Chapitre 21

Probabilités sur un univers quelconque

Au premier semestre, nous avons étudié les probabilités sur un univers fini. Nous allons étendre notre étude à des univers infinis.

I Introduction

1) Ensembles dénombrables

Définition 1.
- *Un ensemble est dit dénombrable s'il est en bijection avec $\mathbb{N}$.*
- *Un ensemble est dit au plus dénombrable s'il est fini ou dénombrable.*

Remarque : Un ensemble dénombrable est infini (sinon il existe une bijection de $\mathbb{N}$ sur $[\![1, n]\!]$ pour un certain $n \in \mathbb{N}^*$, ce qui est bien évidemment absurde).

Exemples :
- *L'ensemble $\mathbb{N}$ est dénombrable (il est en bijection avec lui-même).*
- *Si A est une partie infinie de $\mathbb{N}$, alors A est dénombrable.*

  En effet, par récurrence, on construit une suite $(a_n)_{n\in\mathbb{N}}$ par $a_0 = \min(A)$, $a_1 = \min(A\backslash\{a_0\})$ et, pour tout $n \in \mathbb{N}$, $a_{n+1} = \min(A\backslash\{a_0, a_1, \dots, a_n\})$ (ce qui est possible puisque A est infini donc $A\backslash\{a_0, a_1, \dots, a_n\}$ est non vide). L'application $n \longmapsto a_n$ est une bijection de $\mathbb{N}$ dans A.
- *Comme nous l'avons vu dans la feuille d'exercice n^o 7, les ensembles $\mathbb{Z}$, $\mathbb{N}\times\mathbb{N}$ et $\mathbb{Q}$ sont dénombrables.*
- *Le résultat suivant est admis : l'ensemble $\mathbb{R}$ est infini et non dénombrable. C'est aussi le cas de $\mathbb{R}\backslash\mathbb{Q}$ et de tout intervalle non vide et non réduit à un point de $\mathbb{R}$.*

L'étude des probabilités sur des univers non-dénombrables amène de nombreuses difficultés techniques (mais dont les subtilités ne sont pas au programme) et nous allons notamment introduire la notion de tribu pour y remédier.

2) Tribu et événements

Une expérience aléatoire est une expérience renouvelable dont le résultat ne peut être prévu à l'avance et qui, renouvelée dans des conditions identiques, ne donne pas forcément le même résultat à chaque fois. Comme dans le cas fini, on travaille avec un ensemble Ω, appelé univers, dont les éléments sont les issues de l'expérience aléatoire.

Exemple : *Si on lance une infinité de fois un dé, alors on peut choisir l'univers $\Omega = [\![1, 6]\!]^{\mathbb{N}^*}$. Un élément $\omega = (\omega_n)_{n\in\mathbb{N}^*}$ de Ω est une suite réelle telle que, pour tout $n \in \mathbb{N}^*$, ω_n est le numéro obtenu au $n^{\text{ème}}$ lancer. Le fait de tomber sur 6 à chaque fois est représenté par l'éventualité $\omega = (6, 6, 6, \dots)$.*

Une fois l'univers fixé, on va s'intéresser à un certain nombre d'événements qui sont des parties de Ω (et que l'on définit souvent avec une phrase qui peut être vraie ou fausse selon le résultat de l'expérience). Dans le cas des univers finis, on a travaillé systématiquement avec tous les événements possibles, c'est-à-dire avec $\mathcal{P}(\Omega)$. Mais en fait, on peut se restreindre à une classe de parties $\mathcal{A}$ de Ω, plus petite que $\mathcal{P}(\Omega)$ qui contient suffisamment d'informations pour décrire l'expérience.

Exemple : *Si on lance un dé, alors on travaille avec l'univers $\Omega = [\![1, 6]\!]$. Si on ne s'intéresse qu'à l'événement A : « obtenir un 6 », alors les seuls événements qui nous intéressent vraiment sont $A = \{6\}$ et $\overline{A} = [\![1, 5]\!]$. On peut donc se contenter de travailler avec l'ensemble d'événements $\mathcal{A} = \{\varnothing, \{6\}, [\![1, 5]\!], \Omega\}$.*

L'ensemble d'événements $\mathcal{A}$ que l'on considère modélise l'information que l'on peut obtenir à partir des résultats de l'expérience aléatoire. Plus cet espace est gros, plus on dispose d'information.

Une fois que l'on s'est donné une liste d'événements qui nous intéressent, on est amené à étudier des événements définis à partir de ces événements. Ainsi on exige que $\mathcal{A}$ possède certaines propriétés de stabilité :
- Si $A \in \mathcal{A}$, alors $\overline{A} \in \mathcal{A}$: on souhaite en effet pouvoir observer la non réalisation des événements qui nous intéressent.
- Si $(A, B) \in \mathcal{A}^2$, alors $A \cap B \in \mathcal{A}$ et $A \cup B \in \mathcal{A}$: on souhaite en effet pouvoir observer la réalisation de deux événements en même temps ou d'au moins un événement. Plus généralement on exige une stabilité par union et intersection dénombrable : si $(A_i)_{i\in I}$ est une famille d'événements de $\mathcal{A}$, indexée par une partie I (finie ou infinie) de $\mathbb{N}$, alors $\displaystyle\bigcup_{i\in I} A_i \in \mathcal{A}$ et $\displaystyle\bigcap_{i\in I} A_i \in \mathcal{A}$.

Par exemple, si on s'intéresse au nombre de lancers nécessaires pour obtenir le premier 6, alors nous allons naturellement travailler avec les événements A_n : « obtenir le premier 6 au $n^{\text{ième}}$ lancer ». Un autre événement qui peut nous intéresser est E : « le premier 6 intervient au bout d'un nombre pair de lancers ». On a $E = \bigcup_{n\in\mathbb{N}^*} A_{2n}$.

- Enfin on demande aussi que $\Omega \in \mathcal{A}$ et $\varnothing \in \mathcal{A}$: on souhaite en effet pouvoir observer l'événement certain et l'événement impossible.

Une collection de parties $\mathcal{A}$ vérifiant ces propriétés est appelée tribu ou σ-algèbre.

La restriction à $\mathcal{A}$ peut sembler inutile au premier abord. Dans le cas d'un univers fini, nous avions travaillé systématique avec $\mathcal{A} = \mathcal{P}(\Omega)$ sans rencontrer de difficultés. Mais lorsque Ω est infini, surtout lorsqu'il n'est pas dénombrable, l'ensemble $\mathcal{P}(\Omega)$ est très gros et il n'est pas raisonnable d'envisager d'observer tous ses éléments.

Pensons aux parties de $\mathbb{R}$... on connait bien les intervalles et les unions dénombrables d'intervalles. Mais il existe une infinité d'autres types de parties qui nous sont impossibles à décrire. On travaillera donc usuellement avec la plus petite tribu contenant tous les intervalles de $\mathbb{R}$, appelée tribu borélienne.

Une fois que l'on s'est fixé un univers Ω et une tribu $\mathcal{A}$, on se donne une probabilité $\mathbb{P}$ sur $\mathcal{A}$. Comme au premier semestre il s'agit d'une application de $\mathcal{A}$ dans $[0,1]$ qui à un événement A de $\mathcal{A}$ associe sa probabilité $\mathbb{P}(A)$ d'occurrence. Elle vérifie :

- $\mathbb{P}(\Omega) = 1$.
- Si A et B sont deux événements incompatibles (c'est-à-dire tels que $A \cap B = \varnothing$), alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. On parle d'additivité. Dans le cas d'un univers quelconque, il est naturel d'exiger un peu plus : si $(A_i)_{i\in I}$ est une famille d'événements de $\mathcal{A}$, indexée par une partie I de $\mathbb{N}$, alors la somme (éventuellement infinie) des A_i, $i \in I$, a un sens (cf. chapitre sur les séries) et $\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mathbb{P}(A_i)$. On parle alors de σ-additivité.

Une autre motivation de la notion de tribu est qu'il n'est pas possible de calculer la probabilité de n'importe quel événement lorsque l'univers est infini non dénombrable.

Considérons l'expérience aléatoire consistant à choisir uniformément un réel entre 0 et 1. On aurait envie de considérer l'univers $\Omega = [0,1]$ et une probabilité $\mathbb{P}$ qui vérifie la propriété suivante :

(P_λ) *Pour tout $(a,b) \in [0,1]^2$ tel que $a \leqslant b$, $\mathbb{P}([a,b]) = b - a$.
C'est-à-dire la probabilité que le réel choisi se trouve entre a et b est égal à la longueur du segment $[a,b]$.*

On peut montrer (mais c'est très compliqué et totalement hors-programme) qu'il n'existe pas de probabilité sur $\mathcal{P}([0,1])$ qui vérifie la propriété (P_λ). Par contre on peut définir une telle probabilité sur une plus petite tribu, appelée tribu borélienne sur $[0,1]$, qui contient notamment tous les intervalles de $[0,1]$.

Alors convaincus ?

⚠ Ce paragraphe sert avant tout d'introduction et de motivation à la notion des tribus. Conformément au programme, aucune difficulté théorique sur les notions de tribu et de probabilité sur des univers infinis ne sera soulevée cette année (ni l'année prochaine d'ailleurs). Les énoncés admettront en général l'existence d'une tribu et d'une probabilité permettant de décrire l'expérience aléatoire donnée. De plus la notion de variable aléatoire permettra plus ou moins de masquer cette difficulté.

II Espaces probabilisables quelconques

Définition 2 (tribu). *Une tribu (ou σ-algèbre) d'un ensemble Ω est un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ qui vérifie :*

1. $\Omega \in \mathcal{A}$,
2. Stabilité par passage au complémentaire : *pour tout $A \in \mathcal{A}$, on a $\overline{A} \in \mathcal{A}$*
3. Stabilité par union dénombrable : *pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, on a $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.*

Définition 3 (espace probabilisable). *Un espace probabilisable est un couple $(\Omega, \mathcal{A})$ tel que*

- *Ω est un ensemble appelé univers.*
- *$\mathcal{A}$ est une tribu sur Ω appelée tribu des événements de Ω.*

On appelle événement tout élément de la tribu $\mathcal{A}$.

Exemples :

Soit $(\Omega, \mathcal{A})$ un espace probabilisable. Rappelons les propriétés de la réunion et l'intersection d'une famille $(A_i)_{i\in I}$ d'événements, indexée par une partie I de $\mathbb{N}$ (ce sont des cas particuliers des propriétés vues au chapitre *Ensembles et applications*) :

- Si $B \in \mathcal{A}$, alors $\left(\bigcup_{i\in I} A_i\right) \cap B = \bigcup_{i\in I}(A_i \cap B)$ (distributivité de $\cap$ par rapport à $\cup$),
- Si $B \in \mathcal{A}$, alors $\left(\bigcap_{i\in I} A_i\right) \cup B = \bigcap_{i\in I}(A_i \cup B)$ (distributivité de $\cup$ par rapport à $\cap$),
- $\overline{\bigcap_{i\in I} A_i} = \bigcup_{i\in I} \overline{A_i}$ et $\overline{\bigcup_{i\in I} A_i} = \bigcap_{i\in I} \overline{A_i}$ (lois de Morgan).

Retenons aussi la formule suivant qui est très utile en pratique : $\bigcap_{i\in I} A_i = \overline{\bigcup_{i\in I} \overline{A_i}}$.

Proposition 4. *Soit $(\Omega, \mathcal{A})$ un espace probabilisable. Alors :*

1. $\varnothing \in \mathcal{A}$.
2. *Si $(A, B) \in \mathcal{A}^2$, alors $A \cup B$, $A \cap B$ et $A \backslash B$ sont des éléments de $\mathcal{A}$ (c'est-à-dire ce sont des événements).*
3. *Soit I une partie (finie ou infinie) de $\mathbb{N}$. Soit $(A_i)_{i\in I}$ une famille d'événements de $\mathcal{A}$ indexée par I. Alors :*
   - $\bigcup_{i\in I} A_i \in \mathcal{A}$. *Il s'agit de l'événement « il existe $i \in I$ tel que A_i est réalisé ».*
   - $\bigcap_{i\in I} A_i \in \mathcal{A}$. *Il s'agit de l'événement « pour tout $i \in I$, A_i est réalisé ».*

DÉMONSTRATION.

□

Remarque : Les résultats précédents restent valables si l'ensemble I des indices est au plus dénombrable (c'est-à-dire fini ou dénombrable) mais pas forcément une partie de $\mathbb{N}$ (une partie de $\mathbb{Z}$ par exemple). Puisque qu'un ensemble au plus dénombrable est en bijection avec une partie de $\mathbb{N}$, alors on peut toujours se ramener à indexer par une partie de $\mathbb{N}$ (quitte à renuméroter).

Exemples : *On dispose d'une urne contenant n boules numérotées de 1 à n. On tire avec remise une infinité de boules dans cette urne. On peut par exemple travailler avec l'univers $[\![1,n]\!]^{\mathbb{N}^*}$. On admet qu'il existe une tribu adéquate pour laquelle, pour tout $k \in \mathbb{N}^*$, A_k : « le numéro de la $k^{\text{ième}}$ boule tirée est impair » est un événement.*

- *L'événement* $\boxed{}$ *est « tirer au moins une boule dont le numéro est impair ».*
- *L'événement* $\boxed{}$ *est « ne tirer que des boules dont le numéro est pair ».*

Définition 5 (système complet d'événements). *Soient $(\Omega, \mathcal{A})$ un espace probabilisable. Un système complet d'événements est une famille $(A_i)_{i\in I}$ d'événements (i.e d'éléments de $\mathcal{A}$) indexée par une partie I (finie ou infinie) de $\mathbb{N}$ vérifiant :*

- *Pour tout $(i,j) \in I^2$ tel que $i \neq j$, on a $A_i \cap A_j = \varnothing$.*
  On dit que les événement de la famille sont deux à deux incompatibles.
- $\displaystyle\bigcup_{i\in I} A_i = \Omega$.

Exemples :

Proposition/Définition 6. *Soient $(\Omega, \mathcal{A})$ un espace probabilisable et $(A_i)_{i\in I}$ un système complet d'événements. Il existe une plus petite (au sens de l'inclusion) tribu contenant tous les A_i, $i \in I$. On l'appelle tribu engendrée par $(A_i)_{i\in I}$ et on la note $\sigma(A_i,\, i \in I)$.*

DÉMONSTRATION. Admis. □

Exemples :

Remarque : Plus généralement, si $(A_i)_{i\in I}$ est une famille quelconque d'événements alors il existe aussi une plus petite tribu contenant les A_i, $i \in I$. On l'appelle encore tribu engendrée par $(A_i)_{i\in I}$.

Par exemple la tribu engendrée par les intervalles ouverts de $\mathbb{R}$ est appelée la tribu borélienne de $\mathbb{R}$ et notée $\mathcal{B}(\mathbb{R})$.

... et dans la pratique ?

- Lorsque Ω est fini, on prendra toujours $\mathcal{A} = \mathcal{P}(\Omega)$ (c'est ce que l'on a fait au premier semestre).
- Lorsque Ω est infini dénombrable (par exemple $\Omega = \mathbb{N}$), on choisira généralement $\mathcal{A} = \mathcal{P}(\Omega)$. Cependant on peut aussi travailler avec une tribu plus petite qui est mieux adaptée à l'expérience considérée.
- Lorsque Ω est infini et non dénombrable, on travaille presque toujours avec des tribus plus petites (par exemple si $\Omega = \mathbb{R}$, alors on travaille avec la tribu borélienne $\mathcal{B}(\mathbb{R})$). En général, la construction explicite de tribu est très difficile, mais l'existence théorique d'un bonne tribu adaptée à une expérience est souvent prouvée. Par soucis de simplicité et conformément au programme d'ECS, on admettra généralement que l'expérience peut être modélisée à l'aide d'une tribu sans l'expliciter.

III Espaces probabilisés quelconques

1) Probabilité sur un espace probabilisable quelconque

a) Définitions

Au premier semestre une probabilité $\mathbb{P}$ sur un espace probabilisable vérifiait une propriété d'additivité finie : si $(A_n)_{n\in\mathbb{N}}$ est une famille d'événements deux à deux incompatibles alors, pour tout $n \in \mathbb{N}$,

$$S_n = \sum_{k=0}^{n} \mathbb{P}(A_k) = \mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) \leqslant \mathbb{P}\left(\bigcup_{k\in\mathbb{N}} A_k\right).$$

Par conséquent $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ est une série à termes positifs dont la suite des sommes partielles est majorée. Nous en déduisons que qu'elle converge et que $\sum_{n=0}^{+\infty} \mathbb{P}(A_n) \leqslant \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$. Si l'univers Ω est fini, alors il y a égalité puisqu'il n'y a en fait qu'un nombre fini d'événements non vides deux à deux incompatibles. Mais si l'univers Ω n'est pas fini, alors on ne peut pas conclure a priori et nous avons besoin d'ajouter une propriété à la définition d'une probabilité afin de garantir que la somme de cette série soit $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.

Définition 7 (probabilité sur un espace probabilisable). *Soit $(\Omega, \mathcal{A})$ un espace probabilisable quelconque. Une probabilité sur $(\Omega, \mathcal{A})$ est une application de $\mathcal{A}$ dans $[0, 1]$ vérifiant les propriétés suivante :*

1. $\mathbb{P}(\Omega) = 1$.
2. *$\mathbb{P}$ est σ-additive, c'est-à-dire pour toute famille dénombrable $(A_n)_{n\in\mathbb{N}}$ d'événements deux à deux incompatibles,*

$$\boxed{\phantom{\sum_{n=0}^{+\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup A_n\right)}}.$$

Pour tout $A \in \mathcal{A}$, le réel $\mathbb{P}(A)$ est appelé la probabilité de l'événement A.

Remarque : Dans la définition ci-dessus, il est sous-entendu (et on a pas besoin de le préciser) que, si $(A_n)_{n\in\mathbb{N}}$ est une famille d'événements deux à deux incompatibles, alors la série $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ converge et que sa somme vaut $\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.

⚠ Attention, si les événements de la famille $(A_n)_{n\in\mathbb{N}}$ ne sont pas deux à deux incompatibles, alors il se peut que la série $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ diverge.

Définition 8 (espace probabilisé). *Un espace probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ tel que*

- *$(\Omega, \mathcal{A})$ est un espace probabilisable.*
- *$\mathbb{P}$ est une probabilité sur $(\Omega, \mathcal{A})$.*

b) Premières propriétés

Proposition 9. *Une probabilité $\mathbb{P}$ sur un espace probabilisable $(\Omega, \mathcal{A})$ est additive : si A et B sont deux événements incompatibles, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.*

Plus généralement, si $n \in \mathbb{N}$ et $A_0, \dots, A_n$ sont des événements deux à deux incompatibles, alors $\mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) = \sum_{k=0}^{n} \mathbb{P}(A_k)$.

Démonstration.

□

La proposition suivante découle de l'additivité d'une probabilité sur un espace probabilisable. La démonstration est analogue à ce que l'on a vu au premier semestre dans le cas d'un univers finis.

Proposition 10. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Nous avons :*

1. $\mathbb{P}(\varnothing) = 0$,
2. *Si* $(A, B) \in \mathcal{A}^2$, *alors* $\mathbb{P}(A \backslash B) = \mathbb{P}(A) - \mathbb{P}(B \cap A)$.
3. *Si* $A \in \mathcal{A}$, *alors* $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
4. *Si* $(A, B) \in \mathcal{A}^2$ *et* $B \subset A$, *alors* $\mathbb{P}(B) \leqslant \mathbb{P}(A)$.
5. *Si* $(A, B, C) \in \mathcal{A}^3$, *alors* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *et*

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Maintenant voyons comment étendre la notion de σ-additivité à l'union d'une famille d'événements indexée par une partie I (finie ou infinie) de $\mathbb{N}$. Pour cela, nous avons besoin de définir la somme $\sum_{i \in I} \mathbb{P}(A_i)$ lorsque $(A_i)_{i \in I}$ est composée d'événements deux à deux incompatibles.

- Si I est finie, alors la définition de cette somme ne pose pas de problème (cf. chapitre 2).
- Si $I = \mathbb{N}$, alors il s'agit d'une série convergente.
- Si I est infini, alors I est dénombrable et il existe une bijection $\sigma : \mathbb{N} \longrightarrow I$. Puisque la série $\sum_{k \in \mathbb{N}} \mathbb{P}(A_{\sigma(k)})$ converge, elle converge absolument et $\sum_{i \in I} \mathbb{P}(A_i) = \sum_{k=0}^{+\infty} \mathbb{P}(A_{\sigma(k)})$ (i.e. l'ordre des termes n'a pas d'importance).

Nous en déduisons la proposition suivante :

Proposition 11. *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Soit* I *une partie (finie ou infinie) de* $\mathbb{N}$. *Si* $(A_i)_{i \in I}$ *est une famille d'événements deux à deux incompatibles, alors* $\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i)$.

Proposition 12. *Soient* $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* $(A_i)_{i \in I}$ *un système complet d'événement, indexé par une partie* I *de* $\mathbb{N}$. *Alors*

En particulier .

DÉMONSTRATION.

□

c) Exemple : probabilité sur un univers dénombrable

Au premier semestre, nous avons déjà caractérisé les probabilités sur un univers fini. Le résultat se généralise très bien dans le cas d'univers dénombrable :

Théorème 13 (probabilité sur un univers dénombrable). *Soit* Ω *un univers dénombrable. On peut numéroter ses éléments :* $\Omega = \{\omega_0, \omega_1, \ldots, \omega_n, \ldots\}$.

1. *Soit* $\mathbb{P}$ *est une probabilité sur* $(\Omega, \mathcal{P}(\Omega))$. *Pour tout* $n \in \mathbb{N}$, *notons* $p_n = \mathbb{P}(\{\omega_n\}) \in [0, 1]$. *Alors la série* $\sum_{n \in \mathbb{N}} p_n$ *converge et* $\sum_{n=0}^{+\infty} p_n = 1$.
2. *Réciproquement, si* $\sum_{n \in \mathbb{N}} p_n$ *est une série convergente à termes positifs dont la somme vaut* 1, *alors il existe une unique probabilité* $\mathbb{P}$ *sur* $(\Omega, \mathcal{P}(\Omega))$ *telle que, pour tout* $n \in \mathbb{N}$, $\mathbb{P}(\{\omega_n\}) = p_n$.

Dans ce cas, si A *est un événement, alors* $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \sum_{n \in \mathbb{N},\, \omega_n \in A} p_n$.

DÉMONSTRATION.

□

Remarque : Plus généralement, si I est une famille de $\mathbb{N}$ et $\Omega = \{\omega_i \,|\, i \in I\}$, alors une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est entièrement caractérisée par la donnée des $p_i = \mathbb{P}(\{\omega_i\})$, $i \in I$ (ceci se déduit du théorème en attribuant une probabilité nulle à tout entier naturel n'étant pas dans I).

Exemple : *A quelle condition sur $a \in \mathbb{R}$, existe-t-il une probabilité $\mathbb{P}$ sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ vérifiant, pour tout $n \in \mathbb{N}$, $\mathbb{P}(\{n\}) = \dfrac{a}{3^n}$?*

Terminons par un théorème qui souligne un constat : il n'y a pas d'équiprobabilité sur un univers dénombrable (cela découle immédiatement du théorème précédent).

Théorème 14. *Si Ω est dénombrable, alors il n'existe pas de probabilité uniforme sur $(\Omega, \mathcal{P}(\Omega))$.*

2) Propriété de la limite monotone

Définition 15. *Une famille $(A_n)_{n\in\mathbb{N}}$ d'événements est dite croissante (resp. décroissante) si, pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$ (resp. $A_{n+1} \subset A_n$).*

Théorème 16 (propriété de la limite monotone).

Si $(A_n)_{n\in\mathbb{N}}$ est une famille croissante d'événements, alors

Si $(A_n)_{n\in\mathbb{N}}$ est une famille décroissante d'événements, alors

DÉMONSTRATION.

□

L'utilisation de ce théorème requiert absolument que les familles considérées soient monotones.

Si on est en présence de l'union d'une famille composée d'un nombre dénombrable d'événements quelconques, alors on peut encore utiliser la propriété de la limite monotone de la façon suivante :

Théorème 17. *Soit $(A_n)_{n\in\mathbb{N}}$ une famille d'événements. On a*

et

Démonstration.

□

3) Événement négligeable, événement presque sûr

Définition 18. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilité. Un événement A est dit négligeable ou quasi-impossible si $\mathbb{P}(A) = 0$. On dit aussi A est $\mathbb{P}$-négligeable.*

On a $\mathbb{P}(\varnothing) = 0$. Par contre si A est un événement $\mathbb{P}$-négligeable, alors cela ne veut pas dire que $A = \varnothing$ (cf. exemple ci-dessous).

Définition 19. *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilité. Un événement A est dit presque-sûr ou quasi-certain si $\mathbb{P}(A) = 1$. On dit aussi que A est $\mathbb{P}$-presque sûr et on note $\mathbb{P}$-p.s.*

Si une propriété est vraie sur un ensemble de probabilité 1, alors on dit qu'elle est vraie $\mathbb{P}$-presque sûrement.

On a $\mathbb{P}(\Omega) = 1$. Par contre si A est un événement $\mathbb{P}$-presque sûr, alors cela ne veut pas dire que $A = \Omega$ (cf. exemple ci-dessous).

Exemple : *On dispose d'une pièce de monnaie que l'on lance successivement indéfiniment. On travaille avec l'univers $\Omega = \{0,1\}^{\mathbb{N}}$ qui est l'ensemble des suites ne prenant que 0 ou 1 pour valeurs de telle sorte qu'un événement élémentaire $\omega = (\omega_1, \dots, \omega_n, \dots) \in \Omega$ est tel que, pour tout $n \in \mathbb{N}$,*

$$\omega_n = \begin{cases} 1 & \text{si la pièce tombe sur Pile au } n^{\text{ème}} \text{ lancer,} \\ 0 & \text{si la pièce tombe sur Face au } n^{\text{ème}} \text{ lancer.} \end{cases}$$

On peut montrer que l'univers Ω est non dénombrable (il est même en bijection avec $\mathbb{R}$). On admet qu'il existe :

- *une tribu $\mathcal{A}$ sur Ω telle que, pour tout $n \in \mathbb{N}^*$, notons A_n : « obtenir Pile au $n^{\text{ème}}$ lancer » est un événement.*
- *une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ telle que, pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(A_n) = \dfrac{1}{2}$.*

IV Conditionnement et indépendance

Dans toute cette section $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé quelconque.

1) Probabilité conditionnelle

La définition de la probabilité conditionnelle d'un événement de $\mathcal{A}$ sachant un événement de $\mathcal{A}$ de probabilité non nulle est la même que dans le cas fini et toutes ses propriétés se généralisent.

Définition 20. *Soient A et B deux événements de $\mathcal{A}$ tels que $\mathbb{P}(A) \neq 0$. On appelle probabilité conditionnelle de B sachant A le réel*

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

On la note aussi $\mathbb{P}(B|A)$.

Remarques : Soient A et B deux événements tels que $\mathbb{P}(A) \neq 0$.

- Si $A \subset B$, alors $\mathbb{P}_A(B) = 1$. En particulier $\mathbb{P}_A(A) = 1$.
- Puisque $A \cap B \subset A$, nous avons $\mathbb{P}(A \cap B) \leqslant \mathbb{P}(A)$ et donc $\mathbb{P}_A(B) \in [0, 1]$.
- Le nombre $\mathbb{P}_A(B)$ représente la probabilité de A calculée du point de vue d'un observateur qui arriverait en cours d'expérience au moment où B vient de se réaliser. Il dispose de plus d'informations qu'un observateur qui assiste à l'expérience depuis son début : pour lui l'univers devient A.

Proposition 21. *Soit A un événement de $\mathcal{A}$ de probabilité non nulle. L'application $\begin{array}{rccc} \mathbb{P}_A & : & \mathcal{A} & \longrightarrow & \mathbb{R}_+ \\ & & B & \longmapsto & \mathbb{P}_A(B) \end{array}$ est une probabilité sur $(\Omega, \mathcal{A})$.*

Démonstration. Nous avons déjà vu que, pour tout $B \in \mathcal{A}$, $\mathbb{P}_A(B) \in [0, 1]$. Ensuite

- $\mathbb{P}_A(\Omega) = \dfrac{\mathbb{P}(A \cap \Omega)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$
- Si $(B_i)_{i \in I}$ est une famille d'événements deux à deux incompatibles de $\mathcal{A}$, indexée par une famille I de $\mathbb{N}$, alors

$$\mathbb{P}_A\left(\bigcup_{i \in I} B_i\right) = \mathbb{P}\left(\left(\bigcup_{i \in I} B_i\right) \cap A\right) = \mathbb{P}\left(\bigcup_{i \in I} (B_i \cap A)\right) = \sum_{i \in I} \mathbb{P}(B_i \cap A) = \sum_{i \in I} \mathbb{P}_A(B_i),$$

  car $\mathbb{P}$ est σ-additive et $(B_i \cap A)_{i \in I}$ est une famille d'événements deux à deux incompatibles de $\mathcal{A}$.

L'application $\mathbb{P}_A$ est bien une probabilité sur sur $(\Omega, \mathcal{A})$). □

Corollaire 22. *Si A, B et C sont trois événements de $\mathcal{A}$ tels que $\mathbb{P}(A) \neq 0$, alors*

$$\mathbb{P}_A(B \cup C) = \mathbb{P}_A(B) + \mathbb{P}_A(C) - \mathbb{P}_A(B \cap C).$$

Théorème 23 (formule des probabilités composées). *Soit $n \geqslant 2$ un entier naturel. Soient $A_1, \dots, A_n$ des événements tels que $\mathbb{P}(A_1 \cap \dots \cap A_n) \neq 0$. Alors*

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2)\mathbb{P}_{A_1 \cap A_2}(A_3) \times \dots \times \mathbb{P}_{A_1 \cap \dots \cap A_{n-1}}(A_n).$$

Démonstration. La démonstration est identique à celle vue au premier semestre dans le cas fini (en effet elle ne dépend pas de Ω). □

Théorème 24 (formule des probabilités totales). *Soit A un événement tel que $\mathbb{P}(A) \in]0, 1[$. Pour tout événement B, nous avons*

$$\mathbb{P}(B) = \mathbb{P}_A(B)\mathbb{P}(A) + \mathbb{P}_{\overline{A}}(B)\mathbb{P}(\overline{A}).$$

Démonstration. La démonstration est identique à celle vue au premier semestre dans le cas fini (en effet elle ne dépend pas de Ω). □

Généralisons la formule des probabilités totales pour un système complet d'événements.

Théorème 25. *Soit I une partie de $\mathbb{N}$. Soit $(A_i)_{i \in I}$ un système complet d'événements de $(\Omega, \mathcal{A}, \mathbb{P})$ tels que, pour tout $i \in I$, $\mathbb{P}(A_i) \neq 0$. Alors, pour tout événement B,*

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(A_i \cap B) = \sum_{i \in I} \mathbb{P}_{A_i}(B)\mathbb{P}(A_i).$$

Démonstration.

$\square$

Exemple : *Considérons un opérateur téléphonique d'un standard de SAV. On suppose que, en une heure, il reçoit un nombre $n \in \mathbb{N}$ d'appels avec probabilité $e^{-12}\dfrac{12^n}{n!}$. A chaque appel, il parvient à apporter une solution au problème de trois clients sur quatre (indépendamment des uns des autres). Quelle est la probabilité que l'opérateur résolve le problème d'exactement 8 clients dans l'heure qui suit ?*

On admet que l'expérience aléatoire peut être décrite à l'aide d'un espace probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

- Pour tout $n \in \mathbb{N}$, introduisons T_n l'événement « l'opérateur reçoit n appels pendant l'heure ». Par hypothèse $\mathbb{P}(T_n) = e^{-12}\dfrac{12^n}{n!}$.
- Pour tout $k \in \mathbb{N}$, notons R_k l'événement « l'opérateur résout k problèmes dans pendant l'heure ». Nous avons

Il aurait été plus adéquat de résoudre ce problème à l'aide variables aléatoires (mais c'est l'objet du paragraphe et des chapitres suivants).

Théorème 26 (formule de Bayes). *Soient A et B deux événements de $(\Omega, \mathcal{A}, \mathbb{P})$ qui sont tous les deux de probabilité non nulle. Alors*

$$\mathbb{P}_A(B) = \frac{\mathbb{P}_B(A)\mathbb{P}(B)}{\mathbb{P}(A)}.$$

Démonstration. Il suffit de remarquer que $\mathbb{P}_A(B)\mathbb{P}(A) = \mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}_B(A)\mathbb{P}(B)$. $\square$

De la formule de Bayes et la formule des probabilités totales nous déduisons que, si $(A_i)_{i \in I}$ est un système complet d'événements de $(\Omega, \mathcal{A}, \mathbb{P})$ tels que, pour tout $i \in I$, $\mathbb{P}(A_i) \neq 0$. Si B est un événement tel que $\mathbb{P}(B) \neq 0$, alors

$$\forall i \in I, \qquad \mathbb{P}_B(A_i) = \frac{\mathbb{P}_{A_i}(B)\mathbb{P}(B)}{\displaystyle\sum_{k \in I} \mathbb{P}_{A_k}(B)\mathbb{P}(A_k)}.$$

Exemple : *Reprenons l'expérience précédente. Quelle est la probabilité que l'opérateur ait reçu 10 appels pendant l'heure sachant qu'il a résolu 8 problèmes ?*

2) Indépendance

La définition de l'indépendance de deux événements de $\mathcal{A}$ est la même que dans le cas fini et toutes ses propriétés se généralisent.

Définition 27. *On dit que deux événements A et B de $(\Omega, \mathcal{A}, \mathbb{P})$ sont indépendants si*

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

En pratique l'indépendance ne se démontre pas. C'est une hypothèse de modélisation qui permet de simplifier les calculs. Typiquement si on effectue deux lancers successifs d'une pièce ou d'un dé (ou deux tirages successifs avec remise de boules dans une urne), alors on fera l'hypothèse que les deux lancers sont indépendants : tout événement concernant le résultat du premier lancer seul sera indépendant d'un événement concernant le résultat du deuxième lancer seul.

Les propositions suivant se démontrent de la même façon que dans le cas fini vu au premier semestre (la preuve ne dépend pas de Ω).

Proposition 28. *Soient A et B dans $\mathcal{A}$ tels que $\mathbb{P}(A) \neq 0$. Les événements A et B sont indépendants si et seulement si $\mathbb{P}_A(B) = \mathbb{P}(B)$.*

Proposition 29. *Si A et B sont des événements indépendants, alors*

- *A et $\overline{B}$ sont indépendants.*
- *$\overline{A}$ et B sont indépendants.*
- *$\overline{A}$ et $\overline{B}$ sont indépendants.*

Définition 30 (indépendance mutuelle). *Soit I une partie de $\mathbb{N}$. Soit $(A_i)_{i \in I}$ une famille d'événements de $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que les événements A_i, $i \in I$ sont (mutuellement) indépendants si, pour tout partie finie J de I,*

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

On n'écrit jamais de produit infini de probabilité. Si on se retrouve à devoir calculer la probabilité de $\mathbb{P}\left(\bigcap_{i \in \mathbb{N}} A_i\right)$, alors le seul résultat que l'on peut utiliser est la propriété de la limite monotone.

Exemple : *On dispose d'une urne contenant des boules rouges et des boules bleues. On tire une infinité de boules successivement et avec remise. Pour tout $n \in \mathbb{N}^*$, on note R_n l'événement « la $n^{\text{ème}}$ boule tirée est rouge ». Alors les événements R_n, $n \in \mathbb{N}^*$, sont mutuellement indépendants (on admet encore une fois l'existence d'une espace probabilisé adéquat).*

Théorème 31 (des coalitions). *Soit I une partie de $\mathbb{N}$. Si $(A_i)_{i \in I}$ est une famille d'événements mutuellement indépendants.*

1. *Si pour tout $i \in I$, $B_i = A_i$ ou $\overline{A}_i$, alors $(B_i)_{i \in I}$ est une famille d'événements mutuellement indépendants.*
2. *Si $J \subset I$, alors $(A_i)_{i \in J}$ est une famille d'événements mutuellement indépendants.*
3. *Si $J \subset I$, alors tout événement construit à partir de $(A_i)_{i \in J}$ est indépendant de tout événement construit à partir de $(A_i)_{i \notin J}$.*

$\rightsquigarrow$ Exercice.

V Variables aléatoires réelles

1) Définitions et exemples

Définition 32. *Soit $(\Omega, \mathcal{A})$ un espace probabilisable quelconque. Une variable aléatoire réelle sur Ω est une application $X : \Omega \longrightarrow \mathbb{R}$ telle que*

On appelle support (ou univers image) de X l'ensemble $X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$ des valeurs prises par la variable aléatoire X.

Bien sûr nous allons très vite être amenés à calculer la probabilité qu'une variable aléatoire prenne ses valeurs dans un intervalle de $\mathbb{R}$ (c'est-à-dire s'intéresser à sa loi). Cela implique de munir $(\Omega, \mathcal{A})$ d'une probabilité mais, comme on l'a évoqué, ce n'est pas possible en général lorsque $\mathcal{A} = \mathcal{P}(\Omega)$ avec Ω non-dénombrable. Pour contourner ce problème on se place sur une plus petite tribu mais alors se pose la difficulté de vérifier que les applications de Ω dans $\mathbb{R}$ sont bien des variables aléatoires réelles (c'est-à-dire, pour tout $x \in \mathbb{R}$, $[X \leqslant x] \in \mathcal{A}$). En première année d'ECS, on supposera toujours que les variables aléatoire réelles rencontrées sont bien des variables aléatoires...

Exemple : *On dispose d'un dé et on note X le nombre de lancers nécessaires pour obtenir 6 pour la première fois. On peut travailler avec $\Omega = [\![1, 6]\!]^{\mathbb{N}^*}$ et X est alors l'application qui à $(x_n)_{n\in\mathbb{N}}$ associe $\inf\{n \in \mathbb{N}^* \mid x_n = 6\}$. Il s'agit d'une variable aléatoire réelle*

> ... du moins si Ω est muni d'une tribu $\mathcal{A}$ adéquate. C'est le cas si $\mathcal{A} = \mathcal{P}(\Omega)$ mais la tribu discrète est trop grosse pour être envisagée dans ce cas, puisque Ω est non dénombrable. On admet qu'il existe une tribu plus adéquate telle que X est bien une variable aléatoire.

Le support de X est $X(\Omega) = \mathbb{N}^*$.

Rappelons les notations introduites au premier semestre :

Définition 33. *Soient $(a, b) \in \mathbb{R}^2$ tel que $a < b$ et X une variable aléatoire réelle sur Ω. On note*

$$[X = a] = \{\omega \in \Omega \mid X(\omega) = a\}, \qquad [X > a] = \{\omega \in \Omega \mid X(\omega) > a\},$$
$$[X < a] = \{\omega \in \Omega \mid X(\omega) < a\}, \qquad [X \geqslant a] = \{\omega \in \Omega \mid X(\omega) \geqslant a\},$$
$$[X \leqslant a] = \{\omega \in \Omega \mid X(\omega) \leqslant a\}, \qquad [a \leqslant X \leqslant b] = \{\omega \in \Omega \mid a \leqslant X(\omega) \leqslant b\},$$
$$\ldots$$

Plus généralement, si I est un intervalle de $\mathbb{R}$, alors on note $[X \in I] = \{\omega \in \Omega \mid X(\omega) \in I\}$.

Poursuivons par un petit lemme qui nous sera utile à plusieurs reprises :

Lemme 34. *Soient X une variable aléatoire réelle et $x \in \mathbb{R}$. On a* $\varnothing = \bigcap\limits_{n\in\mathbb{N}} [X \leqslant -n]$,

$$\Omega = \bigcup_{n\in\mathbb{N}} [X \leqslant n], \qquad [X \leqslant x] = \bigcap_{n\in\mathbb{N}^*} \left[X \leqslant x + \frac{1}{n}\right] \qquad \text{et} \qquad [X < x] = \bigcup_{n\in\mathbb{N}^*} \left[X \leqslant x - \frac{1}{n}\right].$$

Démonstration. Montrons le dernier point (les autres sont laissés en exercice).

□

Proposition 35. *Soit X une variable aléatoire réelle sur un espace probabilisable $(\Omega, \mathcal{A})$. Alors, pour tout intervalle I de $\mathbb{R}$, $[X \in I]$ est un événement.*

Démonstration.

□

2) Fonction de répartition d'une variable aléatoire réelle

Définition 36. *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé quelconque et une X une variable aléatoire réelle sur Ω. L'application* $\begin{array}{rccc} F_X & : & \mathbb{R} & \longrightarrow & [0,1] \\ & & x & \longmapsto & \mathbb{P}(X \leqslant x) \end{array}$ *est appelée la fonction de répartition de X.*

Proposition 37. *Si X est une variables aléatoire réelle, alors sa fonction de répartition F_X vérifie :*

1. *F_X est croissante sur $\mathbb{R}$.*
2. $\lim_{x \to -\infty} F_X(x) = 0$ *et* $\lim_{x \to +\infty} F_X(x) = 1$.
3. *F_X est continue à droite en tout point : pour tout $x \in \mathbb{R}$, $\lim_{y \to x^+} F_X(y) = F_X(x)$.*
4. *F_X admet une limite à gauche en tout point. Plus précisément :*

$$\forall x \in \mathbb{R}, \qquad \lim_{y \to x^-} F_X(y) = \mathbb{P}(X < x) = F_X(x) - \mathbb{P}(X = x).$$

5. *F_X est continue en un réel x si et seulement si $\mathbb{P}(X = x) = 0$.*

Démonstration.

□

3) Loi d'une variable aléatoire réelle

Définition 38. *Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé quelconque et une X une variable aléatoire réelle sur Ω. On appelle loi de X l'application*[1] $\mathbb{P}_X$ *qui à tout intervalle I de $\mathbb{R}$ associe la probabilité $\mathbb{P}(X \in I)$.*

Proposition 39. *Si X est une variables aléatoire réelle alors, pour tout $(a, b) \in \mathbb{R}^2$ tel que $a < b$, $\mathbb{P}(a < X \leqslant b) = \mathbb{F}_X(b) - F_X(a)$.*

Démonstration.

□

Théorème 40 (la fonction de répartition caractérise la loi). *Si X et Y sont deux variables aléatoires réelles ayant la même fonction de répartition, alors elles ont la même loi.*

Démonstration. Admis (mais la démonstration dans le cas dénombrable est tout à fait accessible et nous la verrons dans le chapitre suivant). □

Remarque : Ce résultat est fondamental : il suffit de connaître la probabilité qu'une variable aléatoire réelle X appartiennent à chaque intervalle de la forme $]-\infty, x]$, $x \in \mathbb{R}$, pour caractériser complètement sa loi.

Dans le cadre du programme, nous allons étudier deux types particuliers de variables aléatoires réelles :

- les v.a. discrètes, qui sont celles qui prennent un nombre au plus dénombrable de valeurs réelles. Elles font l'objet du chapitre suivant.
- les v.a. à densité, qui sont celles dont la fonction de répartition est continue sur $\mathbb{R}$ et de classe C^1 (sauf éventuellement en un nombre fini de points) de $\mathbb{R}$.

1. En fait la définition usuelle d'une variable aléatoire réelle est la suivante : On appelle tribu borélienne, la tribu sur $\mathbb{R}$, notée $\mathcal{B}(\mathbb{R})$, engendrée par les intervalles ouverts de $\mathbb{R}$. On appelle loi de probabilité d'une variable aléatoire X sur $(\Omega, \mathcal{A}, \mathbb{P})$ l'application $\begin{array}{cccc} \mathbb{P}_X & : & \mathcal{B}(\mathbb{R}) & \longrightarrow & [0,1] \\ & & A & \longmapsto & \mathbb{P}(X \in A) \end{array}$.