PROBABILITÉS (révisions du cours de 1ère année)

I. Probabilités sur un univers fini.

a. Le vocabulaire de base

On considère une **expérience aléatoire** (lancer d'un dé, tirage d'une boule dans une urne). On appelle **univers** et on note habituellement Ω l'ensemble des **issues** (ou **résultats**, ou **réalisations**) possibles.

Exemples. Pour le lancer d'un dé, on choisira $\Omega = [\![1,6]\!]$. Pour le tiercé, s'il y a 18 chevaux au départ, on choisira l'ensemble des triplets d'éléments distincts dans l'intervalle $[\![1,18]\!]$ et on aura donc $|\Omega| = 18 \times 17 \times 16$.

Toute partie de l'ensemble Ω est appelée un **événement**. Un événement A est donc un élément de $\mathcal{P}(\Omega)$. Les **événements élémentaires** sont les singletons $\{\omega\}$ avec $\omega \in \Omega$.

Si A est un événement, on définit son **événement contraire**, noté $\overline{A}$, qui est le complémentaire de A dans Ω, c'est donc l'événement qui se réalise si et seulement si A ne se réalise pas. On peut noter aussi $\overline{A} = \Omega \setminus A$, ou encore $\overline{A} = A^c$.

L'ensemble vide $\emptyset$ est l'**événement impossible**.

Si A et B sont deux événements, on définit l' événement "A **et** B" qui se réalise si et seulement si les événements A et B sont tous deux réalisés. Dans le vocabulaire ensembliste, il s'agit de l'intersection $A \cap B$. On définit aussi l'événement "A **ou** B" qui se réalise lorsqu'au moins un des deux événements A ou B se réalise, il correspond à la réunion $A \cup B$.

Deux événements A et B sont dits **disjoints** ou **incompatibles** lorsque $A \cap B = \emptyset$.

Enfin, on appelle **système complet d'événements** toute famille finie $(A_1, \cdots, A_n)$ d'événements deux à deux disjoints (ou "incompatibles") et dont la réunion est Ω.

Le programme de première année se limite aux univers finis.

Exemple. On tire une carte au hasard dans un jeu de 52 cartes. Alors $|\Omega| = 52$, les événements $P =$"c'est un pique", $C =$"c'est un cœur", $K =$"c'est un carreau", $T =$"c'est un trèfle" constituent un système complet.

b. Espaces probabilisés finis

On appelle **probabilité** sur un univers fini Ω toute application $P : \mathcal{P}(\Omega) \to [0,1]$ telle que $P(\Omega) = 1$ et telle que, si A et B sont deux parties de Ω disjointes, on ait la relation $P(A \cup B) = P(A) + P(B)$. Le couple (Ω, P) est alors appelé un **espace probabilisé fini**.

On appelle **distribution de probabilités** sur un ensemble fini E toute famille $(p_x)_{x\in E}$ de réels positifs indexée par E et de somme 1, i.e. telle que

$$\big(\forall x \in E \quad p_x \geq 0\big) \quad \text{et} \quad \sum_{x\in E} p_x = 1 \ .$$

Proposition. Une probabilité P sur un univers fini Ω est entièrement déterminée par les images des singletons (ou "événements élémentaires"), autrement dit par la distribution de probabilités $\Big(P\big(\{\omega\}\big)\Big)_{\omega\in\Omega}$. Cela signifie que, si $(p_\omega)_{\omega\in\Omega}$ est une famille de réels positifs telle que $\displaystyle\sum_{\omega\in\Omega} p_\omega = 1$, alors il existe une unique probabilité P sur Ω telle que, pour tout $\omega \in \Omega$, on ait $P\big(\{\omega\}\big) = p_\omega$. Cette probabilité P est alors définie par

$$\forall A \in \mathcal{P}(\Omega) \qquad P(A) = \sum_{\omega\in A} p_\omega = \sum_{\omega\in A} P\big(\{\omega\}\big) \ .$$

Définition. Si $|\Omega| = n$, la **probabilité uniforme** sur Ω est celle définie par $p_\omega = \dfrac{1}{n}$ pour tout $\omega \in \Omega$. On a alors $P(A) = \dfrac{|A|}{|\Omega|}$ pour tout événement A. On parle de situation d'**équiprobabilité**. La probabilité de l'événement A est alors “le nombre de cas favorables” $|A|$ divisé par “le nombre de cas possibles” (i.e. le cardinal de l'univers) $|\Omega|$. Les calculs de probabilités se ramènent dans ce cas à de simples problèmes de dénombrement.

Propriétés. Soit P une probabilité sur un univers fini Ω. On a alors les propriétés suivantes:

- Probabilité de l'événement contraire: $\forall A \in \mathcal{P}(\Omega) \qquad P(\overline{A}) = 1 - P(A)$;
- Croissance: si $A \subset B$, alors $P(A) \leq P(B)$;
- Probabilité de la réunion de deux événements: si A et B sont deux événements,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \ .$$

- Probabilité de la différence de deux événements: si A et B sont deux événements,

- dans le cas où $B \subset A$, alors $A \setminus B$ est le **complémentaire** de B dans A, et A est la réunion **disjointe** de B et de $A \setminus B$, donc

$$P(A \setminus B) = P(A) - P(B) \ .$$

- dans le cas général, la différence $A \setminus B$ est le complémentaire de $A \cap B$ dans A, donc

$$P(A \setminus B) = P(A) - P(A \cap B) = P(A \cup B) - P(B) \ .$$

En conséquence, on a $P(\emptyset) = 0$ et, si $A_1, \cdots, A_n$ sont des événements deux à deux incompatibles, alors $P\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} P(A_k)$.

c. Probabilités conditionnelles

Définition. Si A et B sont deux événements, avec $P(B) > 0$, le nombre $\dfrac{P(A \cap B)}{P(B)}$ est appelé **probabilité conditionnelle de** A **sachant** B, on le note $P(A|B)$ ou encore $P_B(A)$.

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

Proposition. Si B est un événement tel que $P(B) > 0$, l'application $P_B : \mathcal{P}(\Omega) \to [0,1]$ est une probabilité sur Ω.

Formule des probabilités composées. Si A et B sont deux événements avec $P(B) > 0$, on a

$$P(A \cap B) = P(A|B)\, P(B) \ .$$

Plus généralement, si $A_1, \cdots, A_n$ sont des événements tels que $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, alors

$$P(A_1 \cap \cdots \cap A_n) = P(A_n|A_1 \cap \cdots \cap A_{n-1})\, P(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}) \cdots P(A_2|A_1)\, P(A_1) \ .$$

Formule des probabilités totales. Si $(A_1, \cdots, A_n)$ est un système complet d'événements, si B est un événement quelconque, on a

$$P(B) = \sum_{k=1}^{n} P(B \cap A_k) \; .$$

Si, de plus, chacun des A_k a une probabilité non nulle, on peut écrire

$$P(B) = \sum_{k=1}^{n} P(B|A_k)\, P(A_k) \; .$$

On pourra illustrer l'utilisation de ces formules sur des arbres de probabilité.

Exemple. Deux urnes X et Y contiennent des boules blanches et noires. Un mécanisme fait qu'on tire dans l'urne X avec probabilité $\frac{1}{3}$, et bien sûr dans l'urne Y avec probabilité $\frac{2}{3}$. La proportion de boules blanches étant de 60% dans l'urne X et de 75% dans l'urne Y, quelle est la probabilité de tirer une boule blanche ?

Formules de Bayes.

(1) : Si A et B sont deux événements de probabilités non nulles, alors

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

(2) : Si $(A_1, \cdots, A_n)$ est un système complet d'événements de probabilités non nulles, si B est un événement de probabilité non nulle, alors

$$\forall k \in [\![1, n]\!] \qquad P(A_k|B) = \frac{P(B|A_k)\, P(A_k)}{\displaystyle\sum_{i=1}^{n} P(B|A_i)\, P(A_i)} \; .$$

Convention. La formule des probabilités totales, ainsi que la deuxième formule de Bayes, restent vraies si l'on convient que, dans le cas où l'un des A_i est de probabilité nulle, on pose $P(B|A_i)\, P(A_i) = 0$.

d. Événements indépendants

Définition 1. Dans un espace probabilisé fini (Ω, P), deux événements A et B sont dits **indépendants** lorsque $P(A \cap B) = P(A)\, P(B)$.
Si $P(B) > 0$, l'indépendance de A et B se traduit par $P(A|B) = P(A)$: la réalisation ou non de B n'influe pas sur la probabilité de réalisation de A.

Remarque. Il faudra se garder d'une interprétation intuitive du genre “absence de lien de cause à effet”, puisque la notion d'indépendance dépend du choix de la probabilité sur l'univers Ω.

Proposition. Si (A, B) est un couple d'événements indépendants, il en est de même des couples $(A, \overline{B})$, $(\overline{A}, B)$ et $(\overline{A}, \overline{B})$.

Définition 2. Des événements A_1, $\cdots$, A_n sont dits (**mutuellement**) **indépendants** lorsque, pour toute partie J de l'intervalle entier $[\![1, n]\!]$, on a $P\Big(\bigcap_{i\in J} A_i\Big) = \prod_{i\in J} P(A_i)$, autrement dit lorsque la probabilité de l'intersection d'un nombre quelconque de ces événements est égale au produit de leurs probabilités. On peut écrire cela aussi en disant que, si $k \in [\![1, n]\!]$, si $i_1, \cdots, i_k$ sont des entiers tels que $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, on a

$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}) .$$

Remarque. Si $(A_1, \cdots, A_n)$ est une famille finie d'événements indépendants, alors ces événements sont indépendants deux à deux. Par contre, la réciproque est fausse: l'indépendance deux à deux n'entraîne pas l'indépendance mutuelle.

Remarque très personnelle: c'est un peu comme les sous-espaces vectoriels en somme directe: si $F_1, \cdots, F_m$ sont des s.e.v. de E en somme directe, alors ils sont deux à deux en somme directe ($F_i \cap F_j = \{0_E\}$ si $i \neq j$), mais la réciproque est fausse!

Exemple. On lance deux dés, l'univers est naturellement $\Omega = [\![1, 6]\!]^2$, que l'on munit de l'équiprobabilité. Les événements:

A_1 : le premier dé donne un résultat pair ;

A_2 : le deuxième dé donne un résultat pair ;

A_3 : la somme des deux résultats est un nombre pair

sont indépendants deux à deux, mais pas mutuellement indépendants.

Proposition. Soient $A_1, \cdots, A_n$ des événements indépendants. Soient $B_1, \cdots, B_n$ des événements tels que, pour tout $i \in [\![1, n]\!]$, on ait $B_i = A_i$ ou $B_i = \overline{A_i}$. Alors les événements $B_1, \cdots, B_n$ sont indépendants.

II. Variables aléatoires et lois.

a. Définitions et notations

Définition. On appelle **variable aléatoire** (en abrégé v.a.) toute application définie sur l'univers Ω et à valeurs dans un ensemble E. En particulier, une **variable aléatoire réelle** (en abrégé v.a.r.) sur Ω est une application de Ω vers $\mathbb{R}$.

Notations. Si $X : \Omega \to E$ est une variable aléatoire, si F est une partie de E, l'image réciproque $X^{-1}(F) = \{\omega \in \Omega \mid X(\omega) \in F\}$ est souvent noté $\{X \in F\}$ ou encore $(X \in F)$. De le même façon, si $x \in E$, on notera $\{X = x\}$ ou encore $(X = x)$ l'ensemble $X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}$. Si $E \subset \mathbb{R}$, et si $x \in \mathbb{R}$, on notera aussi $\{X \leq x\}$ ou encore $(X \leq x)$ l'ensemble $X^{-1}(]-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\}$ et ainsi de suite.

Si P est une probabilité sur l'univers Ω, si $F \subset E$, on écrira par exemple $P(X \in F)$ pour $P(X^{-1}(F))$. On verra aussi des notations du style $P(X = x)$, $P(X \leq x)$, $P(a \leq X \leq b)$, etc.

Exemple. On lance successivement deux dés. L'univers est naturellement $\Omega = [\![1, 6]\!]^2$, muni de la probabilité uniforme. On note $X : \Omega \to \mathbb{R}$, $(a, b) \mapsto a + b$, autrement dit la variable aléatoire X correspond à la somme des résultats des deux lancers. Calculer $P(X = 10)$, $P(X \leq 4)$, $P(X \in \{2; 12\})$.

Définition. Soit X une variable aléatoire définie sur un espace probabilisé fini (Ω, P), et à valeurs dans un ensemble fini E. On a donc $X(\Omega) \subset E$. L'application $P_X : \mathcal{P}(E) \to [0, 1]$ définie par $P_X(F) = P(X \in F) = P(X^{-1}(F))$ pour tout $F \in \mathcal{P}(E)$, est appelée **loi de la variable aléatoire** X. *Elle dépend bien sûr de la probabilité* P *choisie sur l'univers* Ω.

Propriété. Cette application P_X est une probabilité sur l'ensemble fini E, et elle est déterminée par la distribution de probabilités $\big(P(X=x)\big)_{x\in E}$. Par souci de simplification, dans le cas de ces "événements élémentaires", le réel $P_X\big(\{x\}\big)=P(X=x)$, pour $x\in E$, est souvent noté abusivement $P_X(x)$.

Remarque. On prendra souvent $E=X(\Omega)$. L'ensemble $X(\Omega)$ étant fini, si k est son cardinal, on peut le noter $X(\Omega)=\{x_1,\cdots,x_k\}$. Connaître la loi P_X de la variable X revient à se donner une liste de couples (x_i,p_i), $1\leq i\leq k$, avec $p_i\in[0,1]$ tels que $\displaystyle\sum_{i=1}^{k}p_i=1$. Ainsi, pour tout $i\in[\![1,k]\!]$, $P_X\big(\{x_i\}\big)=P(X=x_i)=p_i$.

Exemple. Reprenons les deux lancers de dés consécutifs ci-dessus. On a $X(\Omega)=[\![2,12]\!]$. On aura par exemple $P_X(12)=\dfrac{1}{36}$ mais $P_X(7)=\dfrac{6}{36}=\dfrac{1}{6}$. La loi de X est donnée par les couples $\big(x,P_X(x)\big)$ avec $2\leq x\leq 12$, soit

$$\left\{\left(2,\frac{1}{36}\right)\left(3,\frac{2}{36}\right)\left(4,\frac{3}{36}\right)\left(5,\frac{4}{36}\right)\left(6,\frac{5}{36}\right)\left(7,\frac{6}{36}\right)\left(8,\frac{5}{36}\right)\left(9,\frac{4}{36}\right)\left(10,\frac{3}{36}\right)\left(11,\frac{2}{36}\right)\left(12,\frac{1}{36}\right)\right\}.$$

Notation. Lorsque deux variables aléatoires X et Y, définies sur le même espace probabilisé fini (Ω,P), ont la même loi, i.e. lorsque $X(\Omega)=Y(\Omega)$ et $P_X=P_Y$, on note alors $X\sim Y$.

Image d'une variable aléatoire. Soit $X:\Omega\to E$ une variable aléatoire, soit $f:E\to E'$ une application, alors l'application composée $Y=f\circ X$ est une variable aléatoire sur l'univers Ω à valeurs dans E', on la notera $Y=f(X)$, on l'appellera **image** de la v.a. X par la fonction f. On dit aussi que la variable aléatoire $Y=f(X)$ est une **fonction** de la variable aléatoire X. La loi P_Y de Y est alors définie par

$$\forall F\in\mathcal{P}(E')\qquad P_Y(F)=P(Y\in F)=P\big(f(X)\in F\big)=P\big(X\in f^{-1}(F)\big)=P_X\big(f^{-1}(F)\big).$$

Conséquence. Si $X_1\sim X_2$, alors $f(X_1)\sim f(X_2)$.

Notion de loi conditionnelle. Soit (Ω,P) un espace probabilisé fini, soit $A\in\mathcal{P}(\Omega)$ un événement tel que $P(A)>0$. On peut alors considérer un nouvel espace probabilisé fini (Ω,P_A) où P_A est la "probabilité conditionnelle sachant A", i.e.

$$\forall B\in\mathcal{P}(\Omega)\quad P_A(B)=P(B|A)=\frac{P(B\,\cap\,A)}{P(A)}.$$

Si X est une variable aléatoire sur Ω, i.e. une application de Ω vers un ensemble fini E, on définit la **loi conditionnelle de X sachant l'événement** A comme étant la loi de la variable aléatoire X sur l'espace probabilisé (Ω,P_A). C'est donc l'application $P_{X,A}$ (*il n'y a pas vraiment de notation officielle*) définie par

$$P_{X,A}:\begin{cases}\mathcal{P}(E)\ \to\ [0,1]\\[2ex] \quad F\quad\mapsto\ P_A(X\in F)=\dfrac{P\big((X\in F)\,\cap\,A\big)}{P(A)}\end{cases}.$$

b. Lois usuelles

• Soit E un ensemble fini non vide, soit $n = \mathrm{Card}(E) \in \mathbb{N}^*$, notons alors $E = \{x_1, \cdots, x_n\}$ où les x_i sont distincts. Une variable aléatoire X sur un espace probabilisé fini (Ω, P) est dite **variable uniforme sur** E si on a $X(\Omega) = E$ et, pour tout $i \in [\![1, n]\!]$, $P(X = x_i) = \dfrac{1}{n}$. On dit encore que X suit la loi uniforme sur E et on note $X \sim \mathcal{U}(E)$.

• Soit $p \in [0, 1]$. On dit qu'une v.a. X sur (Ω, P) est une **variable de Bernoulli de paramètre** p si on a $X(\Omega) = \{0, 1\}$, $P_X(1) = p$ et $P_X(0) = 1 - p$. On dit aussi que la variable X suit la loi de Bernoulli de paramètre p, et on note $X \sim \mathcal{B}(p)$.

Cette loi de Bernoulli correspond à toutes les situations "binaires" (échec ou succès d'une expérience) comme un tirage à pile ou face, un choix entre deux éventualités.

Si $A \in \mathcal{P}(\Omega)$ est un événement, on appelle **fonction indicatrice** de A l'application $\mathbb{1}_A : \Omega \to \{0, 1\}$ définie par $\mathbb{1}_A(\omega) = 1$ si $\omega \in A$, $\mathbb{1}_A(\omega) = 0$ sinon. Cette application $\mathbb{1}_A$ peut être considérée comme une variable aléatoire de Bernoulli de paramètre $p = P(A)$.

• Soient $n \in \mathbb{N}^*$ et $p \in [0, 1]$. On dit qu'une v.a. X sur (Ω, P) est une **variable binomiale de paramètres** n **et** p si $X(\Omega) = [\![0, n]\!]$ et si, pour tout $k \in [\![0, n]\!]$, on a

$$P_X(k) = \binom{n}{k} p^k (1 - p)^{n-k} \ .$$

On dit encore que la v.a. X suit la loi binomiale de paramètres n et p, ce que l'on note $X \sim \mathcal{B}(n, p)$. On en verra bientôt une interprétation dans une situation concrète.

c. Couples de variables aléatoires

Soit (Ω, P) un espace probabilisé fini. Soient $X : \Omega \to E$ et $Y : \Omega \to F$ deux variables aléatoires. On peut alors considérer une nouvelle variable aléatoire $U = (X, Y) : \Omega \to E \times F$ telle que $U(\omega) = \big(X(\omega), Y(\omega)\big)$ pour tout $\omega \in \Omega$. La loi P_U de cette variable aléatoire U est appelée **loi conjointe** de X et Y. Les lois P_X et P_Y des variables aléatoires X et Y sont alors appelées **lois marginales** du couple (X, Y).

La loi conjointe $P_U = P_{(X,Y)}$ **détermine les lois marginales** P_X **et** P_Y. En effet, si $x \in X(\Omega)$, on a $\{X = x\} = \displaystyle\bigsqcup_{y \in Y(\Omega)} \big(\{X = x\} \cap \{Y = y\}\big) = \bigsqcup_{y \in Y(\Omega)} \big\{U = (x, y)\big\}$ (*j'utilise la notation* $\bigsqcup$ *pour désigner une réunion disjointe*), donc

$$P_X(x) = P(X = x) = \sum_{y \in Y(\Omega)} P\big((X, Y) = (x, y)\big) = \sum_{y \in Y(\Omega)} P_U(x, y) \ .$$

De même, si $y \in Y(\Omega)$, $P_Y(y) = \displaystyle\sum_{x \in X(\Omega)} P_U(x, y)$.

Remarque. La notation correcte pour $P_U(x, y)$ serait $P_U\Big(\big\{(x, y)\big\}\Big)$ mais bôf!

La probabilité $P\big(\{X = x\} \cap \{Y = y\}\big) = P\big(U = (x, y)\big) = P_U(x, y)$ est couramment notée $P(X = x, Y = y)$. *À ne pas confondre avec la probabilité conditionnelle* $P(X = x | Y = y)$! On a ainsi $P(X = x) = \displaystyle\sum_{y \in Y(\Omega)} P(X = x, Y = y)$ et $P(Y = y) = \displaystyle\sum_{x \in X(\Omega)} P(X = x, Y = y)$.

On peut aussi remarquer que, si A est une partie de E et B une partie de F, alors

$$P_X(A) = P_U(A \times F) \qquad \text{et} \qquad P_Y(B) = P_U(E \times B) \ ,$$

soit

$$P(X \in A) = P\big((X,Y) \in A \times F\big) \qquad \text{et} \qquad P(Y \in B) = P\big((X,Y) \in E \times B\big) \ .$$

En revanche, les lois marginales ne déterminent pas la loi conjointe. Voici un exemple illustrant cette remarque et permettant de comprendre le terme de “lois marginales”.

Il est possible de généraliser à des n-uplets de variables aléatoires: si $X_1 : \Omega \to E_1$, $\cdots$, $X_n : \Omega \to E_n$ sont des v.a. sur le même espace probabilisé fini (Ω, P), on peut considérer la variable aléatoire $X = (X_1, \cdots, X_n)$, à valeurs dans le produit cartésien $E_1 \times \cdots \times E_n$, telle que, pour tout $\omega \in \Omega$, on ait $X(\omega) = \big(X_1(\omega), \cdots, X_n(\omega)\big)$. On parle aussi de **loi conjointe** (celle de X) et de **lois marginales** (celles des X_i, $1 \leq i \leq n$). **La loi conjointe détermine les lois marginales, mais la réciproque est fausse.**

d. Variables aléatoires indépendantes

Définition. Deux variables aléatoires X et Y définies sur le même espace probabilisé fini (Ω, P) sont dites **indépendantes**, et on note $X \perp\!\!\!\perp Y$, lorsque, pour tout $A \in \mathcal{P}\big(X(\Omega)\big)$ et tout $B \in \mathcal{P}\big(Y(\Omega)\big)$, les événements $(X \in A)$ et $(Y \in B)$ sont indépendants, autrement dit si

$$P\big((X,Y) \in A \times B\big) = P(X \in A) \cdot P(Y \in B) \ ,$$

soit encore

$$P(X \in A \ , \ Y \in B) = P(X \in A) \cdot P(Y \in B) \ .$$

Proposition. Les variables aléatoires X et Y sont indépendantes **si et seulement si** la distribution de probabilités du couple (X, Y) est donnée par

$$\forall (x,y) \in X(\Omega) \times Y(\Omega) \qquad P\big((X,Y) = (x,y)\big) = P(X = x)\, P(Y = y) \ .$$

Proposition. Si X et Y sont deux variables aléatoires indépendantes, si f et g sont des applications définies respectivement sur $X(\Omega)$ et $Y(\Omega)$, alors les variables $f(X)$ et $g(Y)$ sont indépendantes. Autrement dit, toute fonction de X est indépendante de toute fonction de Y. Schématiquement, $X \perp\!\!\!\perp Y \Longrightarrow f(X) \perp\!\!\!\perp g(Y)$.

Extension aux n-uplets de variables aléatoires. Des v.a. $X_1, \cdots, X_n$ sur (Ω, P) sont dites **(mutuellement) indépendantes** si, quel que soit $(A_1, \cdots, A_n) \in \prod_{i=1}^{n} \mathcal{P}\big(X_i(\Omega)\big)$, les événements $(X_i \in A_i)$, $1 \leq i \leq n$, sont (mutuellement) indépendants.

Proposition. Les v.a. $X_1, \cdots, X_n$ sont mutuellement indépendantes **si et seulement si**, quel que soit $(A_1, \cdots, A_n) \in \prod_{i=1}^{n} \mathcal{P}\big(X_i(\Omega)\big)$, on a $P\Big(\bigcap_{i=1}^{n} \{X_i \in A_i\}\Big) = \prod_{i=1}^{n} P(X_i \in A_i)$.

Proposition. Les variables aléatoires $X_1, \cdots, X_n$ sont indépendantes *si et seulement si*, pour tout n-uplet $(x_1, \cdots, x_n) \in \prod_{i=1}^{n} X_i(\Omega)$, on a $P\Big(\bigcap_{i=1}^{n} \{X_i = x_i\}\Big) = \prod_{i=1}^{n} P(X_i = x_i)$.

Remarque. Il est immédiat que l'indépendance (mutuelle) des v.a. $X_1, \cdots, X_n$ entraîne leur indépendance deux à deux, mais la réciproque est fausse.

On pourra modéliser la succession de n expériences aléatoires indépendantes par une suite finie $(X_1, \cdots, X_n)$ de variables aléatoires indépendantes. On rencontrera notamment la situation dite **schéma de Bernoulli**, dans laquelle on répéte indépendamment n expériences aléatoires ayant deux issues ("succès" avec probabilité p, "échec" avec probabilité $q = 1-p$). On associe à la i-ème expérience une variable aléatoire X_i qui prend la valeur 1 en cas de succès, la valeur 0 en cas d'échec. Chacune de ces v.a. X_i suit la loi de Bernoulli $\mathcal{B}(p)$, et elles sont mutuellement indépendantes. Leur somme $S_n = X_1 + \cdots + X_n$, qui comptabilise le nombre de succès, est alors une variable binomiale de paramètres n et p, i.e. $S_n \sim \mathcal{B}(n, p)$. On résume donc:

Proposition. Si $X_1, \cdots, X_n$ sont indépendantes de loi $\mathcal{B}(p)$, alors leur somme $X_1 + \cdots + X_n$ suit la loi binomiale $\mathcal{B}(n, p)$.

Lemme des coalitions. Si les variables aléatoires $X_1, \cdots, X_n$ sur (Ω, P) sont indépendantes, si $m \in [\![1, n-1]\!]$, si f et g sont deux fonctions définies sur $X_1(\Omega) \times \cdots \times X_m(\Omega)$ et sur $X_{m+1}(\Omega) \times \cdots \times X_n(\Omega)$ respectivement, alors les variables $Y = f(X_1, \cdots, X_m)$ et $Z = g(X_{m+1}, \cdots, X_n)$ sont indépendantes.

Remarque. On peut étendre ce résultat au cas de plus de deux "coalitions", c'est-à-dire regroupements. Par exemple, si X_1, X_2, X_3, X_4, X_5, X_6 sont six variables indépendantes sur (Ω, P), alors les trois variables $U = f(X_1, X_2)$, $V = g(X_3, X_4)$ et $W = h(X_5, X_6)$ sont elles aussi indépendantes.

III. Espérance et variance.

a. Espérance

Dans tout ce paragraphe, les variables aléatoires considérées seront à valeurs réelles ou complexes.

Définition. Soit X une variable aléatoire réelle ou complexe sur un espace probabilisé fini (Ω, P). On définit son **espérance** $\mathrm{E}(X)$ comme étant le réel

$$\mathrm{E}(X) = \sum_{x \in X(\Omega)} x \cdot P_X(x) = \sum_{x \in X(\Omega)} x \cdot P(X = x)\ .$$

L'espérance $\mathrm{E}(X)$ peut être considérée comme une **moyenne pondérée**: c'est la moyenne des valeurs prises par X, pondérées par leurs probabilités d'obtention respectives. L'espérance est un **indicateur de position**, i.e. elle donne la "position moyenne" de la valeur de X sur la droite réelle (ou dans le plan complexe). On dit que la variable X est **centrée** si $\mathrm{E}(X) = 0$.

Proposition. En réordonnant la somme, on obtient la relation

$$\mathrm{E}(X) = \sum_{\omega \in \Omega} X(\omega)\, P\big(\{\omega\}\big) \ .$$

Exemples. Une variable aléatoire X constante de valeur C a pour espérance $\mathrm{E}(X) = C$.

Si X est une variable de Bernoulli de paramètre p, alors $\mathrm{E}(X) = p$. En particulier, si A est un événement (une partie de Ω), alors $\mathrm{E}(\mathbb{1}_A) = P(A)$.

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

Propriétés. (*on notera que l'on retrouve les propriétés d'une intégrale.*)

- **Linéarité.** Si X et Y sont deux v.a. sur (Ω, P), si a et b sont deux scalaires, alors

$$\mathrm{E}(aX + bY) = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y) \ .$$

- **Positivité.** Si X est une v.a. positive (i.e. à valeurs dans $\mathbb{R}_+$), alors $\mathrm{E}(X) \geq 0$.
- **Croissance.** Si X et Y sont deux v.a. réelles telles que $X \leq Y$, alors $\mathrm{E}(X) \leq \mathrm{E}(Y)$.
- **Inégalité triangulaire.** On a toujours $\big|\mathrm{E}(X)\big| \leq \mathrm{E}\big(|X|\big)$.

Formule de transfert. Si X est une v.a. réelle ou complexe sur (Ω, P), et si f est une fonction à valeurs réelles ou complexes définie sur $X(\Omega)$, alors $Y = f(X)$ est une v.a. sur (Ω, P) et

$$\mathrm{E}\big(f(X)\big) \ = \sum_{y \in Y(\Omega)} y \cdot P(Y = y) \ = \sum_{x \in X(\Omega)} f(x) \cdot P(X = x) \ .$$

Extension. Cette formule de transfert s'applique aussi aux couples et aux n-uplets. Ainsi,

- si X et Y sont deux v.a. à valeurs dans $\mathbb{K}$, et si $f : X(\Omega) \times Y(\Omega) \to \mathbb{K}$ est une application, alors $\mathrm{E}\big(f(X,Y)\big) = \displaystyle\sum_{(x,y) \in X(\Omega) \times Y(\Omega)} f(x,y) \cdot P(X = x, Y = y)$.
- si $X_1, \cdots, X_n$ sont des v.a. à valeurs dans $\mathbb{K}$, et si $f : X_1(\Omega) \times \cdots \times X_n(\Omega) \to \mathbb{K}$ est une application, alors

$$\mathrm{E}\big(f(X_1, \cdots, X_n)\big) = \sum_{(x_1, \cdots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)} f(x_1, \cdots, x_n) \cdot P(X_1 = x_1, \cdots, X_n = x_n) \ .$$

Théorème. Si X et Y sont deux v.a. indépendantes, alors $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.
Mais la réciproque est fausse (cette relation ne caractérise pas les v.a. indépendantes).

Extension. Si $X_1, \cdots, X_n$ sont n variables aléatoires mutuellement indépendantes sur (Ω, P), alors $\mathrm{E}\Big(\displaystyle\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathrm{E}(X_i)$.

Exercice. Un dé est pipé de telle sorte que la probabilité de sortie de la face k est proportionnelle à k, $1 \leq k \leq 6$. On note X le résultat obtenu lors d'un lancer. Calculer l'espérance de la variable aléatoire $Y = \dfrac{1}{X}$.

Exercice. Soit $n \in \mathbb{N}^*$. On choisit au hasard deux entiers p et q entre 0 et n (choix indépendants, avec équiprobabilité). On pose $X(p,q) = |p-q|$ et $Y(p,q) = \min\{p,q\}$. Déterminer les lois des variables aléatoires X et Y. Calculer les espérances $\mathrm{E}(X)$ et $\mathrm{E}(Y)$ puis, avec $n = 10$ par exemple, écrire un programme Python permettant de vérifier expérimentalement ces valeurs.

b. Variance, écart-type et covariance.

Définition. Soit X une variable aléatoire réelle (v.a.r.) sur un espace probabilisé fini (Ω, P). Sa **variance** est le nombre

$$\mathrm{V}(X) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big) = \sum_{x \in X(\Omega)} \big(x - \mathrm{E}(X)\big)^2 \, P(X = x) \, .$$

On a bien sûr $\mathrm{V}(X) \geq 0$.

L'**écart-type** de X est la racine carrée de la variance: $\sigma(X) = \sqrt{\mathrm{V}(X)}$.

La variance de X est donc la moyenne (pondérée) des carrés des écarts à la moyenne (espérance). C'est un **indicateur de dispersion** de la variable aléatoire X.

On dit qu'une v.a.r. est **réduite** lorsque $\mathrm{V}(X) = 1$, i.e. lorsque $\sigma(X) = 1$. Si X est une v.a.r. de variance non nulle, alors la variable $Y = \dfrac{X - \mathrm{E}(X)}{\sigma(X)}$ est centrée réduite.

Formule de Koenig-Huygens. On a la relation

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

Propriétés. Si X est une v.a.r., si a et b sont deux réels, on a $\mathrm{V}(aX + b) = a^2 \, \mathrm{V}(X)$.

Intuitivement, l'ajout d'une constante b décale la moyenne, i.e. l'espérance, mais ne modifie en rien la dispersion autour de cette moyenne, alors que la multiplication par le facteur a multiplie par a l'espérance, mais multiplie aussi la dispersion par un facteur $|a|$, donc la variance par a^2.

Lois usuelles. Si X soit la loi de Bernoulli $\mathcal{B}(p)$, alors $\mathrm{V}(X) = p(1-p)$.

Si X suit la loi binomiale $\mathcal{B}(n,p)$, alors $\mathrm{V}(X) = np(1-p)$.

Proposition. Si X et Y sont deux v.a.r. indépendantes, on a $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$.

En effet, on a $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$, donc en utilisant la linéarité de l'espérance:

$$\begin{aligned}
\mathrm{V}(X+Y) &= \mathrm{E}\big[(X+Y)^2\big] - \mathrm{E}(X+Y)^2 \\
&= \mathrm{E}(X^2 + 2XY + Y^2) - \Big(\mathrm{E}(X) + \mathrm{E}(Y)\Big)^2 \\
&= \mathrm{E}(X^2) + 2\,\mathrm{E}(XY) + \mathrm{E}(Y^2) - \mathrm{E}(X)^2 - 2\,\mathrm{E}(X)\,\mathrm{E}(Y) - \mathrm{E}(Y)^2 \\
&= \mathrm{V}(X) + \mathrm{V}(Y) \, .
\end{aligned}$$

Définition. Soient X et Y deux variables aléatoires réelles (v.a.r.) sur (Ω, P). On définit leur **covariance** par

$$\mathrm{Cov}(X,Y) = \mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)\,\big(Y - \mathrm{E}(Y)\big)\Big) \; .$$

Propriété. La covariance est une forme bilinéaire symétrique positive sur le $\mathbb{R}$-espace vectoriel des variables aléatoires réelles sur (Ω, P).

On a donc

$$\begin{aligned} \mathrm{Cov}(Y,X) &= \mathrm{Cov}(X,Y) && (\textit{symétrie}) \; , \\ \mathrm{Cov}(aX_1 + X_2, Y) &= a\,\mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y) && (\textit{linéarité à gauche}) \; , \\ \mathrm{Cov}(X, aY_1 + Y_2) &= a\,\mathrm{Cov}(X,Y_1) + \mathrm{Cov}(X,Y_2) && (\textit{linéarité à droite}) \; , \\ \mathrm{Cov}(X,X) &= \mathrm{V}(X) \geq 0 \; . && (\textit{posivitié}) \; . \end{aligned}$$

Définition. Deux v.a.r. dont la covariance est nulle sont dites **décorrélées**.

Formule de Koenig-Huygens généralisée. On a la relation

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

Conséquence. Si deux v.a.r. sont indépendantes, alors elles sont décorrélées. Autrement dit,

$$X \perp\!\!\!\perp Y \Longrightarrow \mathrm{Cov}(X,Y) = 0 \; .$$

La réciproque est fausse.

Proposition. Si $X_1, \cdots, X_n$ sont des variables aléatoires réelles, on a

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

C'est une conséquence immédiate de la bilinéarité de la covariance.

En particulier, si les v.a.r. X_i sont décorrélées (deux à deux), alors $\mathrm{V}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{V}(X_i)$.

Conséquence. Si les X_i sont deux à deux indépendantes, alors $\mathrm{V}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{V}(X_i)$.

On retrouve ainsi la variance d'une variable binomiale car, si $S_n = \sum_{i=1}^{n} X_i$ avec $X_i \sim \mathcal{B}(p)$ pour tout i, les X_i étant **mutuellement** indépendantes, on sait que $S_n \sim \mathcal{B}(n,p)$, puis

$$\mathrm{V}(S_n) = \sum_{i=1}^{n} \mathrm{V}(X_i) = np(1-p) \; .$$

c. Inégalités probabilistes.

Inégalité de Markov. Si X est une variable aléatoire **positive** sur (Ω, P), alors

$$\forall a \in \mathbb{R}_+^* \qquad P(X \geq a) \leq \frac{\mathrm{E}(X)}{a} \; .$$

Inégalité de Bienaymé-Tchebychev. Si X est une variable aléatoire réelle définie sur un espace probabilisé fini (Ω, P), on a l'inégalité

$$P\Big(\big|X - \mathrm{E}(X)\big| \geq \varepsilon\Big) \leq \frac{\mathrm{V}(X)}{\varepsilon^2}\ .$$

Si, par exemple, X représente une grandeur scalaire attachée à une population d'individus (la taille d'une population de bipèdes, la note obtenue à un DS de math par les mêmes êtres bipèdes), cette inégalité donne une majoration de la probabilité d'obtenir un résultat éloigné de la moyenne.

On peut utiliser ces deux résultats de cours pour obtenir des **inégalités de concentration**, i.e. des inégalités majorant la probabilité qu'une variable aléatoire dévie d'une certaine valeur (souvent son espérance).

Exemple. Si X est une v.a. réelle sur (Ω, P), en appliquant l'inégalité de Markov aux variables positives e^{aX} et e^{-aX}, on obtient les **inégalités de Chernoff**:

$$\forall t \in \mathbb{R} \qquad P(X \geq t) \leq e^{-at}\,\mathrm{E}(e^{aX}) \qquad \text{et} \qquad P(X \leq t) \leq e^{at}\,\mathrm{E}(e^{-aX})\ .$$

On peut aussi aller vers la **loi faible des grands nombres**: si $X_1, \cdots, X_n$ sont des variables réelles indépendantes de même loi sur (Ω, P), si on considère leur moyenne $M_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$, on peut appliquer l'inégalité de Bienaymé-Tchebychev à la variable M_n. Si on pose $m = \mathrm{E}(X_1)$ et $\sigma = \sigma(X_1)$, on a $\mathrm{E}(M_n) = \dfrac{1}{n}\,nm = m$ par linéarité de l'espérance, et par l'indépendance,

$$\mathrm{V}(M_n) = \frac{1}{n^2}\,\mathrm{V}\Big(\sum_{k=1}^{n} X_k\Big) = \frac{1}{n^2}\,\sum_{k=1}^{n}\mathrm{V}(X_k) = \frac{1}{n^2}\,n\sigma^2 = \frac{\sigma^2}{n}\ .$$

Donc, pour tout $\varepsilon > 0$, on a $P\big(|M_n - m| \geq \varepsilon\big) \leq \dfrac{\mathrm{V}(M_n)}{\varepsilon^2} = \dfrac{\sigma^2}{n\varepsilon^2}$ **(*)**.
En particulier, cette probabilité tend vers 0 lorsque n tend vers l'infini.

On peut donner une **interprétation fréquentiste** de cela: on répète n fois, de façon indépendante, une même expérience aléatoire se traduisant par la mesure d'une grandeur (par exemple le nombre de "Pile" obtenus lors de 100 lancers d'une pièce, ou le nombre de points fixes d'une permutation de 100 nombres générée aléatoirement), et note $X_1, \cdots, X_n$ les résultats obtenus. Les variables X_k ont la même loi donc la même espérance m que l'on appellera **moyenne théorique** de l'expérience. La valeur de M_n correspond à la **moyenne observée** lors de cette réalisation de l'expérience. L'inégalité **(*)** majore la probabilité que la moyenne observée s'écarte de la moyenne théorique de plus que la valeur ε préalablement choisie. Pour expliquer le terme "fréquentiste", on peut aussi s'intéresser à la fréquence d'apparition d'un 6 lors d'une suite de lancers de dés. Si le dé n'est pas pipé, la fréquence théorique d'apparition est $\dfrac{1}{6}$, alors que la fréquence observée est le nombre de "succès" divisé par le nombre de lancers. Dans ce dernier cas, X_k serait la variable indicatrice de l'événement: "le k-ième lancer donne 6".