Introduction aux probabilités

Benjamin Charlier

10 février 2016

Table des matières

Chapitre 1

Espace probabilisé

1.1 Introduction

La théorie des probabilités utilise le vocabulaire ensembliste pour modéliser le résultat d'expériences aléatoires (*i.e.* soumis au **hasard**).

Remarque 1. Le mot *aléatoire* vient du latin *alea* qui signifie dé, la chance (`http://www.cnrtl.fr/etymologie/al%C3%A9atoire`). Le mot *hasard* vient de l'arable *al-zahr* qui signifie dé. À rapprocher aussi de *hazard* risque, danger (`http://www.cnrtl.fr/etymologie/hasard`).

But : prédire les fréquences d'apparition d'événements et non le résultat précis d'une expérience.

1.2 Ensembles et opérations sur les ensembles

1.2.1 Définitions

Un ensemble Ω est constitué de points ω tous distincts. L'expression logique

$$\omega \in \Omega$$

signifie que ω appartient à Ω. On dit qu'un ensemble A est inclu dans Ω si

$$\omega \in A \Rightarrow \omega \in \Omega.$$

On note $A \subset \Omega$ et on dit que A est un **sous-ensemble** de Ω.

> **Définition 1.2.1.** On note $\emptyset$ l'ensemble vide (qui ne contient aucun élément).

1.2.2 Opérations

> **Définition 1.2.2.** Soit $A, B, C \subset \Omega$:
> 1. ***A* inter *B*** est noté $A \cap B = \{\omega \in \Omega | \omega \in A \text{ et } \omega \in B\}$.
> 2. ***A* union *B*** est noté $A \cup B = \{\omega \in \Omega | \omega \in A \text{ ou } \omega \in B\}$.

3. **$\boldsymbol{A}$ complémentaire** est noté $A^c = \{\omega \in \Omega | \omega \notin A\}$.

On retrouve aussi d'autres notations :

1. **$\boldsymbol{A}$ privé de $\boldsymbol{B}$** noté $A \setminus B = A \cap B^c$.

2. **différence symétrique entre $\boldsymbol{A}$ et $\boldsymbol{B}$** noté $A \Delta B = (A \cap B^c) \cup (A^c \cap B) = (A \cup B) \setminus (A \cap B)$.

Proposition 1.2.1. Les opérations $\cup$ et $\cap$ sont :

1. commutatives : $A \cup B = B \cup A$ et $A \cap B = B \cap A$
2. associatives : $(A \cup B) \cup C = A \cup (B \cup C)$ et $(A \cap B) \cap C = A \cap (B \cap C)$
3. élément neutre :$A \cup \emptyset = A$ et $A \cap \Omega = A$
4. distributives : $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ et $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

On peut alors en déduire des règles de calculs suivantes :

Proposition 1.2.2. 1. **Loi de De Morgan :** $(A \cap B)^c = A^c \cup B^c$, $(A \cup B)^c = A^c \cap B^c$

2. **Partition** $A \cup A^c = \Omega$ et $A \cap A^c = \emptyset$
3. $A \cup \Omega = \Omega$ et $A \cap \emptyset = \emptyset$
4. $A \cup A = A$ et $A \cap A = A$

Démonstration.

□

1.2.3 Fonctions indicatrices

Définition 1.2.3. Soit $A \subset \Omega$. La **fonction indicatrice** de A est définie par

$$\begin{aligned} \mathbb{1}_A : \Omega &\to \{0,1\} \\ \omega &\mapsto \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{sinon} \end{cases} \end{aligned}$$

Proposition 1.2.3. Soit $A, B \subset \Omega$ on a

$$\begin{aligned} \mathbb{1}_{A\cup B} &= \max\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A\mathbb{1}_B \\ \mathbb{1}_{A\cap B} &= \min\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_A\mathbb{1}_B \\ \mathbb{1}_{A^c} &= 1 - \mathbb{1}_A \end{aligned}$$

Démonstration.

$\square$

Définition 1.2.4. L'ensemble des parties de Ω est noté $\mathcal{P}(\Omega)$.

Exemple 1.2.1. Si $\Omega = \{1, 2\}$ alors $\mathcal{P}(\Omega) = \{\{1\}, \{2\}, \{1, 2\}, \emptyset\}$.

1.2.4 Cardinalité

Définition 1.2.5. Le nombre d'éléments d'un ensemble Ω est le **cardinal** de Ω et est noté $\text{Card}(\Omega)$ ou $|\Omega|$.

1. Lorsque $\text{Card}(\Omega) < +\infty$ on parle d'**ensemble fini**.
2. Lorsque $\text{Card}(\Omega) = +\infty$ on distingue les cas suivants
   (a) si Ω peut être mis en bijection avec $\mathbb{N}$, on parle d'**ensemble dénombrable**.
   (b) dans le cas contraire on parle d'**ensemble non-dénombrable**.

Exemple 1.2.2.

1.2.5 Dénombrement

Dans cette section, les ensembles considérés sont finis.

Principe de bijection

Lorsque l'on veut compter les éléments d'un ensemble, on montre que cet ensemble est en bijection avec un ensemble dont on connaît le cardinal. Le reste de cette section énonce un certains nombre de résultats qu'il faut connaître.

Principe d'indépendance

Soit A et B deux ensembles finis, on a $A \times B = \{(a, b) | a \in A, b \in B\}$. On a alors

$$|A \times B| = |A| \cdot [B|$$

Principe de partition

On dit que les ensembles $(A_i)_{i\in I}$ forme une partition de A si $A = \bigcup_{i\in I} A_i$ et $i \neq j \Rightarrow A_i \cap A_j = \emptyset$. On a alors

$$|A| = \sum_{i\in I} |A_i|$$

Lemme des bergers

Ce résultat généralise la procédure de comptage suivante : Quand un berger veut compter ses moutons, il compte le nombre de pattes puis divise par quatre.

Proposition 1.2.4 (Lemme des bergers). Soit $\phi : D \to A$ une application surjective. On suppose qu'il existe un entier $a \geq 1$ tel que pour tout $y \in A$

$$|\{x \in D | \phi(x) = y\}| = a$$

C'est à dire si tout $y \in A$ a exactement a antécédent dans D. Alors on a

$$|A| = \frac{|D|}{a}$$

Démonstration. On applique le principe de partition à l'ensemble D : les ensembles $(D_y)_{y\in A}$ où $D_y = \{x \in D | \phi(x) = y\}$ forment une partition de D. D'où

$$|D| = \sum_{y\in A} |D_y| = \sum_{y\in A} a = a|A|. \qquad \square$$

Quelques résultats incontournables

Nombre d'applications de D dans A Il existe exactement $|A|^{|D|}$ applications de D dans A, ce qui peut s'écrire

$$|A^D| = |A|^{|D|}$$

Remarque 2. Posons $|D| = p$ et $|A| = n$. Un cas particulier important est celui où $D = \{1, \cdots, p\}$. Or un p-uplet $(x_1, \cdots, x_p)$ d'éléments de A peut être vu comme le graphe d'une application de $\{1, \cdots, p\} \to A$. Le nombre de p-uplets $(x_1, \cdots, x_p)$ dont les composantes sont des éléments de A est n^p (on retrouve le principe d'indépendance et le nombre de tirage avec remise).

Exemple 1.2.3. Un professeur note chaque étudiant d'une promotion de 300 étudiants avec une note entière entre 0 et 20. Combien de résultats sont possibles ?

Nombre de permutations Soit Ω un ensemble fini avec $|\Omega| = n$. Le nombre de permutations de Ω est

$$n! = n(n-1)\cdots 1$$

Remarque 3. Une permutation de Ω est une bijection de Ω dans lui même.

Exemple 1.2.4. Quel est ne nombre de manières de mélanger un jeu de 32 cartes ?

Nombre d'injection de D dans A On pose $|D| = p$ et $|A| = n$. Il existe une injection de D vers A si et seulement si $p \leq n$. Dans ce cas le nombre d'injection de D dans A est

$$A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}.$$

Remarque 4. Lorsque $n = p$ on a $A_n^n = n!$. En fait un injection entre deux ensembles de même cardinal est une bijection.

Exemple 1.2.5. On a 3500 candidats à un concours. Seules 300 places sont disponibles. Combien de palmarès possible y a-t-il ? (On suppose qu'il n'y a pas d'*ex æquo*)

Nombre de parties a p éléments dans Ω On pose $|\Omega| = n$. Le nombre de parties à p éléments dans un ensemble à n éléments est

$$\binom{n}{p} = \frac{n!}{(n-p)!p!}.$$

Démonstration. On applique le lemme des bergers à

— D l'ensemble des injections de $\{1, \cdots, p\}$ dans Ω

— A est l'ensemble des parties à p éléments dans Ω

— $\phi : D \to A$ définie par $\phi(f) = \{f(k); k = 1, \cdots, p\}$

On a vu que $|D| = \frac{n!}{(n-p)!}$. Il n'est pas difficile de voir que ϕ est surjective. De plus, étant donnée une partie $\{x_1, \cdots, x_p\}$ de Ω, le nombre d'injections $f : \{1, \cdots, p\} \to \Omega$ telles que $\{f(1), \cdots, f(p)\} = \{x_1, \cdots, x_p\}$ est $p!$ (le nombre de permutation de $\{1, \cdots, p\}$). On applique le lemme des bergers avec $a = p!$. $\square$

Exemple 1.2.6. On a 3500 candidats à un concours. Seules 300 places sont disponibles. Combien de listes alphabétiques de reçus possible y a-t-il ? (On suppose que tous les noms sont différents)

Nombre totale de partie Si Ω est de cardinal fini avec $|\Omega| = n$ alors $\mathrm{Card}(\mathcal{P}(\Omega)) = 2^n$.

Démonstration.

□

Exemple 1.2.7. On a 3500 candidats à un examen.Combien de listes alphabétiques de reçus possible y a-t-il ? (On suppose que tous les noms sont différents)

1.3 Évènement et probabilités

1.3.1 Expérience aléatoire et événements

Une **expérience aléatoire** est une expérience renouvelable (au moins en théorie) dont on ne peut prédire l'issue de manière précise (renouvelée dans des conditions identiques, ne donne pas à chaque essai le même résultat). On connaît cependant l'ensemble des résultats possibles *a priori*.

Exemple 1.3.1. Voici quelques exemples d'expériences aléatoires :

1. lancer deux dés ;

2. distribuer les cartes au tarot, c'est-à-dire répartir le paquet de 78 cartes entre les 3, 4 ou 5 joueurs ;
3. observer la formation d'un caractère génétique d'un individu à partir des caractères correspondants de ses parents ;
4. observer la désintégration d'un noyau atomique radioactif ;
5. attendre le tram à la station Universités, à partir de 18h.

Définition 1.3.1. Étant donnée une expérience aléatoire, l'**univers** noté Ω est l'ensemble des résultats possibles (ou **événement élémentaire**). Un **événement** est un sous-ensemble de Ω (réunion d'événements élémentaires).

Exemple 1.3.2. Expérience aléatoire : jet d'un dé à 6 faces.

On a le "dictionnaire" suivant entre les vocabulaires ensemblistes et probabilistes :

Terminologie probabiliste	Terminologie ensembliste
univers, événement certain	Ω
événement impossible	$\emptyset$
résultat possible, événement élémentaire	$\{\omega\}$ où ω élément de Ω
événement	A, sous-ensemble de Ω
A est réalisé	$\omega \in A$ ou $\{\omega\} \subset A$
A implique B	$A \subset B$
A ou B	$A \cup B$
A et B	$A \cap B$
contraire de A	A^c, complémentaire de A
A et B sont incompatibles	$A \cap B = \emptyset$

1.3.2 Variable aléatoire

Quand on étudie un phénomène aléatoire, on est amené à étudier des grandeurs numériques (ou vectorielles) liées à celui-ci. En termes vagues, une variable aléatoire est un nombre ou un vecteur (une variable) dont la valeur dépend de l'issue ω d'une expérience aléatoire. Mathématiquement, il s'agit donc d'une fonction sur l'ensemble Ω. Si cette fonction est à valeurs dans $\mathbb{R}$, on parle de **variable aléatoire réelle**.

Exemple 1.3.3. Expérience aléatoire : jet d'un dé à 6 faces.

1.4 Tribus et probabilités

Définition 1.4.1. Une famille $\mathcal{A}$ de sous-ensembles d'un ensemble Ω est une **tribu** ou **σ-algèbre** sur Ω si elle satisfait aux trois axiomes suivants :
(i) $\Omega \in \mathcal{A}$.
(ii) Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.
(iii) Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, on a $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Exemple 1.4.1.

Définition 1.4.2. On appelle **espace probabilisable** un couple $(\Omega, \mathcal{A})$ où Ω est un ensemble et $\mathcal{A}$ une tribu sur l'ensemble Ω. Quand un espace probabilisable est fixé, on dit que $\mathcal{A}$ est **la tribu des événements**.

Exemple 1.4.2. Jet d'un dé : $\Omega = \Big\{\{\text{⚀}\}, \cdots, \{\text{⚅}\}\Big\}$. On a $\mathcal{A} = \mathcal{P}(\Omega)$.

La modélisation d'un phénomène aléatoire et d'une famille d'événements commence par le choix d'un espace probabilisable qui rend compte de l'ensemble des réalisations envisagées et des événements qui peuvent être sujets de l'étude.

Définition 1.4.3. Soit $(\Omega, \mathcal{A})$ un espace probabilisable. Une **probabilité** sur cet espace est une application $\mathbb{P} : \mathcal{A} \to [0, 1]$ satisfaisant :
(i) $\mathbb{P}(\Omega) = 1$.
(ii) Propriété de **σ-additivité** : pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments **disjoints** deux à deux de $\mathcal{A}$,

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

Exemple 1.4.3. Jet d'un dé : $\mathbb{P}(\{\text{face du dessus ⚀}\}) = \cdots = \mathbb{P}(\{\text{face du dessus ⚅}\}) = \frac{1}{6}$.

Proposition 1.4.1. *(i)* Pour tout $A \in \mathcal{A}$, on a : $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. En particulier : $\mathbb{P}(\emptyset) = 0$.

(ii) Si $A, B \in \mathcal{A}$ et si $A \subseteq B$, on a

$$\mathbb{P}(A) \leq \mathbb{P}(B) \quad \text{et} \quad \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A).$$

(iii) Pour toute suite croissante $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, c'est-à-dire tels que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$, on a :

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(iv) Pour toute suite décroissante $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, c'est-à-dire tels que $A_{n+1} \subset A_n$ pour tout $n \in \mathbb{N}$, on a :

$$\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Démonstration. On montre uniquement *(i)* et *(ii)*

□

Proposition 1.4.2. Pour toute suite finie $(A_i)_{1\leq i\leq n}$ d'éléments disjoints deux à deux de $\mathcal{A}$,

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Démonstration. Découle directement de la σ-additivité. □

Remarque 5. Cette formule est particulièrement utile pour le calcul des probabilités. Dans le cas $n = 2$ on a pour $A, B \subset \Omega$: $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$.

Proposition 1.4.3 (Formule de Poincaré ou du crible). Si $n \geq 2$, pour toute suite finie $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{A}$, on a :

$$\begin{aligned}\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = &\sum_{i=1}^{n} \mathbb{P}(A_i)\\ &- \sum_{(i,j)|1\leq i<j\leq n} \mathbb{P}(A_i \cap A_j)\\ &+ \sum_{(i,j,k)|1\leq i<j<k\leq n} \mathbb{P}(A_i \cap A_j \cap A_k)\\ &- \cdots\\ &+ (-1)^{n-1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).\end{aligned}$$

Démonstration. On se contente de montrer et comprendre la formule sur le cas $n = 2$ et $n = 3$. La formule générale se montre par récurrence.

$\square$

1.5 Espaces probabilisés discrets

1.5.1 Définition

Définition 1.5.1. On appelle **espace probabilisable discret** un espace probabilisable $(\Omega, \mathcal{A})$ où Ω est dénombrable et où $\mathcal{A} = \mathcal{P}(\Omega)$, l'ensemble des parties de Ω.

Une probabilité $\mathbb{P}$ sur un espace probabilisable discret $\big(\Omega, \mathcal{P}(\Omega)\big)$ est définie par les probabilités des réalisations élémentaires $\{\omega\} \subset \Omega$. On a :

$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\})$$

pour tout sous-ensemble A de Ω car les réalisations ω sont incompatibles deux à deux.

1.5.2 Probabilité uniforme sur un ensemble fini

La probabilité uniforme sur un ensemble fini Ω attribue la même probabilité à chaque réalisation élémentaire $\{\omega\}$:

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|},$$

où $|\Omega|$ désigne le cardinal (nombre d'éléments) de Ω. Par conséquent, on a :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \tag{1.1}$$

pour tout événement $A \subset \Omega$.

Remarque 6. Lorsque l'on parle de choix "au hasard" sur un ensemble fini, on sous-entend que ce choix est fait au moyen de la probabilité uniforme, c'est-à-dire en donnant à chaque élément de l'ensemble les mêmes chances d'être choisi. On résume souvent la formule (1.1) par

$$\frac{\text{nombre de cas favorables}}{\text{nombre total de cas}}.$$

Cela suggère que le calcul des probabilité revient à faire des dénombrements. On s'aperçoit également qu'une probabilité uniforme sur un ensemble non fini n'est pas définie.

Exemple 1.5.1. On lance n fois un dé à 6 faces équilibré et on cherche la probabilité d'obtenir k fois ($k \leq n$) la face ⚅.

Chapitre 2

Indépendance et conditionnement

2.1 Indépendance d'événements et de variables aléatoires

2.1.1 Événements indépendants

Définition 2.1.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

(i) Deux événements $A \in \mathcal{A}$ et $B \in \mathcal{A}$ sont **indépendants** si :

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

(ii) Soit $(A_i)_{i\in I}$ une famille quelconque d'événements. Ces événements sont dits **indépendants** (dans leur ensemble) si, pour toute partie non vide J de I, on a :

$$\mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbb{P}(A_j).$$

Remarque 7.
1. L'indépendance des événements $A_1, A_2, \cdots, A_n$ représente $2^n - n - 1$ conditions (nombre de parties de $I = \{1, \cdots, n\}$ de cardinal ≥ 2).
2. Si n événements sont indépendants dans leur ensemble, ils sont indépendants deux à deux. La réciproque est fausse en générale.

Exemple 2.1.1. On lance deux dés et on pose
— A : le résultat du lancer du dé numéro 1 est pair,
— B : le résultat du lancer du dé numéro 2 est pair,
— C : la somme des résultats des 2 lancers est impaire.

Dire que A et B sont indépendants signifie que la réalisation ou la non-réalisation de B n'influe pas sur la probabilité de voir A se réaliser.

Proposition 2.1.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Si A et B sont deux événements indépendants, les événements A et B^c d'une part, A^c et B d'autre part, A^c et B^c enfin, sont indépendants.

Démonstration.

$\square$

2.1.2 Variables aléatoires indépendantes

Généralités

Définition 2.1.2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_1, \mathcal{E}_1)$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_2, \mathcal{E}_2)$ deux variables aléatoires. Les variables aléatoires X et Y sont **indépendantes** si pour tous ensembles $A_1 \in \mathcal{E}_1$ et $A_2 \in \mathcal{E}_2$, les événements $\{X \in A_1\}$ et $\{Y \in A_2\}$ sont indépendants.

Proposition 2.1.2. Avec les notations de la définition précédente. Les variables aléatoires X et Y sont indépendantes si et seulement si on a

$$\mathbb{P}\Big(\{X \in A_1\} \cap \{Y \in A_2\}\Big) = \mathbb{P}(\{X \in A_1\})\mathbb{P}(\{Y \in A_2\}).$$

pour tous ensembles $A_1 \in \mathcal{E}_1$ et $A_2 \in \mathcal{E}_2$.

Fonctions de variables aléatoires

On note $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $(E_i, \mathcal{E}_i)$ ou $(F_i, \mathcal{F}_i)$ désignent des espaces probabilisables.
Soit une variable aléatoire $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$ et une fonction $f : (E, \mathcal{E}) \to (F, \mathcal{F})$ mesurable (c'est à dire telle que pour tout $A \in \mathcal{F}$ on ait $f^{-1}(A) \in \mathcal{E}$ où $f^{-1}(A) \subset F$ est l'image réciproque de A par f). Alors $f(X) = f \circ X$ est une variable aléatoire de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(F, \mathcal{F})$.

Proposition 2.1.3. Soient $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_1, \mathcal{E}_1)$ et $Y : (\Omega, \mathcal{A}, \mathbb{P}) \to (E_2, \mathcal{E}_2)$ deux variables aléatoires indépendantes. Soient deux fonctions $f : (E_1, \mathcal{E}_1) \to (F_1, \mathcal{F}_1)$ et $g : (E_2, \mathcal{E}_2) \to (F_2, \mathcal{F}_2)$. Alors les variables aléatoires $f_1(X)$ et $f_2(Y)$ sont indépendantes.

Exemple 2.1.2.

2.2 Probabilités conditionnelles

Dans la suite, $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé quelconque.

2.2.1 Définitions. Formule des probabilités totales

Définition 2.2.1. Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. Soit $A \in \mathcal{A}$ un événement ; on appelle **probabilité de A conditionnée par** B(ou **sachant B**) le nombre réel, noté $\mathbb{P}(A|B)$ ou $\mathbb{P}^B(A)$, défini par :

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

Proposition 2.2.1. Soit $B \in \mathcal{A}$ un événement tel que $\mathbb{P}(B) > 0$. L'application $\mathbb{P}^B$ de $\mathcal{A}$ dans $\mathbb{R}^+$ qui à tout $A \in \mathcal{A}$ fait correspondre $\mathbb{P}^B(A)$ est une probabilité sur l'espace probabilisable $(\Omega, \mathcal{A})$. Elle est appelée **probabilité conditionnée à** B (ou **probabilité conditionnelle sachant** B).

Démonstration. On a d'abord

$$\mathbb{P}^B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

On montre ensuite que l'application $\mathbb{P}^B$ est σ-additive : en effet, pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'événements disjoints deux à deux on a

$$\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) \cap B = \bigcup_{i\in\mathbb{N}} (A_i \cap B),$$

les $A_i \cap B$ étant disjoints deux à deux, et donc

$$\mathbb{P}^B\Big(\bigcup_{i\in\mathbb{N}} A_i\Big) = \frac{\sum_{i\in\mathbb{N}} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i\in\mathbb{N}} \mathbb{P}^B(A_i).$$

□

Remarque 8.
1. Cette proposition est très importante puisqu'elle dit que toutes les propriétés établies jusqu'à maintenant pour une probabilité quelconque sont vraies pour la probabilité conditionnelle $\mathbb{P}^B$.
2. Pour que les événements A et B soient **indépendants**, il faut et il suffit que : $\mathbb{P}^B(A) = \mathbb{P}(A)$.

Exemple 2.2.1. On jette deux fois un dé à 6 faces. Soit A l'événement "*on obtient un 6 au premier jet*", et soit $B_k, 2 \le k \le 12$, les événements "*la somme des deux résultats est k*". On modélise les deux lancers de dé par l'espace probabilisé $\big(\Omega, \mathcal{P}(\Omega), \mathbb{P}\big)$, où $\Omega = \{1, 2, \cdots, 6\}^2$ et $\mathbb{P}$ est la probabilité uniforme.

D'après la définition des probabilités conditionnelles on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A)$. On a la généralisation suivante :

Proposition 2.2.2 (Probabilités conditionnelles en cascade). Soit $(A_i)_{1\le i\le n}$ une suite finie d'événements tels que $\mathbb{P}\Big(\bigcap_{1\le i\le n-1} A_i\Big) > 0$. On a :

$$\mathbb{P}\Big(\bigcap_{1\le i\le n} A_i\Big) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Démonstration. On remarque que toutes les probabilités conditionnelles introduites sont bien définies puisque, pour tout $j \in \{1, \cdots, n-1\}$, on a :

$$\mathbb{P}\Big(\bigcap_{1\le i\le j} A_i\Big) \ge \mathbb{P}\Big(\bigcap_{1\le i\le n-1} A_i\Big) > 0.$$

La formule s'obtient ensuite par simple récurrence. □

Exemple 2.2.2. On tire successivement 4 cartes d'un jeu de 52 cartes. Quelle est la probabilité de tirer les 4 as ?

Définition 2.2.2. Soit $(A_i)_{i\in I}$ une famille **dénombrable** d'événements disjoints deux à deux telle que

$$\mathbb{P}\Big(\bigcup_{i\in I} A_i\Big) = 1$$

Une telle famille est appelée **système complet d'événements**.

Soit N le complémentaire de $\bigcup_{i\in I} A_i$ dans Ω. On a $\mathbb{P}(N) = 0$. Les événements $(A_i)_{i\in I}$ et N forment une partition de Ω. On traduit cette situation en termes probabilistes : avec probabilité 1, l'un des événements A_i et un seul se réalise.

Exemple 2.2.3. On lance une pièce de monnaie jusqu'à ce qu'on obtienne un "pile". Pour tout $i \in \mathbb{N} \setminus \{0\}$ on note l'événement A_i : "*pile apparaît pour la première fois au i-ème lancer*".

Théorème 2.2.1 (Formule des probabilités totales). Soit $(A_i)_{i\in I}$ un système complet d'événements tel que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $A \in \mathcal{A}$:

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|A_i)\,\mathbb{P}(A_i).$$

Démonstration. Soit N le complémentaire de $\bigcup_{i\in I} A_i$.

$$A = \Big(\bigcup_{i\in I}(A\cap A_i)\Big) \cup (A\cap N),$$

les deux événements étant disjoints. Comme on a $\mathbb{P}(A\cap N) \leq \mathbb{P}(N) = 0$, on obtient immédiatement

$$\mathbb{P}(A) = \mathbb{P}\Big(\bigcup_{i\in I}(A\cap A_i)\Big) = \sum_{i\in I}\mathbb{P}(A\cap A_i)$$

car les $A\cap A_i$ sont eux aussi disjoints deux à deux. Il ne reste plus alors qu'à utiliser la définition des probabilités conditionnelles. □

Remarque 9. Un cas particulier de système complet est le système (B, B^c) où B est un événement tel que $0 < \mathbb{P}(B) < 1$. La formule des probabilités totales devient dans ce cas

$$\mathbb{P}(A) = \mathbb{P}(A|B)\,\mathbb{P}(B) + \mathbb{P}(A|B^c)\,\mathbb{P}(B^c).$$

Exemple 2.2.4. Une urne U_1 (respectivement U_2) contient b_1 (respectivement b_2) boules blanches et n_1 (respectivement n_2) boules noires. On choisit au hasard une urne et on tire ensuite une boule dans cette urne. On cherche la probabilité de tirer une boule noire.

2.2.2 Formule de Bayes

Proposition 2.2.3. Soit deux événements $A_1, A_2 \in \mathcal{A}$ de probabilités non nulles. Alors :

$$\mathbb{P}(A_2|A_1) = \mathbb{P}(A_1|A_2)\,\frac{\mathbb{P}(A_2)}{\mathbb{P}(A_1)}.$$

Cette proposition s'obtient juste en utilisant la définition d'une probabilité conditionnelle.

Théorème 2.2.2 (Formule de Bayes). Soit $(A_i)_{i\in I}$ un système complet d'événements tel que $\mathbb{P}(A_i) > 0$ pour tout $i \in I$. On a alors, pour tout événement $A \in \mathcal{A}$ de probabilité non nulle et pour tout $i \in I$:

$$\mathbb{P}(A_i|A) = \frac{\mathbb{P}(A|A_i)\,\mathbb{P}(A_i)}{\sum_{j\in I} \mathbb{P}(A|A_j)\,\mathbb{P}(A_j)}.$$

Démonstration. Il résulte de la proposition 2.2.3 que, pour tout $i \in I$:

$$\mathbb{P}(A_i|A) = \mathbb{P}(A|A_i)\,\frac{\mathbb{P}(A_i)}{\mathbb{P}(A)}.$$

Il suffit alors d'appliquer le théorème 2.2.1. □

Exemple 2.2.5. On reprend l'exemple des deux urnes. On cherche maintenant la probabilité *a posteriori* d'avoir tiré dans l'urne U_i, sachant qu'une boule noire a été tirée.

Chapitre 3

Variables aléatoires discrètes

3.1 Loi de probabilité

De manière générale si $f : A \to B$ est une application. L'image réciproque de $\beta \subset B$ est l'ensemble $f^{-1}(\beta) = \{a \in A | f(a) \in \beta\}$. On note aussi $f^{-1}(\beta) = \{f = \beta\} \subset A$.

3.1.1 Généralités

Définition 3.1.1. Soit $(\Omega, \mathcal{A})$ un espace probabilisable et soit E un ensemble. On dit qu'une application $X : \Omega \to E$ est une **variable aléatoire discrète** si les deux conditions sont satisfaites :
(i) L'ensemble $X(\Omega)$ des valeurs prises par X est dénombrable.
(ii) Pour tout $x \in E$, on a $X^{-1}(\{x\}) \in \mathcal{A}$.

Notation : Pour alléger l'écriture, on utilise la notation $\{X = x\}$ au lieu de $X^{-1}(\{x\}) \doteq \{\omega \in \Omega | X(\omega) = x\}$. Les variables aléatoires sont toujours notées en majuscules tandis que les valeurs (déterministes) prises par ces variables aléatoires sont notées en minuscules.

Remarque 10. La plupart du temps, on aura $E = \mathbb{N}$ ou $\mathbb{Z}$ et on munira E de la tribu $\mathcal{E} = \mathcal{P}(E)$.

Définition - Proposition 3.1.1. Soit X une variable aléatoire définie sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans l'espace probabilisable $(E, \mathcal{E})$. L'application

$$\begin{aligned} \mathbb{P}_X : \mathcal{E} &\to [0, 1] \\ A &\mapsto \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A) \end{aligned}$$

est une probabilité sur l'espace probabilisable $(E, \mathcal{E})$ appelée **loi (de probabilité)** de la variable aléatoire X.

Démonstration. La mesure $\mathbb{P}_X$ est bien de masse 1 : en effet, $\mathbb{P}_X(E) = \mathbb{P}(X \in E) = \mathbb{P}(\Omega) = 1$. Elle est bien σ-additive car si les ensembles $E_1, E_2 \in \mathcal{E}$ sont disjoints, alors $X^{-1}(E_1)$ et $X^{-1}(E_2)$

le sont aussi. On a alors pour toute suite $(E_n)_{n\in\mathbb{N}}$ d'éléments disjoints deux à deux de $\mathcal{E}$

$$\mathbb{P}_X\Big(\bigcup_{n\in\mathbb{N}} E_n\Big) = \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} X^{-1}(E_n)\Big) = \sum_{n=0}^{\infty} \mathbb{P}(X^{-1}(E_n)) = \sum_{n=0}^{\infty} \mathbb{P}_X(E_n). \qquad \square$$

La loi d'une variable aléatoire est donc une probabilité sur l'espace des valeurs prises par celle-ci.

Définition - Proposition 3.1.2. Si X est une variable aléatoire discrète à valeurs dans $(E, \mathcal{E})$, sa loi est déterminée entièrement par la fonction

$$\begin{aligned} f^X : E &\to [0,1] \\ x &\mapsto \mathbb{P}^X(\{x\}) = \mathbb{P}(\{X = x\}) \end{aligned}$$

appelée **fonction de masse** de la variable X. L'ensemble $\{x \in E | \mathbb{P}(\{X = x\}) > 0\} \subset X(\Omega)$ est appelé **support** de la variable aléatoire X.

Notation : Pour alléger l'écriture, on écrit $\mathbb{P}(X = x)$ au lieu de $\mathbb{P}(\{X = x\})$.

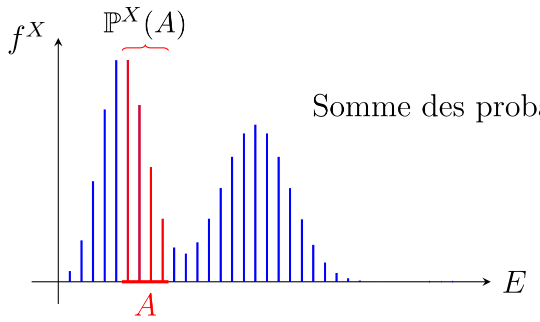
Démonstration. En effet, pour tout $A \in \mathcal{E}$, on a :

$$\mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}\left(\bigcup_{x\in X(\Omega)\cap A} X^{-1}(\{x\})\right) = \sum_{x\in X(\Omega)\cap A} \mathbb{P}\left(X^{-1}(\{x\})\right) = \sum_{x\in A} \mathbb{P}(\{X = x\}). \qquad \square$$

Représentation : Lorsque l'on demande de donner la loi d'une variable aléatoire X, on s'attend soit à :

— la formule de la fonction de masse : $\mathbb{P}(X = x) = f^X(x)$ pour tout x dans le support de X (en général, on ne note que les probabilités des éléments du support car les autres éléments de E ont une probabilité nulle d'apparition).

— la représentation du graphe de f^X :



— (si E est de cardinal n fini), un tableau contenant les couples $(x, \mathbb{P}(X = x) = p_x)$

x	x_1	$\cdots$	x_p
$\mathbb{P}(X = x)$	p_{x_1}	$\dots$	p_{x_n}

3.1.2 Quelques lois usuelles

Loi uniforme

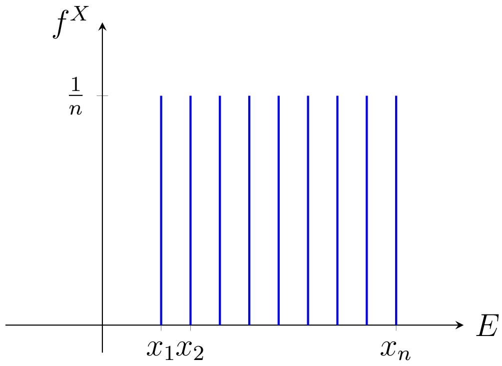
Comme nous l'avons vu précédemment, cette loi accorde la même probabilité à chaque élément d'un ensemble fini, que nous noterons $\{x_1, \cdots, x_n\}$. Si X est l'un des éléments de cet ensemble pris au hasard, c'est une variable aléatoire discrète à valeurs dans $E = \{x_1, \cdots, x_n\}$. On dit que

X est une variable aléatoire de loi uniforme sur l'ensemble E et on écrit $X \sim \mathcal{U}(\{x_1, \cdots, x_n\})$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = \frac{1}{n}$$

pour $x \in E = \{x_1, \cdots, x_n\}$. On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{i=1}^{n} \mathbb{P}(X = x_i) = \sum_{i=1}^{n} \frac{1}{n} = 1.$$



Exemple 3.1.1.
1. On effectue un tirage avec un dé à 6 faces équilibré. Soit X le résultat : il suit une loi uniforme sur $\{1, \cdots, 6\}$.
2. Une urne est remplie de 10 boules numérotées de 1 à 10. On tire une boule et on note X son numéro : X suit une loi uniforme sur $\{1, \cdots, 10\}$.

Loi de Bernoulli

Cette loi de probabilité sert à coder de manière numérique la réalisation ou non d'un événement. Soit A un événement aléatoire de probabilité $\mathbb{P}(A) = p \in]0, 1[$. On définit la variable aléatoire

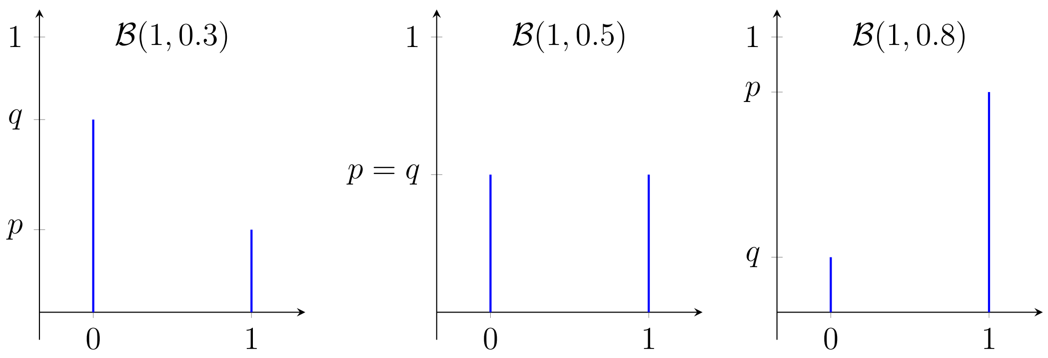
$$X = \mathbb{1}_A = \begin{cases} 1 \text{ si } A \text{ se réalise (succès)}, \\ 0 \text{ si } A^c \text{ se réalise (échec)}. \end{cases}$$

Lorsque $X = 1$ (*i.e.* A se réalise) ont parle de **succès**. Ainsi $p = \mathbb{P}(A)$ est la probabilité de succès. Lorsque $X = 0$ (*i.e.* A^c se réalise) on parle d'**échec**.
On dit que X suit la loi de Bernoulli de paramètre p et on écrit : $X \sim \mathcal{B}(1, p)$. Le support de X est $X(\Omega) = \{0, 1\}$. Sa fonction de masse est pour $x \in E = \{0, 1\}$

$$\mathbb{P}(X = x) = p^x q^{1-x} = \begin{cases} p \text{ si } x = 1 \\ q \text{ si } x = 0 \end{cases}$$

avec $q = 1 - p$. On vérifie que

$$\mathbb{P}_X(E) = \mathbb{P}(X \in E) = \sum_{x \in E} \mathbb{P}(X = x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = p + q = 1.$$



Exemple 3.1.2. On peut toujours se ramener à un schéma de Bernoulli dès lors que l'on ne considère que les deux issues "A" et "non A" :

1. On lance une pièce de monnaie équilibrée. Soit A l'événement "face". Ici $X = 1$ si la pièce tombe sur face et 0 sinon. On a $p = \frac{1}{2}$.
2. On lance un dé équilibré. Soit A l'événement "résultat < 3". Ici $X = 1$ si le résultat est ⚀ ou ⚁ et $X = 0$ sinon. On a $p = \frac{1}{3}$.
3. Observation du sexe du prochain bébé qui naîtra dans une maternité : $A =$ "fille". Ici $X = 1$ si c'est une fille et 0 sinon. Selon des statistiques de 2008 (`http://www.bartleby.com/151/fields/31.html`), en France, $p = \frac{1}{2.05} \approx 0.4878$. Noter que p varie (peu) selon les pays (voir le site internet).

Loi binomiale

Cette loi de probabilité sert à modéliser le nombre de réalisations d'un événement : C'est la loi de la variable aléatoire du nombre de succès en n répétitions indépendantes d'un schéma de Bernoulli.

Une telle variable aléatoire X s'écrit donc comme la somme de n variables aléatoires de loi de Bernoulli indépendantes. Soit donc $Y_1, \cdots, Y_n$ des variables aléatoires $\mathcal{B}(p)$ indépendantes, on a :

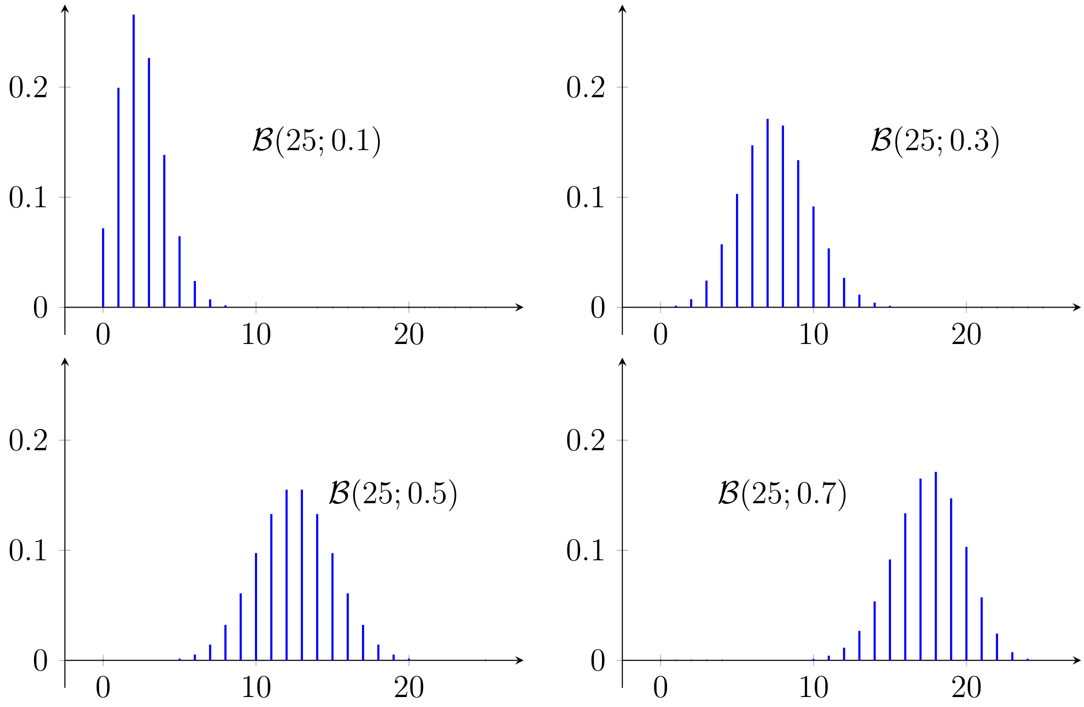
$$\mathbb{P}_X = \mathbb{P}_{Y_1+Y_2+\cdots+Y_n}.$$

On dit que X est une variable aléatoire de loi binomiale de paramètres n et p et on écrit $X \sim \mathcal{B}(n,p)$. Le support de X est $X(\Omega) = \{0, 1, \cdots, n\}$. Sa fonction de masse est pour $x \in \{0, 1, \cdots, n\}$,

$$\mathbb{P}(X = x) = \binom{n}{x} p^x q^{n-x}$$

où $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ est le nombre de combinaisons de x éléments parmi n. On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1.$$



Exemple 3.1.3. 1. On lance une pièce de monnaie équilibrée 6 fois de suite. On s'intéresse à la probabilité d'obtenir, au cours de ces 6 lancers, 2 fois l'événement $A =$ "face". Si on note X le nombre de "face" sur 6 lancers, on a

2. On effectue des tirages avec remise dans une urne remplie de boules, dont 1/3 de blanches et 2/3 de rouges. Soit X le nombre de boules blanches tirées après 10 lancers : on a $X \sim \mathcal{B}(10, \frac{1}{3})$.
3. On estime qu'en France à l'instant t, la probabilité qu'un individu soit un fumeur est de 0.2. Un bus contient 6 personnes. Quelle est la probabilité que 4 ou plus de ces passagers soient fumeurs ? On suppose leur indépendance.

Loi géométrique

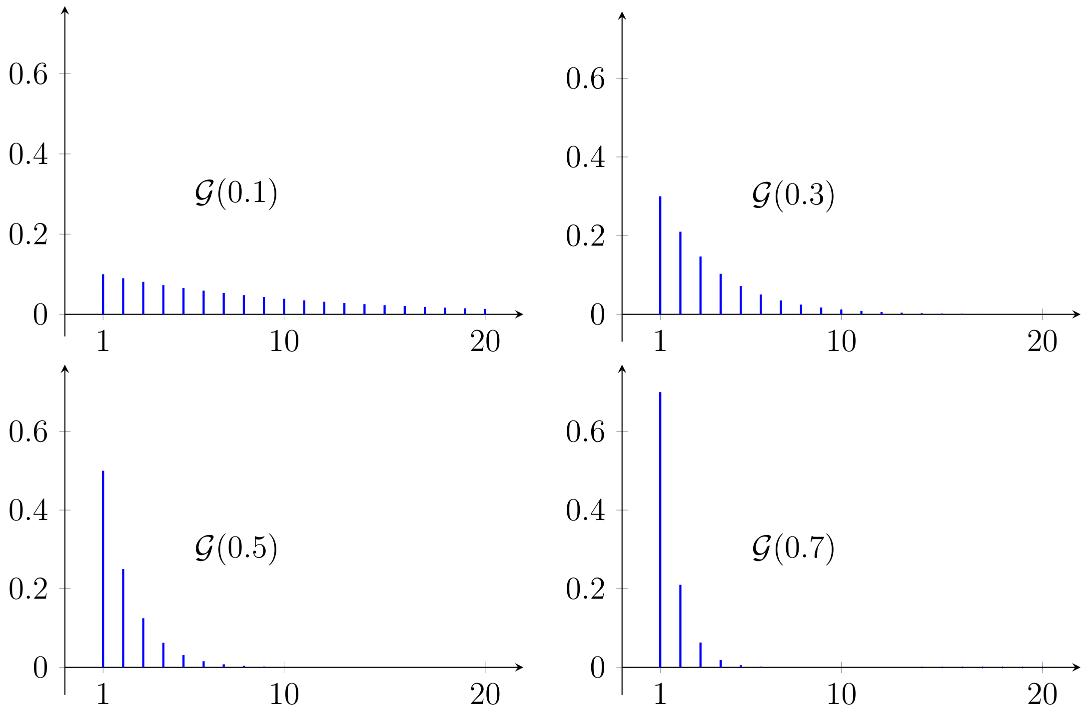
Cette loi de probabilité sert à modéliser le nombre d'essais nécessaires à la réalisation d'un événement. C'est la loi de la variable aléatoire du rang du premier succès dans la répétition indépendante d'un schéma de Bernoulli de paramètre p (probabilité de succès).

Une telle variable aléatoire X suit une loi géométrique de paramètre p et on écrit $X \sim \mathcal{G}(p)$. Le support de X est $X(\Omega) = \mathbb{N} \setminus \{0\} = \mathbb{N}^*$. Sa fonction de masse est pour $x \in E = \mathbb{N}^*$

$$\mathbb{P}(X = x) = pq^{x-1}$$

On vérifie que

$$\sum_{x \in E} \mathbb{P}(X = x) = \sum_{x=1}^{+\infty} pq^{x-1} = p\frac{1}{1-q} = 1.$$



Exemple 3.1.4. 1. On lance un dé à 6 faces jusqu'à l'obtention d'un 6. Si on note X le nombre de lancers nécessaires, on a

2. On lance une pièce de monnaie équilibrée jusqu'à la réalisation de l'événement A = "face". Si on note X le nombre de lancers nécessaires, on a $X \sim \mathcal{G}(\frac{1}{2})$ de sorte que :

$$\mathbb{P}(X = x) = \frac{1}{2^x}.$$

Loi hypergéométrique

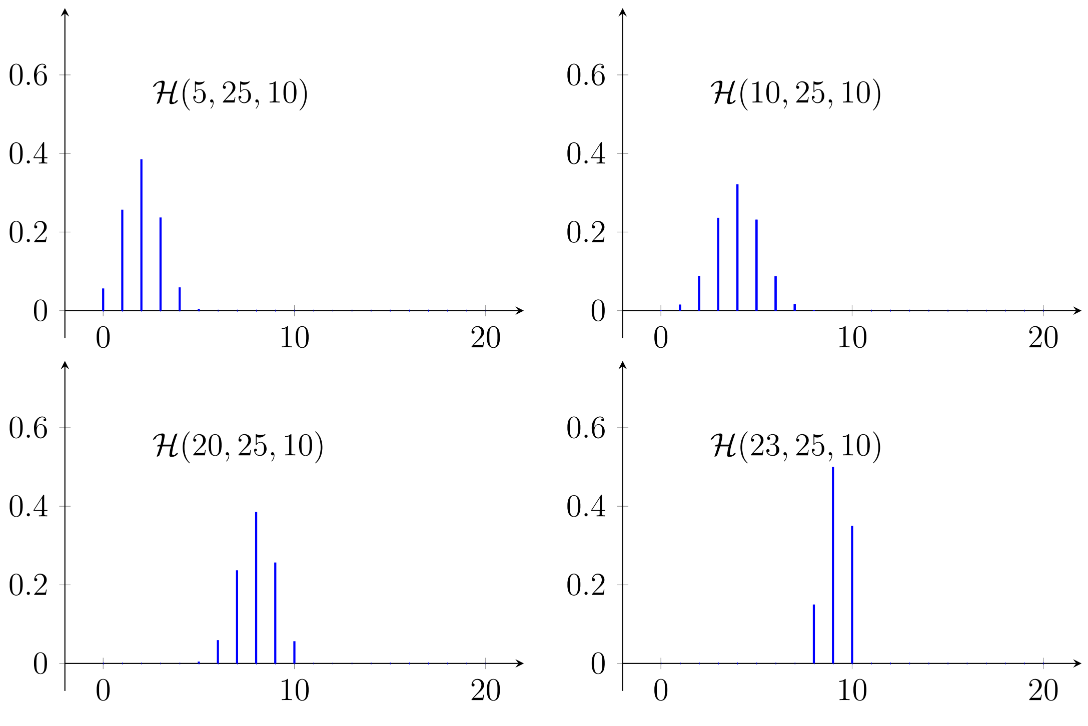
Cette loi de probabilité sert à modéliser le résultat d'un tirage *sans remise*. Soit U un ensemble de cardinal r partitionné en deux sous-ensembles U_1 et U_2, de cardinaux respectifs r_1 et $r_2 = r - r_1$. On se propose de choisir "au hasard" n éléments de U ($n \leq r$) et on note X le nombre d'éléments de U_1 parmi ces n. C'est une variable aléatoire discrète de support $X(\Omega) = \{\max(0, n - r_2), \cdots, \min(n, r_1)\}$. On dit que X est une variable aléatoire de loi hypergéométrique de paramètres n, r et r_1 et on écrit $X \sim \mathcal{H}(n, r, r_1)$. Sa fonction de masse est :

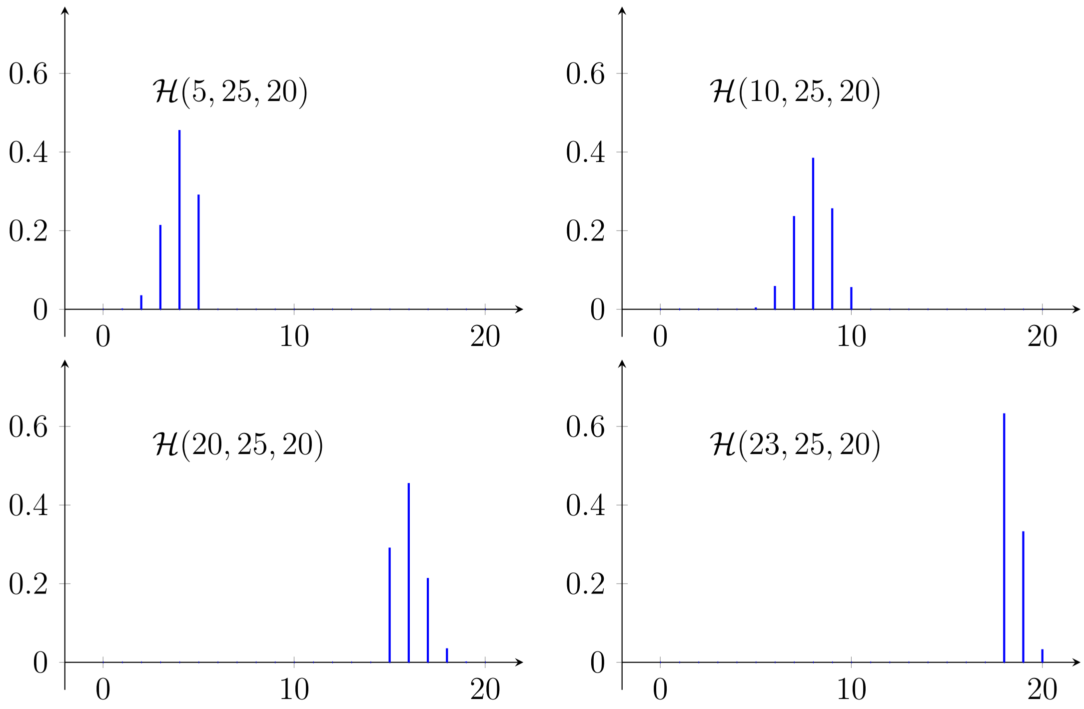
$$\mathbb{P}(X = x) = \frac{\binom{r_1}{x}\binom{r_2}{n-x}}{\binom{r}{n}}$$

pour $x \in \{\max(0, n - r_2), \cdots, \min(n, r_1)\}$. On vérifie que

$$\begin{aligned}\sum_{x \in E} \mathbb{P}(X = x) &= \sum_{x=\max(0,n-r_2)}^{\min(n,r_1)} \frac{\binom{r_1}{x}\binom{r_2}{n-x}}{\binom{r}{n}} \\ &= \frac{1}{\binom{r}{n}} \sum_{x=\max(0,n-r_2)}^{\min(n,r_1)} \binom{r_1}{x}\binom{r_2}{n-x} = \frac{1}{\binom{r_1+r_2}{n}} \sum_{x=0}^{r_1} \binom{r_1}{x}\binom{r_2}{n-x} = 1\end{aligned}$$

d'après l'identité de Vandermonde.





Exemple 3.1.5. 1. Un lac contient r poissons, dont r_1 sont d'une espèce intéressante. On pêche n poissons, en supposant que toutes les espèces se laissent aussi facilement attraper, et on note X le nombre de poissons de l'espèce intéressante parmi les poissons attrapés : X suit une loi hypergéométrique.

2. Ce modèle hypergéométrique est à la base de la théorie des sondages.

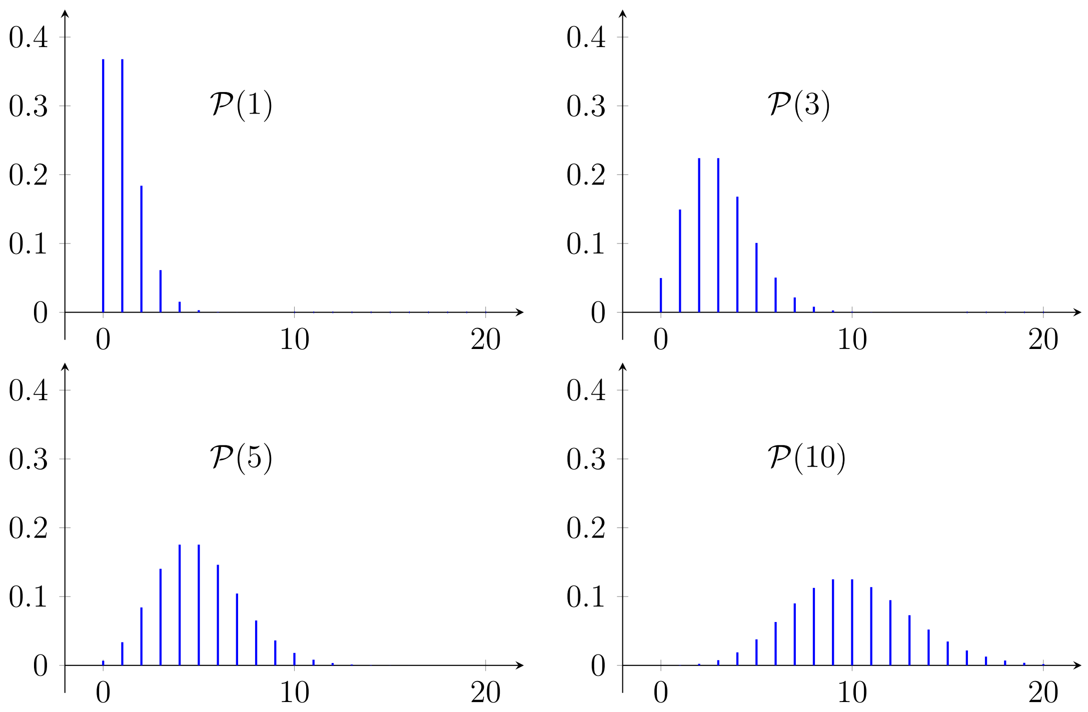
Loi de Poisson

Cette loi de probabilité décrit le comportement du nombre d'événements se produisant dans un laps de temps fixé, si ces événements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l'événement précédent. Soit X ce nombre d'occurrences. C'est une variable aléatoire discrète de support $X(\Omega) = \mathbb{N}$. On dit que X est une variable aléatoire de loi de Poisson de paramètre $\lambda > 0$ et on écrit $X \sim \mathcal{P}(\lambda)$. Sa fonction de masse est :

$$\mathbb{P}(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

pour $x \in \mathbb{N}$. On vérifie que

$$\sum_{x\in E}\mathbb{P}(X = x) = \sum_{x=0}^{+\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{+\infty}\frac{\lambda^x}{x!} = 1.$$



Exemple 3.1.6. Actuellement, on utilise beaucoup cette loi dans les télécommunications (pour compter le nombre de communications dans un intervalle de temps donné), le contrôle de qualité statistique (nombre de défauts), la description de certains phénomènes liés à la désintégration radioactive, la biologie (mutations), la météorologie, la finance pour modéliser la probabilité de défaut d'un crédit, le Yield Management (pour estimer la demande de passagers).

3.2 Moments

Les moments d'une variable aléatoire sont des paramètres numériques qui donnent des renseignements sur la loi de cette variable aléatoire (sans toutefois, en général, la déterminer complètement).

3.2.1 Moyenne (espérance mathématique)

Définition 3.2.1. Soit X une variable aléatoire réelle discrète à valeurs dans E. Si la somme $\sum_{x\in E} |x|\, \mathbb{P}(X = x)$ est finie, on dit que la variable aléatoire X possède une **moyenne** ou **espérance mathématique**

$$\mathbb{E}(X) = \sum_{x\in E} x\, \mathbb{P}(X = x).$$

L'espérance $\mathbb{E}(X)$ est la **moyenne** des valeurs prises par la variable aléatoire X, pondérées par leur probabilité d'apparition.

Premières propriétés de l'espérance

Proposition 3.2.1. On peut montrer les propriétés suivantes :

(i) Soit un réel a. Si la variable aléatoire X est telle que $\mathbb{P}(X = a) = 1$, elle admet une moyenne égale à a et on écrit
$$\mathbb{E}(a) = a.$$

(ii) Toute variable aléatoire discrète bornée admet une moyenne.

(iii) Si $X \geq 0$ alors $\mathbb{E}(X) \geq 0$. Si $X \geq 0$ et $\mathbb{E}(X) = 0$ alors $X = 0$ avec probabilité 1.

(iv) Si X et Y possèdent une moyenne et vérifient $X \leq Y$ alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

(v) Si X possède une moyenne, on a :
$$|\mathbb{E}(X)| \leq \mathbb{E}|X|.$$

Exemple 3.2.1. 1. Si $X \sim \mathcal{U}(\{x_1, \cdots, x_n\})$, sa moyenne est la moyenne arithmétique des x_i :
$$\mathbb{E}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

2. Si $X \sim \mathcal{B}(1, p)$, sa moyenne est
$$\mathbb{E}(X) = 1 \times p + 0 \times q = p.$$

Espérance d'une fonction de variable aléatoire

Théorème 3.2.1 (Théorème de transfert). Soit X une variable aléatoire discrète à valeurs dans E. Soit f une application de E dans $\mathbb{R}$. L'application composée $Y = f \circ X = f(X)$ est une variable aléatoire réelle discrète. Pour qu'elle admette une moyenne, il faut et il suffit que la somme
$$\sum_{x \in E} |f(x)|\, \mathbb{P}(X = x)$$
soit finie. Dans ce cas, on a :
$$\mathbb{E}\big(f(X)\big) = \sum_{x \in E} f(x)\, \mathbb{P}(X = x).$$

Exemple 3.2.2. Quel est la moyenne du carrée du tirage d'un dé équilibré ?

Proposition 3.2.2. Soit X et Y deux variables aléatoires réelles discrètes et a et b deux réels. On a :
$$\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y).$$

Démonstration. Cette égalité s'obtient en appliquant le théorème de transfert avec $Z = (X, Y)$, variable aléatoire discrète, et $f : (x, y) \mapsto ax + by$.

□

Moyennes des lois discrètes classiques

Loi binomiale Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n,p)$.

Loi géométrique Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$.

Loi de Poisson Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$.

Remarque 11. Si l'on sait qu'une variable aléatoire suit une loi géométrique ou une loi de Poisson, pour identifier entièrement sa loi, il suffit de connaître sa moyenne, ce qui n'est pas le cas pour la loi binomiale.

3.2.2 Moments d'ordre deux et variance

Généralités

Définition 3.2.2. Soit X une variable aléatoire réelle discrète et p un entier supérieur ou égal à 1. Si $\mathbb{E}(|X|^p) < +\infty$, on dit que X admet un **moment d'ordre p**, qui n'est autre que le nombre réel $\mathbb{E}(X^p)$.

Proposition 3.2.3. (Inégalité de Cauchy-Schwarz). Soit X et Y deux variables aléatoires réelles discrètes admettant un moment d'ordre deux. On a :

$$[\mathbb{E}(XY)]^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2).$$

Démonstration.

□

Définition 3.2.3. (i) Si X admet un moment d'ordre 1, la variable aléatoire

$$\dot{X} = X - \mathbb{E}(X)$$

est appelée **variable aléatoire centrée** associée à X. On a $\mathbb{E}(\dot{X}) = 0$.

(ii) Si X admet un moment d'ordre 2, X admet une moyenne et le réel

$$\mathbb{V}(X) = \mathbb{E}\Big[\big(X - \mathbb{E}(X)\big)^2\Big]$$

est appelé **variance** de X. Sa racine carrée est appelée **écart-type** de X et notée σ_X. On a donc $\mathbb{V}(X) = \sigma_X^2$.

(iii) Si X admet un moment d'ordre 2 et que $\sigma_X \neq 0$, la variable aléatoire

$$\tilde{X} = \frac{X - \mathbb{E}(X)}{\sigma_X}$$

est appelée **variable aléatoire centrée réduite** associée à X. On a $\mathbb{E}(\tilde{X}) = 0$ et $\mathbb{V}(\tilde{X}) = 1$.

D'après le théorème de transfert, la variance de X s'obtient de la manière suivante :

$$\mathbb{V}(X) = \sum_{x \in E} [x - \mathbb{E}(X)]^2 \, \mathbb{P}(X = x).$$

Proposition 3.2.4. Si X admet un moment d'ordre deux, on a :

(i)

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

(ii) Pour tous réels a et b, on a :

$$\mathbb{V}(aX + b) = a^2 \mathbb{V}(X).$$

Démonstration.

$\square$

Covariance de deux variables aléatoires

Définition 3.2.4. Si X et Y admettent un moment d'ordre 2, il résulte de l'inégalité de Cauchy-Schwarz que la variable aléatoire $\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)$ admet une moyenne, que nous appelons **covariance** de X et Y :

$$\operatorname{cov}(X, Y) = \mathbb{E}\Big[\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\Big].$$

La proposition suivante permet en général un calcul plus simple de la covariance de deux variables aléatoires et permet également de calculer la variance d'une somme de variables aléatoires.

Proposition 3.2.5. Si X et Y admettent un moment d'ordre 2, on a :

(i)

$$\operatorname{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y).$$

(ii)

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\ \operatorname{cov}(X, Y).$$

Démonstration.

$\square$

Proposition 3.2.6. Si X et Y admettent un moment d'ordre 1 et sont **indépendantes**, on a :

$$\mathbb{E}(X\,Y) = \mathbb{E}(X)\,\mathbb{E}(Y).$$

Démonstration. On considère que X est à valeurs dans E_X et Y à valeurs dans E_Y. La variable aléatoire XY est donc discrète, à valeurs dans $E_Z = \{z = xy \in \mathbb{R} : x \in E_X, y \in E_Y\}$. On a :

$$\begin{aligned}
\mathbb{E}(X\,Y) &= \sum_{z\in E_Z} z\,\mathbb{P}(XY = z)\\
&= \sum_{z\in E_Z} z \sum_{x\in E_X, y\in E_Y | xy=z} \mathbb{P}(X = x, Y = y)\\
&= \sum_{z\in E_Z} z \sum_{x\in E_X, y\in E_Y | xy=z} \mathbb{P}(X = x)\,\mathbb{P}(Y = y)\\
&= \sum_{z\in E_Z} \sum_{x\in E_X, y\in E_Y | xy=z} x\,y\,\mathbb{P}(X = x)\,\mathbb{P}(Y = y)\\
&= \sum_{x\in E_X, y\in E_Y} x\,y\,\mathbb{P}(X = x)\,\mathbb{P}(Y = y)\\
&= \sum_{x\in E_X} x\,\mathbb{P}(X = x) \sum_{y\in E_Y} y\,\mathbb{P}(Y = y)\\
&= \mathbb{E}(X)\,\mathbb{E}(Y).
\end{aligned}$$

$\square$

Ainsi, si X et Y sont **indépendantes**, on a :

$$\mathrm{cov}(X, Y) = 0 \text{ et } \mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

Par extension, la variance d'une somme de variables aléatoires indépendantes est la somme de leurs variances. Attention ! L'égalité $\mathrm{cov}(X, Y) = 0$ **n'entraîne pas** l'indépendance entre X et Y.

Variance des lois discrètes classiques

Loi de Bernoulli Soit X une variable aléatoire suivant une loi de Bernoulli $\mathcal{B}(1, p)$. On a :

Loi binomiale Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$.

Loi géométrique Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$.

Loi de Poisson Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$.

On remarque que, pour une variable aléatoire de Poisson, le paramètre représente à la fois la moyenne et la variance.

3.3 Couples de variables aléatoires

3.3.1 Lois jointes et marginales

Définition 3.3.1. Soit X et Y deux variables aléatoires discrètes définies sur l'espace probabilisable $(\Omega, \mathcal{A})$, et à valeurs respectivement dans E et F. Le couple (X, Y) définit ce que l'on peut appeler une variable aléatoire discrète à valeurs dans $E \times F$: à tout ω de Ω, il associe en effet le vecteur $\big(X(\omega), Y(\omega)\big)$.

Proposition 3.3.1. La loi de probabilité du couple (X, Y), aussi appelée **loi conjointe** de (X, Y), est déterminée entièrement par la fonction de masse $(x, y) \mapsto \mathbb{P}(X = x, Y = y)$ sur $E \times F$.

Exemple 3.3.1. On lance deux dés à 6 faces équilibrés. Soit X la somme des résultats et Y leur différence en valeur absolue : le couple (X, Y) est à valeurs dans

Définition 3.3.2. On appelle **lois marginales** de X et de Y les lois respectives de X et de Y, que l'on peut calculer à partir de la loi conjointe :

$$\forall x \in E, \mathbb{P}(X = x) = \sum_{y \in F} \mathbb{P}(X = x, Y = y).$$
$$\forall y \in F, \mathbb{P}(Y = y) = \sum_{x \in E} \mathbb{P}(X = x, Y = y).$$

Exemple 3.3.2. En gardant l'exemple précédent, on veut calculer $\mathbb{P}(X = 7)$:

3.3.2 Lois conditionnelles

Définition 3.3.3. Soit X et Y deux variables aléatoires discrètes définies sur l'espace probabilisable $(\Omega, \mathcal{A})$, et à valeurs respectivement dans les espaces probabilisables $(E, \mathcal{E})$ et $(F, \mathcal{F})$. Soit $x \in E$ tel que $\mathbb{P}(X = x) > 0$. La **loi de Y conditionnelle à l'événement** $\{X = x\}$, aussi appelée **loi de Y sachant que** $\{X = x\}$, est la probabilité sur $(F, \mathcal{F})$ définie par l'application

$$B \mapsto \mathbb{P}^{(X=x)}(Y \in B) = \mathbb{P}(Y \in B | X = x)$$

et elle est notée $\mathbb{P}_Y^{(X=x)}$.

Exemple 3.3.3. En gardant l'exemple précédent, la loi conditionnelle $\mathbb{P}_Y^{(X=7)}$ est donnée par :

Proposition 3.3.2. La famille des probabilités $\mathbb{P}_Y^{(X=x)}$, où x décrit l'ensemble des éléments de E tels que $\mathbb{P}(X = x) > 0$, et la loi $\mathbb{P}_X$ de X déterminent entièrement la loi du couple aléatoire (X, Y)

Démonstration.

□

Proposition 3.3.3. Deux variables aléatoires X et Y sont indépendantes si et seulement si, pour tout $x \in E$ tel que $\mathbb{P}(X = x) > 0$, la loi conditionnelle $\mathbb{P}_Y^{(X=x)}$ est égale à la loi marginale $\mathbb{P}_Y$ de Y.

Définition 3.3.4. La **moyenne conditionnelle** et la **variance conditionnelle** de Y sachant que $X = x$ sont la moyenne et la variance calculées à partir de la loi conditionnelle $\mathbb{P}_Y^{(X=x)}$.

Exemple 3.3.4. En gardant l'exemple précédent, la moyenne conditionnelle de Y sachant $X = 7$ est :

et la variance conditionnelle de Y sachant $X = 7$ est :

3.3.3 Somme de variables aléatoires

On a très souvent besoin de calculer la somme de n variables aléatoires :

$$Y = X_1 + \cdots + X_n.$$

Quand les variables aléatoires $X_1, \cdots, X_n$ sont indépendantes, la loi de leur somme peut se calculer à partir des lois marginales. Nous donnons la méthode de calcul, dite de **convolution**, dans le cas où $n = 2$.

Proposition 3.3.4. Soit X_1 et X_2 deux variables aléatoires discrètes **indépendantes** définies sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $\mathbb{Z}$. Alors la variable aléatoire $Y = X_1 + X_2$ est discrète et sa loi est donnée par :

$$\forall x \in \mathbb{Z}, \mathbb{P}(Y = y) = \sum_{x_1 \in \mathbb{Z}} \mathbb{P}(X_1 = x_1)\, \mathbb{P}(X_2 = y - x_1).$$

Démonstration. On a équivalence entre les événements suivants :

$$\{Y = y\} = \bigcup_{x_1 \in \mathbb{Z}} \{(X_1 = x_1) \cup (X_2 = y - x_1)\}.$$

On utilise ensuite le fait que les événements $\Big((X_1 = x_1) \cup (X_2 = y - x_1), x_1 \in \mathbb{Z}\Big)$, sont incompatibles deux à deux et on conclut en utilisant l'indépendance de X_1 et X_2. □

Exemple 3.3.5. On lance deux dés équilibrés. Calculer la loi de la somme des résultats.

Définition 3.3.5. On dit que la loi de probabilité de Y, obtenue à partir de $\mathbb{P}_{X_1}$ et $\mathbb{P}_{X_2}$, est le **produit de convolution** (ou simplement la convolution) des deux probabilités $\mathbb{P}_{X_1}$ et $\mathbb{P}_{X_2}$: elle est notée $\mathbb{P}_{X_1} * \mathbb{P}_{X_2}$.

3.4 Fonctions génératrices

Généralités

Définition 3.4.1. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. La **fonction génératrice** de X est la fonction

$$\begin{aligned} G_X : \mathbb{R} &\to \mathbb{R} \\ s &\mapsto G_X(s) = \mathbb{E}(s^X) \end{aligned}$$

lorsque cette quantité existe.

Proposition 3.4.1. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$, de fonction génératrice G_X et de loi déterminée par :

$$\mathbb{P}(X = n) = p_n, \quad \forall n \in \mathbb{N}.$$

(i) Le domaine de définition de G_X contient l'intervalle $[-1, 1]$. On a :

$$\forall s \in [-1, 1], \quad |G_X(s)| \leq 1 \text{ et } G_X(1) = 1.$$

(ii) Pour tout $s \in [-1, 1]$, on a :

$$G_X(s) = \sum_{n=0}^{+\infty} p_n s^n.$$

(iii) La fonction G_X est continue sur $[-1, 1]$, et $\mathcal{C}^\infty$ sur $]-1, 1[$.

(iv) La fonction génératrice G_X caractérise la loi de X puisque, pour tout $k \in \mathbb{N}$, on a :

$$\mathbb{P}(X = k) = \frac{G_X^{(k)}(0)}{k!}.$$

Fonctions génératrices des lois discrètes classiques

Loi binomiale Soit X une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, p)$. On a, $\forall s \in [-1, 1]$:

Loi géométrique Soit X une variable aléatoire suivant une loi géométrique $\mathcal{G}(p)$. On a, $\forall s \in [-1, 1]$:

Loi de Poisson Soit X une variable aléatoire suivant une loi de Poisson $\mathcal{P}(\lambda)$. On a, $\forall s \in [-1, 1]$:

Fonctions génératrices et moments

La fonction génératrice d'une variable aléatoire déterminant sa loi, il est naturel qu'elle donne ses moments lorsque ceux-ci existent. Nous noterons, pour $r \in \mathbb{N}^*$, $G_X^{(r)}(1^-)$ la $r^{\text{ième}}$ dérivée à gauche de G_X en 1 lorsqu'elle existe.

> **Proposition 3.4.2.** Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Pour que X admette un moment d'ordre $r \in \mathbb{N}^*$, il faut et il suffit que sa fonction génératrice G_X soit r fois dérivable à gauche en 1 et dans ce cas, on a
>
> $$G_X^{(r)}(1^-) = \sum_{k=r}^{+\infty} k(k-1)\cdots(k-r+1)p_k = \mathbb{E}[X(X-1)\cdots(X-r+1)].$$