Chapitre 23 : Variables aléatoires

MPSI Lycée Camille Jullian

24 mai 2023

Un mèdecin annonce à un de ses patients :
« J'ai une bonne et une mauvaise nouvelle, je commence par la mauvaise.
Vous avez une maladie grave dont on ne guérit qu'avec une probabilité $\frac{1}{10}$*. »*
« Et la bonne nouvelle docteur ? »
« Mes neuf derniers patients sont morts... »

Pour introduire cette nouvelle notion, absolument fondamentale en probabilités (tellement d'ailleurs que vous n'entendrez plus parler que de ça jusqu'à la fin de votre prépa, à peu de choses près), prenons un exemple très classique : on lance simultanément (ou successivement, ça ne change pas grand chose) quatre pièces équilibrées. L'univers Ω des résultats de l'expérience est un ensemble à $2^4 = 16$ éléments constitué des suites de quatre « Pile ou Face » (on pourra ainsi noter PPFP le résultat de l'expérience consistant à tirer Pile au premier lancer, à nouveau Pile au deuxième, Face au troisième et Pile au quatrième). On peut naturellement déjà se poser plein de questions concernant cet univers, mais il arrive qu'on ait envie de considérer des résultats qui ne sont pas directement ceux de l'expérience. Par exemple, on veut étudier plus particulièrement le nombre de Piles obtenus lors de ces quatre lancers de pièces. Ce nombre de Piles est un entier directement associé au résultat de l'expérience (si vous connaissez le résultat, vous connaissez le nombre de Piles). Eh bien, une variable aléatoire, c'est exactement ça : une application qui, à chaque élément de Ω, associe un nombre réel. Ces variables aléatoires donnent fréquemment lieu à des développements mathématiques faisant intervenir des techniques de calcul qui pour nous se limiteront cette année aux calculs de sommes (et l'an prochain aux calculs de séries, quand vous aurez étendu la notion de variable aléatoire à des univers infinis). Par ailleurs, certaines situations classiques en probabilités donnent lieu à l'apparition de variables aléatoires usuelles (schémas de Bernoulli notamment) pour lesquelles nous démontrerons quelques résultats dans ce chapitre qui nous éviteront de refaire en permanence les mêmes calculs (d'espérance ou de variance notamment).

Objectifs du chapitre :

- être capable d'appliquer ses connaissances en dénombrement pour déterminer correctement la loi d'une variable aléatoire.
- maitriser les techniques de calcul de l'espérance, de la variance et de la covariance.
- savoir repérer sans hésitation les variables aléatoires suivant une loi binômiale, et celles qui n'en suivent pas une.

1 Variables aléatoires finies.

Comme dans le chapitre de probabilités précédent, on se contentera de travailler sur des univers finis dans tout ce chapitre.

1.1 Définition, notations.

Définition 1. Soit Ω l'univers associé à une expérience aléatoire. Une **variable aléatoire** (réelle) X sur Ω est une application $X : \Omega \to \mathbb{R}$.

Remarque 1. On note $X(\Omega)$ l'**univers image** de la variable aléatoire X, c'est-à-dire l'ensemble des valeurs prises par X (qui est bien l'image de l'ensemble Ω par l'application X).

Exemple : L'application qui à chaque français associe sa taille est une variable aléatoire sur l'ensemble de la population française. On a ici $X(\Omega) \subset [0, 3]$ (j'ai pris large). Sur le même univers, on pourrait définir une autre variable aléatoire associant par exemple à chaque français son groupe de musique préféré. Mais ce ne serait bien sûr pas une variable aléatoire réelle (les résultats ne seraient même pas numériques) et n'aurait donc aucun intérêt mathématique (calculer la moyenne du groupe de musique préféré des français n'aurait par exemple absolument aucun sens, alors que calculer une taille moyenne est nettement plus envisageable).

Les trois exemples numérotés qui suivent seront repris dans l'ensemble du chapitre pour illustrer les différentes notions abordées et donner des exemples de calculs (d'espérance et de variance notamment).

Exemple 1 : On lance deux dés équilibrés à six faces, et on note X la somme des résultats obtenus sur les deux dés. On aura $X(\Omega) = \{2, 3, \dots, 12\}$.

Exemple 2 : On reprend l'exemple explicité en introduction (lancer de pièces équilibrées) mais avec six pièces, en notant X le nombre de Piles obtenus. X est une variable aléatoire, et $X(\Omega) = \{0, 1, 2, 3, 4, 5, 6\}$.

Exemple 3 : Dans une urne se trouvent cinq jetons numérotés de 1 à 5. On en tire 3 simultanément et on note X le plus petit des trois numéros tirés. On a ici $X(\Omega) = \{1, 2, 3\}$ (si on tire trois jetons simultanément, le plus petit ne peut pas être plus grand que 3, ce serait bien sûr différent si on effectuait des tirages successifs avec remise).

Définition 2. Soit X une variable aléatoire sur un univers Ω. On note habituellement $(X = k)$, l'événement $\{\omega \in \Omega \mid X(\omega) = k\}$. On utilisera de même la notation $(X \leqslant k)$ pour l'événement $\{\omega \in \Omega \mid X(\omega) \leqslant k\}$ (et $(X \geqslant k)$, $(X < k)$ et $(X > k)$ pour des évènements similaires). Plus généralement, on peut noter $(X \in A)$ (on croise aussi $\{X \in A\}$) l'évènement $\{\omega \in \Omega \mid X(\Omega) \in A\}$ pour tout sous-ensemble A de $X(\Omega)$. Autrement dit, $(X \in A) = X^{-1}(A)$.

Exemple : Ainsi, si on reprend l'exemple du lancer de six pièces (et toujours avec X le nombre de Piles), on pourra écrire $\mathbb{P}(X = 1)$ pour désigner la probabilité de l'événement « On a obtenu exactement une fois Pile lors de nos six lancers de pièce », raccourci extrêmement pratique. Ici, en l'occurrence, $\mathbb{P}(X = 1) = \dfrac{6}{2^6} = \dfrac{6}{64}$ (il y a six cas sur les 64 possibles pour lesquels on obtient un seul Pile, il faut choisir quelle pièce est tombée sur Pile). De façon similaire, $\mathbb{P}(X \geqslant 5) = \dfrac{7}{64}$ (sept cas valables sur 64, les six où on a obtenu cinq Piles et un Face et l'unique tirage ayant donné six Piles).

Proposition 1. Soient X et Y deux variables aléatoires sur un même univers Ω, alors $X + Y$, XY, λX (où λ est un réel quelconque), $\max(X, Y)$ et $\min(X, Y)$ sont également des variables aléatoires. Plus généralement, X^2, X^3 ou n'importe quelle expression de la forme $f(X)$, où f est une fonction définie sur $X(\Omega)$, est une variable aléatoire sur Ω.

Pas de démonstration, c'est évident, ce sont aussi des applications de Ω dans $\mathbb{R}$. Il faut simplement avoir à l'esprit qu'on peut « effectuer des calculs » avec des variables aléatoires, ce qui interviendra très régulièrement dans les calculs de variance.

1.2 Loi d'une variable aléatoire.

Pour pouvoir dire des choses intéressantes sur une variable aléatoire, la première chose à étudier est la probabilité d'apparition de chacun des résultats possibles :

Définition 3. Soit X une variable aléatoire, l'application $\mathbb{P}_X : \left\{ \begin{array}{ccc} \mathcal{P}(X(\Omega)) & \to & [0,1] \\ A & \mapsto & \mathbb{P}(X \in A) \end{array} \right.$ définit une loi de probabilité sur l'univers-image $X(\Omega)$ appelée **loi de probabilité** de la variable X. Elle est caractérisée par la donnée des probabilités $\mathbb{P}(X = k)$, pour toutes les valeurs k appartenant à $X(\Omega)$.

Remarque 2. Pour calculer la loi d'une variable aléatoire, il suffit donc de déterminer toutes les valeurs qu'elle peut prendre, puis de calculer la probabilité de chaque résultat, ce qui revient en fait à définir une distribution de probabilités $(\mathbb{P}(X = k))$ sur l'ensemble $X(\Omega)$ des valeurs prises par X. On présente souvent les résultats sous forme d'un tableau, comme nous allons le faire en reprenant nos trois exemples cités plus haut. Dans certains cas, on peut également donner une formule exprimant $\mathbb{P}(X = k)$ en fonction de k, mais une telle formule valable pour toutes les valeurs possibles de k n'est pas toujours évidente à déterminer (quand il y en a une !). Ce sera le cas des variables aléatoires usuelles que nous étudierons à la fin de ce chapitre.

Définition 4. Deux variables aléatoires X et Y ayant la même loi (et donc le même univers image) sont dites **identiquement distribuées**, on le note $X \sim Y$.

Remarque 3. La relation ainsi définie est de façon évidente une relation d'équivalence sur l'ensemble des variables aléatoires définies sur un même univers Ω.

Exemple 1 : Pour le lancer de deux dés, la somme des résultats obtenus a pour loi :

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

Démonstration. C'est un calcul de probabilités tout à fait classique. Lorsqu'on lance deux dés à six faces, on a $6^2 = 36$ résultats équiprobables possibles (en considérant bien sûr ici comme résultats les paires d'entiers compris entre 1 et 6, par exemple $(2, 4)$ ou $(5, 5)$). Sur ces 36 résultats, un seul donne une valeur de X égale à 2 (on a tiré un double 1), deux donnent une valeur de X égale à 3 (on a tiré un 1 sur un des deux dés et un 2 sur l'autre, ce qui correspond bien à deux cas différents), etc. $\square$

Exemple 2 : Reprenons notre exemple du nombre de Piles sur six lancers de pièces. On peut présenter la loi de X sous la forme d'un tableau :

k	0	1	2	3	4	5	6
$\mathbb{P}(X = k)$	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{15}{64}$	$\frac{20}{64}$	$\frac{15}{64}$	$\frac{6}{64}$	$\frac{1}{64}$

Démonstration. On a déjà fait une bonne partie du travail plus haut : il y a $2^6 = 64$ résultats différents possibles, un seul (FFFFFF) qui correspond à l'évènement $X = 0$, six qui donnent $X = 1$ (on doit choisir la pièce qui tombe sur Pile) etc. Les plus observateurs d'entre vous remarqueront qu'on peut ici trouver une formule générale : $\mathbb{P}(X = k) = \dfrac{\binom{6}{k}}{2^6}$. On retrouvera ce type de formule plus loin dans le chapitre quand nous définirons les lois binomiales. $\square$

Exemple 3 : Pour déterminer la loi, le plus simple est de dénombrer tous les cas possibles (il n'y en a que 10), même si on peut exprimer les probabilités à l'aide de coefficients binomiaux (par exemple, pour avoir $X = 1$, il faut tirer le jeton 1 puis deux autres parmi les 4 restants, soit $\binom{4}{2}$ tirages favorables sur les $\binom{5}{3}$, donc 6 tirages sur 10). On obtient en tout cas :

k	1	2	3
$\mathbb{P}(X = k)$	$\frac{6}{10}$	$\frac{3}{10}$	$\frac{1}{10}$

Proposition 2. Les événements $(X = k)_{k \in X(\Omega)}$ forment un système complet d'événements. On a donc $\sum_{k \in X(\Omega)} \mathbb{P}(X = k) = 1$.

Démonstration. Ces événements sont incompatibles (on ne peut pas avoir à la fois $X(\omega) = k$ et $X(\omega) = k'$ pour des valeurs différentes de k et k'). Leur réunion est bien Ω tout entier puisque chaque élément ω de Ω a une image par X. □

2 Moments d'une variable aléatoire.

Lorsqu'on s'intéresse à une variable aléatoire pouvant prendre un grand nombre de valeurs (et même dans les autres cas !), il peut être intéressant de donner, en plus de la loi de la variable qui ne sera pas toujours une donnée très lisible (imaginez par exemple qu'on reprenne l'étude de la somme des résultats d'un lancer de dés de notre exemple 1, mais avec dix dés au lieu de deux, bon courage pour trouver la loi de X), des caractéristiques d'ensemble de cette loi, comme la moyenne des valeurs prises (pondérées par leur probabilité d'apparition). Ces paramètres sont les mêmes que ceux qu'on étudie en statistiques, et se rangent dans deux catégories donnant des informations complémentaires :

- les paramètres de **position** donnent une idée globale des valeurs prises par la variable aléatoire. Dans cette catégorie se situent la moyenne (qu'on privilégiera en pratique dans nos calculs) mais aussi par exemple la médiane (valeur qui sépare en deux parties égales l'ensemble des valeurs prises par la variable aléatoire, en tenant compte des probabilités d'apparition de ces valeurs). Dans certains cas, connaitre la médiane est plus pertinent que la moyenne (par exemple, si on étudie les salaires des français, un calcul de moyenne est peu représentatif car les plus gros salaires sont tellement énormes qu'ils contribuent « trop » à augmenter la moyenne, quand la médiane sera plus représentative du salaire d'un français « moyen »).
- les paramètres de **dispersion** servent à mesurer la répartition des valeurs prises par la variable aléatoire autour du paramètre de position (moyenne ou médiane). C'est le cas de l'écart-type (associé en général au calcul de moyenne) ou de l'écart absolu moyen (associé au calcul de médiane, que nous n'étudierons pas). À moyenne fixée, plus l'écart-type est gros, plus les valeurs prises par la variable sont (en moyenne !) « loin » de la moyenne.

2.1 Espérance.

Définition 5. L'**espérance** d'une variable aléatoire X est définie par $\mathbb{E}(X) = \sum_{k \in X(\Omega)} k\mathbb{P}(X = k)$.

Remarque 4. Il s'agit bel et bien d'un calcul de moyenne pondérée, les coefficients étant égaux aux probabilités $\mathbb{P}(X = k)$. La somme des coefficients vaut donc 1, ce qui explique qu'il n'y ait pas de division à faire pour calculer cette moyenne. Si on dispose de la loi de X sous forme de tableau, le calcul de l'espérance est donc très simple, on multiplie chaque valeur par la probabilité qui se trouve juste en-dessous et on fait la somme.

Exemple 1 : On calcule un peu laborieusement l'espérance de la somme de nos deux chiffres lors du lancer de dés : $\mathbb{E}(X) = \dfrac{2+6+12+20+30+42+40+36+30+22+12}{36} = \dfrac{252}{36} = 7$. Ce résultat tout simple vous semble logique ? On reviendra dessus un peu plus loin (il y a une méthode plus simple pour retrouver le résultat). On ne pourrait d'ailleurs sûrement pas calculer par cette méthode l'espérance de la somme des chiffres obtenus lorsqu'on lance dix dés (on serait incapables de donner la loi) alors qu'elle devrait logiquement valoir 35.

Exemple 2 : Reprenons l'exemple de six lancers de pièce, où X était le nombre de Piles obtenu. On aura $\mathbb{E}(X) = \dfrac{0+6+30+60+60+30+6}{64} = \dfrac{192}{64} = 3$. Le résultat est bien conforme à l'intuition qu'on a de la moyenne de la variable aléatoire X.

Exemple 3 : Pour nos cinq jetons dans l'urne, l'espérance vaut $\mathbb{E}(X) = \dfrac{6+6+3}{10} = \dfrac{3}{2}$. Tout ce qu'on peut constater c'est que c'est cohérent puisque l'espérance se situe entre les valeurs extrêmes prises par la variable X (et plus proche de 1 qui est la valeur qu'on obtiendra la majorité du temps).

Proposition 3. *La variable aléatoire constante $X : \omega \mapsto a$, $a \in \mathbb{R}$, a une espérance $\mathbb{E}(X) = a$.*

Démonstration. C'est évidemment complètement trivial puisque dans ce cas la variable (qui ne l'est pas vraiment) prend comme unique valeur a, avec probabilité 1. En fait, cette propriété sert simplement à supprimer les espérances quand elles sont autour d'une constante dans certains calculs formels. □

Définition 6. Soit A un évènement inclus dans notre univers Ω. La **variable indicatrice de l'évènement** A, notée $\mathbf{1}_A$, est la variable aléatoire définie par $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$, et $\mathbf{1}_A(\omega) = 0$ sinon.

Proposition 4. *L'espérance de la variable aléatoire indicatrice $\mathbf{1}_A$ vaut $\mathbb{P}(A)$.*

Démonstration. C'est bien parce que je suis consciencieux que je fais une preuve. La loi de $\mathbf{1}_A$ est extrêmement simple, elle prend la valeur 1 si $\omega \in A$, c'est-à-dire avec probabilité $\mathbb{P}(A)$, et 0 sinon, donc avec probabilité $1 - \mathbb{P}(A)$. L'espérance vaut donc bien $\mathbb{P}(A)$. En fait, ces variables indicatrices jouent un rôle essentiel dans le principe des schémas de Bernoulli que nous évoquerons dans la dernière partie du cours, et permettent aussi de calculer plus simplement certaines espérances, comme nous allons le voir ci-dessous. □

Théorème 1. Linéarité de l'espérance.

Si X et Y sont deux variables aléatoires définies sur le même univers Ω, et a, b deux réls, on a $\mathbb{E}(aX+bY) = a\mathbb{E}(X)+b\mathbb{E}(Y)$. En particulier, on aura toujours $\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y)$, et $\mathbb{E}(aX) = a\mathbb{E}(X)$, ou encore en utilisant l'espérance d'une variable constante calculée plus haut, $\mathbb{E}(X+b) = \mathbb{E}(X)+b$.

Démonstration. La preuve est très facile en écrivant légèrement différemment la formule définissant l'espérance : $\mathbb{E}(X) = \sum_{k \in X(\Omega)} k \times \mathbb{P}(X = k) = \sum_{\omega \in \Omega} X(\omega) \times \mathbb{P}(\omega)$ (on calcule exactement la même somme dans les deux cas, la fomule usuelle regroupant simplement en un seul terme tous les résultats de l'expérience aléatoire donnant la même valeur pour la variable X, ce qui est en pratique nettement plus simple pour effectuer le calcul puisqu'on a beaucoup moins de termes à gérer). Sous la dernière forme écrite, $\mathbb{E}(aX + bY) = \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega))\mathbb{P}(\omega) = a \sum_{\omega \in \Omega} X(\omega) \times \mathbb{P}(\omega) + b \sum_{\omega \in \Omega} Y(\omega) \times \mathbb{P}(\omega) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. □

Remarque 5. Cette linéarité est en fait « évidente » d'un point de vue intuitif. Prenons un exemple classique et toujours réjouissant, celui des notes qu'on attribue à nos chers élèves. Lors d'une interro de maths particulièrement peu révisée, j'obtiens à la fin de ma correction une moyenne brute égale à 5 sur 20. Je décide dans un élan de générosité d'ajouter trois points à tout le monde, quel effet cela aura-t-il sur la moyenne ? Ça va bien sûr l'augmenter de trois points. C'est exactement ce que dit la propriété $\mathbb{E}(X + b) = \mathbb{E}(X) + b$. Si au lieu d'augmenter tout le monde de trois points, je décide carrément de doubler toutes les notes, la moyenne sera multipliée par deux. C'est ce qu'affirme la formule $\mathbb{E}(aX) = a\mathbb{E}(X)$. Enfin, si je décide de faire une deuxième interro pour rattraper la première, deuxième interro où on atteint une moyenne de 7, et que j'additionne les deux notes pour en former une nouvelle, la moyenne de la classe sera de 12. C'est ce qui découle de la dernière formule $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Exemple : Cette propriété très simple est mine de rien extrêmement utile (c'est même la propriété fondamentale à maîtriser sur l'espérance). On lance par exemple successivement 100 dés équilibrés à six faces (tant qu'à faire, maintenant qu'on a une méthode efficace, on peut se le permettre). On note X la somme des résultats obtenus. Calculer l'espérance directement demande un certain courage (la loi de X est une horreur absolue), mais on peut en fait s'en sortir beaucoup plus simplement ! Notons X_i la variable aléatoire donnant le résultat du lancer du dé numéro i. On peut constater que $X = X_1 + X_2 + \cdots + X_{100}$. Or, toutes les variables X_i ont la même espérance, celle de la variable donnant le résultat d'un lancer de dé à six faces, qui vaut $\frac{7}{2}$ (cette variable prend chacune des valeurs 1, 2, 3, 4, 5 et 6 avec probabilité $\frac{1}{6}$, je vous laisse détailler le calcul si nécessaire). On a donc $\mathbb{E}(X) = \mathbb{E}(X_1 + X_2 + \cdots + X_{100}) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_{100}) = \frac{7}{2} + \cdots + \frac{7}{2} = 350$ (résultat intuitivement évident, soit dit en passant).

Proposition 5. Une variable aléatoire positive (c'est-à-dire que $X(\Omega) \subset \mathbb{R}^+$) a nécessairement une espérance positive.

Si X, Y sont deux variables aléatoires telles que $X \leqslant Y$ (c'est-à-dire que $\forall \omega \in \Omega$, $X(\omega) \leqslant Y(\omega)$), alors $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$.

Démonstration. C'est une fois de plus évident. Tous les termes intervenant dans le calcul de l'espérance de X étant positifs, la somme sera nécessairement positive. Pour la deuxième propriété, on peut utiliser une ruse classique : si $X \leqslant Y$, la variable aléatoire $Y - X$ est positive, donc $\mathbb{E}(Y - X) \geqslant 0$. Or, par linéarité de l'espérance, $\mathbb{E}(Y - X) = \mathbb{E}(Y) - \mathbb{E}(X)$, ce qui nous donne l'inégalité voulue. On remarquera en passant que les propriétés fondamentales de l'espérance (linéarité, positivité) rappellent fortement celles de l'intégrale. Ce n'est bien entendu pas du tout un hasard. □

Définition 7. Une variable aléatoire X est **centrée** si $\mathbb{E}(X) = 0$.

Proposition 6. Si X est une variable aléatoire d'espérance m, la variable aléatoire $X-m$ est centrée. On l'appelle **variable aléatoire centrée associée à** X.

Démonstration. Par linéarité, $\mathbb{E}(X-m) = \mathbb{E}(X) - \mathbb{E}(m) = m - m = 0$. □

Théorème 2. (théorème du transfert) Soit X une variable aléatoire et $f : \mathbb{R} \to \mathbb{R}$ une fonction, alors on a $\mathbb{E}(f(X)) = \sum_{k \in X(\Omega)} f(k)\mathbb{P}(X=k)$.

Démonstration. Encore une fois, la démonstration est immédiate en utilisant la formulation de l'espérance utilisée pour la démonstration de la linéarité. □

Remarque 6. Ce théorème signifie qu'on peut calculer très facilement des espérances de variables aléatoires obtenues en appliquant une fonction à la variable X. Ainsi, par exemple, $\mathbb{E}(X^2) = \sum_{k \in X(\Omega)} k^2\mathbb{P}(X=k)$ (autrement dit, on élève les valeurs au carré et on garde les probabilités intactes par rapport au calcul de l'espérance de X). Un principe simple à retenir quand on fait ce genre de calcul : on ne modifie jamais les probabilités, mais uniquement les valeurs, dans le calcul d'espérance.

2.2 Variance et moments d'ordre supérieur.

Définition 8. Soit X une variable aléatoire et r un entier strictement positif, le **moment d'ordre** r de X, noté $m_r(X)$, est l'espérance de la variable aléatoire X^r. Autrement dit (en utilisant le théorème du transfert) $m_r(X) = \sum_{k \in X(\Omega)} k^r\mathbb{P}(X=k)$.

Remarque 7. Le moment d'ordre 1 de X n'est autre que l'espérance de X.

Définition 9. La **variance** $\mathbb{V}(X)$ d'une variable aléatoire X est le moment d'ordre 2 de la variable aléatoire centrée associée à X. Autrement dit, $\mathbb{V}(X) = \mathbb{E}((X-\mathbb{E}(X))^2)$. L'écart-type σ de la variable aléatoire X est défini par $\sigma(X) = \sqrt{\mathbb{V}(X)}$.

Que représente cette variance ? Il s'agit, techniquement, d'une moyenne de carrés d'écarts à la moyenne. Pourquoi prendre le carré ? Tout simplement car la moyenne des écarts à la moyenne est nulle. Pour réellement mesurer ces écarts, il faut « les rendre positifs », ce qui se fait bien en les élevant au carré. On pourrait également penser à prendre leur valeur absolue, mais cela aurait moins de propriétés intéressantes pour le calcul. Par contre, pour « effacer » la mise au carré, on reprend ensuite la racine carrée du résultat obtenu pour définir l'écart-type. L'écart-type représente donc (comme son nom l'indique) un écart moyen entre les valeurs prises par X et la moyenne de X (plus il est grand, plus les valeurs prises par X sont étalées).

Proposition 7. La variance d'une variable aléatoire est toujours positive.

On a la formule $\mathbb{V}(aX+b) = a^2\mathbb{V}(X)$. En découle $\sigma(aX+b) = |a|\sigma(X)$.

Démonstration. La première propriété découle immédiatement de la définition du moment d'ordre 2, qui est une somme de termes positifs. Pour la deuxième, c'est du calcul un peu formel. Il faut calculer l'espérance de $(aX+b-\mathbb{E}(aX+b))^2$. Or, par linéarité de l'espérance, $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$ donc l'expression précédente vaut $(aX+b-a\mathbb{E}(X)-b)^2 = a^2(X-\mathbb{E}(X))^2$, dont l'espérance vaut $a^2\mathbb{E}((X-\mathbb{E}(X))^2) = a^2\mathbb{V}(X)$. □

Remarque 8. Une variable aléatoire a une variance (et un écart-type) nulle si et seulement si elle est constante.

Théorème 3. Formule de König-Huygens : $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Démonstration. C'est à nouveau un calcul très formel : $(X-\mathbb{E}(X))^2 = X^2 - 2\mathbb{E}(X)X + (\mathbb{E}(X))^2$, donc, par linéarité de l'espérance, $\mathbb{V}(X) = \mathbb{E}((X-\mathbb{E}(X))^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(\mathbb{E}(X))^2 = \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. □

Remarque 9. En pratique, c'est à peu près systématiquement via la formule de König-Huygens que nous effectuerons nos calculs de variance, car il est plus simple de calculer $\mathbb{E}(X^2)$ que de calculer directement $\mathbb{E}((X-\mathbb{E}(X))^2)$.

Définition 10. Une variable aléatoire est dite **réduite** si son écart-type (et donc sa variance) vaut 1.

Proposition 8. Si X est une variable aléatoire, la **variable aléatoire centrée réduite associée à** X est $X^* = \dfrac{X-\mathbb{E}(X)}{\sigma(X)}$ (qui est, vous vous en seriez doutés, centrée et réduite).

Démonstration. On a déjà vu plus haut que $X-\mathbb{E}(X)$ était centrée, la diviser par l'écart-type ne va pas changer cela. De plus, $\mathbb{V}\left(\dfrac{X-\mathbb{E}(X)}{\sigma(X)}\right) = \left(\dfrac{1}{\sigma(X)}\right)^2 \mathbb{V}(X) = 1$. □

Exemple 1 : Pour vous montrer qu'un calcul d'écart-type à la main est en général très fastidieux, prenons l'exemple classique du lancer de deux dés, où l'on note X la somme des deux chiffres obtenus. Un calcul pénible donne $\mathbb{E}(X^2) = \dfrac{2^2 \times 1 + 3^2 \times 2 + \cdots + 12^2 \times 1}{36} = \dfrac{1974}{36}$ donc (formule de König-Huygens) $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \dfrac{35}{6}$. L'écart-type vaut donc $\sigma(X) = \sqrt{\dfrac{35}{6}} \simeq 2.415$.

Exemple 2 : Dans le cas du lancer de six pièces, c'est moins affreux. On a $\mathbb{E}(X^2) = \dfrac{6+60+180+240+150+36}{64} = \dfrac{672}{64} = \dfrac{21}{2}$, donc $\mathbb{V}(X) = \dfrac{21}{2} - 9 = \dfrac{3}{2}$.

Exemple 3 : Dernier calcul pour nos histoires de jetons dans une urne, $\mathbb{E}(X^2) = \dfrac{6+4\times 3+9}{10} = \dfrac{27}{10}$, donc $\mathbb{V}(X) = \dfrac{27}{10} - \dfrac{9}{4} = \dfrac{9}{20}$. Sans surprise, on obtient ici un écart-type faible puisque les valeurs que peut prendre X sont limitées, et que la variable va en pratique prendre presque tout le temps la valeur 1 ou 2 (dont être à une « distance » $\dfrac{1}{2}$ de sa moyenne).

2.3 Inégalité de Bienaymé-Tchebychev.

Théorème 4. Inégalité de Markov.

Soit X une variable aléatoire positive et a un réel strictement positif, alors $\mathbb{P}(X \geqslant a) \leqslant \dfrac{\mathbb{E}(X)}{a}$.

Démonstration. C'est intuitivement assez évident (si ce n'était pas le cas, X prendrait des valeurs plus grandes que a avec une probabilité plus grande que $\dfrac{\mathbb{E}(X)}{a}$, ce qui contribuerait à l'espérance pour une quantité plus grande que $\mathbb{E}(X)$), mais prouvons-le rigoureusement. Pour cela, on définit une deuxième variable Y de la façon suivante : $Y(\omega) = a$ si $X(\Omega) \geqslant a$, et $Y(\omega) = 0$ sinon. Par construction, $\mathbb{E}(Y) = a \times \mathbb{P}(X \geqslant a)$. Or, par construction également, on a toujours $Y \leqslant X$. Les propriétés de positivité de l'espérance assurent alors que $\mathbb{E}(X) \geqslant \mathbb{E}(Y) = a \times \mathbb{P}(X \geqslant a)$, ce qui est exactement l'inégalité de Markov. □

Théorème 5. Inégalité de Bienaymé (Irénée-Jules) - Tchebychev (Pafnouti).

Soit X une variable aléatoire d'espérance m et d'écart-type σ, alors, pour tout réel strictement positif a, on a l'inégalité $\mathbb{P}(|X - m| \geqslant a) \leqslant \dfrac{\sigma^2}{a^2}$.

Démonstration. Appliquons l'inégalité de Markov à la variable $(X - m)^2$, qui est positive, et à a^2. On sait que $\mathbb{E}((X - m)^2) = \mathbb{V}(X) = \sigma^2$, donc $\mathbb{P}((X - m)^2 \geqslant a^2) \leqslant \dfrac{\sigma^2}{a^2}$. Or, la probabilité de gauche est la même que celle de $|X - m| \geqslant a$, ce qui prouve l'inégalité de Bienaymé-Tchebychev. □

Remarque 10. Cette inégalité permet de majorer la probabilité d'obtenir des valeurs éloignées de la moyenne pour une variable aléatoire (ce qui a notamment un lien avec la notion d'intervalle de confiance en statistiques), mais elle est hélas très imprécise, et ne donne pas de bons résultats en pratique sauf pour des valeurs réellement extrêmes de la variable aléatoire. En fait, son principal intérêt est de permettre de démontrer rapidement la loi faible des grandes nombres, mais cette dernière n'est pas à votre programme cette année !

3 Lois usuelles finies.

Certaines lois de probabilité interviennent suffisamment régulièrement lorsqu'on étudie des variables aléatoires dans des cas classiques (lancers de dés ou de pièces, tirages de boules dans des urnes, bref toutes les bêtises qu'on aime bien vous infliger dans les exercices de probas) pour qu'il soit intéressant de les étudier une bonne fois pour toutes (et accessoirement de leur donner un nom) et d'en retenir les caractéristiques (espérance et variance notamment). Nous en étudierons trois dans ce chapitre (dont l'une est un cas particulier d'une autre), et vous en verrez quelques autres l'an prochain (sur des univers infinis).

3.1 Loi uniforme.

Exemple fondamental : Dans une urne se trouvent n boules numérotées de 1 à n. On en tire une au hasard et on note X le numéro obtenu.

Définition 11. On dit que la variable aléatoire X suit une **loi uniforme sur** $\{1,\dots,n\}$, et on note $X \sim \mathcal{U}(\{1,\dots,n\})$, si $X(\Omega) = \{1,\dots,n\}$ et $\forall k \in \{1,\dots,n\}$, $\mathbb{P}(X=k) = \dfrac{1}{n}$.

Proposition 9. Si $X \sim \mathcal{U}(\{1,\dots,n\})$, on a $\mathbb{E}(X) = \dfrac{n+1}{2}$ et $\mathbb{V}(X) = \dfrac{n^2-1}{12}$.

Démonstration. Pour l'espérance, on a $\mathbb{E}(X) = \displaystyle\sum_{k=1}^{k=n} k \times \frac{1}{n} = \frac{1}{n}\sum_{k=1}^{k=n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$. Pour la variance, on va comme toujours utiliser la formule de König-Huygens. On a $\mathbb{E}(X^2) = \displaystyle\sum_{k=1}^{k=n} k^2 \times \frac{1}{n} = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$ donc $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \dfrac{(n+1)(2n+1)}{6} - \dfrac{(n+1)^2}{4} = \dfrac{2(n+1)(2n+1) - 3(n+1)^2}{12} = \dfrac{4n^2+6n+2-3n^2-6n-3}{12} = \dfrac{n^2-1}{12}$. □

Remarque 11. À partir d'une loi uniforme prenant ses valeurs entre 1 et n, on construit facilement une loi dont la probabilité est uniforme entre deux entiers m et p (il suffit d'ajouter une constante). La loi ainsi construite a une espérance égale à $\dfrac{m+p}{2}$ et une variance égale à $\dfrac{(m-p+1)^2-1}{12}$.

3.2 Loi de Bernoulli.

Exemple fondamental : On lance une pièce mal équilibrée pour laquelle la probabilité d'obtenir Pile vaut p et on note X la variable aléatoire valant 1 si on tombe sur Pile et 0 si on tombe sur face.

Définition 12. On dit que la variable aléatoire X suit une **loi de Bernoulli de paramètre** p (avec $p \in [0,1]$) si $X(\Omega) = \{0,1\}$, $\mathbb{P}(X=1) = p$ et $\mathbb{P}(X=0) = 1-p$. On le note $X \sim \mathcal{B}(1,p)$.

Remarque 12. Cette loi est aussi appelée loi indicatrice de paramètre p, puisqu'elle apparait essentiellement dans le cas où X est la variable aléatoire indicatrice d'un événement.

Proposition 10. Si $X \sim \mathcal{B}(1,p)$, alors $\mathbb{E}(X) = p$ et $\mathbb{V}(X) = p(1-p)$.

Démonstration. Pour l'espérance, on a déjà fait le calcul un peu plus haut. On a par ailleurs de la même façon $\mathbb{E}(X^2) = p$, donc $\mathbb{V}(X) = p - p^2 = p(1-p)$. □

Remarque 13. On utilise surtout en pratique des sommes de variables aléatoires suivant des lois de Bernoulli. On peut voir une variable aléatoire de Bernoulli comme un exemple d'expérience aléatoire où on a seulement deux issues possibles, le « succès » (représenté par la valeur 1 pour la variable aléatoire) et l'« échec » (la variable prend la valeur 0). Ces variables basées sur la dualité succès-échec sont à la base du principe du **schéma de Bernoulli** : si on répète plusieurs fois de suite une expérience de Bernoulli (la « même » expérience au sens où la probabilité de succès doit rester toujours la même, et où les différentes expériences doivent être indépendantes les unes des autres) et qu'on compte le nombre de succès obtenus lors de ces n essais, on construit une variable aléatoire qui va suivre une loi binômiale, que nous allons à présent décrire.

3.3 Loi binômiale

Exemple fondamental : Une urne contient des boules blanches et noires, avec une proportion p de boules blanches (et donc une proportion $1-p$ de boules noires). On tire n boules **avec remise** dans l'urne et on note X le nombre de boules blanches obtenues. Par rapport à la remarque concluant le paragraphe précédent, on répète bien ici n fois une expérience ayant deux issues possibles, le succès étant ici représenté par le tirage d'une boule blanche et l'échec par le tirage d'une boule noire. Il est essentiel que les tirages s'effectuent **avec remise** pour que la probabilité de tirer une boule blanche reste la même à chaque tirage, sinon on obtiendra un tout autre type de loi (les différents tirages n'étant alors plus indépendants).

Définition 13. On dit qu'une variable aléatoire X suit une **loi binomiale de paramètre** (n,p) (avec $n \in \mathbb{N}^*$ et $p \in [0,1]$) si $X(\Omega) = \{0,\dots,n\}$ et $\forall k \in \{0,\dots,n\}$, $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$. On le note $X \sim \mathcal{B}(n,p)$.

Remarque 14. Si $n=1$, la loi binomiale de paramètre $(1,p)$ n'est autre que la loi de Bernoulli de paramètre p, ce qui justifie l'emploi de la même notation.

Remarque 15. Comme on l'a déjà dit, une loi binômiale est obtenue lorsqu'on compte le nombre de réussites après avoir tenté n fois de suite (de façon indépendante) une expérience aléatoire ayant une probabilité p de réussir. C'est même ainsi qu'on doit repérer les lois binômiales dans un exercice. Attention tout de même, ce n'est pas parce qu'un schéma de Bernoulli est présent dans l'énoncé que toutes les variables aléatoires vont nécessairement suivre des lois binômiales, on peut très bien créer d'autres variables à partir de cette même situation. Par exemple, on lance dix fois de suite une pièce (équilibrée) à Pile ou Face. Si on note X le nombre de Faces obtenues, la variable X suivra une loi binômiale de paramètre $\left(10, \frac{1}{2}\right)$ (on répète dix fois une expérience ayant une probabilité $\frac{1}{2}$ de réussir et on compte simplement le nombre de succès). Mais si, à partir de la même expérience, on note Y le numéro du lancer où on a obtenu pour la première fois Pile (en admettant qu'on peut poser $Y=0$ si on n'obtient que des Faces), la variable Y ne suivra pas du tout une loi binômiale.

Proposition 11. Si $X \sim \mathcal{B}(n,p)$, alors $\mathbb{E}(X) = np$ et $\mathbb{V}(X) = np(1-p)$ (on note parfois $q = 1-p$, auquel cas on a $\mathbb{V}(X) = npq$).

Démonstration. On a $\mathbb{E}(X) = \sum_{k=1}^{k=n} k \binom{n}{k} p^k (1-p)^{n-k}$. On aimerait bien appliquer le binôme de Newton, mais il faut pour cela faire disparaitre le k, ce qui est par exemple possible grace à la formule sans nom $k\binom{n}{k} = n\binom{n-1}{k-1}$. On a donc $\mathbb{E}(X) = \sum_{k=1}^{k=n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} = n \sum_{j=0}^{j=n-1} \binom{j}{k-1} p^{j+1} (1-p)^{n-1-j} = np(p+1-p)^{n-1} = np$.
Pour la variance, on ne va pas calculer $\mathbb{E}(X^2)$ directement, mais passer par $\mathbb{E}(X(X-1))$, ce qui va permettre d'utiliser la formule $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$ (obtenue en appliquant deux fois de suite la formule sans nom). Un calcul extrêmement similaire au précédent donne alors $\mathbb{E}(X(X-1)) = n(n-1)p^2$, donc $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1)+X) - \mathbb{E}(X)^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p)$. $\square$

Remarque 16. La somme de n variables suivant chacune de façon indépendante une loi de Bernoulli de paramètre p sera toujours une variable binômiale de paramètre (n,p). C'est même exactement le principe du schéma de Bernoulli.

Remarque 17. Pour les curieux, que se passe-t-il si on reprend le dispositif de notre exemple fondamental (une urne contient des boules blanches en proportion p et noires en proportion $1 - p$) mais qu'on effectue n tirages successifs **sans** remise (en notant toujours X le nombre de boules blanches tirées après ces n tirages) ? On obtient en fait un autre type de loi classique (mais qui n'est pas à votre programme !), appelée loi hyper-géométrique (quel beau nom). On a aussi des formules pour l'espérance et la variance de telles lois, qui se démontrent en utilisant une belle formule de dénombrement (pas à votre programme non plus !) appelée formule de Vandermonde. Si jamais vous tombez sur une variable de ce type dans un exercice, vous referez bien sûr tous les calculs à la main.

4 Couples de variables aléatoires.

4.1 Lois d'un couple de variables aléatoires.

Le pluriel dans le titre n'est pas une erreur, la principale chose à bien assimiler ici est le vocabulaire concernant les différents types de lois qu'on peut associer à un couple de variables aléatoires. Comme dans la première partie de ce chapitre, trois exemples numérotés seront rapidement introduits pour servir d'illustration aux concepts définis.

Définition 14. Un couple de variables aléatoires (X, Y) est la donnée de deux variables aléatoires définies sur le même espace probabilisé Ω. Une façon plus technique de voir les choses est de dire qu'un couple est une application $(X, Y) : \Omega \to \mathbb{R}^2$.

Remarque 18. On peut bien entendu généraliser le principe et définir des n-uplets de variables aléatoires définies sur le même univers Ω pour tout entier $n \geqslant 2$.

Exemple 1 : On lance simultanément deux dés (ça faisait longtemps), et on note X le plus grand des deux chiffres obtenus et Y le plus petit (au sens large).

Exemple 2 : Une urne contient 3 boules rouges, 4 boules vertes et 2 boules bleues. On tire 3 boules simultanément dans l'urne et on note X le nombre de boules rouges obtenues et Y le nombre de boules vertes.

Exemple 3 : On effectue une suite de lancers avec une pièce déséquilibrée pour laquelle la probabilité d'obtenir Pile vaut $\dfrac{3}{4}$ (et celle d'obtenir face vaut donc $\dfrac{1}{4}$). On note X la longueur de la première chaîne obtenue, c'est-à-dire le nombre de tirages initiaux donnant le même résultat que le premier tirage. On note Y la longueur de la deuxième chaîne (nombre de tirages identiques à partir du lancer où a eu lieu le premier changement). Une chaîne peut très bien être de longueur 1. Ainsi, si les premiers tirages donnent $PPPPFFP$ (peu importe la suite), on aura $X = 4$ (on a commencé par Pile et on en a obtenu quatre de suite avant de changer pour tomber sur Face) et $Y = 2$ (la deuxième chaîne est la chaîne constituée des deux Face obtenus au cinquième et au sixième lancer).

Définition 15. La **loi conjointe** d'un couple de variables aléatoires (X, Y) est la donnée de toutes les probabilités $\mathbb{P}((X = i) \cap (Y = j))$, pour $i \in X(\Omega)$ et $j \in Y(\Omega)$.

Remarque 19. Cette loi est souvent présentée sous forme d'un tableau à double entrée, les valeurs prises par X apparaissant par exemple en ligne et celles prises par Y en colonne. C'est l'équivalent « à deux dimensions » des tableaux de loi qu'on faisait déjà pour les variables aléatoires classiques. On pourrait bien sûr imaginer étudier de la même façon des triplets de variables aléatoires (X, Y, Z), mais la représentation de la loi complète d'un tel triplet nécessiterait un tableau à trois dimensions, ce qui pose des problèmes évidents en pratique (même si on peut toujours représenter un tel tableau « tranche par tranche » sur une feuille de papier). Bien sûr, il arrivera aussi qu'on trouve une formule générale pour $\mathbb{P}((X = i) \cap (Y = j))$ qui évitera de faire le tableau de la loi (et ce sera même nécessaire dans notre troisième exemple où les variables constituant le couple ne sont pas finies, ce qui constitue d'ailleurs un léger débordement par rapport au programme de première année).

Remarque 20. Certaines des probabilités $\mathbb{P}((X=i)\cap(Y=j))$ peuvent être nulles, même si $\mathbb{P}(X=i)$ et $\mathbb{P}(Y=j)$ sont toutes les deux non nulles. Il arrivera donc régulièrement que certaines cases du tableau représentant la loi conjointe contiennent la valeur 0. En pratique, c'est d'ailleurs la chose qu'on essaie de repérer en premier (on place tous les 0 dans le tableau avant de calculer les probabilités correspondant aux cases restantes).

Exemple 1 : On a $X(\Omega)=Y(\Omega)=\{1,2,3,4,5,6\}$, et la loi conjointe se calcule ici sans difficulté : $\mathbb{P}((X=i)\cap(Y=j))=0$ si $i<j$ (le plus grand nombre ne peut pas être inférieur au plus petit), $\mathbb{P}((X=i)\cap(Y=j))=\dfrac{1}{36}$ si $i=j$ (le seul tirage favorable sur les 36 possibles est le tirage (i,i)), et $\mathbb{P}((X=i)\cap(Y=j))=\dfrac{1}{18}$ si $i>j$ (les deux tirages (i,j) et (j,i) sont possibles), ce qu'on peut résumer par le tableau suivant (X en colonne, Y en ligne) :

$Y\backslash X$	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$
2	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$
3	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$
4	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$
6	0	0	0	0	0	$\frac{1}{36}$

Remarque 21. La somme de toutes les probabilités présentes dans le tableau doit logiquement être égale à 1.

Exemple 2 : On a ici $X(\Omega)=\{0,1,2,3\}$ et $Y(\Omega)=\{0,1,2,3\}$ (si on avait noté Y le nombre de boules bleues et pas vertes, les deux variables ne prendraient pas les mêmes valeurs, ce qui ne pose aucun problème). On ne peut bien sûr pas avoir $X+Y>3$ puisqu'on ne tire que trois boules, ce qui permet déjà de mettre des 0 dans toutes les cases « en-dessous de la diagonale » dans le tableau. On aura également $\mathbb{P}((X=0)\cap(Y=0))=0$, puisqu'on ne peut pas tirer trois boules bleues (il n'y en a que deux dans l'urne !). Lorsque cela a un sens, on a en fait $\mathbb{P}((X=i)\cap(Y=j))=\dfrac{\binom{3}{i}\binom{4}{j}\binom{2}{3-i-j}}{\binom{9}{3}}$ (puisqu'il faut choisir les i boules rouges parmi les trois disponibles, les j boules vertes parmi les quatre disponibles, et il reste $3-i-j$ boules bleues à prendre parmi les deux que compte l'urne ; le dénominateur représente bien sûr le nombre total de tirages possibles de 3 boules dans une urne qui en contient 9, en l'occurrence égal à 84). Par exemple, $\mathbb{P}((X=1)\cap(Y=1))=\dfrac{\binom{3}{1}\times\binom{4}{1}\times\binom{2}{1}}{84}=\dfrac{24}{84}=\dfrac{2}{7}$ (même si on ne simplifiera pas les fractions dans le tableau de la loi). On obtient le tableau suivant :

$Y\backslash X$	0	1	2	3
0	0	$\frac{3}{84}$	$\frac{6}{84}$	$\frac{1}{84}$
1	$\frac{4}{84}$	$\frac{24}{84}$	$\frac{12}{84}$	0
2	$\frac{12}{84}$	$\frac{18}{84}$	0	0
3	$\frac{4}{84}$	0	0	0

Remarque 22. On note parfois l'évenement $(X=i)\cap(Y=j)$ sous la forme $(X,Y)=(i,j)$.

Définition 16. Si (X,Y) est un couple de variables aléatoires, les lois de X et de Y (en tant que variables aléatoires « classiques » étudiées indépendamment l'une de l'autre) sont appelées **lois marginales** du couple.

Proposition 12. Si (X,Y) est un couple de variables aléatoires discrètes, on peut obtenir les lois marginales à partir de la loi conjointe : $\forall i \in X(\Omega)$, $\mathbb{P}(X = i) = \sum_{j \in Y(\Omega)} \mathbb{P}((X = i) \cap (Y = j))$ (et symétriquement pour la loi de Y).

Démonstration. Cela découle immédiatement du fait que les événements $Y = j$ forment un système complet d'événements. C'est en fait exactement la formule des probabilités totales appliquée à ce système particulier. □

Exemple 1 : Pour connaitre les lois marginales à partir du tableau précédemment établi, c'est très simple, il suffit de faire les sommes des lignes du tableau (pour la loi de Y) ou des colonnes (pour celle X), ce qu'on matérialise en pratique en ajoutant une ligne et une colonne au tableau :

	1	2	3	4	5	6	$\mathbb{P}(Y=j)$
1	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{11}{36}$
2	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{9}{36}$
3	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{7}{36}$
4	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{5}{36}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{3}{36}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$
$\mathbb{P}(X=i)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	

Exemple 2 : De façon tout à fait similaire, on va retrouver la loi de X sur la dernière ligne du tableau et celle de Y dans la dernière colonne :

	0	1	2	3	$\mathbb{P}(Y=j)$
0	0	$\frac{3}{84}$	$\frac{6}{84}$	$\frac{1}{84}$	$\frac{10}{84}$
1	$\frac{4}{84}$	$\frac{24}{84}$	$\frac{12}{84}$	0	$\frac{40}{84}$
2	$\frac{12}{84}$	$\frac{18}{84}$	0	0	$\frac{30}{84}$
3	$\frac{4}{84}$	0	0	0	$\frac{4}{84}$
$\mathbb{P}(X=i)$	$\frac{20}{84}$	$\frac{45}{84}$	$\frac{18}{84}$	$\frac{1}{84}$	

Remarque 23. On vient de voir qu'on pouvait déduire les lois marginales de la loi conjointe. Le contraire n'est pas possible en général.

Exemple 3 : ici, $X(\Omega) = Y(\Omega) = \mathbb{N}^*$ (théoriquement d'ailleurs, si on tire toujours le même côté de la pièce, Y n'a même pas de valeur puisque la première chaîne ne s'arrête jamais ; en pratique ce cas a une probabilité nulle, tout comme le fait que Y prenne une valeur infinie, on peut ne pas les prendre en compte pour les calculs qui suivent). On n'est donc pas dans le cadre de ce cours (les variables sont infinies), mais cet exemple va permettre de présenter un principe essentiel de l'utilisation des couples de variables aléatoires : passer par la loi conjointe et la formule des probabilités totales pour retrouver la loi marginale de la deuxième variable, qu'on aurait beaucoup de mal à trouver directement. Vous retrouverez **très** souvent ce genre de calculs l'an prochain, il est essentiel de bien en comprendre le principe.

Commençons par calculer la loi (marginale) de la variable X (sans s'intéresser pour l'instant au couple) : on aura $X = k$ si la suite de lancers commence par k Pile suivis d'un Face ou par k Face suivis d'un Pile (peu importe ce qui se produira après), ces deux cas sont incompatibles et $\mathbb{P}(\underbrace{PP\dots PP}_{k \text{ fois}}F) = \left(\frac{3}{4}\right)^k \times \frac{1}{4}$. De même, $\mathbb{P}(\underbrace{FF\dots FF}_{k \text{ fois}}P) = \left(\frac{1}{4}\right)^k \times \frac{3}{4}$, donc $\mathbb{P}(X = k) = \left(\frac{3}{4}\right)^k \times$

$\dfrac{1}{4}+\left(\dfrac{1}{4}\right)^k\times\dfrac{3}{4}=\dfrac{3^k+3}{4^{k+1}}$. Profitons-en pour vérifier que la somme de toutes ces probabilités est égale à 1 (ce qui revient à dire que le cas où la première chaîne est infinie a une probabilité nulle). Il s'agit d'un calcul de somme géométrique (ou plutôt de série géométrique puisque la somme est infinie) que nous sommes tout à fait capables de faire. Il faut juste faire attention au fait que les sommes commencent à $k=1$: $\displaystyle\sum_{k=1}^{n}\mathbb{P}(X=k)=\frac{1}{4}\sum_{k=1}^{n}\left(\frac{3}{4}\right)^k+\frac{3}{4}\sum_{k=1}^{n}\left(\frac{1}{4}\right)^k=\frac{1}{4}\left(\frac{1-(\frac{3}{4})^{n+1}}{1-\frac{3}{4}}-1\right)+\frac{3}{4}\left(\frac{1-\frac{1}{4^{n+1}}}{1-\frac{1}{4}}-1\right)$ qui admet pour limite quand n tend vers $+\infty$ la valeur $\dfrac{3}{4}+\dfrac{1}{4}=1$. Pour les calculs qui suivront, on se permettra de manipuler des sommes géométriques infinies et en particulier d'appliquer la formule $\displaystyle\sum_{k=0}^{+\infty}q^k=\frac{1}{1-q}$ (avec bien sûr $q\in]-1,1[$) qui n'est qu'un simple passage à la limite de la formule classique de calcul des sommes géométriques.

Déterminons maintenant la loi conjointe du couple (X,Y). Le principe est exactement le même que pour le calcul de la loi de X : on aura $(X,Y)=(i,j)$ si on débute par i Pile, suivis de j Face et à nouveau d'un Pile ; ou bien par i Face, suivis de j Pile et d'un Face. Tout cela s'additionne pour donner une probabilité $\mathbb{P}((X=i)\cap(Y=j))=\left(\dfrac{3}{4}\right)^i\times\left(\dfrac{1}{4}\right)^j\times\dfrac{3}{4}+\left(\dfrac{1}{4}\right)^i\times\left(\dfrac{3}{4}\right)^j\times\dfrac{1}{4}=\dfrac{3^{i+1}+3^j}{4^{i+j+1}}$. Il est bien sûr hors de question ici de présenter tout ça sous forme de tableau, on est très contents d'avoir une formule générale pas trop horrible.

Il reste à calculer la loi marginale de Y. Je vous laisse vous convaincre que calculer directement la probabilité de l'événement $Y=j$ est essentiellement impossible (car elle dépend profondément de ce qui s'est passé avant le début de la deuxième chaîne). Par contre, on peut obtenir cette probabilité à l'aide de la formule des probabilités totales : les événements $X=i$ forment un système complet d'événements (il faut bien sûr tous les prendre, ce qui fait un système constitué d'une infinité d'événements, mais ça ne change rien à la formule ; notons d'ailleurs que le système est « presque complet » dans la mesure où on oublie le cas d'une première chaîne infinie, mais comme ce cas a une probabilité nulle ça n'a pas d'importance). On écrit donc $\displaystyle\mathbb{P}(Y=j)=\sum_{i=1}^{+\infty}\mathbb{P}(X=i)\times P_{X=i}(Y=j)=\sum_{i=1}^{+\infty}\mathbb{P}((X=i)\cap(Y=j))=\sum_{i=1}^{+\infty}\frac{3^{i+1}+3^j}{4^{i+j+1}}=\frac{1}{4^j}\sum_{i=1}^{+\infty}\left(\frac{3}{4}\right)^{i+1}+\left(\frac{3}{4}\right)^j\sum_{i=1}^{+\infty}\frac{1}{4^{i+1}}=\frac{1}{4^j}\left(\frac{3}{4}\right)^2\frac{1}{1-\frac{3}{4}}+\left(\frac{3}{4}\right)^j\frac{1}{4^2}\frac{1}{1-\frac{1}{4}}=\frac{3^2+3^{j-1}}{4^{j+1}}$. On constate que la loi de Y n'est pas du tout la même que celle de X (ce qui n'était pas totalement évident a priori). On peut là aussi vérifier que la somme des probabilités est égale à 1 : $\displaystyle\sum_{j=1}^{+\infty}\mathbb{P}(X=j)=\frac{9}{4}\sum_{j=1}^{+\infty}\frac{1}{4^j}+\frac{1}{12}\sum_{j=1}^{+\infty}\left(\frac{3}{4}\right)^j=\frac{9}{4}\times\left(\frac{1}{1-\frac{1}{4}}-1\right)+\frac{1}{12}\times\left(\frac{1}{1-\frac{3}{4}}-1\right)=\frac{3}{4}+\frac{1}{4}=1$.

Définition 17. La **loi conditionnelle de** X **sachant** $Y=j$ est la loi de la variable Z définie par $\forall i\in X(\Omega)$, $\mathbb{P}(Z=i)=\mathbb{P}_{Y=j}(X=i)$. On définit de même les lois conditionnelle de Y sachant $X=i$.

Exemple 1 : si on reprend l'exemple 1 et qu'on souhaite obtenir par exemple la loi conditionnelle de Y sachant $X=4$, on se place (dans le tableau bidimensionnel fait auparavant) « dans la colonne » correspondant à $X=4$. On constate alors que les seules valeurs possibles pour Y sont 1, 2, 3 et 4, avec des probabilités respectives $\dfrac{1}{18}$, $\dfrac{1}{18}$, $\dfrac{1}{18}$ et $\dfrac{1}{36}$. Mais attention, ces probabilités sont des probabilités **d'intersection** et non des probabilités **conditionnelles** (d'ailleurs, leur somme n'est pas égale à 1, ce qui n'est pas possible pour une loi de variable aléatoire). Pour obtenir notre loi conditionnelle, on doit donc diviser ces probabilités par $\mathbb{P}(X = 4)$, c'est-à-dire en fait par leur somme égale à $\dfrac{7}{36}$. La loi conditionnelle recherchée serait donc donnée par le tableau suivant :

k	1	2	3	4
$\mathbb{P}_{X=4}(Y = k)$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

Exemple 3 : Un dernier calcul intéressant sur cette expérience (quoique ne faisant pas vraiment intervenir le fait qu'on travaille sur un couple) est celui des espérances de X et de Y. Malheureusement, il fait intervenir une formule de calcul de série un peu plus compliquée que celle utilisée dans les calculs précédents, qu'on donnera donc sans justification pour l'instant (on la démontrera dans le chapitre que nous consacrerons bientôt aux séries : si $q \in]-1, 1[$, $\displaystyle\sum_{k=0}^{n} kq^{k-1}$ converge vers la limite $\dfrac{1}{(1-q)^2}$ (notons qu'on peut faire démarrer la somme à $k = 1$ sans en changer la valeur ici). Comme précédemment, on se permettra d'écrire plus directement ce résultat sous la forme $\displaystyle\sum_{k=1}^{+\infty} kq^{k-1} = \frac{1}{(1-q)^2}$. À l'aide de cette formule, on calcule aisément $\mathbb{E}(X) = \displaystyle\sum_{k=1}^{+\infty} k \times \frac{3^k + 3}{4^{k+1}} = \frac{3}{4^2}\sum_{k=1}^{+\infty} k\left(\frac{3}{4}\right)^{k-1} + \frac{3}{4^2}\sum_{k=1}^{+\infty} k\left(\frac{1}{4}\right)^{k-1} = \frac{3}{4^2}\left(\frac{1}{(1-\frac{3}{4})^2} + \frac{1}{(1-\frac{1}{4})^2}\right) = 3 + \frac{1}{3} = \frac{10}{3}$.

Pour Y, on a $\mathbb{E}(Y) = \displaystyle\sum_{k=1}^{+\infty} k \times \frac{3^2 + 3^{k-1}}{4^{k+1}} = \frac{3^2}{4^2}\sum_{k=1}^{+\infty} k\left(\frac{1}{4}\right)^{k-1} + \frac{1}{4^2}\sum_{k=1}^{+\infty} k\left(\frac{3}{4}\right)^{k-1} = \frac{3^2}{4^2} \times \frac{1}{(1-\frac{1}{4})^2} + \frac{1}{4^2} \times \frac{1}{(1-\frac{3}{4})^2} = 1 + 1 = 2$. Encore une fois, les courageux pourront faire le calcul avec une pièce équilibrée et constater que dans ce cas les deux variables ont pour espérance 2. Tout cela est en fait très normal. Quand la pièce est déséquilibrée, on a plus de chances d'avoir une première chaîne constituée de Pile que de Face. Mais le fait qu'on commence par un Pile augmente la probabilité que la chaîne soit longue puisqu'on a ensuite à chaque tirage plus de chances d'obtenir Pile que Face. Au contraire, la deuxième chaîne étant plus souvent constituée de Face, elle sera en moyenne plus courte que la première. Ce qui est (beaucoup !) moins évident, c'est le fait que cette deuxième chaîne ait une longueur moyenne « normale », c'est-à-dire identique à celles des chaînes obtenues avec une pièce équilibrée (pour laquelle il n'y a logiquement pas de différence entre première et deuxième chaîne). En fait, plus on déséquilibre la pièce (en faisant tendre la probabilité de tomber sur Pile vers 1, par exemple), plus la longueur moyenne de la première chaîne va devenir grande, mais celle de la deuxième chaîne sera toujours égale à 2. Bien sûr, la troisième chaîne aura à nouveau une longueur moyenne égale à celle de la première chaîne, la quatrième une longueur moyenne égale à 2 etc.

4.2 Indépendance de variables aléatoires.

Définition 18. Deux variables aléatoires sont dites **indépendantes** si tous les couples d'événements $X = i$, $Y = j$ sont indépendants. Autrement dit, $\forall i \in X(\Omega)$, $\forall j \in Y(\Omega)$, $\mathbb{P}((X = i) \cap (Y = j)) = \mathbb{P}(X = i)\mathbb{P}(Y = j)$. On le note $X \perp\!\!\!\perp Y$.

Exemple : Un exemple idiot pour illustrer. Si on tire deux dés simultanément et qu'on note X le résultat du premier dé et Y celui du deuxième dé, X et Y sont des variables aléatoires indépendantes. On a $\mathbb{P}((X = i) \cap (Y = j)) = \dfrac{1}{36}$ pour tout $1 \leqslant i \leqslant 6$ et $1 \leqslant j \leqslant 6$.

Exemple : Par contre, toujours dans le cas de l'inépuisable lancer de deux dés, si on prend pour X la somme des deux dés et pour Y leur produit, les deux variables ne sont pas indépendantes. On a par exemple $\mathbb{P}(X = 8) = \dfrac{5}{36}$, $\mathbb{P}(Y = 15) = \dfrac{1}{18}$, et $\mathbb{P}((X = 8) \cap (Y = 15)) = \dfrac{1}{18}$.

Remarque 24. Deux variables aléatoires sont indépendantes si et seulement si toutes les lois conditionnelles de X sachant $Y = j$ sont identiques à la loi de X. Dans le cas où X et Y sont des variables aléatoires indépendantes, on peut obtenir la loi conjointe du couple (X, Y) à partir des deux lois marginales. On peut bien entendu étendre facilement la notion d'indépendance à un n-uplet de variables aléatoires.

Exemples : Dans nos deux premiers exemples, on constate sans difficulté et sans surprise que les variables X et Y ne sont pas indépendantes (le fait qu'on ait des 0 dans le tableau de la loi conjointe suffit à imposer qu'il n'y ait pas indépendance).

Le troisième exemple est moins intuitif. Le calcul des lois conditionnelles, qui sont distinctes de la loi marginale de X, prouve qu'il n'y a pas non plus indépendance. Une autre façon de voir les choses est de trouver un contre-exemple à l'indépendance des évènements $X = i$ et $Y = j$. On peut ici calculer simplement $\mathbb{P}(X = 1) = \dfrac{3+3}{4^2} = \dfrac{3}{8}$, $\mathbb{P}(Y = 1) = \dfrac{3^2+1}{4^2} = \dfrac{5}{8}$ et $\mathbb{P}((X = 1) \cap (Y = 1)) = \dfrac{3^2+3}{4^3} = \dfrac{3}{16} \neq \dfrac{3}{8} \times \dfrac{5}{8}$. Les courageux vérifieront qu'en effectuant la même expérience avec une pièce équilibrée, on obtiendrait deux variables aléatoires indépendantes.

Proposition 13. Lemme des coalitions.

Si $(X_1, X_2, \dots, X_n)$ sont des variables aléatoires indépendantes définies sur le même univers Ω, et $f : \mathbb{R}^k \to \mathbb{R}$ et $g : \mathbb{R}^{n-k} \to \mathbb{R}$ deux fonctions, les variables $f(X_1, X_2, \dots, X_k)$ et $g(X_{k+1}, \dots, X_n)$ sont également indépendantes.

Démonstration. Le nom de ce théorème est du au fait qu'on appelle souvent coalitions des variables de la forme $(X_1, \dots, X_k)$ qui ne sont pas à valeurs réelles mais à valeurs dans $\mathbb{R}^k$. Le résultat est en fait complètement évident si on comprend ce que signifie la propriété : les variables $Y = (X_1, \dots, X_k)$ et $Z = (X_{k+1}, \dots, X_n)$ sont indépendantes d'après l'hypothèse faite sur la liste complète des variables X_i, et l'application des fonctions f et g n'y change rien. □

Proposition 14. Si X et Y sont deux variables aléatoires indépendantes, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Démonstration. Notons $f : (x, y) \mapsto xy$, alors $\mathbb{E}(XY) = \displaystyle\sum_{(k,l) \in X(\Omega) \times Y(\Omega)} f(x, y)\mathbb{P}((X = k) \cap (Y = l)) = \displaystyle\sum_{(k,l) \in X(\Omega) \times Y(\Omega)} x\mathbb{P}(X = k) \times y\mathbb{P}(Y = l) = \mathbb{E}(X)\mathbb{E}(Y)$ (on applique seulement le théorème du transfert et l'indépendance des variables). □

Proposition 15. Soient X et Y deux variables aléatoires indépendantes suivant des lois binomiales $\mathcal{B}(m, p)$ et $\mathcal{B}(n, p)$ alors $X + Y$ suit une loi binomiale de paramètre $\mathcal{B}(m + n, p)$.

Si une famille de variables indépendantes X_i vérifie $X_i \sim \mathcal{B}(1, p)$, alors $\displaystyle\sum_{i=1}^{n} X_i \sim \mathcal{B}(n, p)$.

Démonstration. C'est intuitivement évident (la deuxième propriété est une simple description du principe du schéma de Bernoulli, et la premièere dit en gros que si on effectue m fois de suite une expérience suivant un schéma de Bernoulli, puis qu'on recommence n fois de suite, et qu'on compte à la fin le nombre total de succès, on obtiendra la même chose qu'en regroupant les $m+n$ essais), mais pas si facile à démontrer rigoureusement (si on tente un calcul brutal, il nécessite la formule de Vandermonde, formule de dénombrement qui n'est pas à votre programme). On ne le fera donc pas. □

4.3 Covariance d'un couple de variables aléatoires.

Définition 19. La **covariance** d'un couple de variables aléatoires (X, Y) est le nombre réel $\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$.

Proposition 16. Propriétés de la covariance.

- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ (formule de König-Huygens)
- $\text{Cov}(Y, X) = \text{Cov}(X, Y)$ (symétrie de la covariance)
- $\text{Cov}(aX + Z, Y) = a\,\text{Cov}(X, Y) + \text{Cov}(Z, Y)$ (bilinéarité de la covariance)
- $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\text{Cov}(X, Y)$
- si X et Y sont indépendantes, $\text{Cov}(X, Y) = 0$ et $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$

Démonstration. Il s'agit de calculs formels assez simples utilisant bien sûr la linéarité de l'espérance :

- en développant, $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ (même calcul que pour König-Huygens avec une seule variable)
- la symétrie est complètement triviale
- là aussi c'est très facile, par exemple en appliquant König-Huygens : $\text{Cov}(aX + Z, Y) = \mathbb{E}(aXY + ZY) - \mathbb{E}(aX + Z)\mathbb{E}(Y) = a\mathbb{E}(XY) + \mathbb{E}(ZY) - a\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Z) = a\,\text{Cov}(X, Y) + \text{Cov}(X, Z)$
- par définition, $\mathbb{V}(X) = \text{Cov}(X, X)$, donc à l'aide de la bilinéarité (par symétrie, on peut développer à la fois par rapport au membre de gauche et au membre de droite) $\mathbb{V}(X + Y) = \text{Cov}(X + Y, X + Y) = \text{Cov}(X, X) + 2\,\text{Cov}(X, Y) + \text{Cov}(Y, Y) = \mathbb{V}(X) + 2\,\text{Cov}(X, Y) + \mathbb{V}(Y)$
- on a vu plus haut dans ce cours que, si X et Y sont indépendantes, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Il suffit d'appliquer ce résultat dans la formule de König-Huygens pour obtenir $\text{Cov}(X, Y) = 0$, et donc la linéarité de la variance en reprenant la formule démontrée juste avant.

□

Remarque 25. On peut généraliser la dernière propriété (sur la linéarité de la variance de variables indépendantes) à un ensemble de n variables indépendantes. On peut alors calculer facilement la variance d'une loi binomiale en la décomposant en somme de n variables de Bernoulli indépendantes, qui ont chacune pour variance $p(1 - p)$. La variance de la binomiale vaut donc $np(1 - p)$.

Définition 20. Le **coefficient de corrélation** de deux variables X et Y est le réel $\dfrac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$, qui appartient toujours à l'intervalle $[-1, 1]$. Les variables sont **positivement corrélées** si ce coefficient est positif, **négativement corrélées** s'il est négatif. Elles sont **non corrélées** (mais pas forcément indépendantes) s'il est nul.

Remarque 26. Les propriétés formelles de la covariance démontrées juste avant cette définition sont en fait formellement celles d'un produit scalaire (on reparlera notamment de symétrie et de bilinéarité quand on définira la notion de produit scalaire dans les espaces vectoriels). Si on poursuit cette analogie géométrique, la variance d'une variable aléatoire X correspond alors au carré de sa norme (la norme elle-même étant l'écart-type). Toutes les propriétés classiques des normes vectorielles sont en fait vérifiées par la variance, notamment l'inégalité de Cauchy-Schwartz : $|\operatorname{Cov}(X,Y)| \leqslant \sigma(X)\sigma(Y)$ (l'équivalent de la propriété classique $|u.v| \leqslant \|u\| \times \|v\|$) qui explique pourquoi le coefficient de corrélation linéaire est toujours compris entre -1 et 1. Ce coefficient mesure en fait le lien existant entre deux variables aléatoires. Plus on s'éloigne de 0, « moins les variables sont indépendantes ». Plus précisément, si le coefficient est proche de 1, cela signifie que Y aura une forte tendance à avoir le même comportement que X (plus X prend de grandes valeurs, plus Y fera de même) : les variables sont « pratiquement colinéaires de même sens » en reprenant l'analogie géométrique. Au contraire, si le coefficient est négatif et proche de -1, les variables auront un comportement opposé (plus X est grand, plus Y aura tendance à diminuer) : les variables sont « pratiquement colinéaires de sens opposé ».

Exemple : On lance deux dés équilibrés à quatre faces (pour simplifier un peu les calculs) et on note X la somme des deux résultats obtenus, et Y le produit des deux résultats obtenus. La variable X a pour univers-image $\{2,3,4,5,6,7,8\}$, et sa loi est donnée par le tableau suivant (on ne détaille pas les calculs très similaires à ceux effectués pour des deux à six faces en début de chapitre) :

k	2	3	4	5	6	7	8
$\mathbb{P}(X=k)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Par symétrie, on aura $\mathbb{E}(X)=5$, et on calcule sans difficulté $\mathbb{E}(X^2)$ $= \dfrac{4+18+48+100+108+98+64}{16} = \dfrac{440}{16} = \dfrac{55}{2}$, puis $\mathbb{V}(X) = \dfrac{55}{2} - 25 = \dfrac{5}{2}$.

La variable Y peut prendre toutes les valeurs obtenues en faisant le produit de deux entiers inférieurs ou égaux à 4, soit $Y(\Omega) = \{1,2,3,4,6,8,9,12,16\}$. Les probabilités correspondantes se calculent simplement en dénombrant tous les cas (il n'y en a que seize). On peut par exemple obtenir $Y=8$ de deux façons : 2×4 ou 4×2 :

k	1	2	3	4	6	8	9	12	16
$\mathbb{P}(Y=k)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

On calcule alors $\mathbb{E}(Y) = \dfrac{1+4+6+12+12+16+9+24+16}{16} = \dfrac{100}{16} = \dfrac{25}{4}$ (on pouvait aussi utiliser le résultat sur le produit des espérances de variables indépendantes pour obtenir cette valeur plus rapidement), puis $\mathbb{E}(Y^2) = \dfrac{1+8+18+48+72+128+81+288+256}{16} = \dfrac{900}{16} = \dfrac{225}{4}$. Enfin, on obtient $\mathbb{V}(X) = \dfrac{900}{16} - \dfrac{625}{16} = \dfrac{275}{16}$.

Si on veut effectuer le calcul de covariance, il ne reste plus qu'à réussir à calculer $\mathbb{E}(XY)$. Pas vraiment d'autre chois hélas que de faire un calcul très moche : la variable XY (produit de la somme et du produit des résultats obtenus lors des deux lancers de dés) prend les valeurs 2, 6, 12, 16, 20, 30, 48, 54, 84, et 128, avec la loi suivante :

k	2	6	12	16	20	30	48	54	84	128
$\mathbb{P}(X=k)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Un calcul sans intérêt donne alors $\mathbb{E}(XY) = \dfrac{2+12+24+16+40+60+96+54+168+128}{16} = \dfrac{600}{16} = \dfrac{75}{2}$. On en déduit donc (en appliquant comme toujours la formule de König-Huygens) que $\operatorname{Cov}(X,Y) = \dfrac{75}{2} - 5\times\dfrac{25}{4} = \dfrac{25}{4}$. Ou encore que le coefficient de corrélation entre les variables X et Y est égal à $\dfrac{25}{4}\times\sqrt{\dfrac{2}{5}\times\dfrac{16}{275}} = \dfrac{25}{4}\times\dfrac{4\sqrt{2}}{5\sqrt{55}} = \sqrt{\dfrac{10}{11}}$. Sans surprise, ce coefficient est très proche de 1.