Chapitre 11

Probabilités

(version mise à jour le 6 janvier 2020)

Sommaire

11.1 Espaces probabilisés

11.1.1 Ensembles dénombrables

Définition 11.1.1 (ensemble dénombrable)
Un ensemble E est dit *dénombrable* s'il existe une bijection de $\mathbb{N}$ sur E.

Remarques

– Les ensembles suivants sont, de façon évidente, dénombrables :
  - l'intervalle d'entiers $[\![a, \infty[$ pour tout a de $\mathbb{Z}$ (penser à l'application $n \mapsto a + n$)
  - l'ensemble des entiers pairs (avec l'application $n \mapsto 2n$)
  - l'ensemble celui des entiers impairs (avec l'application $n \mapsto 2n + 1$).

– Soit E un ensemble dénombrable.
Un ensemble F est en bijection avec E si et seulement si il est lui-même dénombrable.

– Un ensemble dénombrable est infini.
La réciproque est fausse : l'ensemble $\mathbb{R}$ est infini mais n'est pas dénombrable (voir plus loin).

Proposition 11.1.1 (les parties de $\mathbb{N}$ sont finies ou dénombrables)
Toute partie de $\mathbb{N}$ est soit finie, soit dénombrable.
Plus généralement toute partie d'un ensemble dénombrable est soit finie, soit dénombrable.

– On rappelle que toute partie non vide de $\mathbb{N}$ possède un élément minimum.
Soit A une partie infinie de $\mathbb{N}$. Montrons que A est dénombrable.
On forme $f : \mathbb{N} \to A$, par récurrence, en posant : $\forall\, n \in \mathbb{N},\ f(n) = \min(A \setminus \{f(0), \ldots, f(n-1)\})$.
En particulier, $f(0) = \min(A \setminus \emptyset) = \min(A)$. L'application f est à valeurs dans A.
On va montrer que f est une bijection de $\mathbb{N}$ sur A.
L'application f est injective : en effet, si $m > n$, alors $f(m) = \min(A \setminus \{f(0), \ldots, f(m-1)\})$ donc m n'est pas dans $\{f(0), \ldots, f(m-1)\}$, et a fortiori pas dans $\{f(0), \ldots, f(n)\}$, et a fortiori $f(m) \neq f(n)$.
Ainsi f réalise une bijection de $\mathbb{N}$ sur $f(\mathbb{N}) \subset A$ (en particulier $f(\mathbb{N})$ est infini).
Supposons par l'absurde que $f : \mathbb{N} \to A$ ne soit pas surjective.
Il existe alors un élément a de A qui n'a pas d'antécédent par f.
Pour tout $n \geqslant 0$, l'elément a est donc dans $A \setminus \{f(0), \ldots, f(n-1)\}$, ensemble dont le minimum est $f(n)$.
Ainsi $f(n) \leqslant a$ pour tout n de $\mathbb{N}$, donc $f(\mathbb{N}) \subset [\![0, a]\!]$ (absurde car $f(\mathbb{N})$ est infini).
Conclusion : f est une bijection de $\mathbb{N}$ sur A, donc A est dénombrable.

– Soit E un ensemble dénombrable. Soit F une partie de E.
On sait qu'il existe une bijection f de $\mathbb{N}$ sur E.
La restriction de f^{-1} à F réalise une bijection de F sur une partie A de $\mathbb{N}$.
L'ensemble A, en tant que partie de $\mathbb{N}$, est fini ou dénombrable : il en est donc de même de F. $\square$

Proposition 11.1.2 (caractérisation des ensembles finis ou dénombrables)
Soit E un ensemble. Les deux propositions suivantes sont équivalentes :

- *l'ensemble E est fini ou dénombrable*
- *il existe une application surjective de $\mathbb{N}$ sur E*

Ainsi E est fini ou dénombrable si et seulement si on peut le décrire « en extension » : $E = \{x_n,\ n \in \mathbb{N}\}$.

Remarque : dans la proposition précédente, on peut remplacer $\mathbb{N}$ par un ensemble dénombrable connu.

- Si E est dénombrable il existe une bijection (donc une surjection) de $\mathbb{N}$ sur E.
  Si E est fini de cardinal m, il existe une bijection f de $[\![0, m-1]\!]$ sur E.
  Celle-ci se prolonge (de façon quelconque) en une surjection de $\mathbb{N}$ sur E.
- Réciproquement, on suppose qu'il existe une surjection f de $\mathbb{N}$ sur E.
  Pour tout x de E, on note $g(x)$ le plus petit antécédent de x par f.
  On définit ainsi une application $g : E \to \mathbb{N}$ telle que $f \circ g = \mathrm{Id}_E$ (par construction).
  L'application $f \circ g$ étant injective, il en est de même de g.
  L'application g réalise donc une bijection de E sur une partie de $\mathbb{N}$.
  Celle-ci étant finie ou dénombrable, il en est de même de E. □

Remarque : un ensemble E est fini ou dénombrable si on peut l'écrire $E = \{x_n,\ n \in J\}$, où J est une partie de $\mathbb{N}$ et où les x_n sont distincts deux à deux (ce point de vue sera notamment adopté dans la définition 11.2.4 et les propositions 11.2.2 et 11.2.3).

Proposition 11.1.3 (l'ensemble $\mathbb{Z}$ est dénombrable)
L'ensemble $\mathbb{Z}$ des entiers relatifs est dénombrable.

Pour tout n de $\mathbb{N}$, on pose $\begin{cases} f(2n) = n \\ f(2n+1) = -n-1 \end{cases}$, qu'on peut résumer en : $\forall n \in \mathbb{N},\ f(n) = (-1)^n \left\lceil \dfrac{n}{2} \right\rceil$.
On vérifie facilement que f est une bijection de $\mathbb{N}$ sur $\mathbb{Z}$. □

Proposition 11.1.4 (produit cartésien de deux ensembles dénombrables)
L'ensemble $\mathbb{N} \times \mathbb{N}$ est dénombrable.
Plus généralement, si E et F sont dénombrables, leur produit cartésien $E \times F$ est dénombrable.

- Tout entier n strictement positif s'écrit d'une manière unique $n = 2^p(2q+1)$, avec $(p, q) \in \mathbb{N}^2$.
  En effet, p est l'exposant maximum k tel que $2^k \mid n$, et $2q+1$ est l'entier (nécessairement impair) résultant du quotient exact de n par 2^p.
  L'application $(p, q) \mapsto 2^p(2q+1)$ est donc une bijection de $\mathbb{N}^2$ sur $\mathbb{N}^*$.
  Comme $\mathbb{N}^*$ est dénombrable, il en résulte que $\mathbb{N}^2$ est dénombrable.
- Soit E et F deux ensembles dénombrables.
  Soit $f : \mathbb{N} \to E$ et $g : \mathbb{N} \to F$ deux bijections.
  Alors l'application $h : \mathbb{N}^2 \to E \times F$ définie par $h(m, n) = (f(m), g(n))$ est une bijection.
  Comme on sait que $\mathbb{N}^2$ est dénombrable, il en résulte que l'ensemble $E \times F$ est dénombrable. □

Proposition 11.1.5 (non dénombrabilité de l'ensemble des suites à valeurs dans $\{0,1\}$)
L'ensemble $E=\{0,1\}^{\mathbb{N}}$ de toutes les suites à valeurs dans $\{0,1\}$ n'est pas dénombrable.

Soit f une application quelconque de $\mathbb{N}$ dans E. On va montrer que f n'est pas surjective.
Pour tout k de $\mathbb{N}$, $f(k)$ désigne donc une suite à valeurs dans $\{0,1\}$.
Pour chaque n de $\mathbb{N}$, on note $a_n^{(n)}$ (égal à 0 ou à 1) le terme d'indice n de la suite $f(n)$.
Par exemple, $a_0^{(0)}$ est le terme d'indice 0 de la suite $f(0)$, $a_1^{(1)}$ est le terme d'indice 1 de la suite $f(1)$.
On définit alors une suite $u=(u_n)_{n\geqslant 0}$ de E en posant : $\forall n\in\mathbb{N},\ u_n=1-a_n^{(n)}$.
Pour tout n de $\mathbb{N}$, le terme d'indice n de la suite u est $u_n=1-a_n^{(n)}$ donc est différent de $a_n^{(n)}$ qui, lui, désigne le terme d'indice n de la suite $f(n)$. Les suites u et $f(n)$ sont donc distinctes, pour tout n de $\mathbb{N}$.
La suite u de E n'a donc pas d'antécédent par f, ce qui prouve que f n'est pas surjective.
Ainsi il n'existe pas de surjection de $\mathbb{N}$ sur E, donc E n'est pas dénombrable (cf proposition 11.1.2). □

Proposition 11.1.6 (non dénombrabilité de l'ensemble des réels)
L'ensemble $\mathbb{R}$ est infini non dénombrable.

Il suffit de montrer que $[0,1[$ n'est pas dénombrable.
Soit f une application quelconque de $\mathbb{N}^*$ sur $[0,1[$. Il suffit de montrer qu'elle n'est pas surjective.
Pour tout $n\geqslant 1$, soit a_n le n-ième « chiffre après la virgule » dans le développement décimal du réel $f(n)$.
Pour tout $n\geqslant 1$, on pose $x_n=0$ si $a_n\neq 0$, et $x_n=1$ si $a_n=0$.
Soit x le réel de $[0,1[$ dont le développement décimal est $0,x_1x_2x_3\cdots$.
Par construction, les réels $f(n)$ et x n'ont pas le même n-ième chiffre après la virgule, donc $x\neq f(n)$.
Le réel x n'a donc pas d'antécédent par d'antécédent par f : l'application f n'est pas surjective.
Puisqu'il n'existe pas de surjection de $\mathbb{N}^*$ (dénombrable) sur $[0,1[$, celui-ci n'est pas dénombrable.
A fortiori, l'ensemble $\mathbb{R}$ n'est pas dénombrable (cf proposition 11.1.1). □

Exercice 11.1 (⇝ **corrigé**)
Montrer que l'ensemble $\mathcal{P}(\mathbb{N})$ de toutes les parties de $\mathbb{N}$ n'est pas dénombrable.

▷ Unions et intersections indicées par $\mathbb{N}$

Soit $(A_n)_{n\in\mathbb{N}}$, une suite de parties d'un ensemble E. On pose :

$$\bigcup_{n=0}^{+\infty}A_n=\bigcup_{n\in\mathbb{N}}A_n=\{x\in E,\ \exists\, n\in\mathbb{N},\ x\in A_n\}\ \text{et}\ \bigcap_{n=0}^{+\infty}A_n=\bigcap_{n\in\mathbb{N}}A_n=\{x\in E,\ \forall\, n\in\mathbb{N},\ x\in A_n\}$$

Les notations $\displaystyle\bigcup_{n\in\mathbb{N}}A_n$ et $\displaystyle\bigcap_{n\in\mathbb{N}}A_n$ font ressortir que le résultat ne dépend pas de l'ordre d'énumération.

Avec ces notations, on a alors les propriétés suivantes :

– Passage au complémentaire : $\displaystyle\overline{\bigcup_{n\in\mathbb{N}}A_n}=\bigcap_{n\in\mathbb{N}}\overline{A_n}$ et $\displaystyle\overline{\bigcap_{n\in\mathbb{N}}A_n}=\bigcup_{n\in\mathbb{N}}\overline{A_n}$.

– Distributivité : $\displaystyle B\cap\Big(\bigcup_{n\in\mathbb{N}}A_n\Big)=\bigcup_{n\in\mathbb{N}}\big(B\cap A_n\big)$ et $\displaystyle B\cup\Big(\bigcap_{n\in\mathbb{N}}A_n\Big)=\bigcap_{n\in\mathbb{N}}\big(B\cup A_n\big)$

– Union croissante, et intersection décroissante :

On pose $\displaystyle C_n=\bigcup_{0\leqslant k\leqslant n}A_k$ et $\displaystyle D_n=\bigcap_{0\leqslant k\leqslant n}A_k$. Alors $\displaystyle\bigcup_{n\in\mathbb{N}}A_n=\bigcup_{n\in\mathbb{N}}C_n$ et $\displaystyle\bigcap_{n\in\mathbb{N}}A_n=\bigcap_{n\in\mathbb{N}}D_n$.

- Remarquons que pour tout n, on a $C_n \subset C_{n+1}$ (suite croissante) et $D_{n+1} \subset D_n$ (suite décroissante). Il s'agit donc de prouver qu'on peut remplacer une union (resp. une intersection) de parties de E (indicée par $\mathbb{N}$) par une union croissante (resp. une intersection décroissante).
- Pour tout n de $\mathbb{N}$, on a $A_n \subset C_n$, et il en résulte $\bigcup\limits_{n\in\mathbb{N}} A_n \subset \bigcup\limits_{n\in\mathbb{N}} C_n$.

  Ensuite, soit x un élément de $\bigcup\limits_{n\in\mathbb{N}} C_n$. Il existe donc un entier n_0 tel que x soit dans $C_{n_0} = \bigcup\limits_{0\leqslant k\leqslant n_0} A_k$.

  Mais cela signifie qu'il existe $n_1 \in [\![0, n_0]\!]$ tel que x soit dans A_{n_1}. Ainsi x est dans $\bigcup\limits_{n\in\mathbb{N}} A_n$.

  On a donc obtenu l'égalité $\bigcup\limits_{n\in\mathbb{N}} A_n = \bigcup\limits_{n\in\mathbb{N}} C_n$.
- Pour l'intersection, on peut écrire une démonstration analogue mais on peut aussi utiliser des passages au complémentaire dans E et le résultat précédent sur l'union :

$$\overline{\bigcap_{n\in\mathbb{N}} D_n} = \bigcup_{n\in\mathbb{N}} \overline{D_n} = \bigcup_{n\in\mathbb{N}} \overline{\bigcap_{0\leqslant k\leqslant n} A_k} = \bigcup_{n\in\mathbb{N}} \bigcup_{0\leqslant k\leqslant n} \overline{A_k} = \bigcup_{n\in\mathbb{N}} \overline{A_n} = \overline{\bigcap_{n\in\mathbb{N}} A_n} \quad \text{donc} \quad \bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n\in\mathbb{N}} D_n \qquad \square$$

Exercice 11.2 (⇝ **corrigé**)

On se donne une suite $(A_n)_{n\geqslant 0}$ de parties d'un ensemble E.

Décrire (uniquement avec des réunions ou des intersections) les sous-ensembles suivants de E :
- B est l'ensemble des éléments de E qui appartiennent à une infinité de A_n.
- C est l'ensemble des éléments de E qui n'appartiennent qu'à un nombre fini de A_n.
- D est l'ensemble des éléments de E qui appartiennent à tous les A_n à partir d'un certain rang.

▷ **Fonctions à valeurs discrètes**

Définition 11.1.2 (Fonctions à valeurs discrètes)

On dit qu'une fonction $X : \Omega \to E$ est *à valeurs discrètes* (ou simplement « est discrète ») si son ensemble image $X(\Omega)$ est fini ou dénombrable

- Ici l'adjectif « discret » évoque la possibilité de « séparer » ou d'« isoler » les éléments par une simple énumération avec des indices dans $\mathbb{N}$).
- Évidemment cette définition n'a d'intérêt que si l'ensemble d'arrivée E est infini non dénombrable.
- Cette notion sera reprise dans la section 11.2.1 (« Variables aléatoires discrètes »).

11.1.2 Tribu, espace probabilisé

Dans tout ce qui suit, on va étudier des situations dans le déroulement desquelles intervient le hasard (on parlera d'« expériences aléatoires ») et à l'issue desquelles on pourra dire si tel ou tel événement particulier s'est, ou ne s'est pas, produit. Pour estimer les « chances » que possède un événement de se produire effectivement (et avant même que l'expérience aléatoire n'ait débuté), on cherchera à lui attribuer une probabilité.

Il est souvent possible (mais pas toujours nécessaire, ni même utile) de décrire l'ensemble de toutes les « issues » de l'expérience (on parle aussi de « résultats élémentaires »). Cet ensemble Ω est appelé « univers » ou « ensemble des possibles ». Avec ce vocabulaire, un événement peut être vu comme une collection particulière d'issues de l'expérience.

Notre objectif, dans un premier temps, est de fournir un modèle mathématique qui rende compte du concept d'expérience aléatoire, d'événement au sein de cette expérience, et qui permette de définir précisément ce qu'est la probabilité qu'un tel événement « se produise ».

Définition 11.1.3 (tribu de parties d'un ensemble)
Soit Ω un ensemble, et soit $\mathscr{A}$ un sous-ensemble de $\mathcal{P}(\Omega)$.
On dit que $\mathscr{A}$ est une *tribu* sur Ω si les propriétés suivantes sont vérifiées :
a) l'ensemble Ω est lui-même un élément de $\mathscr{A}$.
b) pour tout A de $\mathscr{A}$, son complémentaire $\overline{A} = \Omega \setminus A$ est encore un élément de $\mathscr{A}$.
c) si $(A_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathscr{A}$, la réunion $\displaystyle\bigcup_{n=0}^{+\infty} A_n$ est encore un élément de $\mathscr{A}$.

Remarques

- On exprime la propriété (c) en disant qu'une tribu est *stable par union dénombrable*.
- Si $\mathscr{A}$ est une tribu, alors la partie vide $\emptyset$ est un *élément* de $\mathscr{A}$ (utiliser (b) avec $A = \Omega$).
- Les ensembles $\{\emptyset, \Omega\}$ et $\mathcal{P}(\Omega)$ sont des tribus de parties de Ω.
  On dit souvent que $\{\emptyset, \Omega\}$ est la *tribu grossière* (c'est en fait la plus petite tribu possible).
  On dit également que $\mathcal{P}(\Omega)$ est la *tribu discrète*.
- Si $A \subset \Omega$, alors $\mathscr{A} = \{\emptyset, A, \overline{A}, \Omega\}$ est une tribu (qui se réduit à $\{\emptyset, \Omega\}$ si $A = \emptyset$ ou $A = \Omega$).

▷ Complément (hors-programme) : tribu engendrée

On vérifie facilement qu'une intersection quelconque de tribus sur Ω est encore une tribu sur Ω.
Si $\mathscr{X}$ est une partie *quelconque* de $\mathcal{P}(\Omega)$, il existe donc une plus petite tribu $\mathscr{A}$ contenant $\mathscr{X}$.
On dit que $\mathscr{A}$ est la tribu *engendrée* par $\mathscr{X}$.
Par exemple, si $A \subset \Omega$, la tribu engendrée par $\mathscr{X} = \{A\}$ est $\mathscr{A} = \{\emptyset, A, \overline{A}, \Omega\}$.

C'est déjà un peu plus compliqué si $\mathscr{X} = \{A, B\}$, avec $A \subset \Omega$ et $B \subset \Omega$.
En effet, une tribu contenant $\mathscr{X}$ doit contenir : $X_1 = A \cap \overline{B}$, $X_2 = \overline{A} \cap B$, $X_3 = A \cap B$, $X_4 = \overline{A} \cap \overline{B}$.
Elle doit donc contenir les 16 parties suivantes $Y_J = \displaystyle\bigcup_{j\in J} X_j$, où J est une partie quelconque de $[\![1, 4]\!]$.
Réciproquement, on vérifie que $\mathscr{A} = \{Y_J,\ J \subset [\![1, 4]\!]\}$ est une tribu sur $\mathcal{P}(\Omega)$.

Proposition 11.1.7 (stabilité d'une tribu par intersection dénombrable)
Soit $\mathscr{A}$ une tribu de parties d'un ensemble Ω.
Si $(A_n)_{n\geqslant 0}$ est une suite d'éléments de $\mathscr{A}$, l'intersection $\displaystyle\bigcap_{n=0}^{+\infty} A_n$ est encore un élément de $\mathscr{A}$.

On utilise ici les parties (b) et (c) de la définition 11.1.3.
On sait que les $\overline{A_n}$ sont dans $\mathscr{A}$. Il est donc de même de $B = \displaystyle\bigcup_{n=0}^{+\infty} \overline{A_n}$, donc de $\overline{B} = \displaystyle\bigcap_{n=0}^{+\infty} A_n$. □

▷ Stabilité par union et intersection finie

Une tribu $\mathscr{A}$ est stable par *union finie*, et par *intersection finie*.
En d'autres termes, si les ensembles $A_0, A_1, \ldots, A_m$ sont dans $\mathscr{A}$, alors $\displaystyle\bigcup_{n=0}^{m} A_n$ et $\displaystyle\bigcap_{n=0}^{m} A_n$ sont dans $\mathscr{A}$.
Il suffit pour s'en convaincre d'utiliser la version dénombrable en posant $A_n = A_m$ pour $n > m$.
Si pour chaque n on pose $B_n = A_n$ ou $B_n = \overline{A_n}$, les ensembles $\displaystyle\bigcup_{n=0}^{m} B_n$ et $\displaystyle\bigcap_{n=0}^{m} B_n$ sont encore dans $\mathscr{A}$.

▷ Terminologie des expériences aléatoires

– On fait le choix d'un ensemble Ω, appelé *univers*, et qui décrit de façon assez informelle l'ensemble des *issues possibles* (ou *résultats possibles* ou *réalisations*) de l'expérience aléatoire étudiée. Il n'est souvent ni nécessaire ni même utile d'en donner une définition précise.

– On fait le choix d'une tribu $\mathscr{A}$ de parties de Ω.

Si l'ensemble Ω est fini ou dénombrable, on choisira le plus souvent $\mathscr{A} = \mathcal{P}(\Omega)$.

Les éléments A de $\mathscr{A}$ (ce sont donc des parties de Ω) sont appelés *événements*. Le choix de la tribu des événements est une partie importante de la phase de modélisation de l'expérience aléatoire. L'ensemble des événements étudiés doit donc satisfaire aux conditions de la définition 11.1.3.

– On dit parfois que le couple $(\Omega, \mathscr{A})$ est un *espace probabilisable*.

– Quand l'expérience aléatoire « se déroule », elle se conclut par une issue ω (un élément de Ω).

Si A est un événement et si cette issue ω est dans A, on dit alors que l'« événement A est réalisé ».

– L'ensemble $\emptyset$ est appelé *événement impossible*, et l'ensemble Ω est appelé *événement certain*.

Si A est un événement, on dit que $\overline{A}$ est l'*événement contraire* de A.

– Soit A et B deux événements :

- si $A \subset B$, on dit que A *implique* B ;
- si $A \cap B = \emptyset$, on dit que A et B sont *incompatibles* (ou tout simplement : disjoints) ;
- l'ensemble $A \cap B$ est appelé « événement A et B » ;
- l'ensemble $A \cup B$ est appelé « événement A ou B ».

▷ Rappels du programme de première année

– Le programme de première année se limite aux univers Ω finis, et à $\mathscr{A} = \mathcal{P}(\Omega)$.

Dans ce cas, les événements sont donc *toutes* les parties de Ω.

Les singletons $\{\omega\}$, avec ω dans Ω, sont appelés *événements élémentaires*.

– On dit qu'une famille $(A_i)_{0 \leqslant i \leqslant n}$ de Ω est un *système complet d'événements* si elle vérifie les conditions :

- (facultatif) les événements $A_0, \ldots, A_n$ sont tous non vides : $\forall\, i \in [\![0, n]\!]$, $A_i \neq \emptyset$.
- les événements $A_0, \ldots, A_n$ sont deux à deux incompatibles : $\forall\,(i,j) \in [\![0, n]\!]^2$, $i \neq j \Rightarrow A_i \cap A_j = \emptyset$.
- la réunion des événements $A_0, \ldots, A_n$ est l'événement certain : $\displaystyle\bigcup_{n=0}^{m} A_n = \Omega$.

Avec cette définition, et pour toute issue ω de l'expérience, un et un seul des événements A_i est réalisé.

Définition 11.1.4 (probabilité sur $(\Omega, \mathscr{A})$; espace probabilisé)
Soit $\mathscr{A}$ une tribu sur un univers Ω.
On appelle *probabilité* sur $(\Omega, \mathscr{A})$ toute application $\mathbb{P}$ définie sur $\mathscr{A}$, à valeurs dans $[0, 1]$ telle que :
a) $\mathbb{P}(\Omega) = 1$ (probabilité de l'événement certain)
b) pour toute suite $(A_n)_{n \geqslant 0}$ d'événements deux à deux incompatibles, $\mathbb{P}\Big(\displaystyle\bigcup_{n=0}^{+\infty} A_n\Big) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$
Le triplet $(\Omega, \mathscr{A}, \mathbb{P})$ est alors appelé un *espace probabilisé*.

La propriété (b) est appelée « σ-additivité ».

Elle affirme d'abord la convergence de la série $\sum \mathbb{P}(A_n)$ avant d'en donner la valeur.

On admet que la valeur de la somme $\sum_{n=0}^{+\infty} \mathbb{P}(A_n)$ ne dépend pas de l'ordre d'énumération.

▷ Quelques propriétés immédiates

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.

– Probabilité de l'événement impossible : on a $\boxed{\mathbb{P}(\emptyset) = 0}$

> Dans la partie (b) de la définition 11.1.4, on choisit $A_n = \emptyset$ pour tout n de $\mathbb{N}$.
> La convergence de la série $\sum \mathbb{P}(\emptyset)$ exige $\mathbb{P}(\emptyset) = 0$. □

– Si les événements A, B sont incompatibles : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
Plus généralement si $(A_n)_{0 \leqslant n \leqslant m}$ sont deux à deux incompatibles, alors $\mathbb{P}\Big(\bigcup_{n=0}^{m} A_m\Big) = \sum_{n=0}^{m} \mathbb{P}(A_n)$.

> Il suffit en effet d'utiliser la σ-additivité avec $A_n = \emptyset$ pour tout $n \geqslant m$. □

– Si A et $\overline{A}$ sont deux événements contraires, alors $\boxed{\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)}$

> Puisque A et $\overline{A}$ sont incompatibles, c'est la traduction de $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = \mathbb{P}(A \cup \overline{A}) = \mathbb{P}(\Omega) = 1$. □

– Si A et B sont deux événements, et si $A \subset B$ alors $\mathbb{P}(A) \leqslant \mathbb{P}(B)$ (croissance de la probabilité).

> On a $B = A \cup (B \setminus A)$, avec $A \cap (B \setminus A) = \emptyset$, donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geqslant \mathbb{P}(A)$. □

– Si A et B sont deux événements, alors $\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)}$

> On a $\begin{cases} A \cup B = A \cup (B \setminus A) \\ B = (A \cap B) \cup (B \setminus A) \end{cases}$ (union d'événements incompatibles) donc $\begin{cases} \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \\ \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) \end{cases}$
> Par différence, on trouve $\mathbb{P}(A \cup B) - \mathbb{P}(B) = \mathbb{P}(A) - \mathbb{P}(B \setminus A)$, ce qu'il fallait démontrer. □

– Si A, B, C sont trois événements, alors :

$$\begin{aligned} \mathbb{P}(A \cup B \cup C) &= \mathbb{P}\big((A \cup B) \cup C\big) = \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}\big((A \cap C) \cup (B \cap C)\big) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \\ &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \end{aligned}$$

Exercice 11.3 (⇝ **corrigé**)
Soit A, B, C trois événements. Montrer que $\mathbb{P}(A \cup B) + \mathbb{P}(\overline{A} \cup B) + \mathbb{P}(A \cup \overline{B}) + \mathbb{P}(\overline{A} \cup \overline{B}) = 3$.

Exercice 11.4 (⇝ **corrigé**)
Soit $(A_i)_{1 \leqslant i \leqslant n}$ une famille de n événements. On note $\mathcal{J}_n$ l'ensemble des 2^n-uplets $(B_1, \ldots, B_n)$ obtenus en posant, indifféremment $B_i = A_i$ ou $B_i = \overline{A_i}$. Montrer que $\sum_{\mathcal{J}_n} \mathbb{P}(B_1 \cup \cdots \cup B_n) = 2^n - 1$.

Exercice 11.5 (formule du crible, ⇝ **corrigé**)

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Soit $(A_i)_{1\leqslant i\leqslant n}$ une famille de n événements.

Montrer que $\mathbb{P}\Big(\bigcup\limits_{1\leqslant i\leqslant n} A_i\Big) = \sum\limits_{k=1}^{n}(-1)^{k-1}\sum\limits_{\substack{J\subset[\![1,n]\!]\\ \mathrm{Card}(J)=k}} \mathbb{P}\Big(\bigcap\limits_{j\in J} A_j\Big)$ $\left(\begin{array}{l}\text{somme interne étendue à l'ensemble des}\\ \text{parties } J \text{ de } [\![1,n]\!] \text{ telles que } \mathrm{Card}(J)=k\end{array}\right)$

NB : la démonstration, très technique, peut être passée en première lecture. On verra plus tard une preuve plus simple et qui utilise la notion de « fonction indicatrice ».

Exercice 11.6 (⇝ **corrigé**)

Une urne contient n boules indiscernables, numérotées de 1 à n. On les extrait, une à une, sans remise. On note A_i l'événement « la boule apparue lors du $i^{\text{ème}}$ tirage porte le numéro i ».

a) Calculer $\mathbb{P}\Big(\bigcap\limits_{j\in J} A_j\Big)$, où J est une partie non vide de $[\![1,n]\!]$ de cardinal k.

b) En utilisant le résultat de l'exercice 11.5, calculer la probabilité de l'événement B_n : « à aucun moment, le tirage n°i ne fait apparaître la boule n°i ». Préciser $\lim\limits_{n\to+\infty} \mathbb{P}(B_n)$.

Exercice 11.7 (⇝ **corrigé**)

On dispose au hasard quatre pions sur trois cases (assez grandes pour contenir tous les pions). Quelle est la probabilité qu'une case au moins soit vide ?

▷ Rappels du programme de première année

- Dans le programme de première année, on se limite aux univers finis Ω et à la tribu $\mathscr{A} = \mathcal{P}(\Omega)$.
  La partie (b) de la définition 11.1.4 se simplifie alors en :
  « pour toutes parties disjointes A et B de Ω, $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ ».
  On en déduit $\mathbb{P}\Big(\bigcup\limits_{n=0}^{m} A_m\Big) = \sum\limits_{n=0}^{m} \mathbb{P}(A_n)$ quand $(A_n)_{0\leqslant n\leqslant m}$ sont deux à deux incompatibles.
  Dans le cas Ω fini, le recours à des unions d'événements indicées par $\mathbb{N}$ est évidemment sans objet.
- Si $(A_i)_{0\leqslant i\leqslant n}$ est un système complet d'événements, alors $\sum\limits_{i=0}^{n} \mathbb{P}(A_i) = 1$. En particulier $\sum\limits_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$.
- Réciproquement, une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ est caractérisée par la donnée des $\big(\mathbb{P}(\{\omega\})\big)_{\omega\in\Omega}$, c'est-à-dire par la probabilité des événements élémentaires, avec bien sûr $\sum\limits_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$.
- Une situation courante est l'*hypothèse d'équiprobabilité*, où on pose : $\forall\,\omega\in\Omega,\ \mathbb{P}(\{\omega\}) = \dfrac{1}{\mathrm{Card}(\Omega)}$.
  Pour toute partie A de Ω, on a alors $\mathbb{P}(A) = \dfrac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}$.

▷ Cas d'un univers dénombrable

Ou suppose ici $\Omega = \{\omega_n,\ n\in\mathbb{N}\}$, où les ω_n sont distincts deux à deux.

On suppose également qu'on a choisi la tribu $\mathscr{A} = \mathcal{P}(\Omega)$.

Les singletons $\{\omega_n\}$, avec n dans $\mathbb{N}$, sont appelés *événements élémentaires*.

- Soit $\mathbb{P}$ une probabilité sur $(\Omega, \mathscr{A})$.
  Pour toute partie A de Ω (donc pour tout événement), on a $A = \bigcup\limits_{\omega_n\in A}\{\omega_n\}$, donc $\mathbb{P}(A) = \sum\limits_{\omega_n\in A}\mathbb{P}(\{\omega_n\})$.

– Réciproquement, soit $(p_n)_{n\geqslant 0}$ une suite de $[0,1]$, avec $\displaystyle\sum_{n=0}^{+\infty} p_n = 1$.

  On définit une (unique) probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$ en posant : $\forall\, n \in \mathbb{N},\ \mathbb{P}(\{\omega_n\}) = p_n$.

Une probabilité sur Ω dénombrable est donc caractérisée par les probabilité des événements élémentaires.

▷ Complément : un exemple d'univers non dénombrable

On considère l'expérience qui consiste à lancer « indéfiniment » (ce qui consiste déjà en soi une vue de l'esprit) une pièce de monnaie (supposée bien équilibrée) et à observer le résultat « Pile » ou « Face ».

On traduit le résultat « Pile » par l'entier 1, et le résultat « Face » par l'entier 0.

On pourrait considérer qu'un résultat élémentaire ω est une suite à valeurs dans $\{0,1\}$, ou encore un « mot », de longueur infinie, formé des seuls symboles 0 et 1.

On sait que l'univers $\Omega = \{0,1\}^{\mathbb{N}}$ n'est pas dénombrable (cf proposition 11.1.5).

Pour créer une probabilité sur cet univers il faut d'abord choisir une tribu $\mathscr{A}$ de parties de Ω (en d'autres termes, il faut décider quels événements méritent notre attention).

La première idée (comme dans le cas d'un univers fini ou dénombrable) est de choisir $\mathscr{A} = \mathcal{P}(\Omega)$.

Soit $\omega = (\omega_n)_{n\in\mathbb{N}}$ un résultat élémentaire fixé, donc une suite de 0 pour Face et de 1 pour Pile.

Soit N donné dans $\mathbb{N}$, et soit $A_{N,\omega}$ l'événement formé des u de Ω tels que : $\forall\, n \in [\![0, N-1]\!],\ u_n = \omega_n$.

Dire que $A_{N,\omega}$ est réalisé, c'est donc imposer le résultat (Pile ou Face) des N premiers lancers.

L'intuition (la pièce est équilibrée) veut qu'on attribue la probabilité $\dfrac{1}{2^N}$ à l'événement $A_{N,\omega}$.

Mais il est clair que $\{\omega\} \subset A_{N,\omega}$ donc $\mathbb{P}(\{\omega\}) \leqslant \dfrac{1}{2^N}$ pour tout N de $\mathbb{N}$, donc $\mathbb{P}(\{\omega\}) = 0$.

Ainsi chaque événement élémentaire (c'est-à-dire chaque succession indéfinie particulère de résultats Pile ou Face) est-elle un événement négligeable (de probabilité nulle). Grâce à la propriété de σ-additivité (voir la définition 11.1.4), tout événement dénombrable est aussi de probabilité nulle.

Dans cet exemple, on voit donc qu'une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ ne peut pas être définie par la seule connaissance des probabilités des événements élémentaires $\{\omega\}$.

Il y a de toutes façons des difficultés : l'égalité $P(\Omega) = \sum P(\{\omega\})$ par exemple n'a en effet plus aucun sens (pour trois raisons : on ne connait pas la signification d'une somme sur un ensemble non dénombrable d'indices, les probabilités $P(\{\omega\})$ sont nulles et la probabilité $\mathbb{P}(\Omega)$ vaut 1).

Plutôt que de choisir la tribu $\mathcal{P}(\Omega)$, on procédera donc de la façon suivante :

- Pour toute partie J de $\mathbb{N}$, pour tout ω de Ω, on pose $E_{J,\omega} = \{u \in \Omega,\ \forall\, n \in J,\ u_n = \omega_n\}$.
- Si J est fini de cardinal k, on considère que $E_{J,\omega}$ est un événement, et que sa probabilité est $\dfrac{1}{2^k}$.

  C'est légitime puisque réaliser $E_{J,\omega}$ c'est fixer, dans la succession indéfinie de lancers, le résultat (Pile ou Face) des k lancers dont la position est un élément de J.
- On choisira alors (et on admet son existence) la plus petite tribu $\mathscr{A}$ contenant les événements $E_{J,\omega}$ (avec J une partie finie quelconque de $\mathbb{N}$ et ω un élément quelconque de Ω), donc contenant les réunions finies ou dénombrables de ces événements.
- Dans ce schéma, et si J est une partie dénombrable de $\mathbb{N}$, l'ensemble $E_{J,\omega}$ est un événement, et il est de probabilité nulle (si $J = \mathbb{N}$, on retrouve $\mathbb{P}(\{\omega\} = 0)$).

Définition 11.1.5 (événements négligeables, événements presque sûrs)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.
On dit qu'un événement A est *négligeable* si $\mathbb{P}(A) = 0$, et *presque sûr* si $\mathbb{P}(A) = 1$.

Remarques

– L'événement impossible $\emptyset$ est évidemment négligeable, de même que l'événement certain Ω est presque sûr. Réciproquement, il peut exister des événements qui sont négligeables sans être impossibles (obtenir toujours Pile quand on lance indéfiniment une pièce honnête !) ou presque sûrs sans être certains.
– Tout événement qui contient un événement presque sûr est lui-même presque sûr.
Tout événement inclus dans un événement négligeable est lui-même négligeable.
– Si A est presque sûr, et si B est un événement quelconque, alors $\mathbb{P}(B \cap A) = \mathbb{P}(B)$.

L'événement $B \cup A$ contient A, donc est presque sûr. Ainsi $\mathbb{P}(A) = \mathbb{P}(B \cup A) = 1$.
L'égalité $\mathbb{P}(B \cup A) = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(B \cap A)$ donne alors $\mathbb{P}(B) = \mathbb{P}(B \cap A)$. □

Proposition 11.1.8 (propriété de continuité croissante)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ *un espace probabilisé.*
Soit $(A_n)_{n \geqslant 0}$ *une suite croissante d'événements, c'est-à-dire telle que :* $\forall n \in \mathbb{N},\ A_n \subset A_{n+1}$.
Alors on a l'égalité : $\lim\limits_{n \to +\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} A_n\Big)$.

Par hypothèse : $\forall n \in \mathbb{N},\ \mathbb{P}(A_n) \leqslant \mathbb{P}(A_{n+1}) \leqslant 1$. La suite $(\mathbb{P}(A_n))_{n \geqslant 0}$, croissante et majorée, converge.
On pose $B_0 = A_0$ et, pour tout n de $\mathbb{N}^*$: $B_n = A_n \setminus A_{n-1}$.
Pour $m < n$ on a $B_m \subset A_m \subset A_{n-1}$ donc $B_m \cap B_n = \emptyset$: les événements B_n sont deux à deux incompatibles.
On a $A_0 = B_0$ et $A_n = A_{n-1} \cup B_n$ si $n \geqslant 1$. Ainsi $A_n = \bigcup\limits_{k=0}^{n} B_k$ et $\mathbb{P}(A_n) = \sum\limits_{k=0}^{n} \mathbb{P}(B_k)$ pour tout n.
On sait que $\bigcup\limits_{n=0}^{+\infty} B_n = \bigcup\limits_{n=0}^{+\infty} \Big(\bigcup\limits_{k=0}^{n} B_k\Big)$ (cf « unions et intersections indicées par $\mathbb{N}$ » dans 11.1.1).
Ainsi $\bigcup\limits_{n=0}^{+\infty} B_n = \bigcup\limits_{n=0}^{+\infty} A_n$, donc $\mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} A_n\Big) = \mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} B_n\Big) = \sum\limits_{n=0}^{+\infty} \mathbb{P}(B_n) = \lim\limits_{n \to +\infty} \sum\limits_{k=0}^{n} \mathbb{P}(B_k) = \lim\limits_{n \to +\infty} \mathbb{P}(A_n)$. □

▷ Exemple : suites stationnaires de lancers d'une pièce

On reprend ici l'étude qui a été faite sur l'expérience qui consiste à lancer « indéfiniment » une pièce de monnaie équilibrée et à observer le résultat (1 pour « Pile », ou 0 pour « Face »).
On va montrer que l'ensemble S des suites stationnaires de lancers est un événement négligeable.
En effet, $S = \bigcup\limits_{m=0}^{+\infty} S_m$, où S_m est l'ensemble des lancers stationnaires pour $n \geqslant m$.
Pour tout m de $\mathbb{N}$, on a $S_m \subset S_{m+1}$.
Or $S_m = S_{m,P} \cup S_{m,F}$, où $S_{m,P}$ (resp. $S_{m,F}$) désigne les lancers renvoyant Pile (resp. Face) pour $n \geqslant m$.
On sait que les ensembles $S_{m,P}$ et $S_{m,F}$ (qui consistent à fixer le résultat du lancer pour un ensemble dénombrable de positions) sont des événements négligeables. Il en est donc de même de S_m.
Mais $S = \bigcup\limits_{m=0}^{+\infty} S_m$ donc S est un événement, et $\mathbb{P}(S) = \mathbb{P}\Big(\bigcup\limits_{m=0}^{+\infty} S_m\Big) = \lim\limits_{m \to +\infty} \mathbb{P}(S_m) = 0$.

Proposition 11.1.9 (propriété de continuité décroissante)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.
Soit $(A_n)_{n\geqslant 0}$ une suite décroissante d'événements, c'est-à-dire telle que : $\forall n \in \mathbb{N},\ A_{n+1} \subset A_n$.
Alors on a l'égalité : $\lim\limits_{n\to+\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcap\limits_{n=0}^{+\infty} A_n\Big)$.

Il suffit d'utiliser la propriété de continuité croissante sur la suite (croissante) d'événements $n \mapsto \overline{A_n}$.
L'égalité $\lim\limits_{n\to+\infty} \mathbb{P}(\overline{A_n}) = \mathbb{P}\Big(\bigcup\limits_{n\geqslant 0} \overline{A_n}\Big)$ devient en effet : $\lim\limits_{n\to+\infty} (1 - \mathbb{P}(A_n)) = \mathbb{P}\Big(\overline{\bigcap\limits_{n\geqslant 0} A_n}\Big) = 1 - \mathbb{P}\Big(\bigcap\limits_{n\geqslant 0} A_n\Big)$.
Il en résulte immédiatement : $\lim\limits_{n\to+\infty} \mathbb{P}(A_n) = \mathbb{P}\Big(\bigcap\limits_{n=0}^{+\infty} A_n\Big)$. □

Proposition 11.1.10 (propriété de sous-additivité)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, et soit $(A_n)_{n\geqslant 0}$ une suite d'événements.
Alors on a l'inégalité : $\mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} A_n\Big) \leqslant \sum\limits_{n=0}^{+\infty} \mathbb{P}(A_n)$.

L'inégalité précédente ne signifie pas que la série $\sum \mathbb{P}(A_n)$ converge !
D'ailleurs, si cette série diverge, ou si sa somme est supérieure à 1, le résultat est sans intérêt.
La proposition précédente montre en particulier qu'une réunion dénombrable d'événements négligeables est un événement négligeable.

On sait que $\mathbb{P}(A_0 \cup A_1) = \mathbb{P}(A_0) + \mathbb{P}(A_1) - \mathbb{P}(A_0 \cap A_1) \leqslant \mathbb{P}(A_0) + \mathbb{P}(A_1)$.
Par une récurrence finie immédiate, il en résulte que pour tout entier N, on a : $\mathbb{P}\Big(\bigcup\limits_{n=0}^{N} A_n\Big) \leqslant \sum\limits_{n=0}^{N} \mathbb{P}(A_n)$ $(\star)$.
En posant $B_n = \bigcup\limits_{k=0}^{n} A_k$, on définit une suite croissante $(B_n)_{n\geqslant 0}$ d'événements.
Il est clair qu'on a les égalités $\bigcup\limits_{n=0}^{N} A_n = \bigcup\limits_{n=0}^{N} B_n$, et $\bigcup\limits_{n=0}^{+\infty} A_n = \bigcup\limits_{n=0}^{+\infty} B_n$ (cf sous-section 11.1.1).
L'inégalité $(\star)$ devient $\mathbb{P}\Big(\bigcup\limits_{n=0}^{N} B_n\Big) \leqslant \sum\limits_{n=0}^{N} \mathbb{P}(A_n)$ et, par passage à la limite : $\lim\limits_{N\to+\infty} \mathbb{P}\Big(\bigcup\limits_{n=0}^{N} B_n\Big) \leqslant \sum\limits_{n=0}^{+\infty} \mathbb{P}(A_n)$.
Par continuité croissante, cela s'écrit $\mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} B_n\Big) \leqslant \sum\limits_{n=0}^{+\infty} \mathbb{P}(A_n)$, donc : $\mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} A_n\Big) \leqslant \sum\limits_{n=0}^{+\infty} \mathbb{P}(A_n)$. □

Exercice 11.8 ($\rightsquigarrow$ **corrigé**)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, et soit $(A_n)_{n\geqslant 0}$ une suite d'événements presque sûrs.
Montrer que l'événement $\bigcap\limits_{n=0}^{+\infty} A_n$ est presque sûr.

Exercice 11.9 (premier lemme de Borel-Cantelli, $\rightsquigarrow$ **corrigé**)
Soit $(A_n)_{n\geqslant 0}$ une suite d'événements, telle que la série $\sum \mathbb{P}(A_n)$ converge. On pose $B = \bigcap\limits_{n=0}^{+\infty} \Big(\bigcup\limits_{k=n}^{+\infty} A_k\Big)$.

a) Montrer que (B est réalisé) $\Leftrightarrow$ (une infinité d'événements A_n sont réalisés).

b) Prouver que l'événement B est négligeable.

11.1.3 Conditionnement

Un contexte choisi une fois pour toutes

Tous les énoncés des sous-sections "Conditionnement" et "Indépendance" commencent par la même phrase « Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé ».

Proposition 11.1.11 (probabilité conditionnelle de A sachant B)

Soit B un événement tel que $\mathbb{P}(B) > 0$. Pour tout événement A, on pose $\mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Alors $\mathbb{P}_B$ est une probabilité sur $(\Omega, \mathscr{A})$, dite « application probabilité conditionnée à B ».

Il est clair que $\mathbb{P}_B$ est à valeurs dans $[0, 1]$, et que $\mathbb{P}_B(\Omega) = 1$.

Pour toute suite $(A_n)_{n\geqslant 0}$ d'événements, on a $\left(\bigcup\limits_{n=0}^{+\infty} A_n\right) \cap B = \bigcup\limits_{n=0}^{+\infty} (A_n \cap B)$.

On suppose que les A_n sont deux à deux incompatibles. Il en est donc de même des $A_n \cap B$.

On en déduit alors facilement la σ-additivité de $\mathbb{P}_B$:

$$\mathbb{P}_B\Big(\bigcup_{n=0}^{+\infty} A_n\Big) = \frac{1}{\mathbb{P}(B)}\,\mathbb{P}\Big(\Big(\bigcup_{n=0}^{+\infty} A_n\Big) \cap B\Big) = \frac{1}{\mathbb{P}(B)}\,\mathbb{P}\Big(\bigcup_{n=0}^{+\infty} (A_n \cap B)\Big) = \frac{1}{\mathbb{P}(B)}\sum_{n=0}^{+\infty} \mathbb{P}(A_n \cap B) = \sum_{n=0}^{+\infty} \mathbb{P}_B(A_n) \qquad \square$$

Le réel $\mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (aussi noté $\mathbb{P}(A \mid B)$) est la « probabilité conditionnelle de A sachant B ».

Avec cette définition, on a l'égalité évidente (mais importante) : $\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}_B(A)}$.

Convention d'écriture

Si $\mathbb{P}(B) = 0$, qui implique $\mathbb{P}(A \cap B) = 0$, on convient que : $\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A\,|\,B) = 0$.
Cependant, dans ce cas, on évitera d'utiliser la notation $\mathbb{P}_B$.

Exercice 11.10 ($\rightsquigarrow$ **corrigé**)
Mme Michu a deux enfants dont au moins une fille. Probabilité qu'elle ait au moins un garçon ?

Exercice 11.11 ($\rightsquigarrow$ **corrigé**)
Mme Michu cherche son parapluie qui se trouve (peut être !) dans l'un des 7 étages de son immeuble.
Soit p (avec $0 < p < 1$) la probabilité que le parapluie se trouve effectivement dans l'immeuble.
Pour $1 \leqslant i \leqslant 7$, on note E_i l'évènement : « Le parapluie se trouve à l'étage i ».

1. On suppose que la probabilité que le parapluie soit à l'étage i ne dépend pas de i.
   On suppose également que Mme Michu a exploré en vain les i premiers étages.
   Quelle est la probabilité p_i que le parapluie se trouve dans un des étages non visités ?
   Comment évolue cette probabilité lorqu'on s'élève dans l'immeuble sans trouver le parapluie ?
2. On suppose que la probabilité de retrouver le parapluie à l'étage i est proportionnelle à i :
   Reprendre la question précédente avec cette nouvelle hypothèse (mais en notant q_i plutôt que p_i).
3. Comparer les probabilités p_i et q_i obtenues dans les deux questions précédentes.

Proposition 11.1.12 (formule des probabilités composées)
Soit $(A_n)_{n\geqslant 0}$ *une suite d'événements. Pour tout* n *de* $\mathbb{N}$, *posons* $B_n = \bigcap\limits_{0\leqslant k\leqslant n} A_k$.
Il est clair que la suite $n \mapsto \mathbb{P}(B_n)$ *est décroissante.*
Alors, pour tout entier n *tel que* $\mathbb{P}(B_n) > 0$, *on a l'égalité* $\mathbb{P}(B_{n+1}) = \mathbb{P}(A_0)\prod\limits_{k=0}^{n}\mathbb{P}_{B_k}(A_{k+1})$.

Pour $n = 0$, la formule précédente s'écrit : $\mathbb{P}(A_0 \cap A_1) = \mathbb{P}(A_0)\,\mathbb{P}_{A_0}(A_1)$ (à condition que $\mathbb{P}(A_0) > 0$).
Pour $n = 1$, elle s'écrit : $\mathbb{P}(A_0 \cap A_1 \cap A_2) = \mathbb{P}(A_0)\,\mathbb{P}_{A_0}(A_1)\,\mathbb{P}_{A_0\cap A_1}(A_2)$ (à condition que $\mathbb{P}(A_0 \cap A_1) > 0$).
Pour $n = 2$, elle s'écrit (sous la condition que $\mathbb{P}(A_0 \cap A_1 \cap A_2) > 0$) :

$$\mathbb{P}(A_0 \cap A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_0)\,\mathbb{P}_{A_0}(A_1)\,\mathbb{P}_{A_0\cap A_1}(A_2)\,\mathbb{P}_{A_0\cap A_1\cap A_2}(A_3)$$

En utilisant l'autre notation pour les probabilités conditionnelles, cela s'écrit :

$$\mathbb{P}(A_0 \cap A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_0)\;\mathbb{P}(A_1 \mid A_0)\;\mathbb{P}(A_2 \mid A_0 \cap A_1)\;\mathbb{P}(A_3 \mid A_0 \cap A_1 \cap A_2)$$

Pour tout n de $\mathbb{N}$, on a $B_{n+1} = B_n \cap A_{n+1} \subset B_n$, ce qui justifie la décroissance de la suite $(\mathbb{P}(B_n))_{n\geqslant 0}$.
On se donne un entier m tel que $\mathbb{P}(B_m) > 0$. On a alors $P(B_n) \geqslant \mathbb{P}(B_m) > 0$ pour tout n de $[\![0, m]\!]$.
On a prouver l'égalité (E_n) : $\mathbb{P}(B_{n+1}) = \mathbb{P}(A_0)\prod\limits_{k=0}^{n}\mathbb{P}(A_{k+1} \mid B_k)$ par récurrence finie sur n, avec $0 \leqslant n \leqslant m$.
On sait que $B_0 = A_0$ et $B_1 = A_0 \cap A_1$.
(E_0) est vraie car elle se réduit à $\mathbb{P}(B_1) = \mathbb{P}(A_0)\,\mathbb{P}(A_1 \mid B_0)$, c'est-à-dire $\mathbb{P}(A_0 \cap A_1) = \mathbb{P}(A_0)\,\mathbb{P}(A_1 \mid A_0)$.
On se donne n dans $[\![1, m]\!]$ et on suppose que (E_{n-1}) est vraie.
On a $B_{n+1} = B_n \cap A_{n+1}$ et $\mathbb{P}(B_n > 0)$. On en déduit :

$$\mathbb{P}(B_{n+1}) = \mathbb{P}(B_n)\,\mathbb{P}(A_{n+1} \mid B_n) = \Big(\mathbb{P}(A_0)\prod_{k=0}^{n-1}\mathbb{P}(A_{k+1} \mid B_k)\Big)\,\mathbb{P}(A_{n+1} \mid B_n) = \mathbb{P}(A_0)\prod_{k=0}^{n}\mathbb{P}(A_{k+1} \mid B_k)$$

ce qui prouve l'égalité (E_n) et achève la récurrence. □

Exercice 11.12 (⇝ **corrigé**)
On considère trois urnes : U_1 contient deux boules blanches et trois boules rouges ; U_2 contient deux boules vertes et quatre boules blanches ; U_3 contient cinq boules noires et deux boules rouges.
On tire une boule dans U_1, que l'on place dans U_2.
Puis on tire une boule dans U_2, que l'on place dans U_3. Enfin on tire une boule dans U_3.
Quelle est la probabilité que les trois boules tirées soient de couleurs différentes ?

La définition suivante complète la notion de système complet (fini) d'événements, vue en 1ère année :

Définition 11.1.6 (système complet dénombrable d'événements)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé.
On dit qu'une suite $(A_n)_{n\geqslant 0}$ d'éléments de $\mathscr{A}$ est un *système complet dénombrable d'événements* si :
- (facultatif) les événements A_n sont tous non vides : $\forall\, n \in \mathbb{N},\ A_n \neq \emptyset$.
- les événements A_n sont deux à deux incompatibles : $\forall\,(n, m) \in \mathbb{N}^2,\ n \neq m \Rightarrow A_n \cap A_m = \emptyset$.
- la réunion des événements A_n est l'événement certain : $\bigcup\limits_{n\geqslant 0} A_n = \Omega$.

Avec cette définition, et pour toute issue ω de l'expérience, un et un seul des événements A_n est réalisé.

Proposition 11.1.13 (formule des probabilités totales)
Soit $(A_n)_{n\in\mathbb{N}}$ un système complet dénombrable d'événements. Soit B un événement.
Alors la série $\sum \mathbb{P}(B\cap A_n)$ converge et : $\mathbb{P}(B)=\displaystyle\sum_{n=0}^{+\infty}\mathbb{P}(B\cap A_n)=\sum_{n=0}^{+\infty}\mathbb{P}(A_n)\,\mathbb{P}(B\,|\,A_n)$.

Rappel : si $\mathbb{P}(A_n)=0$, on convient que $\mathbb{P}(A_n)\,\mathbb{P}(B\,|\,A_n)=0$.

On a $\displaystyle\bigcup_{n=0}^{+\infty}A_n=\Omega$ donc $B=B\cap\Big(\bigcup_{n=0}^{+\infty}A_n\Big)=\bigcup_{n=0}^{+\infty}(B\cap A_n)$ et les $B\cap A_n$ sont deux à deux incompatibles.
Par σ-additivité (voir définition 11.1.4), on en déduit la convergence de $\sum\mathbb{P}(B\cap A_n)$ et les égalités :
$\mathbb{P}(B)=\mathbb{P}\Big(\displaystyle\bigcup_{n=0}^{+\infty}(B\cap A_n)\Big)=\sum_{n=0}^{+\infty}\mathbb{P}(B\cap A_n)=\sum_{n=0}^{+\infty}\mathbb{P}(A_n)\,\mathbb{P}(B\,|\,A_n)$. □

La proposition 11.1.13 reste vraie si les A_n sont deux à deux incompatibles et tels que $\displaystyle\sum_{n=0}^{+\infty}\mathbb{P}(A_n)=1$.

En effet, les A_n étant deux à deux incompatibles, la σ-additivité donne : $\mathbb{P}\Big(\displaystyle\bigcup_{n=0}^{+\infty}A_n\Big)=\sum_{n=0}^{+\infty}\mathbb{P}(A_n)$.
Donc dire que $\displaystyle\sum_{n=0}^{+\infty}\mathbb{P}(A_n)=1$, c'est dire que $A=\displaystyle\bigcup_{n=0}^{+\infty}A_n$ est *presque sûr*.
On alors $\mathbb{P}(B)=\mathbb{P}(B\cap A)$ (voir la troisième remarque après la définition 11.1.5).
On termine alors la démonstration comme précédemment. □

Rappel : la formule des probabilités totales est au programme de première année, pour un espace probabilisé $(\Omega,\mathcal{P}(\Omega),P)$ (avec Ω fini) et un système complet fini d'événements $(A_k)_{0\leqslant k\leqslant n}$.

Exercice 11.13 (⇝ **corrigé**)
N personnes (numérotées de 1 à N) se transmettent (dans cet ordre) une information (cette information possède un caractère binaire : elle est vraie ou fausse, c'est tout).

La personne n°1 possède l'information initiale. Elle la transmet (fidèlement ou non) à la personne n°2, et ainsi de suite.
Plus précisément, chaque personne transmet l'information reçue de son prédécesseur : fidèlement avec la probabilité p $(0<p<1)$ et en son contraire (avec la probabilité $q=1-p$).
Quelle est la probabilité f_N que l'information arrive fidèlement à la $N^{\text{ième}}$ personne ?
Que devient cette probabilité quand N tend vers $+\infty$?

Exercice 11.14 (⇝ **corrigé**)
1. Calculer l'intégrale $J_n=\displaystyle\int_0^1 x^n(1-x)^n\,\mathrm{d}x$, avec $n\in\mathbb{N}$.
2. Soit $p\in\mathbb{N}^*$. On dispose de p urnes numérotées de 1 à p.
   Pour $k\in[[1,p]]$, l'urne numéro k contient k boules noires et $p-k$ boules blanches.
   On choisit au hasard une urne puis on y effectue des tirages avec remise.
   On note $A_{n,p}$: « après $2n$ tirages, on a obtenu n boules noires ». Calculer $\mathbb{P}(A_{n,p})$, puis $\displaystyle\lim_{p\to+\infty}\mathbb{P}(A_{n,p})$.

Proposition 11.1.14 (formule de Bayes, première version)
Soit A et B deux événements tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$. Alors $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A)\,\mathbb{P}(B \mid A)}{\mathbb{P}(B)}$.

C'est une simple réécriture de : $\mathbb{P}(B)\,\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A)$. □

Exercice 11.15 (⇝ **corrigé**)
Une maladie atteint une personne sur 1000 dans une population donnée. Un test sanguin détecte cette maladie avec une fiabilité de 99% si elle est effectivement présente. Mais on obtient aussi un résultat faussement positif pour 0.2% des personnes saines testées. Quelle est la probabilité qu'une personne soit malade lorsque son test est positif ?

Exercice 11.16 (⇝ **corrigé**)
On dispose de 100 dés dont 25 sont truqués.
Pour chaque dé pipé, la probabilité d'obtenir le chiffre 6 lors d'un lancer vaut $\frac{1}{2}$.
On tire un dé au hasard parmi les 100 dés. On lance ce dé n fois et on obtient n fois le chiffre 6.
Quelle est la probabilité p_n que ce dé soit truqué ?

Proposition 11.1.15 (formule de Bayes, deuxième version)
Soit $(A_n)_{0 \leqslant n \leqslant m}$ un système complet fini d'événements.
Soit B un événement tel que $\mathbb{P}(B) > 0$. Pour tout n de $[\![0, m]\!]$, on a : $\mathbb{P}(A_n \mid B) = \dfrac{\mathbb{P}(A_n)\,\mathbb{P}(B \mid A_n)}{\displaystyle\sum_{0 \leqslant k \leqslant m} \mathbb{P}(A_k)\,\mathbb{P}(B \mid A_k)}$

On applique la formule précédente, et on utilise $\mathbb{P}(B) = \displaystyle\sum_{0 \leqslant k \leqslant m} \mathbb{P}(A_k)\,\mathbb{P}(B \mid A_k)$. □

Rappel : les deux premières versions de la formule de Bayes sont au programme de première année, dans le cas particulier d'un espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$ sur un univers fini Ω.

Proposition 11.1.16 (formule de Bayes, troisième version)
Soit $(A_n)_{n \geqslant 0}$ un système complet dénombrable d'événements. Soit B un événement tel que $\mathbb{P}(B) > 0$.
Pour tout n de $\mathbb{N}$, on a : $\mathbb{P}(A_n \mid B) = \dfrac{\mathbb{P}(A_n)\,\mathbb{P}(B \mid A_n)}{\displaystyle\sum_{k \geqslant 0} \mathbb{P}(A_k)\,\mathbb{P}(B \mid A_k)}$

On utilise la proposition 11.1.14, avec $\mathbb{P}(B) = \displaystyle\sum_{k=0}^{+\infty} \mathbb{P}(A_k)\,\mathbb{P}(B \mid A_k)$ (formule des probabilités totales). □

Rappel : si $\mathbb{P}(A_k) = 0$, on convient que $\mathbb{P}(A_k)\,\mathbb{P}(B \mid A_k) = 0$.

11.1.4 Indépendance

Définition 11.1.7 (indépendance de deux événements)
On dit que deux événements A et B sont *indépendants* si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

Interprétation importante : si $\mathbb{P}(B) > 0$, l'indépendance de A et B équivaut à l'égalité $\mathbb{P}(A \,|\, B) = \mathbb{P}(A)$.

Exercice 11.17 (⇝ **corrigé**)
On lance trois honnêtes. Calculer la probabilité des évènements :

1. A = « obtenir au moins un 6 »
2. B = « obtenir au moins deux fois le même chiffre »
3. C = « la somme des points marqués est paire »
4. $D = B \cap C$. Les évènements B et C sont-ils indépendants ?

Définition 11.1.8 (indépendance mutuelle d'une famille finie d'événements)
Soit $(A_i)_{0 \leqslant i \leqslant n}$ une famille finie d'événements.
On dit que $A_0, \ldots, A_n$ sont *mutuellement indépendants* si : $\forall\, J \subset [\![0, n]\!]$, $\mathbb{P}\Big(\bigcap\limits_{j \in J} A_j\Big) = \prod\limits_{j \in J} \mathbb{P}(A_j)$.

Il est clair que toute sous-famille d'événements mutuellement indépendants est à son tour constituée d'événements mutuellement indépendants.

▷ **Rapport entre « indépendance mutuelle » et « indépendance deux à deux »**

– Soit $(A_i)_{0 \leqslant i \leqslant n}$ une famille finie d'événements.
  S'ils sont mutuellement indépendants, ils sont indépendants deux à deux.
– À partir de trois événements la réciproque est fausse.
  On retiendra donc que « l'indépendance deux à deux n'entraîne pas l'indépendance mutuelle ».

  Il suffit pour s'en convaincre de connaître l'exemple de référence suivant :

  On lance deux fois de suite une pièce équilibrée. On considère les événements A : « le 1$^{\text{er}}$ lancer a renvoyé Pile », B : « le 2$^{\text{nd}}$ lancer a renvoyé Face » et C : « les deux lancers ont renvoyé le même résultat ».
  Chacun est de probabilité $\dfrac{1}{2}$, et leurs intersections deux à deux sont toutes de probabilité $\dfrac{1}{4}$.
  Mais $A \cap B \cap C = A \cap B$ donc $\mathbb{P}(A \cap B \cap C) = \dfrac{1}{4}$ est différent de $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \dfrac{1}{8}$.
  Les événements A, B, C sont donc indépendants deux à deux, mais pas mutuellement indépendants. □

Remarques

– Si l'événement A est négligeable (c'est-à-dire si $\mathbb{P}(A) = 0$) ou s'il est presque sûr (c'est-à-dire si $\mathbb{P}(A) = 1$) alors A est indépendant de tout événement B (réciproque vraie).

  Si $\mathbb{P}(A) = 0$, alors $\mathbb{P}(A \cap B) = 0$ et on a bien l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
  Si $\mathbb{P}(A) = 1$, on sait que $\mathbb{P}(A \cap B) = \mathbb{P}(B)$ et on a bien l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
  Si l'événement A est indépendant de tout événement B, il est indépendant de lui-même : on a donc les égalités $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$ et il en résulte $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$. □

– Si A et B sont indépendants, alors A et $\overline{B}$ le sont également.

En effet $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$ donc $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(\overline{B})$. □

– Soit $(A_i)_{0 \leqslant i \leqslant n}$ une famille finie d'événements, mutuellement indépendants.

Pour tout $i \in [\![0, n]\!]$, soit $B_i = A_i$ ou $B_i = \overline{A_i}$. Alors $B_0, \ldots, B_n$ sont mutuellement indépendants.

L'ordre dans lequel on considère les événements A_i n'a évidemment aucune importance.

Si dans la famille $(A_i)_{0 \leqslant i \leqslant n}$ on change *un* des A_i en son contraire, l'indépendance mutuelle est conservée (définition de l'indépendance mutuelle et remarque précédente).

Il suffit ensuite de procéder par une récurrence finie sur le nombre de A_i qu'on a remplacé par leur contraire. □

Exercice 11.18 (deuxième lemme de Borel-Cantelli, ⇝ **corrigé**)

Soit $(A_n)_{n \geqslant 0}$ une suite d'événements mutuellement indépendants.

On note $p_n = \mathbb{P}(A_n)$, et on suppose que la série $\sum\limits_{n \geqslant 0} p_n$ diverge.

1. Interpréter l'événement $B = \bigcap\limits_{n=0}^{+\infty} \Big(\bigcup\limits_{k=n}^{+\infty} A_k \Big)$.
2. Exprimer $\mathbb{P}\Big(\bigcup\limits_{k=n}^{m} A_k \Big)$ en fonction des p_k.
3. Montrer que la série de terme général $\ln(1 - p_n)$ est divergente.
4. En déduire que pour tout $n \geqslant 0$ on a : $\mathbb{P}\Big(\bigcup\limits_{k=n}^{+\infty} A_k \Big) = 1$.
5. En déduire que $\mathbb{P}(B) = 1$.

Exercice 11.19 (⇝ **corrigé**)

Soit $(A_k)_{1 \le k \le n}$ des événements mutuellement indépendants.

Montrer que la probabilité qu'aucun des A_k ne se réalise est inférieure à $\exp\Big(- \sum\limits_{k=1}^{n} \mathbb{P}(A_k) \Big)$.

11.2 Variables aléatoires discrètes

11.2.1 Variable aléatoire discrète X sur $(\Omega, \mathscr{A})$

Définition 11.2.1 (variable aléatoire discrète sur $(\Omega, \mathscr{A})$)
Soit $\mathscr{A}$ une tribu de parties d'un ensemble Ω.

Soit X une application définie sur Ω et à valeurs dans un ensemble E.

On dit que X est une *variable aléatoire discrète* si :

a) l'ensemble $\mathrm{X}(\Omega)$ des valeurs prises par X est fini ou dénombrable ;

b) pour tout $x \in \mathrm{X}(\Omega)$, l'ensemble $\mathrm{X}^{-1}(\{x\}) = \{\omega \in \Omega,\ \mathrm{X}(\omega) = x\}$ est un événement, noté $(\mathrm{X} = x)$.

Remarques

– Le cadre naturel pour la définition d'une variable aléatoire est celui des *espaces probabilisables* $(\Omega, \mathscr{A})$.

À ce stade, il n'est donc pas nécessaire de se placer dans un *espace probabilisé* $(\Omega, \mathscr{A}, \mathbb{P})$.

En revanche, si X est une variable aléatoire sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$, on sera interessé à la connaissance des probabilités des événements $(\mathrm{X} = x)$ (à suivre...).

– On considérera le plus souvent des *variables aléatoires réelles*, c'est-à-dire à valeurs dans $\mathbb{R}$.

– La proposition (b) est équivalente à : $\forall\, x \in E,\ \mathrm{X}^{-1}(\{x\})$ est un événement.
En effet, si x n'est pas dans $\mathrm{X}(\Omega)$, alors $\mathrm{X}^{-1}(\{x\})$ est l'événement vide.

– Toute application constante sur Ω est une variable aléatoire discrète.

– Si l'ensemble Ω est fini (programme de première année), ou même dénombrable, la proposition (a) est sans objet, car automatiquement vérifiée.

De même si $\mathscr{A} = \mathcal{P}(\Omega)$ (ce qui est quasiment toujours le cas si Ω est fini ou dénombrable), la proposition (b) est sans objet.

Retenons que si Ω est fini ou dénombrable, toute application $\mathrm{X} : \Omega \to E$ est une variable aléatoire sur $(\Omega, \mathcal{P}(\Omega))$.

– Le programme de deuxième année se limite aux variables aléatoires *discrètes*, c'est-à-dire dont l'ensemble image est fini ou dénombrable.Il existe des variables aléatoires X dont l'ensemble image $\mathrm{X}(\Omega)$ est infini non dénombrable (par exemple un intervalle de $\mathbb{R}$ d'intérieur non vide : on parle alors de variable aléatoire réelle *continue*) mais cette notion est hors-programme.

Dans tout ce qui suit, l'expression « variable aléatoire » signifie donc « variable aléatoire discrète ».

Définition 11.2.2 (variable indicatrice d'un événement)
Soit $\mathscr{A}$ une tribu de parties d'un ensemble Ω.

Soit A un événement.
On appelle *indicatrice de l'événement* A la variable aléatoire réelle définie par $\begin{cases} \forall\, \omega \in A,\ \mathrm{X}(\omega) = 1 \\ \forall\, \omega \notin A,\ \mathrm{X}(\omega) = 0 \end{cases}$
Elle est souvent notée $\mathbb{1}_A$.

Exercice 11.20

Soit $\mathscr{A}$ une tribu de parties d'un ensemble Ω.

Soit A et B deux événements.

Exprimer à l'aide de $\mathbb{1}_A$ et $\mathbb{1}_B$ les variables indicatrices de $\overline{A}$, $A \cap B$, $A \cup B$, $A \setminus B$ et $A \Delta B$.

On trouve : $\mathbb{1}_{\overline{A}} = 1 - \mathbb{1}_A$, $\mathbb{1}_{A\cap B} = \mathbb{1}_A \mathbb{1}_B$, $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B$ et $\mathbb{1}_{A\setminus B} = \mathbb{1}_A - \mathbb{1}_A \mathbb{1}_B$.
Enfin, on peut écrire indifféremment : $\mathbb{1}_{A\Delta B} = \mathbb{1}_A + \mathbb{1}_B - 2\mathbb{1}_A \mathbb{1}_B = (\mathbb{1}_A - \mathbb{1}_B)^2 = |\mathbb{1}_A - \mathbb{1}_B|$

Proposition 11.2.1 (événements notés $(\mathrm{X} \in U)$, $\{\mathrm{X} \in U\}$)
Soit $\mathrm{X} : \Omega \to E$ *une variable aléatoire discrète sur* $(\Omega, \mathscr{A})$.
Soit U *une partie de l'ensemble image* $\mathrm{X}(\Omega)$.
Alors l'ensemble $\mathrm{X}^{-1}(U) = \{\omega \in \Omega,\ \mathrm{X}(\omega) \in U\}$ *est un événement.*
Cet événement est noté $(\mathrm{X} \in U)$, *ou encore* $\{\mathrm{X} \in U\}$.

Tout comme $\mathrm{X}(\Omega)$, l'ensemble U est fini ou dénombrable.
On peut donc l'écrire $U = \{x_n,\ n \in J \subset \mathbb{N}\}$, où les x_n sont distincts deux à deux.
Ainsi $\mathrm{X}^{-1}(U) = \displaystyle\bigcup_{n\in J} \mathrm{X}^{-1}(\{x_n\})$ est un événement (union finie ou dénombrable d'événements). □

Remarques

– La proposition précédente est vraie pour toute partie U de E. En effet $\mathrm{X}^{-1}(U) = \mathrm{X}^{-1}(U \cap \mathrm{X}(\Omega))$.
– Soit X une variable aléatoire réelle sur $(\Omega, \mathscr{A})$.
  On considère souvent les événements notés $(\mathrm{X} \leqslant a)$, $(a \leqslant \mathrm{X} \leqslant b)$, $(\mathrm{X} > a)$, etc.

11.2.2 Loi et fonction de répartition d'une variable aléatoire

Définition 11.2.3 (loi d'une variable aléatoire, première définition)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Soit $\mathrm{X} : \Omega \to E$ une variable aléatoire.
On munit l'ensemble (fini ou dénombrable) $\mathrm{X}(\Omega)$ de la tribu discrète $\mathcal{P}(\mathrm{X}(\Omega))$.
Dans ces conditions, l'application $\mathbb{P}_\mathrm{X}$ définie par : $\forall A \subset \mathrm{X}(\Omega)$, $\mathbb{P}_\mathrm{X}(A) = \mathbb{P}(\mathrm{X} \in A)$ est une probabilité.
Cette probabilité est appelée *loi de la variable aléatoire* X.

Définition 11.2.4 (loi d'une variable aléatoire, deuxième définition)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Soit $\mathrm{X} : \Omega \to E$ une variable aléatoire.
On sait que l'ensemble $\mathrm{X}(\Omega)$ est fini ou dénombrable.
On peut donc l'écrire $\mathrm{X}(\Omega) = \{x_n,\ n \in J \subset \mathbb{N}\}$, les x_n étant ici distincts deux à deux.
Les événements $(\mathrm{X} = x_n)$ constituent alors un système complet (fini ou dénombrable) d'événements.
On appelle loi de X la donnée des probabilités $p_n = \mathbb{P}(\mathrm{X} = x_n)$. On a bien sûr $\displaystyle\sum_{n\in J} p_n = 1$.

Équivalence des deux définitions

Les deux définitions précédentes sont deux façons équivalentes de se donner la loi de X.

- En effet, si on se donne la loi de X au sens de la première définition, on se donne également les probabilités $p_n = \mathbb{P}(\mathrm{X} = x_n)$ (considérer les singletons $A = \{x_n\}$ de $\mathrm{X}(\Omega)$).
- Réciproquement, si on se donne les probabilités $p_n = \mathbb{P}(\mathrm{X} = x_n)$ pour chacun des x_n de $\mathrm{X}(\Omega)$), alors on connaît la probabilité $\mathbb{P}(\mathrm{X} \in A)$ pour chaque partie A de $\mathrm{X}(\Omega)$. Il suffit en effet d'écrire $\mathbb{P}(A) = \sum\limits_{x_n \in A} \mathbb{P}(\mathrm{X} = x_n)$ (somme finie ou dénombrable).

▷ Variable aléatoire « presque constante »

En s'inspirant de la définition 11.1.5, on dira qu'une variable aléatoire X est « presque constante », ou encore « presque sûre », s'il existe un élément x de $\mathrm{X}(\Omega)$ pour lequel $\mathbb{P}(\mathrm{X} = x) = 1$. C'est notamment le cas si X est la variable indicatrice (voir définition 11.2.2) d'un événement A presque sûr.

La proposition suivante (dont la démonstration est hors-programme) est intéressante, car elle montre qu'on peut fabriquer une probabilité $\mathbb{P}$ sur le couple $(\Omega, \mathscr{A})$ à partir d'une fonction discrète $\mathrm{X} : \Omega \to E$ et de la seule donnée des $p_n = \mathbb{P}(\mathrm{X} = x_n)$.

Proposition 11.2.2 (probabilité sur $(\Omega, \mathscr{A})$ à partir des $p_n = \mathbb{P}(\mathrm{X} = x_n)$)

Soit $\mathscr{A}$ une tribu sur un ensemble Ω. Soit $\mathrm{X} : \Omega \to E$ une variable aléatoire discrète.

Notons $\mathrm{X}(\Omega) = \{x_n,\ n \in J \subset \mathbb{N}\}$, les x_n étant distincts deux à deux.

On se donne une famille $(p_n)_{n \in J}$ d'éléments du segment $[0, 1]$, telle que $\sum\limits_{n \in J} p_n = 1$.

Alors il existe (au moins) une probabilité $\mathbb{P}$ sur $(\Omega, \mathscr{A})$ telle que : $\forall\, n \in J,\ \mathbb{P}(\mathrm{X} = x_n) = p_n$.

La proposition suivante (qui n'est pas explicitement au programme) va plus loin, car elle dit qu'on peut fabriquer tout un espace probabilisé $(\Omega, \mathscr{A}, P)$ à partir d'une fonction discrète $\mathrm{X} : \Omega \to E$ et de la donnée des $p_n = \mathbb{P}(\mathrm{X} = x_n)$.

Proposition 11.2.3 (espace probabilisé $(\Omega, \mathscr{A}, P)$ déduit d'une fonction $\mathrm{X} : \Omega \to E$ discrète)

Soit un univers Ω et soit X une application définie sur Ω à valeurs dans un ensemble E.

On suppose que son ensemble image $\mathrm{X}(\Omega)$ est fini ou dénombrable.

Notons $\mathrm{X}(\Omega) = \{x_n,\ n \in J \subset \mathbb{N}\}$, les x_n étant distincts deux à deux.

On se donne enfin une famille $(p_n)_{n \in J}$ d'éléments du segment $[0, 1]$, telle que $\sum\limits_{n \in J} p_n = 1$.

On définit alors un espace probabilisé $(\Omega, \mathscr{A}, P)$ de la manière suivante :

- *les événements sont les images réciproques des parties de E, donc : $\mathscr{A} = \{\mathrm{X}^{-1}(U),\ U \subset E\}$; l'ensemble $\mathscr{A}$ est effectivement une tribu, au sens de la définition 11.1.3 ;*
- *l'application $\mathrm{X} : \Omega \to E$ devient alors une variable aléatoire, au sens de la définition 11.2.1 ;*
- *pour tout événement $A = \mathrm{X}^{-1}(U)$, on pose : $\mathbb{P}(A) = \sum\limits_{n,\ x_n \in U} p_n$ (somme finie ou dénombrable) ; l'application $\mathbb{P}$ est alors une probabilité sur $(\Omega, \mathscr{A})$, au sens de la définition 11.1.4.*

▷ **Modélisation d'une expérience aléatoire à partir d'une variable discrète**

Voici ce qu'on peut retenir des résultats précédents.

– Dans l'étude d'une expérience aléatoire, il arrive souvent qu'on soit interessé uniquement par des mesures (numériques, en général) réalisées à l'issue de l'expérience. On se place ici dans l'hypothèse où l'ensemble des valeurs qu'on peut ainsi mesurer est fini ou dénombrable.

– Soit $V = \{x_n,\ n \in J\}$, avec $J \subset \mathbb{N}$, l'ensemble de ces valeurs (distinctes deux à deux).
  À chaque résultat élémentaire ω de l'expérience, est attachée une valeur x_n, pour un certain n de J.
  Bien sûr, de nombreux résultats élémentaires peuvent conduire à une même valeur x_n.
  On dispose ainsi d'une fonction X définie sur Ω, et dont l'ensemble image est $\mathrm{X}(\Omega) = V$.

– On décide alors que les seuls événements à prendre en considération sont les images réciproques des parties de V (c'est-à-dire les collections de résultats élémentaires ω pour lesquels la « mesure effectuée » se situe dans un certain sous-ensemble de V).
  L'ensemble $\mathscr{A}$ de ces événements est alors une tribu du Ω, et la fonction X gagne par la même occasion son statut de variable aléatoire.

– On appelle « événements élémentaires liés à X » les images réciproques $\mathrm{X}^{-1}(\{x_n\})$ des singletons $\{x_n\}$, avec n dans J (c'est-à-dire les collections de résultats élémentaires ω pour lesquels la mesure effectuée $\mathrm{X}(\omega)$ prend une valeur particulière x_n).
  On note bien sûr $(\mathrm{X} = x_n)$ un tel événement élémentaire lié à X.

– Dans notre effort de modélisation de l'expérience aléatoire, on attribue une probabilité p_n à chaque événement élémentaire lié à X. Il faut bien sûr que les p_n soient dans $[0, 1]$ et que leur somme (finie ou dénombrable) soit égale à 1. On est alors en mesure de définir la probabilité d'un événement quelconque A, comme indiqué à la fin de la proposition 11.2.3

Exercice 11.21 (⇝ **corrigé**)
Soit quatre cases numérotées de 1 à 4. À l'étape $n = 0$, un jeton est placé sur la case 1. Si à l'étape n, le jeton se trouve sur une case numérotée $k \in \{2, 3, 4\}$, il est déplacé sur la case $k - 1$; sinon, on le place au hasard (de manière équiprobable) sur l'une des cases $2, 3, 4$. On note, pour tout n, X_n la variable aléatoire donnant la position du jeton à la n-ième étape et U_n le vecteur colonne dont les coefficients sont $P(X_n = i)$ pour $1 \leq i \leq 4$.

1. Écrire une fonction d'argument n et renvoyant les positions successives d'une réalisation de cette expérience. Tracer ces positions pour $n = 10$, $n = 50$ et $n = 100$. Que peut-on conjecturer ?
2. Montrer qu'il existe $A \in \mathcal{M}_4(\mathbb{R})$ telle que $U_{n+1} = AU_n$ pour tout n.
3. Diagonaliser A ; en déduire que la suite (U_n) converge et préciser sa limite.
4. Pour $i \in \{1, 2, 3, 4\}$ et $n \in \mathbb{N}$, soit $Y_n(i) = \operatorname{card}\{k \leq n, X_k = i\}$. Chercher la loi de $Y_n(i)$.

Définition 11.2.5 (fonction de répartition d'une variable aléatoire réelle)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, et soit $\mathrm{X} : \Omega \to \mathbb{R}$ une variable aléatoire réelle.
La fonction $\mathrm{F}_{\mathrm{X}} : \mathbb{R} \to [0, 1]$ définie par $\mathrm{F}_{\mathrm{X}}(x) = \mathbb{P}(\mathrm{X} \leqslant x)$ est appelée *fonction de répartition* de X.

Proposition 11.2.4 (croissance et limites en $\pm\infty$ d'une fonction de répartition)
Soit F_{X} *la fonction de répartition d'une variable aléatoire* *réelle* X.
Alors la fonction F_{X} *est croissante sur* $\mathbb{R}$. *Elle vérifie* $\lim\limits_{x\to-\infty} \mathrm{F}_{\mathrm{X}}(x) = 0$ *et* $\lim\limits_{x\to+\infty} \mathrm{F}_{\mathrm{X}}(x) = 1$.

Si $x \leqslant y$, on a $(\mathrm{X} \leqslant x) \subset (\mathrm{X} \leqslant y)$ donc $\mathbb{P}(\mathrm{X} \leqslant x) \leqslant \mathbb{P}(\mathrm{X} \leqslant y)$: c'est la croissance de l'application $\mathrm{F_X}$.
Cette monotonie implique l'existence des limites $\lim\limits_{x\to-\infty} \mathrm{F_X}(x)$ et $\lim\limits_{x\to+\infty} \mathrm{F_X}(x)$.
Pour tout n de $\mathbb{N}$, on pose $A_n = (X \leqslant n)$ et $B_n = (X \leqslant -n)$. Ainsi $\mathrm{F_X}(n) = \mathbb{P}(A_n)$ et $\mathrm{F_X}(-n) = \mathbb{P}(B_n)$.
La suite $(A_n)_{n\geqslant 0}$ est croissante et $\displaystyle\bigcup_{n=0}^{+\infty} A_n = \Omega$. La suite $(B_n)_{n\geqslant 0}$ est décroissante et $\displaystyle\bigcap_{n=0}^{+\infty} B_n = \emptyset$.
Avec les propriétés de continuité croissante (cf 11.1.8) et de continuité décroissante (cf 11.1.9), on trouve :

$$\lim_{n\to+\infty} P(A_n) = \mathbb{P}\Big(\bigcup_{n=0}^{+\infty} A_n\Big) = \mathbb{P}(\Omega) = 1 \quad \text{et} \quad \lim_{n\to+\infty} P(B_n) = \mathbb{P}\Big(\bigcap_{n=0}^{+\infty} B_n\Big) = \mathbb{P}(\emptyset) = 0$$

Ainsi : $\lim\limits_{x\to+\infty} \mathrm{F_X}(x) = \lim\limits_{n\to+\infty} \mathrm{F_X}(n) = \lim\limits_{n\to+\infty} P(A_n) = 1$ et $\lim\limits_{x\to-\infty} \mathrm{F_X}(x) = \lim\limits_{n\to+\infty} \mathrm{F_X}(-n) = \lim\limits_{n\to+\infty} P(B_n) = 0$ □

▷ **Rappels de première année**

– On se place ici sur un espace probabilisé $(\Omega, \mathcal{P}(\Omega), P)$, avec Ω fini.
  Soit X une variable aléatoire réelle et soit $\mathrm{F_X}$ sa fonction de répartition.
  L'ensemble $\mathrm{X}(\Omega)$ est fini, et on peut l'écrire $\mathrm{X}(\Omega) = \{x_0, x_1, \ldots, x_n\}$, avec $x_0 < x_1 < \cdots < x_n$.

– La fonction de répartition $\mathrm{F_X}$ est en escaliers sur $\mathbb{R}$.
  Plus précisément : $\begin{cases} \forall\, x < x_0,\ \mathrm{F_X}(x) = 0 \\ \forall\, x \geqslant x_n,\ \mathrm{F_X}(x) = 1 \end{cases}$, et $\mathrm{F_X}(x) = \displaystyle\sum_{j=0}^{k} \mathbb{P}(\mathrm{X}=x_j)$ sur $[x_k, x_{k+1}[$ (avec $k \in [\![0, n[\![$).
  Inversement : $\mathbb{P}(x_0) = \mathrm{F_X}(x_0)$ et : $\forall\, k \in [\![1, n]\!]$, $\mathbb{P}(\mathrm{X}=x_k) = \mathrm{F_X}(x_k) - \mathrm{F_X}(x_{k-1})$.
  Cette égalité permet donc de retrouver la loi de X à partir de sa fonction de répartition.
  La fonction $\mathrm{F_X}$ est discontinue en chaque point x_k où $\mathbb{P}(\mathrm{X}=x_k) > 0$, et la valeur de $\mathbb{P}(\mathrm{X}=x_k)$ est alors le « saut de discontinuité » de la fonction $\mathrm{F_X}$ en ce point.

Remarques

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, et soit $\mathrm{X} : \Omega \to \mathbb{R}$ une variable aléatoire réelle.
(rappel : conformément au programme, toutes les variables aléatoires sont à image finie ou dénombrable).
Soit $\mathrm{F_X}$ la fonction de répartition de X.

– Pour $a < b$ on a : $\mathbb{P}(\mathrm{X} > a) = 1 - \mathrm{F_X}(a)$ et $\mathrm{F_X}(b) - \mathrm{F_X}(a) = \mathbb{P}(a < \mathrm{X} \leqslant b)$.

  C'est évident car $(\mathrm{X} > a) = \overline{(\mathrm{X} \leqslant a)}$ donc $\mathbb{P}(\mathrm{X} > a) = 1 - \mathbb{P}(\mathrm{X} \leqslant a) = 1 - \mathrm{F_X}(a)$.
  Ensuite $(\mathrm{X} \leqslant b) = (\mathrm{X} \leqslant a) \cup (a < \mathrm{X} \leqslant b)$ (union disjointe).
  Il en résulte : $\mathbb{P}(\mathrm{X} \leqslant b) - \mathbb{P}(\mathrm{X} \leqslant a) = \mathbb{P}(a < \mathrm{X} \leqslant b)$. □

– Pour tout a de $\mathbb{R}$, on a : $\mathbb{P}(\mathrm{X}=a) = \mathrm{F_X}(a) - \lim\limits_{x\to a^-} \mathrm{F_X}(x)$.

  Puisque $\mathrm{F_X}$ est monotone, on sait qu'elle a des limites finies à gauche et à droite en a.
  On pose $A_n = \Big(X \leqslant a + \dfrac{1}{n}\Big)$ et $B_n = \Big(X \leqslant a - \dfrac{1}{n}\Big)$. Ainsi $\mathbb{P}(A_n) = \mathrm{F_X}\Big(a + \dfrac{1}{n}\Big)$ et $\mathbb{P}(B_n) = \mathrm{F_X}\Big(a - \dfrac{1}{n}\Big)$.
  $(A_n)_{n\geqslant 0}$ est décroissante et $\displaystyle\bigcap_{n=0}^{+\infty} A_n = (X \leqslant a)$. La suite $(B_n)_{n\geqslant 0}$ est croissante et $\displaystyle\bigcup_{n=0}^{+\infty} B_n = (X < a)$.
  Par continuité décroissante : $\lim\limits_{x\to a+} \mathrm{F_X} = \lim\limits_{n\to+\infty} \mathrm{F_X}\Big(a+\dfrac{1}{n}\Big) = \lim\limits_{n\to+\infty} P(A_n) = \mathbb{P}\Big(\displaystyle\bigcap_{n=0}^{+\infty} A_n\Big) = \mathbb{P}(\mathrm{X}\leqslant a) = \mathrm{F_X}(a)$.

Par continuité croissante : $\lim\limits_{x\to a-} \mathrm{F_X} = \lim\limits_{n\to+\infty} \mathrm{F_X}\Big(a-\dfrac{1}{n}\Big) = \lim\limits_{n\to+\infty} P(B_n) = \mathbb{P}\Big(\bigcup\limits_{n=0}^{+\infty} B_n\Big) = \mathbb{P}(\mathrm{X} < a)$.

On constate que $\lim\limits_{x\to a+} \mathrm{F_X} = \mathrm{F_X}(a)$: la fonction de répartition est continue à droite en tout point.

Enfin, avec $\mathbb{P}(\mathrm{X} \leqslant a) = \mathbb{P}(\mathrm{X} = a) + \mathbb{P}(\mathrm{X} < a)$, on a obtenu : $\mathbb{P}(\mathrm{X} = a) = \mathrm{F_X}(a) - \lim\limits_{x\to a-} \mathrm{F_X}$. □

On peut donc a priori retrouver la loi de X à partir de sa fonction de répartition.

Les « sauts de discontinuité » de $\mathrm{F_X}$ correspondent aux valeurs du réel a telles que $\mathbb{P}(\mathrm{X}=a) > 0$.

▷ Image d'une variable aléatoire par une fonction, loi associée

On étend ici des résultats abordés en première année.

Proposition 11.2.5 (fonction d'une variable aléatoire)
Soit $\mathrm{X} : \Omega \to E$ *une variable aléatoire discrète sur* $(\Omega, \mathscr{A})$.
Soit φ *une fonction définie sur* E *(en tout cas sur* $\mathrm{X}(\Omega)$*) et à valeurs dans un ensemble* F.
L'application $\mathrm{Y} = \varphi \circ \mathrm{X}$ *(par convention notée* $\varphi(\mathrm{X})$*) est une variable aléatoire discrète sur* $(\Omega, \mathscr{A})$.

Tout comme $\mathrm{X}(\Omega)$, l'ensemble $\mathrm{Y}(\Omega) = \varphi(\mathrm{X}(\Omega))$ est fini ou dénombrable.
Pour tout $y \in \mathrm{Y}(\Omega)$: $\mathrm{Y}^{-1}(\{y\}) = \mathrm{X}^{-1}(\varphi^{-1}\{y\}) = \mathrm{X}^{-1}(U)$ avec $U = \mathrm{X}(\Omega) \cap \varphi^{-1}\{y\}$.
Mais U est une partie de $\mathrm{X}(\Omega)$, donc $\mathrm{X}^{-1}(U)$ est un événement (voir proposition 11.2.1).
Ainsi $\mathrm{Y}^{-1}(\{y\})$ est un événement pour tout y de $\mathrm{Y}(\Omega)$, ce qu'il fallait démontrer. □

On sait que la loi $\mathbb{P}_\mathrm{Y}$ de Y est une probabilité sur $(\mathrm{Y}(\Omega), \mathcal{P}(\mathrm{Y}(\Omega)))$ (voir définition 11.2.3).

Pour tout $B \subset \mathrm{Y}(\Omega)$, on a : $\mathbb{P}_\mathrm{Y}(B) = \mathbb{P}(\mathrm{Y} \in B) = \mathbb{P}(\varphi(\mathrm{X}) \in B) = \mathbb{P}(\mathrm{X} \in \varphi^{-1}(B)) = \mathbb{P}_\mathrm{X}(\varphi^{-1}(B))$.

En particulier : $\forall\, y_n \in \mathrm{Y}(\Omega),\ \mathbb{P}(\mathrm{Y}=y_n) = \sum\limits_{\varphi(x_k)=y_n} \mathbb{P}(\mathrm{X}=x_k)$.

Les cas usuels de fonctions d'une variable aléatoire sont $\mathrm{Y} = a\mathrm{X} + b$, $\mathrm{Y} = \mathrm{X}^2$, $\mathrm{Y} = |\mathrm{X}|$, etc.
Si $\mathrm{Y} = \mathrm{X}^2$, par exemple, on écrira : $\forall\, a \in \mathbb{R}^{+*},\ \mathbb{P}(\mathrm{Y} = a) = \mathbb{P}(\mathrm{X} = \sqrt{a}) + \mathbb{P}(\mathrm{X} = -\sqrt{a})$.

11.2.3 Trois lois usuelles (rappels de première année)

Dans les exemples qui suivent, on considère une variable aléatoire réelle X attachée à une expérience aléatoire dont on note Ω l'« univers des possibles ».

Il n'est pas utile de spécifier l'espace probabilisé $(\Omega, \mathscr{A}, P)$ sur lequel on travaille. Celui qu'on peut former à partir de X, comme on l'a vu dans la proposition 11.2.3, convient très bien.

▷ Loi uniforme sur un intervalle de n entiers consécutifs

Soit $[\![a, b]\!]$ un intervalle de n entiers consécutifs (donc $b = a + n - 1$)

Le plus souvent, il s'agit de $[\![0, n-1]\!]$ ou de $[\![1, n]\!]$.

On dit que X suit la loi uniforme sur $[\![a, b]\!]$, et on note $\mathrm{X} \hookrightarrow \mathcal{U}([\![a, b]\!])$ si $\begin{cases} \mathrm{X}(\Omega) = [\![a, b]\!] = [\![a, a+n-1]\!] \\ \forall\, k \in [\![a, b]\!],\ \mathbb{P}(\mathrm{X}=k) = \dfrac{1}{n} \end{cases}$

Remarque : si $\mathrm{X} \hookrightarrow \mathcal{U}([\![a, b]\!])$, alors $\mathrm{Y} = \mathrm{X}-a+1 \hookrightarrow \mathcal{U}([\![1, n]\!])$.

▷ **Loi de Bernoulli** $\mathcal{B}(p)$ **avec** $0 \leqslant p \leqslant 1$

Soit p dans $[0, 1]$. Dans ce contexte, on note souvent $q = 1 - p$.

On dit que X suit la loi de Bernoulli de paramètre p, et on note $X \hookrightarrow \mathcal{B}(p)$ si $\begin{cases} X(\Omega) = \{0, 1\} \\ \mathbb{P}(X=1) = p \\ \mathbb{P}(X=0) = 1 - p \end{cases}$

On modélise ainsi une « épreuve de Bernoulli de paramètre p », c'est-à-dire une expérience aléatoire qui ne possède deux issues possibles : le « succès » (ici la valeur 1) avec la probabilité p, et l'« échec » (ici la valeur 0) avec la probabilité $q = 1 - p$.

Avec ce modèle, on peut dire que X est la variable indicatrice de l'événement « l'expérience s'est soldée par un succès » (voir définition 11.2.2).

Remarque : si $X \hookrightarrow \mathcal{B}(p)$, alors $Y = 1 - X \hookrightarrow \mathcal{B}(q)$.

▷ **Loi binomiale** $\mathcal{B}(n, p)$ **de paramètres** $n \in \mathbb{N}^*$ **et** $p \in [0, 1]$

Soit n dans $\mathbb{N}^*$. Soit p dans $[0, 1]$ (dans ce contexte, on note souvent $q = 1 - p$).

On dit que X suit la loi binomiale de paramètres n, p, et on note $X \hookrightarrow \mathcal{B}(n, p)$ si :

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

On modélise ainsi une expérience aléatoire consistant en la répétition de n épreuves de Bernoulli indépendantes de même paramètre p (l'indépendance est une notion qui sera évoquée plus loin et on la prend ici dans son sens intuitif : les conditions de l'expérience sont telles que le résultat d'une de ces épreuves de Bernoulli « n'influe » pas sur celui des autres épreuves).

Avec ce modèle, on peut dire que X est la variable qui compte le nombre de succès dans cette répétition de n épreuves de Bernoulli indépendantes de même paramètre p.

Un autre modèle où intervient la loi $\mathcal{B}(n, p)$ est l'expérience aléatoire consistant à effectuer n tirages successifs « avec remise » d'une boule dans une urne contenant une proportion p de boules blanches, et de compter le nombre de boules blanches obtenues après ces n tirages. Le fait que le tirage s'effectue avec remise est important pour garantir l'indépendance des résultats des tirages successifs.

Remarque : si $X \hookrightarrow \mathcal{B}(n, p)$, alors $Y = n - X \hookrightarrow \mathcal{B}(n, q)$ (on rappelle que $q = 1 - p$).

11.3 Couples de variables aléatoires discrètes

11.3.1 Couple de variables aléatoires discrètes

Proposition 11.3.1 (couple formé par deux variables aléatoires discrètes)
Soit $\mathscr{A}$ une tribu sur un ensemble Ω.
Soit $\mathrm{X} : \Omega \to E$ et $\mathrm{Y} : \Omega \to F$ deux variables aléatoires discrètes.
En posant : $\forall\, \omega \in \Omega,\ \mathrm{Z}(\omega) = (\mathrm{X}(\omega), \mathrm{Y}(\omega))$, on définit une variable aléatoire discrète $\mathrm{Z} : \Omega \to E \times F$.

- Tout d'abord les ensembles $\mathrm{X}(\Omega)$ et $\mathrm{Y}(\Omega)$ sont finis ou dénombrables.
  Il en est donc de même de $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$, et a fortiori de $\mathrm{Z}(\Omega)$, qui est lui-même inclus dans $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$.
- Soit $z = (x, y)$ un élément de $\mathrm{Z}(\Omega)$.
  Pour tout ω de Ω, on a $\mathrm{Z}(\omega) = z \Leftrightarrow \begin{cases} \mathrm{X}(\omega) = x \\ \mathrm{Y}(\omega) = y \end{cases} \Leftrightarrow \omega \in \mathrm{X}^{-1}(\{x\}) \cap \mathrm{Y}^{-1}(\{y\})$.
  Mais $\mathrm{X}^{-1}(\{x\})$ et $\mathrm{Y}^{-1}(\{y\})$ sont des événements. Il en est donc de même de $\mathrm{Z}^{-1}(\{z\})$
- Conclusion : Z est une variable aléatoire discrète sur $(\Omega, \mathscr{A})$ (revoir la définition 11.2.1). □

La proposition précédente admet une réciproque :

Proposition 11.3.2 (composantes d'une variable aléatoire à valeurs dans $E \times F$)
Soit $\mathscr{A}$ une tribu sur un ensemble Ω.
Soit $\mathrm{Z} : \Omega \to E \times F$ une variable aléatoire discrète.
Soit $\mathrm{X} : \Omega \to E$ et $\mathrm{Y} : \Omega \to F$ les composantes de Z, définies par : $\forall\, \omega \in \Omega,\ \mathrm{Z}(\omega) = (\mathrm{X}(\omega), \mathrm{Y}(\omega))$.
Alors les applications X et Y sont des variables aléatoires discrètes.

Soit π_1 la projection de $\mathrm{X} \times \mathrm{Y}$ sur X, définie par $\pi_1(x, y) = x$.
De même, soit π_2 la projection de $\mathrm{X} \times \mathrm{Y}$ sur Y.
On a les égalités $\mathrm{X} = \pi_1(\mathrm{Z})$ et $\mathrm{Y} = \pi_2(\mathrm{Z})$, et il suffit d'appliquer la proposition 11.2.5. □

Définition 11.3.1 (couple de variables aléatoires)
Soit $\mathscr{A}$ une tribu sur un ensemble Ω.
Dans les propositions 11.3.1 et 11.3.2, on dit que $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ est un *couple de variables aléatoires.*

On retiendra qu'il importe peu que X, Y préexistent à Z ou au contraire s'en déduisent.
On remarquera également que $\mathrm{Z}(\Omega)$ est une partie (souvent stricte) de $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$.
Pour tout (x, y) de $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$, l'événement $(\mathrm{Z} = (x, y))$ sera noté $(\mathrm{X}{=}x, \mathrm{Y}{=}y)$.

▷ Combinaisons linéaires de variables aléatoires réelles discrètes

Soit $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ un couple de variables aléatoires réelles sur $(\Omega, \mathscr{A})$.
Pour tous scalaires λ, μ, l'application $\lambda\mathrm{X} + \mu\mathrm{Y}$ est une variable aléatoire réelle.
Il suffit en effet d'appliquer à $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ la fonction φ définie sur $\mathbb{R}^2$ par $\varphi(x, y) = \lambda x + \mu y$.
L'ensemble des variables aléatoires réelles sur $(\Omega, \mathscr{A})$ est donc un $\mathbb{R}$-espace vectoriel.
Remarque : pour des raisons similaires, l'application XY est une variable aléatoire réelle sur $(\Omega, \mathscr{A})$.

▷ **Vecteurs aléatoires**

On peut généraliser ce qui précède à une application de Ω dans un produit cartésien $F_1 \times \cdots \times F_n$.

Si on note $\mathrm{Z} = (\mathrm{X}_1, \ldots, \mathrm{X}_n)$, dire que Z est une variable aléatoire discrète sur $(\Omega, \mathscr{A})$ équivaut à dire que chacune des applications $\mathrm{X}_i : \Omega \to F_i$ est elle-même une variable aléatoire discrète.

On dit alors que Z est un vecteur de variables aléatoires.

On peut aussi définir une variable aléatoire fonction $\varphi(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)$ des n variables aléatoires X_i.

11.3.2 Loi conjointe, loi marginale, loi conditionnelle

Définition 11.3.2 (loi conjointe d'un couple de variables aléatoires discrètes)
Soit X et Y deux variables aléatoires sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.
On appelle *loi conjointe* de X et Y la loi de la variable aléatoire $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$.
C'est donc la probabilité $\mathbb{P}_\mathrm{Z}$ définie sur $(\mathrm{Z}(\Omega), \mathcal{P}(\mathrm{Z}(\Omega)))$ par : $\forall\, A \subset \mathrm{Z}(\Omega),\ \mathbb{P}_\mathrm{Z}(A) = \mathbb{P}(\mathrm{Z}^{-1}(A))$.

▷ **Autres présentations de la loi conjointe**

On a utilisé ici la loi d'une variable aléatoire au sens de la définition 11.2.3.

Si on utilise la définition 11.2.4, on peut donc dire :

$\boxed{\textit{La loi conjointe de } \mathrm{X} \textit{ et } \mathrm{Y} \textit{ est la donnée des probabilités } \mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y) \textit{ pour tout } (x,y) \textit{ de } \mathrm{Z}(\Omega)}$

On se souvient que $\mathrm{Z}(\Omega)$ n'est qu'une partie de $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$.

Par ailleurs, il est clair que si (x, y) est dans le complémentaire de $\mathrm{Z}(\Omega)$ dans $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$, l'événement $(\mathrm{X}{=}x, \mathrm{Y}{=}y)$ est impossible donc de probabilité nulle.

On peut donc également adopter la définition suivante :

$\boxed{\textit{La loi conjointe de } \mathrm{X} \textit{ et } \mathrm{Y} \textit{ est la donnée des } \mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y) \textit{ pour tout } (x,y) \textit{ de } \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)}$

Définition 11.3.3 (lois marginales d'un couple de variables aléatoires discrètes)
Soit $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ un couple de variables aléatoires sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.
Les lois de X et de Y sont appelées *lois marginales* de Z.

▷ **La loi conjointe détermine les lois marginales**

Soit $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ un couple de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Proposition 11.3.3 (de la loi conjointe aux lois marginales)
Soit $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ un couple de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.
On se donne la loi conjointe de X et Y, donc les probabilités $\mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y)$ pour $(x, y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$.
Pour tout x de $\mathrm{X}(\Omega)$, on a alors : $\mathbb{P}(\mathrm{X}{=}x) = \displaystyle\sum_{y \in \mathrm{Y}(\Omega)} \mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y)$ (première loi marginale de Z).
De même : $\forall\, y \in \mathrm{Y}(\Omega),\ \mathbb{P}(\mathrm{Y}{=}y) = \displaystyle\sum_{x \in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y)$ (deuxième loi marginale de Z).

Les événements $(X{=}x)$ (quand x décrit $X(\Omega)$) forment un système complet (fini ou dénombrable).
Il en est de même des événements $(Y{=}y)$ quand y décrit $Y(\Omega)$.
Avec la formule des probabilités totales (voir proposition 11.1.13), on obtient effectivement :

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

On voit donc qu'on obtient les deux lois marginales de $Z = (X, Y)$ à partir de la loi conjointe de X et Y, par « sommation dans les marges » (d'où le nom). □

▷ Les lois marginales ne déterminent pas la loi conjointe

Soit $Z = (X, Y)$ un couple de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.

On suppose connues les lois marginales de Z, c'est-à-dire les lois de X et de Y, c'est-à-dire les probabilités $\mathbb{P}(X{=}x)$ pour tout x de $X(\Omega)$ et $\mathbb{P}(Y{=}y)$ pour tout y de $Y(\Omega)$.

Ces informations ne sont pas suffisantes pour retrouver la loi conjointe de X et de Y.

Un exemple simple : on pose $Z = (X, Y)$ avec $X(\Omega) = Y(\Omega) = \{0, 1\}$ et la loi conjointe :

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

On voit que $\mathbb{P}(X{=}0) = \mathbb{P}(X{=}1) = \dfrac{1}{2}$ et $\mathbb{P}(Y{=}0) = \mathbb{P}(Y{=}1) = \dfrac{1}{2}$.

Ici la simple connaissance des lois marginales du couple (X, Y) (toutes les deux uniformes sur $\{0, 1\}$) ne permet pas de reconstituer la loi conjointe de X et Y (car le paramètre p est quelconque dans $[0, \frac{1}{2}]$).

Définition 11.3.4 (loi conditionnelle de Y sachant $(X{=}x)$)
Soit $Z = (X, Y)$ un couple de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Soit x un élément de $X(\Omega)$ tel que $\mathbb{P}(X{=}x) > 0$.
Pour tout élément y de $Y(\Omega)$, on pose $\mathbb{P}_{(X=x)}(Y{=}y) = \dfrac{\mathbb{P}(X{=}x, Y{=}y)}{\mathbb{P}(X{=}x)}$.

Alors $\mathbb{P}_{(X=x)}$ est une probabilité sur $Y(\Omega)$, dite « probabilité conditionnelle de Y sachant $(X{=}x)$ ».

Remarques

– On est ici dans une situation analogue à celle abordée dans la proposition 11.1.11

– On peut bien sûr utiliser la notation : $\mathbb{P}(Y{=}y \mid X{=}x) = \dfrac{\mathbb{P}(X{=}x, Y{=}y)}{\mathbb{P}(X{=}x)}$

– Soit $y \in Y(\Omega)$ tel que $\mathbb{P}(Y{=}y) > 0$. La probabilité conditionnelle de X sachant $(Y{=}y)$ est :
$\forall\, x \in X(\Omega),\ \mathbb{P}_{(Y=y)}(X{=}x) = \mathbb{P}(X{=}x \mid Y{=}y) = \dfrac{\mathbb{P}(X{=}x, Y{=}y)}{\mathbb{P}(Y{=}y)}$.

Proposition 11.3.4 (des lois conditionnelles aux lois marginales)
On va réécrire la proposition 11.3.3 en utilisant les lois conditionnelles.

Soit $Z = (X, Y)$ un couple de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Pour tout x de $X(\Omega)$, on a : $\mathbb{P}(X{=}x) = \displaystyle\sum_{y\in Y(\Omega)} \mathbb{P}(Y{=}y)\,\mathbb{P}(X{=}x \mid Y{=}y)$.

Pour tout y de $Y(\Omega)$, on a : $\mathbb{P}(Y{=}y) = \displaystyle\sum_{x\in X(\Omega)} \mathbb{P}(X{=}x)\,\mathbb{P}(Y{=}y \mid X{=}x)$.

Voir les remarques qui suivent la proposition 11.1.11, notamment sur l'égalité $\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A \mid B)$.
On a les égalités $\mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y) = \mathbb{P}(\mathrm{Y}{=}y)\,\mathbb{P}(\mathrm{X}{=}x \mid \mathrm{Y}{=}y) = \mathbb{P}(\mathrm{X}{=}x)\,\mathbb{P}(\mathrm{Y}{=}y \mid \mathrm{X}{=}x)$.
Il s'agit donc ici d'une simple réécriture de la proposition 11.3.3. □

Rappel : si $\mathbb{P}(\mathrm{Y}{=}y) = 0$ on convient que $\mathbb{P}(\mathrm{Y}{=}y)\,\mathbb{P}(\mathrm{X}{=}x \mid \mathrm{Y}{=}y) = 0$.
Cette convention évite d'avoir à supposer que toutes les probabilités $\mathbb{P}(\mathrm{Y}{=}y)$ sont strictement positives dans la première des égalités de la proposition précédente.
C'est bien sûr la même chose avec la seconde égalité à chaque fois que $\mathbb{P}(\mathrm{X}{=}x) = 0$.

11.3.3 Couples de variables aléatoires indépendantes

Définition 11.3.5 (indépendance de deux variables aléatoires)
Soit $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ un couple de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.
On dit que X et Y sont *indépendantes* si, $\forall\,(x, y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega),\ \mathbb{P}(\mathrm{X}{=}x, \mathrm{Y}{=}y) = \mathbb{P}(\mathrm{X}{=}x)\,\mathbb{P}(\mathrm{Y}{=}y)$.

Remarques

– L'indépendance de X et Y implique que, pour tout x de $\mathrm{X}(\Omega)$ tel que $\mathbb{P}(\mathrm{X}{=}x) > 0$, la loi conditionnelle de Y sachant $(\mathrm{X}{=}x)$ est égale à la loi de Y.
Autrement dit : $\forall\,(x, y) \in \mathrm{X}(\omega)$ avec $\mathbb{P}(\mathrm{X}{=}x) > 0$, $\forall\, y \in \mathrm{Y}(\omega),\ \mathbb{P}_{(\mathrm{X}=x)}(\mathrm{Y}{=}y) = \mathbb{P}(\mathrm{Y}{=}y)$.
– L'indépendance de X et Y signifie que, pour tout (x, y) de $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$, les événements $(\mathrm{X}{=}x)$ et $(\mathrm{Y}{=}y)$ sont indépendants.
– En général, l'indépendance de deux variables aléatoires résulte du modèle décrivant l'expérience. C'est plus une hypothèse a priori qu'une conséquence d'un calcul.
– Si une variable aléatoire X est constante (ou seulement « presque constante », cf sous-section 11.2.2) elle est indépendante de toute autre variable aléatoire (on pourra se reporter à ce sujet à la deuxième des trois remarques concluant la sous-section 11.1.4).

Proposition 11.3.5 (caractérisation de l'indépendance de deux variables aléatoires)
Soit $\mathrm{Z} = (\mathrm{X}, \mathrm{Y})$ *un couple de variables aléatoires indépendantes sur* $(\Omega, \mathscr{A}, \mathbb{P})$.
Alors X, Y *sont indépendantes* $\Leftrightarrow\ \ \forall\, A \subset \mathrm{X}(\Omega),\ \forall\, B \subset \mathrm{Y}(\Omega),\ \mathbb{P}(\mathrm{X}{\in}A, \mathrm{Y}{\in}B) = \mathbb{P}(\mathrm{X}{\in}A)\,\mathbb{P}(\mathrm{Y}{\in}B)$.

– On suppose que $\mathbb{P}(\mathrm{X}{\in}A, \mathrm{Y}{\in}B) = \mathbb{P}(\mathrm{X}{\in}A)\,\mathbb{P}(\mathrm{Y}{\in}B)$ pour tout $A \subset \mathrm{X}(\omega)$ et tout $B \subset \mathrm{Y}(\omega)$.
C'est vrai en particulier pour les singletons $A = \{x\}$ et $B = \{y\}$ pour tout (x, y) de $\mathrm{X}(\Omega) \times \mathrm{Y}(\Omega)$.
On a donc : $\forall\,(x, y) \in \mathrm{X}(\Omega) \times \mathrm{Y}(\Omega),\ \mathbb{P}(\mathrm{X}=x, \mathrm{Y}=y) = \mathbb{P}(\mathrm{X}=x)\,\mathbb{P}(\mathrm{Y}=y)$, ce qui est la définition de l'indépendance de X et Y.
– Réciproquement, on suppose que X et Y sont indépendantes, au sens de la définition 11.3.5.
Soit A une partie de $\mathrm{X}(\Omega)$ et soit B une partie de $\mathrm{Y}(\Omega)$.
Tout comme $\mathrm{X}(\Omega)$ et $\mathrm{Y}(\Omega)$, les ensembles A et B sont finis ou dénombrables.
On pose $A = \{x_j,\ j \in J \subset \mathbb{N}\}$ (les x_j distincts) et $B = \{y_k,\ k \in K \subset \mathbb{N}\}$ (les y_k distincts).
On alors, successivement :

$$\begin{aligned}
\mathbb{P}(\mathrm{X}\in A,\mathrm{Y}\in B) &= \mathbb{P}\Big(\bigcup_{j\in J}(\mathrm{X}=x_j)\ \cap\ \bigcup_{k\in K}(\mathrm{Y}=y_k)\Big) = \mathbb{P}\Big(\bigcup_{j\in J}\Big(\bigcup_{k\in K}(\mathrm{X}=x_j)\cap(\mathrm{Y}=y_k)\Big)\Big)\\
&= \sum_{j\in J}\mathbb{P}\Big(\bigcup_{k\in K}(\mathrm{X}=x_j)\cap(\mathrm{Y}=y_k)\Big) \quad \text{(par utilisation de la σ-additivité sur l'indice j)}\\
&= \sum_{j\in J}\sum_{k\in K}\mathbb{P}(\mathrm{X}=x_j,\mathrm{Y}=y_k) \quad \text{(par utilisation de la σ-additivité sur l'indice k)}\\
&= \sum_{j\in J}\sum_{k\in K}\mathbb{P}(\mathrm{X}=x_j)\,\mathbb{P}(\mathrm{Y}=y_k) \quad \text{(indépendance de X et Y)}\\
&= \sum_{j\in J}\mathbb{P}(\mathrm{X}=x_j)\ \sum_{k\in K}\mathbb{P}(\mathrm{Y}=y_k) = \mathbb{P}(\mathrm{X}\in A)\,\mathbb{P}(\mathrm{Y}\in B) \qquad \square
\end{aligned}$$

Proposition 11.3.6 (fonctions de deux variables aléatoires indépendantes)
Soit $\mathrm{Z}=(\mathrm{X},\mathrm{Y})$ *un couple de variables aléatoires indépendantes sur* $(\Omega,\mathscr{A},\mathbb{P})$.
Soit $\mathrm{X}'=\varphi(\mathrm{X})$ *une fonction de* X *et* $\mathrm{Y}'=\psi(\mathrm{Y})$ *une fonction de* Y *(voir proposition 11.2.5).*
Alors les variables aléatoires X' *et* Y' *sont indépendantes.*

On se donne deux parties A' de $\mathrm{X}'(\Omega)$ et B' de $\mathrm{Y}'(\Omega)$. On a successivement :

$$\begin{aligned}
\mathbb{P}(\mathrm{X}'\in A',\mathrm{Y}'\in B') &= \mathbb{P}(\mathrm{X}\in\varphi^{-1}(A'),\mathrm{Y}\in\varphi^{-1}(B'))\\
&= \mathbb{P}(\mathrm{X}\in\varphi^{-1}(A'))\,\mathbb{P}(\mathrm{Y}\in\varphi^{-1}(B')) = \mathbb{P}(\mathrm{X}'\in A')\,\mathbb{P}(\mathrm{Y}'\in B') \qquad \square
\end{aligned}$$

Exercice 11.22 (⇝ **corrigé**)
X et Y sont deux variables aléatoires indépendantes et à valeurs dans $\mathbb{N}$.
Elles suivent la même loi définie par : $\forall\, k\in\mathbb{N},\ P(X=k)=P(Y=k)=pq^k$ où $p\in\,]0,1[$ et $q=1-p$.
On considère alors les variables U et V définies par $U=\sup(X,Y)$ et $V=\inf(X,Y)$.
Déterminer la loi du couple (U,V), et les lois marginales. U et V sont-elles indépendantes ?

11.3.4 Variables mutuellement indépendantes

Définition 11.3.6 (variables aléatoires mutuellement indépendantes)
Soit $(\mathrm{X}_i)_{1\leqslant i\leqslant n}$ une suite finie de variables aléatoires sur $(\Omega,\mathscr{A},\mathbb{P})$.
On dit que ces variables sont *mutuellement indépendantes* si :
$\forall\,(x_1,x_2,\ldots,x_n)\in\mathrm{X}_1(\Omega)\times\mathrm{X}_2(\Omega)\times\cdots\times\mathrm{X}_n(\Omega)$, les événements $(\mathrm{X}_1=x_1),(\mathrm{X}_2=x_2),\ldots,(\mathrm{X}_n=x_n)$ sont mutuellement indépendants.
Rappelons que cela signifie : pour toute partie J de $[\![1,n]\!]$, $\mathbb{P}\Big(\bigcap_{j\in J}(\mathrm{X}_j=x_j)\Big)=\prod_{j\in J}\mathbb{P}(\mathrm{X}_j=x_j)$.

Proposition 11.3.7 (caractérisation de l'indépendance mutuelle de n variables aléatoires)
Soit $(\mathrm{X}_i)_{1\leqslant i\leqslant n}$ *une suite finie de variables aléatoires sur* $(\Omega,\mathscr{A},\mathbb{P})$.
Elles sont mutuellement indépendantes si et seulement si :

$$\begin{cases}\forall\,A_1\subset\mathrm{X}_1(\Omega),\ \forall\,A_2\subset\mathrm{X}_2(\Omega),\ \ldots,\ \forall\,A_n\subset\mathrm{X}_n(\Omega)\\ \textit{pour toute partie } J \textit{ de l'intervalle } [\![1,n]\!]\end{cases}\quad \textit{on a l'égalité : } \mathbb{P}\Big(\bigcap_{j\in J}(\mathrm{X}_j\in A_j)\Big)=\prod_{j\in J}\mathbb{P}(\mathrm{X}_j\in A_j).$$

C'est une généralisation facile (par récurrence finie) de la proposition 11.3.5. □

▷ La loi binomiale comme somme de variables de Bernoulli indépendantes

Soit X une variable aléatoire réelle suivant la loi binomiale $\mathcal{B}(n,p)$ de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$.

On sait que X représente le nombre de succès dans une répétition de n épreuves de Bernoulli indépendantes de même paramètre p.

Soit X_i la variable de Bernoulli de paramètre p associée à l'épreuve n°i.

Avec ces notations, on a $\mathrm{X} = \mathrm{X}_1 + \mathrm{X}_2 + \cdots + \mathrm{X}_n$.

Remarques

– Dans la pratique, si on réalise n expériences aléatoires mutuellement indépendantes (les résultats d'une ou plusieurs d'entre elles n'affectent pas le résultat des autres : ce sont les conditions générales de ces expériences qui permettent d'émettre cette hypothèse) et si on attache à chacune de ces expériences une variable aléatoire X_i, alors les X_i sont mutuellement indépendantes.

– Si $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ sont mutuellement indépendantes, il est de même des $(\mathrm{X}_j)_{j\in J}$, avec $J \subset [\![1,n]\!]$.
En particulier les variables aléatoires X_i sont indépendantes deux à deux.

– Dire que deux événements A et B sont indépendants, c'est dire que leurs variables indicatrices $\mathbb{1}_A$ et $\mathbb{1}_B$ sont indépendantes (se reporter aux définitions 11.1.7 et 11.2.2).

Écrire que $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$ c'est écrire $\mathbb{P}(\mathbb{1}_A = 1, \mathbb{1}_B = 1) = \mathbb{P}(\mathbb{1}_A = 1)\mathbb{P}(\mathbb{1}_B = 1)$.

Mais cela implique $\begin{cases} \mathbb{P}(\mathbb{1}_A = 1, \mathbb{1}_B = 0) = \mathbb{P}(\mathbb{1}_A = 1)\mathbb{P}(\mathbb{1}_B = 0) \\ \mathbb{P}(\mathbb{1}_A = 0, \mathbb{1}_B = 1) = \mathbb{P}(\mathbb{1}_A = 0)\mathbb{P}(\mathbb{1}_B = 1) \\ \mathbb{P}(\mathbb{1}_A = 0, \mathbb{1}_B = 0) = \mathbb{P}(\mathbb{1}_A = 0)\mathbb{P}(\mathbb{1}_B = 0) \end{cases}$ d'où l'indépendance de $\mathbb{1}_A$ et $\mathbb{1}_B$.

Réciproquement, si les deux variables indicatrices $\mathbb{1}_A$ et $\mathbb{1}_B$ sont indépendantes, alors on a l'égalité $\mathbb{P}(\mathbb{1}_A = 1, \mathbb{1}_B = 1) = \mathbb{P}(\mathbb{1}_A = 1)\mathbb{P}(\mathbb{1}_B = 1)$ c'est-à-dire $\mathbb{P}(A\cap B) = \mathbb{P}(A)\mathbb{P}(B)$. □

De même, dire que les événements $(A_i)_{1\leqslant i\leqslant n}$ sont mutuellement indépendants (voir définition 11.1.8) c'est dire que leurs variables indicatrices $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_n}$ sont mutuellement indépendantes.

– De même que l'indépendance « deux à deux » des événements $A_1, \ldots, A_n$ (où $n \geqslant 3$) n'implique pas leur indépendance mutuelle (voir proposition 11.1.4), l'indépendance « deux à deux » de n variables $\mathrm{X}_1, \ldots, \mathrm{X}_n$ n'implique pas leur indépendance mutuelle (prendre les variables indicatrices $\mathbb{1}_{A_i}$).

À partir de variables aléatoires mutuellement indépendantes, on peut en former d'autres :

Proposition 11.3.8 (fonctions de variables mutuellement indépendantes)
Soit $(\mathrm{X}_i)_{1\leqslant i\leqslant n}$ une suite finie de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$, mutuellement indépendantes.
Pour tout i de $[\![1,n]\!]$, soit $\mathrm{Y}_i = \varphi_i(\mathrm{X}_i)$ une fonction de la variable aléatoire X_i.
Alors les variables aléatoires $(\mathrm{Y}_i)_{1\leqslant i\leqslant n}$ sont mutuellement indépendantes.

C'est une généralisation facile (par récurrence finie) de la proposition 11.3.6. □

Autre exemple : si $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$ sont mutuellement indépendantes, et si J et K sont deux parties disjointes de $[\![1,n]\!]$, toute fonction Y des $(\mathrm{X}_j)_{j\in J}$ est indépendante de toute fonction Z des $(\mathrm{X}_k)_{k\in K}$.

Définition 11.3.7 (suite de variables aléatoires indépendantes)
Soit $(\mathrm{X}_n)_{n\geqslant 0}$ une suite de variables aléatoires sur $(\Omega, \mathscr{A}, \mathbb{P})$.
On dit que cette suite est formée de variables aléatoires indépendantes si, pour tout m de $\mathbb{N}$, les variables $\mathrm{X}_0, \mathrm{X}_1, \ldots, \mathrm{X}_m$ sont mutuellement indépendantes.

Il revient au même de dire que toute sous-famille finie de la suite $(\mathrm{X}_n)_{n\geqslant 0}$ est formée de variables mutuellement indépendantes.
On peut imaginer d'effectuer une répétition indéfinie d'expériences aléatoires (le plus souvent identiques et mutuellement indépendantes) et, pour tout n de $\mathbb{N}$, associer une variable aléatoire X_n à l'expérience de rang n (le plus souvent, les X_n suivent la même loi usuelle).
On pourra par exemple répéter indéfiniment le même jeu consistant à lancer un dé jusqu'à obtenir 6. Lors du n-ième jeu, on pourra noter X_n le temps d'attente du premier « 6 ».

11.4 Espérance et variance

11.4.1 Espérance (rappels de première année)

Les propriétés rappelées dans cette sous-section sont au programme de première année.

▷ Espérance d'une variable aléatoire réelle (univers fini)

– Soit $\mathrm{X} : \Omega \to E$ une variable aléatoire réelle sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, avec Ω un ensemble fini.
La quantité $\mathrm{E}(\mathrm{X}) = \displaystyle\sum_{x\in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}=x)\, x$ est appelée *espérance* de la variable aléatoire réelle X.
Cette somme finie est la *moyenne pondérée* des valeurs x que la variable X est susceptible de prendre, le poids affecté à chacune de ces valeurs x étant la probabilité de l'événement $(\mathrm{X}=x)$.

– Plutôt que de sommer sur les valeurs x de $\mathrm{X}(\Omega)$, on peut sommer sur les résultats élémentaires ω.
On obtient alors la relation : $\mathrm{E}(\mathrm{X}) = \displaystyle\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\, \mathrm{X}(\omega)$.

L'événement $(X = x)$ est la réunion disjointe des singletons $\{\omega\}$ tels que $X(\omega) = x$.
Ainsi $\mathbb{P}(X = x) = \displaystyle\sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} \mathbb{P}(\{\omega\})$ et on en déduit :

$$\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\, \mathrm{X}(\omega) = \sum_{x\in \mathrm{X}(\Omega)} \Big(\sum_{\substack{\omega\in\Omega\\ \mathrm{X}(\omega)=x}} \mathbb{P}(\{\omega\})\, \mathrm{X}(\omega)\Big) = \sum_{x\in \mathrm{X}(\Omega)} \Big(x \sum_{\substack{\omega\in\Omega\\ \mathrm{X}(\omega)=x}} \mathbb{P}(\{\omega\})\Big) = \sum_{x\in \mathrm{X}(\Omega)} x\, \mathbb{P}(\mathrm{X}=x) = \mathrm{E}(\mathrm{X})$$

Remarque : ce calcul ne nécessite pas de justification car les sommes sont finies. □

▷ Espérance de lois usuelles (univers fini)

On se reportera aux définitions rappelées dans la sous-section 11.2.3.

– Espérance d'une loi constante : si X est constante, de valeur a, alors $\mathrm{E}(\mathrm{X}) = a$.

Évidentissime : si $\mathrm{X}(\Omega) = \{a\}$, alors $\mathbb{P}(\mathrm{X}=a) = 1$ donc $\mathrm{E}(\mathrm{X}) = \displaystyle\sum_{x\in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}=x)\, x = \mathbb{P}(\mathrm{X}=a)\, a = a$. □

– Espérance d'une variable indicatrice : si X est l'indicatrice d'un événement A, alors $\mathrm{E}(\mathrm{X}) = \mathbb{P}(A)$.

Évident : on a $(X=1) = A$ et $(X=0) = \overline{A}$ donc $\mathrm{E}(\mathrm{X}) = \mathbb{P}(\mathrm{X}=0)\, 0 + \mathbb{P}(\mathrm{X}=1)\, 1 = \mathbb{P}(\mathrm{X}=1) = \mathbb{P}(A)$. □

– Si X suit la loi de Bernoulli $\mathcal{B}(p)$ de paramètre p, alors $\text{E}(\text{X}) = p$.

Trop facile : $\text{E}(\text{X}) = \mathbb{P}(\text{X}=0)\cdot 0 + \mathbb{P}(\text{X}=1)\cdot 1 = (1-p)\cdot 0 + p \cdot 1 = p$. □

– Si X suit la loi uniforme $\mathcal{U}([\![a,b]\!])$ sur $[\![a,b]\!]$, alors $\text{E}(\text{X}) = \dfrac{a+b}{2}$.

Très facile : on a $\text{E}(\text{X}) = \displaystyle\sum_{k=a}^{b} k\,\mathbb{P}(\text{X}=k) = \frac{1}{n}\sum_{k=a}^{b} k = \frac{1}{n}\Big(n\frac{a+b}{2}\Big) = \frac{a+b}{2}$. □

– Si X suit la loi binomiale $\mathcal{B}(n,p)$ de paramètres n,p, alors $\text{E}(\text{X}) = np$.

Easy : $\text{E}(\text{X}) = \displaystyle\sum_{k=0}^{n} k\,\mathbb{P}(\text{X}=k) = \sum_{k=1}^{n}\binom{n}{k} k\,p^k q^{n-k} = np\sum_{k=1}^{n}\binom{n-1}{k-1} p^{k-1}q^{n-k} = np\,(p+q)^{n-1} = np$. □

▷ Théorème du transfert (univers fini)

La propriété suivante est très importante.

On considère une fonction $\text{Y} = f(\text{X})$ d'une variable aléatoire X.

Pour calculer $\text{E}(\text{Y})$, il faut théoriquement connaître la loi de Y, c'est-à-dire les probabilités $\mathbb{P}(Y=y)$.

En fait il suffit de connaître, et d'utiliser, la loi de X :

Proposition 11.4.1 (théorème du transfert, dans le cas fini)
Soit $\text{X} : \Omega \to E$ *une variable aléatoire sur* $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$*, avec* Ω *fini.*
Soit $f : E \to \mathbb{R}$ *une application. Alors* $\text{E}(f(\text{X})) = \displaystyle\sum_{x\in \text{X}(\Omega)} \mathbb{P}(\text{X}=x) f(x)$.

L'événement $(\text{Y}=y)$ est la réunion disjointe des événements $(\text{X}=x)$ pour les x de $\text{X}(\Omega)$ tels que $f(x) = y$.

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

L'interversion $\displaystyle\sum_{y\in\text{Y}(\Omega)}\sum_{\substack{x\in\text{X}(\Omega)\\ f(x)=y}} = \sum_{x\in\text{X}(\Omega)}\sum_{\substack{y\in\text{Y}(\Omega)\\ y=f(x)}}$ est justifiée (sommes finies), et $\displaystyle\sum_{\substack{y\in\text{Y}(\Omega)\\ y=f(x)}}$ se réduit à un seul terme. □

▷ Croissance et linéarité de l'espérance (univers fini)

Soit X et Y deux variables aléatoires réelles sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, avec Ω un ensemble fini.

– On suppose que $\text{X}(\omega) \leqslant \text{Y}(\omega)$ pour tout ω de Ω. Alors $\text{E}(\text{X}) \leqslant \text{E}(\text{Y})$.

Évident : $\text{E}(\text{X}) = \displaystyle\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\,\text{X}(\omega) \leqslant \sum_{\omega\in\Omega}\mathbb{P}(\{\omega\})\,\text{Y}(\omega) = \text{E}(\text{Y})$. □

– Si $X \geqslant 0$ alors on ne peut avoir $\text{E}(\text{X}) = 0$ que si $\mathbb{P}(\text{X}=0) = 1$ (c'est-à-dire : X est « presque nulle »).

La somme $\text{E}(\text{X}) = \displaystyle\sum_{x\in\text{X}(\Omega)} \mathbb{P}(\text{X}=x)\,x$ (à termes positifs ou nuls) est nulle donc tous ses termes sont nuls.
Cela implique $\mathbb{P}(\text{X}=x) = 0$ si $x > 0$, donc $\mathbb{P}(\text{X}=0) = 1$. □

– Pour tous réels a, b, on a : $\mathrm{E}(a\mathrm{X}+b) = a\mathrm{E}(\mathrm{X})+b$.

On applique le théorème du transfert à $\mathrm{Y} = a\mathrm{X}+b$:

$\mathrm{E}(a\mathrm{X}+b) = \sum\limits_{x\in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}=x)(ax+b) = a\sum\limits_{x\in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}=x)x + \sum\limits_{x\in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}=x) = a\,\mathrm{E}(\mathrm{X})+b.$ □

– Plus généralement, on a l'égalité : $\mathrm{E}(a\mathrm{X}+b\mathrm{Y}) = a\mathrm{E}(\mathrm{X})+b\mathrm{E}(\mathrm{Y})$.

$\mathrm{S} = a\mathrm{X}+b\mathrm{Y}$ est une fonction du couple $\mathrm{Z} = (\mathrm{X},\mathrm{Y})$ (en fait $\mathrm{S} = \varphi(\mathrm{Z})$, où $\varphi(x,y) = ax+by$).

On en déduit : $\mathrm{E}(\mathrm{S}) = \sum\limits_{(x,y)} (ax+by)\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y) = a\sum\limits_{(x,y)} x\,\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y) + b\sum\limits_{(x,y)} y\,\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y)$.

Mais $\sum\limits_{(x,y)} x\,\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y) = \sum\limits_{x} x\Big(\sum\limits_{y}\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y)\Big) = \sum\limits_{x} x\,\mathbb{P}(\mathrm{X}=x) = \mathrm{E}(\mathrm{X})$.

De même $\sum\limits_{(x,y)} y\,\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y) = \sum\limits_{y} y\Big(\sum\limits_{x}\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y)\Big) = \sum\limits_{y} y\,\mathbb{P}(\mathrm{Y}=y) = \mathrm{E}(\mathrm{Y})$.

On a finalement obtenu : $\mathrm{E}(a\mathrm{X}+b\mathrm{Y}) = a\mathrm{E}(\mathrm{X})+b\mathrm{E}(\mathrm{Y})$. □

– Par récurrence évidente : pour toutes variables $\mathrm{X}_1,\ldots,\mathrm{X}_n$, on a : $E\Big(\sum\limits_{i=1}^{n}\lambda_i\mathrm{X}_i\Big) = \sum\limits_{i=1}^{n}\lambda_i\mathrm{E}(\mathrm{X}_i)$.

▷ **Où l'on retrouve l'espérance de la loi binomiale $\mathcal{B}(n,p)$**

On sait que si $\mathrm{X}\hookrightarrow\mathcal{B}(n,p)$ alors $\mathrm{X} = \sum\limits_{i=1}^{n}\mathrm{X}_i$ où les $\mathrm{X}_i\hookrightarrow\mathcal{B}(p)$ sont mutuellement indépendantes.

Cette représentation de X permet de retrouver : $\mathrm{E}(\mathrm{X}) = \sum\limits_{i=1}^{n}\underbrace{\mathrm{E}(\mathrm{X}_i)}_{=p} = np$.

Remarque : l'indépendance des X_i n'est pas utilisée ici.

▷ **Espérance du produit de deux variables indépendantes (univers fini)**

Soit X et Y deux variables aléatoires réelles sur $(\Omega,\mathcal{P}(\Omega),\mathbb{P})$, avec Ω un ensemble fini.

On suppose que X et Y sont indépendantes. Alors : $\mathrm{E}(\mathrm{XY}) = \mathrm{E}(\mathrm{X})\,\mathrm{E}(\mathrm{Y})$.

On utilise le théorème du transfert, appliqué à XY en tant que fonction du couple (X,Y).

$\mathrm{E}(\mathrm{XY}) = \sum\limits_{(x,y)} xy\,\mathbb{P}(\mathrm{X}=x,\mathrm{Y}=y) = \sum\limits_{(x,y)} xy\,\mathbb{P}(\mathrm{X}=x)\mathbb{P}(\mathrm{Y}=y) = \sum\limits_{x} x\,\mathbb{P}(\mathrm{X}=x)\sum\limits_{y} y\,\mathbb{P}(\mathrm{Y}=y) = \mathrm{E}(\mathrm{X})\mathrm{E}(\mathrm{Y})$. □

Inversement, l'égalité $\mathrm{E}(\mathrm{XY}) = \mathrm{E}(\mathrm{X})\,\mathrm{E}(\mathrm{Y})$ n'implique pas l'indépendance de X et Y en général.

11.4.2 Variance, écart-type (rappels de première année)

Soit X une variable aléatoire réelle sur $(\Omega,\mathcal{P}(\Omega),\mathbb{P})$, avec Ω fini. Posons $m = \mathrm{E}(\mathrm{X})$.

La variable positive $\mathrm{Y} = (\mathrm{X}-m)^2$ mesure la « dispersion » de X autour de son espérance m.

L'espérance de Y est appelé *variance* de X et est notée $\mathrm{V}(\mathrm{X})$. Ainsi $\mathrm{V}(\mathrm{X}) = \mathrm{E}((\mathrm{X}-\mathrm{E}(\mathrm{X}))^2)$.

La quantité $\sigma(\mathrm{X}) = \sqrt{\mathrm{V}(\mathrm{X})}$ est appelée *écart-type* de la variable aléatoire X.

L'utilisation de l'écart-type est justifiée pour des raisons d'homogénéité par rapport aux valeurs de X.

▷ Propriétés de la variance et de l'écart-type (univers fini)

– On a la *formule de König-Huyghens* : $\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}^2(\mathrm{X})$.

> Posons $m = \mathrm{E}(\mathrm{X})$.
> On développe : $\mathrm{V}(\mathrm{X}) = \mathrm{E}((\mathrm{X}-m)^2) = \mathrm{E}(\mathrm{X}^2 - 2m\mathrm{X} + m^2) = \mathrm{E}(\mathrm{X}^2) - 2m\overbrace{\mathrm{E}(\mathrm{X})}^{=m} + m^2 = \mathrm{E}(\mathrm{X}^2) - m^2$. □

– Pour tous réels a, b on a : $\mathrm{V}(a\mathrm{X}+b) = a^2\mathrm{V}(\mathrm{X})$, donc $\sigma(a\mathrm{X}+b) = |a|\,\sigma(\mathrm{X})$.

> Si $\mathrm{Y} = a\mathrm{X}+b$, alors $\mathrm{E}(\mathrm{Y}) = a\mathrm{E}(\mathrm{X}) + b$ donc $\mathrm{Y} - \mathrm{E}(\mathrm{Y}) = a(\mathrm{X} - \mathrm{E}(\mathrm{X}))$.
> Ainsi $\mathrm{V}(\mathrm{Y}) = \mathrm{E}((\mathrm{Y}-\mathrm{E}(\mathrm{Y}))^2) = \mathrm{E}(a^2(\mathrm{X}-\mathrm{E}(\mathrm{X}))^2) = a^2\mathrm{E}((\mathrm{X}-\mathrm{E}(\mathrm{X}))^2) = a^2\mathrm{V}(\mathrm{X})$. □

– On a toujours $\mathrm{V}(\mathrm{X}) \geqslant 0$ et on ne peut avoir $\mathrm{V}(\mathrm{X}) = 0$ que si X est « presque constante ».

> Posons $m = \mathrm{E}(\mathrm{X})$. La variable $(\mathrm{X}-m)^2 \geqslant 0$ est à valeurs positives.
> Ainsi : $\mathrm{V}(\mathrm{X}) = 0 \Leftrightarrow \mathrm{E}((\mathrm{X}-m)^2) = 0 \Leftrightarrow \mathbb{P}((\mathrm{X}-m)^2 = 0) = 1 \Leftrightarrow \mathbb{P}(\mathrm{X} = m) = 1$. □

▷ Variance de lois usuelles (univers fini)

On se reportera aux définitions rappelées dans la sous-section 11.2.3.

– Variance d'une loi constante : si X est constante, alors $\mathrm{V}(\mathrm{X}) = 0$.

> Si X est constante de valeur a, alors X^2 est constante de valeur a^2.
> Ainsi $\mathrm{E}(\mathrm{X}^2) = a^2$ et $\mathrm{E}(\mathrm{X}) = a$, donc $\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}(\mathrm{X})^2 = 0$. □

– Variance d'une variable indicatrice : si $\mathrm{X} = \mathbb{1}_A$, alors $\mathrm{V}(\mathrm{X}) = \mathbb{P}(A)\,\mathbb{P}(\overline{A})$.

> En effet, on a encore $\mathrm{X}^2 = \mathbb{1}_A$, donc $\mathrm{E}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) = \mathbb{P}(A)$.
> Ainsi $\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}(\mathrm{X})^2 = \mathbb{P}(A)^2 - \mathbb{P}(A) = \mathbb{P}(A)(1 - \mathbb{P}(A) = \mathbb{P}(A)\,\mathbb{P}(\overline{A})$. □

– Si X suit la loi de Bernoulli $\mathcal{B}(p)$ de paramètre p, alors $\mathrm{V}(\mathrm{X}) = pq$.

> Ici X et X^2 suivent la même loi, donc $\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}) - \mathrm{E}(\mathrm{X})^2 = p - p^2 = pq$. □

– Si X suit la loi uniforme $\mathcal{U}([\![a, b]\!])$ (n entiers successifs), alors $\mathrm{V}(\mathrm{X}) = \dfrac{n^2-1}{12}$.

> Par translation, on se ramène à $X \hookrightarrow \mathcal{U}([\![1, n]\!])$.
> $\mathrm{E}(\mathrm{X}^2) = \displaystyle\sum_{k=1}^{n} k^2\,\mathbb{P}(\mathrm{X}=k) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{(n+1)(2n+1)}{6}$. D'autre part $\mathrm{E}(\mathrm{X}) = \dfrac{n+1}{2}$.
> Ainsi : $\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}^2(\mathrm{X}) = \dfrac{(n+1)(2n+1)}{6} - \dfrac{(n+1)^2}{4} = \dfrac{n+1}{12}(2(2n+1) - 3(n+1)) = \dfrac{n^2-1}{12}$ □

– Si X suit la loi binomiale $\mathcal{B}(n, p)$ de paramètres n, p, alors $\mathrm{V}(\mathrm{X}) = npq$.

> On sait déjà que $\mathrm{E}(X) = np$. Pour calculer $\mathrm{V}(\mathrm{X})$, on calcule $\mathrm{E}(\mathrm{X}(\mathrm{X}-1))$.
> $$\begin{aligned}\mathrm{E}(\mathrm{X}(\mathrm{X}-1)) &= \sum_{k=0}^{n} k(k-1)\,\mathbb{P}(\mathrm{X}=k) = \sum_{k=2}^{n}\binom{n}{k}k(k-1)\,p^k q^{n-k}\\ &= n(n-1)p^2\sum_{k=2}^{n}\binom{n-2}{k-2}p^{k-2}q^{n-k} = n(n-1)p^2\,(p+q)^{n-2} = n(n-1)p^2\end{aligned}.$$
> D'autre part $\mathrm{E}(\mathrm{X}) = np$.
> Ainsi $\mathrm{V}(\mathrm{X}) = \mathrm{E}(\mathrm{X}(\mathrm{X}-1)) + \mathrm{E}(\mathrm{X}) - \mathrm{E}^2(\mathrm{X}) = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq$. □

▷ Covariance de deux variables aléatoires réelles (univers fini, MPSI)

La *covariance de* X *et* Y est l'espérance, notée $\mathrm{Cov}(X,Y)$, de la variable $(X-E(X))(Y-E(Y))$.

On remarque que $V(X) = \mathrm{Cov}(X,X)$.

On a l'égalité $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.

Soit $m = E(X)$ et $m' = E(Y)$: $\begin{cases} \mathrm{Cov}(X,Y) &= E(X-m)(Y-m') = E(XY - mY - m'X + mm') \\ &= E(XY) - mE(Y) - m'E(X) + mm' = E(XY) - mm' \end{cases}$ □

Si X et Y sont indépendantes, elles sont de covariance nulle (réciproque fausse).

On a $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ (symétrie) et $\mathrm{Cov}(aX+bX',Y) = a\,\mathrm{Cov}(X,Y) + b\,\mathrm{Cov}(X',Y)$ (bilinéarité).

La symétrie est évidente. Quant à la linéarité à gauche (qui devient bilinéarité par symétrie) :

$$\begin{aligned}\mathrm{Cov}(aX+bX',Y) &= E((aX+bX')Y) - E(aX+bX')\,E(Y) = E(aXY+bX'Y) - (aE(X)+bE(X'))\,E(Y)\\ &= aE(XY) + bE(X'Y) - aE(X)E(Y) - bE(X')E(Y)\\ &= a(E(XY)-E(X)E(Y)) + b(E(X'Y)-E(X')E(Y)) = a\,\mathrm{Cov}(X,Y) + b\,\mathrm{Cov}(X',Y).\end{aligned}$$ □

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

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

On en déduit que si X et Y sont indépendantes, alors $V(X+Y) = V(X) + V(Y)$.

Plus généralement si $X_1,\ldots,X_n$ sont deux à deux indépendantes, on a : $V\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} V(X_i)$.

Par bilinéarité : $V\Big(\sum_{i=1}^{n} X_i\Big) = \mathrm{Cov}\Big(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\Big) = \sum_{i=1}^{n} \overbrace{\mathrm{Cov}(X_i,X_i)}^{=V(X_i)} + 2\sum_{i<j} \overbrace{\mathrm{Cov}(X_i,X_j)}^{=0} = \sum_{i=1}^{n} V(X_i)$ □

On remarque que dans le résultat précédent, il n'est pas nécessaire que les variables E_i soient mutuellement indépendantes : il suffit qu'elles soient indépendantes deux à deux.

▷ Où l'on retouve la variance de la loi binomiale $\mathcal{B}(n,p)$

On sait que si $X \hookrightarrow \mathcal{B}(n,p)$ alors $X = \sum_{i=1}^{n} X_i$ où les $X_i \hookrightarrow \mathcal{B}(p)$ sont mutuellement indépendantes.

Cette représentation de X permet de retrouver : $V(X) = \sum_{i=1}^{n} \overbrace{V(X_i)}^{=pq} = npq$.

Remarque : seule l'indépendance deux à deux des X_i est utilisée ici, pas leur indépendance mutuelle.

▷ Coefficient de corrélation linéaire (univers fini, MPSI)

Soit X, Y deux variables aléatoires réelles de variance non nulle.

On appelle coefficient de corrélation linéaire de X et Y la quantité $\rho(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$.

Si X et Y sont indépendantes, on a bien sûr $\rho(X,Y) = 0$.

On a toujours $|\rho(\mathrm{X},\mathrm{Y})| \leqslant 1$.

D'autre part : $\begin{cases} \rho(\mathrm{X},\mathrm{Y}) & = 1 \quad \Leftrightarrow \ (\exists\, a > 0, \exists\, b \in \mathbb{R},\ \mathbb{P}(\mathrm{Y}{=}a\mathrm{X}{+}b) = 1) \\ \rho(\mathrm{X},\mathrm{Y}) & = -1 \ \Leftrightarrow \ (\exists\, a < 0, \exists\, b \in \mathbb{R},\ \mathbb{P}(\mathrm{Y}{=}a\mathrm{X}{+}b) = 1) \end{cases}$

En termes plus suggestifs, dire que $\rho(\mathrm{X},\mathrm{Y})$ vaut ± 1 (ou en est très proche) c'est dire que chacune des deux variables X et Y dépend « quasiment » de l'autre de façon affine.

C'est une inégalité de Cauchy-Schwarz...
Pour tout réel λ, on a $A(\lambda) = \mathrm{V}(\lambda \mathrm{X} - Y) = \lambda^2 \mathrm{V}(\mathrm{X}) - 2\lambda \operatorname{Cov}(\mathrm{X},\mathrm{Y}) + \mathrm{V}(\mathrm{Y})$.
Ce trinôme $A(\lambda)$ reste positif, donc son discriminant $\Delta = \operatorname{Cov}(\mathrm{X},\mathrm{Y})^2 - \mathrm{V}(\mathrm{X})\mathrm{V}(\mathrm{Y})$ est négatif ou nul.
Mais écrire $\Delta \leqslant 0$, c'est écrire $|\rho(\mathrm{X},\mathrm{Y})| \leqslant 1$.
Ensuite : $|\rho(\mathrm{X},\mathrm{Y})| = 1 \ \Leftrightarrow\ \Delta = 0 \ \Leftrightarrow\ \exists\, a \in \mathbb{R},\ \mathrm{V}(a\mathrm{X} - Y) = 0$.
Avec ces notations, la racine double de $A(\lambda)$ est $a = \dfrac{\operatorname{Cov}(\mathrm{X},\mathrm{Y})}{\mathrm{V}(\mathrm{X})}$ donc a et $\rho(\mathrm{X},\mathrm{Y})$ sont de même signe.
Mais $\mathrm{V}(a\mathrm{X} - \mathrm{Y}) = 0$ signifie que $a\mathrm{X} - \mathrm{Y}$ est « presque sûre » càd : $\exists\, b \in \mathbb{R},\ \mathbb{P}(\mathrm{Y} = a\mathrm{X} + b) = 1$. □

▷ Inégalités de Markov et de Bienaymé-Tchebychev (univers fini)

Soit X une variable aléatoire *réelle positive* sur $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$, avec Ω un ensemble fini.

– Pour tout $a > 0$, on a l'inégalité de Markov : $\mathbb{P}(\mathrm{X} \geqslant a) \leqslant \dfrac{\mathrm{E}(\mathrm{X})}{a}$.

  On décompose $\mathrm{E}(\mathrm{X}) = \displaystyle\sum_{x \in \mathrm{X}(\Omega)} \mathbb{P}(\mathrm{X}{=}x)\, x$ en $\mathrm{E}(\mathrm{X}) = \displaystyle\sum_{x<a} \mathbb{P}(\mathrm{X}{=}x)\, x + \sum_{x \geqslant a} \mathbb{P}(\mathrm{X}{=}x)\, x$.
  La première somme ne contient que des termes positifs.
  Ainsi $\mathrm{E}(\mathrm{X}) \geqslant \displaystyle\sum_{x \geqslant a} \mathbb{P}(\mathrm{X}{=}x) x \geqslant a \sum_{x \geqslant a} \mathbb{P}(\mathrm{X}{=}x)$, c'est-à-dire $\mathrm{E}(\mathrm{X}) \geqslant a\, \mathbb{P}(\mathrm{X} \geqslant a)$. □

– Pour tout $\varepsilon > 0$, on a : $\mathbb{P}\Big(|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geqslant \varepsilon\Big) \leqslant \dfrac{\mathrm{V}(\mathrm{X})}{\varepsilon^2}$ (inégalité de Bienaymé-Chebychev).

  On applique l'inégalité de Markov à la variable positive $\mathrm{Y} = (\mathrm{X} - \mathrm{E}(\mathrm{X}))^2$, dont l'espérance est $\mathrm{V}(\mathrm{X})$.
  Pour tout $\varepsilon > 0$, on a : $\mathbb{P}(\mathrm{Y} \geqslant \varepsilon^2) \leqslant \dfrac{\mathrm{V}(\mathrm{X})}{\varepsilon^2}$, c'est-à-dire : $\mathbb{P}(|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geqslant \varepsilon) \leqslant \dfrac{\mathrm{V}(\mathrm{X})}{\varepsilon^2}$. □

11.4.3 Espérance d'une variable aléatoire discrète

On va maintenant reprendre et généraliser les résultats vus en première année.
Le cas est maintenant celui des espaces probabilisés $(\Omega, \mathscr{A}, \mathbb{P})$.

– Si Ω est fini ou dénombrable (donc si on peut écrire $\Omega = \{\omega_n,\ n \in \mathbb{N}\}$) on utilise en général $\mathscr{A} = \mathcal{P}(\Omega)$.
  La probabilité $\mathbb{P}$ est alors caractérisée par les $p_n = \mathbb{P}(\{\omega_n\})$, où $0 \leqslant p_n \leqslant 1$ et $\sum p_n = 1$.
  (on se reportera au paragraphe « cas d'un univers dénombrable » dans la sous-section 11.1.2)

– Il se peut que l'ensemble Ω ne soit pas dénombrable (par exemple l'univers $\{0,1\}^{\mathbb{N}}$ attaché à la répétition indéfinie du lancer d'une pièce).
  Dans ce cas, et si X est une application discrète (à image finie ou dénombrable) de Ω dans un ensemble E, on sait qu'on peut former une tribu $\mathscr{A}$ de parties de Ω comportant tous les événements qui « intéressent » la variable X (voir la proposition 11.2.3).

Dans tous les cas, on étudie une variable discrète Ω sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

Si $X(\Omega)$ est infini dénombrable, on écrira $X(\Omega) = \{x_n,\ n \in \mathbb{N}\}$, avec des x_n distincts deux à deux.

Contexte : dans cette sous-section, X et Y désignent deux variables aléatoires réelles sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Définition 11.4.1 (espérance d'une variable aléatoire réelle discrète)
Posons $X(\Omega) = \{x_n,\ n \in \mathbb{N}\}$, où les x_n sont distincts deux à deux.
On dit que X est « d'espérance finie » si la série $\sum \mathbb{P}(X{=}x_n)\, x_n$ est absolument convergente.
Si tel est le cas, la somme $E(X) = \displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(X{=}x_n)\, x_n$ est appelée *espérance de* X.

Remarques

– On admet que l'absolue convergence fait que $\displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(X{=}x_n)\, x_n$ ne dépend pas de l'ordre d'énumération.
Cette indépendance permet de noter l'espérance de X sous la forme $E(X) = \displaystyle\sum_{n\in\mathbb{N}} \mathbb{P}(X{=}x_n) x_n$.

– On peut unifier les cas « $X(\Omega)$ fini » et « $X(\Omega)$ dénombrable » en posant $X(\Omega) = \{x_n,\ n \in J \subset \mathbb{N}\}$, où les x_n sont supposés distincts deux à deux.
On notera alors $E(X) = \displaystyle\sum_{n\in J} x_n\, \mathbb{P}(X = x_n)$, ou encore $E(X) = \displaystyle\sum_{x\in X(\Omega)} x\, \mathbb{P}(X = x)$, sachant qu'on doit d'abord s'assurer de la convergence de la somme avant de l'utiliser quand $X(\Omega)$ est dénombrable.

– Le problème de l'existence de l'espérance ne se pose évidemment pas si l'ensemble $X(\Omega)$ est fini.
Tout ce qui a été dit dans la sous-section 11.4.1 reste donc valable.

– En particulier, si X est constante, de valeur a, alors $E(X) = a$. Le résultat reste valable si X est « presque constante », c'est-à-dire s'il existe a tel que $\mathbb{P}(X{=}a) = 1$.
Si X est la variable indicatrice d'un événement A, alors $E(X) = \mathbb{P}(A)$.

– On dit qu'une variable aléatoire est *centrée* si son espérance est nulle (snif).
Si X est d'espérance finie, alors $X - E(X)$ est une variable centrée.

On généralise le théorème du transfert (cf proposition 11.4.1), quand $X(\Omega)$ est dénombrable.
NB : la démonstration dans le cas dénombrable est *hors-programme*.

Proposition 11.4.2 (expression de l'espérance d'une variable à valeurs dans $\mathbb{N}$)
Soit X *une variable aléatoire à valeurs dans* $\mathbb{N}$.
Dire que E *est d'espérance finie équivaut à dire que la série* $\sum \mathbb{P}(X \geqslant n)$ *est convergente.*
Dans ce cas, on a alors l'égalité : $E(X) = \displaystyle\sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n)$.

On a l'union disjointe d'événements : $(X \geqslant k) = (X{=}k) \cup (X \geqslant k+1)$.
Il en résulte $\mathbb{P}(X{=}k) = \mathbb{P}(X \geqslant k) - \mathbb{P}(X \geqslant k{+}1)$. On en déduit, pour tout n de $\mathbb{N}$:

$$\begin{aligned}\sum_{k=0}^{n} \mathbb{P}(X{=}k)\, k &= \sum_{k=0}^{n} (\mathbb{P}(X \geqslant k) - \mathbb{P}(X \geqslant k{+}1))\, k = \sum_{k=0}^{n} \mathbb{P}(X \geqslant k)\, k - \sum_{k=0}^{n} \mathbb{P}(X \geqslant k{+}1)\, k \\ &= \sum_{k=0}^{n} \mathbb{P}(X \geqslant k)\, k - \sum_{k=1}^{n+1} \mathbb{P}(X \geqslant k)\,(k{-}1) = \sum_{k=1}^{n+1} \mathbb{P}(X \geqslant k) - n\,\mathbb{P}(X \geqslant n{+}1) \quad (\star)\end{aligned}$$

– On suppose que X est d'espérance finie.

On a les inégalités : $0 \leqslant n\,\mathbb{P}(\mathrm{X} \geqslant n+1) = n \displaystyle\sum_{k=n+1}^{+\infty} \mathbb{P}(\mathrm{X}=k) \leqslant \sum_{k=n+1}^{+\infty} k\,\mathbb{P}(\mathrm{X}=k) = R_n$.

Or la suite $(R_n)_{n\geqslant 0}$ converge vers 0 (suite des restes d'une série convergente).

On fait alors tendre n vers $+\infty$ dans le résultat $(\star)$ et on obtient l'égalité : $\mathrm{E}(\mathrm{X}) = \displaystyle\sum_{n=1}^{+\infty} \mathbb{P}(\mathrm{X} \geqslant n)$.

– Inversement, si $\sum \mathbb{P}(\mathrm{X} \geqslant n)$ converge, alors $(\star)$ implique $\displaystyle\sum_{k=0}^{n} \mathbb{P}(X=k)\,k \leqslant \sum_{k=1}^{n+1} \mathbb{P}(\mathrm{X} \geqslant k) \leqslant \sum_{k=1}^{+\infty} \mathbb{P}(\mathrm{X} \geqslant k)$.

Ainsi $\sum \mathbb{P}(X=k)\,k$ converge, donc $\mathrm{E}(\mathrm{X})$ existe, et on est ramené au sens direct. □

Exercice 11.23 (⇝ **corrigé**)

Une urne contient $2n$ boules numérotées de 1 à $2n$. On tire les boules successivement et sans remise.

1. Probabilité de tirer les boules impaires dans l'ordre, non nécessairement consécutivement.
2. Probabilité de tirer les boules impaires dans l'ordre et consécutivement.
3. Soit X le rang de la dernière boule impaire tirée. Calculer son espérance.

Exercice 11.24 (⇝ **corrigé**)

On considère une urne contenant n boules numérotées de 1 à n.

On tire aléatoirement k boules en une seule prise. On note X le plus petit numéro obtenu.

Déterminez la loi de X. Calculer la valeur de $S_{n,k} = \displaystyle\sum_{i=1}^{n-k+1} \binom{n-i}{k-1}$, puis $E(X)$.

Proposition 11.4.3 (théorème du transfert, dans le cas dénombrable)

Posons $\mathrm{X}(\Omega) = \{x_n,\ n \in \mathbb{N}\}$, *où les* x_n *sont distincts deux à deux.*

Soit f *une application à valeurs réelles, définie sur* $\mathrm{X}(\Omega)$.

Les conditions suivantes sont équivalentes :

– *la variable* $f(\mathrm{X})$ *est d'espérance finie ;*

– *la série* $\sum \mathbb{P}(\mathrm{X}=x_n)\,f(x_n)$ *est absolument convergente.*

Dans ce cas, on a alors : $\mathrm{E}(f(\mathrm{X})) = \displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=x_n)\,f(x_n)$.

Pour calculer $\mathrm{E}(f(\mathrm{X}))$, il suffit donc de connaître et d'utiliser la loi de X.

On voit ici une ébauche de démonstration, où manque une justification concernant la possibilité d'intervertir deux sommations de séries.

Posons $\mathrm{X}(\Omega) = \{x_k,\ k \in \mathbb{N}\}$, et $\mathrm{Y}(\Omega) = \{y_n,\ n \in \mathbb{N}\}$ (les x_k distincts deux à deux, les y_n aussi).

L'événement $(\mathrm{Y}=y_n)$ est la réunion disjointe des événements $(\mathrm{X}=x_k)$ pour les k de $\mathbb{N}$ tels que $f(x_k) = y_n$.

$$\mathrm{E}(\mathrm{Y}) = \sum_{n\in\mathbb{N}} y_n\,\mathbb{P}(\mathrm{Y}=y_n) = \sum_{n\in\mathbb{N}}\Big(\sum_{\substack{k\in\mathbb{N}\\ f(x_k)=y_n}} y_n\,\mathbb{P}(\mathrm{X}=x_k)\Big) = \sum_{k\in\mathbb{N}}\Big(\sum_{\substack{n\in\mathbb{N}\\ y_n=f(x_k)}} y_n\,\mathbb{P}(\mathrm{X}=x_k)\Big) = \sum_{k\in\mathbb{N}} f(x_k)\,\mathbb{P}(X=x_k)$$

La somme $\displaystyle\sum_{\substack{n\in\mathbb{N}\\ y_n=f(x_k)}} y_n\,\mathbb{P}(\mathrm{X}=x_k)$ se réduit au seul terme $f(x_k)\,\mathbb{P}(X=x_k)$.

Le problème est que l'interversion $\displaystyle\sum_{n\in\mathbb{N}} \sum_{\substack{k\in\mathbb{N}\\ f(x_k)=y_n}} = \sum_{k\in\mathbb{N}} \sum_{\substack{n\in\mathbb{N}\\ y_n=f(x_k)}}$ est admise (limitations du programme). □

Proposition 11.4.4 (linéarité de l'espérance)
On suppose que les variables réelles X *et* Y *sont d'espérance finie.*
Alors il en est de même des variables $\lambda \mathrm{X}+\mu$ *et* $\lambda \mathrm{X}+\mu \mathrm{Y}$ *pour tous réels* λ *et* μ.
Plus précisément, on a $\mathrm{E}(\lambda \mathrm{X}+\mu)=\lambda \mathrm{E}(\mathrm{X})+\mu$ *et* $\mathrm{E}(\lambda \mathrm{X}+\mu \mathrm{Y})=\lambda \mathrm{E}(\mathrm{X})+\mu \mathrm{E}(\mathrm{Y})$.
Plus généralement encore, si $\mathrm{X}_1, \ldots, \mathrm{X}_n$ *sont d'espérance finie, on a :* $E\Big(\sum_{i=1}^{n} \lambda_i \mathrm{X}_i\Big)=\sum_{i=1}^{n} \lambda_i \mathrm{E}(\mathrm{X}_i)$.

Interprétation :
– l'ensemble des variables aléatoires réelles d'espérance finie sur $(\Omega, \mathscr{A}, \mathbb{P})$ est un $\mathbb{R}$-espace vectoriel.
– l'espérance est une forme linéaire sur cet espace vectoriel.

Pour $\mathrm{E}(\lambda \mathrm{X}+\mu \mathrm{Y})=\lambda \mathrm{E}(\mathrm{X})+\mu \mathrm{E}(\mathrm{Y})$, la démonstration est admise.

On se limite donc ici à la démonstration concernant $\mathrm{S}=\lambda \mathrm{X}+\mu$.

On sait que $\mathrm{S}=\lambda \mathrm{X}+\mu$ est une variable aléatoire réelle.

Tout comme $\sum \mathbb{P}(\mathrm{X}=x_n)\, x_n$ et $\sum \mathbb{P}(\mathrm{X}=x_n)$, la série $\sum \mathbb{P}(\mathrm{X}=x_n)\,(\lambda x_n+\mu)$ converge absolument.

Donc S est d'espérance finie et, d'après le théorème du transfert :

$\mathrm{E}(\mathrm{S})=\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=x_n)\,(\lambda x_n+\mu)=\lambda \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=x_n)\, x_n+\mu \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=x_n)=\lambda \mathrm{E}(\mathrm{X})+\mu.$ □

Exercice 11.25 (⇝ **corrigé**)
Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé. Soit $(A_i)_{1 \leqslant i \leqslant n}$ une famille de n événements.

Montrer que $\mathbb{P}\Big(\bigcup_{1 \leqslant i \leqslant n} A_i\Big)=\sum_{k=1}^{n}(-1)^{k-1} \sum_{\substack{J \subset [\![1, n]\!] \\ \operatorname{Card}(J)=k}} \mathbb{P}\Big(\bigcap_{j \in J} A_j\Big)$ (la somme interne étant étendue à l'ensemble des parties J de $[\![1, n]\!]$ telles que $\operatorname{Card}(J)=k$)

Indication : exprimer la variable indicatrice de $\bigcup_{1 \leqslant i \leqslant n} A_i$ en fonction de celles des événements A_i.

Proposition 11.4.5 (positivité et croissance de l'espérance)
On suppose que les variables X *et* Y *sont d'espérance finie.*
Si X *est à valeurs positives ou nulles, alors* $\mathrm{E}(\mathrm{X}) \geqslant 0$.
Si $\mathrm{X}(\omega) \leqslant \mathrm{Y}(\omega)$ *pour tout* ω *de* Ω, *alors* $\mathrm{E}(\mathrm{X}) \leqslant \mathrm{E}(\mathrm{Y})$.

Si X est à valeurs dans $\mathbb{R}^+$, alors $\mathrm{E}(\mathrm{X})=\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=x_n)\, x_n \geqslant 0$ (tous les termes sont dans $\mathbb{R}^+$).

Si $\mathrm{Y} \geqslant \mathrm{X}$ alors $\mathrm{Y}-\mathrm{X} \geqslant 0$, donc $\mathrm{E}(\mathrm{Y}-\mathrm{X}) \geqslant 0$ donc $\mathrm{E}(\mathrm{Y}) \geqslant \mathrm{E}(\mathrm{X})$. □

Remarques et propriétés

– Si $X \geqslant 0$ alors on ne peut avoir $\mathrm{E}(\mathrm{X})=0$ que si $\mathbb{P}(\mathrm{X}=0)=1$ (c'est-à-dire : X est « presque nulle »).

C'est quasiment la même démonstration que dans le cas des univers finis.

La somme $\mathrm{E}(\mathrm{X})=\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=x_n)\, x_n$ (à termes positifs ou nuls) est nulle donc tous ses termes sont nuls.

Cela implique $\mathbb{P}(\mathrm{X}=x_n)=0$ si $x_n>0$, donc $\mathbb{P}(\mathrm{X}=0)=1$. □

– Si une variable aléatoire réelle X est bornée, alors elle admet une espérance.

On suppose par exemple $|X(\omega)| \leqslant M$ pour tout ω de Ω.

Alors la série $\sum_{n=0}^{+\infty} \mathbb{P}(X=x_n)\,|x_n|$ est majorée par $\sum_{n=0}^{+\infty} \mathbb{P}(X=x_n)M$ dont la somme est M. □

– Si X et Y sont d'espérance finie, il en est de même de $\min(X, Y)$ et de $\max(X, Y)$.

On sait que $X - Y$ donc $|X - Y|$ sont d'espérance finie.

Le résultat découle de : $\min(X, Y) = \frac{1}{2}(X + Y - |X - Y|)$ et $\max(X, Y) = \frac{1}{2}(X + Y + |X - Y|)$. □

– Dire que X est d'espérance finie, c'est dire que $|X|$ est d'espérance finie.

Cela équivaut aussi à dire que X^+ et X^- sont d'espérance finie. On rappelle que $\begin{cases} X^+ = \max(X, 0) \\ X^- = \max(-X, 0) \end{cases}$

Dans la définition 11.4.1, rien ne distingue en effet X de $|X|$.

Si X (donc $|X|$) est d'espérance finie, alors $X^+ = \max(X, 0)$ et $X^- = \max(-X, 0)$ sont d'espérance finie.

Réciproquement, si X^+, X^- sont d'espérance finie, il en est de même de $X = X^+ - X^-$ donc de $|X|$. □

– Si $|X(\omega)| \leqslant Y(\omega)$ pour tout ω, et si Y est d'espérance finie, X est d'espérance finie et $|E(X)| \leqslant E(Y)$.

En particulier, si X est d'espérance finie, on a : $|E(X)| \leqslant E(|X|)$.

Posons $Z = (X, Y)$ et $Z(\Omega) = \{z_n = (x_n, y_n),\ n \in \mathbb{N}\}$.

Pour tout $n \geqslant 0$, il existe $\omega \in \Omega$ tel que : $z_n = Z(\omega)$ càd $(x_n, y_n) = (X(\omega), Y(\omega))$ et donc $|x_n| \leqslant y_n$.

Soit $\pi : \mathbb{R}^2 \to \mathbb{R}$ définie par $\pi(x, y) = x$. Avec cette notation, on a $X = \pi(Z)$.

Dire que $|X|$ est d'espérance finie, c'est dire que la série $\sum_{n=0}^{+\infty} |\pi(z_n)|\,\mathbb{P}(Z=z_n)$ est convergente.

Mais son terme général est $|x_n|\,\mathbb{P}(X=x_n, Y=y_n) \leqslant y_n\,\mathbb{P}(X=x_n, Y=y_n) \leqslant y_n\,\mathbb{P}(Y=y_n)$.

C'est fini car $\sum y_n\,\mathbb{P}(Y=y_n)$ converge, et on obtient $E(|X|) \leqslant E(Y)$ par croissance de l'espérance. □

Proposition 11.4.6 (espérance du produit de deux variables aléatoires réelles indépendantes)
Soit X *et* Y *deux variables indépendantes et d'espérance finie.*
Alors la variable XY *est d'espérance finie et on a :* $E(XY) = E(X)\,E(Y)$.

Démonstration admise. □

11.4.4 Variance d'une variable aléatoire discrète

Dans cette sous-section, X et Y sont des variables aléatoires (discrètes) réelles sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Proposition 11.4.7 (si X^2 est d'espérance finie, alors X est d'espérance finie)
Soit X *une variable aléatoire réelle, telle que* X^2 *soit d'espérance finie.*
Alors la variable X *est elle-même d'espérance finie.*

Pour tout réel x, on a $|x| \leqslant 1 + x^2$ (considérer $|x| \leqslant 1$ et $|x| \geqslant 1$).

Par hypothèse, la série $\sum_{n\geqslant 0} \mathbb{P}(X = x_n) x_n^2$ est convergente (et bien sûr $\sum_{n\geqslant 0} \mathbb{P}(X = x_n)$ aussi !).

Par majoration, il en résulte que $\sum_{n\geqslant 0} \mathbb{P}(X = x_n)\,|x_n|$ converge : ainsi X est d'espérance finie. □

Définition 11.4.2 (variance d'une variable aléatoire discrète)
Soit X une variable aléatoire réelle, telle que X^2 (donc X) soit d'espérance finie. Soit $m = E(X)$.
La variable $Y = (X - m)^2 = X^2 - 2mX + m^2$ est donc d'espérance finie, et $E(Y) = E(X^2) - m^2$.
On appelle *variance* de X le réel : $V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$.
La quantité $\sigma(X) = \sqrt{V(X)}$ est appelée *écart-type* de la variable aléatoire X.

Remarques

– La phrase « X est de variance finie » est un synonyme commode de « X^2 est d'espérance finie ».
On exprimera parfois cette situation en disant que « X possède un moment d'ordre 2 ».
On retiendra que si X est de variance finie alors X est d'espérance finie.
Rappelons enfin cette évidence : le problème « être de variance finie » et « être d'espérance finie » ne se pose pas pour les variables aléatoires réelles X dont l'ensemble image $X(\Omega)$ est fini !
– La variance et l'écart-type mesure la « dispersion » autour de l'espérance.
– Soit X une variable aléatoire réelle, telle que X^2 (donc X) soit d'espérance finie.
Pour tous réels a, b on a : $V(aX + b) = a^2 V(X)$, donc $\sigma(aX + b) = |a|\,\sigma(X)$.

C'est la même démonstration que dans le cas des univers finis.
Si $Y = aX + b$, alors $E(Y) = aE(X) + b$ donc $Y - E(Y) = a(X - E(X))$.
Ainsi $V(Y) = E((Y - E(Y))^2) = E(a^2(X - E(X))^2) = a^2 E((X - E(X))^2) = a^2 V(X)$. □

– On utilise l'écart-type pour des raisons d'homogénéité par rapport aux valeurs de X.

Proposition 11.4.8 (caractérisation de la variance nulle)
Une variable aléatoire est de variance nulle si et seulement si elle est presque constante.

C'est la même démonstration que dans le cas des univers finis.
Posons $m = E(X)$. La variable $(X - m)^2 \geqslant 0$ est à valeurs positives.
Ainsi : $V(X) = 0 \Leftrightarrow E((X - m)^2) = 0 \Leftrightarrow \mathbb{P}((X - m)^2 = 0) = 1 \Leftrightarrow \mathbb{P}(X = m) = 1$. □

NB : la notion de variable presque constante a été introduite dans la sous-section 11.2.2.

11.4.5 Variance et indépendance, covariance

Dans cette sous-section, X et Y sont des variables aléatoires réelles sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Proposition 11.4.9 (espérance du produit de deux variables aléatoires de variance finie)
Soit X *et* Y *deux variables aléatoires, toutes les deux de variance finie.*
Alors la variable XY *est d'espérance finie et* $|E(XY)| \leqslant \sqrt{E(X^2)}\sqrt{E(Y^2)}$ *(Cauchy-Schwarz).*

On admet l'existence de l'espérance de XY. Posons $A(\lambda) = \mathrm{E}((\lambda \mathrm{X} - \mathrm{Y})^2)$.

Si $\mathrm{E}(\mathrm{X}^2) = 0$, alors X^2, donc X, donc XY, sont presque nulles, et l'inégalité demandée est vraie.

On suppose donc $\mathrm{E}(\mathrm{X}^2) > 0$.

Pour tout réel λ, on a : $A(\lambda) = \mathrm{E}(\lambda^2\mathrm{X}^2 - 2\lambda\mathrm{XY} + \mathrm{Y}^2) = \lambda^2\mathrm{E}(\mathrm{X}^2) - 2\lambda\,\mathrm{E}(\mathrm{XY}) + \mathrm{E}(\mathrm{Y}^2)$.

Ce trinôme $A(\lambda)$ reste positif, donc son discriminant $\Delta = \mathrm{E}^2(\mathrm{XY}) - \mathrm{E}(\mathrm{X}^2)\mathrm{E}(\mathrm{Y}^2)$ est négatif ou nul.

Mais écrire $\Delta \leqslant 0$, c'est écrire $|\mathrm{E}(\mathrm{XY})| \leqslant \sqrt{\mathrm{E}(\mathrm{X}^2)}\sqrt{\mathrm{E}(\mathrm{Y}^2)}$.

Il y a égalité si et seulement si $\Delta = 0 \;\Leftrightarrow\; \exists\, a \in \mathbb{R},\ \mathrm{E}((a\mathrm{X} - Y)^2) = 0$.

Mais $\mathrm{E}((a\mathrm{X} - Y)^2) = 0$ signifie que $a\mathrm{X} - \mathrm{Y}$ est « presque nulle » càd : $\mathbb{P}(\mathrm{Y} = a\mathrm{X}) = 1$. □

Définition 11.4.3 (covariance de deux variables de variance finie)
Soit X et Y deux variables aléatoires, toutes les deux de variance (donc d'espérance) finie.
On nomme *covariance de* X *et* Y l'espérance, notée $\mathrm{Cov}(\mathrm{X}, \mathrm{Y})$, de la variable $(\mathrm{X}-\mathrm{E}(\mathrm{X}))(\mathrm{Y}-\mathrm{E}(\mathrm{Y}))$.
Par un simple développement, on trouve : $\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = \mathrm{E}(\mathrm{XY}) - \mathrm{E}(\mathrm{X})\,\mathrm{E}(\mathrm{Y})$.

Posons $\alpha = \mathrm{E}(\mathrm{X})$ et $\beta = \mathrm{E}(\mathrm{Y})$.
La variable $\mathrm{Z} = (\mathrm{X}-\alpha)(\mathrm{Y}-\beta)$ se développe en $\mathrm{Z} = \mathrm{XY} - \alpha\mathrm{Y} - \beta\mathrm{X} + \alpha\beta$.
Par hypothèse, X et Y sont de variance finie, donc X, Y, XY, et donc Z sont d'espérance finie.
Par linéarité :
$\mathrm{E}(\mathrm{Z}) = \mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = \mathrm{E}(\mathrm{XY} - \alpha\mathrm{Y} - \beta\mathrm{X} + \alpha\beta) = \mathrm{E}(\mathrm{XY}) - \alpha\mathrm{E}(\mathrm{Y}) - \beta\mathrm{E}(\mathrm{X}) + \alpha\beta = \mathrm{E}(\mathrm{XY}) - \alpha\beta$. □

Remarques

- Si X et Y sont indépendantes, leur covariance est nulle (réciproque fausse).

  Simple conséquence de la définition de la covariance et de la proposition 11.4.6. □

- En particulier, si Y est constante (ou « presque constante », cf 11.2.2), alors $\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = 0$.
- Sous réserve d'existence, on a $\begin{cases} \mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = \mathrm{Cov}(\mathrm{Y}, \mathrm{X}) \\ \mathrm{Cov}(a\mathrm{X}+b\mathrm{Y}, Z) = a\,\mathrm{Cov}(\mathrm{X}, Z) + b\,\mathrm{Cov}(\mathrm{Y}, Z) \end{cases}$

  Simple conséquence de la définition de la covariance et de la linéarité de l'espérance (proposition 11.4.4).
  La symétrie et la linéarité à gauche impliquent la linéarité à droite, donc la bilinéarité. □

Proposition 11.4.10 (variance d'une somme de deux variables aléatoires)
Soit X *et* Y *deux variables aléatoires, toutes les deux de variance (donc d'espérance) finie.*
Alors $a\mathrm{X} + b\mathrm{Y}$ *est de variance finie et* $\mathrm{V}(a\mathrm{X} + b\mathrm{Y}) = a^2\,\mathrm{V}(\mathrm{X}) + 2\,ab\,\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) + b^2\,\mathrm{V}(\mathrm{Y})$.
En particulier, si X *et* Y *sont indépendantes, on a :* $\mathrm{V}(\mathrm{X} + \mathrm{Y}) = \mathrm{V}(\mathrm{X}) + \mathrm{V}(\mathrm{Y})$ *(réciproque fausse).*

C'est la même démonstration que dans le cas des univers finis.

$$\begin{aligned} \mathrm{V}(a\mathrm{X} + b\mathrm{Y}) &= \mathrm{Cov}(a\mathrm{X} + b\mathrm{Y}, a\mathrm{X} + b\mathrm{Y}) = a^2\mathrm{Cov}(\mathrm{X}, \mathrm{X}) + ab\,\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) + ba\,\mathrm{Cov}(\mathrm{Y}, \mathrm{X}) + b^2\mathrm{Cov}(\mathrm{Y}, \mathrm{Y}) \\ &= a^2\mathrm{V}(\mathrm{X}) + 2ab\,\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) + b^2\mathrm{V}(\mathrm{Y}). \end{aligned}$$

Si X et Y sont indépendantes, on a $\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) = 0$ donc $\mathrm{V}(\mathrm{X} + \mathrm{Y}) = \mathrm{V}(\mathrm{X}) + \mathrm{V}(\mathrm{Y})$ □

Interprétation vectorielle

- L'ensemble des variables aléatoires réelles de variance finie sur $(\Omega, \mathscr{A}, \mathbb{P})$ est un $\mathbb{R}$-espace vectoriel ; C'est un sous-espace de l'espace des variables aléatoires réelles d'espérance finie sur $(\Omega, \mathscr{A}, \mathbb{P})$.
- La covariance est une forme bilinéaire symétrique positive sur l'espace des variables de variance finie.

On peut généraliser la proposition 11.4.10 à une somme finie de variables aléatoires :

Proposition 11.4.11 (variance d'une somme de plusieurs variables aléatoires)
Soit $(\mathrm{X}_i)_{1 \leqslant i \leqslant n}$ *une suite finie de variables aléatoires de variance finie.*
Alors on a l'égalité $\mathrm{V}\Big(\sum_{i=1}^{n} \mathrm{X}_i\Big) = \sum_{i=1}^{n} \mathrm{V}(\mathrm{X}_i) + 2\sum_{i<j} \mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_j)$.
En particulier si les X_i *sont indépendantes deux à deux, on a* $\mathrm{V}\Big(\sum_{i=1}^{n} \mathrm{X}_i\Big) = \sum_{i=1}^{n} \mathrm{V}(\mathrm{X}_i)$.

Clair par développement, bilinéarité et symétrie de la covariance, et les égalités $\mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_i) = \mathrm{V}(\mathrm{X}_i)$. □

Exercice 11.26 (⇝ **corrigé**)
Soient $a \in \mathbb{N}^*$, $n \geqslant 2$. On répartit au hasard, et de façon équiprobable, an boules dans n urnes.
Pour tout $i \in [\![1, n]\!]$, soit T_i la variable indicatrice de l'événement « l'urne numéro i est vide ».
On note Y_n le nombre d'urnes restant vides après la répartition des N boules, et on pose $S_n = \dfrac{Y_n}{n}$.

1. Donner la loi, l'espérance et la variance de T_i.
2. Déterminer l'espérance et la variance de S_n ainsi que leur limite quand n tend vers $+\infty$.

Définition 11.4.4 (coefficient de corrélation de deux variables de variance > 0)
Soit X, Y deux variables aléatoires réelles de variance non nulle.
On appelle coefficient de corrélation (linéaire) de X et Y la quantité : $\rho(\mathrm{X}, \mathrm{Y}) = \dfrac{\mathrm{Cov}(\mathrm{X}, \mathrm{Y})}{\sigma(\mathrm{X})\sigma(\mathrm{Y})}$.

Proposition 11.4.12 (inégalité de Cauchy-Schwarz)
Soit X, Y *deux variables aléatoires réelles de variance non nulle.*
On a toujours $|\rho(\mathrm{X}, \mathrm{Y})| \leqslant 1$ *(inégalité de Cauchy-Schwarz).*
On a l'égalité $|\rho(\mathrm{X}, \mathrm{Y})| = 1$ *si et seulement si il existe* a *dans* $\mathbb{R}^*$ *(du signe de* $\rho(\mathrm{X}, \mathrm{Y})$*) et* b *dans* $\mathbb{R}$ *tels que l'événement* $(\mathrm{Y} = a\mathrm{X} + b)$ *soit presque sûr.*

Dire que l'événement $(\mathrm{Y} = a\mathrm{X} + b)$ est presque sûr, c'est dire la variable $\mathrm{Y} - a\mathrm{X} - b$ est presque nulle.

C'est strictement la même démonstration que dans le cas des univers finis.
Pour tout réel λ, on a $A(\lambda) = \mathrm{V}(\lambda\mathrm{X} - Y) = \lambda^2\mathrm{V}(\mathrm{X}) - 2\lambda\,\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) + \mathrm{V}(\mathrm{Y})$.
Ce trinôme $A(\lambda)$ reste positif, donc son discriminant $\Delta = \mathrm{Cov}(\mathrm{X}, \mathrm{Y})^2 - \mathrm{V}(\mathrm{X})\mathrm{V}(\mathrm{Y})$ est négatif ou nul.
Mais écrire $\Delta \leqslant 0$, c'est écrire $|\rho(\mathrm{X}, \mathrm{Y})| \leqslant 1$.
Ensuite : $|\rho(\mathrm{X}, \mathrm{Y})| = 1 \Leftrightarrow \Delta = 0 \Leftrightarrow \exists\, a \in \mathbb{R},\ \mathrm{V}(a\mathrm{X} - Y) = 0$.
Avec ces notations, la racine double de $A(\lambda)$ est $a = \dfrac{\mathrm{Cov}(\mathrm{X}, \mathrm{Y})}{\mathrm{V}(\mathrm{X})}$ donc a et $\rho(\mathrm{X}, \mathrm{Y})$ sont de même signe.
Mais $\mathrm{V}(a\mathrm{X} - \mathrm{Y}) = 0$ signifie que $a\mathrm{X} - \mathrm{Y}$ est « presque sûre » càd : $\exists\, b \in \mathbb{R},\ \mathbb{P}(\mathrm{Y} = a\mathrm{X} + b) = 1$. □

Exercice 11.27 (⇝ **corrigé**)

Soit X, Y deux v.a.r. indépendantes suivant la loi binomiale de paramètres n et $\frac{1}{2}$.

1. Calculer l'espérance et la variance de $Z = X - Y$. Donner la loi de Z.
2. Les variables X et Z sont-elles indépendantes ?
   Calculer la covariance et le coefficient de corrélation de X et Z.

11.5 Variables aléatoires à valeurs dans $\mathbb{N}$

Définition 11.5.1 (fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$)
Soit X une variable aléatoire sur $(\Omega, \mathscr{A}, \mathbb{P})$, telle que $\mathrm{X}(\Omega) \subset \mathbb{N}$.
On appelle *fonction génératrice* de X la série entière $\sum \mathbb{P}(\mathrm{X}=n)\, t^n$ de la variable réelle t.
On note $\mathrm{G_X}(t)$ la somme de cette série entière.

▷ Cas des lois usuelles à image finie

Si $\mathrm{X}(\Omega) \subset \mathbb{N}$ est fini, alors $\mathrm{G_X}$ est un simple polynôme (le « rayon de convergence » est infini).

- si X est constante, de valeur $a \in \mathbb{N}$, alors $\mathrm{G_X}(t) = t^a$.
- si X suit la loi de Bernoulli de paramètre p (donc $\mathrm{X} \hookrightarrow \mathcal{B}(p)$), alors $\mathrm{G_X}(t) = pt + q$ (avec $q = 1 - p$).
- si X suit la loi uniforme sur $[\![a, b]\!] \subset \mathbb{N}^2$, alors $\mathrm{G_X}(t) = \dfrac{1}{b-a+1}(t^a + t^{a+1} + \cdots + t^b)$.
- si X suit la loi binomiale $\mathcal{B}(n, p)$ de paramètres n, p, alors $\mathrm{G_X}(t) = (pt + q)^n$ (en posant $q = 1 - p$).

On a : $\mathrm{X}(\Omega) = [\![0, n]\!]$ et : $\forall k \in [\![0, n]\!],\ \mathbb{P}(\mathrm{X}=k) = \dbinom{n}{k} p^k q^{n-k}$.

Ainsi $\mathrm{G_X}(t) = \displaystyle\sum_{k=0}^{n} \mathbb{P}(\mathrm{X}=k)\, t^k = \sum_{k=0}^{n} \binom{n}{k} (pt)^k q^{n-k} = (pt+q)^n$. □

La notion de fonction génératrice ne prend tout son sens que quand l'ensemble $\mathrm{X}(\Omega)$ est dénombrable.

▷ Remarques importantes

- D'après le théorème du transfert (voir proposition 11.4.3), on a : $\mathrm{G_X}(t) = \mathrm{E}(t^{\mathrm{X}})$.
- Les séries $\displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=n)(-1)^n$ et $\displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=n)$ convergent absolument (la 2$^{\text{nde}}$ est de somme 1).
  Il en découle que le rayon de convergence de la fonction génératrice de X est au moins égal à 1.
  Il en découle aussi que la série $t \mapsto \mathrm{G_X}(t) = \displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=n)\, t^n$ est normalement convergente sur $[-1, 1]$.

Proposition 11.5.1 (la fonction génératrice caractérise la loi d'une variable aléatoire)
Soit $\mathrm{G_X}$ *la fonction génératrice d'une variable aléatoire* X *à valeurs dans* $\mathbb{N}$.
Pour tout n *de* $\mathbb{N}$, *on a l'égalité :* $\mathbb{P}(\mathrm{X}=n) = \dfrac{1}{n!}\mathrm{G_X^{(n)}}(0)$.

On sait que la somme d'une série entière réelle $S(t) = \sum_{n=0}^{+\infty} a_n t^n$ et de rayon $r > 0$ est $\mathcal{C}^\infty$ sur $]-r, r[$.

On sait également que : $\forall n \in \mathbb{N},\ a_n = \dfrac{S^{(n)}(0)}{n!}$ (voir propositions 8.2.2 et 8.2.9).

Ainsi, pour tout n de $\mathbb{N}$, on a l'égalité : $\mathbb{P}(\mathrm{X}=n) = \dfrac{\mathrm{G}_\mathrm{X}^{(n)}(0)}{n!}$. □

Morale : *La loi d'une variable* X *telle que* $\mathrm{X}(\Omega) \subset \mathbb{N}$ *est caractérisée par sa fonction génératrice* G_X.

Exercice 11.28 (⇝ **corrigé**)

Soit $(X_1, \ldots, X_n)$ une famille de variables indépendantes de même loi définie par :

$$\mathbb{P}(X_k = 0) = \frac{8}{27},\quad \mathbb{P}(X_k = 1) = \frac{12}{27},\quad \mathbb{P}(X_k = 2) = \frac{6}{27},\quad \mathbb{P}(X_k = 3) = \frac{1}{27}$$

Déterminer la loi de $S = X_1 + \cdots + X_n$.

Exercice 11.29 (⇝ **corrigé**)

Pierre et Marie effectuent une suite de parties indépendantes d'un jeu, numérotées 1, 2, $\cdots$.

La probabilité que Pierre gagne une partie est $p \in [0,1]$, donc celle de Marie est $q = 1 - p$.

On note a_{2n} la probabilité qu'il y a ait égalité à la $(2n)$-ième partie.

On note b_{2n} la probabilité que la première égalité soit à la $(2n)$-ième partie.

On considère les séries entières : $A(x) = \sum_{n=1}^{+\infty} a_{2n} x^{2n}$ et $B(x) = \sum_{n=1}^{+\infty} b_{2n} x^{2n}$.

1. Exprimer a_{2n} en fonction de n.
2. Déterminer le rayon de convergence R de la série entière définissant $A(x)$.
   Pour quelles valeurs de p la fonction A est-elle définie en 1 ?
3. Monter que : $\forall\, x \in]-R, R[,\ A(x) = \dfrac{1}{\sqrt{1 - 4pqx^2}} - 1$
4. Établir une identité vérifiée par A et B, puis expliciter $B(x)$.
5. Déterminer la probabilité qu'il n'y ait jamais égalité.

Proposition 11.5.2 (lien entre espérance et dérivée en 1 de la fonction génératrice)

Soit X *une variable aléatoire sur* $(\Omega, \mathscr{A}, \mathbb{P})$, *à valeurs dans* $\mathbb{N}$.

Alors X *est d'espérance finie si et seulement si* G_X *est dérivable en* 1.

Si tel est le cas, on a alors $\mathrm{E}(\mathrm{X}) = \mathrm{G}'_\mathrm{X}(1)$.

Remarquons que dire que X est d'espérance finie, c'est dire que la série $\mathrm{E}(\mathrm{X}) = \sum_{n\geqslant 1} \mathbb{P}(\mathrm{X}=n)\, n$ converge.

- Il y a une preuve simple de l'égalité $\mathrm{E}(\mathrm{X}) = \mathrm{G}'_\mathrm{X}(1)$ mais avec des hypothèses plus fortes.

  On suppose que le rayon de convergence R de la série entière $\mathrm{G}_\mathrm{X}(t) = \sum_{n\geqslant 0} \mathbb{P}(\mathrm{X}=n)\, t^n$ vérifie $R > 1$.

  Alors sur $]-R, R\,[$, G_X est $\mathcal{C}^\infty$ et, $\mathrm{G}'_\mathrm{X}(t) = \sum_{n\geqslant 1} n\, \mathbb{P}(\mathrm{X}=n)\, t^{n-1}$.

  En particulier la série $\mathrm{G}'_\mathrm{X}(1) = \sum_{n\geqslant 1} \mathbb{P}(\mathrm{X}=n)\, n$ converge donc $\mathrm{E}(\mathrm{X}) = \mathrm{G}'_\mathrm{X}(1)$.

- On revient aux hypothèses de départ.

Tout ce qu'on peut dire sur le rayon de convergence R de $\mathrm{G_X}(t) = \sum_{n\geqslant 0} \mathbb{P}(\mathrm{X}=n)\, t^n$ est $R \geqslant 1$.

On suppose que X est d'espérance finie, c'est-à-dire que $\sum_{n\geqslant 1} \mathbb{P}(\mathrm{X}=n)\, n$ converge.

La série $t \mapsto \sum_{n\geqslant 1} \mathbb{P}(\mathrm{X}=n)\, n\, t^{n-1}$ est alors normalement, donc uniformément convergente, sur $[0, 1]$.

On peut alors appliquer le théorème 7.4.4 de dérivation des séries de fonctions de classe $\mathcal{C}^1$.

Sur $I = [0, 1]$, la fonction $t \mapsto \mathrm{G_X}(t) = \sum_{n=0}^{+\infty} P(\mathrm{X}=n)\, t^n$ est donc $\mathcal{C}^1$ et $\mathrm{G'_X}(t) = \sum_{n=1}^{+\infty} n\, \mathbb{P}(\mathrm{X}=n)\, t^{n-1}$.

En particulier : $\mathrm{G_X}$ est dérivable (à gauche) en 1 et $\mathrm{G'_X}(1) = \sum_{n=1}^{+\infty} n\, \mathbb{P}(\mathrm{X}=n) = \mathrm{E(X)}$.

– Réciproquement, on suppose que l'application $\mathrm{G_X}$ est dérivable (à gauche) en 1.

La fonction $t \mapsto \mathrm{G'_X}(t) = \sum_{n=1}^{+\infty} n\, \mathbb{P}(\mathrm{X}=n) t^{n-1}$ est positive croissante sur $[0, 1[$.

Elle possède donc une limite ℓ quand $t \to 1^-$, et nécessairement $\ell = \mathrm{G'_X}(1)$.

Pour $N \in \mathbb{N}$, et $0 \leqslant t < 1$, on a : $\sum_{n=1}^{N} n\, \mathbb{P}(\mathrm{X}=n) t^{n-1} \leqslant \mathrm{G'_X}(t) \leqslant \mathrm{G'_X}(1)$, donc $\sum_{n=1}^{N} n\, \mathbb{P}(\mathrm{X}=n) \leqslant \mathrm{G'_X}(1)$.

La série positive $\sum_{n=1}^{+\infty} n\, \mathbb{P}(\mathrm{X}=n)$ est donc convergente, ce qui ramène au sens direct. □

Exercice 11.30 (⇝ **corrigé**)

Soit X une variable aléatoire telle que $\forall\, k \in \mathbb{N}^*$, $\mathbb{P}(X = k) = \dfrac{k-1}{2^k}$.

1. Vérifier par le calcul que $\sum_{k=1}^{+\infty} \mathbb{P}(X = k) = 1$.
2. Donner la fonction génératrice de X. Quel est son rayon de convergence ?
3. La variable X admet-elle une espérance finie ? Si oui, que vaut-elle ?

Proposition 11.5.3 (lien entre variance et dérivée seconde en 1 de la fonction génératrice)

La variable aléatoire X *est de variance finie si et seulement si* $\mathrm{G_X}$ *est deux fois dérivable en* 1.

Si tel est le cas, on a alors : $\mathrm{G''_X}(1) = \mathrm{E(X(X-1))}$ *donc* $\mathrm{E(X^2)} = \mathrm{G''_X}(1) + \mathrm{G'_X}(1)$.

On en déduit enfin $V(\mathrm{X}) = \mathrm{E(X^2)} - \mathrm{E^2(X)} = \mathrm{G''_X}(1) + \mathrm{G'_X}(1) - (\mathrm{G'_X}(1))^2$.

Soit R le rayon de convergence des séries $\sum_{n\geqslant 0} \mathbb{P}(\mathrm{X}=n)\, t^n$, $\sum_{n\geqslant 1} n\mathbb{P}(\mathrm{X}=n)\, t^{n-1}$ et $\sum_{n\geqslant 2} n(n-1)\mathbb{P}(\mathrm{X}=n)\, t^{n-2}$

On sait que $R \geqslant 1$, et on admet le résultat si $R = 1$. On suppose donc $R > 1$.

La dérivation terme à terme est donc possible au voisinage de 1.

En particulier $\mathrm{G'_X}(1) = \sum_{n\geqslant 0} \mathbb{P}(\mathrm{X}=n)\, n = \mathrm{E(X)}$ et $\mathrm{G''_X}(1) = \sum_{n\geqslant 0} \mathbb{P}(\mathrm{X}=n)\, n(n-1) = \mathrm{E(X^2)} - \mathrm{E(X)}$.

On trouve donc $\mathrm{V(X)} = \mathrm{E(X^2)} - \mathrm{E^2(X)} = \mathrm{G''_X}(1) + \mathrm{G'_X}(1) - (\mathrm{G'_X}(1))^2$. □

Proposition 11.5.4 (série génératrice d'une somme de deux variables indépendantes)

Soit X *et* Y *deux variables aléatoires indépendantes sur* $(\Omega, \mathscr{A}, \mathbb{P})$*, à valeurs dans* $\mathbb{N}$.

Soit r *le minimum des rayons de convergence des fonctions génératrices* $\mathrm{G_X}$ *et* $\mathrm{G_Y}$.

Alors la variable X + Y *(qui est à valeurs dans* $\mathbb{N}$*) vérifie :* $\forall\, t \in]-r, r[$, $\mathrm{G_{X+Y}}(t) = \mathrm{G_X}(t)\, \mathrm{G_Y}(t)$.

On voit deux démonstrations.

– D'abord avec un produit de Cauchy de deux séries entières (voir proposition 8.1.10).

Posons $s_n = \mathbb{P}(X+Y=n)$, $x_n = \mathbb{P}(X=n)$ et $y_n = \mathbb{P}(Y=n)$.

On a $s_n = \sum\limits_{p+q=n} \mathbb{P}(X=p, Y=q) = \sum\limits_{p+q=n} \mathbb{P}(X=p)\,\mathbb{P}(Y=q) = \sum\limits_{p+q=n} x_p\, y_q$ (par indépendance de X et Y).

Pour $|t| < r$, on a alors l'égalité : $\sum\limits_{n=0}^{+\infty} s_n t^n = \Bigl(\sum\limits_{n=0}^{+\infty} x_n t^n\Bigr)\Bigl(\sum\limits_{n=0}^{+\infty} y_n t^n\Bigr)$, c'est-à-dire $G_{X+Y}(t) = G_X(t)\,G_Y(t)$.

– Deuxième démonstration : X et Y sont indépendantes donc t^X et t^Y sont indépendantes (cf prop.11.3.6).

On finit avec 11.4.6 : si $|t| < r$, $G_{X+Y}(t) = E(t^{X+Y}) = E(t^X\, t^Y) = E(t^X)\, E(t^Y) = G_X(t)\, G_Y(t)$. □

11.6 Loi géométrique et loi de Poisson

11.6.1 Loi géométrique de paramètre p, notation $X \hookrightarrow \mathcal{G}(p)$

Définition 11.6.1 (loi géométrique de paramètre p)

Soit X une variable aléatoire sur $(\Omega, \mathscr{A}, \mathbb{P})$, et $0 < p < 1$. Soit $q = 1 - p$.

On dit que X suit une loi géométrique de paramètre p (on note $X \hookrightarrow \mathcal{G}(p)$) si $\begin{cases} X(\Omega) = \mathbb{N}^* \\ \forall\, n \geqslant 1,\ \mathbb{P}(X=n) = pq^{n-1} \end{cases}$

Interprétation et remarques

– **Interprétation** : la loi géométrique $\mathcal{G}(p)$ peut être interprétée comme le rang du premier succès dans une suite illimitée d'épreuves de Bernoulli indépendantes et de même paramètre p.

– Pour être précis, on a $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$, en considérant l'événement E_∞ : « toutes les épreuves de Bernoulli se soldent par un échec ». Mais cet événement, non impossible, est négligeable.

– Si $X \hookrightarrow \mathcal{G}(p)$, la variable $Y = X - 1$ représente le « nombre d'échecs avant un premier succès ».
On a bien sûr $Y(\Omega) = \mathbb{N}$ et : $\forall\, n \in \mathbb{N},\ \mathbb{P}(Y=n) = \mathbb{P}(X=n+1) = p\, q^n$.

Proposition 11.6.1 (propriétés de la loi géométrique de paramètre p)

Soit X une variable aléatoire suivant la loi géométrique $\mathcal{G}(p)$ de paramètre $0 < p < 1$. Soit $q = 1 - p$.

– *sa fonction génératrice G_X est de rayon de convergence $\dfrac{1}{q}$ et : $\forall\, t \in \Bigl]-\dfrac{1}{q}, \dfrac{1}{q}\Bigr[$, $G_X(t) = \dfrac{pt}{1-qt}$.*

– *son espérance est $\dfrac{1}{p}$, et sa variance est $\dfrac{q}{p^2}$.*

On a $G_X(t) = \sum\limits_{n=1}^{+\infty} \mathbb{P}(X=n)\, t^n = \sum\limits_{n=1}^{+\infty} p\, q^{n-1} t^n = p\, t \sum\limits_{n=1}^{+\infty} (qt)^{n-1} = \dfrac{pt}{1-qt}$ avec convergence si $q\,|t| < 1$.

Ici le rayon de convergence est $\dfrac{1}{q} > 1$, et on peut appliquer les propositions 11.5.2 et 11.5.3.

On a $G_X(t) = \dfrac{p}{q}\Bigl(\dfrac{1}{1-qt} - 1\Bigr)$ donc $G'_X(t) = \dfrac{p}{(1-qt)^2}$ et $G''_X(t) = \dfrac{2pq}{(1-qt)^3}$.

Ainsi $E(X) = G'_X(1) = \dfrac{1}{p}$ et $G''_X(1) = E(X(X-1)) = \dfrac{2q}{p^2}$ donc $E(X^2) = G''_X(1) + E(X) = \dfrac{2q}{p^2} + \dfrac{1}{p} = \dfrac{1+q}{p^2}$.

On en déduit enfin : $V(X) = E(X^2) - E^2(X) = \dfrac{1+q}{p^2} - \dfrac{1}{p^2} = \dfrac{q}{p^2}$. □

Définition 11.6.2 (loi « sans mémoire »)
Soit X une variable aléatoire sur $(\Omega, \mathscr{A}, \mathbb{P})$, à valeurs dans $\mathbb{N}^*$.
Pour tout n de $\mathbb{N}$, on suppose que l'événement $\mathbb{P}(X > n)$ est de probabilité strictement positive.
On dit que X est « sans mémoire » si : $\forall\, d \in \mathbb{N},\ \forall\, n \in \mathbb{N},\ \mathbb{P}(X > d+n \mid X > d) = \mathbb{P}(X > n)$.

Proposition 11.6.2 (caractérisation des lois géométriques comme étant les lois sans mémoire)
Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$. On pose $p = \mathbb{P}(X=1)$ et on suppose $0 < p < 1$.
Alors X est sans mémoire et si seulement si elle suit la loi géométrique de paramètre p.

- On suppose que $X \hookrightarrow \mathcal{G}(p)$, avec $0 < p < 1$.

  Pour tout n de $\mathbb{N}$, on a $\mathbb{P}(X > n) = \displaystyle\sum_{k=n+1}^{+\infty} \mathbb{P}(X=k) = \sum_{k=n+1}^{+\infty} pq^{k-1} = pq^n \frac{1}{1-q} = q^n$.

  Ainsi : $\mathbb{P}(X > d+n \mid X > d) = \dfrac{\mathbb{P}((X > d+n) \cap (X > d))}{\mathbb{P}(X > d)} = \dfrac{\mathbb{P}(X > d+n)}{\mathbb{P}(X > d)} = \dfrac{q^{d+n}}{q^d} = q^n = \mathbb{P}(X > n)$.

  La loi géométrique $\mathcal{G}(p)$ de paramètre p (avec $0 < p < 1$) est donc sans mémoire.

- Réciproquement, on se donne une variable aléatoire X à valeurs dans $\mathbb{N}^*$.

  On pose $p = \mathbb{P}(X=1)$. On suppose que $0 < p < 1$ et que X est sans mémoire.

  On note $q = 1 - p = 1 - \mathbb{P}(X=1) = \mathbb{P}(X>1)$.

  Pour tout entier strictement positif d, on a : $q = \mathbb{P}(X > 1) = \dfrac{\mathbb{P}((X > n+1) \cap (X > n))}{\mathbb{P}(X > n)} = \dfrac{\mathbb{P}(X > n+1)}{\mathbb{P}(X > n)}$.

  Ainsi la suite $n \mapsto \mathbb{P}(X > n)$ (de premier terme q) est géométrique de raison q.

  Il en découle : $\mathbb{P}(X > n) = q^n$, donc $\mathbb{P}(X = n) = \mathbb{P}(X > n-1) - \mathbb{P}(X > n-1) = q^{n-1} - q^n = pq^{n-1}$.

  On a donc vérifié que X suit la loi $\mathcal{G}(p)$. □

Exercice 11.31 (⇝ **corrigé**)
Soit X et Y deux variables indépendantes de loi géométrique de paramètre p.
Déterminer la loi et l'espérance de $Z = |X - Y|$.

Exercice 11.32 (⇝ **corrigé**)
Soit N une variable aléatoire telle que $N+1$ suive une loi géométrique de paramètre p.
On considère une urne contenant une boule bleue et une boule verte.
On effectue N tirages avec remise et on note X le nombre de boules vertes tirées.
1) Trouver le rayon de convergence et l'expression de $f : x \mapsto \displaystyle\sum_{n=m}^{+\infty} \binom{n}{m} x^n$. 2) Trouver la loi de X.

11.6.2 Loi de Poisson de paramètre λ, notation $X \hookrightarrow \mathcal{P}(p)$

Définition 11.6.3 (loi de Poisson de paramètre λ)
Soit X une variable aléatoire sur $(\Omega, \mathscr{A}, \mathbb{P})$, et soit λ dans $\mathbb{R}^{+*}$.
On dit que X suit la loi de Poisson de paramètre λ et on note $X \hookrightarrow \mathcal{P}(\lambda)$ si $\begin{cases} X(\Omega) = \mathbb{N} \\ \forall\, n \in \mathbb{N},\ \mathbb{P}(X=n) = e^{-\lambda} \dfrac{\lambda^n}{n!} \end{cases}$

Proposition 11.6.3 (propriétés de la loi de Poisson de paramètre λ)
Soit X *une variable aléatoire suivant la loi de Poisson* $\mathcal{P}(\lambda)$, *avec* λ *dans* $\mathbb{R}^{+*}$.
– *sa fonction génératrice* $\mathrm{G_X}$ *est de rayon de convergence* $+\infty$ *et :* $\forall t \in \mathbb{R},\ \mathrm{G_X}(t) = \mathrm{e}^{\lambda(t-1)}$.
– *son espérance est* λ, *et sa variance est* λ.

On a $\mathrm{G_X}(t) = \displaystyle\sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X}=n)\, t^n = \mathrm{e}^{-\lambda} \sum_{n=0}^{+\infty} \frac{(\lambda t)^n}{n!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda t} = \mathrm{e}^{\lambda(t-1)}$ avec rayon de convergence $R = +\infty$.

Ici $R > 1$ donc on peut appliquer les propositions 11.5.2 et 11.5.3.

Pour tout réel t, on a : $\mathrm{G}'_{\mathrm{X}}(t) = \lambda \mathrm{e}^{\lambda(t-1)}$ et $\mathrm{G}''_{\mathrm{X}}(t) = \lambda^2 \mathrm{e}^{\lambda(t-1)}$.

Ainsi $\mathrm{E(X)} = \mathrm{G}'_{\mathrm{X}}(1) = \lambda$ et $\mathrm{G}''_{\mathrm{X}}(1) = \mathrm{E}(\mathrm{X}(\mathrm{X}-1)) = \lambda^2$ donc $\mathrm{E}(\mathrm{X}^2) = \lambda^2 + \lambda$.

On en déduit enfin : $\mathrm{V(X)} = \mathrm{E}(\mathrm{X}^2) - \mathrm{E}^2(\mathrm{X}) = \lambda^2 + \lambda - \lambda^2 = \lambda$. □

Interprétation :
Les lois de Poisson permettent de modéliser le nombre de réalisations d'un « événement rare » E sur une longue période, si on connait le nombre moyen de réalisations de E par unité de temps (et si on considère que les réalisations de E sont indépendantes les unes des autres).

Cette interprétation est validée par le résultat suivant :

Proposition 11.6.4 (approximation de la loi binomiale par la loi de Poisson)
Soit $(\mathrm{X}_n)_{n\geqslant 0}$ *une suite de variables telle que* $\mathrm{X}_n \hookrightarrow \mathcal{B}(n, p_n)$, *avec* $\displaystyle\lim_{n\to+\infty} n\, p_n = \lambda$.
Alors, pour tout k *de* $\mathbb{N}$, *on a :* $\displaystyle\lim_{n\to+\infty} \mathbb{P}(\mathrm{X}_n = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$

Posons $\lambda_n = np_n$. Par hypothèse, on a $\displaystyle\lim_{n\to+\infty} \lambda_n = \lambda$.

Pour tout k de $\{0, \ldots, n\}$, on a : $\mathbb{P}(\mathrm{X}_n = k) = \displaystyle\binom{n}{k} p_n^k\, (1-p_n)^{n-k} = \frac{1}{k!}(1-p_n)^{n-k}\, p_n^k \prod_{j=0}^{k-1}(n-j)$.

Or le produit $\displaystyle\prod_{j=0}^{k-1}(n-j)$ est équivalent à n^k quand $n \to +\infty$, donc $\displaystyle\lim_{n\to+\infty} p_n^k \prod_{j=0}^{k-1}(n-j) = \lambda^k$.

D'autre part : $(n-k)\ln(1-p_n) \sim -np_n \to -\lambda$, donc $\displaystyle\lim_{n\to+\infty} (1-p_n)^{n-k} = \mathrm{e}^{-\lambda}$.

Finalement, on trouve $\displaystyle\lim_{n\to+\infty} \mathbb{P}(\mathrm{X}_n = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$. □

Proposition 11.6.5 (somme de deux lois de Poisson indépendantes)
Soit X, Y *deux variables de Poisson indépendantes, avec* $\mathrm{X} \hookrightarrow \mathcal{P}(\lambda)$ *et* $\mathrm{Y} \hookrightarrow \mathcal{P}(\mu)$.
Alors $\mathrm{X} + \mathrm{Y}$ *suit la loi de Poisson de paramètre* $\lambda + \mu$.

Les variables X et Y sont indépendantes donc t^{X} et t^{Y} sont indépendantes.

D'après la proposition 11.5.4 : $\forall t \in \mathbb{R},\ \mathrm{G_{X+Y}}(t) = \mathrm{G_X}(t)\, \mathrm{G_Y}(t) = \mathrm{e}^{\lambda(t-1)}\mathrm{e}^{\mu(t-1)} = \mathrm{e}^{(\lambda+\mu)(t-1)}$.

Il en découle que $\mathrm{X} + \mathrm{Y} \hookrightarrow \mathcal{P}(\lambda + \mu)$ (on a utilisé ici la proposition 11.5.1). □

Exercice 11.33 (⇝ **corrigé**)
Soient λ, μ dans $\mathbb{R}^{+*}$, et les variables aléatoires $X \rightsquigarrow \mathcal{P}(\lambda)$ et $Y \rightsquigarrow \mathcal{P}(\mu)$, supposées indépendantes.
1. On note $Z = X + Y$. Déterminer la loi de Z.
2. En déduire la loi de X sachant $(Z = n)$.

Exercice 11.34 (⇝ **corrigé**)
Soit X une variable aléatoire aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$.
Si la valeur prise par X est paire, on pose $Y = X/2$, et $Y = 0$ sinon.
Donner la loi de Y, son espérance et sa variance.

11.7 Loi faible des grands nombres

Dans les deux propositions suivantes, X désigne une variable aléatoire réelle sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Proposition 11.7.1 (inégalité de Markov)
On suppose que X *est à valeurs dans* $\mathbb{R}^+$, *et d'espérance finie. Alors :* $\forall\, a > 0,\ \mathbb{P}(\mathrm{X} \geqslant a) \leqslant \dfrac{\mathrm{E}(\mathrm{X})}{a}$.

La variable $\mathrm{Y} = \min(\mathrm{X}, a)$ est d'espérance finie, et on a : $0 \leqslant \mathrm{Y} \leqslant \mathrm{X}$ donc $0 \leqslant \mathrm{E}(\mathrm{Y}) \leqslant \mathrm{E}(\mathrm{X})$.
D'autre part : $\mathrm{E}(\mathrm{Y}) = \displaystyle\sum_{y\in \mathrm{Y}(\Omega)} \mathbb{P}(\mathrm{Y}=y)\, y = \mathbb{P}(\mathrm{Y}=a)\, a + \sum_{y\in \mathrm{Y}(\Omega),\, y\neq a} \mathbb{P}(\mathrm{Y}=y)\, y \geqslant \mathbb{P}(\mathrm{Y}=a)\, a$.
Mais $(\mathrm{Y}=a) = (\mathrm{X} \geqslant a)$ et $\mathrm{E}(\mathrm{Y}) \leqslant \mathrm{E}(\mathrm{X})$. On a donc obtenu $\mathrm{E}(\mathrm{X}) \geqslant \mathbb{P}(\mathrm{X}\geqslant a)\, a$. □

Exercice 11.35 (⇝ **corrigé**)
Soit X une variable aléatoire discrète positive qui admet un moment d'ordre 2 non nul.
Montrer que, pour tout $\lambda \in]0, 1[$, $\mathbb{P}(X \geq \lambda E(X)) \geq (1 - \lambda)^2 \dfrac{E(X)^2}{E(X^2)}$.

Proposition 11.7.2 (inégalité de Bienaymé-Tchebychev)
On suppose que X^2 *(donc* X*) est d'espérance finie.*
Pour tout réel $\varepsilon > 0$, *on a l'inégalité :* $\mathbb{P}\Big(|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geqslant \varepsilon\Big) \leqslant \dfrac{\mathrm{V}(\mathrm{X})}{\varepsilon^2}$.

C'est la même démonstration que dans le cas fini.
On applique l'inégalité de Markov à la variable positive $\mathrm{Y} = (\mathrm{X} - \mathrm{E}(\mathrm{X}))^2$, dont l'espérance est $\mathrm{V}(\mathrm{X})$.
Pour tout $\varepsilon > 0$, on a : $\mathbb{P}(\mathrm{Y} \geqslant \varepsilon^2) \leqslant \dfrac{\mathrm{V}(\mathrm{X})}{\varepsilon^2}$, c'est-à-dire : $\mathbb{P}(|\mathrm{X} - \mathrm{E}(\mathrm{X})| \geqslant \varepsilon) \leqslant \dfrac{\mathrm{V}(\mathrm{X})}{\varepsilon^2}$. □

Exercice 11.36 (⇝ **corrigé**)
1. Soit $\alpha > 0$. Montrer l'existence d'une variable aléatoire X telle que $X(\Omega) = \mathbb{N}$ et $G_X(t) = \dfrac{1}{(2 - t)^\alpha}$.
2. Donner un équivalent de $\mathbb{P}(X = n)$ quand $n \to +\infty$ dans le cas où $\alpha \in \mathbb{N}^*$.
3. Soit $\lambda > 0$. Montrer que $\mathbb{P}(X \geq \lambda + \alpha) \leq \dfrac{2\alpha}{\lambda^2}$.

Exercice 11.37 (⇝ **corrigé**)

Soit (X_n) une suite de de variables aléatoires.

On dit qu'elle converge vers X en probabilité si : $\forall\, \varepsilon > 0,\ \lim\limits_{n\to+\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$.

1. On suppose les X_n indépendantes, de même espérance, de même variance. Soit $M_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$.

   Montrer qu'il existe une constante λ telle que (M_n) converge en probabilité vers λ.
2. On suppose que $E(X_n) \to E(X)$ et que $V(X_n - X) \to 0$.
   Montrer que (X_n) converge en probabilité vers X.
3. Soit (S_n) et (X_n) deux suites de v.a.r. telles que $S_n \rightsquigarrow \mathcal{B}(n,p)$ et $X_n = \exp(S_n/n)$.

   Calculer l'espérance et la variance de X_n.

   Existe-t-il une variable X telle que (X_n) converge en probabilité vers X ?

Proposition 11.7.3 (loi faible des grands nombres)

Soit $(\mathrm{X}_n)_{n\geqslant 1}$ *une suite de variables aléatoires réelles, deux à deux indépendantes, de même loi.*

Soit m *leur espérance, et* σ *leur écart-type. Pour tout* n *de* $\mathbb{N}^*$*, on pose* $\mathrm{S}_n = \displaystyle\sum_{k=1}^{n} \mathrm{X}_k$.

La variable $\dfrac{\mathrm{S}_n}{n}$ *(moyenne arithmétique de* $\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n$*) est d'espérance* m *et d'écart-type* $\dfrac{\sigma}{\sqrt{n}}$.

Pour tout $\varepsilon > 0$*, on a :* $P\left(\left|\dfrac{1}{n}S_n - m\right| \geqslant \varepsilon\right) \leqslant \dfrac{\sigma^2}{n\,\varepsilon^2}$ *, donc :* $\lim\limits_{n\to+\infty} P\left(\left|\dfrac{1}{n}S_n - m\right| \geqslant \varepsilon\right) = 0$

Par linéarité de l'espérance : $\mathrm{E}\Big(\dfrac{\mathrm{S}_n}{n}\Big) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} \mathrm{E}(\mathrm{X}_k) = m$.

Les X_k étant deux à deux indépendantes, on a $\mathrm{V}\Big(\dfrac{\mathrm{S}_n}{n}\Big) = \dfrac{1}{n^2}\displaystyle\sum_{k=1}^{n} \mathrm{V}(\mathrm{X}_k) = \dfrac{\sigma^2}{n}$, donc $\sigma\Big(\dfrac{\mathrm{S}_n}{n}\Big) = \dfrac{\sigma}{\sqrt{n}}$.

L'inégalité de Bienaymé-Tchebychev, appliquée à $\dfrac{\mathrm{S}_n}{n}$, s'écrit donc : $\mathbb{P}\left(\left|\dfrac{\mathrm{S}_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \dfrac{\sigma^2}{n\,\varepsilon^2}$. □