Chapitre 26 - Probabilités finis II : Lois d'une variable aléatoire et indépendance

1 Loi d'une variable aléatoire

1.1 Loi d'une variable aléatoire

On rappelle qu'une variable aléatoire sur un univers fini Ω n'est rien d'autre qu'une fonction $X : \Omega \to E$, où E est un ensemble. Nous allons voir comment la donnée d'une probabilité P sur Ω se transporte sur l'ensemble (fini) des valeurs $X(\Omega)$ que peut renvoyer la variable aléatoire X. Dans la pratique, ce sont les valeurs mesurées par la variable aléatoire X qui ont de l'importance, et dont on cherche à mesurer la probabilité d'occurence. Cependant, l'espace probabilisé (Ω, P), modélisant l'expérience aléatoire, doit être étudié avec soin car c'est lui qui dicte les probabilités des valeurs renvoyées par X.

Dans ce qui suit, (Ω, P) est un espace probabilisé fini.

Définition-théorème 1 - Loi de probabilité d'une variable aléatoire. Soit X une variable aléatoire sur Ω. On appelle *loi (de probabilité) de* X l'application $\mathrm{P}_X : X(\Omega) \longrightarrow [0\,,1]$, définie par

$$P_X : x \longmapsto \mathrm{P}(X = x)\,.$$

Alors P_X est une probabilité sur $X(\Omega)$, en particulier $\displaystyle\sum_{x \in X(\Omega)} \mathrm{P}(X = x) = 1$.

On définit ainsi une probabilité P_X directement sur l'ensemble $X(\Omega)$ – souvent plus simple et plus intéressant que Ω – naturellement issue de la probabilité P déjà connue sur Ω.

La loi de X mesure la vraisemblance des valeurs de X. Les valeurs possibles de X, tout en étant possibles, n'ont en effet pas forcément les mêmes chances de réalisation et c'est précisément cette information que P_X mesure.

✎ **En pratique** ✎ Expliciter la loi P_X d'une variable aléatoire finie revient à déterminer :

(i) son support $X(\Omega)$;

(ii) la valeur de $\mathrm{P}(X = x)$, pour tout $x \in X(\Omega)$.

Si $X(\Omega)$ est de cardinal raisonnable, on peut présenter cela à l'aide d'un tableau.

Exemple 2 Reprenons le lancer de deux dés successifs, et les variables aléatoires X_1 : « valeur du premier dé » et S : « somme des deux dés ». Les lois de probabilité des variables aléatoires X_1 et S sont résumées par les tableaux ci-dessous.

i	1	2	3	4	5	6
$\mathrm{P}(X_1 = i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

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

Le théorème suivant indique qu'il est possible de définir la loi d'une variable aléatoire *ex nihilo*, c'est-à-dire sans devoir décrire complètement l'expérience aléatoire sous-jacente. Ainsi, on peut escamoter l'espace probabilisé $(\Omega, \mathscr{P}(\Omega), \mathrm{P})$ au profit de la seule loi de la variable aléatoire. Ce résultat justifie qu'on s'intéresse à certaines variable aléatoirequi reviennent naturellement dans des expériences aléatoires a priori différentes.

Théorème 3 - Générer la probabilité d'une variable aléatoire réelle finie. Soit I un sous-ensemble fini de $\mathbb{R}$. Une famille de réels $(p_i)_{i\in I}$ définit la loi de probabilité d'une variable aléatoire réelle finie X de support I si et seulement si

$$\forall i \in I, \quad p_i \geqslant 0 \qquad \text{et} \qquad \sum_{i\in I} p_i = 1.$$

Démonstration. ... ■

La définition suivante est essentielle, elle est la clef pour comprendre pourquoi il est important d'étudier la loi d'une variable aléatoire pour certain cas particuliers :

Définition 4 - Deux variables aléatoires pour une seule loi. Soient $X : \Omega \to E$ et $Y : \Omega' \to E$ deux variables aléatoires. On note $X \sim Y$ la relation $P_X = P_Y$.

Avez-vous vu la subtilité ? Les expériences Ω et Ω' peuvent être différentes, et pourtant on peut arriver à comparer les lois des variables aléatoires X et Y.

1.2 Fonction d'une variable aléatoire finie

Une variable aléatoire réelle étant notamment une fonction à valeurs réelles, on peut opérer sur l'ensemble des variables aléatoires réelles. Ainsi, si X et Y sont des variables aléatoires réelles, il en va alors de même de $X+Y$, XY, λX, avec $\lambda \in \mathbb{R}$, etc.Nous allons également considérer des variables aléatoires à valeurs vectorielles, que nous pourrons toujours ajouter.

Définition-théorème 5 - Image d'une variable aléatoire par une fonction.
Soit $X : \Omega \to E$ une variable aléatoire sur Ω, F un ensemble, et $f : X(\Omega) \longrightarrow F$ une fonction. La fonction $f \circ X : \Omega \longrightarrow F$ est aussi une variable aléatoire, notée simplement $f(X)$.
La loi $\mathrm{P}_{f(X)}$ de $f(X)$ est entièrement déterminée par f et la loi de X. Précisément, pour tout $y \in f(X)(\Omega)$,

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

En particulier, si Y est une autre variable aléatoire, alors on a

$$X \sim Y \implies f(X) \sim f(Y).$$

Démonstration. Pour tout $y \in f(X)(\Omega) = f\left(X\left(\Omega\right)\right)$, il suffit de remarquer que $(f(X) = y) = \bigsqcup_{\substack{x\in X(\Omega)\\ y=f(x)}} (X = x)$. ■

✎ **En pratique** ✎ **Détermination de la loi de $Y = f(X)$**

(i) On commence par déterminer le support de Y, qui s'obtient comme l'image direct du support $X(\Omega)$ de X par la fonction f : $Y(\Omega) = f(X(\Omega))$.

(ii) Pour chaque valeur $y \in Y(\Omega)$, on détermine $\mathrm{P}(Y = y)$, en décomposant l'événement $(Y = y)$ à l'aide du système complet d'événements $((X = x))_{x\in X(\Omega)}$ associé à X. En pratique, cela revient à déterminer les antécédents de y dans $X(\Omega)$ par la fonction f.

Exemple 6 On choisit un entier X au hasard entre 1 et $2n$ et on pose $Y = \left\lfloor \frac{X}{2} \right\rfloor$. Quelle est la loi de Y ?

En effet, On choisit pour univers Ω l'ensemble $[\![1\,,2n]\!]$, que l'on munit de la probabilité uniforme, notée P. Le support de X est Ω et, pour tout $i \in [\![1\,,2n]\!]$, puisque $(X = i) = \{i\}$, $\mathrm{P}(X = i) = \dfrac{1}{2n}$.

Quant à Y, son support est l'ensemble $[\![0\,,n]\!]$ et, pour tout $j \in [\![0\,,n]\!]$, $\mathrm{P}(Y = j) = \displaystyle\sum_{1\leqslant i\leqslant 2n,\left\lfloor \frac{i}{2}\right\rfloor = j} \mathrm{P}(X = i)$.

Ainsi, $\mathrm{P}(Y=0)=\mathrm{P}(X=1)=\dfrac{1}{2n}$, $\mathrm{P}(Y=n)=\mathrm{P}(X=2n)=\dfrac{1}{2n}$ et, pour tout $j\in[\![1\,,n-1]\!]$,

$$\mathrm{P}(Y=j)=\mathrm{P}(X=2j)+\mathrm{P}(X=2j+1)=\frac{1}{n}.$$

1.3 Loi conditionnelle

Définition 7 - Loi conditionnelle d'une variable aléatoire.
Soit X une variable aléatoire finie sur Ω et $A\in\mathscr{P}(\Omega)$ tel que $\mathrm{P}(A)\neq 0$.
On appelle *loi conditionnelle de X sachant A* l'application $x\longmapsto \mathrm{P}_A(X=x)$ de $X(\Omega)$ dans $[0\,,1]$.

Exemple 8 Reprenons l'exemple du lancer d'un dé à 6 faces 2 fois. on peut choisir pour univers Ω l'ensemble $[\![1\,,6]\!]^2$ des 2-listes de $[\![1\,,6]\!]$ et pour probabilité sur Ω la probabilité uniforme. On avait noté X_1 (resp. X_2) la « valeur obtenue au premier (resp. deuxième) lancer » et $S=X_1+X_2$ « la somme des deux faces obtenues ». La loi conditionnelle de la variable aléatoire X_1 sachant $(S=4)$ est donnée par le tableau ci-contre.

k	1	2	3	4	5	6
$\mathrm{P}_{(S=4)}(X_1=k)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0	0

2 Lois usuelles finies

Dans l'ensemble de cette section, (Ω,P) désigne un espace probabilisé fini et X une variable aléatoire réelle finie sur cet espace.

2.1 Loi uniforme

Définition 9 - Loi uniforme. Soit E un ensemble fini non vide. Une variable aléatoire X suit la *loi uniforme sur A* lorsque

$$X(\Omega)=E \qquad \text{et} \qquad \forall x\in E,\quad \mathrm{P}(X=x)=\frac{1}{|E|}.$$

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

On définit ainsi correctement une loi de probabilité (cf. théorème 4), puisque

- pour tout $x\in E$, on a $\mathrm{P}(X=x)=\dfrac{1}{|E|}\geqslant 0$;
- $\displaystyle\sum_{x\in E}\mathrm{P}(X=x)=|E|\times\frac{1}{|E|}=1.$

La loi uniforme est la loi pour laquelle les événements $(X=x)$, avec $x\in E$, sont équiprobables, elle modélise ainsi les situations d'équiprobabilité.

Exemple 10

- La variable aléatoire égal au résultat obtenu suite au lancer d'un dé équilibré à 6 faces suit la loi uniforme sur $[\![1\,,6]\!]$.
- On dispose d'une urne contenant n boule numérotées de 1 à n et indiscernables. On tire une boule au hasard. La variable aléatoire donnant le numéro de la boule tirée suit la loi uniforme sur $[\![1\,,n]\!]$.

Exemple 11 On choisit un entier X au hasard entre 1 et $2n$. Quelle est la loi conditionnelle de X sachant $(X\leqslant n)$? Quelle est la loi de $(-1)^X$?

2.2 Loi de Bernoulli

La loi de Bernoulli[†] apparait dans toute expérience de type « succès/échec ». Précisément, on appelle *épreuve de Bernoulli de paramètre* $p \in [0\,,1]$ toute expérience aléatoire à deux issues : une nommée « succès », de probabilité p, et l'autre « échec », de probabilité $1-p$, avec $p \in [0\,,1]$.

Exemple 12 L'expérience aléatoire consistant à lancer une pièce équilibrée avec pour succès l'obtention de la face « pile » est une épreuve de Bernoulli de paramètre $p = 1/2$.

Définition 13 - Loi de Bernoulli. Soit $p \in [0\,,1]$. Une variable aléatoire X suit la *loi de Bernoulli de paramètre* p lorsque

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

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

On définit ainsi correctement une loi de probabilité, puisque $p \geqslant 0$, $1-p \geqslant 0$ et $p+(1-p)=1$ (cf. théorème 4).

Une variable aléatoire X suit la loi de Bernoulli de paramètre p lorsqu'elle est liée à une épreuve de Bernoulli de paramètre p en associant la valeur 1 au « succès » et la valeur 0 à l'« échec ».

Exemple 14 Lors du lancer d'un dé, Alice gagne 1 euro lorsque le résultat est supérieur à 5 et rien sinon. On note G son gain. Alors la variable aléatoire G suit une loi de Bernoulli de paramètre $\frac{1}{3}$.
En effet, $G(\Omega) = \{0,1\}$ et $\mathrm{P}(G=1) = \mathrm{P}(\{5,6\}) = \frac{2}{6} = \frac{1}{3}$. Ici le « succès » est l'événement « le résultat est supérieur à 5 ».

Exemple 15 Soit $p \in [0\,,1]$ et $k \in \mathbb{N}^*$. Si $X \sim \mathcal{B}(p)$, alors $X^k \sim \mathcal{B}(p)$.
En effet, $X^k(\Omega) = \{0,1\}$ et $\mathrm{P}\big(X^k = 1\big) = \mathrm{P}(X=1) = p$.

Remarque 16 La loi de Bernoulli $\mathcal{B}(\frac{1}{2})$ se confond avec la loi uniforme $\mathcal{U}(\llbracket 0\,,1 \rrbracket)$.

Exemple 17 - Exemple fondamental. Pour tout événement $A \in \mathscr{P}(\Omega)$, on note $\mathbb{1}_A$ la variable aléatoire définie par

$$\forall \omega \in \Omega, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

On a alors $\mathbb{1}_A \sim \mathcal{B}(\mathrm{P}(A))$.

2.3 Loi binomiale

Un *schéma binomiale de paramètres* $n \in \mathbb{N}^*$ *et* $p \in [0\,,1]$ est une expérience aléatoire consistant à répéter n fois de manière INDÉPENDANTE une même épreuve de Bernoulli de paramètre p. On observe ainsi une succession de succès et d'échecs mutuellement indépendants et on s'intéresse à la variable aléatoire X comptant le nombre de succès observés suite aux n répétitions. Quelle est sa loi ?

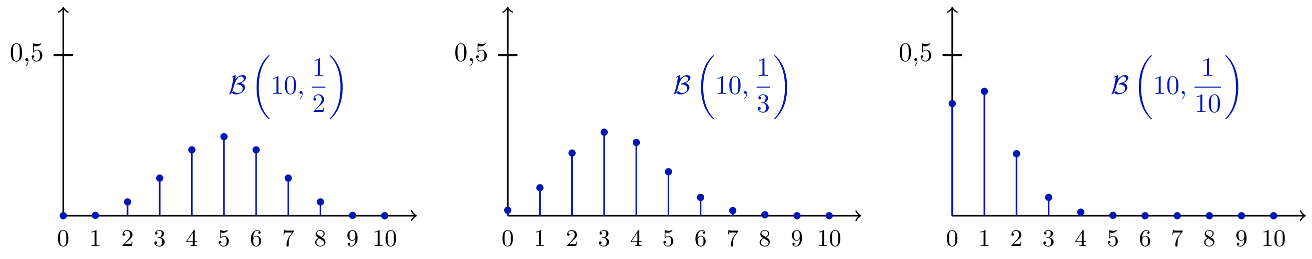
Son support est $X(\Omega) = \llbracket 0\,,n \rrbracket$ et, pour tout $k \in \llbracket 0\,,n \rrbracket$, l'événement $(X=k)$ est réalisé si et seulement si l'on a obtenu k succès et $n-k$ échecs. Or il y a $\binom{n}{k}$ manières de distribuer ces k succès parmi les n répétitions et chacune de ces possibilités a, par indépendance, la même probabilité $p^k(1-p)^{n-k}$ de se réaliser, ainsi $\mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$.

†. En l'honneur de Jacques Bernoulli (1654 à Bâle - 1705 à Bâle) mathématicien et physicien suisse ayant notamment contribué à l'étude des probabilités, des séries et de certaines équations différentielles.

Définition 18 - Loi binomiale ou loi des tirages avec remise. Soit $n \in \mathbb{N}^*$ et $p \in [0\,,1]$.
Une variable aléatoire X suit la *loi binomiale de paramètres n et p* lorsque

$$X(\Omega) = [\![0\,,n]\!] \qquad \text{et} \qquad \forall k \in [\![0\,,n]\!]\,, \quad \mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

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



On définit bien ainsi une loi de probabilité (cf. théorème 4), puisque $\mathrm{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \geqslant 0$, pour tout $k \in [\![0\,,n]\!]$, et, d'après la formule du binôme,

$$\sum_{k=0}^{n} \mathrm{P}(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (\cancel{p} + (1-\cancel{p}))^n = 1^n = 1.$$

Lorsque l'on répète n fois indépendamment une expérience aléatoire à deux issues de probabilité p pour l'issue favorable, le NOMBRE D'ISSUES FAVORABLES suit la loi binomiale $\mathcal{B}(n,p)$.

Remarque 19

- L'expression « loi des tirages AVEC remise » par laquelle on décrit parfois la loi binomiale trouve son explication dans l'exemple suivant : lorsque l'on tire AVEC remise – donc indépendamment – n boules dans une urne contenant des boules blanches en proportion p et des boules noires en proportion $1-p$, la variable aléatoire comptant le nombre de boules blanches tirées suit la loi binomiale $\mathcal{B}(n,p)$.
- Clairement, la loi binomiale $\mathcal{B}(1,p)$ n'est autre que la loi de Bernoulli $\mathcal{B}(p)$!

Exemple 20 On lance 5 fois un dé équilibré à 6 faces dont 2 blanches et 4 noires. Avec quelle probabilité obtient-on exactement 3 fois une face noire ?

Exemple 21 Pour toute variable aléatoire X suivant une loi binomiale de paramètre (n,p), la variable aléatoire $n-X$ suit une loi binomiale de paramètre $(n,1-p)$.

3 Espérance et variance d'une variable aléatoire réelle finie

3.1 Définition de l'espérance

Définition 22 - Espérance d'une variable aléatoire réelle. Soit X une variable aléatoire réelle finie sur Ω.

- On appelle *espérance de X* le réel, noté $\mathbf{E}(X)$, défini par $\mathbf{E}(X) = \displaystyle\sum_{x \in X(\Omega)} \mathrm{P}(X=x)\,x$.
- La variable aléatoire X est dite *centrée* lorsque $\mathbf{E}(X) = 0$.

L'espérance de X est une moyenne des valeurs de X, donc un *indicateur de position*. Précisément, chaque valeur x, x décrivant $X(\Omega)$, s'y trouve comptabilisé en proportion de sa probabilité d'occurrence $\mathrm{P}(X=x)$: plus $\mathrm{P}(X=x)$ est proche de 1, plus x a d'importance dans la somme.

Exemple 23 Reprenons l'exemple du lancer d'un dé à 6 faces 2 fois. on peut choisir pour univers Ω l'ensemble $[\![1\,,6]\!]^2$ des 2-listes de $[\![1\,,6]\!]$ et pour probabilité sur Ω la probabilité uniforme. On avait noté X_1 (resp. X_2) la « valeur obtenue au premier (resp. deuxième) lancer » et $S = X_1 + X_2$ « la somme des deux faces obtenues »
Alors pour la variable aléatoire X_1, on a $\mathbf{E}(X_1) = \dfrac{7}{2}$ (nous verrons bientôt comment calculer $\mathbf{E}(S)$).

Remarque 24 Dans un jeu de hasard, l'espérance représente la chance de gain (algébrique) du joueur. Un jeu est dit *équitable* lorsque l'espérance de gain d'un joueur est nulle.

Théorème 25 - Propriétés de l'espérance. Soit X et Y deux variables aléatoires réelles finies sur Ω.

- **(i) Expression alternative.** $\mathbf{E}(X) = \sum_{\omega\in\Omega} \mathrm{P}(\{\omega\})\, X(\omega)$.
- **(ii) Linéarité.** Pour tous $a, b \in \mathbb{R}$, $\mathbf{E}(aX + bY) = a\,\mathbf{E}(X) + b\,\mathbf{E}(Y)$.
- **(iii) Positivité.** Si $X \geqslant 0$, alors $\mathbf{E}(X) \geqslant 0$. **Croissance.** Si $X \leqslant Y$, alors $\mathbf{E}(X) \leqslant \mathbf{E}(Y)$.
- **(iv) Inégalité triangulaire.** $|\mathbf{E}(X)| \leqslant \mathbf{E}\left(|X|\right)$.

Démonstration.

(i) La famille $(X = x)_{x\in X(\Omega)}$ est un système complet d'événements, ainsi

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

(ii) D'après l'expression précédente,

$$\mathbf{E}(aX + bY) = \sum_{\omega\in\Omega} \mathrm{P}(\{\omega\})\,(aX + bY)(\omega) = a \sum_{\omega\in\Omega} \mathrm{P}(\{\omega\})\, X(\omega) + b \sum_{\omega\in\Omega} \mathrm{P}(\{\omega\})\, Y(\omega) = a\,\mathbf{E}(X) + b\,\mathbf{E}(Y).$$

(iii) La positivité est évidente dans la définition de l'espérance. Pour la croissance, si $X \leqslant Y$, alors $Y - X \geqslant 0$ et donc, par linéarité, $\mathbf{E}(Y) - \mathbf{E}(X) = \mathbf{E}(Y - X) \geqslant 0$.

(iv) Conséquence de la croissance et de la linéarité, sachant que $-|X| \leqslant X \leqslant |X|$.

■

Exemple 26 Soit $m \in \mathbb{R}$. Si l'on considère m comme une variable aléatoire constante sur Ω, alors $\mathbf{E}(m) = m$.
En effet, notons X la variable aléatoire constante de valeur m sur Ω. Alors $X(\Omega) = \{m\}$ et $\mathrm{P}(X = m) = \mathrm{P}(\Omega) = 1$, ainsi $\mathbf{E}(X) = 1 \times m = m$.

Définition-théorème 27 - Variable centrée associée à une variable aléatoire. Soit X une variable aléatoire réelle finie. On appelle *variable centrée associée à* X la variable aléatoire centrée $X^+ = X - \mathbf{E}(X)$.

Démonstration. Par linéarité de l'espérance, $\mathbf{E}(X^+) = \mathbf{E}(X) - \mathbf{E}(\mathbf{E}(X)) = 0$. ■

Exemple 28 L'espérance de la variable aléatoire S de l'exemple 23 (somme des deux faces de dé obtenues) est 7.

3.2 Espérance des lois usuelles

Théorème 29 - Espérance de la loi uniforme sur $[\![1\,,\mathbf{n}]\!]$. Soit X une variable aléatoire telle que $X \sim \mathcal{U}\left([\![1\,,n]\!]\right)$. Alors

$$\mathbf{E}(X) = \frac{n+1}{2}.$$

Démonstration. ... ■

Exemple 30 - Espérance d'une loi uniforme translatée. Soit $a, b \in \mathbb{Z}$ avec $a \leqslant b$.
Soit X une variable aléatoire telle que $X \sim \mathcal{U}(\llbracket a\,, b \rrbracket)$. En étudiant $Y = X - a + 1$, montrer que

$$\mathbf{E}(X) = \frac{b+a}{2}$$

Théorème 31 - Espérance d'une loi de Bernoulli. Soit X une variable aléatoire telle que $X \hookrightarrow \mathcal{B}(p)$, avec $p \in [0\,, 1]$. Alors

$$\mathbf{E}(X) = p$$

Démonstration. ... ■

Exemple 32 - Exemple fondamental. Pour tout événement $A \in \mathscr{P}(\Omega)$, on rappelle que $\mathbb{1}_A$ est la variable aléatoire définie par

$$\forall \omega \in \Omega, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

On a alors $\mathbf{E}(\mathbb{1}_A) = \mathrm{P}(A)$.

Théorème 33 - Espérance d'une loi binomiale. Soit X une variable aléatoire telle que $X \sim \mathcal{B}(n, p)$, avec $n \in \mathbb{N}^*$ et $p \in [0\,, 1]$. Alors

$$\mathbf{E}(X) = np$$

Exemple 34 On lance 5 fois un dé équilibré à 6 faces dont 2 blanches et 4 noires. Combien de faces blanches peut-on espérer avec tirées au total, « en moyenne ».

3.3 Formules pour l'espérance de fonctions de variables aléatoires

Théorème 35 - Formule de transfert. Soit X une variable aléatoire finie et $f : X(\Omega) \longrightarrow \mathbb{R}$ une fonction. L'espérance de la variable aléatoire réelle finie $f(X)$ s'obtient comme

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

Démonstration. La famille $(X = x)_{x \in X(\Omega)}$ est un système complet d'événements, ainsi

$$\begin{aligned}
\mathbf{E}(f(x)) &= \sum_{\omega \in \Omega} \mathrm{P}(\{\omega\})\, f(X(\omega)) \\
&= \sum_{x \in X(\Omega)} \sum_{\omega \in (X = x)} \mathrm{P}(\{\omega\})\, f(X(\omega)) \\
&= \sum_{x \in X(\Omega)} \left(\sum_{\omega \in (X = x)} \mathrm{P}(\{\omega\}) \right) f(x) \\
&= \sum_{x \in X(\Omega)} \mathrm{P}(X = x)\, f(x).
\end{aligned}$$

■

✎ **En pratique** ✎ L'espérance de la variable aléatoire $f(X)$ est entièrement déterminée par f et la loi de X. On peut donc calculer $\mathbf{E}(f(X))$ directement à partir de la loi de X, sans avoir à déterminer celle de $f(X)$.

Exemple 36 On considère à nouveau le lancer d'un dé à 6 faces 2 fois (cf. exemple 23), pour lequel nous avons noté X_1 la valeur obtenue au premier lancer. Alors $\mathbf{E}(X_1^2) = \dfrac{91}{6}$.

3.4 Variance d'une variable aléatoire réelle finie : mesure d'une dispersion

Définition 37 - Variance/écart-type d'une variable aléatoire réelle.
Soit X une variable aléatoire réelle finie.

- On appelle *variance de* X le réel positif, noté $\mathbf{V}(X)$, défini par

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

- On appelle *écart-type de* X le réel $\sigma(X) = \sqrt{\mathbf{V}(X)}$.
- La variable X est dite *réduite* lorsque $\mathbf{V}(X) = 1$.

Remarque 38

- L'espérance de X est un indicateur de position, on peut alors se demander si les valeurs de X sont plutôt proches de cette valeur moyenne ou plutôt éloignées. Toute mesure de cette proximité à la moyenne est appelée un *indicateur de dispersion*. À proprement parler, l'écart naturel entre X et son espérance est $|X - \mathbf{E}(X)|$, ainsi l'écart moyen de X à sa moyenne est $\mathbf{E}(|X - \mathbf{E}(X)|)$, mais ce n'est pas l'indicateur de dispersion que les mathématiciens ont choisi d'utiliser couramment. Ces derniers lui ont préféré la variance, c'est-à-dire l'écart QUADRATIQUE moyen à la moyenne – « quadratique » parce que l'on passe au carré. Ce choix, moins naturel au premier abord, se justifie par une plus grande simplicité de manipulation. D'ailleurs, l'écart type ne ressemble-t-il pas à une norme euclidienne ?
- Quelle différence entre écart-type et variance ? Si X représente par exemple une longueur, $\mathbf{V}(X)$ représente une longueur AU CARRÉ. Ainsi il n'est pas possible de comparer directement la position moyenne $\mathbf{E}(X)$ et sa dispersion $\mathbf{V}(X)$ – un physicien dirait que l'espérance et la variance ne sont pas « homogène ». L'écart-type, au contraire, est homogène à une longueur, donc comparable à l'espérance, d'où son intérêt.

Théorème 39 - Propriétés de la variance. Soit X une variable aléatoire réelle finie.

(i) Formule de König-Huygens. $\mathbf{V}(X) = \mathbf{E}\left(X^2\right) - (\mathbf{E}(X))^2$.

(ii) Nullité. $\mathbf{V}(X) = 0$ si et seulement si $\mathrm{P}(X = \mathbf{E}(X)) = 1$, *i.e.* si et seulement si X est « presque sûrement » constante.

(iii) Effet d'une transformation affine. Pour tous $a, b \in \mathbb{R}$, $\mathbf{V}(aX + b) = a^2\,\mathbf{V}(X)$.

En particulier, si $\sigma(X) \neq 0$, alors la variable aléatoire $X^\star = \dfrac{X - \mathbf{E}(X)}{\sigma(X)}$ est centrée-réduite.

Démonstration.

(i) Par linéarité de l'espérance,

$$\mathbf{V}(X) = \mathbf{E}\left(X^2 - 2\,\mathbf{E}(X)X + (\mathbf{E}(X))^2\right) = \mathbf{E}\left(X^2\right) - \cancel{2}\,\mathbf{E}(X)\,\mathbf{E}(X) + \cancel{(\mathbf{E}(X))^2} = \mathbf{E}\left(X^2\right) - (\mathbf{E}(X))^2.$$

(ii) D'après la formule de transfert $\mathbf{V}(X) = \displaystyle\sum_{x \in X(\Omega)} \mathrm{P}(X = x)\,(x - \mathbf{E}(X))^2$, somme à termes positifs, ainsi

$$\begin{aligned}\mathbf{V}(X) = 0 \quad &\iff \quad \forall x \in X(\Omega), \quad \mathrm{P}(X = x)\,(x - \mathbf{E}(X))^2 = 0 \\ &\iff \quad \forall x \in X(\Omega), \quad \mathrm{P}(X = x) = 0 \quad \text{ou} \quad x = \mathbf{E}(X)\end{aligned}$$

$$\iff \quad \forall x \in X(\Omega)\backslash\{\mathbf{E}(X)\}, \quad \mathrm{P}(X = x) = 0$$
$$\iff \quad \mathrm{P}(X = \mathbf{E}(X)) = 1 \qquad \text{car} \sum_{x\in X(\Omega)} \mathrm{P}(X = x) = 1.$$

(iii) $(aX + b - \mathbf{E}(aX + b))^2 \underset{\substack{\text{Linéarité}\\ \text{de } \mathbf{E}}}{=} \left(aX + \cancel{b} - \left(a\,\mathbf{E}(X) + \cancel{b}\right)\right)^2 = (a(X - \mathbf{E}(X)))^2 = a^2(X - \mathbf{E}(X))^2$. Ainsi

$$\mathbf{V}(aX + b) = \mathbf{E}\left((aX + b - \mathbf{E}(aX + b))^2\right) = \mathbf{E}\left(a^2(X - \mathbf{E}(X))^2\right) \underset{\substack{\text{Linéarité}\\ \text{de } \mathbf{E}}}{=} a^2\,\mathbf{E}\left((X - \mathbf{E}(X))^2\right) = a^2\,\mathbf{V}(X).$$

Enfin, si $\sigma(X) \neq 0$,

$$\mathbf{E}(X^\star) = \mathbf{E}\left(\frac{X - \mathbf{E}(X)}{\sigma(X)}\right) = \frac{\mathbf{E}(X) - \mathbf{E}(X)}{\sigma(X)} = 0 \qquad \text{et} \qquad \mathbf{V}(X^\star) = \mathbf{V}\left(\frac{X - \mathbf{E}(X)}{\sigma(X)}\right) = \frac{\mathbf{V}(X)}{\sigma(X)^2} = 1.$$

■

Remarque 40

- Le résultat ci-dessus, concernant l'effet d'une transformation affine sur la variance, se comprend aisément si l'on a à l'esprit que la variance mesure la dispersion de la loi d'une variable aléatoire réelle. En effet, lorsque l'on translate une variable aléatoire d'une constante b, on ne modifie pas la dispersion de sa loi et la variance est donc constante : $\mathbf{V}(X + b) = \mathbf{V}(X)$.
- Centrer et réduire permet d'obtenir des variables représentant des données indépendantes des unités ou de l'échelle choisies et des variables ayant même moyenne et même dispersion. Autrement dit, des variables plus facilement comparables. Il s'agit d'un procédé de normalisation.

Exemple 41 Pour la variable aléatoire X_1 de l'exemple 23 (valeur obtenue au premier lancer d'un dé), $\mathbf{V}(X_1) = \dfrac{161}{12}$.

3.5 Variance des lois usuelles

Les formules ci-dessous sont à connaître, il faut en outre savoir les montrer car ce sont des cas modèles.

Théorème 42 - Variance de la loi uniforme sur $[\![1\,,n]\!]$. Soit X une variable aléatoire telle que $X \sim \mathcal{U}([\![1\,,n]\!])$. Alors

$$\mathbf{V}(X) = \frac{n^2 - 1}{12}.$$

Théorème 43 - Variance d'une loi de Bernoulli. Soit X une variable aléatoire telle que $X \sim \mathcal{B}(p)$, avec $p \in [0\,,1]$. Alors

$$\mathbf{V}(X) = p(1 - p).$$

Théorème 44 - Variance d'une loi binomiale. Soit X une variable aléatoire telle que $X \sim \mathcal{B}(n,p)$, avec $n \in \mathbb{N}^*$ et $p \in [0\,,1]$. Alors

$$\mathbf{V}(X) = np(1 - p)$$

Puisqu'une variable aléatoire de loi binomiale est la somme de n variables aléatoires indépendante de loi de bernouilli, on retrouvera cette dernière formule lorsque nous aurons étudié ce qu'est la « variance d'une somme ».

Formulaire – Lois usuelles discrètes

	Notation	Paramètres	$X(\Omega)$	$\mathrm{P}(X=k)$	$\mathbf{E}(X)$	$\mathbf{V}(X)$
Loi certaine		$a \in \mathbb{R}$	$\{a\}$	1	a	0
Loi uniforme	$\mathcal{U}(\llbracket 1, n \rrbracket)$	$n \in \mathbb{N}^*$	$\llbracket 1, n \rrbracket$	$\dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$
	$\mathcal{U}(\llbracket a, b \rrbracket)$	$(a,b) \in \mathbb{Z}^2$, $a \leqslant b$	$\llbracket a, b \rrbracket$	$\dfrac{1}{b-a+1}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a+1)^2-1}{12}$
Loi de Bernoulli	$\mathcal{B}(p)$	$p \in [0,1]$	$\{0,1\}$	$\begin{cases} p & \text{si } k=1 \\ 1-p & \text{si } k=0 \end{cases}$	p	$p(1-p)$
Loi binomiale	$\mathcal{B}(n,p)$	$n \in \mathbb{N}$, $p \in [0,1]$	$\llbracket 0, n \rrbracket$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$

Remarque 45 L'espérance et la variance d'une variable aléatoire suivant une loi binomiale de paramètres (n,p) sont respectivement égales à l'espérance et à la variance d'une variable aléatoire suivant une loi Bernoulli de paramètre p multipliée par n.

4 Couple de variables aléatoires

Jusque là on a construit une théorie des variables aléatoires à valeurs dans un ensemble quelconque. Dans ce chapitre, on va considérer des couples de variables aléatoires, qui forment des variables aléatoires à valeurs dans un produit cartésien, l'exemple le plus emblématique étant celui des couples de variables aléatoires réelles, qui forment donc une variable aléatoire à valeur dans $\mathbb{R}^2$, c'est-à-dire à valeurs vectorielles.

4.1 Différentes loi pour un couple de variables aléatoires

Définition 46 - Couple de variables aléatoires. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires sur Ω. Alors on appelle couple des variables aléatoires X et Y la variable aléatoire $Z : \Omega \to E \times F$, notée $Z = (X, Y)$, et définie par

$$Z : \omega \longmapsto (X(\omega), Z(\omega)).$$

Notons que pour $(x, y) \in E \times F$, l'événement $(X, Y) = (x, y)$ se réécrit $\{X = x\} \cap \{Y = y\}$.

Définition 47 - Loi conjointe de deux variables aléatoires. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires sur Ω. On appelle *loi conjointe* de X et Y la loi du couple (X, Y), c'est-à-dire l'application $P_{(X,Y)}$: $X(\Omega) \times Y(\Omega) \to [0, 1]$ définie, pour $(x, y) \in$ par

$$P_{(X,Y)} : (x, y) \longmapsto P\left(\{X = x\} \cap \{Y = y\}\right).$$

On note également

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

En particulier, si $A \subset E \times F$, alors on a $P_{X,Y}(A) = P((X, Y) \in A)$.

Définition 48 - Lois marginales d'un couple de variables aléatoires. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires sur Ω. On appelle *première loi marginale de* (X, Y) la loi de X (notée P_X) et *deuxième loi marginale de* (X, Y) la loi de Y (notée P_Y).

La proposition suivante montre comment déduire les lois marginales à partir de la loi conjointe :

Proposition 49 - Lois marginales d'un couple de variables aléatoires. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires sur Ω. On note $X(\Omega) = \{x_1, \dots, x_n\}$ et $Y(\Omega) = \{y_1, \dots, y_p\}$. Alors on a

$$\forall x \in X(\Omega), \quad P(X = x) = \sum_{k=1}^{p} P(X = x, Y = y_k)$$

et

$$\forall y \in Y(\Omega), \quad P(Y = y) = \sum_{i=1}^{n} P(X = x_i, Y = y)$$

✎ **En pratique** ✎ Une loi conjointe peut être représentée par un tableau à double entrée (voir exemple). La proposition précédente indique qu'on retrouve les lois marginales en sommant les lignes et les colonnes.

Exemple 50 - Un exemple de loi conjointe. Une urne contient 2 boules rouges, 5 boules noires et 3 boules vertes. On tire simultanément 3 boules. Soit X la variable aléatoire comptant le nombre de boules rouges obtenues et Y celle comptant le nombre de boules noires obtenues. Remplir le tableau suivant, décrivant la loi conjointe. Retrouver les lois marginales.

$X \backslash Y$	0	1	2	3	P_X
0					
1					
2					
P_Y					

Remarque 51 - Quel univers pour ce tirage ?. Notez qu'on ne prend plus vraiment la peine de définir Ω systématiquement. Ceci est courant en probabilité, où la loi d'une variable aléatoire (ce qui inclut son ensemble d'arrivée) est plus intéressante que l'univers Ω des possibles. Dans l'exemple précédent, Ω pourrait être l'ensemble des parties à 3 éléments de l'ensemble des 10 boules (numérotées au préalable). On a alors $|\Omega| = \binom{10}{3}$.

Définition 52 - Extension aux n-uplets . Soient $(E_i)_{i=1,\dots,n}$ et soient $X_i : \Omega \to E_i$ une famille de n variables aléatoires sur Ω. On peut définir le n-uplets $Z = (X_1, \dots, X_n)$, à valeurs dans $E_1 \times \dots \times E_n$. La loi conjointe des variables aléatoires $X_1, \dots, X_n$ est la loi de Z, tandis que la i-ème loi marginale est la loi de de X_i.

4.2 Variables aléatoires indépendantes

L'indépendance de deux variables aléatoires finies se décalque sur celle d'événements.

Définition 53 - Paire de variables aléatoires indépendantes. Deux variables aléatoires réelles discrètes X et Y sur Ω sont dites *indépendantes* lorsque, pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, les événements $(X = x)$ et $(Y = y)$ sont indépendants, *i.e.*

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

On note alors $X \perp\!\!\!\perp Y$.

L'interprétation en terme de la loi conjointe de (X, Y) est importante, même si ce n'est qu'une paraphrase de la définition :

✎ En pratique ✎

- En général, les lois marginales du couple (X, Y) ne déterminent pas sa loi conjointe, mais cela devient le cas lorsque X et Y sont indépendantes. Plus précsiément, la loi de distribution de probabilité conjointe de la variable aléatoire(X, Y) consiste à décrire les valeurs de $\mathrm{P}(X = x, Y = y)$ (cad de $\mathrm{P}((X = x) \cap (Y = y))$), pour $(x, y) \in X(\Omega) \times Y(\Omega)$. Dans le cas où les variables aléatoires X et Y sont indépendantes, ces valeurs sont données par $\mathrm{P}(X = x)\,\mathrm{P}(Y = y)$, autrement dit les lois marginales permettent de déterminer les lois conjointes.
- En conséquence, lorsque deux variables aléatoires sont indépendantes, les « lignes » (resp. « colonnes ») de leur loi mutuelle sont toutes proportionnelles.

Remarque 54 Observons que, pour $x \in X(\Omega)$,

$$\mathrm{P}((X = x) \cap (Y = y)) = \mathrm{P}(X = x)\,\mathrm{P}(Y = y) \iff \mathrm{P}(Y = y) = \frac{\mathrm{P}((X = x) \cap (Y = y))}{\mathrm{P}(X = x)} = \mathrm{P}_{(X=x)}(Y = y)\,.$$

Ainsi loi de Y sachant $(X = x)$ ne dépend aucunement de l'événement$(X = x)$, ce qui traduit que la connaissance de X n'apporte aucune information concernant Y, et inversement. C'est bien là la signification de l'indépendance.

Exemple 55 Soient X et Y deux variables aléatoires indépendantes suivant toutes les deux la loi uniforme sur $[\![1\,,n]\!]$. Déterminons la loi de $Z = X + Y$, dite *loi triangulaire*.

Le support du couple (X, Y) est $[\![1\,,n]\!]^2$ et, pour tout $(x, y) \in [\![1\,,n]\!]^2$, par indépendance,

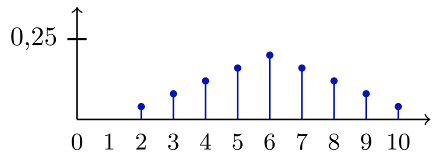
$$\mathrm{P}((X = x) \cap (Y = y)) = \mathrm{P}(X = x)\,\mathrm{P}(Y = y) = \frac{1}{n^2}.$$

Le support de Z est clairement $[\![2\,,2n]\!]$ et, pour tout $z \in Z(\Omega)$,

$$\mathrm{P}(Z = z) = \sum_{x=1}^{n} \mathrm{P}((X = x) \cap (Y = z - x)) = \sum_{x=1}^{n} \mathrm{P}(X = x)\,\mathrm{P}(Y = z - x)\,,$$

sachant que $\mathrm{P}(Y = z - x) = 0$ lorsque $z - x \notin [\![1\,,n]\!]$. Or $1 \leqslant z - x \leqslant n \iff x \leqslant z - 1$ et $z - n \leqslant x$.

- Si $z \leqslant n + 1$, la seconde condition est toujours vérifiée, puisque $x \geqslant 1$, et
$$\mathrm{P}(Z = z) = \sum_{x=1}^{z-1} \mathrm{P}(X = x)\,\mathrm{P}(Y = z - x) = \sum_{x=1}^{z-1} \frac{1}{n^2} = \frac{z-1}{n^2}.$$
- Si $z > n + 1$, la première condition est toujours vérifiée, et
$$\mathrm{P}(Z = z) = \sum_{x=z-n}^{n} \mathrm{P}(X = x)\,\mathrm{P}(Y = z - x) = \sum_{x=z-n}^{n} \frac{1}{n^2} = \frac{2n - z + 1}{n^2}.$$



Loi triangulaire pour $n = 5$.

Théorème 56 - Caractérisation de l'indépendance d'une paire de variables aléatoires. Deux variables aléatoires réelles discrètes X et Y sur Ω sont indépendantes si et seulement si, pour toutes parties A de $X(\Omega)$ et B de $Y(\Omega)$, les événements $(X \in A)$ et $(Y \in B)$ sont indépendants, *i.e.*

$$\mathrm{P}((X \in A) \cap (Y \in B)) = \mathrm{P}(X \in A)\,\mathrm{P}(Y \in B)\,.$$

Remarque 57 Il y a un moyen simple de repérer des variables aléatoires qui ne sont pas indépendantes. En effet, si X et Y sont indépendantes,

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \quad \mathrm{P}((X = x) \cap (Y = y)) = \mathrm{P}(X = x)\,\mathrm{P}(Y = y) \neq 0.$$

Ainsi, un couple de variables aléatoires dont la loi conjointe contient des « termes » nuls ne saurait être composé de variables aléatoires indépendantes.

Plus généralement, pour un ensemble fini de variables aléatoires, à l'instar de l'indépendance d'événements, il faut distinguer l'indépendance mutuelle et l'indépendance deux à deux.

Définition-théorème 58 - Ensemble fini de variables aléatoires mutuellement indépendantes.
Soit $X_1, \ldots, X_n$ des variables aléatoires finies sur Ω.

- Les variables aléatoires $X_1, \ldots, X_n$ sont dites *(mutuellement) indépendantes* lorsque, pour tous $x_1 \in X_1(\Omega)$, ..., $x_n \in X_n(\Omega)$,
$$\mathrm{P}((X_1 = x_1) \cap \ldots \cap (X_n = x_n)) = \mathrm{P}(X_1 = x_1) \ldots \mathrm{P}(X_n = x_n)\,.$$
Cette définition équivaut à ce que, pour tous $x_1 \in X_1(\Omega)$, ..., $x_n \in X_n(\Omega)$, les événements $(X_1 = x_1)$, ..., $(X_n = x_n)$ soient (mutuellement) indépendants.
- Les variables aléatoires $X_1, \ldots, X_n$ sont (mutuellement) indépendantes si et seulement si, pour toutes parties $A_1 \subset X_1(\Omega)$, ..., $A_n \subset X_n(\Omega)$, les événements $(X_1 \in A_1)$, ..., $(X_n \in A_n)$ sont (mutuellement) indépendants. En particulier, le cas échéant,
$$\mathrm{P}((X_1 \in A_1) \cap \ldots \cap (X_n \in A_n)) = \mathrm{P}(X_1 \in A_1) \ldots \mathrm{P}(X_n \in A_n)\,.$$

L'équivalence évoquée résulte de ce que tout sous-ensemble d'un ensemble fini de variables aléatoires (mutuellement) indépendantes est bien sûr un ensemble de variables aléatoires (mutuellement) indépendantes.

✖ Attention ! ✖ $\boxed{\text{(Mutuellement) indépendants} \quad \Longrightarrow \quad \text{Deux à deux indépendants}}$ – mais la réciproque est fausse !

Exemple 59 Soit $p_1, \ldots, p_n \in [0\,,1]$ et $X_1, \ldots, X_n$ des variables aléatoires mutuellement indépendantes de lois respectives $\mathscr{B}(p_1), \ldots, \mathscr{B}(p_n)$. Alors $X_1 \cdots X_n \sim \mathscr{B}(p_1 \ldots p_n)$.
En effet, la variable aléatoire $X_1 \cdots X_n$ a pour image $\{0, 1\}$ et, ayant $\mathrm{P}(X_1 \cdots X_n = 0) = 1 - \mathrm{P}(X_1 \cdots X_n = 1)$, il suffit de vérifier que

$$\mathrm{P}(X_1 \cdots X_n = 1) = \mathrm{P}((X_1 = 1) \cap \ldots \cap (X_n = 1)) \underset{\text{Indép.}}{=} \mathrm{P}(X_1 = 1) \cdots \mathrm{P}(X_n = 1) = p_1 \ldots p_n.$$

Le résultat suivant ouvre une connexion importante entre deux lois usuelles :

Théorème 60 - Somme de loi de Bernouilli.
Soient $(X_i)_{i=1,\ldots,n}$ une famille de n variables aléatoires de Bernoulli de paramètre de paramètre $p \in [0, 1]$. Alors

$$X_1 + \ldots + X_n \sim \mathcal{B}(n, p).$$

Il s'agit d'une redite de ce qui a déjà été dit : la variable aléatoire$X_1 + \ldots + X_n$ compte le nombre de succés lors de la répétition (indépendante) de n expériences aléatoires de type Bernoulli, c'est-à-dire chacune pouvant donner lieu à un succès ou à un échec, avec probabilité p et $1 - p$.

Théorème 61 - Indépendances des images de variables aléatoires indépendantes par des fonctions.
Soient X et Y deux variables aléatoires réelles indépendantes, ainsi que f et g deux fonctions de $\mathbb{R}$ dans $\mathbb{R}$. Alors les variables aléatoires $f(X)$ et $g(Y)$ sont indépendantes.

Le résultat précédent se généralise au cas d'un n-uplets de variables aléatoires :

Théorème 62 - Lemme des coalitions : le cas d'une famille.
Soit $X_1, \ldots, X_n$ des variables aléatoires réelles discrètes sur Ω, un entier m tel que $1 < m < n$, et $f : \mathbb{R}^m \longrightarrow \mathbb{R}$ et $g : \mathbb{R}^{n-m} \longrightarrow \mathbb{R}$ deux fonctions.
Si $X_1, \ldots, X_n$ sont indépendantes, il en va alors de même de $f(X_1, \ldots, X_m)$ et $g(X_{m+1}, \ldots, X_n)$.

Exemple 63 Soit X, Y et Z trois variables aléatoires indépendantes sur Ω. Alors $X + Y$ et Z le sont aussi.
Ici, $f : \mathbb{R}^2 \longrightarrow \mathbb{R}, (x, y) \longmapsto x + y$ et $g : \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto x$.

Démonstration. Contentons-nous de traiter le cas particulier de deux variables aléatoires indépendantes X et Y sur Ω et montrons que pour deux fonctions f et g définies sur $\mathbb{R}$, $f(X)$ et $g(Y)$ sont également indépendantes. Or, pour tous $A \subset f(X)(\Omega)$ et $B \subset g(Y)(\Omega)$,

$$\begin{aligned}
\mathrm{P}((f(X) \in A) \cap (g(Y) \in B)) &= \mathrm{P}\big((X \in f^{-1}(A)) \cap (Y \in g^{-1}(B))\big) \\
&= \mathrm{P}\big(X \in f^{-1}(A)\big)\,\mathrm{P}\big(Y \in g^{-1}(B)\big) && \text{indépendance de } X \text{ et } Y \\
&= \mathrm{P}(f(X) \in A)\,\mathrm{P}(g(Y) \in B)\,.
\end{aligned}$$

■

Terminons par un résultat concernant l'espérance de deux variables aléatoires indépendantes.

Théorème 64 - Espérance du produit de deux variables aléatoires réelles indépendantes.
Soient X et Y deux variables aléatoires réelles indépendantes sur Ω alors

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

✖ Attention ! ✖ L'identité « $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$ » est fausse en général. Par exemple, pour toute variable aléatoire X de loi uniforme sur $\{-1, 1\}$, $\mathbf{E}(X) = \frac{1}{2} \times 1 + \frac{1}{2} \times (-1) = 0$ et donc $\mathbf{E}(X)^2 = 0$, tandis que $\mathbf{E}\left(X^2\right) = \mathbf{E}(1) = 1$.

4.3 Covariance de variables aléatoires

La *covariance* de deux variables aléatoires réelles est destinée à mesurer le défaut par rapport à l'indépendance.

Définition 65 - Covariance. Soit X et Y deux variables aléatoires réelles finies sur Ω.
On appelle *covariance* de X et Y le réel suivant :

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

Deux variables aléatoires de covariance nulle sont dites *décorrélées*.

Théorème 66 Soient X, Y, Z, X_1,..., X_n des variables aléatoires réelles finies sur Ω. Alors on a :

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

(ii) (Positivité) $\mathrm{Cov}(X,X) = \mathbf{V}(X) \geqslant 0$. **(iii) (Symétrie)** $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.

(iv) (Linéarité à gauche/droite)

$$\mathrm{Cov}(aX+bY,Z) = a\,\mathrm{Cov}(X,Z) + b\,\mathrm{Cov}(Y,Z) \quad \text{et} \quad \mathrm{Cov}(Z,aX+bY) = a\,\mathrm{Cov}(Z,X) + b\,\mathrm{Cov}(Z,Y).$$

(v) (Lien avec la variance) $\mathbf{V}(X+Y) = \mathbf{V}(X) + 2\,\mathrm{Cov}(X,Y) + \mathbf{V}(Y)$.

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

(vi) Si X et Y sont INDÉPENDANTES, alors $\mathrm{Cov}(X,Y)=0$ et $\mathbf{V}(X+Y) = \mathbf{V}(X) + \mathbf{V}(Y)$.

Plus généralement, si $X_1,\dots,X_n$ sont DEUX À DEUX indépendantes, alors $\mathbf{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbf{V}(X_i)$.

Le résultat du point (vi) traduit bien l'idée que la covariance mesure la dépendance entre X et Y.

✖ **ATTENTION !** ✖ La réciproque de (vi) est fausse ! La covariance de X et Y peut être nulle sans que X et Y soient indépendantes. Considérons par exemple une variable aléatoire X suivant la loi uniforme sur $\{-1,0,1\}$ et la variable aléatoire $Y = 1 - X^2$ de support $Y(\Omega) = \{0,1\}$. Les variables X et Y ne sont pas indépendantes, puisque

$$\mathrm{P}((X=1)\cap(Y=1)) = \mathrm{P}((X=1)\cap(X=0)) = \mathrm{P}(\varnothing) = 0 \quad \text{et} \quad \mathrm{P}(X=1)\,\mathrm{P}(Y=1) = \mathrm{P}(X=1)\,\mathrm{P}(X=0) = \frac{1}{9}.$$

En revanche, puisque $XY = X - X^3$ est constante égale à 0 et puisque X est d'espérance nulle,

$$\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) = \mathbf{E}(0) - 0 \times \mathbf{E}(Y) = 0.$$

Démonstration.

(i) $(X-\mathbf{E}(X))\,(Y-\mathbf{E}(Y)) = XY - \underbrace{\mathbf{E}(Y)}_{\text{constante}}\,X - \underbrace{\mathbf{E}(X)}_{\text{constante}}\,Y + \underbrace{\mathbf{E}(X)\,\mathbf{E}(Y)}_{\text{constante}}$, ainsi, par linéarité de $\mathbf{E}$,

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

(iii) La symétrie est évidente, au vu de la définition.

(iv) Découle de la linéarité du produit et de l'espérance.

(v) Par définition de $\mathbf{V}$ et par linéarité de $\mathbf{E}$:

$$\mathbf{V}(X+Y) = \mathbf{E}\left((X-\mathbf{E}(X)+Y-\mathbf{E}(Y))^2\right)\dots$$
$$\dots = \mathbf{E}\left((X-\mathbf{E}(X))^2 + 2\,(X-\mathbf{E}(X))\,(Y-\mathbf{E}(Y)) + (Y-\mathbf{E}(Y))^2\right) = \mathbf{V}(X) + 2\,\mathrm{Cov}(X,Y) + \mathbf{V}(Y).$$

(vi) Si X et Y sont indépendantes, alors $\mathrm{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\,\mathbf{E}(Y) = 0$ et conclut finalement grâce à (iii). Plus généralement, dans la formule $(\star)$ de (v), toutes les covariances $\mathrm{Cov}(X_i,X_j)$ sont nulles.

■

Remarque 67 La formule ($\star$) du point (v) rappelle fortement la formule classique

$$\forall x_1, \ldots, x_n \in \mathbb{R}, \quad \left(\sum_{k=1}^{n} x_k\right)^2 = \sum_{k=1}^{n} x_k^2 + 2 \sum_{1 \leqslant i < j \leqslant n} x_i x_j,$$

qui elle-même est une généralisation de l'identité remarquable $(x+y)^2 = x^2 + 2xy + y^2$. Il s'agit ainsi de retenir que

« Le carré de la somme est égal à la somme des carrés plus la somme des doubles produits. »

Voici un exemple classique montrant la force de ces outils :

Exemple 68 Retrouver la variance d'une loi binomiale en la voyant comme une somme de loi de Bernoulli indépendantes.

Voyons sur un exemple plus dur comment les formules précédentes permettent en pratique de calculer des variances ou covariances en évitant de recourir à la formule de transfert.

Exemple 69 Une urne contient des boules de N types différents (bleues, jaunes, ...) mais indiscernables au toucher. Le type i est présent en proportion p_i. On tire n boules au hasard avec remise et on note X_i le nombre de boules de type i obtenues. Pour tous $1 \leqslant i < j \leqslant N$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$.

En effet, pour tout $1 \leqslant i \leqslant N$, les tirages ayant lieu avec remise, la variable aléatoire X_i suit la loi binomiale $\mathscr{B}(n, p_i)$. En outre, pour tous $1 \leqslant i < j \leqslant N$, $X_i + X_j$ représente le nombre de boules de type « i ou j » tirées. Ces dernières étant en proportion $p_i + p_j$, la variable aléatoire $X_i + X_j$ suit la loi binomiale $\mathscr{B}(n, p_i + p_j)$. Or,

$$\mathbf{V}(X_i + X_j) = \mathbf{V}(X_i) + 2\,\mathrm{Cov}(X_i, X_j) + \mathbf{V}(X_j),$$

d'où

$$\mathrm{Cov}(X_i, X_j) = \frac{1}{2}\left[\mathbf{V}(X_i + X_j) - \mathbf{V}(X_i) - \mathbf{V}(X_j)\right] = \frac{n}{2}\left[(p_i + p_j)(1 - p_i - p_j) - p_i(1 - p_i) - p_j(1 - p_j)\right]$$

et le résultat annoncé, après simplification.

5 Inégalités probabilistes et conséquences

Nous entrons dans un nouveaux domaine, celui des résultats généraux impliquant l'espérance et la variance d'une variable aléatoire réelle sans pour autant connaîre sa loi !

Théorème 70 - Inégalité de Markov. Soient X une variable aléatoire réelle finie positive sur Ω et $a > 0$. Alors on a :

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

Exemple 71 Une usine produit en moyenne 50 pièces par jour, mais la production peut varier selon les aléas. Que dire de la probabilité que la production dépasse 75 pièces un jour donné ?

Théorème 72 - Inégalité de Bienaymé-Tchebytchev. Soient X une variable aléatoire réelle finie sur Ω et $a > 0$. Alors on a :

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

Corollaire 73 - Loi faible des grands nombres. Soient $(X_i)_{i=1,\dots,n}$ une famille de variables aléatoires réelles finies sur Ω, indépendantes mutuellement, et de même loi. Notons $E(X)$ et $V(X)$ la variance et l'espérance communes à ces variables aléatoires . Soit $\varepsilon > 0$, alors on a

$$\mathrm{P}(|\frac{X_1 + \ldots + X_n}{n} - E(X)| \geqslant \varepsilon) \leqslant \frac{V(X)}{n\varepsilon^2}.$$

En particulier

$$\lim_{n\to+\infty} \mathrm{P}(|\frac{X_1 + \ldots + X_n}{n} - E(X)| \geqslant \varepsilon) = 0.$$

Exemple 74 On effectue une suite de lancers d'un dé à 6 faces, quel nombre de lancer suffit-il de faire pour pouvoir affirmer avec un risque d'erreur inférieur à 5% que la fréquence d'apparaition du 6 est comprise entre $\frac{1}{6} - 0.01$ et $\frac{1}{6} + 0.01$?