29

Variables aléatoires sur un univers fini

Nous poursuivons dans ce chapitre l'étude des probabilités amorcée au chapitre 26, toujours en nous limitant aux univers finis.
Si nous avons déjà défini ce que sont un événement et sa probabilité, nous allons ici nous intéresser à des fonctions qui, à une issue de l'expérience considérée associent un nombre.
Par exemple :

1. si on lance deux dés, la somme des deux dés.
2. si on tire n boules numérotées dans une urne, le plus grand numéro obtenu.
3. si on lance n fois une pièce, le nombre de lancers nécessaires avant d'obtenir le premier *pile*[1].

Et en particulier, on souhaiterait donner un sens à la notion de valeur moyenne : «quand je lance deux dés, en moyenne leur somme vaut 7».
Vous avez déjà rencontré (brièvement) ces objets, les variables aléatoires, ainsi que les notions d'espérance et de variance, mais il s'agit ici de tout redéfinir proprement[2].

[1] Mais que fait-on alors si on obtient que des *faces* ?

[2] Finalement, c'est ce à quoi nous avons passé l'année...

Dans tout le chapitre, sans plus de précision, $(\Omega, \mathbf{P})$ désigne un espace probabilisé fini.

29.1 Variables aléatoires

29.1.1 Définition

Définition 29.1 – Soit Ω un ensemble fini, et soit E un ensemble. On appelle **variable aléatoire sur Ω à valeurs dans E** toute application $X : \Omega \to E$.
Dans le cas où $E = \mathbf{R}$, on parle de **variable aléatoire réelle.**

Difficile
Il n'est pas toujours évident de comprendre en quoi une fonction (quelque chose qui est on ne peut plus déterministe) nous aide à étudier les phénomènes aléatoires. C'est effectivement surprenant au premier abord, mais un passage obligé...

Pour l'instant, la notion de variable aléatoire n'a évidemment pas l'air d'un grand intérêt, puisque toute application sur Ω est une variable aléatoire, mais un peu de patience. Notons qu'on n'a même pas besoin de mettre une probabilité sur $\mathbf{P}$ pour parler d'une variable aléatoire. Mais là aussi patience, les variables aléatoires prendront tout leur intérêt une fois qu'on les couplera à des probabilités.

Profitons-en pour faire la différence entre les événements et les variables aléatoires. Un événement est un ensemble d'événements élémentaires, c'est-à-dire d'issues de l'expérience.
Par exemple[3] lorsqu'on lance deux dés, et qu'on modélise ceci par $\Omega = [\![1,6]\!]^2$, l'événement A : «la somme des deux dés vaut 3» est $\{(1,2),(2,1)\}$.
Une variable aléatoire associe un nombre[4] à **chaque** issue de l'expérience.
Par exemple, la variable aléatoire «somme des deux dés» est définie quel que soit l'issue de l'expérience. Cette variable aléatoire n'est rien d'autre que $X : \left| \begin{array}{ccc} [\![1,6]\!]^2 & \longrightarrow & \mathbf{R} \\ (i,j) & \longmapsto & i+j \end{array} \right.$.
Il y a bien un lien avec des événements, que nous allons expliciter dans un instant, par exemple l'événement A ci-dessus n'est rien d'autre que l'événement «la variable X vaut 3». Mais il est important de bien comprendre la différence de nature entre une variable aléatoire et un événement. Notamment pour savoir ce qu'on peut écrire et ce qu'on ne peut pas écrire. Par exemple, un événement est réalisé ou ne l'est pas, ce qui n'a pas de sens pour une variable aléatoire. Et on ne parle donc pas du contraire d'une variable aléatoire.

[3] Et c'est de loin mon *toy model* préféré, à utiliser sans modération lorsque vous avez un doute sur la signification des notions de base en probas.

[4] Au moins dans le cas d'une variable aléatoire réelle.

Sur le même espace probabilisé existent d'autres variables aléatoires, par exemple «le plus grand numéro des deux dés», qui est $Y : \left| \begin{array}{ccc} [\![1,6]\!]^2 & \longrightarrow & \mathbf{R} \\ (i,j) & \longmapsto & \max(i,j) \end{array} \right.$, ou encore le numéro du second dé qui est $Z : \left| \begin{array}{ccc} [\![1,6]\!]^2 & \longrightarrow & \mathbf{R} \\ (i,j) & \longmapsto & j \end{array} \right.$.

Pour reprendre un exemple de l'introduction, lorsqu'on lance n fois une pièce (que l'on modélise par $\{P,F\}^n$), la variable aléatoire «le numéro du premier *pile*» n'est pas bien définie car on ne lui attribue pas de valeur dans le cas où on n'obtient que des *faces*.

En revanche, $X : \left| \begin{array}{ccc} \Omega & \longrightarrow & \mathbf{N} \cup \{+\infty\} \\ (a_1,\dots,a_n) & \longmapsto & \begin{cases} +\infty & \text{si } (a_1,\dots,a_n) = (F,F,\dots,F) \\ \min\{i \in [\![1,n]\!] \mid a_i = P\} & \text{sinon} \end{cases} \end{array} \right.$

est bien une variable aléatoire.

Donnons un exemple de variable aléatoire qui n'est pas une variable aléatoire réelle, mais dans la suite nous ne parlerons quasiment que de variables à valeurs réelles.

Si on tire sans remise deux boules dans une urne qui contient 3 boules rouges, 2 noires et une blanche, on peut considérer la variable aléatoire à valeurs dans $\{R,N,B\}$ qui donne la couleur de la seconde boule.

À l'aide d'une variable aléatoire, il est possible de former de nombreux événements, et c'est là le premier intérêt des variables aléatoires : nous aider à nommer facilement des événements.

Définition 29.2 – Soit $X : \Omega \to E$ une variable aléatoire. Pour $A \subset E$, on note $[X \in A]$ l'événement[5] défini par

$$[X \in A] = X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}.$$

Lorsque $A = \{a\}$, on note $[X = a]$ au lieu de $[X \in \{a\}]$.
Et lorsque X est une variable aléatoire réelle, on note $[X \leqslant a] = X^{-1}(]-\infty, a])$, $[X < a] = X^{-1}(]-\infty, a[)$, et de même, on utilise les notations $[X \geqslant a]$, $[X > a]$, $[a \leqslant X \leqslant b] = X^{-1}([a,b])$, etc.

[5] C'est une partie de Ω.

Exemple 29.3

Toujours en lançant deux dés, notons X le résultat du premier dé, Y le résultat du second, et $Z = X + Y$ la somme des deux.
Alors l'événement «le premier dé vaut 1» est $[X \leqslant 1]$, l'événement «le second dé est pair» est $[Y = 2] \cup [Y = 4] \cup [Y = 6]$.
Et l'événement $[Z > 10]$ est $\{(5,6),(6,5),(6,6)\}$.

Entendons-nous bien : un événement peut avoir une probabilité, pas une variable aléatoire !
Donc si X est une variable aléatoire, $\mathbf{P}([X = 2])$, $\mathbf{P}([X \geqslant 3])$ et $\mathbf{P}([X \geqslant 3] \cup [X < -2])$ ont un sens. En revanche, $\mathbf{P}(X)$ **ne veut rien dire.**
Pas plus que $\mathbf{P}([X \geqslant 3]) \cup \mathbf{P}([X < -2])$ ou $\mathbf{P}([X \geqslant 3] + [X < -2])$, mais c'est une autre histoire[6].

Rédaction
En général on oublie les crochets quand on écrit des probas : $\mathbf{P}(X \leqslant a)$ et pas $\mathbf{P}([X \leqslant a])$.

[6] On peut considérer l'union d'événements, ou la somme de probas, pas le contraire !

Définition 29.4 – Si $X : \Omega \to E$ est une variable aléatoire, alors son image $X(\Omega)$ (qu'il faut interpréter comme l'ensemble des valeurs prises par X) est appelé **support de X.**
Puisque Ω est fini, il s'agit toujours d'un ensemble fini.

Proposition 29.5 : *Soit X une variable aléatoire sur Ω. Alors $\{[X=x], x\in X(\Omega)\}$ est un système complet d'événements, appelé système complet d'événements associé à X.*

Démonstration. Il suffit de se rappeler que $[X=x]=\{\omega\in\Omega \mid X(\omega)=x\}$.
Il est clair que les $[X=x]$ sont deux à deux incompatibles[7].
Et puisque $X(\Omega)$ est l'image de X, on a bien[8]

$$\bigcup_{x\in X(\Omega)} [X=x]=\Omega.$$

□

[7] Disjoints.
[8] Tout élément de Ω a son image dans $X(\Omega)$.

L'idée est tout simplement qu'on distingue plusieurs cas suivant la valeur que prend X, et comme elle prend toujours une valeur, on a bien ainsi distingué tous les cas.
Un cas très particulier de variable aléatoire est celle de l'indicatrice d'un événement A, qui rappelons-le, est définie par $\mathbb{1}_A : \left|\begin{array}{ccc}\Omega & \longrightarrow & \{0,1\} \\ \omega & \longmapsto & \begin{cases}1 & \text{si } \omega\in A \\ 0 & \text{si } \omega\notin A\end{cases}\end{array}\right.$.
Autrement dit, c'est la variable aléatoire qui vaut 1 si A est réalisé et 0 sinon.
Et on a donc $A=[\mathbb{1}_A=1]$.

Exemple 29.6

Toujours en lançant deux dés, on note A l'événement «les deux dés ont même parité».
Alors $\mathbb{1}_A$ est la variable aléatoire qui vaut 1 si et seulement si les deux dés ont même parité.
Donc $[\mathbb{1}_A=1]=\{(1,3),(3,1),(1,5),(2,6),\dots\}$.

29.1.2 Loi d'une variable aléatoire

Venons-en enfin aux probabilités : lorsque Ω est muni d'une probabilité $\mathbf{P}$, on peut parler de probabilité d'un événement, et en particulier de tous les événements qui sont associés à une variable aléatoire.
Par exemple, lorsqu'on joue au tiercé, en notant X le numéro du cheval gagnant, pour parier intelligemment il faut connaître les $\mathbf{P}([X=1]), \mathbf{P}([X=2])$, etc. La donnée de toutes ces probabilités est ce qu'on appelle la loi de X.

Cote/probas
La probabilité qu'un cheval gagne est grosso modo (c'est-à- dire modulo la commission du bookmaker) l'inverse de sa cote.

Définition 29.7 – Soit X une variable aléatoire sur $(\Omega,\mathbf{P})$. La **loi de** X est alors l'application $\mathbf{P}_X : \left|\begin{array}{ccc}\mathscr{P}(X(\Omega)) & \longrightarrow & [0,1] \\ A & \longmapsto & \mathbf{P}(X\in A)=\mathbf{P}(X^{-1}(A))\end{array}\right.$.

Proposition 29.8 : *La loi $\mathbf{P}_X$ de X est une probabilité sur $\mathscr{P}(X(\Omega))$.*

Démonstration. Il est clair que $\mathbf{P}_X$ est à valeurs dans $[0,1]$, et

$$\mathbf{P}_X(X(\Omega))=\mathbf{P}\left(X^{-1}(X(\Omega))\right)=\mathbf{P}(\Omega)=1.$$

Si A et B sont deux parties disjointes de $X(\Omega)$, alors

$$\mathbf{P}_X(A\cup B)=\mathbf{P}_X(A\cup B)=\mathbf{P}\left(X^{-1}(A\cup B)\right)=\mathbf{P}\left(X^{-1}(A)\cup X^{-1}(B)\right)$$

et ces deux événements étant incompatibles,

$$\mathbf{P}_X(A\cup B)=\mathbf{P}(X^{-1}(A))+\mathbf{P}(X^{-1}(B))=\mathbf{P}_X(A)+\mathbf{P}_X(B).$$

□

La loi d'une variable aléatoire dépend de la probabilité qu'on utilise sur Ω.
Par exemple, si on lance deux fois un dé, et qu'on note X le résultat du premier lancer, alors que le dé soit équilibré ou non, on peut modéliser l'expérience par $\Omega = [\![1,6]\!]^2$. Et alors X est toujours l'application $(i,j) \mapsto i$.
Par contre, si le dé est équilibré, on va utiliser la probabilité uniforme sur $[\![1,6]\!]$ et on vérifie alors que $\mathbf{P}_X$ est la probabilité uniforme sur $[\![1,6]\!]$.
En revanche, si le dé est pipé, on va utiliser un autre $\mathbf{P}$, ce qui donnera une autre loi pour X.

Proposition 29.9 : *Soit X une variable aléatoire sur Ω. Alors la loi de X est entièrement déterminée par les $\mathbf{P}_X(\{x\}) = \mathbf{P}(X = x)$, $x \in X(\Omega)$.*
Plus précisément, pour $A \in \mathscr{P}(X(\Omega))$, on a

$$\mathbf{P}_X(A) = \sum_{x \in A} \mathbf{P}(X = x).$$

Démonstration. Il suffit de noter que pour $A \in \mathscr{P}(X(\Omega))$, $A = \bigcup_{x \in A} \{x\}$ et donc

$$\mathbf{P}_X(A) = \mathbf{P}\left(X^{-1}(A)\right) = \mathbf{P}\left(X^{-1}\left(\bigcup_{x\in A}\{x\}\right)\right) = \mathbf{P}\left(\bigcup_{x\in A} X^{-1}(\{x\})\right) = \sum_{x\in A} \mathbf{P}_X(\{x\}) = \sum_{x \in A} \mathbf{P}(X = x).$$

□

Ainsi, lorsqu'on déterminera la loi d'une variable aléatoire X, on n'explicitera pas la probabilité $\mathbf{P}_X$ (et heureusement car il s'agit d'une application définie sur $\mathscr{P}(X(\Omega))$), mais on se contentera de déterminer :

- son support $X(\Omega)$
- les $\mathbf{P}(X = x)$, pour $x \in X(\Omega)$.

Autrement dit, on donnera les valeurs prises par la variable, et les probabilités qu'elle prenne chacune de ces valeurs.

Exemple 29.10

Une urne contient N boules numérotées de 1 à N.
On tire simultanément n ($n \leqslant N$) boules dans cette urne, et on note X le plus grand numéro ainsi obtenu.
Alors $X(\Omega) = [\![n, N]\!]$, et pour déterminer la loi de X, il nous faut donner en plus les $\mathbf{P}(X = k)$, $k \in [\![n, N]\!]$.
Il y a en tout $\binom{N}{n}$ tirages possibles, tous équiprobables. Parmi ceux-ci, ceux qui donnent une plus grande boule portant le numéro k sont au nombre de $\displaystyle\binom{k-1}{n-1}$.
En effet, un tel tirage est formé de la boule numéro k, et de $n-1$ boules de numéro strictement inférieur à k.
Donc $\mathbf{P}(X = k) = \dfrac{\binom{k-1}{n-1}}{\binom{N}{n}}$.

⚠ Deux variables définies sur un même espace et de même loi ne sont pas égales pour autant.
Prenons encore et toujours le lancer de deux dés équilibrés, et notons X le résultat du premier dé, et Y le résultat du second.
Alors X et Y ont même loi (que nous nommerons loi uniforme), mais ne sont pas égales, par exemple car $X(1,2) = 1$ alors que $Y(1,2) = 2$.
De même, si on lance n fois une pièce équilibrée, et que l'on note X (resp. Y) le nombre de *piles* (resp. de *faces*), alors X et Y ont même loi[9] sans pour autant être égales.

[9] Contentons-nous de l'affirmer, mais cela doit vous sembler intuitif.

Puisque $\{[X = x], x \in X(\Omega)\}$ est un système complet d'événements, on a évidemment $\displaystyle\sum_{x\in X(\Omega)} \mathbf{P}(X = x) = 1$.
Il existe une réciproque à ce résultat.

Déjà vu ? C'est le sacro-saint «la somme des probas vaut 1».

Proposition 29.11 : *Soit $n \geqslant 1$, soient $x_1, \dots, x_n$ des éléments deux à deux distincts d'un ensemble E, et soient $p_1, \dots, p_n$ des réels positifs, tels que $\displaystyle\sum_{i=1}^{n} p_i = 1$.*
Alors il existe un ensemble fini Ω, une probabilité $\mathbf{P}$ sur Ω et une variable aléatoire X définie sur Ω et à valeurs dans E telle que

$$\forall i \in [\![1, n]\!],\ \mathbf{P}(X = x_i) = p_i.$$

Remarque Les p_i étant positifs et de somme 1, ils sont automatiquement dans $[0, 1]$.

Démonstration. Prouvons un résultat plus fort, à savoir que sur tout ensemble fini Ω de cardinal supérieur ou égal à n, il existe une probabilité $\mathbf{P}$ et une variable aléatoire X définie sur Ω telle que $\forall i \in [\![1, n]\!]$, $\mathbf{P}(X = x_i) = p_i$.
Soit donc Ω un ensemble de cardinal $\geqslant n$ et soient $\omega_1, \dots, \omega_n$ n éléments distincts de Ω.
Alors par la proposition 28.12, il existe une (unique) probabilité $\mathbf{P}$ sur Ω telle que

$$\mathbf{P}(\{\omega\}) = \begin{cases} p_i & \text{si } \omega = \omega_i \\ 0 & \text{si } \omega \notin \{\omega_1, \dots, \omega_n\} \end{cases}$$

Et alors en posant $X : \left|\begin{array}{ccc} \Omega & \longrightarrow & E \\ \omega & \longmapsto & \begin{cases} x_i & \text{si } \omega = \omega_i \\ 0 & \text{sinon} \end{cases} \end{array}\right.$, alors X est une application définie sur Ω, donc une variable aléatoire sur Ω.
Et on a alors clairement $X(\Omega) = \{x_1, \dots, x_n\}$, et pour tout $i \in [\![1, n]\!]$,

$$\mathbf{P}(X = x_i) = \mathbf{P}(X^{-1}(x_i)) = \mathbf{P}(\{\omega_i\}) = p_i.$$

□

Cette proposition est en réalité très importante, sans qu'il soit vraiment nécessaire d'en maîtriser l'énoncé[10]. En effet, elle nous dispense, lorsqu'on manipule des variables aléatoires, de s'interroger sur l'espace probabilisé $(\Omega, \mathbf{P})$ que l'on utilise.
Par exemple, si on souhaite modéliser le lancer d'un dé pipé qui tombe sur 1 avec probabilité $1/4$, sur 2 avec probabilité $1/8$, sur 3 avec probabilité $1/8$ et sur 4, 5 ou 6 avec probabilité $1/6$ chacun, nul besoin de construire explicitement un univers Ω et une probabilité $\mathbf{P}$ sur laquelle une variable aléatoire dont la loi est celle que l'on veut, il suffit de vérifier que $\frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$, ce qui est bien le cas.

[10] Dont vous n'aurez jamais besoin.

C'est la raison pour laquelle beaucoup d'exercices commencent par «soit X une variable aléatoire suivant la loi *blabla*[11]». On sait qu'alors il existe un espace probabilisé sur lequel ceci a bien un sens, et cela nous suffit, nous n'avons pas besoin d'en savoir plus au sujet de cet espace.
De manière générale, la connaissance de l'univers Ω ne peut avoir d'intérêt que dans les cas où on peut faire du dénombrement, c'est-à-dire quand il y a équiprobabilité.

[11] Par exemple une loi uniforme ou une loi binomiale, voir ci-dessous.

Enfin, il faudra bien comprendre que les énoncés du type : «montrer que les réels $(p_i)_{1\leqslant i\leqslant n}$ définissent bien une loi de probabilité sur $\{x_1, \dots, x_n\}$» signifient juste «prouver que les p_i sont des réels positifs de somme 1».

Exemple 29.12

À quelle condition sur $\lambda \in \mathbf{R}$ existe-t-il une variable aléatoire X à valeurs dans $[\![1, n]\!]$ telle que $\forall i \in [\![1, n]\!]$, $\mathbf{P}(X = i) = \lambda i$?
Il faut bien évidemment que λ soit positif pour que les λi le soient, et il faut de plus

que
$$\sum_{i=1}^{n} \lambda i = 1 \Leftrightarrow \lambda = \frac{2}{n(n+1)}.$$
Puisque ce nombre est bien positif, on en déduit que $\dfrac{2}{n(n+1)}$ est la seule solution.

29.1.3 Lois usuelles

Si toute famille de réels positifs de somme 1 définit la loi d'une variable aléatoire, certaines, que l'on rencontre fréquemment portent un nom.

La plus simple est aussi la moins intéressante, mais on ne peut l'ignorer :

Définition 29.13 – Une variable aléatoire $X : \Omega \to E$ suit la **loi certaine égale à** $a \in E$ si $a \in X(\Omega)$ et $\mathbf{P}(X = a) = 1$.

On a évidemment envie de voir une loi certaine comme une constante[12], mais une petite distinction peut se produire : on autorise X à prendre d'autres valeurs que a, mais seulement avec probabilité nulle.
Ceci peut notamment se produire lorsqu'un espace probabilisé n'est pas complètement adapté à la description d'une expérience.
Par exemple, on peut modéliser le lancer d'un dé équilibré à 6 faces par $\Omega = [\![1, 7]\!]$, avec $\mathbf{P}(\{i\}) = \begin{cases} \frac{1}{6} & \text{si } 1 \leqslant i \leqslant 6 \\ 0 & \text{si } i = 7 \end{cases}$.
Et on peut alors prendre comme variable aléatoire symbolisant le résultat du premier dé la variable $X : \omega \mapsto \omega$.
Alors $X(7) = 7$, de sorte que $7 \in X(\Omega)$. Mais en revanche, $\mathbf{P}(X = 7) = 0$.
Et alors, la variable aléatoire indicatrice de l'événement «le nombre tiré est strictement inférieur à 7» est une variable certaine égale à 1, mais n'est pas constante égale à 1, car elle vaut 0 pour $\omega = 7$.

[12] Et d'ailleurs les fonctions constantes sur Ω sont des variables aléatoires qui suivent une loi certaine.

Définition 29.14 (Loi de Bernoulli) – Soit $p \in [0, 1]$. Une variable aléatoire X définie sur un espace probabilisé $(\Omega, \mathbf{P})$ suit la **loi de Bernoulli de paramètre** p si
$$X(\Omega) = \{0, 1\},\ \mathbf{P}(X = 1) = p,\ \mathbf{P}(X = 0) = 1 - p.$$
Lorsque c'est le cas, on note alors $X \sim \mathcal{B}(p)$.

Remarques. ► Une telle variable existe bien car $p + (1 - p) = 1$.
► La situation typique de loi de Bernoulli est le *pile* (1) ou *face* (0), la pièce tombant sur *pile* avec probabilité p. Autrement dit, la variable aléatoire X qui vaut 1 en cas de *pile* et 0 en cas de *face* suit la loi $\mathcal{B}(p)$.
Mais plus généralement, toute expérience à deux issues[13] donne lieu à une variable de Bernoulli, 1 désignant l'une des deux issues de l'expérience et 0 l'autre.
► Toute variable aléatoire qui ne prend que les valeurs 0 et 1 suit une loi de Bernoulli.
En particulier, la variable indicatrice $\mathbb{1}_A$ d'un événement A suit une loi de Bernoulli. Et puisque $\mathbf{P}(\mathbb{1}_A = 1) = \mathbf{P}(A)$, $\mathbb{1}_A$ suit la loi de Bernoulli de paramètre $\mathbf{P}(A)$.
Inversement, si $X \sim \mathcal{B}(p)$, alors $X = \mathbb{1}_{[X=1]}$.
► Enfin, une remarque qui nous servira plus tard : si $X \sim \mathcal{B}(p)$, alors $X^2 = X$.

[13] Qu'on appelle aussi expériences de Bernoulli.

Remarque

Pour une variable qui suit une loi de Bernoulli, la donnée de $\mathbf{P}(X = 1)$ suffit à caractériser entièrement la loi, puisqu'on a alors $\mathbf{P}(X = 0) = 1 - \mathbf{P}(X = 1)$. Ceci vaut plus généralement pour toute variable aléatoire qui ne prend que deux valeurs.

La formule du binôme nous donne une famille de réels positifs dont la somme vaut 1 : si $p \in [0, 1]$, $\displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = (p + (1 - p))^n = 1$.
Faisons-en donc une loi de variable aléatoire[14] !

[14] C'est le sens de la proposition 29.11 : toute famille de réels positifs de somme 1 permet de définir la loi d'une variable aléatoire.

Définition 29.15 (Loi binomiale) – Soit $p \in]0,1[$, et soit $n \in \mathbf{N}^*$. On dit qu'une variable aléatoire suit la **loi binomiale de paramètres n et p** si

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

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

Le contexte où l'on reconnaît naturellement une loi binomiale $\mathscr{B}(n,p)$ est celui de n répétitions **indépendantes** d'épreuves de Bernoulli de probabilité de succès p.
Alors la variable aléatoire qui compte le nombre de succès lors des n répétitions suit la loi binomiale $\mathscr{B}(n,p)$.
Si nous aurons bientôt un argument plus rapide pour prouver ce résultat, expliquons d'où il vient, en notant S_i (resp. E_i) l'événement «la $i^{\text{ème}}$ répétition de l'expérience a amené un succès (resp. un échec).
Soit alors X la variable qui compte le nombre de succès. Il est clair que $X(\Omega) = [\![0,n]\!]$.
Et on a $[X = 0] = E_1 \cap E_2 \cap \cdots \cap E_n$, de sorte que par indépendance des répétitions,

$$\mathbf{P}(X = 0) = \mathbf{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbf{P}(E_1)\mathbf{P}(E_2)\cdots\mathbf{P}(E_n) = (1-p)^n = \binom{n}{0} p^0 (1-p)^{n-0}.$$

Détails
Il s'agit de distinguer n cas suivant la position de l'unique succès.

Ensuite, on a $[X = 1] = \displaystyle\bigcup_{i=1}^{n} \left(S_i \cap \bigcap_{\substack{j=1\\ j\neq i}}^{n} E_j \right)$.
Par incompatibilité de ces événements[15], puis par indépendance, il vient

15 Ceux qui forment l'union.

$$\mathbf{P}(X = 1) = \sum_{i=1}^{n} \mathbf{P}\left(S_i \cap \bigcap_{\substack{j=1\\ j\neq i}}^{n} E_j \right) = \sum_{i=1}^{n} p(1-p)^{n-1} = np(1-p)^{n-1} = \binom{n}{1} p^1 (1-p)^{n-1}.$$

Ensuite, $[X = 2] = \displaystyle\bigcup_{1 \leqslant i < j \leqslant n} \left(S_i \cap S_j \cap \bigcap_{\substack{k=1\\ k\neq i,j}}^{n} E_k \right)$.
Par le même type de raisonnement,

$$\mathbf{P}(X = 2) = \sum_{1 \leqslant i < j \leqslant n} p^2 (1-p)^{n-2} = \binom{n}{2} p^2 (1-p)^{n-2}.$$

En effet, il y a autant de manières de choisir un couple (i,j) avec $1 \leqslant i < j \leqslant n$ qu'il y a de manières de choisir une partie à 2 éléments de $[\![1,n]\!]$, car il n'y aura alors qu'une façon d'ordonner ses éléments. D'où le coefficient $\binom{n}{2}$.
Plus généralement, pour $k \in [\![1,n]\!]$,

$$[X = k] = \bigcup_{\substack{I \subset [\![1,n]\!]\\ \mathrm{Card}(I) = k}} \left(\bigcap_{i \in I} S_i \cap \bigcap_{j \notin I} E_i \right)$$

et comme l'union comporte $\binom{n}{k}$ éléments[16] deux à deux incompatibles, tous de probabilité $p^k(1-p)^{n-k}$, il vient bien

16 C'est le nombre de k combinaisons de $[\![1,n]\!]$.

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

Concrètement, c'est la loi que vous aurez le plus souvent à reconnaître directement, simplement en remarquant que la variable à laquelle vous avez affaire compte le nombre de succès lors d'une répétition d'épreuves de Bernoulli indépendantes.

☠ Danger !
Méfiez-vous des intuitions que vous pouvez avoir sur l'indépendance...

Exemple 29.16

Un géantiste[17] enfourche à chaque manche avec probabilité p (auquel cas il est disqualifié), les manches étant supposées indépendantes les unes des autres.
Sachant que la saison compte n courses, quelle est la loi de la variable aléatoire X comptabilisant le nombre de courses de deux manches qu'il a terminées ?

Pour chacune des courses, la probabilité qu'il termine les deux manches est $(1-p)^2$. Et donc X suit la loi binomiale $\mathscr{B}(n,(1-p)^2)$. En revanche, la variable comptant le nombre de courses où le skieur a pris le départ de la seconde manche suit la loi $\mathscr{B}(n,1-p)$.

[17] Sportif spécialisée dans le slalom géant en ski alpin.

Notons que pour $n=1$, la loi $\mathscr{B}(n,p)$ est la loi de Bernoulli $\mathscr{B}(p)$.

Définition 29.17 (Loi uniforme) – Soit E un ensemble fini. Une variable aléatoire $X:\Omega\to E$ suit la **loi uniforme sur** E si

$$X(\Omega)=E \text{ et } \forall x\in E,\ \mathbf{P}(X=x)=\frac{1}{\operatorname{Card}(E)}.$$

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

Les cas particuliers qui nous intéresserons seront ceux où $E=[\![a,b]\!]$, où $a<b$ sont deux entiers relatifs. Alors si $X\sim\mathscr{U}([\![a,b]\!])$, pour tout $k\in[\![a,b]\!]$, $\mathbf{P}(X=k)=\dfrac{1}{b-a+1}$.
En particulier, si $X\sim\mathscr{U}([\![1,n]\!])$, alors $\forall k\in[\![1,n]\!]$, $\mathbf{P}(X=k)=\dfrac{1}{n}$.

Ces lois se rencontrent essentiellement lors de lancers de dés équilibrés/ de tirages dans une urne, lorsqu'on est en situation d'équiprobabilité.

29.2 ESPÉRANCE D'UNE VARIABLE ALÉATOIRE RÉELLE

Jusqu'à présent, la notion de variable aléatoire nous a surtout fourni des moyens pratiques de nommer des événements, mais guère plus. La notion d'espérance, qui vient donner un sens en probabilités à la notion de moyenne va nous fournir une bonne raison d'étudier davantage les variables aléatoires.

29.2.1 Définition

Définition 29.18 – Soit X une variable aléatoire réelle définie sur un espace probabilisé $(\Omega,\mathscr{P})$. On appelle alors **espérance de** X, et on note $\mathbf{E}(X)$ le réel défini par

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

L'espérance correspond à la moyenne de X sur un grand nombre de répétitions de l'expérience sous-jacente.
En effet, considérons une expérience qui donne lieu à une variable aléatoire X, avec $X(\Omega)=\{x_1,\dots,x_n\}$.
Alors si on répète N fois l'expérience, avec N très grand, on s'attend[18] à ce que X prenne $N\times\mathbf{P}(X=x_i)$ fois la valeur x_i, et donc qu'en moyenne sur les N répétitions, X prenne la valeur

[18] En tous cas c'est l'idée intuitive qu'on se fait d'une probabilité.

$$\frac{1}{N}(x_1N\mathbf{P}(X=x_1)+\cdots+x_nN\mathbf{P}(X=x_n))=\sum_{i=1}^n x_i\mathbf{P}(X=x_i)=\mathbf{E}(X).$$

Bien entendu, sur une réalisation de l'expérience, la notion d'espérance ne signifie rien, et le résultat obtenu peut être très éloigné de l'espérance : si je lance 10 fois une pièce, j'obtiens en moyenne 5 fois pile, mais en pratique vous pouvez tout à fait lancer une pièce

et obtenir 10 piles.
Il est même possible que vous effectuiez 5 séries de 10 lancers et que vous obteniez 5 fois 10 piles[19].

[19] Mais vous avez tout de même plus de chances de gagner au loto !

Notons que si X et Y sont deux variables aléatoires qui suivent la même loi, alors elles ont même espérance, puisque $\mathbf{E}(X)$ ne dépend que des $\mathbf{P}(X=x)$ (c'est-à-dire de la loi de X).

Exemple 29.19

Reprenons l'exemple de 29.10. Alors

$$\begin{aligned}\mathbf{E}(X) &= \frac{1}{\binom{N}{n}}\sum_{k=n}^{N} k\binom{k-1}{n-1} = \frac{1}{\binom{N}{n}}\sum_{k=n}^{N} n\binom{k}{n}\\
&= \frac{n}{\binom{N}{n}}\sum_{k=n}^{N}\left(\binom{k+1}{n+1}-\binom{k}{n+1}\right)\\
&= \frac{n}{\binom{N}{n}}\left(\binom{N+1}{n+1}-\underbrace{\binom{n}{n+1}}_{=0}\right)\\
&= \frac{n(N+1)}{n+1}.\end{aligned}$$

Formule de Pascal.

Somme télescopique.

Définition 29.20 – Une variable aléatoire X sur $(\Omega,\mathbf{P})$ est dite **centrée** si $\mathbf{E}(X)=0$.

29.2.2 Espérance des lois usuelles

Les espérances des lois usuelles sont à connaître par cœur.

Théorème 29.21 : *Soit X une variable aléatoire réelle définie sur un espace probabilisé fini $(\Omega,\mathbf{P})$.*

1. *Si X suit la loi certaine égale à a, alors $\mathbf{E}(X)=a$.*
2. *Si $X\sim\mathscr{B}(p)$, $p\in[0,1]$, alors $\mathbf{E}(X)=p$.*
3. *Si $X\sim\mathscr{B}(n,p)$, $n\in\mathbf{N}^*$, alors $p\in]0,1[$, $\mathbf{E}(X)=np$.*
4. *Si $X\sim\mathscr{U}([\![a,b]\!])$, avec $(a,b)\in\mathbf{Z}^2$, $a<b$, alors $\mathbf{E}(X)=\dfrac{a+b}{2}$.*

Intuition
En moyenne, on obtient le milieu de $[\![a,b]\!]$.

Démonstration. 1. $\mathbf{E}(X)=a\mathbf{P}(X=a)=a$.

2. $\mathbf{E}(X)=0\times\mathbf{P}(X=0)+1\times\underbrace{\mathbf{P}(X=1)}_{=p}=p$.
3. Si $X\sim\mathscr{B}(n,p)$, alors

$$\begin{aligned}\mathbf{E}(X) &= \sum_{k=0}^{n} k\mathbf{P}(X=k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}\\
&= \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k}\\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k}\\
&= n\sum_{i=0}^{n-1}\binom{n-1}{i}p^{i+1}(1-p)^{(n-1)-i}\\
&= np\sum_{i=0}^{n-1}\binom{n-1}{i}p^{i}(1-p)^{n-1-i} = np(p+1-p)^{n-1} = np.\end{aligned}$$

Le terme en $k=0$ est nul.

Chgt d'indice
$i=k-1$

4. Enfin, si $X \sim \mathcal{U}(\llbracket a, b \rrbracket)$,

$$\begin{aligned}\mathbf{E}(X) &= \sum_{k=a}^{b} \frac{k}{b-a+1} = \frac{1}{b-a+1} \sum_{i=0}^{b-a} (i+a) \\ &= \frac{1}{b-a+1}\left(\sum_{i=0}^{b-a} i + a(b-a+1)\right) = \frac{1}{b-a+1}\left(\frac{(b-a)(b-a+1)}{2} + a(b-a+1)\right) = \frac{a+b}{2}.\end{aligned}$$

Chgt d'indice
$k = i + a \Leftrightarrow i = k - a$.

□

29.2.3 Propriétés de l'espérance

Lemme 29.22. Soit X une variable aléatoire réelle sur un espace probabilisé fini $(\Omega, \mathbf{P})$. Alors $\mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbf{P}(\{\omega\})$.

Démonstration. Souvenons-nous que $[X = x] = \bigcup_{\substack{\omega \in \Omega \\ X(\omega)=x}} \{\omega\}$, de sorte que [20]

$$\mathbf{P}(X = x) = \sum_{\substack{\omega \in \Omega \\ X(\omega)=x}} \mathbf{P}(\{\omega\}).$$

[20] L'union est évidemment disjointe.

Et donc il vient

$$\mathbf{E}(X) = \sum_{x \in X(\Omega)} \sum_{\substack{\omega \in \Omega \\ X(\omega)=x}} \underbrace{x}_{=X(\omega)} \mathbf{P}(\{\omega\}).$$

Mais chaque élément de Ω possède une et une seule image par X, et donc

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

□

Proposition 29.23 (Linéarité de l'espérance) : *L'espérance est une forme linéaire sur l'ensemble des variables aléatoires réelles sur Ω : si X et Y sont deux variables aléatoires réelles sur Ω, alors pour tout $\lambda \in \mathbf{R}$,*

$$\mathbf{E}(\lambda X + Y) = \lambda \mathbf{E}(X) + \mathbf{E}(Y).$$

Démonstration. C'est essentiellement un calcul reposant sur le lemme précédent :

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

Double application du lemme précédent.

□

Corollaire 29.24 – *Si X est une variable aléatoire réelle, alors pour tout $(a, b) \in \mathbf{R}^2$, $\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$.*

Démonstration. Il suffit de prendre pour Y la variable constante[21] égale à b. □

[21] Qui suit donc une loi certaine.

Corollaire 29.25 – *Si X est une variable aléatoire réelle sur $(\Omega, \mathbf{P})$, alors $X - \mathbf{E}(X)$ est centrée.*

Démonstration.

$$\mathbf{E}(X - \underbrace{\mathbf{E}(X)}_{\in\mathbf{R}}) = \mathbf{E}(X) - \mathbf{E}(X) = 0.$$

□

Proposition 29.26 (Positivité de l'espérance) : *Soit X une variable aléatoire positive sur Ω. Alors $\mathbf{E}(X) \geqslant 0$.*
De plus, $\mathbf{E}(X) = 0$ si et seulement si X suit la loi certaine égale à 0.

Démonstration. Notons $X(\Omega) = \{x_1, \dots, x_n\}$, où les x_i sont deux à deux distincts. Alors

$$\mathbf{E}(X) = \sum_{i=1}^{n} x_i \mathbf{P}(X = x_i) \geqslant 0$$

car somme de nombres positifs.
De plus, une somme de nombres positifs est nulle si et seulement si tous ces nombres sont nuls, soit ici si et seulement si $\forall i \in [\![1, n]\!]$, $x_i \mathbf{P}(X = x_i) = 0$.
Donc si $x_i \neq 0$, alors $\mathbf{P}(X = x_i) = 0$.
Puisqu'on doit avoir $\sum_{i=1}^{n} \mathbf{P}(X = x_i) = 1$, alors $\mathbf{P}(X = 0) = 1$, de sorte que X suit bien la loi certaine égale à 0. □

Corollaire 29.27 (Croissance de l'espérance) – *Soient X et Y deux variables aléatoires définies sur Ω et telles que $X \leqslant Y$.*
Alors $\mathbf{E}(X) \leqslant \mathbf{E}(Y)$.

Inégalité
$X \leqslant Y$ signifie que quelle que soit l'issue de l'expérience, la valeur de X est inférieure à celle de Y.

Démonstration. Il suffit de noter que $Y - X$ est une variable positive, donc $\mathbf{E}(Y - X) \geqslant 0$ et donc par linéarité de l'espérance, $\mathbf{E}(X) \leqslant \mathbf{E}(Y)$. □

En appliquant toute fonction à une variable aléatoire réelle, on obtient une autre variable aléatoire Y, par exemple e^X, $\frac{1}{X}$ ou X^2.

Notons au passage que si $\sim Y$, c'est-à-dire si X et Y ont même loi, et si f est une fonction définie sur $X(\Omega) = Y(\Omega)$, alors $f(X)$ et $f(Y)$ ont même loi.
En effet, pour tout $x \in f(X(\Omega))$,

$$\mathbf{P}(f(X) = x) = \mathbf{P}(X \in f^{-1}(\{x\})) = \mathbf{P}_X\left(f^{-1}(\{x\})\right) = \mathbf{P}_Y\left(f^{-1}(\{x\})\right) = \mathbf{P}(f(Y) = x).$$

Par exemple si X et Y suivent toutes les deux une loi binomiale $\mathscr{B}(n, p)$, alors X^2 et Y^2 suivent la même loi.

Comment calculer l'espérance de telles variables à partir de la loi de X ?
Une option serait de déterminer la loi de Y, mais si on ne veut que l'espérance, le résultat suivant nous permet d'aller plus vite.

Proposition 29.28 (Théorème de transfert) : *Soit X une variable aléatoire*[22] *sur Ω, et soit $g : X(\Omega) \to \mathbf{R}$. Alors*

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

[22] À valeurs réelles ou non.

Démonstration. Toujours par le même lemme,

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

$$\begin{aligned}
&= \sum_{x \in X(\Omega)} \sum_{\substack{\omega \in \Omega \\ X(\omega)=x}} g(x)\mathbf{P}(\{\omega\}) \\
&= \sum_{x \in X(\Omega)} g(x) \sum_{\substack{\omega \in \Omega \\ X(\omega)=x}} \mathbf{P}(\{\omega\}) \\
&= \sum_{x \in X(\Omega)} g(x)\mathbf{P}(X = x).
\end{aligned}$$

□

Détails

On a partitionné Ω en

$$\bigcup_{x \in X(\Omega)} [X = x].$$

Autrement dit, on utilise le système complet d'événements associé à X.

Exemple 29.29

Soit X une variable aléatoire qui suit la loi binomiale $\mathscr{B}(n, p)$, et soit $Y = X(X-1)$. Alors par le théorème de transfert, appliqué avec la fonction $x \mapsto x(x-1)$,

$$\begin{aligned}
\mathbf{E}(Y) &= \sum_{k=0}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1)\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2}p^k(1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{i=0}^{n-2} \binom{n-2}{i}p^i(1-p)^{n-2-i} \\
&= n(n-1)p^2(p+1-p)^{n-2} = n(n-1)p^2.
\end{aligned}$$

⚠ Attention !

L'espérance d'un produit n'est en général pas égale au produit des espérances.

Enfin, terminons par une inégalité, qui fait partie d'une grande famille d'inégalités[23] nommées inégalités de concentration, et qui visent à décrire comment les valeurs prises par une variable aléatoire X se «concentrent» autour de son espérance.

[23] Dont seules deux sont au programme, la suivante étant l'inégalité de Bienaymé-Tchebychev.

Théorème 29.30 (Inégalité de Markov) : *Soit X une variable aléatoire **positive**. Alors pour tout $a > 0$,*

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

Démonstration. Soit $a > 0$, et notons $A = [X \geqslant a]$.
Alors $X \geqslant a\mathbb{1}_A$. En effet, pour $\omega \in \Omega$, on a :

- soit $\omega \in A$, auquel cas $X(\omega) \geqslant a$ et $a\mathbb{1}_A(\omega) = a$
- soit $\omega \notin A$, auquel cas $X(\omega) \geqslant 0$ et $a\mathbb{1}_A(\omega) = 0$.

Et donc par croissance de l'espérance, $\mathbf{E}(X) \geqslant a\mathbf{E}(\mathbb{1}_A)$.
Mais $\mathbb{1}_A$ suit la loi de Bernoulli de paramètre $\mathbf{P}(A)$, et donc $\mathbf{E}(\mathbb{1}_A) = \mathbf{P}(A) = \mathbf{P}(X \geqslant a)$, de sorte que

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

□

29.3 VARIANCE D'UNE VARIABLE ALÉATOIRE RÉELLE

29.3.1 Définition

Définition 29.31 – Soit X une variable aléatoire réelle sur $(\Omega, \mathbf{P})$. On appelle **variance de** X le réel $\mathbf{V}(X)$ défini par

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

En une phrase : la variance est la moyenne des carrés des écarts à la moyenne.
L'intuition, vous l'avez déjà car vous connaissez la notion de variance en statistiques : la variance est une mesure de dispersion.
Plus la variance est élevée, plus X prend des valeurs éloignées de son espérance.
Et plus la variance est faible, plus, au contraire, X prend ses valeurs autour de $\mathbf{E}(X)$.

Comme pour l'espérance, la variance ne dépend que de la loi, au sens où deux variables de même loi auront même variance.
En effet, par le théorème de transfert, si X et Y ont même loi, alors

$$\mathbf{V}(X) = \sum_{x\in X(\Omega)} (x - \mathbf{E}(X))^2 \mathbf{P}(X = x) = \sum_{y\in Y(\Omega)} (y - \mathbf{E}(Y))^2 \mathbf{P}(Y = y) = \mathbf{V}(Y).$$

Définition 29.32 – Soit X une variable aléatoire réelle sur $(\Omega, \mathbf{P})$, et soit $k \in \mathbf{N}^*$.
Le **moment d'ordre k de X** est $m_k(X) = \mathbf{E}\left(X^k\right)$.
Notons que par le théorème de transfert, $m_k(X) = \sum_{x\in X(\Omega)} x^k \mathbf{P}(X = x)$.

Le moment d'ordre 2 est intimement lié à la variance, comme l'indique la formule ci-dessous, qui nous fournira souvent un moyen agréable de calculer une variance.

Proposition 29.33 (Formule de (Kœnig-)Huygens) : *Soit X une variable aléatoire réelle. Alors*

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

> **Astuce**
> Cette formule relie trois quantités. Et cela signifie donc que si vous en connaissez deux vous connaissez la troisième.
> En particulier, pour les lois usuelles, dont vous connaissez déjà l'espérance, vous allez aussi apprendre la variance. Ce qui signifie que vous serez capable de retrouver le moment d'ordre 2 sans avoir à l'apprendre.

Démonstration. On a $(X - \mathbf{E}(X))^2 = X^2 - 2\underbrace{\mathbf{E}(X)}_{\in\mathbf{R}}X + \mathbf{E}(X)^2$.

Donc par linéarité de l'espérance,

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

□

29.3.2 Propriétés de la variance

Proposition 29.34 : *Soit X une variable aléatoire réelle sur $(\Omega, \mathbf{P})$. Alors $\mathbf{V}(X) \geqslant 0$, et $\mathbf{V}(X) = 0$ si et seulement si X suit une loi certaine.*

Démonstration. Puisque $(X - \mathbf{E}(X))^2 \geqslant 0$, par positivité de l'espérance, $\mathbf{V}(X) \geqslant 0$.
De plus, on a alors $\mathbf{V}(X) = 0$ si et seulement si $(X - \mathbf{E}(X))^2$ suit la loi certaine égale à 0.
Soit si et seulement si $X - \mathbf{E}(X)$ suit la loi certaine égale à 0.
Soit encore si et seulement si X suit la loi certaine égale à $\mathbf{E}(X)$. □

Proposition 29.35 : *Si X est une variable aléatoire réelle, alors $\forall (a, b) \in \mathbf{R}^2$, $\mathbf{V}(aX + b) = a^2\mathbf{V}(X)$.*

⚠ Le fait que $\mathbf{V}(aX)$ ne soit en général pas égal à $a\mathbf{V}(X)$ prouve que, contrairement à l'espérance, la variance n'est pas linéaire.
Nous verrons un peu plus loin ce qu'on peut dire de la variance d'une somme, mais ce n'est en général pas égal à la somme des variances.

Démonstration. On a $\mathbf{E}(aX + b) = a\mathbf{E}(X) + b$, et donc

$$(aX + b - \mathbf{E}(aX + b))^2 = (aX + b - a\mathbf{E}(X) - b)^2 = a^2 (X - \mathbf{E}(X))^2.$$

On conclut alors par linéarité de l'espérance. □

Proposition 29.36 (Inégalité de Bienaymé-Tchebychev) : *Soit X une variable aléatoire sur $(\Omega, \mathbf{P})$. Alors*

$$\forall \varepsilon > 0,\ \mathbf{P}\left(|X - \mathbf{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{\mathbf{V}(X)}{\varepsilon^2}.$$

Démonstration. Notons que la variable $(X - \mathbf{E}(X))^2$ est positive, et que par définition, son espérance est $\mathbf{V}(X)$. Appliquons lui l'inégalité de Markov : $\forall \varepsilon > 0$,

$$\mathbf{P}\left((X - \mathbf{E}(X))^2 \geqslant \varepsilon^2\right) \leqslant \frac{\mathbf{V}(X)}{\varepsilon^2}.$$

Mais les événements $\left[(X - \mathbf{E}(X))^2 \geqslant \varepsilon^2\right]$ et $[|X - \mathbf{E}(X)| \geqslant \varepsilon]$ sont égaux, et donc ont même probabilité. □

Définition 29.37 – L'écart-type $\sigma(X)$ d'une variable aléatoire X est $\sigma(X) = \sqrt{\mathbf{V}(X)}$.

Définition 29.38 – Une variable aléatoire X est dite **réduite** si $\mathbf{V}(X) = 1$.

Proposition 29.39 : *Soit X une variable aléatoire qui ne suit pas une loi certaine*[24]. *Alors $Y = \dfrac{X - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)}}$ est une variable centrée réduite, qu'on appelle variable centrée réduite associée à X.*

[24] De sorte que sa variance est non nulle

Démonstration. Par linéarité de l'espérance,

$$\mathbf{E}(Y) = \frac{1}{\sqrt{\mathbf{V}(X)}}\mathbf{E}\left(X - \mathbf{E}(X)\right) = \frac{\mathbf{E}(X) - \mathbf{E}(X)}{\sqrt{\mathbf{V}(X)}} = 0.$$

Donc Y est centrée. Et de plus,

$$\mathbf{V}(Y) = \left(\frac{1}{\sqrt{\mathbf{V}(X)}}\right)^2 \mathbf{V}(X - \mathbf{E}(X)) = \frac{1}{\mathbf{V}(X)}\mathbf{V}(X) = 1$$

donc Y est réduite. □

29.3.3 Variance des lois usuelles

Théorème 29.40 : *Soit X une variable aléatoire sur un espace probabilisé fini $(\Omega, \mathbf{P})$.*

1. *Si X suit une loi certaine, $\mathbf{V}(X) = 0$.*
2. *Si X suit une loi de Bernoulli $\mathcal{B}(p)$, alors $\mathbf{V}(X) = p(1-p)$.*
3. *Si X suit la loi binomiale $\mathcal{B}(n,p)$, alors $\mathbf{V}(X) = np(1-p)$.*
4. *Si X suit la loi uniforme sur $[\![a, b]\!]$, alors $\mathbf{V}(X) = \dfrac{(b-a+1)^2 - 1}{12}$.*

   En particulier, si $X \sim \mathcal{U}([\![1, n]\!])$, alors $\mathbf{V}(X) = \dfrac{n^2 - 1}{12}$.

Démonstration.

1. $X - \mathbf{E}(X)$ est une variable certaine égale à 0, donc il en est de même de son carré, et donc elle est d'espérance nulle.
2. On a $X^2 = X$, et donc $\mathbf{E}\left(X^2\right) = \mathbf{E}(X) = p$, et donc par la formule de Kœnig-Huygens, on a $\mathbf{V}(X) = p - p^2 = p(1-p)$.

3. Plutôt que de calculer le moment d'ordre 2 de X, par le théorème de transfert, remarquons que dans l'exemple qui suit le théorème de transfert, nous avons déjà calculé $\mathbf{E}(X(X-1)) = n(n-1)p^2$.
   Et donc par linéarité de l'espérance,
   $$\mathbf{E}\left(X^2\right) = \mathbf{E}(X(X-1)) + \mathbf{E}(X) = n(n-1)p^2 + np = n^2p^2 + np(1-p).$$
   Par la formule de Huygens, il vient donc
   $$\mathbf{V}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2 = n^2p^2 + np(1-p) - (np)^2 = np(1-p).$$
4. Commençons par traiter le cas de $\mathscr{U}(\llbracket 1, n \rrbracket)$. Alors
   $$\mathbf{E}\left(X^2\right) = \sum_{k=1}^{n} k^2 \frac{1}{n} = \frac{(n+1)(2n+1)}{6}.$$
   Et donc par la formule de Kœnig-Huygens,
   $$\mathbf{V}(X) = \mathbf{E}\left(X^2\right) - \mathbf{E}(X)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(4n+2-3n-3)}{12} = \frac{n^2-1}{12}.$$
   Dans le cas général, si $X \sim \mathscr{U}(\llbracket a, b \rrbracket)$, alors $X - a + 1$ suit la loi $\mathscr{U}(\llbracket 1, b-a+1 \rrbracket)$.
   Et donc $\mathbf{V}(X) = \mathbf{V}(X - a + 1) = \dfrac{(b-a+1)^2 - 1}{12}$.

□

Exemple 29.41

Donnons un exemple d'application de l'inégalité de Bienaymé-Tchebychev : combien de fois faut-il lancer un dé équilibré pour garantir avec moins de 5% d'erreur que la fréquence d'apparition du 2 sera $\frac{1}{6} \pm 0.01$?

Soit n le nombre de lancers, et soit X le nombre de lancers dont le résultat est 2.
La fréquence d'apparition du 2 est $Y = \frac{X}{n}$.
Nous cherchons donc quelle valeur donner à n pour avoir
$$P\left(\left|Y - \frac{1}{6}\right| < \frac{1}{100}\right) > 0.95.$$
Autrement dit, nous voulons $P\left(\left|Y - \frac{1}{6}\right| \geqslant 0.01\right) \leqslant 0.05$.
Puisque X suit une loi binomiale $\mathscr{B}\left(n, \frac{1}{6}\right)$, on a $E(X) = \frac{n}{6}$ et $V(X) = n\frac{5}{36}$.
Et donc $E(Y) = \frac{1}{6}$ et $V(Y) = \frac{1}{n^2}V(X) = \frac{5}{36n}$.
D'après l'inégalité de Bienaymé-Tchebychev, appliquée à Y avec $\varepsilon = 0.01$,
$$P\left(\left|Y - \frac{1}{6}\right| \geqslant 0.01\right) \leqslant \frac{50000}{36n}.$$
Ainsi, pour $n \geqslant \frac{50000}{36 \times 0.05} \Leftrightarrow n \geqslant 27\,778$, on a l'inégalité souhaitée.
Cette inégalité est loin d'être optimale, et un calcul détaillé avec `Python`, en utilisant la loi binomiale, prouve qu'en fait $n \geqslant 5395$ suffit.

29.4 Couples et n-uplets de variables aléatoires

29.4.1 Définitions

Définition 29.42 – Soient E et E' deux ensembles. Si $X : \Omega \to E$ et $Y : \Omega \to E'$ sont deux variables aléatoires sur Ω, alors l'application $(X, Y) : \left| \begin{array}{ccc} \Omega & \longrightarrow & E \times E' \\ \omega & \longmapsto & (X(\omega), Y(\omega)) \end{array} \right.$
est appelée **couple de variables aléatoires sur** Ω.
Et si X, Y sont des variables aléatoires réelles, on dit que (X, Y) est un couple de variables aléatoires réelles.

Autrement dit
Un couple de variables aléatoires est une variable aléatoire à valeurs dans $E \times E'$.

Bref, rien de très surprenant : un couple de variables aléatoires, c'est la donnée de deux variables aléatoires.

Exemple 29.43

On tire deux boules sans remise dans une urne qui en contient n, et on note X le numéro de la plus petite des deux et Y le numéro de la plus grande.
Alors (X, Y) est un couple de variables aléatoires réelles sur l'espace Ω modélisant cette expérience.

Comme pour toute variable aléatoire, à un couple (X, Y) est associé un système complet d'événements, qui est $\{[(X, Y) = (x, y)], (x, y) \in (X, Y)(\Omega)\}$. Si on a toujours $(X, Y)(\Omega)$ inclus dans $X(\Omega) \times Y(\Omega)$, il n'y a pas nécessairement égalité.
Dans l'exemple ci-dessus, $(3, 3) \in X(\Omega) \times Y(\Omega)$, mais n'est pas dans $(X, Y)(\Omega)$.
Quitte à ajouter des événements vides, on préférera souvent manipuler le système complet d'événements $\{[(X, Y) = (x, y)], (x, y) \in X(\Omega) \times Y(\Omega)\}$.
Enfin, notons que $[(X, Y) = (x, y)] = [X = x] \cap [Y = y]$, notation souvent plus explicite.
On pourra donc considérer que le système complet d'événements associé au couple (X, Y) est $\{[X = x] \cap [Y = y], (x, y) \in X(\Omega) \times Y(\Omega)\}$.

29.4.2 Loi conjointe, lois marginales

Définition 29.44 – Si (X, Y) est un couple de variables aléatoires sur Ω, la loi de $(X, Y) : \Omega \to E \times E'$ est appelée **loi conjointe** du couple (X, Y).

Notons que par la proposition 29.9, pour caractériser entièrement la loi conjointe, il suffit de se donner les $\mathbf{P}([X = x] \cap [Y = y])$ pour $(x, y) \in X(\Omega) \times Y(\Omega)$.

Définition 29.45 – Si (X, Y) est un couple de variables aléatoires sur $(\Omega, \mathbf{P})$, la loi de X est appelée première **loi marginale** du couple (X, Y), et la loi de Y est appelée seconde loi marginale.

La proposition qui suit nous dit que la loi conjointe permet de retrouver les lois marginales.
Mais insistons tout de suite sur le fait qu'il n'y a pas de réciproque : deux couples de lois conjointes distinctes peuvent avoir les mêmes lois marginales.
Pour donner un exemple simple, toujours avec deux lancers de dés, où l'on note X le résultat du premier dé et Y le résultat du second.
Dans ce cas, les couples (X, Y) et (X, X) ont mêmes lois marginales, les deux lois marginales de chacun des couples sont des lois uniformes sur $[\![1, 6]\!]$.
Pour autant, il est clair que ces couples n'ont pas même loi conjointe. Osons prononcer le mot même s'il ne sera défini qu'un peu plus tard : dans le premier cas, les deux variables sont *indépendantes*, alors que dans le second elles ne le sont pas.
Plus explicitement : on a $\mathbf{P}([X = 1] \cap [Y = 2]) = \dfrac{1}{36}$ alors que $\mathbf{P}([X = 1] \cap [X = 2]) = 0$.
Donc les lois conjointes ne sont pas identiques.
La subtilité vient du fait que la loi conjointe, «contient» non seulement les lois marginales, mais aussi les éventuels liens qui peuvent exister entre les deux variables.

Proposition 29.46 : *Soit (X, Y) un couple de variables aléatoires. Alors*

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

> **Et pour Y ?**
> Il y a bien entendu une formule analogue pour la loi de Y, en sommant sur le support de X.

Démonstration. Ce n'est rien d'autre que la formule des probabilités totales appliquée au système complet d'événements $\{[Y = y], y \in Y(\Omega)\}$. □

Exemple 29.47

Une urne contient n jetons numérotés de 1 à n (avec $n \geqslant 3$). On tire sans remise 3 jetons de l'urne, et on note X (resp. Y) le plus petit (resp. le plus grand) des numéros ainsi obtenus.
Puisque les tirages sans remise peuvent être assimilés à des tirages simultanés, et que tous les tirages sont équiprobables, on peut déterminer la loi conjointe de (X, Y) par dénombrement.
Il y a en tout $\binom{n}{3}$ tirages possibles.
Sans vouloir décrire complètement le support de (X, Y), notons qu'il est inclus dans $[\![1, n]\!]^2$.
► Si $i \geqslant j$, il n'y a évidemment pas de possibilité d'avoir $[X = i] \cap [Y = j]$, et donc $\mathbf{P}((X, Y) = (i, j)) = 0$.
► De même, si $i = j - 1$, alors il n'y a pas de tirage dont le plus grand numéro serait j et le plus petit serait i (quel numéro devrait alors porter le troisième jeton ?).
► Enfin, si $i < j - 1$, alors il y a $j - i - 1$ tirages dont le plus grand numéro est j et le plus petit est i, soit autant qu'il y a de manières de tirer le troisième jeton dans $[\![i + 1, j - 1]\!]$.
Donc la loi conjointe de (X, Y) est

$$\forall (i, j) \in [\![1, n]\!]^2,\ \mathbf{P}([X = i] \cap [Y = j]) = \mathbf{P}\big((X, Y) = (i, j)\big) = \begin{cases} \dfrac{j - i - 1}{\binom{n}{3}} & i \leqslant j - 2 \\ 0 & \text{sinon} \end{cases}$$

Pour obtenir la loi de X, il suffit d'appliquer la formule des probabilités totales avec le système complet d'événements $\{[Y = j], 2 \leqslant j \leqslant n\}$.

$$\begin{aligned} \mathbf{P}(X = i) &= \sum_{j=1}^{n} \mathbf{P}([X = i] \cap [Y = j]) \\ &= \sum_{j=i+2}^{n} \frac{j - i - 1}{\binom{n}{3}} = \frac{6}{n(n-1)(n-2)} \sum_{k=1}^{n-i-1} k \\ &= \frac{3(n-i)(n-i-1)}{n(n-1)(n-2)} \end{aligned}$$

On obtiendrait de même la seconde loi marginale (la loi de Y).

Remarque
Si on se donne un ensemble qui contient strictement le support, alors certaines des probabilités qui suivent vont être nulles. Mais lorsque cela simplifie la rédaction, il ne faut pas se priver.

Et $j = 1$?
Le système complet d'événements donné commence à 2 puisque Y ne peut pas prendre la valeur 1.
Mais si vous ajoutez $[Y = 1]$ (qui est $\emptyset$), vous avez toujours un système complet d'événements.

29.4.3 Lois conditionnelles

Il n'est pas rare qu'une expérience aléatoire se déroule en plusieurs étapes, et que le résultat de la première influe sur le déroulement des suivantes.

Exemple 29.48

Une urne contient 4 boules numérotées de 1 à 4.
On tire une boule de l'urne, on note X le numéro qu'elle porte, et on ne la remet pas dans l'urne.
On effectue alors n tirages avec remise dans cette urne, et on note Y le nombre de boules portant des numéros impairs ainsi obtenus.
Il est évident que X suit une loi uniforme sur $[\![1, 4]\!]$.
Si $[X = 2]$ ou $[X = 4]$, alors Y suit une loi binomiale de paramètres n et $\dfrac{2}{3}$, alors que si $[X = 1]$ ou $[X = 3]$, alors Y suit la loi binomiale $\mathscr{B}\left(n, \dfrac{1}{3}\right)$.
Mais quelle est la loi de Y ? Entendons-nous bien qu'on ne peut pas faire de

distinction de cas dans notre réponse, et donner un résultat du type :

$$\mathbf{P}(Y = k) = \begin{cases} \star & \text{si } X \text{ est pair} \\ \diamond & \text{si } X \text{ est impair} \end{cases}$$

une proba est **un** réel.
En réalité, les lois données plus haut ne sont pas directement la loi de Y pour la probabilité $\mathbf{P}$, mais des probabilités conditionnelles :

$$\mathbf{P}_{[X=2]}(Y = k) = \binom{n}{k}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n-k}.$$

Définition 29.49 – Soit (X, Y) un couple de variables aléatoires, et soit $x \in X(\Omega)$ tel que $\mathbf{P}(X = x) \neq 0$.
Alors la **loi conditionnelle de Y sachant** $[X = x]$ est la loi de Y pour la probabilité $\mathbf{P}_{[X=x]}$ (et non pour la probabilité $\mathbf{P}$).
Autrement dit, c'est la donnée des $\mathbf{P}_{[X=x]}(Y = y)$, pour $y \in Y(\Omega)$.
On définit de même la loi conditionnelle de X sachant $[Y = y]$.

Hypothèse
La condition qui est là sert juste à légitimer l'emploi de probabilités conditionnelles.

Notons qu'on pourrait plus généralement définir des lois conditionnelles sachant un événement A de probabilité non nulle.

⚠ Il n'y a pas de notation standard pour les lois conditionnelles, n'allez pas en inventer, du type $Y \mid [X = 2] \sim \mathscr{B}\left(n, \frac{2}{3}\right)$.
Cette notation est vraiment incorrecte car elle laisse entendre qu'il existe une variable $Y \mid [X = 2]$, alors que les événements conditionnels n'existent pas.

Comme dans l'exemple ci-dessus, il arrive qu'on reconnaisse directement des lois conditionnelles, par exemple car il s'agit de lois usuelles.
Mais alors pour obtenir la loi (sans condition cette fois), il suffit d'appliquer une formule des probabilités totales :

Exemple 29.50

On reprend l'exemple ci-dessus. Pour $k \in [\![0, n]\!]$, appliquons la formule des probabilités totales avec le système complet d'événements $\{[X = i], 1 \leqslant i \leqslant 4\}$. Alors

$$\begin{aligned}\mathbf{P}(Y = k) &= \sum_{i=1}^{4} \mathbf{P}(X = i)\mathbf{P}_{[X=i]}(Y = k) \\ &= \frac{1}{2}\binom{n}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{n-k} + \frac{1}{2}\binom{n}{k}\left(\frac{2}{3}\right)^k\left(\frac{1}{3}\right)^{n-k} \\ &= \frac{1}{2\cdot 3^n}\binom{n}{k}\left(2^{n-k} + 2^k\right).\end{aligned}$$

Détails
On a regroupé les termes $X = 1$ et $X = 3$, ainsi que $X = 2$ et $X = 4$.

29.4.4 Généralisation aux n-uplets de variables aléatoires

Tout ce qui vient d'être fait pour deux variables peut être généralisé à n variables. Les énoncés sont alors laborieux, je les donne pour que vous puissiez vous assurer que vous les comprenez. Si c'est le cas, c'est que vous avez bien compris le cas des couples. Sinon, c'est sûrement qu'il vous reste des points à clarifier.
Dans les deux cas, il est inutile[25] d'apprendre par cœur les définitions ou formules qui vont suivre.

[25] Contre-productif ?

Définition 29.51 – Si $X_1, X_2, \dots, X_n$ sont des variables aléatoires sur Ω, à valeurs dans respectivement $E_1, E_2, \dots, E_n$, on dit que l'application

$$(X_1, \dots, X_n) : \left| \begin{array}{ccc} \Omega & \longrightarrow & E_1 \times \cdots \times E_n \\ \omega & \longmapsto & (X_1(\omega), \dots, X_n(\omega)) \end{array} \right.$$

est un **n-uplet de variables aléatoires** sur Ω (ou encore un vecteur aléatoire).
La loi de la variable aléatoire[26] $(X_1, \dots, X_n)$ est alors appelée loi conjointe de $(X_1, \dots, X_n)$, et les lois des X_i sont appelée lois marginales du n-uplet $(X_1, \dots, X_n)$.

[26] À valeurs dans le produit $E_1 \times \cdots \times E_n$.

Comme pour les couples, la loi conjointe est caractérisée par la donnée des

$$\mathbf{P}([X_1 = x_1] \cap \cdots \cap [X_n = x_n]),\ (x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega).$$

La formule des probabilités totales permet d'obtenir la $i^{\text{ème}}$ loi marginale à partir de la loi conjointe : $\forall x \in X_i(\Omega)$,

$$\mathbf{P}(X_i = x) = \sum_{\substack{x_1 \in X_1(\Omega), \dots, x_{i-1} \in X_{i-1}(\Omega) \\ x_{i+1} \in X_{i+1}(\Omega), \dots, x_n \in X_n(\Omega)}} \mathbf{P}([X_1 = x_1] \cap \cdots \cap [X_i = x] \cap \cdots \cap [X_n = x_n]).$$

Enfin, comme pour les couples, il existe un système complet d'événements associé à $(X_1, \dots, X_n)$, je vous laisse le soin de l'écrire.

29.5 Indépendance de variables aléatoires

Comme pour l'indépendance des événements, vous devez avoir déjà une intuition de ce que sont deux variables indépendantes : la valeur de l'une n'influe pas sur la valeur de l'autre. Reste à formaliser cela, et ce n'est pas le plus agréable. Toutes les preuves de cette partie peuvent être mises de côté, mais les définitions et les résultats sont évidemment à connaître.

29.5.1 Définition

Définition 29.52 – Soient $X : \Omega \to E$ et $Y : \Omega \to E'$ deux variables aléatoires définies sur $(\Omega, \mathbf{P})$. On dit que **X et Y sont indépendantes** si $\forall A \in \mathcal{P}(E)$ et $\forall B \in \mathcal{P}(E')$, les événements $[X \in A]$ et $[Y \in B]$ sont indépendants :

$$\mathbf{P}([X \in A] \cap [Y \in B]) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B).$$

Les événements $[X \in A]$, c'est-à-dire ceux de la forme $[X = 1]$, $[X \geqslant 3]$, $[|X| > 5]$, etc sont ceux «qui ne dépendent que de X», et idem pour Y.
Par exemple, pour un lancer de deux dés, «X est pair», «X ne vaut ni 1 ni 6» sont des événements de la forme $[X \in A]$ pour un $A \in \mathcal{P}(\Omega)$ bien choisi, mais «la somme des deux dés vaut 7», ou «les deux dés donnent des résultats identiques» ne le sont pas.
L'indépendance des variables aléatoires X et Y peut se résumer en «tout événement ne dépendant que de X est indépendant de tout événement ne dépendant que de Y».

Le principal inconvénient de cette définition est que l'indépendance semble fastidieuse à vérifier, s'il faut le faire pour toutes parties de E et de E'. Penser par exemple au cas[27] de variables aléatoires à valeurs réelles !

[27] Il y a **beaucoup** de parties de $\mathbf{R}$!

Proposition 29.53 : *Deux variables aléatoires X et Y sur $(\Omega, \mathbf{P})$ sont* ***indépendantes*** *si et seulement si $\forall (x, y) \in X(\Omega) \times Y(\Omega)$, les événements $[X = x]$ et $[Y = y]$ sont indépendants :*

$$\mathbf{P}([X = x] \cap [Y = y]) = \mathbf{P}(X = x)\mathbf{P}(Y = y).$$

Conjointe/marginales
Ceci signifie que pour deux variables **indépendantes**, les lois marginales caractérisent entièrement la loi conjointe.

Démonstration. Un des sens est évident, il suffit de prendre $A = \{x\}$ et $B = \{y\}$.
Dans l'autre sens, on a

$$\mathbf{P}(X \in A) = \sum_{x \in A} \mathbf{P}(X = x).$$

Par ailleurs,

$$\begin{aligned}[X \in A] \cap [Y \in B] &= \left(\bigcup_{x\in A}[X=x]\right) \cap \left(\bigcup_{y\in B}[Y=y]\right)\\ &= \bigcup_{x\in A}\bigcup_{y\in B}([X=x]\cap[Y=y]).\end{aligned}$$

Loi de De Morgan.

Détails

L'hypothèse faite par l'énoncé était que cette relation ne valait que pour $x \in X(\Omega)$ et $y \in Y(\Omega)$. Mais si $x \notin X(\Omega)$, alors $[X = x] = \emptyset$ et donc

$$\mathbf{P}([X=x]\cap[Y=y]) = \mathbf{P}(\emptyset) = 0$$

et on a aussi

$$\mathbf{P}(X=x)\mathbf{P}(Y=y) = 0.$$

Il est évident que cette union est disjointe, et donc

$$\begin{aligned}\mathbf{P}([X\in A]\cap[Y\in B]) &= \sum_{x\in A}\sum_{y\in B}\mathbf{P}([X=x]\cap[Y=y])\\ &= \sum_{x\in A}\sum_{y\in B}\mathbf{P}(X=x)\mathbf{P}(Y=y)\\ &= \sum_{x\in A}\mathbf{P}(X=x)\sum_{y\in B}\mathbf{P}(Y=y)\\ &= \sum_{x\in A}\mathbf{P}(X=x)\mathbf{P}(Y\in B) = \mathbf{P}(X\in A)\mathbf{P}(Y\in B).\end{aligned}$$

□

En pratique prouver l'indépendance de variables aléatoires, même à l'aide de cette caractérisation reste fastidieux. Fort heureusement, vous aurez rarement à le faire, l'indépendance étant plus souvent une hypothèse fournie par l'énoncé.
En revanche, il est plus fréquent d'avoir à justifier la non indépendance de deux variables aléatoires. Il est clair qu'il faut alors exhiber un couple (x, y) pour lequel $[X = x]$ et $[Y = y]$ ne sont pas indépendants.
Le plus simple est souvent de trouver un couple (x, y) tel que les événements $[X = x]$ et $[Y = y]$ soient de probabilité non nulles, mais incompatibles.
Car alors $\mathbf{P}([X = x] \cap [Y = y]) = 0 \neq \mathbf{P}(X = x)\mathbf{P}(Y = y)$.

Exemple 29.54

On tire deux boules sans remise dans une urne qui en contient n, numérotées de 1 à n. On note X le numéro de la première et Y le numéro de la seconde.
Alors $\mathbf{P}(X = 1)$ et $\mathbf{P}(Y = 1)$ sont non nuls
En revanche, on ne peut avoir la même boule aux deux tirages, donc $\mathbf{P}([X = 1] \cap [Y = 1]) = 0$, de sorte que X et Y ne sont pas indépendantes.

Remarque

En fait on a immédiatement $\mathbf{P}(X = 1) = \frac{1}{n}$, et il n'est pas beaucoup plus cher de prouver que $\mathbf{P}(Y = 1)$ vaut aussi $\frac{1}{n}$, mais ne perdons pas de temps à le faire, tout ce qui compte est leur non nullité.

Proposition 29.55 : *Soit (X, Y) un couple de variables aléatoires sur $(\Omega, \mathbf{P})$. Alors X et Y sont indépendantes si et seulement si $\forall y \in Y(\Omega)$ tel que $\mathbf{P}(Y = y) \neq 0$, la loi conditionnelle de X sachant $[Y = y]$ est égale à la loi de X.*

Démonstration. $\boxed{\Rightarrow}$ Soit y tel que $\mathbf{P}(Y = y) \neq 0$, et soit $x \in X(\Omega)$. Alors

$$\mathbf{P}_{[Y=y]}(X=x) = \frac{\mathbf{P}([X=x]\cap[Y=y])}{\mathbf{P}(Y=y)} = \frac{\mathbf{P}(X=x)\mathbf{P}(Y=y)}{\mathbf{P}(Y=y)} = \mathbf{P}(X=x).$$

Donc la loi conditionnelle de X sachant $[Y = y]$ est égale à la loi de X.

$\boxed{\Leftarrow}$ Soit $(x, y) \in X(\Omega) \times Y(\Omega)$.
► Si $\mathbf{P}(Y = y) \neq 0$, alors

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

► Si $\mathbf{P}([Y = y])$, alors $[X = x] \cap [Y = y] \subset [Y = y]$, et donc par croissance de la probabilité,

$$0 \leqslant \mathbf{P}([X=x]\cap[Y=y]) \leqslant \mathbf{P}(Y=y) = 0$$

si bien que $\mathbf{P}([X = x] \cap [Y = y]) = 0 = \mathbf{P}(X = x)\mathbf{P}(Y = y)$.
Donc X et Y sont bien indépendantes. □

Là aussi c'est assez intuitif : par exemple lors d'un lancer de deux dés, lorsqu'on dit que le résultat du second dé est indépendant de celui du premier, on entend par là que la connaissance du résultat du premier dé ne change rien au comportement du second dé. Autrement dit que sa loi n'est pas changée par la connaissance du premier dé.

Enfin, notons que l'indépendance est une notion qui dépend de la probabilité $\mathbf{P}$ choisie. Sur un même espace Ω, on peut avoir des variables aléatoires X et Y qui sont indépendantes pour une probabilité $\mathbf{P}_1$ et pas pour une probabilité $\mathbf{P}_2$.

Exemple 29.56

Lançons deux fois une pièce. On peut alors prendre $\Omega = \{P, F\}^2$, la probabilité qu'on met sur cet espace dépendant de la probabilité que notre pièce a de tomber sur chacun de ses côtés.
Soit X la variable indicatrice de «les deux lancers donnent le même résultat», et soit Y la variable indicatrice de «le premier lancer donne *face*».
Alors X et Y suivent des lois de Bernoulli. Donc elles sont indépendantes si et seulement si

$$\forall (x, y) \in \{0, 1\}^2,\ \mathbf{P}([X = x] \cap [Y = y]) = \mathbf{P}(X = x)\mathbf{P}(Y = y).$$

Mais un résultat classique[28] affirme que deux événements A et B sont indépendants si et seulement si tout couple d'événements dans $\{A, \overline{A}\} \times \{B, \overline{B}\}$ est formé de deux événements indépendants. Puisqu'ici $[X = 0] = \overline{[X = 1]}$, il suffit donc de déterminer quand $[X = 1]$ et $[Y = 1]$ sont indépendants.
Notons p la probabilité qu'a notre pièce de tomber sur *face*.
Alors $\mathbf{P}(Y = 1) = p$, et $\mathbf{P}(X = 1) = p^2 + (1 - p)^2$.
Par ailleurs, $\mathbf{P}([X = 1] \cap [Y = 1]) = \mathbf{P}(\{(F_1, F_2)\}) = p^2$.
On vérifie alors facilement[29] qu'on a $\mathbf{P}([X = 1] \cap [Y = 1]) = \mathbf{P}(X = 1)\mathbf{P}(Y = 1)$ si et seulement si $p = \dfrac{1}{2}$.
Donc pour $p = \frac{1}{2}$, X et Y sont indépendantes, alors qu'elles ne le sont pas sinon. Bien que d'un point de vue «fonction sur Ω», ce soient toujours les mêmes objets qui sont en jeu.

[28] Voir l'exercice 26.5.

[29] Il faut étudier un polynôme de degré 2.

On a parfois besoin d'étudier la loi d'une somme de deux variables indépendantes, l'outil indispensable est alors la formule des probabilités totales.

Exemple 29.57 Loi triangulaire

Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathbf{P})$, suivant toutes deux la loi uniforme sur $[\![1, n]\!]$.
Soit alors $Z = X + Y$. On a évidemment $Z(\Omega) = [\![2, 2n]\!]$.
Et pour $k \in [\![2, 2n]\!]$, par la formule des probabilités totales, appliquée au système complet d'événements $\{[X = i],\ 1 \leqslant i \leqslant n\}$, il vient

$$\mathbf{P}(Z = k) = \sum_{i=1}^{n} \mathbf{P}([Z = k] \cap [X = i]) = \sum_{i=1}^{n} \mathbf{P}([X = i] \cap [Y = k - i]).$$

Et alors par indépendance de X et Y, on a

$$\mathbf{P}(Z = k) = \sum_{i=1}^{n} \mathbf{P}(X = i)\mathbf{P}(Y = k-i) = \sum_{i=1}^{n} \mathbf{P}(X = i)\mathbf{P}(Y = k-i) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{P}(Y = k-i).$$

▶ <u>**Si** $k \leqslant n$.</u> Alors pour $i \geqslant k$, on a $k - i \leqslant 0$ et donc $\mathbf{P}(Y = k - i) = 0$.
Il reste donc

$$\mathbf{P}(X + Y = k) = \sum_{i=1}^{k-1} \mathbf{P}(X = i)\mathbf{P}(Y = k - i) = \sum_{i=1}^{k-1} \frac{1}{n}\frac{1}{n} = \frac{k - 1}{n^2}.$$

► Si $k \geqslant n+1$, alors pour $i < k-n$, on a $k-i > n$ et donc $\mathbf{P}(Y = k-i) = 0$.
Et de même, pour $i > n$, $\mathbf{P}(X = i) = 0$. Donc il reste

$$\mathbf{P}(X+Y = k) = \sum_{i=k-n}^{n} \mathbf{P}(X = i)\mathbf{P}(Y = k-i) = \sum_{i=k-n}^{n} \frac{1}{n}\frac{1}{n} = \frac{n-(k-n)+1}{n^2} = \frac{2n-k+1}{n^2}.$$

Ainsi, la loi de $X+Y$ est donnée par

$$\mathbf{P}(X+Y = k) = \begin{cases} \frac{k-1}{n^2} & \text{si } 2 \leqslant k \leqslant n \\ \frac{2n-k+1}{n^2} & \text{si } n+1 \leqslant k \leqslant 2n \end{cases}$$

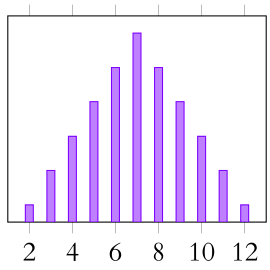

FIGURE 29.1– La loi de Z lorsque $n = 6$.
(c'est la somme de deux dés).

La méthode ci-dessus peut s'adapter à toute variable aléatoire de la forme $Z = g(X, Y)$.

29.5.2 Indépendance de n variables aléatoires

Définition 29.58 – Des variables aléatoires $X_1, \dots, X_n$ définies sur $(\Omega, \mathbf{P})$ sont dites **mutuellement indépendantes** si $\forall A_1 \in \mathscr{P}(X_1(\Omega)), \dots, \forall A_n \in \mathscr{P}(X_n(\Omega))$,

$$\mathbf{P}([X_1 \in A_1] \cap \cdots \cap [X_n \in A_n]) = \prod_{i=1}^{n} \mathbf{P}(X_i \in A_i).$$

Remarque. En fixant certains A_i à $X_i(\Omega)$, on constate que si $X_1, \dots, X_n$ sont mutuellement indépendantes, alors toute sous-famille l'est aussi.
En particulier, les X_i sont deux à deux indépendantes[30].
En revanche, la réciproque est fausse, on peut construire un triplet de variables aléatoires deux à deux indépendantes qui ne sont pas mutuellement indépendantes.

[30] C'est-à-dire que pour $i \neq j$, X_i et X_j sont indépendantes.

Déjà fait ?
Nous avions donné un exemple de trois événements deux à deux indépendants, mais non mutuellement indépendants.
Considérer alors les indicatrices de ces événements.

Proposition 29.59 : *Soient $X_1, \dots, X_n$ des variables aléatoires sur $(\Omega, \mathbf{P})$. Alors $X_1, \dots, X_n$ sont mutuellement indépendantes si et seulement si*

$$\forall x_1 \in X_1(\Omega), \dots, \forall x_n \in X_n(\Omega),\ \mathbf{P}([X_1 = x_1] \cap \cdots \cap [X_n = x_n]) = \prod_{i=1}^{n} \mathbf{P}(X_i = x_i).$$

Démonstration. C'est essentiellement la même que la proposition 29.53, avec n sommes au lieu de deux. □

29.5.3 Le lemme des coalitions

Proposition 29.60 : *Soient $X_1, \dots, X_n$ des variables aléatoires mutuellement indépendantes sur $(\Omega, \mathbf{P})$, et soient $f_1, \dots, f_n$ des applications telles que $\forall i \in [\![1, n]\!]$, f_i soit définie sur $X_i(\Omega)$ (les espaces d'arrivée pouvant être ou ne pas être les mêmes).*
Alors les variables $f_1(X_1), \dots, f_n(X_n)$ sont mutuellement indépendantes.

Autrement dit
Appliquer une fonction (n'importe laquelle : le carré, la valeur absolue, etc) à chacune des variables aléatoires préserve l'indépendance mutuelle.

Démonstration. Pour $i \in [\![1, n]\!]$, et pour A_i partie de $f_i(X_i(\Omega))$, on a

$$[f_i(X_i) \in A_i] = \left[X_i \in f_i^{-1}(A_i)\right].$$

Et donc

$$\begin{aligned}\mathbf{P}\left([f_1(X_1) \in A_1] \cap \cdots \cap [f_n(X_n) \in A_n]\right) &= \mathbf{P}\left([X_1 \in f_1^{-1}(A_1)] \cap \cdots \cap [X_n \in f_n^{-1}(A_n)]\right) \\ &= \prod_{i=1}^{n} \mathbf{P}\left(X_i \in f_i^{-1}(A_i)\right) \\ &= \prod_{i=1}^{n} \mathbf{P}(f_i(X_i) \in A_i).\end{aligned}$$

Indépendance des X_i.

□

Le résultat qui suit, bien que très utile est assez indigeste ne serait-ce qu'à énoncer[31]. Pourtant l'intuition qui se cache derrière est très simple, et vous n'aurez jamais besoin d'énoncer rigoureusement ce résultat.

[31] Et encore, l'énoncé que je donne n'est pas complètement rigoureux.

Proposition 29.61 (Lemme des coalitions) : *Soient $X_1, \dots, X_n$ des variables aléatoires mutuellement indépendantes sur $(\Omega, \mathbf{P})$.*
Soit alors $I_1, \dots, I_p$ une partition de $[\![1, n]\!]$, et pour tout $k \in [\![1, p]\!]$, soit Y_k une variable aléatoire qui est fonction[32] des $X_j, j \in I_k$.
Alors $Y_1, \dots, Y_p$ sont mutuellement indépendantes.

[32] Et c'est là que mon énoncé est flou. Une version plus rigoureuse se trouve dans la preuve, mais je préfère que vous le reteniez sous cette forme imprécise !

Intuitivement, ceci signifie que si on regroupe nos variables aléatoires en p «paquets» disjoints, et que pour chaque paquet on crée une nouvelle variable aléatoire ne dépendant que des X_i dans le paquet, alors les p variables ainsi obtenues sont indépendantes.

Démonstration. Renumérotons les variables de chaque «paquet» et notons

$$\{X_{1,1}, X_{1,2}, \dots, X_{1,i_1}\} = \{X_j, j \in I_1\}, \dots, \{X_{p,1}, \dots, X_{p,i_p}\} = \{X_j,\ j \in I_p\}.$$

Alors pour tout $k \in [\![1, p]\!]$, il existe une fonction f_k telle que $Y_k = f_k(X_{k,1}, \dots, X_{k,i_k})$.
Pour une fois, il est intéressant de voir ici un k-uplet de variables aléatoires comme une variable aléatoire à valeurs dans un produit cartésien.
On prouve alors que les **variables aléatoires** $(X_{1,1}, \dots, X_{1,i_1}), \dots, (X_{p,1}, \dots, X_{p,i_p})$ sont mutuellement indépendantes.
Prouvons-le pour $p = 2$, mais le principe est le même dans le cas général : soient $x_{1,1}, \dots, x_{1,i_1}, x_{2,1}, \dots, x_{2,i_2} \in X_{1,1}(\Omega) \times \cdots \times X_{2,i_2}(\Omega)$. Alors

$$\begin{aligned}
&\mathbf{P}\left([(X_{1,1}, \dots, X_{1,i_1}) = (x_{1,1}, \dots, x_{1,i_1})] \cap [(X_{2,1}, \dots, X_{2,i_2}) = (x_{2,1}, \dots, x_{2,i_2})]\right)\\
=&\mathbf{P}([X_{1,1} = x_{1,1}] \cap \cdots \cap [X_{2,i_2} = x_{2,i_2}])\\
=&\prod_{j=1}^{i_1} \mathbf{P}(X_{1,j} = x_{1,j}) \times \prod_{j=1}^{i_2} \mathbf{P}(X_{2,j} = x_{2,j})\\
=&\mathbf{P}((X_{1,1}, \dots, X_{1,i_1}) = (x_{1,1}, \dots, x_{1,i_1}))\mathbf{P}((X_{2,1}, \dots, X_{2,i_2}) = (x_{2,1}, \dots, x_{2,i_2})).
\end{aligned}$$

Indépendance.

C'est l'indépendance mutuelle de $X_{1,1}, \dots, X_{1,i_1}$.

Et alors il suffit d'appliquer la proposition précédente avec les fonctions f_k. □

Exemple 29.62

Si $(X_1, X_2, X_3, X_4, X_5)$ sont mutuellement indépendantes, alors $(X_3, X_1 + X_5, X_2^2 e^{X_4})$ sont encore mutuellement indépendantes.
En revanche, on ne peut rien dire[33] de l'indépendance éventuelle de $(X_3, X_1 + X_3, X_1 X_2)$, puisque X_3 «figure» dans deux des trois variables, de même que X_1 (les «paquets» ne sont pas disjoints).

[33] Le lemme des coalitions n'est pas un «si et seulement si» !

29.5.4 Un résultat de stabilité

Rappelons une formule qui ne figure pas explicitement au programme, mais que nous avons déjà prouvée deux fois (une avec des polynômes dans le TD17, une autre par dénombrement dans le TD24) : l'identité de Vandermonde[34] :

[34] C'est le même Vandermonde que pour le déterminant.

$$\forall (m, r, n) \in \mathbf{N}^3,\ \sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k} = \binom{n+m}{r}.$$

Proposition 29.63 : *Soient $n_1, n_2 \in \mathbf{N}^*$, soit $p \in]0, 1[$ et soient X_1, X_2 deux variables aléatoires **indépendantes** sur un même espace probabilisé $(\Omega, \mathbf{P})$ telles que $X_1 \sim \mathcal{B}(n_1, p)$ et $X_2 \sim \mathcal{B}(n_2, p)$.*
Alors $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$.

⚠ Attention !
Le second paramètre p doit être le même pour les deux lois.

Démonstration. Puisque X_1 prend des valeurs entre 0 et n_1, et que X_2 prend des valeurs entre 0 et n_2, $X_1 + X_2$ est à valeurs dans $[\![0, n_1 + n_2]\!]$.
Soit $\ell \in [\![0, n_1 + n_2]\!]$. Appliquons alors la formule des probabilités totales avec le système complet d'événements $\{[X_1 = k],\ 0 \leqslant k \leqslant n_1\}$,

$$\begin{aligned}
\mathbf{P}(X_1 + X_2 = \ell) &= \sum_{k=0}^{n_1} \mathbf{P}(X_1 = k)\mathbf{P}(X_2 = \ell - k) \\
&= \sum_{k=0}^{n_1} \binom{n_1}{k} p^k (1-p)^{n_1-k} \binom{n_2}{\ell - k} p^{\ell-k} (1-p)^{n_2-(\ell-k)} \\
&= p^\ell (1-p)^{n_1+n_2-\ell} \sum_{k=0}^{n_1} \binom{n_1}{k}\binom{n_2}{\ell-k} \\
&= p^\ell (1-p)^{n_1+n_2-\ell} \binom{n_1+n_2}{\ell}.
\end{aligned}$$

Identité de Vandermonde.

Donc $X_1 + X_2 \sim \mathscr{B}(n_1 + n_2, p)$. □

Ce résultat est somme toutes assez intuitif : si on se souvient qu'une loi binomiale compte le nombre de succès lors de répétitions indépendantes d'épreuves de Bernoulli, la somme $X_1 + X_2$ est le nombre de succès lors d'une première série de n_1 répétitions indépendantes et du nombre de succès lors d'une seconde série de n_2 répétitions indépendantes, et indépendantes des précédentes[35]
C'est donc en tout le nombre de succès lors de $n_1 + n_2$ répétitions de la même épreuve de Bernoulli de paramètre p, qui suit une loi binomiale $\mathscr{B}(n_1 + n_2, p)$.

[35] C'est l'hypothèse d'indépendance de X_1 et X_2.

Corollaire 29.64 – *Soit $p \in [0,1]$, et soient $X_1, \dots, X_k$ des variables aléatoires sur $(\Omega, \mathbf{P})$, mutuellement indépendantes telles que $X_i \sim \mathscr{B}(n_i, p)$. Alors*

$$\sum_{i=1}^{k} X_i \sim \mathscr{B}\left(\sum_{i=1}^{k} n_i, p\right).$$

Démonstration. Pour $k = 2$, c'est la proposition précédente.
Procédons par récurrence sur n, et supposons le résultat vrai pour k variables aléatoires indépendantes, et soient $X_1, \dots, X_{k+1}$ $k+1$ variables aléatoires indépendantes sur $(\Omega, \mathbf{P})$ telles que $X_i \sim \mathscr{B}(n_i, p)$.
Par le lemme des coalitions, $X_1 + \dots + X_k$ est indépendante de X_{k+1}. De plus, par hypothèse de récurrence,

$$X_1 + \dots + X_k \sim \mathscr{B}\left(\sum_{i=1}^{k} n_i, p\right).$$

Ainsi,

$$(X_1 + \dots + X_k) + X_{k+1} \sim \mathscr{B}\left(\sum_{i=1}^{k} n_i + n_{k+1}, p\right) = \mathscr{B}\left(\sum_{i=1}^{k+1} n_i, p\right).$$

Par le principe de récurrence, la propriété est donc vraie pour tout k. □

Corollaire 29.65 – *En particulier, si $X_1, \dots, X_n$ sont n variables aléatoires indépendantes suivant une loi de Bernoulli de paramètre p, alors $\sum_{i=1}^{n} X_i$ suit la loi binomiale $\mathscr{B}(n, p)$.*

Démonstration. Il suffit de remarquer que la loi de Bernoulli de paramètre p est également la loi binomiale de paramètres 1 et p. □

Remarques. ► Cela permet de justifier un principe connu depuis longtemps : une loi binomiale $\mathscr{B}(n, p)$ correspond au nombre de succès lors d'une répétition de n épreuves de Bernoulli indépendantes de même paramètre p.
► On retrouve ainsi sans calculs le fait que l'espérance d'une loi binomiale est np.

Variance
En réalité, on retrouve aussi la variance, si on sait que la variance d'une somme de variables indépendantes est la somme des variances, cf le paragraphe suivant.

29.6 Covariance

Dans cette partie, on construit une quantité qui en quelque sorte mesure le défaut d'indépendance de deux variables aléatoires.

29.6.1 Espérance d'un produit

Proposition 29.66 : *Soient X et Y deux variables aléatoires réelles sur un espace probabilisé $(\Omega, \mathbf{P})$. Alors*

$$\mathbf{E}(XY) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbf{P}([X=x]\cap[Y=y]).$$

Démonstration. C'est le théorème de transfert appliqué à la variable[36] (X,Y) et à la fonction $(x,y)\mapsto xy$. □

[36] À valeurs dans $\mathbf{R}^2$

Proposition 29.67 : *Si X et Y sont deux variables aléatoires indépendantes sur $(\Omega, \mathbf{P})$, alors $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.*

Démonstration. Par la proposition précédente, on a

$$\begin{aligned}\mathbf{E}(XY) &= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbf{P}([x=x]\cap[Y=y]) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbf{P}(X=x)\mathbf{P}(Y=y)\\ &= \sum_{x\in X(\Omega)}\sum_{y\in Y(\Omega)} xy\mathbf{P}(X=x)\mathbf{P}(Y=y) = \left(\sum_{x\in X(\Omega)} x\mathbf{P}(X=x)\right)\left(\sum_{y\in Y(\Omega)} y\mathbf{P}(Y=y)\right)\\ &= \mathbf{E}(X)\mathbf{E}(Y).\end{aligned}$$

Indépendance de X et Y.

□

⚠ Sans l'hypothèse d'indépendance, cette propriété est absolument fausse.
Par exemple, on n'a pas en général[37] $\mathbf{E}\left(X^2\right) = \mathbf{E}(XX) = \mathbf{E}(X)\mathbf{E}(X)$.

[37] Sauf si X suit une loi certaine.

Corollaire 29.68 – *Si $X_1, \dots, X_n$ sont des variables aléatoires réelles indépendantes sur $(\Omega, \mathbf{P})$, alors $\mathbf{E}\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n \mathbf{E}(X_i)$.*

Démonstration. La preuve se fait par récurrence sur n. Pour $n=2$, c'est la proposition précédente.
Supposons la formule vraie pour un produit de n variables aléatoires indépendantes, et soient $X_1, \dots, X_{n+1}$ des variables aléatoires indépendantes. Alors par le lemme des coalitions, $X_1\times X_2\times\cdots\times X_n$ est indépendante de X_{n+1}. Et alors

$$\mathbf{E}\left((X_1\cdots X_n)X_{n+1}\right) = \mathbf{E}\left(X_1\cdots X_n\right)\mathbf{E}(X_{n+1}) = \mathbf{E}(X_1)\mathbf{E}(X_2)\cdots\mathbf{E}(X_n)\mathbf{E}(X_{n+1}).$$

Donc par le principe de récurrence, la propriété est vraie pour tout n. □

29.6.2 Covariance

Définition 29.69 – Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathbf{P})$. La covariance de X et Y est alors définie par

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

Remarque

Notons que la covariance n'est définie que pour deux variables définies sur le même espace probabilisé.

Remarques. ► Il est possible d'interpréter le signe de la covariance. En effet, s'il est positif, cela signifie qu'en moyenne, $X - \mathbf{E}(X)$ et $Y - \mathbf{E}(Y)$ sont de même signe, donc que lorsque X augmente, en moyenne Y augmente aussi.
En revanche s'il est négatif, cela signifie qu'en moyenne, $X - \mathbf{E}(X)$ et $Y - \mathbf{E}(Y)$ sont de signes opposés, donc que lorsque X augmente, en moyenne Y diminue.
► Notons tout de suite que $\mathrm{Cov}(X, X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right) = \mathbf{V}(X)$.

Proposition 29.70 (Formule de Huygens) : *Si X et Y sont deux variables aléatoires réelles sur $(\Omega, \mathbf{P})$, alors*

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

Remarque
Pour $X = Y$, on retrouve

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

Démonstration.

$$(X - \mathbf{E}(X))(Y - \mathbf{E}(Y)) = XY - X\underbrace{\mathbf{E}(Y)}_{\in\mathbf{R}} - Y\underbrace{\mathbf{E}(X)}_{\in\mathbf{R}} + \mathbf{E}(X)\mathbf{E}(Y).$$

Donc par linéarité de l'espérance, on a

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

□

Proposition 29.71 : *Soient X, Y, Z trois variables aléatoires réelles sur $(\Omega, \mathbf{P})$. Alors*

1. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
2. $\forall \lambda \in \mathbf{R}, \mathrm{Cov}(\lambda X + Y, Z) = \lambda \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$
3. $\forall \lambda \in \mathbf{R}, \mathrm{Cov}(X, \lambda Y + Z) = \lambda \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$
4. $\mathrm{Cov}(X, X) = V(X) \geqslant 0$.

Terminologie
Les points 2 et 3 signifient que la covariance est une **forme bilinéaire** sur le $\mathbf{R}$-espace vectoriel des variables aléatoires réelles sur Ω.
Le premier point s'appelle la **symétrie**, et on dit alors que la covariance est une forme bilinéaire symétrique.
Nous rencontrerons de nouveau ce vocabulaire dans les deux prochains chapitres d'algèbre.

Démonstration. 1. Évident.

2. Soit $\lambda \in \mathbf{R}$. Alors

$$\begin{aligned}\mathrm{Cov}(\lambda X + Y, Z) &= \mathbf{E}((\lambda X + Y)Z) - \mathbf{E}(\lambda X + Y)\mathbf{E}(Z)\\ &= \lambda\mathbf{E}(XZ) + \mathbf{E}(YZ) - (\lambda\mathbf{E}(X) + \mathbf{E}(Y))\,\mathbf{E}(Z)\\ &= \lambda\,(\mathbf{E}(XZ) - \mathbf{E}(X)\mathbf{E}(Z)) + \mathbf{E}(YZ) - \mathbf{E}(Y)\mathbf{E}(Z)\\ &= \lambda\,\mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z).\end{aligned}$$

3. Pour $\lambda \in \mathbf{R}$, on a, en utilisant les deux premières propriétés

$$\mathrm{Cov}(X, \lambda Y+Z) = \mathrm{Cov}(\lambda Y+Z, X) = \lambda\,\mathrm{Cov}(Y, X)+\mathrm{Cov}(Z, X) = \lambda\,\mathrm{Cov}(X, Y)+\mathrm{Cov}(X, Z).$$

4. $\mathrm{Cov}(X, X) = \mathbf{E}\left((X - \mathbf{E}(X))^2\right) = \mathbf{V}(X) \geqslant 0$.

□

Corollaire 29.72 – *Si X et Y sont deux variables aléatoires réelles indépendantes, alors* $\mathrm{Cov}(X, Y) = 0$.

Danger !
La réciproque est **FAUSSE** !
Deux variables aléatoires peuvent avoir une covariance nulle sans être indépendantes.

Démonstration. Puisque X et Y sont indépendantes, alors $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$, et alors par la formule de Huygens, $\mathrm{Cov}(X, Y) = \mathbf{E}(X)\mathbf{E}(Y) - \mathbf{E}(X)\mathbf{E}(Y) = 0$. □

Remarques. ► Si $\mathrm{Cov}(X, Y) = 0$, on dit que X et Y sont **non corrélées**. Ainsi, deux variables indépendantes sont non corrélées, mais la réciproque est fausse.
D'ailleurs un exercice où on vous fait calculer une covariance, qu'elle est nulle, et que la question suivante est «les variables sont elles indépendantes ?» ressemble franchement à un piège dans lequel on attend de voir si vous tombez. Réfléchissez bien avant de répondre, et si malgré tout les deux variables sont indépendantes, ce n'est pas la nullité de la covariance

qui vous permettra de l'affirmer.
▸ En revanche, une covariance non nulle clôt la discussion : les variables ne sont pas indépendantes.

Exemple 29.73

Soit X une variable aléatoire telle que[38]

$$X(\Omega) = \{-1, 1\} \text{ et } \mathbf{P}(X = 1) = \mathbf{P}(X = -1) = \frac{1}{2}.$$

Soit Y une variable aléatoire sur le même espace qui suit la loi de Bernoulli de paramètre p, avec $p \in]0, 1[$.
On suppose que X et Y sont indépendantes, et on note $Z = XY$.
Par indépendance de X et Y, on a $\mathbf{E}(Z) = \mathbf{E}(X)\mathbf{E}(Y) = 0$ car $\mathbf{E}(X) = 0$.
D'autre part, on a $\mathbf{E}(ZY) = \mathbf{E}\left(XY^2\right) = \mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) = 0$.
Et donc, $\mathrm{Cov}(Z, Y) = \mathbf{E}(ZY) - \mathbf{E}(Z)\mathbf{E}(Y) = 0$.
Pourtant Z et Y ne sont pas indépendantes, car

$$\mathbf{P}([Z = 0] \cap [Y = 1]) = 0 \text{ et } \mathbf{P}(Z = 0)\mathbf{P}(Y = 1) = \mathbf{P}(Y = 0)\mathbf{P}(Y = 1) = p(1 - p) \neq 0.$$

[38] Une telle variable est parfois appelée variable de Rademacher

Détails
Puisque Y ne prend que les valeurs 0 et 1, $Y^2 = Y$.

29.6.3 Variance d'une somme

Proposition 29.74 : *Soient X et Y deux variables aléatoires réelles sur $(\Omega, \mathbf{P})$. Alors*

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathrm{Cov}(X, Y).$$

Démonstration. Par bilinéarité de la covariance,

$$\begin{aligned}\mathbf{V}(X + Y) &= \mathrm{Cov}(X + Y, X + Y) = \mathrm{Cov}(X, X + Y) + \mathrm{Cov}(Y, X + Y)\\ &= \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, X) + \mathrm{Cov}(Y, Y)\\ &= \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathrm{Cov}(X, Y).\end{aligned}$$

Linéarité à gauche.

Par symétrie, $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

□

Corollaire 29.75 – *Si X et Y sont indépendantes, alors*

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y).$$

Écart-type
Il n'y a pas de formule analogue pour l'écart-type.
En particulier, on n'a jamais (ou presque jamais) $\sigma_{X+Y} = \sigma_X + \sigma_Y$.

Proposition 29.76 : *Plus généralement, si $X_1, \dots, X_n$ sont n variables aléatoires réelles sur $(\Omega, \mathbf{P})$, alors*

$$\mathbf{V}(X_1 + X_2 + \dots + X_n) = \sum_{i=1}^{n} \mathbf{V}(X_i) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j).$$

En particulier, si $X_1, X_2, \dots, X_n$ sont deux à deux indépendantes[39]

$$\mathbf{V}(X_1 + \dots + X_n) = \mathbf{V}(X_1) + \mathbf{V}(X_2) + \dots + \mathbf{V}(X_n).$$

[39] Et donc aussi si elles sont mutuellement indépendantes.

Démonstration. C'est encore de la bilinéarité couplée à la symétrie :

$$\begin{aligned}\mathbf{V}\left(\sum_{i=1}^{n} X_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i\right)\\ &= \sum_{i=1}^{n} \mathrm{Cov}\left(X_i, \sum_{j=1}^{n} X_j\right)\end{aligned}$$

Rédaction ✎
La deuxième somme a subtilement changé d'indice, et c'était fondamental ici puisque la première somme porte déjà sur i.

$$= \sum_{i=1}^n \sum_{j=1}^n \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^n \mathrm{Cov}(X_i, X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \mathrm{Cov}(X_i, X_j)$$

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

Par symétrie de la covariance, chaque terme apparaissait deux fois dans la somme.

□

Comme mentionné plus haut, dans le cas où les X_i sont mutuellement indépendantes et suivent la même loi $\mathscr{B}(p)$, on retrouve la variance d'une loi binomiale $\mathscr{B}(n, p)$.

29.6.4 Coefficient de corrélation linéaire

Cette partie est hors programme en MPSI, mais également en MP. En revanche elle figure au programme de PSI (dans un cadre plus général que celui abordé ci-dessous).
J'en parle tout de même car il s'agit d'une bonne illustration de l'intérêt de la covariance, mais aussi car dans le prochain chapitre d'algèbre, nous prouverons une célèbre inégalité[40] par une méthode en tous points analogue à celle qui suit. Autant la comprendre dès maintenant.

[40] L'inégalité de Cauchy-Schwarz.

Définition 29.77 – Soient X, Y deux variables aléatoires réelles sur $(\Omega, \mathbf{P})$, de variance non nulle[41]. Le **coefficient de corrélation linéaire de** X **et** Y est

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(Y)}}.$$

[41] Donc ne suivant pas une loi certaine.

Proposition 29.78 : *Sous les hypothèses ci-dessus :*

$$|\rho(X, Y)| \leqslant 1.$$

De plus, $\rho(X, Y) = \pm 1$ si et seulement si il existe un couple (a, b) de réels, avec $a \neq 0$, tels que $\mathbf{P}(Y = aX + b) = 1$. Dans ce cas

- *si $\rho(X, Y) = 1$, alors $a > 0$: X et Y varient dans le même sens.*
- *si $\rho(X, Y) = -1$, alors $a < 0$: X et Y varient dans des sens opposés.*

Démonstration. Considérons la fonction $f(t) = \mathbf{V}(tX - Y)$.
Alors $\forall t \in \mathbf{R}$, $f(t) \geqslant 0$ car une variance est toujours positive.
De plus, $f(t) = \mathbf{V}(tX) + \mathbf{V}(Y) - 2\,\mathrm{Cov}(tX, Y) = t^2\mathbf{V}(X) + \mathbf{V}(Y) - 2t\,\mathrm{Cov}(X, Y)$.
Ainsi, f est une fonction polynôme du second degré, et puisqu'elle est toujours positive, c'est qu'elle ne possède qu'au plus une racine, et son discriminant est donc négatif ou nul.
Ainsi, $4\,\mathrm{Cov}(X, Y)^2 - 4\mathbf{V}(X)\mathbf{V}(Y) \leqslant 0$.
On en déduit que $\mathrm{Cov}(X, Y)^2 \leqslant \mathbf{V}(X)\mathbf{V}(Y)$ et donc

$$|\,\mathrm{Cov}(X, Y)| \leqslant \sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(Y)} \Leftrightarrow \left|\frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(Y)}}\right| \leqslant 1.$$

De plus, on a l'égalité si et seulement si le discriminant est nul, c'est-à-dire si et seulement si f s'annule en un certain réel a.
Mais alors $\mathbf{V}(aX - Y) = 0$, et donc $aX - Y$ suit une loi certaine :

$$\exists b \in \mathbf{R} : \mathbf{P}(aX - Y = -b) = 1 \Leftrightarrow \mathbf{P}(Y = aX + b) = 1.$$

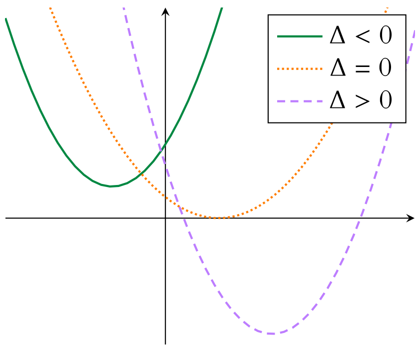

Figure 29.2– Un polynôme de degré 2 de signe constant est de discriminant négatif.

Notons que a est alors non nul car le polynôme f a un terme constant non nul, donc ne s'annule pas en 0.

Dans ce cas, on a alors $\mathbf{V}(Y) = \mathbf{V}(aX+b) = a^2\mathbf{V}(X)$ et donc $\sqrt{\mathbf{V}(Y)} = |a|\sqrt{\mathbf{V}(X)}$.
De même, on a

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(X, aX+b) = a\,\mathrm{Cov}(X,X) + \mathrm{Cov}(X,b) = a\mathbf{V}(X) + \mathrm{Cov}(X,b).$$

Mais, $\mathbf{V}(b) = 0$ et donc $|\mathrm{Cov}(X,b)| \leqslant \sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(b)} = 0$.
Ainsi $\mathrm{Cov}(X,Y) = a\mathbf{V}(X)$. On en déduit que

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbf{V}(X)}\sqrt{\mathbf{V}(Y)}} = \frac{a}{|a|}.$$

Valeur absolue
$\frac{a}{|a|}$ vaut 1 si $a > 0$ et -1 si $a < 0$.

Et donc $\rho(X,Y)$ est du signe de a. □

Le coefficient de corrélation linéaire est donc en quelques sortes une mesure de l'indépendance de deux variables aléatoires. En effet, pour des variables indépendantes il est nul. Et dans le cas où l'une des variables est fonction affine de l'autre, c'est-à-dire à peu près ce qu'on peut imaginer de moins indépendant, il prend les plus grandes valeurs possibles (± 1).

C'est un indicateur très utilisé en pratique, en physique, en biologie, en économie, etc, bref, partout où on a besoin de statistiques.
Notons que comme son nom l'indique, il ne détecte que les corrélations **linéaires**. Si $Y = f(X)$, où f n'est pas linéaire, X et Y ne sont généralement pas indépendantes, mais la valeur de $\rho(X,Y)$ n'a alors pas beaucoup de signification.

Exemple 29.79

Considérons une variable U qui suit la loi uniforme sur $[\![-n,n]\!]$, et soit $V = U^2$.
Alors $\mathbf{E}(UV) = \mathbf{E}\left(U^3\right) = \sum_{k=-n}^{n} k^3 \frac{1}{2n+1}$.
Mais cette somme est clairement nulle par un argument d'imparité[42].
Et donc $\mathrm{Cov}(U,V) = \mathbf{E}(UV) - \mathbf{E}(U)\mathbf{E}(V) = 0$.
Pour autant, n'allons pas affirmer que V «ne dépend pas» de U.

[42] La couper en deux en 0 si vous n'êtes pas convaincu.

29.6.5 Vers la loi faible des grands nombres (hors programme en sup)

La partie qui suit ne figure pas au programme de sup, car un énoncé correct nécessiterait un espace probabilisé infini, ou en tous cas serait quasiment vidé de sa substance[43] sur un espace fini.

[43] Il ne s'appliquerait qu'à des variables certaines...

Considérons donc $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes, qui suivent toute la même loi, et notons $S_n = \sum_{i=1}^n X_i$. Notons également μ l'espérance des X_i et σ^2 leur variance.
Alors par linéarité de l'espérance $\mathbf{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}(X_i) = \frac{1}{n}n\mu = \mu$. On a également

$$\begin{aligned}\mathbf{V}\left(\frac{S_n}{n}\right) &= \frac{1}{n^2}\mathbf{V}(X_1+\cdots+X_n)\\ &= \frac{1}{n^2}\left(\mathbf{V}(X_1)+\mathbf{V}(X_2)+\cdots+\mathbf{V}(X_n)\right)\\ &= \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}.\end{aligned}$$

Par l'inégalité de Bienaymé-Tchebychev appliquée à $\frac{S_n}{n}$, il vient alors, pour tout $\varepsilon > 0$:

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}.$$

Ce n'est pas tant l'inégalité qui est remarquable[44], mais le fait que le majorant tend vers 0 lorsque n tend vers $+\infty$.

[44] La majoration est trop générale pour être bonne pour toutes les lois.

Et c'est cela qu'on appelle la loi faible des grands nombres.

Là où le programme de sup montre ses limites, c'est justement en passant à la limite, car sur un espace fini, il ne sera pas possible de construire une infinité de variables aléatoires indépendantes et de même loi, sauf à considérer des variables certaines. Mais malgré tout, il serait envisageable de construire un espace fini sur lequel il y a un nombre n arbitrairement grand (mais fixé) de variables indépendantes et de même loi. Et alors la majoration ci-dessus vaut toujours, avec un majorant qui est de plus proche de 0 pour n grand.

Très bien, mais tout ceci ne nous dit pas pourquoi cette limite est intéressante...
Cela signifie tout simplement que l'espérance, telle que nous l'avons définie, est bien ce que vous pensez.
En effet, si on répète un grand nombre n de fois une même expérience, qu'on note $X_1, \dots, X_n$ les résultats de l'expérience considérée, et qu'on calcule la moyenne de ces résultats (c'est $\frac{S_n}{n}$), alors la probabilité que cette moyenne soit à distance plus de $\varepsilon > 0$ de $\mu = \mathbf{E}(X_1)$ tend vers 0 lorsque n tend vers $+\infty$.
Donc en résumé, que si on répète un grand nombre de fois l'expérience, la moyenne des résultats va être proche de l'espérance.

La formulation avec ce ε peut sembler laborieuse, mais pourtant on ne peut guère faire mieux : la probabilité que $\dfrac{S_n}{n}$ soit à distance plus de ε de μ a beau tendre vers 0 elle n'est nulle.
Vous savez par exemple que si on joue 10 000 fois la même grille au loto, la probabilité de gagner à chaque fois le gros lot est très[45] faible mais elle n'est pas nulle.
Et alors le gain moyen sur ces 10 000 parties est assez éloigné du gain moyen que peut espérer un joueur de loto, mais la probabilité que ceci se produise est assez faible.

[45] très, très, très, très

En particulier, si les X_i suivent une loi de Bernoulli de paramètre p, et qu'on voit chaque variable de Bernoulli comme la variable indiquant le résultat (échec ou succès) d'une expérience de Bernoulli, alors $\frac{S_n}{n}$ est le nombre moyen de succès lors des n répétitions.
Et donc il est «proche» quand n est grand, de $\mathbf{E}(X_1) = p$.
Le fait de disposer de ce résultat est somme toute assez rassurant, et signifie que tout le formalisme avec lequel nous travaillons depuis le début nous permet bien de modéliser l'intuition qu'on se fait d'une probabilité.

Savoir si «n grand» signifie $n = 100$, $n = 1\,000$ ou $n = 100\,000\,000$ dépasse le cadre des probabilités telles qu'elles sont enseignées en prépa scientifique, mais est tout de même une question intéressante lorsqu'on étudie les statistiques, et vous avez effleuré la question en terminale en parlant d'intervalles de confiance.
Pour un sondeur politique, $n = 1000$ est suffisamment grand. Mais les marges d'erreur sont alors conséquentes[46].
Dans d'autres domaines où le droit à l'erreur n'existe pas, 1 000 est probablement trop petit.
La majoration de $p = \mathbf{P}\left(\left|\dfrac{S_n}{n}\right| \geqslant \varepsilon\right)$ obtenue ci-dessus permet en réalité de trouver, à ε fixé, quelle valeur de n permet de rendre p suffisamment petit, et ça donne parfois lieu à des exercices de concours[47].
Mais en pratique, la majoration donnée par Bienaymé-Tchebychev est trop grossière, et donc les valeurs de n obtenues sont très loin d'être optimales.

[46] Je sais que vous n'étiez pas bien vieux à l'époque, mais avez-vous déjà entendu parler du 21 avril 2002 ?

[47] Par exemple l'exercice 100 de la banque CCP.