$. Soit donc $k \in [\![0\,; n+m]\!]$. Les événements $[X=0], \dots, [X=m]$ forment un système complet d'événements donc, d'après la formule des probas totales :

$$\begin{aligned} P(Z=k) &= \sum_{i=0}^{n} P_{[X=i]}(X+Y=k) P(X=i) \\ &= \sum_{i=0}^{n} P_{[X=i]}(Y=k-i) P(X=i) \\ &= \sum_{i=0}^{n} P(Y=k-i) P(X=i) \end{aligned}$$

Par indépendance de X et Y.

Or, si $i > k$, $k - i < 0$ donc $P(Y = k-i) = 0$ (rappelons qu'une loi binomiale est à valeurs positives) donc cette somme va en réalité de 0 à k. Par conséquent :

$$\begin{aligned}
P(Z=k) &= \sum_{i=0}^{k} P(Y=k-i)P(X=i) \\
&= \sum_{i=0}^{k} \binom{m}{k-i} p^{k-i}(1-p)^{m-k+i}\binom{n}{i}p^{i}(1-p)^{n-i} \\
&= p^k(1-p)^{n+m-k}\sum_{i=0}^{k}\binom{m}{k-i}\binom{n}{i}
\end{aligned}$$ □

Or, d'après la formule de Vandermonde (cf. exercice 37 du chapitre 3 et chapitre 17), la somme de droite est égale à $\binom{n+m}{k}$ ce qui clôt la démonstration du lemme.

On raisonne ensuite par récurrence, et pour l'hérédité, il faut utiliser le lemme de coalitions (pour dire que $X_1 + \cdots + X_n$ est indépendante de X_{n+1}) : $\rightsquigarrow$ EXERCICE..

Remarque : Il pourrait être tentant de simplement définir une variable binomiale comme la somme de n variables de Bernoulli indépendantes, mais cette définition serait trop restrictive. : il est en effet possible de définir des variables suivant une loi binomiale qui ne peuvent pas se décomposer sous cette forme (cf. exercice 48).

VI.4 Espérance et variance

Remarque : Rappelons qu'un couple de variables aléatoires est une variables aléatoire à valeur dans l'espace produit. De même pour un n-uplet de variables aléatoires. Dès lors (comme on l'a déjà vu), on peut appliquer à un couple ou à un n-uplet de variables aléatoires des résultats valables pour une seule variable aléatoire, par exemple le théorème de transfert. Voici une application spectaculaire de ce principe.

Proposition. Soient X et Y deux variables aléatoires réelles indépendantes. Alors $E(XY) = E(X)E(Y)$.

Remarque : ⚠ C'est faux si les variables ne sont pas indépendantes ! Par exemple, si $X \sim B(1/2)$ et $Y = X$, alors $XY = X^2 = X$ (rappelons qu'une loi de Bernoulli ne prend que les valeurs 0 et 1) donc $E(XY) = E(X) = 1/2$ alors que $E(X)E(Y) = 1/4$.

DÉMONSTRATION. Notons Z le couple (X, Y), et soit g la fonction définie sur $X(\Omega) \times Y(\Omega)$ par $g(x, y) = xy$. Alors $E(XY) = E(g(Z))$. Par conséquent, d'après le théorème de transfert,

$$\begin{aligned}
E(XY) &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} g(x,y)P([X=x]\cap[Y=y]) \\
&= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xyP(X=x)\times P(Y=y) \\
&= \left(\sum_{x\in X(\Omega)} xP(X=x)\right)\times\left(\sum_{y\in Y(\Omega)} yP(Y=y)\right) \\
&= E(X)E(Y)
\end{aligned}$$ □

Par indépendance de X et Y.

On peut évidemment généraliser à un nombre quelconque de variables indépendantes.

Proposition. Soient $X_1, \dots, X_n$ des v.a. indépendantes. Alors $E(X_1 \dots X_n) = E(X_1) \cdots E(X_n)$.

Démonstration. Démonstration analogue : $\rightsquigarrow$ Exercice.

VI.5 Covariance de deux variables aléatoires

On se donne dans ce paragraphe deux variables aléatoires X et Y sur Ω et à valeurs réelles.

Définition. On définit la covariance de X et Y, notée $\text{Cov}(X,Y)$, par :

$$\text{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$$

Remarques :

- On a $\text{Cov}(X,X) = V(X)$.
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$: la covariance est symétrique.
- Comme la variance, la covariance est un indicateur de dispersion. Plus précisément, la covariance quantifie l'écart conjoint de X et de Y à leurs espérances respectives.

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

Démonstration. $\text{Cov}(X,Y) = E[XY - XE(Y) - YE(X) + E(X)E(Y)]$. Par linéarité de l'espérance :

$$\begin{aligned}\text{Cov}(X,Y) &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

□

Rappelons encore une fois que $E(X)$ et $E(Y)$ sont des réels i.e. des quantités qui ne sont pas aléatoires.

Définition. Si $\text{Cov}(X,Y) = 0$, on dit que X et Y sont décorrélées.

Proposition. Si X et Y sont indépendantes, alors X et Y sont décorrélées.

Démonstration. Découle du paragraphe précédent.

Remarque : ⚠ La réciproque est fausse ! Si $X \sim U(\{-1;0;1\})$ alors X^3 et X^2 sont décorrélées. En effet, $X^3 = X$ donc $E(X^3) = 0$, et idem pour X^5. Par conséquent :

$$\begin{aligned}\text{Cov}(X^3 \times X^2) &= E(X^5) - E(X^3)E(X^2) \\ &= 0\end{aligned}$$

Cependant, elles ne sont pas indépendantes puisque :

$$P([X^3 = 1] \cap [X^2 = 0]) = 0 \neq P(X^3 = 1) \times P(X^2 = 0) = \frac{1}{9}$$

Mais elle peut être vraie pour certaines variables aléatoires, par exemple celles qui suivent une loi de Bernoulli : cf. exercice 47.

Proposition.

- $V(X+Y) = V(X) + 2\text{Cov}(X,Y) + V(Y)$.
- Si $X_1, \dots, X_n$ sont des v.a. sur Ω à valeurs réelles :

On peut voir ça comme une identité remarquable : d'abord toutes les variances, puis tous les « doubles produits ».

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

Démonstration. D'après la formule de König-Huygens et par linéarité de l'espérance :

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

□

D'où la première égalité. La deuxième se prouve par récurrence : $\rightsquigarrow$ Exercice.

Corollaire.

- Si X et Y sont indépendantes, $V(X+Y) = V(X)+V(Y)$.
- Si $X_1, \dots, X_n$ sont deux à deux indépendantes :

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

On peut remplacer « indépendantes » par « non corrélées ».

Remarques :

- C'est faux en général si les variables ne sont pas indépendantes. Prendre $X \sim B(1/2)$, $Y = 1-X$: on a $X+Y=1$ donc $V(X+Y) = 0 \neq V(X)+V(Y) = 1/2$.
- On retrouve la variance d'une loi $B(n,p)$: on a vu en effet que si $X_1, \dots, X_n$ sont indépendantes et suivent chacune une loi $B(p)$, alors $X_1+\cdots+X_n \sim B(n,p)$. Or, les X_i sont indépendantes donc

$$\begin{aligned}
V(X_1+\cdots+X_n) &= V(X_1)+\cdots+V(X_n) \\
&= p(1-p)+\cdots+p(1-p) \\
&= np(1-p)
\end{aligned}$$

VII Inégalités de concentration

VII.1 Inégalité de Markov

Proposition (Inégalité de Markov). Soit X une v.a. **positive**. Soit $a>0$. Alors :

$$P(X \geqslant a) \leqslant \frac{E(X)}{a}$$

Démonstration. On a successivement :

$$\begin{aligned}
P(X \geqslant a) &= E(\mathbb{1}_{\{X\geqslant a\}}) \\
&= \frac{1}{a} \times aE(\mathbb{1}_{\{X\geqslant a\}}) \\
&= \frac{1}{a} E(a\mathbb{1}_{\{X\geqslant a\}}) \\
&\leqslant \frac{1}{a} \times E(X\mathbb{1}_{\{X\geqslant a\}}) \\
&\leqslant \frac{1}{a} \times E(X)
\end{aligned}$$

□

La probabilité est égale à l'espérance de l'indicatrice.

Linéarité de l'espérance.

Car $a\mathbb{1}_{\{X\geqslant a\}} \leqslant X\mathbb{1}_{\{X\geqslant a\}}$, car on se restreint à l'événement $[X \geqslant a]$, puis croissance de l'espérance.

En effet, si $X \geqslant a$, alors $X\mathbb{1}_{\{X\geqslant a\}} = X$ et sinon, $X\mathbb{1}_{\{X\geqslant a\}} = 0 \leqslant X$ car X est à valeurs positives. Dans tous les cas, on a donc $X\mathbb{1}_{\{X\geqslant a\}} \leqslant X$ et on conclut par croissance de l'espérance.

VII.2 Inégalité de Bienaymé-Tchebychev

Corollaire (Inégalité de Bienaymé-Tchebychev). Soit $a > 0$. Alors :

$$P(|X - E(X)| \geqslant a) \leqslant \frac{V(X)}{a^2}$$

X n'a pas besoin d'être positive ici.

Démonstration. $P(|X - E(X)| \geqslant a) = P\left((X - E(X))^2 \geqslant a^2\right)$ car la fonction carré est strictement croissante sur $\mathbb{R}^+$. Or, $(X - E(X))^2$ est une v.a. positive donc, d'après l'inégalité de Markov :

$$\begin{aligned} P(|X - E(X)| \geqslant a) &\leqslant \frac{E\left[(X - E(X))^2\right]}{a^2} \\ &\leqslant \frac{V(X)}{a^2} \end{aligned}$$

□

Remarque : En pratique la justification ci-dessus (la fonction carré strictement croissante sur $\mathbb{R}^+$) est suffisante, mais il faut savoir montrer l'égalité $P(|X - E(X)| \geqslant a) = P\left((X - E(X))^2 \geqslant a^2\right)$ si nécessaire : soit $\omega \in \Omega$.

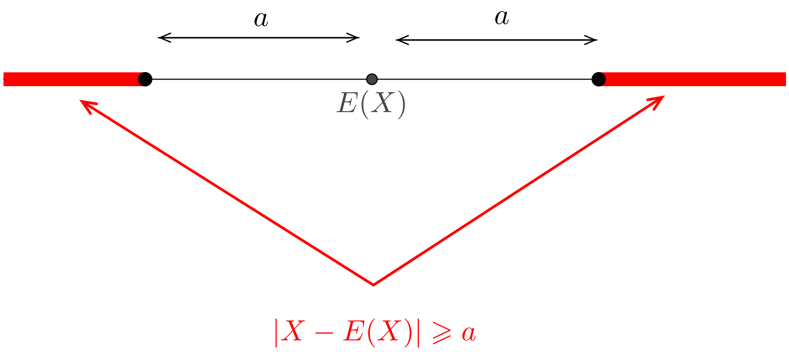
$$\begin{aligned} \omega \in [|X - E(X)| \geqslant a] &\iff |X(\omega) - E(X)| \geqslant a \\ &\iff |X(\omega) - E(X)|^2 \geqslant a^2 \\ &\iff \omega \in \left[|X - E(X)|^2 \geqslant a^2\right] \end{aligned}$$

La deuxième ligne vient du fait que la fonction carré est strictement croissante sur $\mathbb{R}^+$. Il en découle que $[|X - E(X)| \geqslant a] = \left[|X - E(X)|^2 \geqslant a^2\right]$ et en particulier ils ont la même probabilité.

Remarque : Plus généralement, si $f : \mathbb{R} \to \mathbb{R}$ est une fonction croissante, strictement positive, alors $P(X \geqslant a) \leqslant \dfrac{E(f(X))}{f(a)}$ (cf. exercice 68 : attention il n'y a pas forcément égalité si f n'est pas strictement croissante).

VII.3 Interprétation et exemple d'utilisation

L'inégalité de Bienaymé-Tchebychev donne une majoration de la proba, pour une variable aléatoire, de trop « s'écarter » de son espérance, et en particulier de l'erreur faite en faisant l'approximation de X par $E(X)$, la « valeur moyenne ».



Puisque la variance est un indicateur de dispersion d'une v.a. par rapport à son espérance, il est intuitif que plus la variance est faible, plus la variable aléatoire est « concentrée » au voisinage de l'espérance. L'inégalité de Bienaymé-Tchebychev permet de donner une première majoration explicite et de ne plus se contenter de cette approche intuitive.

Exemple : On lance n fois un dé équilibré. On note X_n le nombre de 3 obtenus. Donner un entier n tel que

$$P\left(\left|\frac{X_n}{n}-\frac{1}{6}\right|<\frac{1}{100}\right)\geqslant\frac{95}{100}$$

c'est-à-dire un entier n tel que la fréquence d'apparition du 3 soit égale à $1/6$ plus ou moins 1%, et ce avec une probabilité au moins égale à 95/100.

Réponse : On cherche à appliquer l'inégalité de Bienaymé-Tchebychev mais celle-ci donne une majoration pour une proba du type $P(\cdots\geqslant\dots)$. Il suffit de voir que le problème revient à trouver n tel que

$$P\left(\left|\frac{X_n}{n}-\frac{1}{6}\right|\geqslant\frac{1}{100}\right)\leqslant\frac{5}{100}$$

X_n compte le nombre de succès obtenus (obtenir 3, proba $\frac{1}{6}$) lors de n répétitions indépendantes d'une expérience aléatoire (lancer un dé) donc $X_n\sim B\left(n,\frac{1}{6}\right)$. En particulier, X_n admet une espérance égale à $\frac{n}{6}$ et une variance égale à $\frac{5n}{36}$.

$$\begin{aligned}P\left(\left|\frac{X_n}{n}-\frac{1}{6}\right|\geqslant\frac{1}{100}\right) &= P\left(|X_n-\frac{n}{6}|\geqslant\frac{n}{100}\right)\\ &= P\left(|X_n-E(X_n)|\geqslant\frac{n}{100}\right)\\ &\leqslant \frac{V(X_n)}{\left(\frac{n}{100}\right)^2}\end{aligned}$$

d'après l'inégalité de B-T. On trouve donc que $P\left(\left|\frac{X_n}{n}-\frac{1}{6}\right|\geqslant\frac{1}{100}\right)\leqslant\frac{50000}{36n}$. Il suffit donc de prendre n tel que :

$$\frac{50000}{36n}\leqslant\frac{5}{100}$$

Après calculs, on trouve $n\geqslant 27778$.

Remarque :

- Ce n'est pas le plus petit entier qui convient mais cela permet au moins d'avoir une valeur explicite, ce qui est déjà mieux que la simple intuition : « pour n grand, $\frac{X_n}{n}$ se rapproche de $\frac{1}{6}$ ».
- L'inégalité de Bienaymé-Tchebychev a l'avantage de donner une majoration simple et facile à calculer. Pour cette raison, on l'utilise souvent pour obtenir une première majoration simple et explicite, pour se donner une idée. Cependant, elle montre vite ses limites dès qu'on veut plus de précision, ou dès qu'on travaille avec des valeurs de n assez grandes (on parle alors de « grandes déviations ») ou des « a petits ». On dispose pour ces cas extrêmes d'inégalités plus précises.

Exemple : Si on avait demandé $P(\dots) \leqslant \dfrac{1}{1000}$ au lieu de $\dfrac{5}{100}$, on aurait trouvé (exo) $n \geqslant 1\,388\,889$, tandis qu'avec l'inégalité de concentration de Bernstein (cf DS d'il y a deux ans) :

$$P\left(\left|\frac{X_n}{n} - p\right| \geqslant a\right) \leqslant 2e^{-na^2/4}$$

on aurait obtenu (exo) $n \geqslant 304\,037$.

En conclusion, l'inégalité de Bienaymé-Tchebychev (et celle de Markov dont elle découle) a surtout une importance théorique et est en général utilisée en pratique :

- pour donner une première majoration explicite (pas forcément optimale).
- pour montrer que des probas tendent vers 0 : par exemple, si on a une inégalité $P(X \geqslant 2\lambda) \leqslant \dfrac{1}{\lambda}$ alors celle-ci est suffisante pour montrer que $P(X \geqslant 2\lambda) \xrightarrow[\lambda \to +\infty]{} 0$.
- et surtout : pour montrer la convergence en proba. Cette notion étant HP, nous nous contenterons d'un exemple qui nous permettra également de justifier le bien-fondé du modèle probabiliste que nous avons introduit dans le chapitre précédent.

Théorème (HP - Loi faible des Grands Nombres). Soit $(X_n)_{n\geqslant 1}$ une suite de v.a. indépendantes et $p \in [\,0\,;1\,]$. On suppose que, pour tout $n \geqslant 1$, $X_n \sim B(p)$. Alors, pour tout $\varepsilon > 0$:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - p\right| \geqslant \varepsilon\right) \xrightarrow[n \to +\infty]{} 0$$

Démonstration. Soit $\varepsilon > 0$. Soit $n \geqslant 1$. Notons $S_n = X_1 + \dots + X_n$. Alors $S_n \sim B(n, p)$ si bien que :

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) &= P\left(|S_n - np| \geqslant n\varepsilon\right) \\ &= P\left(|S_n - E(S_n)| \geqslant n\varepsilon\right) \\ &\leqslant \frac{V(S_n)}{n^2\varepsilon^2} \end{aligned}$$

d'après l'inégalité de Bienaymé-Tchebychev. Ainsi,

$$0 \leqslant P\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}$$

□

et le théorème d'encadrement permet de conclure.

Remarque : Cohérence du modèle probabiliste.

On peut voir S_n comme le nombre de succès lors de n répétitions indépendantes d'une expérience aléatoire où la proba de succès est p. $\dfrac{S_n}{n}$ est alors la fréquence d'apparition du succès lors des n premiers lancers, et on vient de montrer que

$$\frac{S_n}{n} \xrightarrow[n \to +\infty]{} p$$

C'est la définition intuitive d'une proba (cf. chapitre 26) : si A est un événement, on veut poser $P(A)$ égale à la limite de la quantité

$$\frac{\text{nombre de fois où } A \text{ est réalisé en } n \text{ tentatives}}{n}$$

et on vient de voir que cette définition intuitive est validée par le calcul !