

Variables aléatoires sur un univers fini

Dans tout ce chapitre $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ désigne un espace probabilisé fini et E désigne un ensemble non vide.

I Notion de variable aléatoire finie

1) Introduction

Lorsqu'on étudie un phénomène, on est amené à étudier des variables liées à ce phénomène. De façon informelle, une variable aléatoire réelle est une quantité liée à une expérience aléatoire dont la valeur dépend exclusivement du résultat ω de cette expérience. On peut donc voir cela comme un objet qui, à une éventualité ω , associe un élément dépendant de ω et qu'on notera $X(\omega)$. Mathématiquement, une variable aléatoire est donc une fonction définie sur l'univers Ω associé à l'expérience.

Quelques exemples de variables aléatoires :

- la somme des chiffres obtenus en lançant deux dés.
- le numéro maximal obtenu lorsqu'on tire des boules numérotées dans une urne.
- le nombre de lancers de dés qui sont nécessaires pour obtenir le premier Pile.
- le nombre moyen d'appels reçus par un standard téléphonique en une heure.
- la durée de vie d'une ampoule.

2) Définition

Définition (variable aléatoire). On appelle variable aléatoire X sur Ω et à valeurs dans E toute application définie sur Ω et à valeurs dans E .

Lorsque $E = \mathbb{R}$, on dit que X est une variable aléatoire réelle.

Définition (univers image ou support). Soit X une variable aléatoire sur Ω . L'ensemble $X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$ des valeurs prises par la variable aléatoire X est appelé univers image ou support de X .

Remarques :

- Contrairement à ce que son nom indique, une variable aléatoire sur Ω et à valeurs dans E n'est pas aléatoire. Ce n'est pas non plus une variable mais une fonction. Elle envoie tout élément de Ω sur un élément fixe de E . L'aléatoire ne va apparaître que lorsqu'on calculera la probabilité qu'elle prenne telle ou telle valeur (cf. paragraphe II).
- Soit X une variable aléatoire sur Ω à valeurs dans E .
 - ★ On a bien sûr $X(\Omega) \subset E$. Et il n'y a aucune raison en général pour que $E = X(\Omega)$, c'est-à-dire pour que X soit surjective.
 - ★ Comme Ω est un ensemble fini, $X(\Omega)$ l'est aussi. En notant $n = \text{card}(X(\Omega))$, on peut écrire $X(\Omega) = \{x_1; \dots; x_n\}$, où les x_i , $1 \leq i \leq n$ sont des éléments de E distincts
 - ★ Si $\omega \in \Omega$, on dit que $X(\omega)$ est une réalisation de X . Mathématiquement c'est juste un élément de E .
- La plupart des variables aléatoires rencontrées dans ce chapitre sont à valeurs dans \mathbb{R} ou plus généralement dans \mathbb{R}^n .

Dans ce chapitre, on se limite aux variables aléatoires définies sur un univers fini et donc ne prenant qu'un nombre fini de valeurs. Nous ne pourrions donc pas traiter les trois derniers exemples cette année (et le dernier n'est même pas au programme de deuxième année).

On note souvent v.a. en abrégé.

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.

On dit qu'une variable aléatoire sur Ω est finie lorsqu'elle ne prend qu'un nombre fini de valeurs. C'est automatique dans ce chapitre puisque Ω est fini.

Exemples :

- On lance deux dés à 6 faces. On peut considérer l'univers $\Omega = \llbracket 1 ; 6 \rrbracket^2$. L'application

est une variable aléatoire réelle qui représente la somme des chiffres des deux faces obtenues. On a alors $X(\Omega) = \llbracket 2 ; 12 \rrbracket$.

- On lance une pièce de monnaie. On peut prendre $\Omega = \{P; F\}$. L'application

est une variable aléatoire réelle. On a $T(\Omega) = \{0; 1\}$.

- Soient p et n des entiers naturels tels que $2 \leq p \leq n$. On dispose d'une urne contenant n boules numérotées de 1 à n . On tire successivement et sans remise p boules. Un univers associé à cette expérience est Ω l'ensemble des p -uplets d'éléments distincts de $\llbracket 1 ; n \rrbracket$. L'application

est une variable aléatoire réelle qui représente le numéro maximal parmi les p boules tirées. On a alors $M(\Omega) = \square$.

- Soient p et n des entiers naturels non nuls. On dispose d'une urne contenant n boules numérotées de 1 à n . On tire successivement et avec remise p boules. Un univers associé à cette expérience est $\Omega = \llbracket 1 ; n \rrbracket^p$. L'application

est une variable aléatoire à valeurs dans \mathbb{R}^2 qui représente le couple formé par le numéro minimal et le numéro maximal parmi les p boules tirées. On a alors $Z(\Omega) = \{(x, y) \in \llbracket 1 ; n \rrbracket^2 \mid x \leq y\}$. En effet :

Une variable aléatoire ne prenant que les deux valeurs 0 et 1 est appelée variable aléatoire de Bernoulli. Nous en reparlerons dans le paragraphe II.2.c.

Dans le pire des cas, on tire les p boules avec les plus petites valeurs et la valeur maximum parmi elles est p .

Remarques :

- Le dernier exemple présente une variable aléatoire à valeurs dans \mathbb{R}^2 . Plus généralement, une variable aléatoire à valeurs dans un produit cartésien est appelé couple de variables aléatoires. On en reparlera plus longuement dans le paragraphe II.5.
- On construit rarement les variables aléatoires explicitement comme on l'a fait dans les exemples ci-dessus. La raison principale est qu'il est souvent difficile de construire Ω (comme on l'a dit dans le chapitre précédent). Souvent les variables aléatoires sont définies par des phrases.

Par exemple :

- Notons S la somme des chiffres des deux faces obtenues en lançant deux dés.
- Notons M le numéro maximal quand on tire p boules sans remise dans une urne contenant n boules numérotées de 1 à n .

Les énoncés se contentent alors d'admettre qu'il s'agit de variables aléatoires. Il n'y a en général pas trop de doute sur le fait qu'une variable aléatoire décrite par une phrase en est bien une. Mais parfois c'est un peu plus subtile que ça :

II Loi d'une variable aléatoire

Dans toute la suite de ce chapitre, X désigne une variable aléatoire sur Ω et à valeurs dans un ensemble E .

Entre en scène à présent la probabilité \mathbb{P} de l'espace probabilisé sur lequel on s'est placé.

1) Notion de loi d'un variable aléatoire

a) Définition

Pour simplifier les notations, lorsque $A \in \mathcal{P}(E)$ et $a \in E$, on note plutôt $\mathbb{P}(X \in A)$ et $\mathbb{P}(X = a)$ au lieu de $\mathbb{P}([X \in A])$ et $\mathbb{P}([X = a])$ respectivement.

Proposition/Définition. L'application

$$\mathbb{P}_X : \begin{cases} \mathcal{P}(X(\Omega)) & \longrightarrow [0; 1] \\ A & \longmapsto \mathbb{P}(X \in A) \end{cases}$$

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



On a donc

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)),$$

pour tout $A \in \mathcal{P}(X(\Omega))$.

DÉMONSTRATION.

- \mathbb{P} étant une probabilité, elle est à valeurs dans $[0; 1]$ donc \mathbb{P}_X aussi.
- $\mathbb{P}_X(X(\Omega)) = \mathbb{P}(X \in X(\Omega)) = \mathbb{P}(\Omega) = 1$.
- Soient A et B des événements de $\mathcal{P}(X(\Omega))$ qui sont incompatibles. Alors $\mathbb{P}_X(A \cup B) = \mathbb{P}(X \in A \cup B)$. On a montré que $X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B)$ et $X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B)$ dans le chapitre 15 si bien que :
 - ★ A et B étant incompatibles, $X^{-1}(A) \cap X^{-1}(B) = X^{-1}(\emptyset) = \emptyset$, c'est-à-dire $X^{-1}(A)$ et $X^{-1}(B)$ sont incompatibles.
 - ★ Par additivité de \mathbb{P} , on a

$$\begin{aligned} \mathbb{P}_X(A \cup B) &= \mathbb{P}(X^{-1}(A) \cup X^{-1}(B)) \\ &= \mathbb{P}(X^{-1}(A)) + \mathbb{P}(X^{-1}(B)) \\ &= \mathbb{P}_X(A) + \mathbb{P}_X(B). \end{aligned}$$

Ainsi \mathbb{P}_X est une probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$. □

b) Premières propriétés

Proposition (caractérisation de la loi d'une variable aléatoire). La loi de X est entièrement déterminée par la distribution de probabilités $(\mathbb{P}(X = x))_{x \in X(\Omega)}$. De plus, pour tout $A \in \mathcal{P}(E)$,

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

DÉMONSTRATION. Comme on l'a vu dans le chapitre précédent, la probabilité \mathbb{P}_X est entièrement caractérisée par son image en les événements élémentaires et

$$\forall A \in \mathcal{P}(X(\Omega)), \quad \mathbb{P}_X(A) = \sum_{x \in A} \mathbb{P}_X(\{x\}) = \sum_{x \in A} \mathbb{P}(X = x).$$

On conclut en remarquant que, si $x \notin X(\Omega)$, alors $\mathbb{P}(X = x) = \mathbb{P}(\emptyset) = 0$. □

Pour donner la loi d'une variable aléatoire X , on peut donc donner une formule générale pour $\mathbb{P}(X = x)$ pour tout $x \in X(\Omega)$. 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 $\mathbb{P}(X = x)$ 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 $\Omega = \llbracket 1; 6 \rrbracket^2$ de la probabilité uniforme). Déterminons sa loi : on a $X(\Omega) = \llbracket 2; 12 \rrbracket$ et, si $k \in \llbracket 2; 12 \rrbracket$,

$$\mathbb{P}(X = k) = \frac{\text{card}([X = k])}{\text{card}(\Omega)} = \frac{\text{card}([X = k])}{36}. \text{ On a}$$

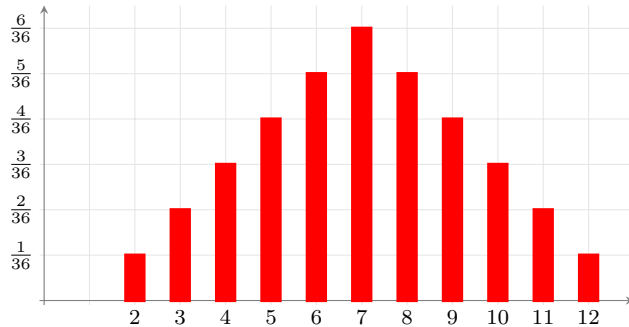
$$[X = 2] = \{(1, 1)\}, \quad [X = 3] = \{(1, 2); (2, 1)\},$$

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

donc on obtient le tableau :

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

On remarque que, pour tout $k \in \llbracket 2; 12 \rrbracket$, $\mathbb{P}(X = k) = \frac{6 - |7 - k|}{36}$. On peut aussi tracer son diagramme en bâtons :



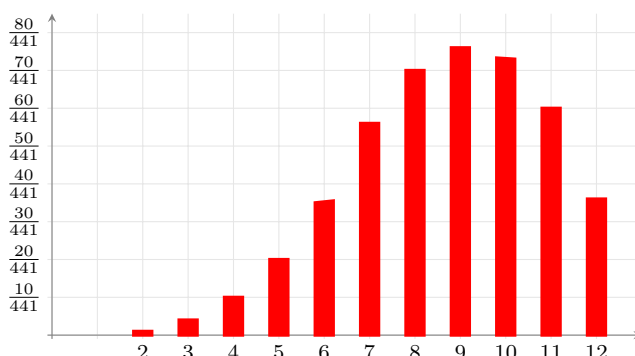
On remarque que le diagramme est symétrique par rapport à 7 : c'est un cas particulier de loi triangulaire que l'on verra en exercices.

! La loi de X dépend de la probabilité \mathbb{P} ! Par exemple, le tableau ci-dessus a été construit en supposant que le dé est équilibré, c'est-à-dire en supposant que \mathbb{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 $k \in \llbracket 1; 6 \rrbracket$ est proportionnelle à k . On avait calculé que, pour tout $k \in \llbracket 1; 6 \rrbracket$, $\mathbb{P}(X = k) = \frac{k}{21}$. Les ensembles $([X = k])_{2 \leq k \leq 12}$ 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
$\mathbb{P}(X = x)$	$\frac{1}{441}$	$\frac{4}{441}$	$\frac{10}{441}$	$\frac{20}{441}$	$\frac{35}{441}$	$\frac{56}{441}$	$\frac{70}{441}$	$\frac{76}{441}$	$\frac{73}{441}$	$\frac{60}{441}$	$\frac{36}{441}$

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 probabilité \mathbb{P} sur $X(\Omega)$ qui a changé.

Revenons maintenant sur la formule

$$\forall A \in \mathcal{P}(E), \quad \mathbb{P}(X \in A) = \sum_{x \in A \cap X(\Omega)} \mathbb{P}(X = x).$$

Si on prend $A = X(\Omega)$, on obtient :

Corollaire. On a $\sum_{x \in X(\Omega)} \mathbb{P}(X = x) = 1$. Autrement dit la famille $(\mathbb{P}(X = x))_{x \in X(\Omega)}$ est une distribution de probabilités sur $X(\Omega)$.

ou encore

$$\sum_{x \in X(\Omega)} \mathbb{P}_X(\{x\}) = 1.$$

Intéressons-nous maintenant au cas particulier où $X(\Omega) \subset \llbracket m; n \rrbracket$ avec m et n des entiers relatifs tels que $m \leq n$. Pour tout $k \in \llbracket m; n \rrbracket$, on a

$$\mathbb{P}(X \leq k) = \boxed{}, \quad \mathbb{P}(X \geq k) = \boxed{}.$$

Mais cela fonctionne même avec un élément n'appartenant pas à $X(\Omega)$.

Par exemple, si $X(\Omega) = \llbracket 1; n \rrbracket$, alors $\mathbb{P}(X \leq \pi) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3)$.

Inversement, on peut retrouver la loi de X en connaissant la famille $(\mathbb{P}(X \leq k))_{m \leq k \leq n}$. En effet, pour tout $k \in \llbracket m; n \rrbracket$, $[X = k] = [X \leq k] \setminus [X \leq k - 1]$ donc

$$\boxed{}$$

Dans le même ordre d'idée, on a $\boxed{}$ et aussi

$$\boxed{}$$

Ces formules sont très utiles si on cherche notamment la loi d'un maximum ou d'un minimum :

Exemple : Soient n et p dans \mathbb{N}^* . On dispose d'une urne contenant n boules numérotées de 1 à n . On tire successivement et avec remise p boules dans l'urne et on note X le plus grand numéro tiré. Déterminons sa loi.

Si X est une variable aléatoire réelle, la fonction $F_X : t \in \mathbb{R} \mapsto \mathbb{P}(X \leq t)$ est appelée fonction de répartition de X . Elle caractérise totalement la loi de X .

Et donc X est l'application qui à un $(\omega_1, \dots, \omega_p) \in \Omega$ associe $\max_{1 \leq i \leq p} \omega_i$.

c) Construction d'une variable aléatoire de loi donnée

On a vu dans le paragraphe précédent que, si X est une variable aléatoire sur Ω , alors la famille $(\mathbb{P}(X = x))_{x \in X(\Omega)}$ est une distribution de probabilités sur $X(\Omega)$. On peut se poser la question inverse : étant donnée une distribution de probabilité sur un ensemble fini E , existe-t-il un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et une variable aléatoire X sur Ω et à valeurs dans E telle que cette distribution de probabilité soit la loi de X ?

Théorème. Soit E un ensemble fini non vide. Soit $(p_e)_{e \in E}$ une distribution de probabilité sur E . Il existe un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ fini et une variable aléatoire $X : \Omega \rightarrow E$ telle que, pour tout $e \in E$, $\mathbb{P}(X = e) = p_e$.

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

DÉMONSTRATION. Il suffit de prendre $\Omega = E$ et de munir Ω de $\mathcal{P}(\Omega)$ et de la probabilité \mathbb{P} telle que, pour tout $e \in E$, $\mathbb{P}(\{e\}) = p_e$ (cela définit entièrement la probabilité \mathbb{P} d'après le chapitre précédent). On pose enfin $X = \text{Id}_E$. Alors X est bien définie sur Ω et à valeurs dans E et, pour tout $e \in \Omega = E$,

$$\mathbb{P}_X(\{e\}) = \mathbb{P}(X = e) = \mathbb{P}(\{e\}) = p_e. \quad \square$$

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

Dans le paragraphe II.2.d, nous dirons que X suit alors une loi binomiale de paramètres n et $\frac{1}{2}$.

$$\forall k \in \llbracket 0; n \rrbracket, \quad \mathbb{P}(X = k) = \frac{1}{2^n} \binom{n}{k}.$$

Comme on l'a dit, dans certains cas, l'univers Ω est un ensemble très compliqué mais heureusement le théorème précédent nous permet d'affirmer son existence sans nous poser de questions : nous n'explicitons plus Ω (sauf cas d'équiprobabilité) dans la suite. Nous savons qu'il existe, et ça nous suffit.

Nous pourrions par exemple dire : soit X une variable aléatoire définie par $\mathbb{P}(X = 0) = p$ et $\mathbb{P}(X = 1) = 1 - p$ avec $p \in [0; 1]$.

d) Égalité en loi

Lorsqu'on étudie une variable aléatoire, on finit par oublier l'univers sur lequel elle est définie et on s'intéresse à sa loi. Mais ce n'est pas parce que deux variables aléatoires ont la même loi qu'elles sont égales en tant qu'application. Illustrons cela par un exemple.

Exemple : Réalisons deux expériences aléatoires. D'une part on lance deux fois une pièce de monnaie : on prend $\Omega_1 = \{P; F\}^2$ muni de $\mathcal{P}(\Omega)$ et de l'équiprobabilité. Notons

$$X : \begin{cases} \Omega_1 & \longrightarrow \mathbb{R} \\ (a, b) & \longmapsto \begin{cases} 1 & \text{si } a = P \text{ et } b = P \\ 2 & \text{si } a = P \text{ et } b = F \\ 3 & \text{si } a = F \text{ et } b = P \\ 4 & \text{si } a = F \text{ et } b = F \end{cases} \end{cases}$$

Dès lors la loi de X est caractérisée par : $X(\Omega) = \llbracket 1; 4 \rrbracket$ et

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \mathbb{P}(X = 3) = \mathbb{P}(X = 4) = \frac{1}{4}.$$

D'autre part, on prend une carte au hasard dans un jeu de 52 cartes. On prend Ω_2 l'ensemble des cartes que l'on munit de $\mathcal{P}(\Omega_2)$ et de l'équiprobabilité. Notons

$$Y : \begin{cases} \Omega_2 & \longrightarrow \mathbb{R} \\ \omega & \longmapsto \begin{cases} 1 & \text{si } \omega \text{ est un coeur} \\ 2 & \text{si } \omega \text{ est un carreau} \\ 3 & \text{si } \omega \text{ est un trèfle} \\ 4 & \text{si } \omega \text{ est un pique} \end{cases} \end{cases}$$

Alors la loi de Y est caractérisée par : $Y(\Omega) = \llbracket 1; 4 \rrbracket$ et

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = 2) = \mathbb{P}(Y = 3) = \mathbb{P}(Y = 4) = \frac{1}{4}.$$

Par conséquent $\mathbb{P}_X = \mathbb{P}_Y$ (puisque une loi est entièrement caractérisée par la donnée des probabilités de chacun des événements élémentaires), autrement dit X et Y ont la même loi. Pourtant, en tant qu'applications, elles ne sont pas égales puisqu'elles ne sont pas définies sur le même ensemble (ici les deux univers n'ont même absolument rien à voir)

Définition. Si X et Y sont deux variables aléatoires (non forcément définies sur le même univers) telles que $\mathbb{P}_X = \mathbb{P}_Y$, alors on note $X \sim Y$.

Remarques :

- « avoir la même loi » est bien sûr une relation d'équivalence.
- Dans l'exemple ci-dessus, X et Y ne sont pas définies sur le même univers. Mais, même si c'est le cas, il se peut tout à fait qu'elles ne prennent jamais la même valeur en même temps.

Exemple : Soit X une variable aléatoire dont la loi est donnée par $X(\Omega) = \{0; 1\}$ et $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$. Posons $Y = 1 - X$. Alors $Y(\Omega) = \{0; 1\}$ et

$$\mathbb{P}(Y = 0) = \mathbb{P}(1 - X = 0) = \mathbb{P}(X = 1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = \frac{1}{2}.$$

Par conséquent $X \sim Y$ alors que X prend la valeur 0 quand Y prend la valeur 1 et vice versa.

2) Lois usuelles

a) Loi certaine

Définition. On dit qu'une variable aléatoire X suit une loi certaine si $X(\Omega)$ est un singleton.

Remarques :

- En notant m l'unique élément de $X(\Omega)$, on a alors $\mathbb{P}(X = m) = 1$.
- Mathématiquement, il s'agit simplement d'une application constante égale à m sur Ω
- Selon les conventions, on dit aussi qu'une variable aléatoire X suit une loi certaine s'il existe $m \in X(\Omega)$ tel que $\mathbb{P}(X = m) = 1$. Dans ce cas, pour tout $x \in X(\Omega) \setminus \{m\}$ (si cet ensemble n'est pas vide), $\mathbb{P}(X = x) = 0$. On parle parfois de variable aléatoire presque certaine ou presque sûrement égale à m . Mais il est plus pratique de « retirer de l'univers image » les éléments ayant une probabilité nulle.

On rappelle que

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

b) Loi uniforme

Définition. Soit E un ensemble fini. On dit qu'une variable aléatoire X suit la loi uniforme sur E si $X(\Omega) = E$ et

$$\forall x \in E, \quad \mathbb{P}(X = x) = \frac{1}{\text{card}(E)}.$$

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

Comme pour la loi certaine, si $X(\Omega) = E$ et s'il existe $A \subsetneq E$ tel que $\mathbb{P}(X = x) = \frac{1}{\text{card}(A)}$ pour tout $x \in A$, alors 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 A et que $X \sim \mathcal{U}(A)$.

Remarque : Si $A \subset E$, alors

$$\mathbb{P}(X \in A) =$$

La loi uniforme est donc la loi que l'on rencontre dans les situations d'équiprobabilité : toutes les valeurs prises par une variable aléatoire X qui suit une loi uniforme sont équiprobables.

Exemples :

- On lance un dé équilibré à 6 faces. La variable aléatoire X qui donne le résultat obtenu suit la loi uniforme sur $\llbracket 1; 6 \rrbracket$.
- Soit $n \in \mathbb{N}^*$. On dispose d'une urne contenant n boules numérotées de 1 à n . On tire une boule au hasard. La variable aléatoire X qui donne le numéro de la boule tirée suit la loi uniforme sur $\llbracket 1; n \rrbracket$.

Exemple : Soit $n \in \mathbb{N}^*$. Si $X \sim \mathcal{U}(\llbracket 1; n \rrbracket)$, alors, pour tout $k \in \llbracket 1; n \rrbracket$,

$$\mathbb{P}(X \leq k) =$$

c) Loi de Bernoulli

On dit qu'une variable aléatoire qui ne prend que deux valeurs 0 et 1 (avec probabilités non nulles) suit une loi de Bernoulli. Plus précisément :

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

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

On note $X \sim \mathcal{B}(p)$. Le paramètre p (respectivement $q = 1 - p$) est appelé probabilité de succès (respectivement d'échec).



Il n'y a qu'à calculer $\mathbb{P}(X = 1)$ pour calculer le paramètre p . Il est automatique que $\mathbb{P}(X = 0) = 1 - p$ ensuite. Le contraire est vrai aussi bien sûr.

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 $X \sim \mathcal{B}(p)$.
- Si $X \sim \mathcal{B}(p)$, alors $X^2 \sim \mathcal{B}(p)$. En effet, comme X ne prend que 0 et 1 pour valeurs, on a $X^2 = X$.
- Si $X \sim \mathcal{B}(p)$, alors $1 - X \sim \mathcal{B}(1 - p)$. En effet :

⋮

- Soit A un événement tel que $\mathbb{P}(A) \in]0; 1[$. On définit $\mathbb{1}_A$ la variable aléatoire définie sur Ω par

$$\mathbb{1}_A : \begin{cases} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases} \end{cases}$$

Alors $\mathbb{1}_A$ est une variable aléatoire de loi de Bernoulli de paramètre $\mathbb{P}(A)$.

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. 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 qui ne prend que les valeurs a et b (avec probabilités non nulles), alors $Y = \frac{X - a}{b - a} \sim \mathcal{B}(p)$ avec $p = \mathbb{P}(X = b)$.

d) Loi binomiale

Définition. Soient $n \in \mathbb{N}^*$ et $p \in]0; 1[$. On dit qu'une variable aléatoire X suit la loi de binomiale de paramètres n et p si $X(\Omega) = \llbracket 0; n \rrbracket$ et :

$$\forall k \in \llbracket 0; n \rrbracket, \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

On note $X \sim \mathcal{B}(n, p)$. Le paramètre p (respectivement $q = 1 - p$) est appelé probabilité de succès (respectivement d'échec).

Remarque : $\mathcal{B}(1, p)$ est la loi $\mathcal{B}(p)$.

Définition. Soit $p \in]0; 1[$. On appelle expérience (ou épreuve ou schéma) de Bernoulli de paramètre p toute expérience aléatoire n'ayant que deux issues possibles :

- l'une appelée succès et dont la probabilité de réalisation est p ,
- l'une appelée échec et dont la probabilité de réalisation est $q = 1 - p$.

Exemple : On lance une pièce de monnaie qui tombe sur pile avec probabilité $p \in]0; 1[$. On gagne le jeu si on tombe sur Pile. L'expérience est alors une épreuve de Bernoulli de probabilité de succès p .

Proposition. Soit $n \in \mathbb{N}^*$. On considère une épreuve de Bernoulli dont la probabilité de succès est $p \in]0; 1[$. Notons X la variable aléatoire comptant le nombre de succès dans n répétitions **indépendantes** de cette épreuve de Bernoulli. Alors X suit la loi binomiale de paramètres n et p .

DÉMONSTRATION.

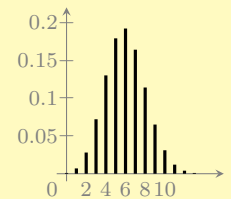


Il s'agit bien d'une loi de probabilité puisque, avec la formule du binôme de Newton, on vérifie que la somme des $\mathbb{P}(X = k)$ pour k allant de 0 à n , est égale à

$$(p + (1 - p))^n = 1.$$



Pour $n = 20$ et $p = 0,3$, on obtient le diagramme en bâtons :



Une telle probabilité existe-t-elle? Oui mais, pour le moment, nous allons l'admettre et nous en reparlerons dans le paragraphe III.1.b.

□

Ainsi, pour montrer qu'une variable aléatoire X suit une loi binomiale, on a deux possibilités :

- Soit on calcule explicitement la loi de X : on montre qu'il existe $n \in \mathbb{N}^*$ et $p \in]0; 1[$ tels que $X(\Omega) = \llbracket 0; n \rrbracket$ et, pour tout $k \in \llbracket 0; n \rrbracket$,


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

La variable aléatoire X suit alors la loi $\mathcal{B}(n, p)$.

- Soit on justifie que X compte le nombre de succès dans n répétitions indépendantes d'épreuves de Bernoulli de probabilité de succès p .

Exemples :

- On dispose d'une pièce de monnaie truquée de telle sorte que la probabilité qu'elle tombe sur Pile soit $p \in]0; 1[$. On lance $n \in \mathbb{N}^*$ fois la pièce successivement et on note X le nombre de fois où la pièce est tombée sur Pile. On peut supposer que les lancers successifs sont indépendants. La variable aléatoire X représente alors le nombre de succès dans n répétitions indépendantes d'épreuves de Bernoulli. Ainsi $X \sim \mathcal{B}(n, p)$.

 Raisonnement classique !

- On dispose d'une urne contenant r boules rouges et b boules bleues (avec $r \in \mathbb{N}^*$ et $b \in \mathbb{N}^*$). On effectue n tirages successifs avec remise de boules dans l'urne et on note X le nombre de boules rouges obtenues. On peut supposer que les tirages successifs sont indépendants (car il y a remise). Tirer une boule rouge dans l'urne est une épreuve de Bernoulli de paramètre $p = \frac{r}{r+b}$. La variable X compte le nombre de succès dans n répétitions indépendantes de cette épreuve de Bernoulli. Ainsi $X \sim \mathcal{B}\left(n, \frac{r}{r+b}\right)$.

Il est impératif de justifier que les épreuves de Bernoulli sont **indépendantes** !

3) Loi d'un transfert de variable aléatoire

Définition. Soit X une variable aléatoire sur Ω et à valeurs dans E . Soit f une fonction définie sur $X(\Omega)$. Alors

$$Y : \begin{cases} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & f(X(\omega)) \end{cases}$$

est une variable aléatoire notée $f(X)$. On dit que $f(X)$ est un transfert de X .

On a aussi $Y = f \circ X$.

Remarque : Si $Y = f(X)$, alors $Y(\Omega) = \{f(x) \mid x \in X(\Omega)\}$ mais il se peut que f ne soit pas injective et donc que les $f(x)$, $x \in X(\Omega)$, ne soient pas tous distincts.

Proposition. Soit X une variable aléatoire sur Ω et à valeurs dans E . Soit f une fonction définie sur $X(\Omega)$. Notons $Y = f(X)$. Alors, pour tout $y \in Y(\Omega)$,

$$\mathbb{P}(Y = y) = \sum_{\substack{x \in X(\Omega) \\ f(x)=y}} \mathbb{P}(X = x).$$



Cette formule n'est pas explicitement au programme mais nous allons l'utiliser plusieurs fois dans de futures démonstrations. On la comprend bien via les exemples ci-dessous.

DÉMONSTRATION.

□

Voyons quelques cas particuliers :

- Si f est bijection sur $X(\Omega)$ alors, pour tout $y \in Y(\Omega)$,

C'est notamment le cas si $f : x \mapsto ax + b$ avec $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$. Si $y \in Y(\Omega)$, alors

- Si $f : x \mapsto |x|$, alors $Y(\Omega) = \{|x| \mid x \in X(\Omega)\}$. Si $y \in \mathbb{R}$, alors

- Si $f : x \mapsto x^2$, alors $Y(\Omega) = \{x^2 \mid x \in X(\Omega)\}$. Si $y \in \mathbb{R}$, alors

Exemple : On lance deux dés à 6 faces bien équilibrés. Notons X la somme des chiffres des dés et Y la différence des chiffres du premier dé avec le second. Posons $Z = X - 7$ et $U = |Y| + 1$. Nous avons $Y(\Omega) = \llbracket -5; 5 \rrbracket$ et $Z(\Omega) = \llbracket -5; 5 \rrbracket$. Pour tout $k \in \llbracket -5; 5 \rrbracket$, $\mathbb{P}(Z = k) = \mathbb{P}(X = 7 + k)$:

x	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\mathbb{P}(Z = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\mathbb{P}(Y = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



Nous laissons le détail du calcul de la loi de Y en exercice. Par exemple il y a 5 façons d'obtenir -1 : $5 - 6$, $4 - 5$, $3 - 4$, $2 - 3$ et $1 - 2$.

On remarque au passage que $Y \sim Z$ ont la même loi. Ensuite $U(\Omega) = \llbracket 1; 6 \rrbracket$, $\mathbb{P}(U = 1) = \mathbb{P}(Y = 0) = \frac{6}{36}$ et, si $k \in \llbracket 2; 6 \rrbracket$,

$$\mathbb{P}(U = k) = \mathbb{P}(|Y| = k - 1) = \mathbb{P}(Y = k - 1) + \mathbb{P}(Y = 1 - k).$$

On obtient alors :

x	1	2	3	4	5	6
$\mathbb{P}(U = x)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Proposition. Soient X et Y deux variables aléatoires à valeurs dans un même ensemble E (mais pas forcément définie sur le même univers). Soit f une fonction définie sur E . Si $X \sim Y$, alors $f(X) \sim f(Y)$.

Autrement dit, si X et Y ont la même loi, alors $f(X)$ et $f(Y)$ ont toujours la même loi.

DÉMONSTRATION. Supposons que $X \sim Y$. On a alors $X(\Omega) = Y(\Omega)$ donc $f(X)(\Omega) = f(Y)(\Omega)$. Soit z dans cet ensemble image commun. On a :

$$\mathbb{P}(f(X) = z) = \sum_{\substack{x \in X(\Omega) \\ f(x)=z}} \mathbb{P}(X = x) = \sum_{\substack{x \in Y(\Omega) \\ f(x)=z}} \mathbb{P}(Y = x) = \mathbb{P}(f(Y) = z).$$

Ainsi $f(X) \sim f(Y)$. □

Sans calculer la loi précisément, on sera amené à utiliser des égalités probabilistes faisant intervenir des transferts :

Exemples :

- Soit X une variable aléatoire réelle. Puisque \exp est strictement croissante sur \mathbb{R} ,

$$\mathbb{P}(e^{-X} \leq 1) = \mathbb{P}(-X \leq 0) = \mathbb{P}(X \geq 0).$$

Aucun problème pour affirmer cela directement mais, dans le doute autant savoir le montrer en raisonnant sur les ensembles :

- Soit X une variable aléatoire réelle. Soit f une fonction croissante sur \mathbb{R} . On a alors $[X \geq 0] \subset [f(X) \geq f(0)]$ donc $\mathbb{P}(X \geq 0) \leq \mathbb{P}(f(X) \geq f(0))$. Là encore on peut affirmer cela directement mais il faut savoir comment le démontrer rigoureusement :

4) Loi conditionnelle

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

$$\mathbb{P}_B : \begin{cases} \mathcal{P}(X(\Omega)) & \longrightarrow [0; 1] \\ A & \longmapsto \mathbb{P}_B(X \in A) \end{cases}$$

Rappelons que pour conditionner par un événement, il est indispensable que celui-ci ait une probabilité non nulle. De plus, d'après le chapitre 34, \mathbb{P}_B est bien une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

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 procède comme une loi « normale » en donnant l'univers image et les probabilités de chaque élément de l'univers image sachant que l'événement B est réalisé (ou on donne la réponse sous la forme d'une loi de référence, cf. exemple ci-dessous).

Exemple : Soit $n \in \mathbb{N}^*$. Soit $X \sim \mathcal{U}(\llbracket 1 ; 2n \rrbracket)$. Déterminons la loi conditionnelle de X sachant que l'événement $[X \leq n]$ est réalisé.

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 $X \leq n$, 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.

5) Loi d'un couple de variables aléatoires

Dans ce paragraphe X et Y désignent des variables aléatoires sur Ω toutes les deux et à valeurs dans E et F respectivement.

a) Loi conjointe d'un couple de variable aléatoire


Le couple de variables aléatoires (X, Y) est une variable aléatoire de Ω à valeurs dans $E \times F$. En effet, l'application $(X, Y) : \omega \rightarrow (X(\omega), Y(\omega))$ est une fonction de Ω dans $E \times F$. Inversement, si Z est une variable aléatoire sur Ω à valeurs dans $E \times F$, alors :

- l'application X qui à tout $\omega \in \Omega$ associe la première coordonnée de $Z(\omega)$ est une variable aléatoire sur Ω à valeurs dans E .
- l'application Y qui à tout $\omega \in \Omega$ associe la deuxième coordonnée de $Z(\omega)$ est une variable aléatoire sur Ω à valeurs dans F .

Puisqu'un couple 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 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 (loi conjointe). La loi d'un couple de variables aléatoires (X, Y) sur Ω est entièrement déterminée par la distribution de probabilité $(\mathbb{P}([X = x] \cap [Y = y]))_{(x,y) \in (X,Y)(\Omega)}$. Cette loi est appelée loi conjointe du couple (X, Y) et, pour tout $(x, y) \in (X, Y)(\Omega)$, on note $\mathbb{P}(X = x, Y = y)$ au lieu de $\mathbb{P}([X = x] \cap [Y = y])$.

DÉMONSTRATION. C'est tout simplement le dernier résultat du paragraphe 1.3 avec (X, Y) à la place de X et en remarquant que $\mathbb{P}((X, Y) = (x, y)) = \mathbb{P}([X = x] \cap [Y = y])$ pour tout $(x, y) \in (X, Y)(\Omega)$. \square

Remarque :  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

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 $\llbracket 1 ; n \rrbracket$.

En d'autres termes, un couple de variables aléatoires est une variable aléatoire dans l'ensemble produit.

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. Aussi, dans la pratique, on précise en général $X(\Omega) \times Y(\Omega)$ et on pose $\mathbb{P}([X = x] \cap [Y = y]) = 0$ lorsque $(x, y) \notin (X, Y)(\Omega)$.

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. Donnons la loi conjointe du couple (C_1, C_2) .

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

C'est une illustration de la remarque précédente : ici (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.

	$C_2(\Omega)$	
	B	N
$C_1(\Omega)$		
B		
N		

Remarque : Pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$ tel que $\mathbb{P}(X = x) \neq 0$, on a

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}_{[X=x]}(Y = y).$$

L'étude de la loi conjointe d'un couple passe donc souvent par l'étude des lois conditionnelles sachant que $[X = x]$ est réalisé, pour tout x tel que $\mathbb{P}(X = x) \neq 0$.

Exemple : Soient $p \in]0; 1[$, $N \in \mathbb{N} \setminus \{0; 1\}$ et $n \in \llbracket 1; N - 1 \rrbracket$. On dispose de N pièces de monnaie non équilibrées, chaque pièce pouvant amener Pile avec la probabilité p . On lance les N pièces de monnaie en l'air toutes ensemble. Un joueur a les yeux bandés et n'a pas assisté au lancer. Il choisit au hasard (sans les regarder bien sûr) n pièces parmi les N . Il gagne celles qui représentent Pile parmi celles qu'il a choisies. On note X le nombre de Piles obtenues en lancer des pièces et Y le nombre de pièces gagnées par le joueur.

Proposition. Soient E et F des ensembles finis non vides. Si $(p_{i,j})_{(i,j) \in E \times F}$ est une famille de réels positifs dont la somme (double) vaut 1, alors il existe un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ fini et deux variables aléatoires X et Y définies sur cet espace telles que $X(\Omega) = E, Y(\Omega) = F$ et :

$$\forall (i, j) \in E \times F, \quad \mathbb{P}(X = i, Y = j) = p_{i,j}.$$

DÉMONSTRATION. C'est simplement la proposition du paragraphe II.1.c appliquée au cas d'un variable aléatoire à valeurs dans $E \times F$. \square

Remarque : Un couple de variable aléatoire étant une variable aléatoire (à valeurs dans un espace produit), on peut tout à fait considérer un transfert : si f est définie sur $E \times F$, alors $Z = f(X, Y)$ est encore une variable aléatoire sur Ω . Pour tout $z \in Z(\Omega)$,

$$\mathbb{P}(Z = z) = \sum_{\substack{(x,y) \in (X,Y)(\Omega) \\ f(x,y)=z}} \mathbb{P}(X = x, Y = y).$$

D'ailleurs les marginales X est Y sont des transferts du couple par les fonctions $(x, y) \mapsto x$ et $(x, y) \mapsto y$ respectivement.

b) Lois marginales

Définition. Les lois de X et Y sont appelées lois marginales du couple (X, Y) .

Pour tout $x \in X(\Omega)$, la formule des probabilités totales appliquée au système complet d'événements associé à Y assure que :

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

Pour tout $y \in Y(\Omega)$, la formule des probabilités totales appliquée au système complet d'événements associé à X assure que :

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

Exemple : Reprenons l'exemple du paragraphe précédent. On connaît la loi de X et, grâce au calcul de la loi du couple, on peut trouver celle de 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.

Ces formules découlent aussi de la formule précédente concernant la loi d'un transfert en prenant tour à tour $f : (x, y) \mapsto x$ et $f : (x, y) \mapsto y$.

Voyons un autre exemple où l'on détermine la loi d'une variable aléatoire en la voyant comme une marginale d'un couple dont la loi conjointe est plus facile à déterminer au premier abord.

Exemple : Soit $n \in \mathbb{N} \setminus \{0; 1\}$. On dispose d'une urne contenant n boules numérotées de 1 à n . On tire deux boules simultanément et on note X le plus petit numéro et Y le plus grand numéro parmi les deux boules tirées. Déterminer les lois de X et Y en déterminant au préalable la loi de (X, Y) .

Parfois on utilise la loi d'une variable aléatoire pour déterminer la loi d'une autre sans forcément expliciter la loi du couple :

Exemple : Soit $n \in \mathbb{N} \setminus \{0; 1\}$. On dispose d'une urne contenant une boule numérotée 1, de 2 boules numérotées 2, de 3 boules numérotées 3, ..., de $n - 1$ boules numérotées $(n - 1)$ et enfin de n boules numérotées n . On tire une boule au hasard dans l'urne. Si c'est la boule numérotée 1, le jeu s'arrête là et on ne gagne rien. Si c'est une boule numérotée k (avec $k \in \llbracket 2; n \rrbracket$), on lance un dé équilibré à k faces. Si la face obtenue porte le numéro j , on gagne $j - 1$ euros. On introduit X la variable aléatoire égale au numéro de la boule tirée et Y la variable aléatoire égale au montant gagné à l'issue du jeu. Déterminons la loi de X , puis celle de Y .

Morale de l'histoire :

Théorème. La connaissance de la loi conjointe d'un couple (X, Y) de variables aléatoires permet toujours d'en déduire les lois marginales du couple.

On peut se demander si la réciproque est vraie, c'est-à-dire si la connaissance des lois marginales permettent de retrouver la loi conjointe. Et bien non en général.

Exemples :

- Si on reprend le premier exemple de ce paragraphe avec X le nombre de Piles obtenues en lançant des pièces et Y le nombre de pièces gagnées. On a obtenu que $X \sim \mathcal{B}(N, p)$ et $Y \sim \mathcal{B}(n, p)$. Pourtant la connaissance de ces deux lois, hors contexte, ne permet pas d'imaginer la loi conjointe obtenue : pour tout $(k, j) \in \llbracket 0; N \rrbracket \times \llbracket 0; n \rrbracket$,

C'est intuitif : connaître les variables aléatoires X et Y séparément ne dit pas comment elle se comportent l'une par rapport à l'autre.

$$\mathbb{P}(X = k, Y = j) = \begin{cases} \binom{n}{j} \binom{N-n}{k-j} p^k (1-p)^{N-k} & \text{si } j \leq k \leq N-n+j \\ 0 & \text{si } k < j \text{ ou } k > N-n+j. \end{cases}$$

Considérons (X', Y') un couple de variables aléatoires telles que $X'(\Omega) = \llbracket 0; N \rrbracket$, $Y'(\Omega) = \llbracket 0; n \rrbracket$ et

$$\forall (k, j) \in \llbracket 0; N \rrbracket \times \llbracket 0; n \rrbracket, \quad \mathbb{P}(X = k, Y = j) = \binom{N}{k} \binom{n}{j} p^{k+j} (1-p)^{N+n-k-j}.$$

On obtient aisément (je vous laisse vérifier) que c'est bien une distribution de probabilité (la somme vaut 1) et que $X' \sim \mathcal{B}(N, p)$ et $Y' \sim \mathcal{B}(n, p)$ (en utilisant la formule des probabilités totales). Ainsi $X \sim X'$, $Y \sim Y'$ alors que (X, Y) n'a pas la même loi que (X', Y') .

La loi conjointe proposée ci-contre est celle obtenue en supposant que X' et Y' sont des variables aléatoires indépendantes (cf. paragraphe II.1.a).

- Donnons un exemple plus simple. Si $X \sim \mathcal{B}(1/2)$ et $Y = X$, alors la loi du couple (X, Y) est donnée par le tableau suivant :

	$Y(\Omega)$	
	0	1
$X(\Omega)$		
0	1/2	0
1	0	1/2

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

	$Y(\Omega)$	
	0	1
$X(\Omega)$		
0	0	1/2
1	1/2	0

Dans les deux cas de figure X et Y suivent une loi $\mathcal{B}(1/2)$ mais les couples n'ont pas la même loi.

c) Généralisation aux n -uplets de variables aléatoires

Soit $n \geq 3$. On peut généraliser au cas d'un n -uplet de variables aléatoires :

- Si, pour tout $i \in \llbracket 1; n \rrbracket$, 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 $E_1 \times \dots \times E_n$.
- La loi de (X_1, \dots, X_n) est entièrement déterminée par la distribution de probabilité

$$\mathbb{P} \left(\bigcap_{i=1}^n [X_i = x_i] \right), \quad (x_1, \dots, x_n) \in (X_1, \dots, X_n)(\Omega). \text{ On l'appelle la loi conjointe.}$$

- On note encore $\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ au lieu de $\mathbb{P}\left(\bigcap_{i=1}^n [X_i = x_i]\right)$ pour tout $(x_1, \dots, x_n) \in (X_1, \dots, X_n)(\Omega)$.
- Là encore on peut faire comme si $(X_1, \dots, X_n)(\Omega) = X_1(\Omega) \times \dots \times X_n(\Omega)$ quitte à attribuer une probabilité nulle à certains éléments.
- On peut retrouver la loi des marginales lorsque l'on connaît la loi du couple à l'aide de la formule des probabilités totales appliquée au système complet d'événements associé aux $(n-1)$ -uplets formé des « autres » marginales. Par exemple, si $n = 3$, pour tout $y \in X_2(\Omega)$,

$$\mathbb{P}(X_2 = y) = \sum_{(x_1, x_3) \in (X_1, X_3)(\Omega)} \mathbb{P}(X_1 = x_1, X_2 = y, X_3 = x_3).$$



Là encore la réciproque est fautive : en général on ne peut pas obtenir la loi conjointe à partir des lois marginales connues séparément.

III Variables aléatoires indépendantes

1) Définition et caractérisations

a) Indépendance de deux variables aléatoires

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



En d'autres termes, deux variables aléatoires sont indépendantes lorsque tout événement relatif à X est indépendant de tout événement relatif à Y : tout ce qui se passe avec X seulement est indépendant de tout ce qui se passe avec Y seulement.

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

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \quad \mathbb{P}([X = x] \cap [Y = y]) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y).$$

⚠ Cas particulier important : Lorsque deux variables aléatoires X et Y sont indépendantes, on peut donc retrouver la loi conjointe du couple (X, Y) à l'aide des lois marginales.



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

DÉMONSTRATION.

□

Exemple : Soient $X \sim \mathcal{B}\left(\frac{1}{2}\right)$ et $Y = 1 - X$. Alors $\mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = \frac{1}{2}$ mais $\mathbb{P}([X = 1] \cap [Y = 1]) = 0 \neq \mathbb{P}(X = 1) \times \mathbb{P}(Y = 1)$: les deux variables aléatoires ne sont pas indépendantes, ce qui est complètement intuitif : si on connaît l'une, alors on connaît l'autre !

On a déjà vu que, pour toute distribution de probabilité indexée par $E \times F$, il existe un espace probabilisé et un couple (X, Y) sur cet espace dont la loi correspond à cette distribution. Mais à quelle condition peut-on construire un couple de variables aléatoires indépendantes ?

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 $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et des variables aléatoires X et Y indépendantes sur Ω à valeurs dans E et F respectivement tels que

$$\forall (e, f) \in E \times F, \quad \mathbb{P}(X = e) = p_e \quad \text{et} \quad \mathbb{P}(Y = f) = q_f.$$



Et donc

$$\mathbb{P}(X = e, Y = f) = p_e q_f.$$

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

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

La famille $(p_e q_f)_{e \in E, f \in F}$ est donc une distribution de probabilités sur l'ensemble $E \times F$ si bien qu'il existe donc un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et une variable aléatoire $(X, Y) : \Omega \rightarrow E \times F$ telle que, pour tout $(e, f) \in E \times F$, $\mathbb{P}([X = e] \cap [Y = f]) = p_e \times q_f$. En appliquant la formule des probabilités totales au système complet d'événements associé à Y , on obtient, pour tout $e \in E$,

$$\mathbb{P}(X = e) = \sum_{f \in F} \mathbb{P}(X = e, Y = f) = \sum_{f \in F} p_e q_f = p_e \times \sum_{f \in F} q_f = p_e \times 1 = p_e.$$

Et de même on obtient $\mathbb{P}(Y = f) = q_f$ pour tout $f \in F$. En particulier, pour tout $(e, f) \in E \times F$, $\mathbb{P}([X = e] \cap [Y = f]) = p_e \times q_f = \mathbb{P}(X = e) \mathbb{P}(Y = f)$ donc X et Y sont indépendantes. □

b) Indépendance mutuelle de n variables aléatoires

Définition. Soit $n \geq 2$. On dit n variables aléatoires X_1, \dots, X_n définies sur le même univers Ω sont (mutuellement) indépendantes si, pour tous $A_1 \in \mathcal{P}(X_1(\Omega)), \dots, A_n \in \mathcal{P}(X_n(\Omega))$, on a :

$$\mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n]) = \mathbb{P}(X_1 \in A_1) \times \dots \times \mathbb{P}(X_n \in A_n).$$


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



Quand on dit que des variables aléatoires sont indépendantes, il est toujours implicite qu'elles le sont mutuellement. C'est ça la bonne notion d'indépendance.

Cette définition semble plus simple que celle des événements indépendantes. En effet (il n'y a pas besoin d'introduire une sous partie J , etc.) mais c'est en fait la même chose :

Proposition. Soit $n \geq 2$. Soient X_1, \dots, X_n des variables aléatoires définies sur Ω . Elles sont mutuellement indépendantes si et seulement si, pour tous $A_1 \in \mathcal{P}(X_1(\Omega)), \dots, A_n \in \mathcal{P}(X_n(\Omega))$, les événements $[X_1 \in A_1], \dots, [X_n \in A_n]$ sont mutuellement indépendants.

 Là encore l'indépendance mutuelle entraîne l'indépendance deux à deux mais la réciproque est fautive.

DÉMONSTRATION. Soient $A_1 \in \mathcal{P}(X_1(\Omega)), \dots, A_n \in \mathcal{P}(X_n(\Omega))$. Supposons que les événements $[X_1 \in A_1], \dots, [X_n \in A_n]$ sont mutuellement indépendants. Alors, en particulier,


$$\mathbb{P}([X_1 \in A_1] \cap \dots \cap [X_n \in A_n]) = \mathbb{P}(X_1 \in A_1) \times \dots \times \mathbb{P}(X_n \in A_n).$$

donc $X_1 \perp \dots \perp X_n$. Réciproquement, supposons que $X_1 \perp \dots \perp X_n$. On se donne alors $A_1 \in \mathcal{P}(X_1(\Omega)), \dots, A_n \in \mathcal{P}(X_n(\Omega))$ puis J une partie finie de $\llbracket 1; n \rrbracket$. Pour tout $i \in J$, notons $B_i = A_i$ et, pour tout $i \notin J$, notons $B_i = X_i(\Omega)$. Alors

$$\mathbb{P}\left(\bigcap_{i \in J} [X_i \in A_i]\right) = \mathbb{P}\left(\bigcap_{i=1}^n [X_i \in B_i]\right) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i) = \prod_{i \in J} \mathbb{P}(X_i \in A_i).$$

Ainsi les événements $[X_1 \in A_1], \dots, [X_n \in A_n]$ sont mutuellement indépendants. \square

Proposition. Les variables aléatoires de toute sous famille de variables aléatoires mutuellement indépendantes sont mutuellement encore indépendantes.

 On utilise le fait que, pour tout $i \notin J$,
 $[X_i \in B_i] = \Omega$
 (donc l'ajouter ou l'enlever de l'union n'a aucun effet) et
 $\mathbb{P}(X_i \in B_i) = 1$
 (donc l'ajouter ou l'enlever du produit n'a aucun effet).

\rightsquigarrow DÉMONSTRATION LAISSÉE EN EXERCICE.


Là aussi, l'indépendance est entièrement caractérisée par ce qui se passe pour les événements élémentaires :

Proposition. Soit $n \geq 2$. Soient X_1, \dots, X_n des variables aléatoires définies sur Ω . Elles sont mutuellement indépendantes si et seulement si, pour tout (x_1, \dots, x_n) dans $X_1(\Omega) \times \dots \times X_n(\Omega)$,

$$\mathbb{P}([X_1 = x_1] \cap \dots \cap [X_n = x_n]) = \mathbb{P}(X_1 = x_1) \times \dots \times \mathbb{P}(X_n = x_n).$$

DÉMONSTRATION. Analogue au cas de deux variables aléatoires. \square


Remarque : En pratique, n variables aléatoires indépendantes servent à modéliser une succession de n expériences aléatoires indépendantes ou n répétitions indépendantes d'une même 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 :

 Par conséquent, comme dans le paragraphe II.2.d, on pourra toujours se donner des variables indépendantes suivant des lois quelconques sans avoir à justifier leur existence.

Théorème. Soient E_1, \dots, E_n des ensembles finis non vides. Pour tout $i \in \llbracket 1; n \rrbracket$, on se donne $(p_{i,e})_{e \in E_i}$ une distribution de probabilités de E_i . Il existe un espace probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et des variables aléatoires X_1, \dots, X_n sur Ω qui sont mutuellement indépendantes sur Ω , à valeurs dans E_1, \dots, E_n et telles que

$$\forall i \in \llbracket 1; n \rrbracket, \quad \forall e \in E_i, \quad \mathbb{P}(X_i = e) = p_{i,e}.$$

DÉMONSTRATION. Par récurrence en utilisant l'analogie pour deux variables indépendantes. \square

 Mais c'est assez technique à faire...

Exemples :

- Reprenons la preuve du fait que le nombre X de succès dans n répétitions indépendantes d'une même expérience de Bernoulli de probabilité de succès p suit une loi $\mathcal{B}(n, p)$. On avait admis l'existence d'une probabilité \mathbb{P} faisant que les S_i : « la $i^{\text{ième}}$ expérience est un succès », $1 \leq i \leq n$, des variables aléatoires mutuellement indépendantes. Il suffit en fait de remarquer que, pour tout $i \in \llbracket 1; n \rrbracket$, $S_i = [X_i = 1]$ avec X_i une variable aléatoire de Bernoulli qui vaut 1 lorsque la $i^{\text{ième}}$ expérience est un succès. Le théorème précédent assure l'existence de X_1, \dots, X_n mutuellement indépendantes et donc l'existence de S_1, \dots, S_n mutuellement indépendants.
- Soient $(p_1, \dots, p_n) \in]0; 1[^n$ et (X_1, \dots, X_n) des variables aléatoires indépendantes de loi respectives $\mathcal{B}(p_1), \dots, \mathcal{B}(p_n)$. Donnons la loi de $Y = X_1 \times \dots \times X_n$.

De telles variables aléatoires existent d'après le théorème précédent.

Définition. Lorsque des variables aléatoires sont (mutuellement) indépendantes et ont toutes la même loi, on dit qu'elles sont *i.i.d.* : indépendantes et identiquement distribuées.

2) Transfert de familles de variables aléatoires indépendantes

a) Transferts indépendants de deux variables aléatoires indépendantes

Proposition. Soient $X : \Omega \rightarrow E$ et $Y : \Omega \rightarrow F$ deux variables aléatoires. 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.

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 . Pour tout $\omega \in \Omega$, on a $X(\omega) \in C$ si et seulement si $f(X(\omega)) = a$ si bien que $[f(X) = a] = [X \in C]$. De même $[g(Y) = b] = [Y \in D]$. Dès lors :

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

car X et Y sont indépendantes. On en déduit que

$$\mathbb{P}([f(X) = a] \cap [g(Y) = b]) = \mathbb{P}(f(X) = a) \times \mathbb{P}(g(Y) = b).$$

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

b) Lemme des coalitions

Lemme (des coalitions). Soit $n \geq 2$. Soient X_1, \dots, X_n des variables aléatoires mutuellement indépendantes sur Ω . Soit $m \in \llbracket 1; n-1 \rrbracket$ et soient f et g définies respectivement sur $X_1(\Omega) \times \dots \times X_m(\Omega)$ et $X_{m+1}(\Omega) \times \dots \times X_n(\Omega)$. Alors $f(X_1, \dots, X_m)$ et $g(X_{m+1}, \dots, X_n)$ sont indépendantes.

DÉMONSTRATION. Notons Z la variable aléatoire (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.

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 $g(Y)$.

On généralise aisément à n variables aléatoires X_1, \dots, X_n . Si f_1, \dots, f_n sont des fonctions définies $X_1(\Omega), \dots, X_n(\Omega)$ respectivement alors $f_1(X_1), \dots, f_n(X_n)$ sont encore mutuellement indépendantes.

Soient $z = (z_1, \dots, z_m)$ dans $X_1(\Omega) \times \dots \times X_m(\Omega)$ et $t = (t_{m+1}, \dots, t_n)$ dans $X_{m+1}(\Omega) \times \dots \times X_n(\Omega)$. Par indépendance de X_1, \dots, X_n :

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

Ainsi Z et T sont indépendantes et la proposition du paragraphe précédent assure que $f(Z)$ et $g(T)$ sont indépendantes, ce que l'on voulait démontrer. \square

Exemples :

- Si X, Y, Z sont des variables aléatoires réelles indépendantes, alors $X^2 + Z^2$ et $\cos(Y)$ sont indépendantes.
- Soient X_1, \dots, X_{2n} des variables aléatoires indépendantes. Pour tout $k \in \llbracket 1; 2n \rrbracket$, notons $S_k = X_1 + \dots + X_k$. Alors

$$S_n = X_1 + \dots + X_n \quad \text{et} \quad S_{2n} - S_n = X_{n+1} + \dots + X_{2n}$$

sont indépendantes par lemme des coalitions. Allons même plus loin : si X_1, \dots, X_{2n} sont i.i.d. alors S_n et $S_{2n} - S_n$ ont la même loi. C'est tout fait à intuitif puisque $X_{n+1} \sim X_1, \dots, X_{2n} \sim X_n$ mais prouvons-le rigoureusement :

Remarque : Le lemme des coalitions s'étend (par récurrence) au cas de plus de deux coalitions.

Par exemple, si X_1, X_2, X_3, X_4, X_5 sont indépendantes, alors $X_1 + X_2, X_3 - X_5, 2X_4$ sont indépendantes.

c) Sommes de deux variables aléatoires indépendantes

Parmi les transferts de loi d'un couple, un cas de figure classique est celui de la somme. Détaillons un peu : on se donne un couple (X, Y) de variables aléatoires réelles (ou complexes) et on pose $Z = X + Y$.

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.

Puisque X et Z jouent le même rôle dans la somme, on peut bien sûr les échanger et sommer sur les $y \in Y(\Omega)$ les

$$\mathbb{P}(Y = y)\mathbb{P}(X = z - y).$$

Voici un cas particulier important :

Lemme (stabilité de la loi binomiale). Soient $p \in]0; 1[$, $n \in \mathbb{N}^*$ et $m \in \mathbb{N}^*$. Si $X \sim \mathcal{B}(n, p)$, $Y \sim \mathcal{B}(m, p)$ et $X \perp\!\!\!\perp Y$, alors $X + Y \sim \mathcal{B}(n + m, p)$.

DÉMONSTRATION.

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 !

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. Ce résultat est donc totalement 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 !

Théorème. Soient $p \in]0; 1[$ et $n \in \mathbb{N}^*$. Si X_1, \dots, X_n sont des variables aléatoires indépendantes de loi $\mathcal{B}(p)$, alors $X_1 + \dots + X_n \sim \mathcal{B}(n, p)$.

DÉMONSTRATION. Raisonnons par récurrence sur le nombre de variables aléatoires.

- Si X_1 et X_2 sont deux variables aléatoires indépendantes de loi $\mathcal{B}(p)$ (rappelons que $\mathcal{B}(p) \sim \mathcal{B}(1, p)$) alors le lemme précédent assure que $X_1 + X_2 \sim \mathcal{B}(2, p)$.
- Soit $n \in \mathbb{N}^*$. Supposons que la somme de n variables aléatoires indépendantes de loi $\mathcal{B}(p)$ suit une loi $\mathcal{B}(n, p)$. On se donne X_1, \dots, X_{n+1} variables indépendantes de loi $\mathcal{B}(p)$. Alors $X_1 + \dots + X_n$ et X_{n+1} sont indépendantes par lemme des coalitions. Comme $X_1 + \dots + X_n$ suit une loi $\mathcal{B}(n, p)$ (par hypothèse de récurrence) et que $X_{n+1} \sim \mathcal{B}(1, p)$, le lemme précédent assure que $X_1 + \dots + X_n + X_{n+1}$ suit une loi $\mathcal{B}(n + 1, p)$.

D'où le théorème par récurrence. □

d) Maximum et minimum de variables aléatoires indépendantes

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 $\mathbb{P}(M \leq n)$ (respectivement $\mathbb{P}(W > n)$). En effet

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

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

$$\mathbb{P}(M = n) = \mathbb{P}(M \leq n) - \mathbb{P}(M \leq n - 1).$$

et $\mathbb{P}(W = n) = \mathbb{P}(W > n - 1) - \mathbb{P}(W > n)$.

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 $\mathcal{U}(\llbracket 1; N \rrbracket)$. Déterminons les lois de $M = \max(X, Y)$ et $W = \min(X, Y)$.

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

On peut aussi calculer $\mathbb{P}(M < n)$ ou $\mathbb{P}(W \geq n)$, c'est analogue. L'idée est que le maximum (resp. minimum) de deux réels est majoré (resp. minoré) par n si et seulement si ces deux réels le sont : le plus grand élève d'une classe fait moins de deux mètres si et seulement si tous les élèves font moins de deux mètres, et le plus petit fait plus d'un mètre si et seulement si tous les élèves font plus d'un mètre.

On généralise sans mal au maximum/minimum d'un nombre quelconque de variables aléatoires mutuellement indépendantes.

IV Espérance et variance

1) Espérance d'une variable aléatoire réelle (ou complexe)

a) Définition et interprétation

Considérons un dé non équilibré et notons, pour tout $k \in \llbracket 1; 6 \rrbracket$, $p_k = \mathbb{P}(\{k\})$. Lançons un grand nombre N de fois ce dé. On s'attend à ce que la face k sorte un nombre de fois environ égal à Np_k pour tout $k \in \llbracket 1; 6 \rrbracket$. Ainsi on s'attend à ce que la face X obtenue prenne en moyenne la valeur

$$\frac{1}{N} (1 \times Np_1 + 2 \times Np_2 + 3 \times Np_3 + 4 \times Np_4 + 5 \times Np_5 + 6 \times Np_6) = \sum_{k=1}^6 kp_k.$$

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

Définition. Soit X une variable aléatoire complexe sur Ω . On appelle espérance de X la quantité

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

Remarques :

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

$$\mathbb{E}(X) = \boxed{}$$

En particulier, si $X(\Omega) = \llbracket 1; n \rrbracket$:

$$\mathbb{E}(X) = \boxed{}$$

- $\mathbb{E}(X)$ n'est rien d'autre qu'une moyenne pondérée : on somme chaque valeur possible pour X pondérée par la probabilité que X prenne cette valeur.
- Si deux variables aléatoires complexes X et Y ont la même loi, alors elles prennent les mêmes valeurs avec les mêmes probabilités donc ont la même espérance.

Exemples :

- Soit $c \in \mathbb{C}$. Si X est une variable aléatoire de loi certaine telle que $X(\Omega) = \{c\}$, alors

$$\mathbb{E}(X) = c \times \mathbb{P}(X = c) = c \times 1 = c.$$

Il n'est pas rare que l'on confonde c avec la variable aléatoire X et donc on pourra écrire $\mathbb{E}(c) = c$.

- Si X est la variable aléatoire donnant le résultat obtenu lors du lancer d'un dé équilibré, alors $X(\Omega) = \llbracket 1; 6 \rrbracket$, $\mathbb{P}(X = k) = \frac{1}{6}$ pour tout $k \in \llbracket 1; 6 \rrbracket$. Ainsi :

$$\mathbb{E}(X) = \sum_{k=1}^6 k \mathbb{P}(X = k) = \sum_{k=1}^6 \frac{k}{6} = \frac{1}{6} \times \frac{6 \times 7}{2} = \frac{7}{2}.$$

Si X désigne plutôt la variable aléatoire donnant le résultat obtenu lors du lancer du dé truqué de sorte que chaque face tombe avec une probabilité proportionnelle à sa valeur, alors $X(\Omega) = \llbracket 1; 6 \rrbracket$, $\mathbb{P}(X = k) = \frac{k}{21}$ et donc

$$\mathbb{E}(X) = \sum_{k=1}^6 k \mathbb{P}(X = k) = \sum_{k=1}^6 \frac{k^2}{21} = \frac{1}{21} \times \frac{6 \times 7 \times (2 \times 6 + 1)}{6} = \frac{13}{3}.$$

Si un événement arrive avec une probabilité p , on s'attend à ce qu'il arrive environ Np fois quand on réalise N fois l'expérience lors de laquelle on peut l'observer (par exemple, on s'attend à obtenir 100 fois la face 6 si on lance 600 fois un dé équilibré).

La somme est bien définie car Ω et donc $X(\Omega)$ sont finis. Ce ne sera plus forcément le cas l'an prochain avec un univers quelconque.

C'est même valable si $X(\Omega) = \llbracket 0; n \rrbracket$ puisque le terme $k \mathbb{P}(X = k)$ est alors nul si $k = 0$.

La réciproque est bien sûr fautive : par exemple, une variable aléatoire certaine égale à $1/2$ et variable aléatoire loi de Bernoulli de paramètre $1/2$ ont la même espérance (cf. paragraphe IV.1.c) mais n'ont pas la même loi !

$\mathbb{E}(X)$ n'est pas la valeur que X prend avec la plus grande probabilité. D'ailleurs $\mathbb{E}(X)$ peut même être une valeur que X ne prend jamais : par exemple l'espérance du résultat d'un dé équilibré est $3,5$.

- Si X donne la somme des chiffres des faces obtenues, alors

$$\mathbb{E}(X) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} + 7 \times \frac{6}{36} + 8 \times \frac{5}{36} + 9 \times \frac{4}{36} + 10 \times \frac{3}{36} + 11 \times \frac{2}{36} + 12 \times \frac{1}{36}.$$

On obtient $\mathbb{E}(X) = \frac{252}{36} = 7$. Par conséquent la somme moyenne des chiffres des faces obtenues en lançant deux dés est égale à 7.

- Reprenons l'exemple de l'urne contenant une boule numérotée 1, de 2 boules numérotées 2, etc. et n boules numérotées n (cf. fin du paragraphe 1.2). On a calculé que, si Y désigne la variable aléatoire égale au montant gagné à l'issue du jeu, alors $Y(\Omega) = \llbracket 0; n-1 \rrbracket$ et, pour tout $j \in \llbracket 0; n-1 \rrbracket$, $\mathbb{P}(Y = j) = \frac{2(n-j)}{n(n+1)}$. Ainsi

On retrouve le double du cas d'un seul dé. C'est attendu et cela sera confirmé par la propriété de linéarité de l'espérance.

Définition (variable centrée). Une variable aléatoire complexe X sur Ω est dite centrée si $\mathbb{E}(X) = 0$.

Exemple : Soit X une variable aléatoire uniforme sur $\{-1; 1\}$. On a alors $\mathbb{E}(X) = (-1) \times \mathbb{P}(X = -1) + 1 \times \mathbb{P}(X = 1) = -\frac{1}{2} + \frac{1}{2} = 0$. Ainsi X est centré.

b) Propriétés

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

DÉMONSTRATION.

A droite, on somme sur tous les éléments de $\omega \in \Omega$ et on multiplie leurs images par X par la probabilité de l'événement élémentaire associé. A gauche, on a juste regroupé les éléments ayant la même image en un seul terme x (l'image en question) multiplié par la somme de toutes les probabilités ($\mathbb{P}(X = x)$) et on somme sur toutes les images possibles. C'est d'ailleurs ce que l'on fait dans la démonstration ci-contre.

Dans le chapitre 7, nous n'avons vu que la formule de sommation par paquets pour deux paquets mais elle se généralise aisément par récurrence : si $(a_i)_{i \in I}$ est une famille de complexes indexée par I et que I est l'union disjointe de r parties J_1, \dots, J_p , alors

$$\sum_{i \in I} a_i = \sum_{k=1}^p \sum_{i \in J_k} a_i$$

on fait p paquets de termes (selon leurs indices), on somme chaque paquets puis on somme les sommes obtenues.

Proposition (linéarité de l'espérance). Soient X et Y deux variables aléatoires complexes sur Ω . Soit $(\lambda, \mu) \in \mathbb{C}^2$. Alors $\mathbb{E}(\lambda X + \mu Y) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y)$.

DÉMONSTRATION. D'après le lemme précédent :

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

Par conséquent :

$$\lambda \mathbb{E}(X) + \mu \mathbb{E}(Y) = \sum_{\omega \in \Omega} (\lambda X(\omega) + \mu Y(\omega)) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} (\lambda X + \mu Y)(\omega) \mathbb{P}(\{\omega\})$$

et donc $\lambda \mathbb{E}(X) + \mu \mathbb{E}(Y) = \mathbb{E}(\lambda X + \mu Y)$.

Un cas particulier important :

Corollaire. Soit X une variable aléatoire complexe sur Ω . Soient $(a, b) \in \mathbb{C}^2$. Alors $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

DÉMONSTRATION. Découle de la linéarité en prenant Y la variable aléatoire constante égale à 1 sur Ω et $(\lambda, \mu) = (a, b)$.

Corollaire (variable centrée associée à une variable aléatoire). La variable aléatoire $X - \mathbb{E}(X)$ est centrée et est appelée variable centrée associée à X .

DÉMONSTRATION. Notons $m = \mathbb{E}(X)$. On applique le résultat précédent avec $a = 1$ et $b = -m$: $\mathbb{E}(X - m) = \mathbb{E}(X) - m = 0$ donc $X - m$ est centrée.

Par récurrence, on obtient :

Corollaire (linéarité de l'espérance). Soient (X_1, \dots, X_n) des variables aléatoires complexes sur Ω . Soient $\lambda_1, \dots, \lambda_n$ des complexes. Alors :

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

Proposition (positivité de l'espérance). Soit X une variable aléatoire sur Ω et à valeurs positives. Alors $\mathbb{E}(X) \geq 0$ avec égalité si et seulement si $\mathbb{P}(X = 0) = 1$ (c'est-à-dire X est presque sûrement constant égale à 0).

DÉMONSTRATION.

⚠ Ce n'est pas vrai avec un produit : on n'a pas forcément $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$! Il faut une condition supplémentaire, cf. paragraphe IV.3.

📎 On a déjà dit qu'on peut assimiler une constante c à la variable aléatoire constante égale à c et on note sans problème $\mathbb{E}(c) = c$.

⚠ Même s'il y a du X dans $\mathbb{E}(X)$, il ne faut pas perdre de vue que $\mathbb{E}(X)$ 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.

⚠ Si X n'est pas positive, on n'a pas forcément $\mathbb{E}(X) \geq 0$, et le cas d'égalité n'est plus valable (par exemple, si $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$, alors $\mathbb{E}(X) = 0$).

□

Proposition (croissance de l'espérance). Soient X et Y deux variables aléatoires réelles sur Ω telles que $X \geq Y$. Alors $\mathbb{E}(X) \geq \mathbb{E}(Y)$ avec égalité si et seulement si $\mathbb{P}(X = Y) = 1$.

DÉMONSTRATION. La variable aléatoire $X - Y$ est alors positive donc $\mathbb{E}(X - Y) \geq 0$ avec égalité si et seulement si $\mathbb{P}(X - Y = 0) = 1$. On conclut avec la linéarité de l'espérance. □

Proposition (inégalité triangulaire). Soit X une variable aléatoire complexe sur Ω . Alors $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.

DÉMONSTRATION. D'après le lemme, par inégalité triangulaire (sur \mathbb{C}) et puisqu'une probabilité est positive :

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

□

Remarque : L'inégalité triangulaire est un cas particulier de ce qu'on appelle l'inégalité de Jensen (pour l'espérance) : si φ est convexe sur \mathbb{R} , alors $\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$ (nous le montrerons dans la feuille d'exercices associée).

Dire que $X \geq Y$ signifie bien sûr que $X(\omega) \geq Y(\omega)$ pour tout $\omega \in \Omega$.
⚠ Cela n'a aucun sens pour des variables complexes !

⚠ Il n'y a pas forcément égalité. Par exemple, si $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$, alors $\mathbb{E}(X) = 0$ et $\mathbb{E}(|X|) = 1$.

c) Espérance des variables aléatoires de lois usuelles

Proposition. Soit $E = \{x_1; \dots; x_n\}$ une partie non vide de \mathbb{C} . Si $X \sim \mathcal{U}(E)$, alors $\mathbb{E}(X) = \frac{1}{n} \sum_{i=1}^n x_i$ est la moyenne arithmétique des éléments de E .

DÉMONSTRATION.

□

Corollaire. Soient a et b des entiers tels que $a < b$. Si $X \sim \mathcal{U}(\llbracket a; b \rrbracket)$, alors $\mathbb{E}(X) = \frac{a+b}{2}$.

DÉMONSTRATION.

Ces deux résultats sur la loi uniforme ne sont pas officiellement au programme mais sont classiques. En particulier, si $X \sim \mathcal{U}(\llbracket 1; n \rrbracket)$ avec $n \in \mathbb{N}^*$, alors $\mathbb{E}(X) = \frac{n+1}{2}$.

C'est intuitif : puisque X prend toutes les valeurs avec la même probabilité, la valeur moyenne est « le milieu » de l'ensemble.

Proposition. Soit $p \in]0; 1[$. Si $X \sim \mathcal{B}(p)$, alors $\mathbb{E}(X) = p$. □

DÉMONSTRATION. On a $\mathbb{E}(X) = 0 \times \mathbb{P}(X = 0) + 1 \times \mathbb{P}(X = 1) = p$. □

Corollaire. Si $A \in \mathcal{P}(\Omega)$, alors $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

DÉMONSTRATION. Soit $A \in \mathcal{P}(\Omega)$. La variable aléatoire $\mathbb{1}_A$ ne prend que deux valeurs 0 et 1 donc suit une loi de Bernoulli. Par ailleurs $\mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A)$ donc $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$. □


Cette formule permet d'exprimer une probabilité comme une espérance.

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

DÉMONSTRATION. Proposons deux démonstrations.

□

2) Théorème de transfert

 On a vu dans le paragraphe IV.1.c que, lorsque X est une variable aléatoire complexe sur Ω , $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$. Cela peut laisser croire que, pour une fonction $f : X(\Omega) \rightarrow \mathbb{C}$ quelconque, on a $\mathbb{E}(f(X)) = f(\mathbb{E}(X))$. C'est totalement faux en général !

Par exemple, si $X \sim \mathcal{B}\left(\frac{1}{2}\right)$, alors $X^2 = X$ donc

$$\mathbb{E}(X^2) = \mathbb{E}(X) = \frac{1}{2} \neq \frac{1}{4} = (\mathbb{E}(X))^2.$$

Pour calculer l'espérance d'un transfert non affine de X , il y a deux options : ou bien on détermine la loi du transfert puis on calcule son espérance. Ou bien on utilise la formule suivante qui permet de calculer directement son espérance uniquement à partir de la loi de X :

Théorème (transfert). Soit X une variable aléatoire sur Ω . Soit $f : X(\Omega) \rightarrow \mathbb{C}$. Alors :

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

X peut être à valeurs dans un ensemble quelconque mais f est à valeurs complexes pour que $f(X)$ le soit.

DÉMONSTRATION.

Exemples :

- Commençons par quelques cas particuliers : soit X est une variable aléatoire sur Ω . Notons $X(\Omega) = \{x_1; \dots; x_n\}$. Alors

★ pour tout $k \in \mathbb{N}$, $\mathbb{E}(X^k) =$

★ $\mathbb{E}(|X|) =$.

★ lorsque $X(\Omega) \subset \mathbb{R}_+^*$, $\mathbb{E}\left(\frac{1}{\sqrt{X}}\right) =$.

- Soient $n \in \mathbb{N}^*$, $p \in [0; 1]$ et $t \in \mathbb{R}$. Calculons $\mathbb{E}(e^{tX})$ lorsque $X \sim \mathcal{B}(n, p)$.

- Soit $n \in \mathbb{N}^*$. Considérons X_n une variable aléatoire de loi uniforme sur $\llbracket 1; n \rrbracket$. Soit f une fonction continue par morceaux sur $]0; 1[$. Calculons $\lim_{n \rightarrow +\infty} \mathbb{E} \left(f \left(\frac{X_n}{n} \right) \right)$.

Cette formule limite permet de calculer des valeurs approchées d'intégrales en utilisant des simulations informatiques de réalisations de variables aléatoires (et la loi des grandes nombres, cf. programme de deuxième année et petit aperçu en fin de chapitre).

Remarque : Notons que la formule de transfert s'applique à n'importe quelle variable aléatoire sur Ω et donc, en particulier, elle s'applique à un couple : si (X, Y) est un couple de variable aléatoires sur Ω et si $f : (X, Y)(\Omega) \rightarrow \mathbb{R}$, alors

$$\mathbb{E}(f(X, Y)) = \sum_{(x, y) \in (X, Y)(\Omega)} f(x, y) \mathbb{P}([X = x] \cap [Y = y]).$$

Cette somme peut être indexée par $X(\Omega) \times Y(\Omega)$ quitte à ce que (X, Y) prenne certaines valeurs avec probabilité nulle (on l'a déjà mentionné plusieurs fois).

3) L'indépendance multiplie les espérances


Poursuivons avec une conséquence spectaculaire de la formule de transfert :

Théorème (l'indépendance multiplie les espérances). Soient X et Y deux variables aléatoires complexes **indépendantes**. Alors


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

DÉMONSTRATION.

□

 Ce résultat est faux en général si les variables ne sont pas indépendantes !

Par exemple, si $X \sim \mathcal{B} \left(\frac{1}{2} \right)$ et $Y = X$, alors $XY = X^2 = X$ (rappelons qu'une loi de Bernoulli ne prend que les valeurs 0 et 1) donc $\mathbb{E}(XY) = \mathbb{E}(X) = \frac{1}{2}$ alors que $\mathbb{E}(X)\mathbb{E}(Y) = \frac{1}{4}$.

 La réciproque de ce théorème est fautive.


Par exemple, considérons $Z \sim \mathcal{U}(\{-1; 0; 1\})$ et posons $X = Z^3$ et $Y = Z^2$. On a $X = Z^3 = Z$ et $XY = Z^5 = Z$ donc $\mathbb{E}(X) = 0$ et $\mathbb{E}(XY) = 0$. Par conséquent $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Pourtant elles ne sont pas indépendantes puisque :

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

De façon analogue (laissée en exercice), on a :

Théorème (l'indépendance multiplie les espérances). Soient X_1, \dots, X_n des variables aléatoires complexes **indépendantes**. Alors

$$\mathbb{E} \left(\prod_{i=1}^n X_i \right) = \prod_{i=1}^n \mathbb{E}(X_i)$$

 Il faut impérativement citer l'indépendance. Par contre le fait que

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

découle de la linéarité de l'espérance et est vrai « tout le temps », y compris si les variables ne sont pas indépendantes (il ne faut donc surtout pas parler d'indépendance).

4) Variance d'une variable aléatoire réelle

Conformément au programme, bien que nous ayons défini l'espérance de variables aléatoires à valeurs complexes, nous ne nous intéressons qu'aux cas des variables réelles pour la variance.

a) Définition et interprétation

L'espérance d'une variable aléatoire réelle X est un indicateur de position. Elle ne permet pas de mesurer la proximité des valeurs prises par X par rapport à l'espérance, la valeur moyenne prise par X . Toute quantité qui mesure la proximité à l'espérance est appelée indicateur de dispersion.

- On pourrait dans un premier temps penser à $\mathbb{E}(X - \mathbb{E}(X))$ mais on a vu que cette quantité est toujours nulle. C'est normal : les valeurs négatives et positives se compensent.
- L'indicateur le plus naturel serait $\mathbb{E}(|X - \mathbb{E}(X)|)$ car il mesure les écarts en valeur absolue donc les termes négatifs ne compensent plus les termes positifs. En notant $m = \mathbb{E}(X)$, le théorème de transfert assure que

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

Le problème est que cette quantité est difficile à calculer dans la pratique.

Nous allons plutôt considérer l'écart quadratique à la moyenne :

Définition. Soit X une variable aléatoire réelle sur Ω . On appelle variance de X la quantité $\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$.

Remarque :

- Notons $X(\Omega) = \{x_1; \dots; x_n\}$ et $m = \mathbb{E}(X)$. On a alors

$$\mathbb{V}(X) = \boxed{\phantom{\sum_{i=1}^n (x_i - m)^2 \mathbb{P}(X = x_i)}}$$

- Puisque $(X - \mathbb{E}(X))^2$ est une variable aléatoire positive, la propriété de positivité de l'espérance assure que $\mathbb{V}(X) \geq 0$.
- Comme la variance est une espérance, deux variables aléatoires de même loi ont la même variance.

Exemples :



Il y a un risque de confusion avec le non moins spectaculaire résultat suivant :

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

concernant la variance et que l'on verra dans le prochain paragraphe.

- Si X est la variable aléatoire donnant le résultat obtenu lors du lancer d'un dé équilibré, alors $X(\Omega) = \llbracket 1; 6 \rrbracket$ et $\mathbb{P}(X = k) = \frac{1}{6}$ pour tout $k \in \llbracket 1; 6 \rrbracket$. On a calculé que $\mathbb{E}(X) = \frac{7}{2}$ donc

$$\mathbb{V}(X) = \sum_{k=1}^6 \left(k - \frac{7}{2}\right)^2 \mathbb{P}(X = k) = \frac{1}{24} \sum_{k=1}^6 (2k - 7)^2 = \frac{1}{24} \sum_{k=1}^6 (4k^2 - 28k + 49).$$

Dans la pratique, on n'utilise presque jamais cette formule mais plutôt la formule de Koenig-Huygens (cf. paragraphe suivant).

Après calcul, on trouve $\mathbb{V}(X) = \frac{35}{12}$.

- Soit $c \in \mathbb{R}$. Si X est une variable aléatoire de loi certaine telle que $X(\Omega) = \{c\}$, alors on a vu que $\mathbb{E}(c) = c$. Ainsi $X - c$ est une variable constante égale à 0 donc $(X - c)^2$ aussi et donc $\mathbb{V}(X) = \mathbb{E}((X - c)^2) = 0$. Réciproquement, si X est une variable aléatoire réelle de variance nulle, alors $\mathbb{P}((X - \mathbb{E}(X))^2 = 0) = 1$ (c'est le cas d'égalité de la propriété de positivité de l'espérance, $(X - \mathbb{E}(X))^2$ étant bien sûr une variable aléatoire positive) et donc $\mathbb{P}(X = \mathbb{E}(X)) = 1$: X est presque sûrement constante.

Remarque : Un physicien dirait que la variance et l'espérance ne sont pas homogènes, puisque la variance mesure le carré des écarts. Par exemple, si X représente une longueur, alors $\mathbb{E}(X)$ est une longueur et $\mathbb{V}(X)$ une longueur au carré. Pour pallier ce problème, on introduit la notion d'écart type :

Définition (écart type). On appelle écart-type d'une variable aléatoire réelle X la quantité $\sigma(X) = \sqrt{\mathbb{V}(X)}$.

Définition (variable réduite). Une variable aléatoire réelle X sur Ω est dite réduite si $\mathbb{V}(X) = 1$.

C'est encore équivalent à dire que $\sigma(X) = 1$.

b) Propriétés de la variance

Dans ce paragraphe, X est une variable aléatoire réelle sur Ω .

Proposition (formule de Koenig-Huygens). $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

DÉMONSTRATION.

En pratique, on utilise toujours la formule de Koenig-Huygens pour calculer une variance.

□

Exemple : Reprenons encore une fois l'exemple de l'urne contenant une boule numérotée 1, de 2 boules numérotées 2, etc. et n boules numérotées n (cf. paragraphe II.5.b) où, pour tout $j \in \llbracket 0; n - 1 \rrbracket$, $\mathbb{P}(Y = j) = \frac{2(n - j)}{n(n + 1)}$. On a calculé dans le paragraphe IV.1.a que $\mathbb{E}(Y) = \frac{n - 1}{3}$.

Proposition. Pour tous réels a et b , on a $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.

DÉMONSTRATION.

□

Corollaire. Pour tous réels a et b , on a $\sigma(aX + b) = |a|\sigma(X)$.

Proposition/Définition. Soit X une variable aléatoire réelle sur Ω telle que $\sigma(X) > 0$. Alors $X^* = \frac{X - \mathbb{E}(X)}{\sigma(X)}$ est une variable aléatoire centrée et réduite, appelée variable aléatoire centrée réduite associée à X .

DÉMONSTRATION. Notons $m = \mathbb{E}(X)$ et $\sigma = \sigma(X)$. Alors $\mathbb{E}\left(\frac{X - m}{\sigma}\right) = \frac{\mathbb{E}(X) - m}{\sigma} = \frac{m - m}{\sigma} = 0$ et $\mathbb{V}\left(\frac{X - m}{\sigma}\right) = \frac{1}{\sigma^2}\mathbb{V}(X - m) = \frac{\mathbb{V}(X)}{\sigma^2} = 1$. □

c) Variance des variables aléatoires de lois usuelles

Proposition. Soit $p \in]0; 1[$. Si $X \sim \mathcal{B}(p)$, alors $\mathbb{V}(X) = p(1 - p)$.

DÉMONSTRATION. On a vu que $\mathbb{E}(X) = p$. Or $X^2 = X$ puisque X ne prend que 0 et 1 pur valeurs. Ainsi $\mathbb{E}(X^2) = \mathbb{E}(X) = p$. La formule de Koenig-Huygens assure enfin que $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = p - p^2 = p(1 - p)$. □

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

DÉMONSTRATION.



Où est donc passé le b ? La variance mesure la moyenne des écarts à la moyenne. Si on décale toutes les valeurs de b , les écarts restent les mêmes donc leur moyenne aussi. En revanche, quand on multiplie par a , on « dilate » les écarts d'un facteur a^2 pour la variance et donc de $|a|$ pour l'écart-type.



Il n'est pas rare qu'un énoncé demande une majoration de la variance d'une variable aléatoire de loi $\mathcal{B}(p)$ qui soit indépendante de p , à savoir $p(1 - p)$. Le premier réflexe est de la majorer par 1 mais, on peut faire en fait bien mieux :

$$p(1 - p) \leq \frac{1}{4}.$$

En effet

$$\begin{aligned} \frac{1}{4} - p(1 - p) &= \frac{1 - 4p + 4p^2}{4} \\ &= \frac{(1 - 2p)^2}{4} \geq 0. \end{aligned}$$

Il y égalité si et seulement si $p = \frac{1}{2}$.

□

d) Covariance

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 et on note $\text{Cov}(X, Y)$ la quantité

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

Remarques :

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

DÉMONSTRATION. On développe

$$(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) = XY - \mathbb{E}(X)Y - \mathbb{E}(Y)X + \mathbb{E}(X)\mathbb{E}(Y).$$

Or $\mathbb{E}(\mathbb{E}(X)Y) = \mathbb{E}(X)\mathbb{E}(Y)$ car $\mathbb{E}(X)$ est un réel et car l'espérance est linéaire. De même $\mathbb{E}(\mathbb{E}(Y)X) = \mathbb{E}(Y)\mathbb{E}(X)$ et $\mathbb{E}(\mathbb{E}(X)\mathbb{E}(Y)) = \mathbb{E}(X)\mathbb{E}(Y)$. Dès lors, par linéarité de l'espérance

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)$$

et donc $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. □

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. Si X et Y sont indépendantes, alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (l'indépendance multiplie les espérances) et donc $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$. □


Proposition.

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

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

 C'est analogue à la formule de Koenig-Huygens.

 On a vu dans le paragraphe IV.3 que la réciproque est fautive : on peut avoir $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (c'est-à-dire $\text{Cov}(X, Y) = 0$) sans que X et Y ne soient indépendantes.

 C'est une formule d'identité remarquable pour la variance.

DÉMONSTRATION. Démontrons tout de suite le cas général. La formule de développement d'un carré (cf. chapitre 7) assure que

$$\left(\sum_{i=1}^n X_i\right)^2 = \sum_{i=1}^n X_i^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j$$

donc, par linéarité de l'espérance,

$$\mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^2\right) = \sum_{i=1}^n \mathbb{E}(X_i^2) + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}(X_i X_j).$$

On a de même

$$\left(\mathbb{E}\left(\sum_{i=1}^n X_i\right)\right)^2 = \left(\sum_{i=1}^n \mathbb{E}(X_i)\right)^2 = \sum_{i=1}^n \mathbb{E}(X_i)^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}(X_i)\mathbb{E}(X_j). \quad \square$$

La formule de Koenig-Huygens entraîne alors que

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

On applique encore la formule de Koenig-Huygens à cette dernière ligne.

Corollaire (l'indépendance somme les variances).

- Si X et Y sont indépendantes, alors $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.
- Si X_1, \dots, X_n sont des variables aléatoires réelles sur Ω qui sont **deux à deux indépendantes**, alors

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

C'est faux en général si les variables ne sont pas indépendantes. Par exemple, si $X \sim \mathcal{B}\left(\frac{1}{2}\right)$ et $Y = 1 - X$, alors $X + Y = 1$ donc $\mathbb{V}(X + Y) = 0 \neq \frac{1}{2} = \mathbb{V}(X) + \mathbb{V}(Y)$.

Remarque : On peut démontrer autrement la formule de variance d'une variable aléatoire X de loi $\mathcal{B}(n, p)$ en utilisant le fait que $X \sim X_1 + \dots + X_n$ avec X_1, \dots, X_n i.i.d. de loi $\mathcal{B}(p)$. On a alors

$$\mathbb{V}(X) = \boxed{\phantom{\sum_{i=1}^n \mathbb{V}(X_i)}}$$

V Inégalités de concentration

1) Inégalité de Markov

Proposition (inégalité de Markov). Soit X une variable aléatoire **positive**. On a alors

$$\forall a > 0, \quad \mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

X est positive signifie que $X(\Omega) \subset \mathbb{R}_+$.

DÉMONSTRATION.

Preuve alternative qui utilise le fait que la probabilité d'un événement est l'espérance de l'indicatrice de cet événement :

$$\begin{aligned}
 a\mathbb{P}(X \geq a) &= a\mathbb{E}(\mathbb{1}_{[X \geq a]}) \\
 &= \mathbb{E}(a\mathbb{1}_{[X \geq a]}) \\
 &\leq \mathbb{E}(X\mathbb{1}_{[X \geq a]})
 \end{aligned}$$

car $a\mathbb{1}_{[X \geq a]} \leq X\mathbb{1}_{[X \geq a]}$ et par croissance de l'espérance. Mais on a aussi $X\mathbb{1}_{[X \geq a]} \leq X$ donc, encore par croissance de l'espérance,

$$a\mathbb{P}(X \geq a) \leq \mathbb{E}(X)$$

2) Inégalité de Bienaymé-Tchebychev

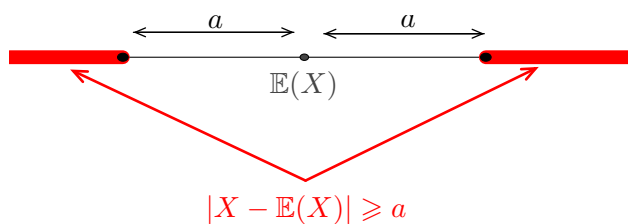
Proposition (inégalité Bienaymé-Tchebychev). Soit X une variable aléatoire réelle. On a alors

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X - \mathbb{E}(X)| \geq \varepsilon) \leq \frac{\mathbb{V}(X)}{\varepsilon^2}.$$

Cette fois l'hypothèse de positivité n'est plus requise.

DÉMONSTRATION.

Remarque : L'inégalité de Bienaymé-Tchebychev donne une majoration de la probabilité, 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 $\mathbb{E}(X)$, la « valeur moyenne ».



On pourrait se demander en quoi une inégalité serait-elle préférable à un calcul exact de la probabilité de gauche. Et bien tout simplement parce qu'en général, on ne sait pas calculer exactement de telles probabilités. Mais on sait très bien les majorer via des inégalités du type Bienaymé-Tchebychev, ce qui est déjà pas mal (comme on le verra dans le prochain paragraphe).

Puisque la variance est un indicateur de dispersion d'une variable aléatoire 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.

3) Exemples d'utilisation

On applique souvent l'inégalité de Bienaymé-Tchebychev à $\frac{S_n}{n}$ où $S_n = X_1 + \dots + X_n$ avec X_1, \dots, X_n de variables aléatoires réelles et i.i.d.

On peut encore majorer par $\frac{1}{4n\varepsilon^2}$ (cf. remarque dans la marge au paragraphe IV.4.c).


Exemples :

- On lance n fois un dé équilibré. On note X_n le nombre de 6 obtenus. Cherchons n de sorte que

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

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

- Un éleveur possède 100 vaches qui se répartissent au hasard entre deux étables contenant chacune k places avec $k \in \llbracket 50 ; 100 \rrbracket$. Déterminons une valeur de k permettant à chaque vache de trouver une place avec une probabilité supérieure ou égale à 95%.

 L'inégalité de Bienaymé-Tchebychev est facile à montrer et facile à utiliser mais elle est en fait assez grossière au sens où elle donne en général une majoration bien trop grande (et peu optimale).

Dans les deux exemples précédents, les probabilités peuvent s'obtenir de manière approchée (mais avec précisions satisfaisante) à l'aide d'un théorème hors-programme mais très important appelé le Théorème Central Limite. On trouve que :

- ★ 54 est la plus petite valeur de n telle que la fréquence d'apparition du 6 soit $1/6$ plus ou moins 1% arrive avec probabilité au moins égale à 95%.
- ★ Il faut au minimum 60 places dans chaque étape pour que la probabilité que chaque vache trouve une place soit supérieure ou égale à 95%.

Une branche entière des probabilités et des statistiques s'intéresse à l'obtention d'inégalités de ce type plus optimales selon telle ou telle situation. On les appelle les inégalités de concentration. Une première piste est d'adapter la preuve de l'inégalité de Bienaymé-Tchebychev. Si on jette un oeil à sa démonstration, on constate que l'on passe au carré avant d'appliquer l'inégalité de Markov. On peut généraliser ce procédé : si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction croissante et à valeurs strictement positives alors, pour tout $a \in \mathbb{R}$, $[X \geq a] \subset [f(X) \geq f(a)]$ et, comme $f(a) > 0$, l'inégalité de Markov entraîne que

$$\mathbb{P}(X \geq a) \leq \mathbb{P}(f(X) \geq f(a)) \leq \frac{\mathbb{E}(f(X))}{f(a)}$$

On peut par exemple considérer $t > 0$ et $f : x \mapsto e^{tx}$. On obtient alors, pour tout $a \in \mathbb{R}$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}} = e^{-ta + \Lambda_X(t)},$$

où $\Lambda_X(t) = \ln(\mathbb{E}(e^{tX}))$. Sous réserve d'existence, on peut considérer la valeur de t qui rend optimale cette inégalité et on obtient l'inégalité de Chernoff :

$$\mathbb{P}(X \geq a) \leq e^{-I(a)} \quad \text{avec} \quad I(a) = \sup_{t \in \mathbb{R}} (ta - \Lambda_X(t)).$$

En considérant des valeurs de t négatives, on obtient la même inégalité pour $\mathbb{P}(X \leq -a)$ et donc $\mathbb{P}(|X| \geq a) \leq 2e^{-I(a)}$. On peut montrer (c'est accessible par le calcul mais c'est

Ce qui est bien moins que 27778.
Ce qui est bien moins que 73 places par étable !

La fonction G_X est appelée (transformée de) log-Laplace de X . La fonction I est appelée transformée de Fenchel-Legendre de la fonction Λ_X .

assez technique) que, si $X \sim \mathcal{B}(n, p)$, alors

$$\forall \varepsilon > 0, \quad \mathbb{P} \left(\left| \frac{X}{n} - p \right| \geq \varepsilon \right) \leq 2e^{-2\varepsilon^2 n}.$$

Exemples :

- Si on applique cette dernière inégalité avec $p = \frac{1}{6}$ et $\varepsilon = \frac{1}{100}$, on obtient

$$\mathbb{P} \left(\left| \frac{X}{n} - \frac{1}{6} \right| \geq \frac{1}{100} \right) \leq 2e^{-n/5000}.$$

C'est précisément la situation de l'exemple avec les dés ci-dessus. Mais cette fois, on a $2e^{-n/5000} \leq 0.05$ si et seulement si $n \geq \left\lfloor \sqrt{5000 \ln(40)} \right\rfloor + 1 = 136$. On gagne en précision par rapport à l'inégalité de Bienaymé-Tchebychev.

- Si on applique cette dernière inégalité avec $n = 100$, $p = \frac{1}{2}$ et $\varepsilon = \frac{k-50}{100}$, on obtient

$$\mathbb{P} (|X - 50| \geq k) \leq 2e^{-(k-50)^2/50}.$$

C'est précisément la situation de l'exemple avec les vaches ci-dessus. Mais cette fois, on a $2e^{-(k-50)^2/50} \leq 0.05$ si et seulement si $k \geq \left\lfloor \sqrt{50 \ln(40)} \right\rfloor + 51 = 63$. On gagne en précision par rapport à l'inégalité de Bienaymé-Tchebychev.

4) Interprétation fréquentiste

Théorème (loi faible des grands nombres – HP). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles i.i.d. Notons m l'espérance de X_1 et σ^2 sa variance. Pour tout $n \in \mathbb{N}^*$, posons $S_n = X_1 + \dots + X_n$. Alors :

$$\forall \varepsilon > 0, \quad \mathbb{P} \left(\left| \frac{S_n}{n} - m \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow +\infty} 0.$$

DÉMONSTRATION. Soit $\varepsilon > 0$. On a déjà montré au début du paragraphe précédent que, dans cette configuration

$$\forall n \in \mathbb{N}^*, \quad \mathbb{P} \left(\left| \frac{S_n}{n} - m \right| \geq \varepsilon \right) \leq \frac{\mathbb{V}(S_n/n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Il suffit de faire tendre n vers $+\infty$ et de conclure avec le théorème d'encadrement. \square

Théorème (loi faible des grands nombres pour une loi binomiale – HP). Soit $p \in]0; 1[$. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes de loi $\mathcal{B}(p)$.

$$\forall \varepsilon > 0, \quad \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - p \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow +\infty} 0.$$

Cohérence du modèle probabiliste. On peut voir $S_n = X_1 + \dots + X_n$ comme le nombre de succès lors de n répétitions indépendantes d'une expérience aléatoire où la probabilité de succès est p . La quantité $\frac{S_n}{n}$ est alors la fréquence d'apparition du succès lors des n premiers lancers, et on vient de montrer que, en un certain sens,

$$\frac{S_n}{n} \xrightarrow{n \rightarrow +\infty} p.$$

C'est la définition intuitive d'une probabilité (cf. chapitre 34) : si A est un événement, on veut poser $\mathbb{P}(A)$ 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 !

L'existence d'une telle suite est un problème mathématique en soi. Il sera abordé l'an prochain.

On dit que la suite $(\frac{S_n}{n})_{n \in \mathbb{N}^*}$ converge en probabilité vers m .

Interprétation : si on joue à un jeu un très grand nombre de fois et de façon indépendante, alors le gain moyen (et non pas le gain total) obtenu sera alors une bonne approximation de l'espérance de gain.