Outils de calculs, probabilités et statistique

UFR des Sciences, Licence 1 info

1 Ensembles et éléments

1.1 Notions fondamentales

Il n'y a pas de définition mathématique précise de ce qu'est un ensemble car il s'agit d'une notion première. On dira qu'un ensemble est une collection d'objets appelés éléments.

On se donne un ensemble Ω, appelé **ensemble universel**, qui contient tous les éléments considérés. L'ensemble ne contenant aucun élément s'appelle l'**ensemble vide** et est noté $\emptyset$. Un ensemble est dit **fini** s'il contient un nombre fini d'objets ; il se dit **infini** s'il en contient une infinité. On appelle **cardinal** d'un ensemble fini le nombre d'éléments qu'il contient. Le cardinal d'un ensemble fini A s'écrit $\text{Card}(A)$ ou $\#A$.

Si a est un élément de l'ensemble A, on dit que a appartient à A et on note $a \in A$. Sinon, on note $a \notin A$.

Pour deux ensembles A et B, on dit que A est un **sous-ensemble** de B (ou que A est inclus dans B) si et seulement si tous les éléments de A appartiennent à B. Cela se note $A \subset B$. Remarquons que nous avons l'équivalence suivante : $A = B \quad \Leftrightarrow \quad A \subset B$ et $B \subset A$.

Travaillant dans l'ensemble universel Ω, pour tout ensemble A, on a en particulier $A \subset \Omega$. L'ensemble des sous-ensembles d'un ensemble A s'appelle la **famille des parties de** A et se note $\mathcal{P}(A)$.

Exemples :

- L'ensemble $\mathbb{R}$ des nombres réels est infini.
- L'ensemble $\mathbb{N}$ des entiers naturels est infini.

— Mars est un élément de l'ensemble fini des mos de l'année, qui est de cardinal 12.
— L'ensemble des jours du week-end {samedi, dimanche} est inclus dans l'ensemble {lundi, mardi, mercredi, jeudi, vendredi, samedi, dimanche}.
— L'ensemble des nombres impairs est un sous-ensemble infini de l'ensemble des entiers naturels.
— Si $A = \{1, 2, 3\}$, alors $\mathcal{P}(A) = \big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\big\}$.

1.2 Opérations sur les ensembles

Pour un ensemble universel Ω et deux ensembles quelconques A et B, nous définissons les opérations suivantes. Sur les illustrations, les résultats de ces opérations sont représentés en bleu.

— **Réunion** : $A \cup B$ est l'ensemble $A \cup B = \{x \in \Omega | x \in A \textit{ ou } x \in B\}$ des éléments appartenant soit à A, soit à B, (et éventuellement aux deux).
Remarque : $A \cup B = A \quad \Leftrightarrow \quad B \subset A$

— **Intersection** : $A \cap B$ est l'ensemble $A \cap B = \{x \in \Omega | x \in A \textit{ et } x \in B\}$ des éléments appartenant à A et à B. Si $A \cap B = \emptyset$, on dit que A et B sont **disjoints**.
Remarque : $A \cap B = A \quad \Leftrightarrow \quad A \subset B$

— **Complémentaire** : le complémentaire de A, noté $\overline{A}$, est l'ensemble $\overline{A} = \{x \in \Omega | x \notin A\}$ des éléments (de Ω) n'appartenant pas à A.
— **Différence** : La différence entre A et B, notée $A - B$ ou $A \backslash B$, est l'ensemble $A - B = \{x \in \Omega | x \in A \textit{ et } x \notin B\}$ des éléments appartenant à A mais pas à B.
Remarque : on a donc $A - B = A \cap \overline{B}$.
— **Produit** : Le produit cartésien de A et B est l'ensemble des couples dont le premier élément appartient à A et le second à B : $A \times B = \{(a, b) | a \in A \textit{ et } b \in B\}$.
Remarque : *il est à noter que pour un couple, l'ordre compte :* $(1, 2) \neq (2, 1)$. *Ainsi, si* $A \neq B$, $A \times B \neq B \times A$.

PROPRIÉTÉS :

- Commutativité :
  $A \cup B = B \cup A$
  $A \cap B = B \cap A$
- Associativité :
  $(A \cup B) \cup C = A \cup (B \cup C)$
  $(A \cap B) \cap C = A \cap (B \cap C)$
- Distributivité :
  $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$
  $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
- Autres relations fondamentales :
  $\overline{A \cup B} = \overline{A} \cap \overline{B}$
  $\overline{A \cap B} = \overline{A} \cup \overline{B}$
  $A \cap \overline{A} = \emptyset$
  $A \cup \overline{A} = \Omega$

Une **partition** d'un ensemble A est une famille de sous-ensembles $\{A_i\}$ non vides de A, deux à deux disjoints et dont l'union est égale à A.

Pour une partition de A en n sous-ensembles $\{A_1, \ldots, A_n\}$, on a donc :

- $\forall i \in \{1, \ldots, n\}$, $A_i \neq \emptyset$.
- $\forall i, j \in \{1, \ldots, n\}$ tel que $i \neq j$ on a : $A_i \cap A_j = \emptyset$.
- $A_1 \cup A_2 \cup \cdots \cup A_n = A$.

EXEMPLE : Soit $A = \{1, 5, 7, 11, 23\}$, $A_1 = \{1\}$, $A_2 = \{5, 23\}$, $A_3 = \{7, 11\}$. Alors $\{A_1, A_2, A_3\}$ est une partition de A.

2 Fonctions

Soient A, B deux ensembles. Une fonction de A vers B est une loi qui associe à **chaque** élément de A un et **un seul** élément de B.

EXEMPLES :

- $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = 1/x$, n'est pas une fonction : elle n'est pas définie en 0.
- $f\colon \mathbb{R}^* \to \mathbb{R}$, $f(x) = 1/x$, est pas une fonction.
- $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = \begin{cases} x+1, & \text{si } x \geq 0 \\ x-1, & \text{si } x \leq 0 \end{cases}$, n'est pas une fonction : elle n'est pas définie en 0.

Une fonction $f\colon A \to B$ est **injective** si $(\forall a_1, a_2 \in A,\ f(a_1) = f(a_2) \Rightarrow a_1 = a_2)$.

Une fonction $f\colon A \to B$ est **surjective** si $(\forall b \in B,\ \exists a \in A \text{ t.q. } f(a) = b)$.

Une fonction $f\colon A \to B$ est **bijective** si elle est à la fois injective et bijective.

EXEMPLES :

— La fonction $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, n'est ni injective ni surjective.

— La fonction $f\colon \mathbb{R} \to \mathbb{R}_+$, $f(x) = x^2$, est surjective mais pas injective.

— La fonction $f\colon \mathbb{R}_+ \to \mathbb{R}_+$, $f(x) = x^2$, est bijective.

Soient $f\colon A \to B$ et $g\colon B \to C$. On définit la **fonction composée** $g \circ f\colon A \to C$ par : $\forall a \in A$, $(g \circ f)(a) = g(f(a))$.

EXEMPLE : Soit $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$ et $g\colon \mathbb{R} \to \mathbb{R}$, $g(x) = x^3$. Alors $g \circ f\colon \mathbb{R} \to \mathbb{R}$ est définie par : $(g \circ f)(x) = (x^2)^3 = x^6$.

Soit $f\colon A \to B$ une fonction et $b \in B$. On appelle **antécédente** de b tout élément de $a \in A$ tel que $f(a) = b$.

Remarque : *f est injective si et seulement si chaque $b \in B$ possède* **au plus** *un antécédente. f est surjective si et seulement si chaque $b \in B$ possède* **au moins** *un antécédente.*

Soit $X \subset A$. On pose $f(X) = \{f(x) \mid x \in X\}$. On appelle $f(A)$ l'**image** de f.

Soit $Y \subset B$. On pose $f^{-1}(Y) = \{x \in A \mid f(x) \in Y\}$.

Remarque : *la fonction f est surjective si et seulement si son image est B. Pour $b \in B$, $f^{-1}(\{b\})$ est l'ensemble des antécédentes de b.*

Soit $f\colon A \to B$ une fonction bijective. On note $f^{-1}\colon B \to A$ sa **fonction inverse**, définie par $f^{-1}(b) = a \Leftrightarrow f(a) = b$.

3 Dénombrement

3.1 Sommes, produits, suites arithmétiques et géométriques

Soit I un ensemble fini, et $(x_i)_{i \in I}$ une famille de réels indexée sur I. La somme des $\mathrm{Card}(I)$ termes de la famille se note $\sum_{i \in I} x_i$ alors que le produit de ces termes s'écrit $\prod_{i \in I} x_i$. Dans le cas où I est un ensemble d'entiers naturels successifs $\{m, \ldots, n\}$, avec $m \leq n$, la somme s'écrit $\sum_{i=m}^{n} x_i$ et le produit $\prod_{i=m}^{n} x_i$.

Définition 3.1. *Une suite $(u_n)_{n \in \mathbb{N}}$ est dite arithmétique de raison r si pour tout $n \in \mathbb{N}$, $u_{n+1} - u_n = r$.*
Une suite arithmétique est totalement définie par la donnée de son terme initial u_0 et

de sa raison r. Le terme général d'une suite arithmétique de raison r est donné par $u_n = n \cdot r + u_0$.

Proposition 3.1. *Soient $(u_n)_{n\in\mathbb{N}}$ une suite arithmétique de raison r, et m, n entiers avec $m \leq n$. On a alors*

$$\sum_{i=m}^{n} u_i = \frac{(n-m+1)(u_m+u_n)}{2} \qquad \left(= \frac{(n-m+1)(2u_0+(n+m)r)}{2}\right).$$

En particulier, on a : $\sum\limits_{i=1}^{n} i = \frac{n(n+1)}{2}$.

Définition 3.2. *Une suite $(u_n)_{n\in\mathbb{N}}$ est dite géométrique de raison r si pour tout $n \in \mathbb{N}$, $u_{n+1} = r \cdot u_n$.*
Une suite géométrique est totalement définie par la donnée de son terme initial u_0 et de sa raison r. Le terme général d'une suite géométrique de raison r est donné par $u_n = u_0 \cdot r^n$.

Proposition 3.2. *Soient $(u_n)_{n\in\mathbb{N}}$ une suite géométrique de raison $r \neq 1$, et $m, n \in \mathbb{N}$ avec $m \leq n$. On a alors*

$$\sum_{i=m}^{n} u_i = \frac{u_{n+1}-u_m}{r-1} \qquad \left(= u_0 \frac{r^{n+1}-r^m}{r-1}\right).$$

3.2 Coefficients binomiaux et dénombrement

Définition 3.3. *Pour tout entier naturel n strictement positif, on appelle « factorielle n» et on note $n!$ la quantité $\prod_{i=1}^{n} i$. On étend cette définition à l'ensemble des entiers en posant, par convention, $0! = 1$.*

Définition 3.4. *Soit $k \in \mathbb{N}^*$. Une collection ordonnée de k éléments est appelée un k-uplet.*

Proposition 3.3. *Soit E un ensemble non vide de n éléments et k un entier naturel inférieur ou égal à n.*

1) *Le nombre de k-uplets d'éléments de E («k-arrangements avec répétition») est égale au nombre*

$$n^k$$

de fonctions $\{1, \ldots, k\} \to \{1, \ldots, n\}$.

2) Le nombre de k-uplets d'éléments **distincts** *de E («k-arrangements») est égale au nombre*
$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$
de fonctions injectives $\{1,\ldots,k\} \to \{1,\ldots,n\}$. On utilise parfois la notation $A_n^k = \frac{n!}{(n-k)!}$.

3) Le nombre de sous-ensemble de E à k-éléments est («k-combinaison») est
$$\binom{n}{k} = \frac{n!}{(n-k)!k!}.$$
On utilise parfois la notation C_n^k à la place de $\binom{n}{k}$.

Un n-uplet d'éléments distincts d'un ensemble E à n éléments (n-arrangements de E) est appelé une **permutation** de E.

Proposition 3.4. *Pour tout k, $n \in \mathbb{N}$ avec $k \leq n$ on a $\binom{n}{k} = \binom{n}{n-k}$. Si de plus $k < n$ alors*
$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}.$$

Proposition 3.5 (Formule du binôme)**.** *Soit $n \in \mathbb{N}$. Pour tous réels a et b on a :*
$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Proposition 3.6. *Soit A est un ensemble fini, de cardinal n. Alors $\mathcal{P}(A)$ est fini et son cardinal est 2^n.*

4 Probabilités : fondements

4.1 Définitions

On appelle **expérience aléatoire** ou **épreuve** une expérience dont le résultat est incertain. Chacun des résultats de cette épreuve est un **événement élémentaire**. L'ensemble Ω des résultats possibles est appelé **ensemble fondamental** ou **univers** : il correspond dans le cadre probabiliste à l'ensemble universel évoqué plus haut. Nous supposerons dans ce cours Ω *fini*.

Un **événement** est un sous ensemble de Ω : il s'agit donc d'un ensemble d'événements élémentaires. L'ensemble vide $\emptyset$ est l'événement impossible alors que Ω est l'événement certain. Soit A un événement ; on note $\overline{A}$ l'événement contraire. Deux événements sont **incompatibles** si leur intersection est vide. Finalement, un **système complet d'événements** est une partition de Ω (*i.e.* un ensemble d'événements pas impossibles, deux à deux incompatibles et dont l'union est égale à Ω).

Définition 4.1. *Une probabilité P est une fonction qui à tout événement associe un nombre compris entre 0 et 1, avec $P(\Omega) = 1$, et qui pour tout ensemble d'événements incompatibles $A_1, \dots, A_n$ vérifie :*

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n).$$

Cela revient à dire que pour des événements incompatibles, la probabilité de l'union est égale à la somme des probabilités.

De cette définition, nous pouvons déduire, pour des événements A et B quelconques, les relations suivantes :

— $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
— $P(\overline{A}) = 1 - P(A)$.
— $P(A - B) = P(A) - P(A \cap B)$.
— Si $(A_1, \dots, A_n)$ est un système complet d'événements, alors $P(A) = P(A \cap A_1) + \dots + P(A \cap A_n)$.
— Si $A \subset B$ alors $P(A) \leq P(B)$.
— $P(\emptyset) = 0$.

Pour un ensemble $\Omega = \{\omega_1, \dots, \omega_n\}$, il suffit de se donner n nombres positifs $(p_1, \dots, p_n)$ vérifiant $\sum_{i=1}^{n} p_i = 1$ pour définir une probabilité en posant, pour tout i, $P(\omega_i) = p_i$. Le choix des p_i est donc ici arbitraire pour définir une probabilité. En pratique, ce sont les conditions de l'expérience qui déterminent ce choix.

Un ensemble fini est dit **équiprobable** si tous les événements élémentaires ont la même probabilité. Dans ce cas, la probabilité d'un événement est égale au rapport de son cardinal et de celui de l'ensemble fondamental :

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

Deux événements quelconques A et B sont dits **indépendants** si et seulement si $P(A \cap B) = P(A) \times P(B)$ et on note alors $A \perp\!\!\!\perp B$.

Remarque : *On peut montrer que :*$A \perp\!\!\!\perp B \Leftrightarrow \overline{A} \perp\!\!\!\perp \overline{B} \Leftrightarrow A \perp\!\!\!\perp \overline{B} \Leftrightarrow \overline{A} \perp\!\!\!\perp B$.

On veillera à ne pas confondre indépendance et incompatibilité. Ainsi, si deux événements sont indépendants, le fait de savoir que l'un s'est réalisé ne nous apporte aucune information sur le fait que l'autre se soit réalisé ou non. En revanche, si A et B sont incompatibles, ils ne peuvent pas se réaliser simultanément : la réalisation de A nous apporte donc une information très précise sur celle de B, B ne pouvant alors pas se réaliser !

EXEMPLES :

- On considère une famille de deux enfants et on suppose que le sexe de l'aîné n'a aucune incidence sur celui du second : chaque enfant a une chance sur deux d'être un garçon et une chance sur deux d'être une fille. Définissons trois événements : $A = \{\text{l'aîné est un garçon}\}$, $B = \{\text{l'aîné est une fille}\}$ et $C = \{\text{le second est une fille}\}$. A et B sont incompatibles alors que A et C sont indépendants.
- On s'intéresse au lancer d'un dé. L'univers est $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $P(i) = \frac{1}{6}$ pour tout $i \in \Omega$ Ω est un ensemble équiprobable et la probabilité de l'événement $\{\text{le résultat est pair}\}$ est bien égale au nombre de cas favorables (3, correspondant au cardinal de l'ensemble $\{2; 4; 6\}$) sur le nombre de cas possibles, $\mathrm{Card}(\Omega) = 6$. Si nous jetons 2 dés et que nous nous intéressons au couple formé par les deux dés, nous avons

$$\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$$

et chacun des 36 résultats possibles a une chance sur 36 de sortir. Cependant, si nous lançons 2 dés et que nous considérons que chaque résultat est la somme des nombres de chaque dé, nous avons alors

$$\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

et les événements élémentaires ne sont plus équiprobables. Effectivement, nous aurons alors

$$\begin{aligned} p_2 = p_{12} = \frac{1}{36}, \qquad & p_3 = p_{11} = \frac{1}{18}, \\ p_4 = p_{10} = \frac{1}{12}, \qquad & p_5 = p_9 = \frac{1}{9}, \\ p_6 = p_8 = \frac{5}{36} \qquad & \text{et } p_7 = \frac{1}{6}. \end{aligned}$$

— Si on lance une monnaie équilibrée, l'espace des résultats possibles est $\Omega = \{pile, face\}$ et on aura $P(pile) = P(face) = 1/2$. L'ensemble est équiprobable. On lance maintenant deux fois une pièce de monnaie équilibrée. L'espace des résultats possibles est alors

$$\Omega = \{(pile, pile), (pile, face), (face, pile), (face, face)\}.$$

La probabilité d'un des événement élémentaire (un des quatre élément de Ω) est $1/4$. L'ensemble est équiprobable. Considérons l'événement «on obtient au moins une fois pile». Cet événement correspond au sous-ensemble $A = \{(pile, pile), (pile, face), (face, pile)\}$ et donc sa probabilité vaut

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{3}{4}.$$

Considérons les événements «on obtient pile au premier lancer» et «on obtient pile au deuxième lancer» qui correspondent aux sous ensembles $A_1 = \{(pile, pile), (pile, face)\}$ et $A_2 = \{(pile, pile), (face, pile)\}$ respectivement. On a alors $A = A_1 \cup A_2$ et

$$P(A) = P(A_1 \cup A_2) = 1 - P(\overline{A_1 \cup A_2}) = 1 - P(\overline{A_1} \cap \overline{A_2}).$$

Les événements $\overline{A_1}$ et $\overline{A_2}$ (qui correspondent à «on obtient face au premier lancer» et à «on obtient face au deuxième lancer») sont indépendants et donc

$$P(\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1})P(\overline{A_2}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

On retrouve $P(A) = 1 - \frac{1}{4} = \frac{3}{4}$.

— Plus généralement, on lance n fois une pièce de monnaie équilibrée :

$$\Omega = \{pile, face\}^n.$$

L'événement A «on obtient au moins une fois pile» est la réunion des événements A_1, ..., A_n, où A_j correspond à «on obtient pile au n-ième lancer». On a alors

$$P(A) = P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - P\left(\overline{\bigcup_{i=1}^{n} A_i}\right) = 1 - P\left(\bigcap_{i=1}^{n} \overline{A_i}\right).$$

L'événement $\bigcap_{i=1}^{n} \overline{A_i} = \{(face, face, \ldots, face\}$ («on obtient face à chaque lancer») est un événement élémentaire, et donc sa probabilité est $\frac{1}{\text{Card}(\Omega)} = \frac{1}{2^n}$. Au total

$$P(A) = 1 - \frac{1}{2^n}.$$

— On joue à la roulette. On considère alors l'ensemble équiprobable

$$\Omega = \{0, 1, 2, \ldots, 36\}.$$

L'événement «il sort un nombre pair» correspond à l'ensemble

$$A = \{0, 2, 4, \ldots, 34, 36\}$$

et sa probabilité est $P(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \frac{19}{37}$. L'événement «il sort un nombre impair > 13» correspond à l'ensemble $B = \{15, 17, \ldots, 33, 35\}$ et sa probabilité est $P(B) = \frac{\mathrm{Card}(B)}{\mathrm{Card}(\Omega)} = \frac{11}{37}$. La probabilité qu'il sort un nombre ou pair ou impair > 13 est alors $\frac{19}{37} + \frac{11}{37} = \frac{30}{37}$, car A et B sont incompatibles.

4.2 Probabilités conditionnelles

On considère un tirage aléatoire uniforme de boules dans une urne qui contient 10 boules numérotés de 1 à 10. On a donc $\Omega = \{1, 2, \ldots, 10\}$. Notons $A = \{1, 2, 3, 4, 5\}$. La probabilité d'extraire une boule dont le numéro est ≤ 5 vaut

$$P(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} = \frac{5}{10} = 0.5.$$

Supposons maintenant que quelqu'un qui a assisté à l'extraction nous dise que la boule extraite porte un numéro paire. Dans ce cas, l'ensemble des résultats possibles n'est plus Ω mais $B = \{2, 4, 6, 8, 10\}$ et la probabilité que la boule extraite ait un numéro ≤ 5 devient

$$\frac{\mathrm{Card}(A \cap B)}{\mathrm{Card}(B)} = \frac{2}{5} = 0.4.$$

Le fait de savoir que l'événement $\{$la boule extraite porte un numéro paire$\}$ est réalisé modifie la probabilité de l'événement $\{$la boule extraite porte un numéro $\leq 5\}$, tout se passant alors comme si nous avions tiré au hasard une boule dans une urne qui contient 5 boules numérotés 2,4,6,8 et 10 : c'est la notion de probabilité conditionnelle qui émerge ici.

Avant de la presenter formellement, faisons un deuxième exemple. Si nous tirons une personne au hasard dans la population française, la probabilité que celle-ci ait le cancer du poumon est égale au nombre de personnes atteintes par ce cancer divisée par la population française. Si nous savons à présent que cette personne fume quotidiennement deux paquets de cigarettes depuis cinquante ans (on parlera ici de « gros fumeur de longue date »), la probabilité sera alors bien supérieure à celle initialement calculée : elle est égale au nombre de gros fumeurs de longue date ayant ce cancer sur le nombre de gros fumeurs de longue date. Le fait de savoir que l'événement $\{$la personne tirée est un gros fumeur de longue date$\}$ est réalisé modifie la probabilité de l'événement {la personne a le cancer du poumon}, tout se passant alors comme si nous avions tiré au hasard une personne parmi les gros fumeurs de longue date, population où l'incidence du cancer du poumon est évidemment supérieure à celle observée dans la population générale.

Soient 2 événements A et B. Si $P(B) \neq 0$, nous définissons la probabilité conditionnelle de A sachant B par :

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Remarque : *L'application P_B qui à tout événement C associe la valeur $P_B(C) = P(C|B)$ est une probabilité : elle est donc à valeurs dans $[0, 1]$ et respecte les règles de calculs énoncée plus haut, vérifiant en particulier, pour tous événements C et D incompatibles, $P_B(C \cup D) = P_B(C) + P_B(D)$.*

Remarque : *Rappelons que deux événement sont indépendants si et seulement si $P(A \cap B) = P(A)P(B)$. Donc, si A et B sont deux événement et $P(B) \neq 0$,*

$$A \text{ et } B \text{ sont indépendants si et seulement si } P(A|B) = P(A).$$

Nous déduisons immédiatement de la définition de probabilité conditionnelle la relation :

Proposition 4.1 (Formule des probabilités totales)**.** *Supposons maintenant A et B tous les deux de probabilité non nulle. On a alors :*

$$P(A|B)P(B) = P(A \cap B) = P(B \cap A) = P(B|A)P(A). \tag{1}$$

Remarque : *Un raisonnement par récurrence nous conduit à la généralisation pour n événements de probabilité non nulle $A_1, \ldots, A_n$ de la relation (1) :*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Théorème 4.1 (Formule de Bayes)**.** *Soient A un événement de probabilité non nulle et $(B_1, \ldots, B_n)$ un système complet d'événements. Alors, pour $i = 1, \ldots, n$,*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + \cdots + P(A|B_n)P(B_n)}.$$

Démonstration. Soit $i \in [1, \ldots, n]$. La relation (1) (avec B remplacé par B_i) nous conduit à :

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}.$$

On remplace ensuite le dénominateur par :

$$P(A) = P(A \cap B_1) + \cdots + P(A \cap B_n) = P(A|B_1)P(B_1) + \cdots + P(A|B_n)P(B_n).$$

□

Exemple : Un signale provient 2 fois sur 5 d'une station émettrice A et 3 fois sur 5 d'une station B. Il peut être long ou court. On sait que A transmet un signale long 2 fois sur 3 et que B transmet un signale long 1 fois sur 4. On reçoit un signale court. Quelle est la probabilité que ce signale soit émis par la station A ?

On considère les événements $A = \{$le signale provient de la station A$\}$, $B = \{$le signale provient de la station B$\}$, $L = \{$le signale est long$\}$, $C = \{$le signale est court$\}$. On a donc

$$P(A) = \frac{2}{5}, \quad P(B) = \frac{3}{5}, \qquad P(L|A) = \frac{2}{3}, \quad P(L|B) = \frac{1}{4}.$$

Mais aussi :

$$P(C|A) = 1 - P(L|A) = \frac{1}{3}, \quad P(C|B) = 1 - P(L|B) = \frac{3}{4}.$$

On nous demande de calculer la probabilité conditionnelle $P(A|C)$. On applique la formule de Bayes :

$$P(A|C) = \frac{P(C|A)P(A)}{P(C|A)P(A) + P(C|B)P(B)} = \frac{\frac{1}{3}\times\frac{2}{5}}{\frac{1}{3}\times\frac{2}{5}+\frac{3}{4}\times\frac{3}{5}} = \frac{\frac{2}{15}}{\frac{2}{15}+\frac{9}{20}} = \frac{2}{15}\times\frac{60}{35} = \frac{8}{35}.$$

Si l'on se rappelle pas de la formule, rien n'empêche de refaire tous les calculs ! Par définition, $P(A|C) = \frac{P(A\cap C)}{P(C)}$. Pour calculer le numérateur, on utilise

$$P(A\cap C) = P(C\cap A) = P(C|A)P(A) = \frac{1}{3}\times\frac{2}{5} = \frac{2}{15}.$$

Et pour calculer $P(C)$, on remarque d'abord que $P(C) = P(C\cap A) + P(C\cap B)$. Nous avons déjà calculé $P(C\cap A) = \frac{2}{15}$. Pour calculer $P(C\cap B)$ on utilise le même procédé :

$$P(C\cap B) = P(C|B)P(B) = \frac{3}{4}\times\frac{3}{5} = \frac{9}{20}.$$

Donc

$$P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{\frac{2}{15}}{\frac{2}{15}+\frac{9}{20}} = \frac{8}{35}.$$

Exemple : Dans la salle des profs 60% sont des femmes ; une femme sur trois porte des lunettes et un homme sur deux porte des lunettes : quelle est la probabilité pour qu'un porteur de lunettes pris au hasard soit une femme ?

Notons les différents événements : F : «être femme», L : «porter des lunettes», H : «être homme». Alors on a $P(F) = 0.6$ et $P(H) = 1 - P(F) = 0.4$. On a aussi $P(L|F) = \frac{1}{3}$: il s'agit de la probabilité conditionnelle de «porter des lunettes» sachant que la personne est une femme. De même, on a $P(L|H) = 0.5$. On cherche la probabilité conditionnelle $P(F|L)$. D'après la formule des probabilités totales on a :

$$P(F|L)P(L) = P(L|F)P(F).$$

On a aussi :

$$P(L) = P(L|F)P(F) + P(L|H)P(H)$$

car $P(L|F)P(F) = P(L \cap F)$, $P(L|H)P(H) = P(L \cap H)$ et L est la réunion des événements incompatibles $L \cap F$ et $L \cap H$. Donc

$$P(F|L) = \frac{P(L|F)P(F)}{P(L)} = \frac{P(L|F)P(F)}{P(L|F)P(F) + P(L|H)P(H)}.$$

(C'est la formule de Bayes !) On calcule : $P(L|F)P(F) = \frac{1}{3} \times 0.6 = 0.2$, $P(L|H)P(H) = 0.5 \times 0.4 = 0.2$ et

$$P(F|L) = \frac{P(L|F)P(F)}{P(L|F)P(F) + P(L|H)P(H)} = \frac{0.2}{0.2 + 0.2} = 0.5.$$

5 Variables aléatoires.

Un programme informatique produit une séquence aléatoire de 4 bits. Par exemple, il peut produire la séquence $\omega_1 = (1, 0, 0, 1)$ ou la séquence $\omega_2 = (1, 1, 0, 1)$. On suppose que l'ensemble fondamental Ω (l'ensemble des séquence de 4 symboles $\in \{0, 1\}$ de cardinal $2^4 = 16$) soit un ensemble équiprobable. La probabilité de chaque événement élémentaire (le fait que le programme produise une certaine séquence) est donc $1/\mathrm{Card}(\Omega) = 1/16$.

Supposons maintenant que le temps utilisé par le programme pour écrire «1» soit de 2 μsec (microsecondes) et que le temps employé pour écrire «0» soit de 3 μsec. Par exemple, le temps nécessaire pour écrire $\omega_1 = (1, 0, 0, 1)$ est de $2+3+3+2 = 10$ μsec, et celui nécéssaire pour écrire ω_2 est de $2 + 2 + 3 + 2 = 9$ μsec. On veut connaître la probabilité que le programme emploie un temps de 9 μsec pour produire une séquence. On s'aperçoit après calcul que les seules séquence qui nécessitent de 9 μsec sont les 4 séquences : $(0, 1, 1, 1)$, $(1, 0, 1, 1)$, $(1, 1, 0, 1)$, $(1, 1, 1, 0)$. La probabilité recherchée est donc $4/16 = 0.25$.

On peut formaliser cette situation en introduisant une *variable aléatoire* (réelle) X définie sur Ω qui associe à chaque séquence le temps nécessaire au programme pour la produire. Par exemple on a : $X(\omega_1) = 10$ et $X(\omega_2) = 9$. Nous sommes intéressés à la probabilité de l'événement $\{\omega \in \Omega \mid X(\omega) = 9\}$ que l'on écrit avec un peu d'abus de notations $\{X = 9\}$. En utilisant ce formalisme et les calculs précédents, on peut écrire

$$P(\{X = 9\}) = 0.25.$$

ou, à nouveau avec encore de l'abus de notations, $P(X = 9) = 0.25$.

De même, la probabilité que le programme emploie un temps > 10 μsec pour produire une séquence s'écrit simplement $P(X > 10)$. On constate que pour tout séquence $\omega \in \Omega$ on a $X(\omega) \leq 12$, qu'il y a un seul séquence ω tel que $X(\omega) = 12$ et 4 séquences ω avec $X(\omega) = 11$. Donc $\{X > 10\}$ est de cardinal $1 + 4 = 5$ et

$$P(X > 10) = \frac{5}{16} = 0.3125.$$

Définition 5.1. *Une variable aléatoire (en abrégé v.a.) réelle définie sur un ensemble fondamental Ω (fini) est une application de Ω dans $\mathbb{R}$: elle associe un nombre à chaque résultat possible.*

Soit X une v.a. Rappelons que (comme pour tous les fonctions) $X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$ est l'image de X. Il s'agit dans notre cas d'un ensemble fini (car l'univers Ω est fini) que l'on appelle *univers image*. On le note usuellement $X(\Omega) = \{x_1, \ldots, x_N\}$ (où les x_i sont deux-à-deux distincts).

Définition 5.2. *On appelle loi de la v.a. X la donnée de tous les $p_i = P(X = x_i)$ lorsque x_i prend toutes les valeurs possibles dans $X(\Omega)$.*

On présent souvent la loi d'une v.a. sous forme de tableau. En notant $x_1, \ldots, x_N$ les éléments de $X(\Omega)$ et $p_i = P(X = x_i)$, on écrira :

x_i	x_1	x_2	$\ldots$	x_N
$P(X = x_i)$	p_1	p_2	$\ldots$	p_N

pour donner la loi de X.

Remarquons que $\sum_{i=1}^{N} p_i = 1$ et donc la loi de la v.a. X définit une probabilité P' (dite *loi de probabilité de* X) sur l'univers $\Omega' = X(\Omega)$.

Définition 5.3. *La fonction de répartition (en abrégé f.d.r.) F d'une v.a. X est la fonction $F\colon \mathbb{R} \to [0, 1]$ définie par :*

$$F(t) = P(X \leq t) \quad \forall t \in \mathbb{R}.$$

Dans notre cas (univers fini) c'est une fonction en escalier.

Reprenons l'exemple du programme informatique qui produit des séquence aléatoires. Rappelons que X associe à chaque séquence le temps nécessaire au programme pour la produire, et que le programme nécessite de 2 μsec pour écrire «1» et de 3

μsec pour écrire «0». On a donc $X(0,0,0,0) = 12$,

ω	$X(\omega)$	ω	$X(\omega)$	ω	$X(\omega)$
$(0,0,0,1)$	11	$(0,0,1,1)$	10	$(0,1,1,1)$	9
$(0,0,1,0)$	11	$(0,1,0,1)$	10	$(1,0,1,1)$	9
$(0,1,0,0)$	11	$(0,1,1,0)$	10	$(1,1,0,1)$	9
$(1,0,0,0)$	11	$(1,0,0,1)$	10	$(1,1,1,0)$	9
		$(1,0,1,0)$	10		
		$(1,1,0,0)$	10		

et $X(1,1,1,1) = 8$, ce qui montre que $X(\Omega) = \{8, 9, 10, 11, 12\}$.

Donc la loi de X est

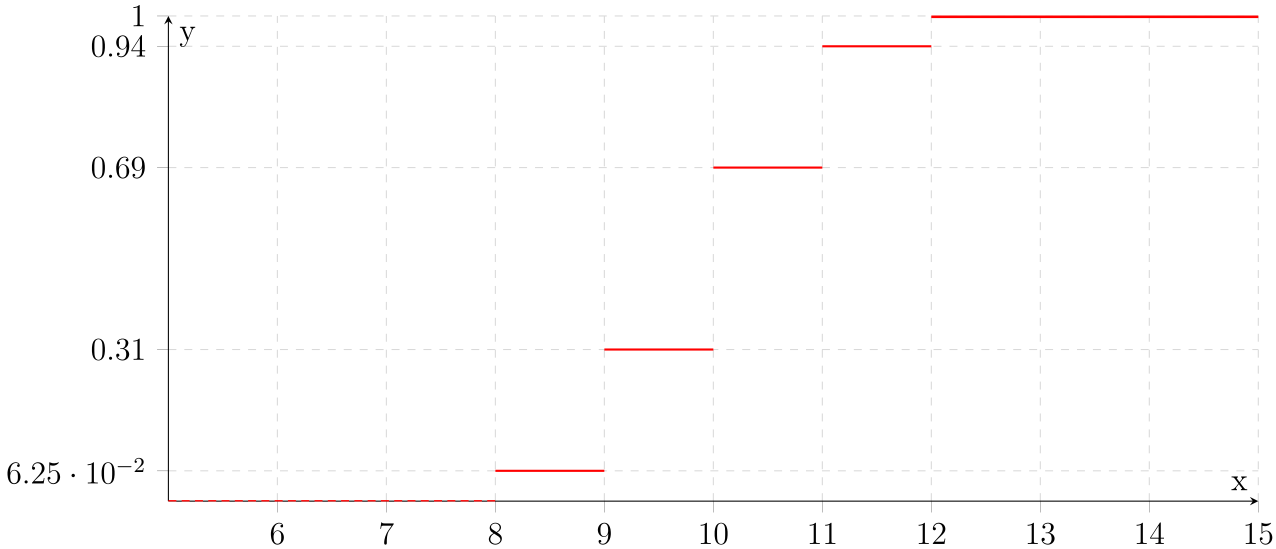
x_i	8	9	10	11	12
p_i	1/16	4/16	6/16	4/16	1/16

(ou $p_i = P(X = x_i)$). On vérifie que $\sum_{i=i}^{5} p_i = 1$.

Enfin, la function de répartition de X est définie par :

$$F(t) = P(X \leq t) = \begin{cases} 0, & \text{si } t < 8; \\ 1/16, & \text{si } 8 \leq t < 9; \\ 5/16, & \text{si } 9 \leq t < 10; \\ 11/16 & \text{si } 10 \leq t < 11; \\ 15/16, & \text{si } 11 \leq t < 12; \\ 1, & \text{si } t \geq 12. \end{cases}$$

Voici sa graphique :



Revenons maintenant au lancer de deux dés, quand on s'intéresse à la somme des nombres de chaque dé. On peut revoir cet example en introduisant une variable aléatoire X «somme des nombres de chaque dé» définie sur l'univers $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$. Les valeurs de $X(i, j)$ sont donnés par le tableau suivant :

$_i\backslash^j$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Ce tableau nous permets de calculer la loi de X :

x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

La variable X induit donc sur l'univers image

$$X(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

la probabilité définie par

$$p_2 = p_{12} = \tfrac{1}{36}, \qquad p_3 = p_{11} = \tfrac{1}{18}, \qquad p_4 = p_{10} = \tfrac{1}{12}$$

$$p_5 = p_9 = \tfrac{1}{9}, \qquad p_6 = p_8 = \tfrac{5}{36}, \qquad p_7 = \tfrac{1}{6}.$$

On retrouve un des exemples au debut du paragraphe 4.1.

Définition 5.4. *Deux variables aléatoires X et Y sont dites indépendantes pour $(x,y) \in X(\Omega) \times Y(\Omega)$ on a*

$$P(X = x \textit{ et } Y = y) = P(X = x)P(Y = y).$$

Cette définition d'introduire la notion de *loi de couple* :

Définition 5.5. *Soient X et Y deux variables aléatoires. On appelle loi de couple (ou loi conjointe) de (X,Y) la donnée de tous les $p_{ij} = P(X = x_i \textit{ et } Y = y_j)$ lorsque (x_i, y_j) prend toutes les valeurs possibles dans $X(\Omega) \times Y(\Omega)$.*

On remarque facilement que pour $x_0 \in X(\Omega)$ on a :

$$\sum_{y \in Y(\Omega)} P(X = x_0 \text{ et } Y = y) = P(X = x_0).$$

De même, pour $y_0 \in Y(\Omega)$ on a :

$$\sum_{x \in X(\Omega)} P(X = x \text{ et } Y = y_0) = P(Y = y_0).$$

On retrouve donc le loi de X et Y (qui prennent en ce context le nom de *lois marginales*) à partir de la loi de couple.

On présent la loi de couple sous forme de tableau. Si $X(\Omega) = \{x_1, \ldots, x_N\}$, $Y(\Omega) = \{y_1, \ldots, y_M\}$ et $p_{ij} = P(X = x_i \text{ et } Y = y_j)$, on écrira :

${}_{x_i}\backslash^{y_j}$	y_1	y_2	$\ldots$	y_M	l.m X
x_1	p_{11}	p_{12}	$\ldots$	p_{1M}	$\sum_j p_{1j}$
x_2	p_{21}	p_{22}	$\ldots$	p_{2M}	$\sum_j p_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_N	p_{N1}	p_{NN}	$\ldots$	p_{NM}	$\sum_j p_{Nj}$
l.m Y	$\sum_i p_{i1}$	$\sum_i p_{i2}$	$\ldots$	$\sum_i p_{iM}$	1

Reprenons à nouveau l'exemple du programme informatique. On s'intéresse maintenant au temps qui nécessite le programme pour écrire les deux premier bits et au temps nécessaire pour les deux derniers. On introduit donc deux nouvelles variables aléatoires X_1 et X_2 qui associent à chaque séquences les temps indiques. Par exemple pour $\omega = (1,1,0,1)$ on aura $X_1(\omega) = 2 + 2 = 4\mu$sec et $X_2(\omega) = 3 + 2 = 5\mu$sec. Notre intuition nous suggère que X_1 et X_2 soient indépendants. On constate que $X_1(\Omega) = X_2(\Omega) = \{4,5,6\}$ et de plus X_1 et X_2 ont la même loi (pour tout x on a $P(X_1 = x) = P(X_2 = x)$), donnée par :

x_i	4	5	6
p_i	1/4	1/2	1/4

(Attention : le fait que deux v.a. ayant la même loi n'implique pas qu'elles soient égales ; et en effet $X_1 \neq X_2$). On calcule aussi facilement le loi de couple de (X, Y) :

$x_i \backslash y_j$	4	5	6	l.m X
4	1/6	1/8	1/16	1/4
5	1/8	1/4	1/8	1/2
6	1/16	1/8	1/16	1/4
l.m Y	1/4	1/2	1/4	1

On peut maintenant vérifier que pour tout x et y on a :

$$P(X_1 = x \text{ et } X_2 = y) = P(X_1 = x)P(X_2 = y).$$

Par exemple $P(X_1 = 4 \text{ et } X_2 = 5) = \frac{1}{8}$ et $P(X_1 = 4)P(X_2 = 5) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$.

5.1 Paramètres d'une variable aléatoire

Dans toute ce paragraphe on continue à supposer que l'univers Ω soit fini.

5.1.1 Espérance

L'*espérance* d'une variable aléatoire X est la moyenne des valeurs de cette variable, pondérées par leurs probabilités. Elle représente intuitivement la valeur que l'on s'attend à trouver en moyenne en répétant pour X un grand nombre de fois la même expérience aléatoire.

Définition 5.6. *Soit X une variable aléatoire de loi (x_i, p_i). On définit l'espérance $E(X)$ de X par :*

$$E(X) = \sum_i p_i x_i.$$

Remarque. On parle de moyenne pondérée car $\sum_i p_i = 1$ et donc $\sum_i p_i x_i = \frac{\sum_i p_i x_i}{\sum_i p_i}$.

L'espérance est un opérateur linéaire :

Proposition 5.1. *Soient X et Y deux v.a. et soit $\lambda \in \mathbb{R}$. Alors*

$$E(X + Y) = E(X) + E(Y), \qquad E(\lambda X) = \lambda E(X).$$

On remarquera que si X est constante $= a$ (donc : $\forall \omega \in \Omega$, $X(\omega) = a$) on a évidemment $E(X) = a$.

On remarquera aussi que E est un opérateur croissante : si $X \leq Y$ (donc : $\forall \omega \in \Omega$, $X(\omega) \leq Y(\omega)$) alors $E(X) \leq E(Y)$. Cette inégalité reste vraie si l'on suppose seulement que $P(X \leq Y) = 1$.

Théorème 5.1. *Soient X et Y deux v.a. indépendantes. Alors $E(XY) = E(X)E(Y)$.*

5.2 Variance et écart type

La *variance* d'une variable aléatoire X est une mesure la «dispersion autour de l'espérance» : c'est la moyen des carrés des écarts à $E(X)$.

Définition 5.7. *Soit X une variable aléatoire de loi (x_i, p_i). On définit la variance $V(X)$ de X par :*

$$V(X) = E\Big(\big(X - E(X)\big)^2\Big) = \sum_i p_i(x_i - E(X))^2$$

(qui est donc toujours positive et nulle si et seulement si X est constant). On définit son écart type comme $\sigma_X = \sqrt{V(X)}$.

Théorème 5.2 (Koenigs)**.** *Soit X une v.a. de loi (x_i, p_i). On a alors :*

$$V(X) = E(X^2) - E(X)^2 = \sum_i p_i x_i^2 - \left(\sum_i p_i x_i\right)^2.$$

Démonstration. Par linéarité on a :

$$\begin{aligned}
V(X) &= E\Big(\big(X - E(X)\big)^2\Big) \\
&= E\Big(X^2 - 2E(X)X + E(X)^2\Big) \\
&= E(X^2) - 2E(X)E(X) + E\Big(E(X)^2\Big) \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}$$

□

Remarque. Pour tout $\lambda \in \mathbb{R}$ on a (par linéarité de l'espérance) :

$$\begin{aligned}
V(\lambda X) &= E\Big(\big(\lambda X - E(\lambda X)\big)^2\Big) \\
&= E\Big(\lambda^2\big(X - E(X)\big)^2\Big) = \lambda^2 V(X)
\end{aligned}$$

et

$$\begin{aligned}
V(X + \lambda) &= E\Big(\big(X + \lambda - E(X + \lambda)\big)^2\Big) \\
&= E\Big(\big(X + \lambda - E(X) - \lambda\big)^2\Big) = V(X).
\end{aligned}$$

Remarque. On considère maintenant plusieurs v.a. $X_1, \dots, X_n$ (pas nécessairement deux comme jusqu'à ici). En utilisant plusieurs fois la linéarité de l'espérance, on déduit que :

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

On dit que $X_1, \dots, X_n$ sont indépendants si $\forall (x_1, \dots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)$ on a

$$P(X_1 = x_1 \text{ et } X_2 = x_2 \text{ et } \cdots \text{ et } X_n = x_n) = P(X = x_1) \times \cdots \times P(X_n = x_n).$$

Pour des v.a. indépendants,

$$V(X_1 + \cdots + X_n) = V(X_1) + \cdots + V(X_n).$$

5.3 Covariance et coefficient de correlation

Définition 5.8. *Soient X et Y deux v.a. On définit la covariance de X et Y comme :*

$$Cov(X, Y) = E\Big((X - E(X)) \times (Y - E(Y))\Big).$$

Remarque. $Cov(X, X) = V(X)$.

Par linéarité de l'espérance on a la généralisation suivante du théorème de Koenigs :

Théorème 5.3. *Soient X, Y deux v.a. On a alors :*

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Démonstration.

$$\begin{aligned} E\Big((X - E(X)) \times (Y - E(Y))\Big) &= E\big(XY - E(X)Y - E(Y)X + E(X)E(Y)\big) \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

□

On sait que si X et Y sont indépendantes alors $E(XY) = E(X)E(Y)$. Donc

Corollary 5.1. *Soient X et Y deux v.a. indépendants. Alors $Cov(X, Y) = 0$.*

Remarque. Attention : la réciproque n'est pas vraie. On ne peut *pas* déduire de $Corr(X, Y) = 0$ que X et Y sont indépendants.

La covariance est un opérateur bilinéaire symétrique :

Proposition 5.2. *Soient X, Y, Z trois v.a. et soit $\lambda \in \mathbb{R}$.*

$$Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z), \qquad Cov(\lambda X, Y) = \lambda Cov(X,Y)$$

et $Cov(X,Y) = Cov(Y,X)$.

La variance n'est *pas* un opérateur linéaire. Néanmoins :

Proposition 5.3. *Soient X, Y deux v.a. On a alors :*

$$V(X+Y) = V(X) + 2Cov(X,Y) + V(Y).$$

En particulier, si X et Y sont indépendantes, $V(X+Y) = V(X) + V(Y)$.

Définition 5.9. *Soient X et Y deux v.a.. On suppose $V(X) \neq 0$ et $V(Y) \neq 0$. On définit alors le coefficient de corrélation de X et Y comme :*

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Remarque. On peut montrer (à l'aide de l'inégalité de Cauchy-Schwarz) que

$$|Corr(X,Y)| \leq 1.$$

Revenons au programme informatique et aux variables X, X_1 et X_2 qui associent à chaque séquences de 4 bits le temps nécessaire pour l'écrire, pour écrire les deux premier bits et pour écrire le deux derniers respectivement. Nous avons constaté que X_1 et X_2 ont la même loi :

x_i	4	5	6
p_i	1/4	1/2	1/4

On en déduit alors (avec $h = 1$ ou $h = 2$: c'est pareil !)

$$E(X_h) = \sum_i p_i x_i = \frac{1}{4} \times 4 + \frac{1}{2} \times 5 + \frac{1}{4} \times 6 = 5$$

$$E(X_h^2) = \sum_i p_i x_i^2 = \frac{1}{4} \times 4^2 + \frac{1}{2} \times 5^2 + \frac{1}{4} \times 6^2 = \frac{51}{2}$$

$$V(X_h) = E(X_h^2) - E(X_h)^2 = \frac{1}{2}$$

$$\sigma_{X_h} = \sqrt{V(X_h)} = \frac{\sqrt{2}}{2}$$

Pour déterminer la covariance de X_1 et X_2, on considère la variable $T = X_1 X_2$. Alors $T(\Omega)$ est l'ensemble de tout les produits de deux nombres $\in \{4,5,6\}$: $T(\Omega) =$

$\{16, 20, 24, 25, 30, 36\}$. On calcule facilement la loi de T. Par exemple, en utilisant que X_1 et X_2 sont indépendants de même loi,

$$\begin{aligned} P(T = 24) &= P(X_1 = 4 \text{ et } X_2 = 6) + P(X_1 = 6 \text{ et } X_2 = 4) \\ &= 2P(X_1 = 4)P(X_2 = 6) = 1/8. \end{aligned}$$

On obtient :

t_i	16	20	24	25	30	36
p_i	1/16	1/4	1/8	1/4	1/4	1/16

D'ou $E(X_1X_2) = \frac{1}{16} \times 16 + \frac{1}{14} \times 20 + \frac{1}{8} \times 24 + \frac{1}{4} \times 25 + \frac{1}{4} \times 30 + \frac{1}{16} \times 36 = 25$ et donc

$$Cov(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = 25 - 5 \times 5 = 0$$

en accord avec le fait que X_1 et X_2 sont indépendants.

La loi de X est

x_i	8	9	10	11	12
p_i	1/16	4/16	6/16	4/16	1/16

ce qui donne

$$\begin{aligned} E(X) &= \sum_i p_i x_i = \frac{1}{16} \times 8 + \frac{4}{16} \times 9 + \frac{6}{16} \times 10 + \frac{4}{16} \times 11 + \frac{1}{16} \times 12 = 10 \\ E(X^2) &= \sum_i p_i x_i^2 = \frac{1}{16} \times 8^2 + \frac{4}{16} \times 9^2 + \frac{6}{16} \times 10^2 + \frac{4}{16} \times 11^2 + \frac{1}{16} \times 12^2 = 101 \\ V(X) &= E(X^2) - E(X)^2 = 1 \\ \sigma_X &= \sqrt{V(X)} = 1 \end{aligned}$$

On aurait pu retrouver la valeur de $E(X)$ en remarquant que $X = X_1 + X_2$, ce qui donne $E(X) = E(X_1) + E(X_2) = 5 + 5 = 10$. De même, *en sachant que X_1, X_2 sont indépendants*, $V(X) = V(X_1) + V(X_2) = 1/2 + 1/2 = 1$.

On se propose maintenant de déterminer la covariance des v.a. X et X_1. On connait leur espérance. Il reste à déterminer l'espérance du produit, et pour cela il nous faut la loi de XX_1. En utilisant le fait que $X = X_1 + X_2$, la linéarité de l'espérance et les calculs précédents on obtient :

$$E(XX_1) = E\big((X_1 + X_2)X_1\big) = E(X_1^2) + E(X_1X_2) = \frac{51}{2} + 25 = \frac{101}{2}.$$

La covariance de X et X_1 est donc :

$$Cov(X, X_1) = E(XX_1) - E(X)E(X_1) = \frac{101}{2} - 10 \times 5 = \frac{1}{2}.$$

Le coefficient de corrélation de X et X_1 est

$$Corr(X, X_1) = \frac{Cov(X, X_1)}{\sigma_X \sigma_{X_1}} = \frac{\frac{1}{2}}{1 \times \frac{\sqrt{2}}{2}} = \frac{\sqrt{2}}{2}.$$

Revenons au lancer de deux dés et à la variable aléatoire X «somme des nombres de chaque dé». Nous avons déjà calculé la loi de X :

x_i	2	3	4	5	6	7	8	9	10	11	12
p_i	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

On on déduit les valeurs de son espérance, de l'espérance de son carré, de sa variance et de son écart type :

$$\begin{aligned}
E(X) &= \sum_i p_i x_i = \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \frac{4}{36} \times 5 + \frac{5}{36} \times 6 \\
&\qquad + \frac{6}{36} \times 7 + \frac{5}{36} \times 8 + \frac{4}{36} \times 9 + \frac{3}{36} \times 10 + \frac{2}{36} \times 11 + \frac{1}{36} \times 12 = 7 \\
E(X^2) &= \sum_i p_i x_i^2 = \frac{1}{36} \times 2^2 + \frac{2}{36} \times 3^2 + \frac{3}{36} \times 4^2 + \frac{4}{36} \times 5^2 + \frac{5}{36} \times 6^2 \\
&\qquad + \frac{6}{36} \times 7^2 + \frac{5}{36} \times 8^2 + \frac{4}{36} \times 9^2 + \frac{3}{36} \times 10^2 + \frac{2}{36} \times 11^2 + \frac{1}{36} \times 12^2 = \frac{329}{6} \\
V(X) &= E(X^2) - E(X)^2 = \frac{35}{6} \approx 5.83 \\
\sigma_X &= \sqrt{V(X)} = \sqrt{\frac{35}{6}} \approx 2.41
\end{aligned}$$

5.4 Quelques lois discrètes usuelles

On considère une v.a. X définie sur un ensemble fini Ω.

5.4.1 Loi discrète uniforme

On dit que X suit la *loi uniforme* si l'univers image est $X(\Omega) = \{1, \ldots, n\}$ et $P(X = k) = 1/n$ pour $k = 1, \ldots, n$. En utilisant les formules :

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \quad \text{et} \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

on obtient :

$$E(X) = \sum_{k=1}^{n} \frac{k}{n} = \frac{1}{n} \times \frac{n(n+1)}{2} = \frac{n+1}{2}$$

et $E(X^2) = \sum_{k=1}^{n} \frac{k^2}{n} = \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$ d'où

$$\begin{aligned} V(X) = E(X^2) - E(X)^2 &= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\ &= \frac{(n+1)\big(2(2n+1) - 3(n+1)\big)}{12} \\ &= \frac{(n+1)(n-1)}{12} \\ &= \frac{n^2-1}{12}. \end{aligned}$$

Exemple : On lance un dé non pipé à 6 faces et on définit la variable aléatoire X égale au résultat du lancer. L'univers est $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $P(i) = \frac{1}{6}$ pour tout $i \in \Omega$. La v.a. X suit donc la loi uniforme (avec $n = 6$) et

$$E(X) = \frac{6+1}{2} = \frac{7}{2}, \qquad V(X) = \frac{6^2-1}{12} = \frac{35}{12}.$$

Nous lançons maintenant k dés. On définit la v.a. X comme la somme des nombres de chaque dé. Soit X_i la v.a. égale au résultat du i-ème lancer. Alors $X_1, \ldots, X_k$ sont des v.a. indépendants qui suivent la loi uniforme (avec $n = 6$) et $X = X_1 + \cdots X_k$. On en déduit (rapidement) le valeurs de l'espérance et de la variance qu'on avait déjà calculé :

$$E(X) = \frac{7k}{2}, \qquad V(X) = \frac{35k}{12}.$$

5.4.2 Loi de Bernoulli

On dit que X suit une *loi de Bernoulli* de paramètre $p \in]0, 1[$ si $X(\Omega) = \{0, 1\}$ (donc X ne peut prendre que 2 valeurs, 0 ou 1), et que $p_1 = P(X = 0) = 1 - p$ et $p_2 = P(X = 1) = p$. On a :

$$E(X) = p_1 x_1 + p_2 x_2 = p$$

et $E(X^2) = E(X)$ (car X est à valeurs en $\{0, 1\}$), d'où

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p).$$

On obtient une loi de Bernoulli dès que l'on s'intéresse à la réalisation ou non d'un événement fixé A de probabilité p.

Exemple : On lance une monnaie équilibrée et l'on s'intéresse à la v.a. X qui prend la valeur 0 si l'on obtient pile et la valeur 1 si l'on obtient face. Alors X suit une loi de Bernoulli de paramètre $p = 1/2$.

5.4.3 Loi binomiale

On dit que X suit une *loi binomiale* de paramètres (n,p) si l'univers image est $X(\Omega) = \{0,1,\dots,n\}$ et

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

pour $k = 0,\dots,n$.

On obtient une loi binomiale dès que l'on répète n fois la même expérience aléatoire, où l'on s'intéresse à la réalisation ou non d'un événement fixé A de probabilité p, les répétitions étant supposées indépendantes. On note X_i la v.a. qui prend la valeur 0 ou 1 selon que à la n-ième répétition l'événement A se réalise et on considère $X = X_1 + \dots + X_n$. La v.a. X suit la loi binomiale de paramètres (n,p). En effet $\{X = k\}$ est la réunion des événements incompatibles $\cap_{j\in J}\{X_j = 1\}$ pour J sous-ensemble à k éléments de $\{1,\dots,n\}$. Or le nombre de ces sous-ensembles vaut $\binom{n}{k}$ et chacun de ces événements à probabilité $p^k(1-p)^{n-k}$.

Les variables X_i suivent toutes une loi de Bernoulli de paramètre p. Donc, par linéarité de l'espérance,

$$E(X) = E(X_1) + \dots + E(X_n) = np.$$

De plus :

$$V(X) = V(X_1) + \dots + V(X_n) = np(1-p)$$

car $X_1,\dots,X_n$ sont indépendants.

5.4.4 Loi hypergéométrique

Soient N et $n \le N$ deux entiers strictement positifs et $p \in [0,1]$ tel que Np soit entier.

On dit que X suit une *loi hypergéométrique* de paramètres (N,n,p) si l'univers image est $X(\Omega) = \{k \in [0,\dots,n] \mid k \le Np,\ n-k \le N-Np\}$ et

$$P(X=k) = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}}$$

pour $k \in X(\Omega)$.

C'est la loi d'une variable aléatoire X donnant le nombre d'individus ayant un caractère donné lors du tirage de n individus dans une population de N individus au sein de laquelle il y a Np personnes ayant le dit caractère. En effet on a :

$$\binom{N}{n}$$

tirages possibles. Et le nombre de tirages de n individus dont précisément k ont le caractère donné vaut

$$\binom{Np}{k}\binom{N-Np}{n-k},$$

car un tel tirage revient à sélectionner un sous-ensemble à k éléments d'un ensemble à Np éléments (l'ensemble des Np individu n'ayant pas le dit caractère) et un sous-ensemble à $n-k$ éléments d'un ensemble à $N-Np$ éléments (l'ensemble des $N-Np$ individu n'ayant pas le dit caractère).

Il s'agit d'un tirage non ordonné et sans remise. Mais la loi reste la même si l'on ordonne le tirage [1]. Dans ce cas, en notant X_i la v.a. qui prend la valeur 1 ou 0 selon que à l'i-ième tirage on tire ou non un individu ayant le caractère souhaité, on a encore $X = X_1+\cdots+X_n$ et les X_i suivent toujours une loi de Bernoulli de paramètre p. Donc :

$$E(X) = E(X_1) + \cdots + E(X_n) = np.$$

Mais cette fois les v.a. X_i ne sont plus indépendants. On peut cependant montrer une formule simple pour la variance :

$$V(X) = np(1-p)\frac{N-n}{N-1}.$$

Exemple : On considère une urne qui contient N boules, dont N_1 rouges et $N-N_1$ blancs. On pose $p = N_1/N$. On tire n boules et on appelle X la v.a. donnant le nombre de boules rouges extraites (au total).

Si on effectue des tirages *avec remise* la v.a. X suit la loi binomiale de paramètres (n,p). Ona aura alors :

$$E(X) = np, \qquad V(X) = np(1-p).$$

Si en revanche on effectue des tirages *sans remise* la v.a. X suivra la loi hypergéométrique de paramètres (N,n,p) et donc

$$E(X) = np, \qquad V(X) = np(1-p)\frac{N-n}{N-1}.$$

On remarque que l'espérance est la même et que, dès que N est grand, la variance dans le cas sans remise «se rapproche» de la variance du cas avec remise, car

$$\lim_{N\to+\infty}\frac{N-n}{N-1} = 1.$$

1. Cela revient à multiplier par $n!$ au numérateur et dénominateur de la quantité $P(X=k)$ de n éléments parmi N

Exemple : Un parc animalier abrite $N = 40$ animaux : $N_1 = 25$ dromadaires et 15 lamas. On prend une photographie sur laquelle $n = 5$ de ces 40 camélidés sont présents, et on suppose que chaque animal avait la même probabilité d'être ainsi immortalisé : on assimilera donc la prise de ce cliché à un tirage équiprobable sans remise de 5 animaux. Nous rappelons que les dromadaires ont une bosse et les lamas aucune mais un regard pénétrant les rendant très photogéniques.

On appelle X la v.a. donnant le nombre de bosses sur l'image. Notons $p = N_1/N = 5/8$. Alors X suivra la loi hypergéométrique de paramètres (N, n, p) :

$$P(X = k) = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}} = \frac{\binom{25}{k}\binom{15}{5-k}}{\binom{40}{5}}.$$

En particulier $E(X) = np = \frac{25}{8} = 3.125$ et

$$V(X) = np(1-p)\frac{N-n}{N-1} = \frac{25}{8}\left(1 - \frac{5}{8}\right)\frac{35}{39} = \frac{875}{832} \approx 1.05.$$