

THÉORIE DES PROBABILITÉS LES BASES (2025)

JUHAN ARU

1

¹Version 2025. Tout retour est apprécié, y compris pour les fautes (petites ou grandes) — juhan.aru@epfl.ch. Il s'agit d'une troisième version de ces notes. Pour la rédaction des versions précédentes, j'ai consulté les notes de I. Manolescu (Fribourg), Y. Velenik (Genève), A. Eberle (Bonn) (toutes disponibles sur leurs sites), ainsi que le livre de R. Dalang et D. Conus publié par EPFL Press.

Contents

	2
0 Introduction	3
1 Cadre de base	4
1.1 Espace probabilisé	4
1.2 Définition mathématique d'un espace de probabilité	6
1.3 Quelques propriétés de base des espaces de probabilité	8
1.4 Variables aléatoires	10
2 Probabilité conditionnelle et indépendance	14
2.1 Probabilité conditionnelle	14
2.2 Indépendance d'événements	17
2.3 Indépendance de variables aléatoires	19
2.4 Indépendance et produits d'espaces de probabilité	20
3 Variables aléatoires et vecteurs aléatoires	26
3.1 La fonction de répartition d'une variable aléatoire	26

SECTION 0

Introduction

Ce cours porte sur la théorie des probabilités : le cadre mathématique qui permet de formaliser nos questions sur les phénomènes aléatoires, et leur étude mathématique.

Lorsque nous souhaitons décrire un phénomène aléatoire du monde réel, nous construisons un modèle mathématique. Choisir un modèle utile — c'est-à-dire un modèle qui nous dit réellement quelque chose sur le monde — implique de nombreuses simplifications bien choisies et des décisions réfléchies. Par exemple, pour modéliser un lancer de pièce, nous écartons habituellement la possibilité qu'elle tombe sur la tranche, ou, en l'absence d'informations supplémentaires, nous considérons pile et face équiprobables, bien que cela puisse ne pas être le cas, ne serait-ce qu'à cause d'une répartition des masses différente. Mais choisir le modèle qui correspond le mieux au monde observé n'est pas le sujet central de ce cours.

Dans ce cours, nous nous concentrerons plutôt sur la mise en place du cadre mathématique général pour l'étude des phénomènes aléatoires, sur la formulation de modèles probabilistes, puis sur les outils mathématiques nécessaires et utiles pour étudier de tels modèles. Nous espérons aussi avoir un peu de temps pour discuter de modèles intéressants et pertinents.

SECTION 1

Cadre de base

Dans ce chapitre, nous discutons quelques notions fondamentales de la théorie des probabilités :

- Espace probabilisé
- Variables aléatoires
- Indépendance

1.1 Espace probabilisé

Notre premier objectif est de motiver la notion moderne d'espace de probabilité, ou de modèle probabiliste. Pour cela, considérons deux exemples :

- (1) Un nombre aléatoire à valeurs dans $\{1, 2, \dots, 12\}$, par exemple celui issu d'une loterie.
- (2) Décrire la météo à Lausanne le lendemain.

Pour décrire ces deux phénomènes aléatoires, nous utiliserons encore le vocabulaire et les intuitions de tous les jours. Ensuite, nous donnerons des définitions mathématiques qui fixeront le vocabulaire pour le reste du cours.

(1) Nombre aléatoire. Dans le cadre moderne, pour décrire mathématiquement un nombre aléatoire, nous utilisons trois ingrédients :

- L'ensemble de tous les résultats possibles : ici $\Omega = \{1, 2, 3, \dots, 12\}$.
- La collection de questions oui/non auxquelles on peut répondre à propos du résultat effectif, c'est-à-dire de ce nombre aléatoire. Par exemple :
 - Ce nombre est-il égal à 3 ?
 - Ce nombre est-il pair ?
 - Ce nombre est-il inférieur à 4 ?

À chacune de ces questions, nous associons le sous-ensemble des résultats pour lesquels la réponse est « oui » : respectivement $\{3\}$, $\{2, 4, 6, 8, 10, 12\}$ ou $\{1, 2, 3\}$. Nous appelons chacun de ces sous-ensembles un *événement*.

- Enfin, à chaque événement $E \subseteq \Omega$, nous souhaitons associer une valeur numérique $\mathbb{P}(E) \in [0, 1]$ que nous appelons *probabilité*. Elle doit correspondre à la proportion de fois où l'événement se produit si l'expérience est répétée un grand nombre de fois, par exemple si la loterie est rejouée de nombreuses fois. ²

Ici, l'ensemble des résultats possibles était facile et directement donné par le problème. Il est aussi naturel de supposer que tout sous-ensemble $E \subseteq \Omega$ est un événement — autrement dit, pour tout E , on peut poser la question : « le nombre est-il dans E ? ». Cela signifie que nous pouvons prendre la collection des événements comme étant l'ensemble de toutes les parties de Ω .

La détermination de la probabilité dépend véritablement de ce que nous voulons modéliser — par exemple, si nous essayons de modéliser la loterie, nous pouvons supposer que tous les nombres sont équiprobables ; nous retrouvons alors le modèle du lycée : on pose $\mathbb{P}(E) =$

²En réalité, on utilise aussi des modèles probabilistes pour décrire des phénomènes qui ne se produisent qu'une seule fois. Dans ce cas, la probabilité mesure en quelque sorte notre degré de croyance.

$|E|/|\Omega|$. En revanche, si nous voulions décrire la somme de deux dés, il faudrait choisir les nombres $\mathbb{P}(E)$ très différemment ! ³

Maintenant, si nous voulons que notre modèle corresponde à la notion intuitive de probabilité et prédise la fréquence lors d'expériences répétées, ces choix ne sont pas totalement libres — il faut ajouter certaines contraintes. Par exemple, nous ne pouvons pas définir une fonction \mathbb{P} arbitraire : en effet, si l'on a deux événements $E_1 \subseteq E_2$, alors on devrait avoir $\mathbb{P}(E_1) \leq \mathbb{P}(E_2)$, puisque chaque fois que E_1 se réalise, E_2 se réalise aussi. On devrait aussi avoir $\mathbb{P}(\Omega) = 1$, puisque « quelque chose » arrive toujours, et $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$ si E et F sont disjoints (pourquoi ?). Bien sûr, toutes ces contraintes ne sont pas indépendantes — certaines peuvent en impliquer d'autres ; lorsque nous donnerons la définition d'un espace de probabilité ci-dessous, nous purifierons ces conditions et n'en choisirons que quelques-unes qui impliqueront mathématiquement toutes les autres.

(2) Météo à Lausanne le lendemain. Dans le cadre moderne que nous allons définir, nous voudrions à nouveau prendre trois décisions, mais ici la tâche est déjà plus difficile dès la première étape. Quel devrait être l'espace d'états ? Un espace naturel pourrait être l'ensemble de tous les états microscopiques possibles de l'atmosphère jusqu'à 20 km d'altitude au-dessus de Lausanne... mais nous avons là de nombreux choix arbitraires — pourquoi 20 km, quelle largeur au-dessus du Léman, etc. ? Et, de toute façon, tout état « naturel » serait d'une complexité impossible !

Heureusement, nous n'avons pas vraiment besoin de nous en préoccuper — dans notre cadre, nous avons seulement à assigner des probabilités à tous les événements de la collection d'événements choisie ! Et nous avons une certaine liberté pour choisir cette collection d'événements — elle peut être déterminée par notre capacité à mesurer les états, par exemple nous pouvons mesurer la température à une certaine précision, ou la densité de molécules de CO_2 ou d'eau à une certaine précision ; cela détermine des sous-ensembles de l'espace d'états. De plus, à mesure que nous nous améliorons dans la mesure, nous pouvons toujours agrandir notre modèle de manière incrémentale !

Cependant, comme pour la fonction de probabilité, il existe aussi des conditions naturelles de cohérence pour la collection d'événements : nous supposons que si l'on peut observer si l'événement E s'est produit, nous devrions aussi être capables de mesurer si son complément E^c s'est produit. Ou encore, si nous sommes capables de dire si E est arrivé, ou si F est arrivé, nous devrions pouvoir dire si l'un des deux est arrivé — c'est-à-dire que $E \cup F$ devrait aussi être un événement. Et il s'avère que c'est tout ce dont nous avons besoin !

Naturellement, la mise en place des probabilités pour ce modèle est aussi horriblement compliquée — il n'y a pas d'hypothèses naturelles de symétrie comme celle utilisée pour la distribution uniforme. Même le meilleur physicien du monde ne sera pas capable de décrire la distribution de probabilité naturelle de tous les états microscopiques de l'atmosphère, d'autant plus qu'elle dépend fortement de ce qui s'est passé juste avant ! Ainsi, notre seul choix consiste essentiellement à essayer d'utiliser une combinaison de nos connaissances sur les processus atmosphériques et de nos observations passées pour établir des estimations pour le modèle ; puis naturellement nous essaierons de l'améliorer chaque jour suivant. Heureusement, cette tâche difficile ne nous incombe pas, mais revient aux services de météo et aux statisticiens !

³Voir feuille d'exercices 1.

Remarque 1.1. *Enfin, avant de donner les définitions mathématiques, soulignons à nouveau que les trois composants du modèle — l'espace d'échantillonnage, l'ensemble des événements et leurs probabilités — sont des entrées que nous choisissons pour construire notre modèle. Lorsqu'on tente de modéliser un phénomène réel, on effectue généralement des simplifications pour chacun de ces choix. Par exemple, pour le lancer de pièce, nous n'utilisons que deux issues : pile et face, même si théoriquement la tranche est possible. Nous fixons aussi habituellement les probabilités à un demi, bien que cela ne soit pas exactement vrai non plus.*

1.2 Définition mathématique d'un espace de probabilité

Nous sommes maintenant prêts à passer par notre filtre mathématique et à donner une définition mathématique d'un espace de probabilité. En fait, nous commençons par utiliser le purificateur mathématique pour présenter une définition dans le cadre restreint où Ω est un ensemble fini, puis nous la généralisons.

Définition 1.2 (Espace de probabilité fini, Kolmogorov 1933). *Un espace de probabilité fini est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où*

- Ω est un ensemble fini, appelé l'espace d'états, espace d'échantillonnage ou l'univers.
- \mathcal{F} est un ensemble de sous-ensembles de Ω , satisfaisant :
 - $\emptyset \in \mathcal{F}$;
 - si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$;
 - si $A_1, A_2 \in \mathcal{F}$, alors aussi $A_1 \cup A_2 \in \mathcal{F}$.

\mathcal{F} est appelée la collection d'événements et tout $A \in \mathcal{F}$ est un événement.

- Enfin, on a une fonction $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfaisant $\mathbb{P}(\Omega) = 1$ et l'additivité pour les ensembles disjoints : si $A_1, A_2 \in \mathcal{F}$ sont disjoints, alors

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Cette fonction \mathbb{P} est appelée la probabilité.

Remarquez que certaines propriétés évoquées plus haut, comme le fait que pour des événements $E_1 \subseteq E_2$, on ait $\mathbb{P}(E_1) \leq \mathbb{P}(E_2)$, découlent directement de la définition.⁴

La plupart des phénomènes du monde réel peuvent être décrits par des ensembles finis simplement parce que nous ne pouvons mesurer les choses qu'à un niveau de précision fini. Cependant, de la même façon que la notion de fonction continue ou dérivable permet de simplifier nos descriptions mathématiques de la réalité et donc d'améliorer notre compréhension, les espaces de probabilité continus rendent aussi les descriptions mathématiques plus nettes, plus simples et, partant, facilitent l'étude des phénomènes aléatoires sous-jacents.

Quelques exemples naturels où des espaces d'échantillonnage infinis apparaissent :

- un point uniforme sur un segment (par exemple issu du bris d'un bâton en plusieurs morceaux) ;
- la position dans la rue où tombe la première goutte de pluie de la journée ;
- l'espace de toutes les suites infinies de lancers de pièce.

Dans tous ces cas, l'espace d'états mathématiquement naturel est même non dénombrable. Des espaces d'états dénombrables apparaissent aussi : par exemple, si l'on modélise le premier instant où des lancers répétés donnent « pile », la valeur peut être 1, 2, 3 ou, avec une

⁴Voir feuille d'exercices 1.

probabilité très, très petite, aussi 10^{10} ; un espace d'états naturel contiendrait donc tous les entiers naturels.

Énonçons donc la définition générale :

Définition 1.3 (Espace de probabilité, Kolmogorov 1933). *Un espace de probabilité est un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, où*

- Ω est un ensemble, appelé l'espace d'états, espace d'échantillonnage ou l'univers.
- \mathcal{F} est un ensemble de sous-ensembles de Ω , satisfaisant :
 - $\emptyset \in \mathcal{F}$;
 - si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$;
 - si $A_1, A_2, \dots \in \mathcal{F}$, alors $\bigcup_{n \geq 1} A_n \in \mathcal{F}$.

\mathcal{F} est appelée la collection d'événements, ou une σ -algèbre, et tout $A \in \mathcal{F}$ est un événement.

- Enfin, on a une fonction $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ satisfaisant $\mathbb{P}(\Omega) = 1$ et l'additivité dénombrable pour des ensembles disjoints : si $A_1, A_2, \dots \in \mathcal{F}$ sont deux à deux disjoints,

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n).$$

Cette fonction \mathbb{P} est appelée la probabilité.

Notez les seules différences : 1) nous ne supposons pas Ω fini ; 2) nous supposons que l'ensemble des événements est stable par unions dénombrables ; 3) nous supposons aussi l'additivité de la probabilité sous unions dénombrables.

Exercice 1.1. *Montrer que tout espace de probabilité « élémentaire » est un espace de probabilité.*

En fait, les espaces de probabilité sont un exemple de la notion générale d'espaces mesurés — les espaces de probabilité ne sont que des espaces mesurés de masse totale égale à 1.

Définition 1.4 (Espace mesuré, Borel 1898, Lebesgue 1901–1903). *Un espace mesuré est un triplet $(\Omega, \mathcal{F}, \mu)$, où*

- Ω est un ensemble, appelé l'espace d'échantillonnage ou l'univers.
- \mathcal{F} est un ensemble de sous-ensembles de Ω , satisfaisant :
 - $\emptyset \in \mathcal{F}$;
 - si $A \in \mathcal{F}$, alors $A^c \in \mathcal{F}$;
 - si $A_1, A_2, \dots \in \mathcal{F}$, alors $\bigcup_{n \geq 1} A_n \in \mathcal{F}$.

\mathcal{F} est appelée une σ -algèbre et tout $A \in \mathcal{F}$ un ensemble mesurable.

- Enfin, on a une fonction $\mu : \mathcal{F} \rightarrow [0, \infty]$ satisfaisant $\mu(\emptyset) = 0$ et l'additivité dénombrable pour des ensembles disjoints : si $A_1, A_2, \dots \in \mathcal{F}$ sont deux à deux disjoints,

$$\mu\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n).$$

Cette fonction μ est appelée une mesure. Si $\mu(\Omega) < \infty$, on appelle μ une mesure finie.

Géométriquement, nous interprétons :

- Ω comme notre espace de points,

- \mathcal{F} comme la collection des sous-ensembles pour lesquels notre notion de volume peut être définie,
- μ comme notre notion de volume : elle donne à chaque ensemble mesurable son volume.

Il est important d'établir ce lien avec la théorie de la mesure, car nombre de propriétés des espaces de probabilité en découlent directement. Il est toutefois bon de garder à l'esprit que la théorie des probabilités n'est pas *que* de la théorie de la mesure — comme l'a bien formulé M. Kac, « La probabilité, c'est la théorie de la mesure avec une âme », et nous adhérons à cette remarque philosophique.

Remarque 1.5. *Vous devriez comparer la définition d'un espace de probabilité / espace mesuré avec celle d'un espace topologique : là aussi, on utilise une collection de sous-ensembles, satisfaisant certaines propriétés, pour doter un ensemble d'une structure. Une question à se poser : pourquoi exigeons-nous exactement des unions et intersections dénombrables pour les événements, et non pas seulement finies, ou bien arbitraires ?*

1.3 Quelques propriétés de base des espaces de probabilité

Commençons par quelques remarques sur la définition d'un espace de probabilité :

Remarque 1.6. *Il vaut la peine de réfléchir aux raisons pour lesquelles on demande la stabilité dénombrable de la σ -algèbre ou l'additivité dénombrable de la probabilité. Même s'il s'agit davantage d'une question méta-mathématique, il est bon de la garder à l'esprit tout au long du cours. Contentons-nous ici de deux observations simples.*

Premièrement, les sommes dénombrables apparaissent naturellement lorsqu'on prend des limites de sommes finies. En fait, l'additivité dénombrable peut être vue comme équivalente à une certaine forme de continuité pour la probabilité (voir ci-dessous).

Deuxièmement, autoriser des unions arbitraires mène rapidement aux ensembles de parties, et des sommes non dénombrables de termes positifs ne peuvent pas être finies (voir la feuille d'exercices).

Exercice 1.2. *Montrer que l'additivité dénombrable dans les axiomes d'un espace de probabilité peut être remplacée par l'additivité finie plus l'énoncé suivant : pour toute suite décroissante d'événements $E_1 \supseteq E_2 \supseteq E_3 \dots$ telle que $\bigcap_{i \geq 1} E_i = \emptyset$, on a $\mathbb{P}(\bigcap_{i=1}^n E_i) \rightarrow 0$ lorsque $n \rightarrow \infty$.*

★ *Cela vaut-il dans un espace mesuré général ?*

Considérons aussi un autre cadre qui explique bien l'utilité des σ -algèbres :

Remarque 1.7. *Dans la vie réelle, nous n'obtenons souvent des informations sur le monde que pas à pas, et si nous souhaitons continuer à travailler sur le même espace de probabilité (ce qui est avantageux car alors \mathbb{P} n'aura besoin que d'être étendue et non redéfinie), nous pouvons considérer une suite de σ -algèbres $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_3 \dots$ appelée filtration — chaque jour, nous pouvons poser davantage de questions oui/non, parce que, par exemple, nous savons déjà ce qui s'est passé la veille et avons peut-être appris quelque chose de nouveau. Toute l'information possible est contenue dans l'ensemble des parties $\mathcal{P}(\Omega)$.*

On classe habituellement les espaces de probabilité en deux types :

Définition 1.8 (Espaces de probabilité discrets et continus). *Les espaces de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ dont l'espace d'états Ω est dénombrable sont appelés discrets, et ceux dont Ω est non dénombrable sont appelés continus.*

Dans ce cours, nous travaillerons principalement avec des espaces discrets, car ils sont techniquement plus simples. Cependant, les espaces continus apparaissent naturellement et nous ne pourrions pas non plus les éviter totalement.

Leur différence technique peut être résumée dans la proposition suivante, dont la preuve (non exigible) est laissée aux enthousiastes.

Proposition 1.9. *Soit Ω dénombrable et \mathcal{F} une σ -algèbre sur Ω . Alors on peut trouver des événements disjoints $E_1, E_2, \dots \in \mathcal{F}$ tels que, pour tout $E \in \mathcal{F}$, on ait la décomposition $E = \cup_{i \in I_E} E_i$.*

Essentiellement, cela dit que, pour tout espace de probabilité discret, il suffit de déterminer $\mathbb{P}(E_i)$ pour une collection dénombrable d'ensembles disjoints E_i , puis, pour tout autre ensemble E , on peut utiliser l'additivité dénombrable pour étendre \mathbb{P} . Notez que cela signifie qu'il est d'abord facile de vérifier si une fonction \mathbb{P} donnée satisfait bien tous les axiomes et, plus important encore, il est facile de vérifier quand deux mesures de probabilité sont égales.

Pour les espaces continus, cela n'est pas nécessairement vrai — les σ -algèbres utiles sont généralement plus compliquées. Pour illustrer pourquoi on ne souhaite pas forcément utiliser l'ensemble des parties, considérons la proposition suivante (sa preuve est en annexe et repose sur l'axiome du choix) :

Proposition 1.10. *Il n'existe pas de probabilité \mathbb{P} sur $([0, 1], \mathcal{P}([0, 1]))$ invariante par translation, c'est-à-dire telle que pour tout $A \in \mathcal{P}([0, 1])$ et $\alpha \in [0, 1)$, on ait $\mathbb{P}(A + \alpha \bmod 1) = \mathbb{P}(A)$, où $A + \alpha \bmod 1 := \{a + \alpha \bmod 1 : a \in A\}$ désigne l'ensemble obtenu en décalant A de α modulo 1.*

En fait, il s'avère que la seule manière de remédier à cette situation consiste à rendre la σ -algèbre pertinente plus petite. Nous souhaitons encore pouvoir répondre « oui » ou « non » à des questions telles que : « mon nombre aléatoire est-il égal à x ? » ou « est-il dans un intervalle (a, b) ? ». Grâce au fait que nous ne disposons que de l'additivité dénombrable, cela n'implique pas que notre σ -algèbre doive être l'ensemble des parties. Et grâce aux propriétés des σ -algèbres, on peut toujours construire au moins une σ -algèbre contenant tous nos ensembles favoris — voir la feuille d'exercices.

Énonçons maintenant quelques conséquences immédiates des définitions sur les σ -algèbres et les mesures de probabilité :

Lemme 1.11 (Stabilité de la σ -algèbre). *Considérons un ensemble Ω muni d'une σ -algèbre \mathcal{F} .*

- (1) *Si $A_1, A_2, \dots \in \mathcal{F}$, alors $\bigcap_{n \geq 1} A_n \in \mathcal{F}$.*
- (2) *On a aussi $\Omega \in \mathcal{F}$ et, si $A, B \in \mathcal{F}$, alors $A \setminus B \in \mathcal{F}$.*
- (3) *Pour tout $n \geq 1$, si $A_1, \dots, A_n \in \mathcal{F}$, alors $A_1 \cup \dots \cup A_n \in \mathcal{F}$ et $A_1 \cap \dots \cap A_n \in \mathcal{F}$.*

Preuve du Lemme 1.11. Par les lois de De Morgan, pour toute famille $(A_i)_{i \in I}$,

$$\bigcap_{i \in I} A_i = \left(\bigcup_{i \in I} A_i^c \right)^c.$$

Le point (1) en découle : si $A_1, A_2, \dots \in \mathcal{F}$, alors, par définition d'une σ -algèbre, $A_1^c, A_2^c, \dots \in \mathcal{F}$ et donc

$$\left(\bigcup_{i \geq 1} A_i^c\right)^c \in \mathcal{F}.$$

Pour (3), à nouveau par De Morgan, il suffit de montrer que $A_1 \cup \dots \cup A_n \in \mathcal{F}$. Mais cela découle de la définition d'une σ -algèbre, car $A_1 \cup \dots \cup A_n = \bigcup_{i \geq 1} A_i$ avec $A_k = \emptyset$ pour $k \geq n + 1$.

Le point (2) est laissé en exercice. \square

De même, les conditions de base sur la mesure donnent lieu à plusieurs propriétés naturelles :

Proposition 1.12 (Propriétés de base d'une probabilité). *Considérons un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Soient $A_1, A_2, \dots \in \mathcal{F}$. Alors*

(1) Pour tout $A \in \mathcal{F}$, on a $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

(2) Pour tout $n \geq 1$, et A_1, \dots, A_n disjoints, on a l'additivité finie :

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

En particulier, si $A_1 \subseteq A_2$, alors $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$.

(3) Si, pour tout $n \geq 1$, $A_n \subseteq A_{n+1}$, alors, quand $n \rightarrow \infty$, on a $\mathbb{P}(A_n) \rightarrow \mathbb{P}(\bigcup_{k \geq 1} A_k)$.

(4) On a la sous-additivité dénombrable (ou union bound) : $\mathbb{P}(\bigcup_{n \geq 1} A_n) \leq \sum_{n \geq 1} \mathbb{P}(A_n)$.

(5) Si, pour tout $n \geq 1$, $A_n \supseteq A_{n+1}$, alors, quand $n \rightarrow \infty$, on a $\mathbb{P}(A_n) \rightarrow \mathbb{P}(\bigcap_{k \geq 1} A_k)$.

Proof. Les propriétés (1), (4) et la seconde partie de (2) figurent sur la feuille d'exercices 1. La première partie de (2) suit comme dans le lemme ci-dessus en prenant $A_{n+1} = A_{n+2} = \dots = \emptyset$ et en utilisant l'additivité dénombrable.

Prouvons (3) : posez $B_1 = A_1$ et, pour $n \geq 2$, $B_n = A_n \setminus A_{n-1}$. Alors les B_n sont disjoints, $\bigcup_{n=1}^N B_n = A_N$ et $\bigcup_{n \geq 1} B_n = \bigcup_{n \geq 1} A_n$.

Par additivité dénombrable,

$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \mathbb{P}\left(\bigcup_{i \geq 1} B_i\right) = \sum_{i \geq 1} \mathbb{P}(B_i).$$

Mais \mathbb{P} est positive, donc

$$\sum_{i \geq 1} \mathbb{P}(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(B_i).$$

Par additivité dénombrable encore,

$$\sum_{i=1}^n \mathbb{P}(B_i) = \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) = \mathbb{P}(A_n),$$

d'où (3). \square

1.4 Variables aléatoires

En réalité, lorsque nous étudions un phénomène aléatoire, nous ne voulons certainement pas nous restreindre à des questions oui/non. Par exemple, dans notre modèle d'un nombre aléatoire parmi $\{1, 2, \dots, 12\}$, la question naturelle n'est pas « est-ce que ce nombre est 5 ? » mais plutôt « quel est ce nombre ? ». De même, dans l'exemple sur la météo, il est plus

naturel de demander « quelle est la température ? », « quelle quantité de pluie y aura-t-il l'après-midi ? ».

Ces observations numériques sur notre phénomène aléatoire seront formalisées sous le nom de *variables aléatoires*. En substance, elles associent un nombre à chaque état et sont donc des fonctions $X : \Omega \rightarrow \mathbb{R}$. Toutefois, nous ne souhaitons pas forcément autoriser toutes ces fonctions pour des raisons de cohérence. En effet, nous voulons pouvoir poser des questions oui/non à propos de nos quantités aléatoires, par exemple « la variable vaut-elle 3 ? », « la température dépasse-t-elle 18 ? ». Or, la réponse oui/non correspond à certains sous-ensembles d'états dans l'univers et, en tant que tels, ces sous-ensembles devraient être des événements de notre modèle. Il existe donc un lien entre la collection d'événements et la collection de fonctions pouvant jouer le rôle de variables aléatoires. Donnons sans plus tarder la définition générale :

Définition 1.13 (Variable aléatoire). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Une fonction $X : \Omega \rightarrow \mathbb{R}$ est appelée une variable aléatoire si, pour tout intervalle (a, b) , l'ensemble $X^{-1}((a, b)) := \{\omega \in \Omega : X(\omega) \in (a, b)\}$ est un événement, c'est-à-dire appartient à \mathcal{F} .*

Il y a une simplification dans le cas des espaces de probabilité discrets :

Lemme 1.14 (Variables aléatoires sur les espaces discrets). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité discret. Alors $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire si et seulement si, pour tout $y \in \mathbb{R}$, on a $X^{-1}(\{y\}) \in \mathcal{F}$.*

Proof. Ceci se vérifie soigneusement à partir des définitions et figurera sur la feuille d'exercices. □

Pour les esprits « structurés », la définition d'une variable aléatoire peut paraître quelque peu arbitraire. En effet, j'ai caché une pièce d'information — la collection naturelle d'événements sur \mathbb{R} esquissée un peu plus haut. Nous l'énoncerons directement sur \mathbb{R}^n .

Définition 1.15 (σ -algèbre borélienne). *La plus petite σ -algèbre sur \mathbb{R}^n contenant toutes les boîtes ouvertes de la forme $(a_1, b_1) \times \cdots \times (a_n, b_n)$ est appelée la σ -algèbre borélienne. On la note \mathcal{F}_B .*

Remarque 1.16. *Cette définition est en fait encore plus générale : étant donné un espace topologique (X, τ) , la plus petite σ -algèbre contenant tous les ouverts est appelée σ -algèbre borélienne. Vous verrez sur la feuille d'exercices que cette définition plus générale se réduit à la précédente dans le cas de \mathbb{R}^n muni de sa topologie euclidienne.*

Sur cette base, une définition équivalente, peut-être plus « structurelle », d'une variable aléatoire est la suivante : une fonction $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire si le préimage de tout borélien de \mathbb{R} par X est un événement.⁵

Une notion importante associée aux variables aléatoires est leur *loi* :

Lemme 1.17 (Loi d'une variable aléatoire). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire.*

Alors il existe une probabilité \mathbb{P}_X induite sur $(\mathbb{R}, \mathcal{F}_B)$ en définissant $\mathbb{P}_X(F) := \mathbb{P}(X^{-1}(F))$ pour tout $F \in \mathcal{F}_B$. Cette probabilité \mathbb{P}_X est appelée la loi (ou distribution) de la variable aléatoire X .

⁵En théorie de la mesure, de telles fonctions sont appelées *mesurables* de (Ω, \mathcal{F}) vers $(\mathbb{R}, \mathcal{F}_B)$; notez la similarité avec la définition de continuité en topologie.

C'est un lemme et non une simple définition, car il faut prouver que \mathbb{P}_X est bien une probabilité sur $(\mathbb{R}, \mathcal{F}_B)$.

Preuve du Lemme. Il faut vérifier les axiomes d'une probabilité :

- $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(\Omega) = 1$;
- de même, $\mathbb{P}_X(F) = \mathbb{P}(X^{-1}(F)) \in [0, 1]$ pour tout $F \in \mathcal{F}_B$;
- enfin, l'additivité dénombrable : si F_1, F_2, \dots sont disjoints dans \mathcal{F}_B , alors

$$\mathbb{P}_X\left(\bigcup_{i \geq 1} F_i\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{i \geq 1} F_i\right)\right) = \mathbb{P}\left(\bigcup_{i \geq 1} X^{-1}(F_i)\right) = \sum_{i \geq 1} \mathbb{P}(X^{-1}(F_i)) = \sum_{i \geq 1} \mathbb{P}_X(F_i).$$

Ici, nous avons utilisé la définition aux première et dernière égalités, les propriétés des préimages à la seconde, et le fait que les $X^{-1}(F_i)$ sont disjoints, ainsi que l'additivité dénombrable, à la troisième.

□

En termes simples, nous avons montré que chaque variable aléatoire X induit une probabilité sur les réels en oubliant tout le contexte et en nous concentrant uniquement sur le nombre observé. Par exemple, dans le cas de la météo à Lausanne, la température fournit une variable aléatoire et, en ne regardant que sa valeur et rien d'autre, nous obtenons simplement un nombre réel aléatoire. Plus simplement encore, si nous lançons deux pièces équilibrées et comptons le nombre de piles, leur somme fournit une variable aléatoire à valeurs dans $\{0, 1, 2\}$. La notion de loi d'une variable aléatoire nous donne ainsi un moyen de comparer des quantités aléatoires provenant de contextes très différents.

Définition 1.18 (Égalité en loi). *Soient X, Y deux variables aléatoires définies éventuellement sur des espaces de probabilité différents. Nous disons que X et Y sont égales en loi (ou égales en distribution), et nous notons $X \sim Y$, si pour tout $E \in \mathcal{F}_B$, on a $\mathbb{P}_X(E) = \mathbb{P}_Y(E)$.*

Nous soulignons que, lorsqu'on regarde la loi d'une variable aléatoire, le contexte est oublié — nous nous concentrons seulement sur la valeur numérique et l'espace de probabilité initial $(\Omega, \mathcal{F}, \mathbb{P})$ ne sert qu'à déterminer \mathbb{P}_X mais ne joue plus de rôle par la suite. Cela permet de relier entre elles différentes situations aléatoires. Par exemple, les fonctions indicatrices de tout événement de probabilité p , quel que soit l'espace de probabilité où elles ont été définies, ont la même loi. Plus concrètement, par exemple, les variables aléatoires suivantes ont la même loi :

- le nombre de piles dans deux lancers indépendants ;
- le nombre de facteurs premiers lorsque l'on choisit uniformément un nombre dans $\{1, 2, 3, 4\}$.

Dans une certaine mesure, une grande partie de ce cours consistera à étudier et décrire les lois de variables aléatoires.

Il existe également d'autres notions d'égalité pour des variables aléatoires :

- Nous disons que deux variables X, Y définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$ sont *partout égales* si, pour tout $\omega \in \Omega$, $X(\omega) = Y(\omega)$.
- Nous disons que deux variables X, Y définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$ sont *presque sûrement égales* si $\mathbb{P}(\{\omega : X(\omega) = Y(\omega)\}) = 1$. Ici, il faut bien sûr d'abord montrer que $\{\omega : X(\omega) = Y(\omega)\}$ est un événement, c'est-à-dire appartient à \mathcal{F} — cela figure sur la feuille d'exercices.

Clairement, l'égalité partout implique l'égalité presque sûre, et la réciproque est fautive — par exemple, considérez les variables $X(\omega) := \omega \mathbf{1}_{\omega \neq 1/2}$ et $Y(\omega) := \omega$ sur l'espace $([0, 1], \mathcal{F}_B, \mathbb{P}_U)$. Il est aussi clair que l'égalité en loi ne peut pas impliquer l'égalité p.s., même si les variables sont définies sur le même espace. Enfin,

Exercice 1.3. *Soient X, Y deux variables aléatoires définies sur le même espace de probabilité et presque sûrement égales. Alors elles sont aussi égales en loi.*

SECTION 2

Probabilité conditionnelle et indépendance

De manière générale, si nous apprenons quelque chose de nouveau sur notre phénomène aléatoire, cette connaissance influence — et modifie souvent — nos prédictions pour le reste du modèle.

- Par exemple, dans le cas d'un nombre uniformément aléatoire entre 1 et 12, si quelqu'un vous dit que ce nombre est pair, alors la probabilité d'observer 1 devient soudain 0, tandis que la probabilité d'observer 2 passe de $1/12$ à $1/6$.
- Dans le cas de la météo à Lausanne, si l'on vous dit qu'il pleut toute la journée, alors il est moins probable qu'il fasse aussi plus de 35 degrés.

Le but de cette section est de fixer le vocabulaire permettant de parler de la manière dont la connaissance d'un événement ou d'une variable aléatoire influence les probabilités que nous devrions attribuer à d'autres événements. Cela nous conduit à parler de probabilités conditionnelles, puis à discuter du cas où les événements ne s'influencent pas, donnant naissance à une notion importante en théorie des probabilités : l'indépendance.

2.1 Probabilité conditionnelle

Nous avons déjà considéré (dans le cours et sur les feuilles d'exercices) de nombreuses situations imprévisibles où plusieurs événements se produisent naturellement soit en même temps, soit successivement : une suite de lancers de pièce ou des pas successifs d'une marche aléatoire, ou encore des liens/arrêtes différents dans un graphe aléatoire. Dans tous ces cas, le fait qu'un événement se soit produit peut facilement influencer les autres. Par exemple, si vous souhaitez modéliser les marchés financiers de demain, il semble plutôt avisé de tenir compte de ce qui s'est passé aujourd'hui. Pour parler du changement des probabilités lorsqu'on a observé quelque chose, on introduit la notion de probabilité conditionnelle :

Définition 2.1 (Probabilité conditionnelle). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $E \in \mathcal{F}$ avec $\mathbb{P}(E) > 0$. Alors, pour tout $F \in \mathcal{F}$, on définit la probabilité conditionnelle de l'événement F sachant E (c.-à-d. sachant que l'événement E a lieu) par*

$$\mathbb{P}(F|E) := \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)}.$$

Rappelez-vous que $E \cap F$ est l'événement « E et F se produisent ». Comme le dénominateur est toujours $\mathbb{P}(E)$, la probabilité conditionnelle sachant E est proportionnelle à $\mathbb{P}(E \cap F)$ pour tout événement F . Voici la justification de la division par $\mathbb{P}(E)$:

Lemme 2.2. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $E \in \mathcal{F}$ avec $\mathbb{P}(E) > 0$. Alors $\mathbb{P}(\cdot|E)$ définit une probabilité sur (Ω, \mathcal{F}) , appelée la probabilité conditionnelle sachant E .*

Proof. Tout d'abord, remarquez que $\mathbb{P}(\cdot|E)$ est bien définie pour tout $F \in \mathcal{F}$. Ensuite, $\mathbb{P}(\emptyset|E) = \mathbb{P}(\emptyset)/\mathbb{P}(E) = 0$ et $\mathbb{P}(\Omega|E) = \mathbb{P}(\Omega)/\mathbb{P}(E) = 1$. Il reste à vérifier l'additivité dénombrable.

Soient $F_1, F_2, \dots \in \mathcal{F}$ disjoints. Alors $E \cap F_1, E \cap F_2, \dots$ sont aussi disjoints. D'où

$$\mathbb{P}\left(\bigcup_{i \geq 1} F_i \mid E\right) = \frac{\mathbb{P}((\bigcup_{i \geq 1} F_i) \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(\bigcup_{i \geq 1} (F_i \cap E))}{\mathbb{P}(E)} = \sum_{i \geq 1} \frac{\mathbb{P}(F_i \cap E)}{\mathbb{P}(E)} = \sum_{i \geq 1} \mathbb{P}(F_i | E),$$

ce qui donne l'additivité dénombrable. \square

On notera que la probabilité conditionnelle d'un événement peut parfois être proche de la probabilité initiale (nous y reviendrons très vite), mais elle peut aussi être très différente. Un exemple un peu « bête », mais instructif, est le suivant :

- La probabilité conditionnelle de E^c , conditionnée par E , est toujours nulle, quelle que soit la probabilité initiale ;
- de même, la probabilité conditionnelle de E , conditionnée par E , est toujours 1.

Ou, pour un exercice plus sensé, considérez ce qui suit :

Exercice 2.1 (Marche aléatoire et probabilités conditionnelles). *Considérez la marche aléatoire simple de longueur n .*

- *Quelle est la probabilité que la marche soit au point n au temps n ? Maintenant, supposez que le premier pas ait été -1 . Quelle est alors la probabilité d'être au point n au temps n ?*
- *Supposons n pair. Quelle est la probabilité que la marche soit au point 0 au temps n ? Maintenant, supposez que le premier pas ait été -1 . Quelle est alors la probabilité d'être au point 0 au temps n ?*

Il faut aussi être très attentif au conditionnement exact, car deux conditionnements qui « se ressemblent » peuvent induire des probabilités conditionnelles très différentes. En général, il nous faut des informations supplémentaires sur la relation entre deux événements pour savoir comment la probabilité de l'un change quand on conditionne par l'autre.

Il existe néanmoins des cas où ces relations, et donc les probabilités conditionnelles, sont simples :

- Lorsque $E \subseteq F$, la probabilité conditionnelle de F sachant E vaut 1.
- Lorsque $F \subseteq E^c$, la probabilité conditionnelle de F sachant E vaut 0.
- Le troisième cas est lorsque F et E sont dits indépendants : dans ce cas $\mathbb{P}(F|E) = \mathbb{P}(F)$, essentiellement par définition (nous y revenons).

En général, il n'existe pas tant d'outils pour calculer des probabilités conditionnelles, mais il en existe un très utile : la formule de Bayes.

2.1.1 Formule de Bayes

Proposition 2.3 (Formule de Bayes). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et E, F deux événements de probabilité strictement positive. Alors*

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(F|E)\mathbb{P}(E)}{\mathbb{P}(F)}.$$

La preuve tient en une ligne : par définition de la probabilité conditionnelle,

$$\mathbb{P}(E|F)\mathbb{P}(F) = \mathbb{P}(E \cap F) = \mathbb{P}(F|E)\mathbb{P}(E).$$

Il s'agit pourtant d'une observation très utile — elle permet non seulement de calculer, mais se trouve aussi au cœur du cadre de la statistique bayésienne / de la pensée bayésienne des probabilités.

Analysons un exemple simple.

Exemple 2.4. *Considérez la situation avec trois pièces : l'une a pile sur les deux faces, l'une a face sur les deux faces, et l'une est équilibrée. Quelqu'un choisit, selon une certaine procédure, l'un des trois types de pièces, vous dit qu'elle a lancé la pièce et que le résultat a été pile. Quelle pièce a-t-elle lancée ?*

L'espace de probabilité pertinent qui contient les trois pièces et trois lancers est le suivant. L'espace d'états est le produit $\{C_h, C_t, C_f\} \times \{H, T\}$ — la première coordonnée décrit le type de pièce, la seconde le résultat du lancer. Comme σ -algèbre, nous prenons la totalité (on peut demander quelle face est sortie et quel était le type de pièce).

Nous savons que, pour un ensemble fini avec l'ensemble des parties, il suffit de définir \mathbb{P} sur chaque singleton de l'espace d'états. D'après les hypothèses, $\mathbb{P}(\{C_h, T\}) = \mathbb{P}(\{C_t, H\}) = 0$ et $\mathbb{P}(\{C_f, T\}) = \mathbb{P}(\{C_f, H\})$. Si l'on pose $p_f = \mathbb{P}(\{\text{coin} = C_f\})$, $p_h = \mathbb{P}(\{\text{coin} = C_h\})$, $p_t = \mathbb{P}(\{\text{coin} = C_t\})$, on doit aussi avoir $p_f + p_t + p_h = 1$, ce qui laisse deux paramètres libres.

Calculons maintenant les probabilités pertinentes. Clairement,

$$\mathbb{P}(\{\text{coin} = C_t\} | \{\text{toss} = H\}) = 0,$$

car la pièce « deux faces » (deux fois T) ne peut pas produire H . Pour les autres cas, on utilise Bayes :

$$\mathbb{P}(\{\text{coin} = C_h\} | \{\text{toss} = H\}) = \frac{\mathbb{P}(\{\text{toss} = H\} | \{\text{coin} = C_h\}) \mathbb{P}(\{\text{coin} = C_h\})}{\mathbb{P}(\{\text{toss} = H\})} = \frac{\mathbb{P}(\{\text{coin} = C_h\})}{\mathbb{P}(\{\text{toss} = H\})}$$

et

$$\mathbb{P}(\{\text{coin} = C_f\} | \{\text{toss} = H\}) = \frac{\mathbb{P}(\{\text{toss} = H\} | \{\text{coin} = C_f\}) \mathbb{P}(\{\text{coin} = C_f\})}{\mathbb{P}(\{\text{toss} = H\})} = \frac{\mathbb{P}(\{\text{coin} = C_f\})}{2 \mathbb{P}(\{\text{toss} = H\})}.$$

Ainsi,

$$\frac{\mathbb{P}(\{\text{coin} = C_h\} | \{\text{toss} = H\})}{\mathbb{P}(\{\text{coin} = C_f\} | \{\text{toss} = H\})} = \frac{2 \mathbb{P}(\{\text{coin} = C_h\})}{\mathbb{P}(\{\text{coin} = C_f\})} = 2p_h/p_f,$$

et, comme

$$\mathbb{P}(\{\text{coin} = C_h\} | \{\text{toss} = H\}) + \mathbb{P}(\{\text{coin} = C_f\} | \{\text{toss} = H\}) = 1,$$

on conclut

$$\mathbb{P}(\{\text{coin} = C_f\} | \{\text{toss} = H\}) = \frac{p_f}{p_f + 2p_h} \quad \text{et} \quad \mathbb{P}(\{\text{coin} = C_h\} | \{\text{toss} = H\}) = \frac{2p_h}{p_f + 2p_h}.$$

Que peut-on conclure ? D'abord, sans connaissance des probabilités a priori de chaque pièce, on ne peut pas dire grand-chose du résultat final, puisqu'il les contient ! Ce que nous supposons sur la probabilité initiale de chaque pièce compte beaucoup : si l'on estime que la pièce « deux piles » est très peu probable par rapport à la pièce équilibrée, disons $p_h = 0,000001 p_f$, alors après avoir observé « pile », notre estimation donne $\mathbb{P}(\{\text{coin} = C_f\} | \{\text{toss} = H\}) \approx 0,999999$. En revanche, si nous n'avons aucune raison de croire qu'une pièce est plus probable qu'une autre (par exemple si la personne a choisi au hasard parmi les

trois), alors $p_f = p_h = p_t = 1/3$ et la formule donne $\mathbb{P}(\{\text{coin} = C_f\} | \{\text{toss} = H\}) = 1/3$ et $\mathbb{P}(\{\text{coin} = C_h\} | \{\text{toss} = H\}) = 2/3$.

Cependant, un point important est que, indépendamment des probabilités initiales, on peut dire comment les probabilités — ou plutôt le rapport des probabilités — a changé : notre estimation que la pièce « deux piles » a été choisie augmente d'un facteur 2 par rapport à la pièce équilibrée. Et, comme vous le verrez sur la feuille d'exercices, si nous observions davantage de lancers, nous deviendrions de plus en plus confiants quant au type de pièce, indépendamment d'une estimation initiale éventuellement mauvaise. C'est aussi l'idée derrière l'approche bayésienne : nous ne connaissons pas forcément tous les paramètres au départ, mais nous pouvons les remplir par des hypothèses puis, à mesure que nous observons le monde, améliorer a posteriori ces hypothèses et raffiner nos modèles.

2.1.2 Formule des probabilités totales

Bien que les probabilités conditionnelles soient souvent délicates, elles sont indispensables — et utiles. Par exemple, elles permettent de décomposer l'espace de probabilité. En effet, le résultat suivant généralise l'idée intuitive : si l'on sait qu'exactly un des trois événements E_1, E_2, E_3 se produit toujours, alors, pour comprendre la probabilité de n'importe quel autre événement F , il suffit de connaître les probabilités conditionnelles $\mathbb{P}(F|E_i)$.

Proposition 2.5 (Formule des probabilités totales). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soit I dénombrable et $(E_i)_{i \in I}$ une famille d'événements disjoints de probabilité strictement positive telle que $\Omega = \bigcup_{i \in I} E_i$. Alors, pour tout $F \in \mathcal{F}$,*

$$\mathbb{P}(F) = \sum_{i \in I} \mathbb{P}(F|E_i)\mathbb{P}(E_i).$$

Proof. Comme $\Omega = \bigcup_{i \in I} E_i$, on a $\mathbb{P}(F) = \mathbb{P}\left(F \cap \left(\bigcup_{i \in I} E_i\right)\right)$.

Or $F \cap \left(\bigcup_{i \in I} E_i\right) = \bigcup_{i \in I} (F \cap E_i)$. Comme les $(E_i)_{i \in I}$ sont disjoints, les $(F \cap E_i)_{i \in I}$ le sont aussi. Par additivité dénombrable,

$$\mathbb{P}(F) = \mathbb{P}\left(\bigcup_{i \in I} (F \cap E_i)\right) = \sum_{i \in I} \mathbb{P}(F \cap E_i).$$

Par définition, $\mathbb{P}(F \cap E_i) = \mathbb{P}(F|E_i)\mathbb{P}(E_i)$, d'où la formule. □

Remarque 2.6. *Presque la même preuve fonctionne si les E_i ne recouvrent pas tout l'espace, mais seulement à probabilité 1, i.e. si $\mathbb{P}(\Omega \setminus (\bigcup_i E_i)) = 0$. Cette généralisation est laissée en exercice.*

2.2 Indépendance d'événements

Les probabilités conditionnelles ne sont bien sûr pas du tout difficiles lorsque la probabilité d'un événement ne change pas sous conditionnement — i.e. lorsque $\mathbb{P}(E|F) = \mathbb{P}(E)$. De telles paires d'événements sont dites indépendantes. En fait, la définition rigoureuse est légèrement différente :

Définition 2.7 (Indépendance de deux événements). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. On dit que deux événements E, F sont indépendants si $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.*

Observez que si $\mathbb{P}(F) > 0$, alors on retrouve l'énoncé intuitif d'indépendance : $\mathbb{P}(E|F) = \mathbb{P}(E)$. En effet, si E et F sont indépendants,

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \frac{\mathbb{P}(E)\mathbb{P}(F)}{\mathbb{P}(F)} = \mathbb{P}(E).$$

Nous avons choisi cette définition pour inclure automatiquement le cas $\mathbb{P}(F) = 0$.

Exemple 2.8. *Considérons le modèle d'un nombre uniforme parmi $\{1, 2, \dots, 12\}$ et les événements $E_1 := \{\text{le nombre vaut } 1\}$, $E_2 := \{\text{le nombre est divisible par } 2\}$, $E_3 := \{\text{le nombre est divisible par } 3\}$. Lesquels sont indépendants ?*

Calcul direct : $\mathbb{P}(E_1) = 1/12$, $\mathbb{P}(E_2) = 1/2$ et $\mathbb{P}(E_3) = 1/3$. D'autre part, $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1 \cap E_3) = 0$, et $\mathbb{P}(E_2 \cap E_3) = \mathbb{P}(\{\text{divisible par } 6\}) = 1/6$. On conclut que E_2 et E_3 sont indépendants, mais ni E_1 et E_2 , ni E_1 et E_3 ne le sont.

Déjà dans cet exemple, nous avons trois événements, et l'on peut se demander s'il existe une notion d'indépendance *conjointe* se généralisant à plusieurs événements. Il y en a en fait deux :

- l'indépendance mutuelle (ou conjointe) ;
- l'indépendance par paires.

La notion la plus forte et la plus importante est l'indépendance mutuelle.

Définition 2.9 (Indépendance mutuelle). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et I un ensemble d'indices. Les événements $(E_i)_{i \in I}$ sont dits mutuellement indépendants si, pour tout sous-ensemble fini $I_1 \subseteq I$,*

$$\mathbb{P}\left(\bigcap_{i \in I_1} E_i\right) = \prod_{i \in I_1} \mathbb{P}(E_i).$$

Parfois, on ne dispose pas de l'indépendance mutuelle (ou on ne sait pas qu'elle tient), mais on peut affirmer l'indépendance *par paires*. Il existe des notions analogues de k -indépendance.

Définition 2.10 (Indépendance par paires). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et I un ensemble d'indices. Les événements $(E_i)_{i \in I}$ sont dits indépendants par paires si, pour tout $i \neq j \in I$, les événements E_i et E_j sont indépendants.*

Il est important de noter que, bien que l'indépendance mutuelle implique l'indépendance par paires, la réciproque est fautive en général :

Exercice 2.2 (Indépendants par paires mais pas mutuellement). *Considérez l'espace de probabilité de deux lancers de pièce indépendants. Soit E_1 l'événement « la première pièce est pile », E_2 l'événement « la seconde pièce est pile » et E_3 l'événement « les deux pièces montrent la même face ». Montrer que E_1, E_2, E_3 sont indépendants par paires mais pas mutuellement indépendants.*

Enfin, on peut aussi parler d'indépendance de collections d'événements. Cela sera important pour généraliser l'indépendance d'événements à celle de variables aléatoires.

Définition 2.11 (Indépendance mutuelle de collections d'événements). *Considérons deux collections d'événements $(E_i)_{i \in I}$ et $(F_j)_{j \in J}$, tous définis sur le même espace de probabilité. On dit qu'elles sont indépendantes si, pour tous $i \in I, j \in J$:*

$$\mathbb{P}(E_i \cap F_j) = \mathbb{P}(E_i)\mathbb{P}(F_j).$$

Dans le cas de plusieurs collections $(E_{j,i})_{i \in I_j}$ pour $j = 1, 2, \dots$, on dit qu'elles sont mutuellement indépendantes si, pour tout sous-ensemble fini $J_1 \subseteq J$ et tout choix d'événements E_{j,i_j} avec $j \in J_1$, on a

$$\mathbb{P}\left(\bigcap_{j \in J_1} E_{j,i_j}\right) = \prod_{j \in J_1} \mathbb{P}(E_{j,i_j}).$$

Autrement dit, on exige que toute sous-famille d'événements, en prenant au plus un événement dans chaque collection, soit mutuellement indépendante.

Avant de passer à l'indépendance des variables aléatoires, voici quelques propriétés de base de l'indépendance pour les événements :

Lemme 2.12 (Propriétés de base). *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.*

- Si E est un événement avec $\mathbb{P}(E) = 1$, alors il est indépendant de tout autre événement.
- Si E, F sont indépendants, alors E^c et F le sont aussi. En particulier, tout événement de probabilité 0 est indépendant de tous les autres.
- Enfin, si un événement est indépendant de lui-même, alors $\mathbb{P}(E) \in \{0, 1\}$.

Proof. Voir la feuille d'exercices. □

2.3 Indépendance de variables aléatoires

Nous formalisons maintenant la notion d'indépendance pour des quantités aléatoires, c'est-à-dire des variables aléatoires. Rappelez-vous que (la loi de) X est caractérisée par tous les événements $\{X \in (a, b)\}$. L'indépendance mutuelle de variables aléatoires est alors définie comme l'indépendance mutuelle de ces familles d'événements. Plus précisément,

Définition 2.13 (Variables aléatoires mutuellement indépendantes). *Soit I un ensemble d'indices et $(X_i)_{i \in I}$ une famille de variables aléatoires définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que ces variables sont mutuellement indépendantes si, pour tout $J \subseteq I$ fini et toute collection d'intervalles $((a_j, b_j))_{j \in J}$, on a*

$$\mathbb{P}\left(\bigcap_{j \in J} \{X_j \in (a_j, b_j)\}\right) = \prod_{j \in J} \mathbb{P}(X_j \in (a_j, b_j)).$$

Remarque 2.14. *La définition plus « structurelle » utiliserait plutôt tous les boréliens $E_j \in \mathcal{F}_B$. C'est peu pratique, et il se trouve (par une théorie de la mesure non triviale) que c'est équivalent à la condition ci-dessus.*

Il existe naturellement d'autres conditions équivalentes. Par exemple, une condition utile (que nous verrons plus tard) est la suivante :

Exercice 2.3. *Soient X_1, X_2, \dots des variables aléatoires définies sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$. Alors X_1, X_2, \dots sont mutuellement indépendantes si et seulement si, pour tout $m \geq 2$ et tous $a_j \in \mathbb{R}$,*

$$\mathbb{P}\left(\bigcap_{1 \leq j \leq m} \{X_j \leq a_j\}\right) = \prod_{1 \leq j \leq m} \mathbb{P}(X_j \leq a_j).$$

Il y a aussi un critère particulièrement simple dans le cas discret.

Lemme 2.15 (Indépendance sur un espace discret). *Soient X_1, \dots, X_n définies sur un espace de probabilité discret. Alors X_1, \dots, X_n sont mutuellement indépendantes si et seulement si, pour tous $s_1, \dots, s_n \in \mathbb{R}$,*

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i = s_i\}\right) = \prod_{i=1}^n \mathbb{P}(X_i = s_i).$$

*Cela reste vrai plus généralement si X_1, \dots, X_n sont définies sur n'importe quel espace de probabilité mais ne prennent qu'un nombre dénombrable de valeurs avec probabilité 1, c'est-à-dire s'il existe, pour chacune, un ensemble dénombrable S_i tel que $\mathbb{P}(X_i \in S_i) = 1$.*⁶

Proof. Exercice. □

Comme vérification élémentaire, on voit alors facilement que, pour les espaces discrets (et en fait en général !), les indicatrices $1_E, 1_F$ de deux événements sont indépendantes si et seulement si E et F sont indépendants comme événements : en effet $\mathbb{P}(\{1_E = x\} \cap \{1_F = y\})$ vaut

$$1_{x=1}1_{y=1}\mathbb{P}(E)\mathbb{P}(F) + 1_{x=1}1_{y=0}\mathbb{P}(E)\mathbb{P}(F^c) + 1_{x=0}1_{y=1}\mathbb{P}(E^c)\mathbb{P}(F) + 1_{x=0}1_{y=0}\mathbb{P}(E^c)\mathbb{P}(F^c),$$

ce qui se réécrit

$$(1_{x=1}\mathbb{P}(E) + 1_{x=0}\mathbb{P}(E^c))(1_{y=1}\mathbb{P}(F) + 1_{y=0}\mathbb{P}(F^c)) = \mathbb{P}(\{1_E = x\})\mathbb{P}(\{1_F = y\}).$$

Exercice 2.4 (Marche aléatoire simple). *Montrer que, pour une marche aléatoire simple de longueur n , tous les incréments $\Delta_i = S_i - S_{i-1}$ pour $i = 1, \dots, n$ sont mutuellement indépendants.*

La notion de variables aléatoires indépendantes est très importante et largement utilisée — souvent aussi parce que, sinon, il est très difficile de mener des calculs !

Remarque 2.16 (v.a. i.i.d.). *On parle souvent d'une famille $(X_j)_{j \in J}$ i.i.d. : cela signifie qu'elles sont mutuellement indépendantes (le premier « i ») et identiquement distribuées (le « $i.d.$ »). Intuitivement, cela correspond à répéter exactement la même expérience aléatoire, encore et encore.*

Nous avons commencé le cours en construisant des espaces de probabilité, puis en y définissant des variables aléatoires. Mais il existe des cas naturels où l'on souhaite procéder dans l'autre sens : d'après l'observation ou l'expérience, nous voulons étudier un ensemble de variables aléatoires indépendantes — comment construire un espace de probabilité sur lequel elles « vivent » ? Cela peut sembler un peu trivial, mais mathématiquement la question n'est pas si facile ! Nous l'aborderons en partie dans la sous-section suivante.

2.4 Indépendance et produits d'espaces de probabilité

Bien que l'indépendance soit une notion probabiliste, elle est liée à une structure des espaces mesurés : le *produit*.

Exemple 2.17 (L'espace de n lancers de pièce équilibrée). *Nous avons vu qu'on peut modéliser l'espace de n lancers de pièce équilibrée en prenant comme espace d'états Ω l'ensemble*

⁶De telles variables sont appelées *discrètes*, comme nous le verrons bientôt.

des n -uplets $\{x_1, \dots, x_n\}$ avec $x_i \in \{H, T\}$, puis \mathcal{F} comme l'ensemble des parties, et enfin la probabilité de chaque singleton (chaque n -uplet) égale à 2^{-n} .

Regardons cela ainsi :

- Chaque n -uplet est un élément de l'espace produit $\{H, T\} \times \dots \times \{H, T\}$, si bien qu'on peut prendre Ω comme espace produit. Notons $\Omega_0 = \{H, T\}$ l'espace d'un seul lancer.
- L'ensemble des parties de Ω est aussi la plus petite σ -algèbre contenant tous les ensembles de la forme $E_1 \times \dots \times E_n$ avec E_i partie de $\{H, T\}$.
- La probabilité uniforme sur Ω satisfait, par définition,

$$\mathbb{P}(E_1 \times \dots \times E_n) = \mathbb{P}_0(E_1) \dots \mathbb{P}_0(E_n),$$

où \mathbb{P}_0 est la probabilité uniforme sur un seul lancer.

- Enfin, le fait que les lancers soient indépendants revient à dire : pour tout i , les événements « la i -ème coordonnée est dans E_i », c.-à-d. $F_i = \Omega_0 \times \dots \times E_i \times \dots \times \Omega_0$, sont mutuellement indépendants. Le calcul est celui attendu :

$$\mathbb{P}(F_i \cap F_j) = \mathbb{P}(\Omega_0 \times \dots \times E_i \times \dots \times E_j \times \dots \times \Omega_0) = \mathbb{P}_0(E_i)\mathbb{P}_0(E_j) = \mathbb{P}(F_i)\mathbb{P}(F_j).$$

On voit donc que la structure de produit va de pair avec l'indépendance. En effet, c'est la règle générale : l'indépendance mutuelle de variables aléatoires est naturellement liée aux produits d'espaces de probabilité.

Poursuivons cela de façon mathématique, en parlant d'abord des espaces produits en général, puis de la construction d'espaces de probabilité pour des variables indépendantes.

2.4.1 Construction d'espaces produits

Considérons des espaces de probabilité $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ pour $i = 1, 2, \dots$. Pour construire l'espace produit, on a besoin d'une σ -algèbre produit et d'une mesure produit.

- (1) La σ -algèbre produit \mathcal{F}_Π : c'est la plus petite σ -algèbre contenant tous les ensembles $E_{i_1} \times \dots \times E_{i_n}$ avec $E_{i_j} \in \mathcal{F}_{i_j}$ pour $j = 1, \dots, n$, et $\{i_j\}_{j=1}^n$ un sous-ensemble fini de \mathbb{N} . On souligne qu'elle n'est pas égale à l'ensemble de *tous* les produits « rectangulaires » $E_{i_1} \times \dots \times E_{i_n}$, même sur un produit fini.⁷
- (2) La mesure produit \mathbb{P}_Π de $\mathbb{P}_1, \mathbb{P}_2, \dots$ sur $(\prod_{i \geq 1} \Omega_i, \mathcal{F}_\Pi)$: elle est (intuitivement) la seule probabilité telle que

$$\mathbb{P}(E_{i_1} \times \dots \times E_{i_n}) = \prod_{j=1}^n \mathbb{P}_{i_j}(E_{i_j})$$

pour tous les rectangles mesurables. La construction et l'unicité, même pour des produits finis, sont techniques en général et hors du cadre de ce cours.

Nous énonçons donc le théorème suivant sans preuve (voir mesure/proba avancée) :

Théorème 2.18 (Mesure produit // admis). *Pour $i \in \mathbb{N}$, soit $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ des espaces de probabilité. Il existe une unique probabilité \mathbb{P}_Π sur $(\prod_{i \in \mathbb{N}} \Omega_i, \mathcal{F}_\Pi)$ telle que, pour tout sous-ensemble fini $J \subset \mathbb{N}$ et tout événement E de la forme $E = \prod_{i \in \mathbb{N}} F_i$ avec $F_i = \Omega_i$ pour $i \notin J$*

⁷Même dans l'exemple ci-dessus avec 2 lancers, on peut vérifier que $\{(H, H), (T, T)\}$ n'est pas un « rectangle ». Un phénomène analogue se produit pour la topologie produit.

et $F_i = E_i \in \mathcal{F}_i$ pour $i \in J$, on ait

$$(2.1) \quad \mathbb{P}_\Pi(E) = \prod_{i \in J} \mathbb{P}_i(E_i).$$

On appelle une telle mesure la mesure produit de $((\Omega_i, \mathcal{F}_i, \mathbb{P}_i))_{i \geq 1}$.

Remarque 2.19. Nous l'utiliserons surtout lorsque $(\Omega_i, \mathcal{F}_i)$ sont $(\mathbb{R}, \mathcal{F}_B)$. Pour de tels produits finis, la preuve apparaît en Analyse IV et peut aussi se déduire de l'existence de la mesure de Lebesgue sur $([0, 1], \mathcal{F}_B)$ (partie non exigible, feuille d'exercices).

Dans le cas discret et fini, l'existence et l'unicité sont faciles ; formulons-le ainsi :

Lemme 2.20 (Espaces produits discrets). Soient $(\Omega_i, \mathcal{P}(\Omega_i), \mathbb{P}_i)$, $i = 1, \dots, n$, des espaces de probabilité discrets. Alors la mesure produit \mathbb{P}_Π sur $(\prod_{i=1}^n \Omega_i, \mathcal{F}_\Pi)$ existe et est unique.

Proof. Sur la feuille d'exercices. □

2.4.2 Espaces de probabilité pour variables indépendantes

Poursuivons l'idée annoncée :

- Si l'on connaît les lois de variables aléatoires et que l'on souhaite construire un espace commun sur lequel elles soient définies et *mutuellement indépendantes*, alors on utilisera des *produits*.

Nous donnons l'énoncé dans une généralité un peu plus large que la preuve.

Théorème 2.21 (Existence d'un espace portant des v.a. indépendantes // partiellement admis). Considérons des variables aléatoires $(X_i)_{i \geq 1}$. On peut trouver un espace commun $(\Omega, \mathcal{F}, \mathbb{P})$ et des variables $(\tilde{X}_i)_{i \geq 1}$ définies dessus telles que

- pour tout $i \geq 1$, \tilde{X}_i a la même loi que X_i ;
- les variables $(\tilde{X}_i)_{i \geq 1}$ sont mutuellement indépendantes.

Exemple 2.22. Supposez que vous ayez une pièce biaisée qui donne pile avec probabilité $p \in (0, 1)$. Comment modéliser une suite de n lancers indépendants ?

L'hypothèse « toutes les suites ont la même probabilité » n'a plus de sens (par exemple quand p est proche de 1, la suite « tout pile » et « tout face » ne peuvent pas avoir la même probabilité). En revanche, l'hypothèse d'indépendance mutuelle et le lien avec les mesures produit sont utiles.

On définit l'espace comme suit :

- on prend le produit de n copies de $(\{0, 1\}, \mathcal{P}(\{0, 1\}), \mathbb{P}_p)$, où $\mathbb{P}_p(\{1\}) = p$, $\mathbb{P}_p(\{0\}) = 1 - p$.

Dans cet espace, la probabilité d'une suite fixée avec m piles et $n - m$ faces est exactement $p^m(1 - p)^{n - m}$. Si l'on veut la probabilité d'avoir exactement m piles (positions quelconques), il faut sommer sur toutes les suites à m piles, ce qui donne $\binom{n}{m} p^m(1 - p)^{n - m}$. Vérifiez que $\sum_{m=0}^n \binom{n}{m} p^m(1 - p)^{n - m} = 1$!

Esquisse de preuve du Thm. 2.21 dans le cas discret et fini. Voir la feuille d'exercices pour les détails : on prend les espaces porteurs individuels et on forme le produit (Lemme 2.20), puis l'on définit \tilde{X}_i par projection : $\tilde{X}_i(\omega_1, \dots, \omega_n) = X_i(\omega_i)$. On vérifie alors l'égalité en loi et l'indépendance par la propriété caractéristique de la mesure produit (2.1). □

Terminons cette section par un exemple important.

2.4.3 Graphe aléatoire d'ErdHos–Rényi

Nous souhaitons décrire et étudier des graphes aléatoires. Les graphes sont des structures mathématiques simples qui aident à décrire des réseaux : réseaux sociaux, logistiques, ou le réseau de neurones dans le cerveau.

Définition 2.23 (Graphe simple). *Soit $n \in \mathbb{N}$. Un graphe simple est une paire $G = (V, E)$ où $V = \{v_1, \dots, v_n\}$ est l'ensemble des sommets, et E est un sous-ensemble de $\{\{v_i, v_j\} : (v_i, v_j) \in V \times V, i \neq j\}$, c'est-à-dire un ensemble de paires non ordonnées de sommets distincts, appelées arêtes.*

On peut représenter le graphe en plaçant les n sommets v_1, \dots, v_n dans le plan et en traçant un segment entre v_i et v_j si et seulement si $\{v_i, v_j\} \in E$.

Si les réseaux sont très grands (cerveau, Facebook), il est à la fois impraticable et impossible de les décrire en détail. De plus, ils ressemblent souvent à certains réseaux *aléatoires*. Pour comprendre leurs propriétés, on étudie des modèles simplifiés de réseaux aléatoires.

Le modèle le plus simple est le graphe d'ErdHos–Rényi, où l'on inclut chaque arête avec probabilité $p > 0$.

Exemple 2.24 (Graphe d'ErdHos–Rényi). *Pour $n \in \mathbb{N}$, considérez un ensemble de sommets V de taille n , et E l'ensemble de toutes les arêtes non orientées possibles entre ces sommets.*

Le graphe aléatoire $G_{n,p}$ de taille n et paramètre d'arête $p \in [0, 1]$ est défini en incluant chaque arête indépendamment avec probabilité p .

Pour définir l'espace de probabilité, on pose :

- *L'espace d'états contient tous les graphes possibles sur V . On peut l'encoder par les configurations d'arêtes : $\Omega = \{0, 1\}^E$ (on interprète 1 comme « arête présente »).*
- *On suppose qu'on peut vérifier pour chaque arête si elle est présente ou non ; on prend donc $\mathcal{F} = \mathcal{P}(\Omega)$.*
- *Enfin, on place chaque arête indépendamment avec probabilité p . Pour $\omega \in \Omega$, on pose*

$$\mathbb{P}_p(\{\omega\}) := p^{|\omega|}(1-p)^{|E|-|\omega|},$$

où $|\omega|$ est le nombre d'arêtes présentes dans la configuration ω .

On identifie ω au graphe $G_{n,p}(\omega) = (V, E(\omega))$.

Quelles questions se poser ? Grosse maille : décrire l'aspect du graphe quand n est très grand ($n \rightarrow \infty$). On peut aussi s'intéresser à n petit, mais alors on peut tout expliciter.

Pour décrire l'aspect du graphe, on peut demander :

- (1) Combien d'arêtes sont présentes ?
- (2) Le graphe est-il connexe (pour tout v, w , existe-t-il un chemin $v = e_0 - e_1 - \dots - e_k = w$) ?
- (3) Si oui, quelle est la distance maximale entre deux sommets ?
- (4) Si non, combien de composantes connexes y a-t-il ?
- (5) Quelle est la plus grande composante connexe ?
- (6) ...

Chacune de ces questions porte sur un graphe, i.e. une configuration ω . Dans le modèle, elles correspondent à un événement ou à une variable aléatoire, dont on peut étudier la probabilité ou la loi.

Par exemple, $N_E(\omega) := |\omega|$ (nombre d'arêtes) répond à (1). L'événement $F := \{\omega : \omega \text{ est connexe}\}$ répond à (2). Il y a des questions plus complexes lorsque l'on combine plusieurs questions.

On s'intéresse au comportement lorsque p est fixé et $n \rightarrow \infty$, mais aussi à la manière dont ce comportement change avec p . A priori, p peut dépendre de n : on peut considérer une suite $G_{n,p(n)}$.

L'étude de $G_{n,p}$ est un domaine très actif (des centaines / milliers d'articles). Nous n'en donnerons qu'un très bref aperçu.

Concentrons-nous sur la connexité et regardons des scénarios. Notez que si $p = 1$, le graphe est connexe p.s., et s'il $p = 0$, il est discontinu p.s. Que se passe-t-il pour $p_n \in (0, 1)$ dépendant de n ?

Assertion 2.25. *Soit $p \in (0, 1)$ fixé. Lorsque $n \rightarrow \infty$, la probabilité que le graphe soit connexe converge vers 1.*

Ce n'est pas si surprenant : avec p fixé, on aura beaucoup d'arêtes — on s'attend à environ $pn(n-1)/2$ arêtes !

Proof. On montre que $\mathbb{P}_p(\{G_{n,p} \text{ non connexe}\}) \rightarrow 0$. D'abord,

$$\{G_{n,p} \text{ non connexe}\} = \bigcup_{v \neq w \in V} \{v, w \text{ non reliés par un chemin}\}.$$

Par l'union bound,

$$\mathbb{P}_p(\{G_{n,p} \text{ non connexe}\}) \leq \frac{1}{2} \sum_{v \neq w \in V} \mathbb{P}_p(\{v, w \text{ non reliés par un chemin}\}),$$

le facteur $1/2$ évitant le double comptage. Par symétrie, chaque paire (v, w) a la même probabilité ; le membre de droite vaut donc $n(n-1)/2 \cdot \mathbb{P}_p(\{v, w \text{ non reliés}\})$.

Bornons $\mathbb{P}_p(\{v, w \text{ non reliés}\})$. Regarder seulement l'arête $\{v, w\}$ ne suffit pas (elle est absente avec probabilité $1-p$ qui ne tend pas vers 0). Mais il existe beaucoup d'autres chemins.

Considérons les chemins de longueur 2 via un sommet z : si v et w ne sont pas reliés, alors, pour tout z , on n'a pas simultanément $\{v, z\}$ et $\{z, w\}$ présentes. Donc

$$\begin{aligned} \mathbb{P}_p(\{v, w \text{ non reliés}\}) &\leq \prod_{z \in V \setminus \{v, w\}} \mathbb{P}_p(\{\{v, z\} \notin E \cup \{z, w\} \notin E\}) \\ &= \prod_z (1 - \mathbb{P}_p(\{\{v, z\} \in E, \{z, w\} \in E\})) = (1 - p^2)^{n-2}. \end{aligned}$$

Cela tend clairement vers 0 quand $n \rightarrow \infty$. Donc deux sommets fixés sont reliés avec proba $\rightarrow 1$.

Revenons à la probabilité globale :

$$\mathbb{P}_p(\{G_{n,p} \text{ non connexe}\}) \leq \frac{n(n-1)}{2} (1 - p^2)^{n-2} \xrightarrow{n \rightarrow \infty} 0.$$

□

En regardant la preuve, on voit que l'énoncé reste vrai tant que $p = p(n)$ décroît « assez lentement ». La même preuve donne :

Assertion 2.26. Soit $(p_n)_{n \geq 1}$ telle que $p_n \geq n^{-1/4}$. Alors, lorsque $n \rightarrow \infty$, la probabilité que le graphe soit connexe converge vers 1.

Proof. Comme ci-dessus, $\frac{n(n-1)}{2}(1 - p_n^2)^{n-2} \rightarrow 0$ si $p_n \geq n^{-1/4}$. □

D'un autre côté :

Assertion 2.27. Soit $(p_n)_{n \geq 1}$ telle que $p_n \leq n^{-2}$. Alors, lorsque $n \rightarrow \infty$, la probabilité que le graphe soit connexe converge vers 0.

(Ceci figure sur la feuille d'exercices.) Notez le phénomène de *seuil*. Si p_n décroît très vite, la probabilité de connexité tend vers 0 ; s'il décroît lentement, elle tend vers 1. Pourquoi ne tend-elle pas vers une valeur entre 0 et 1 ? Où est le seuil exact ? Un théorème non trivial dit que le seuil est précisément $p_n = \frac{\log n}{n}$!

SECTION 3

Variables aléatoires et vecteurs aléatoires

Dans ce chapitre, nous regardons de plus près les variables aléatoires et les n -uplets de variables aléatoires, appelés *vecteurs aléatoires*.

3.1 La fonction de répartition d'une variable aléatoire

Rappelons que nous disons que deux variables sont égales en loi lorsque les probabilités qu'elles induisent sur $(\mathbb{R}, \mathcal{F}_B)$ sont égales — ceci permet de comparer des variables définies sur des espaces différents et issues de contextes différents.

Notre premier objectif est de voir comment classifier et comparer les variables plus facilement. Jusqu'ici, la loi d'une variable est décrite par la probabilité de *tous* les événements, ce qui est peu maniable.

Il s'avère que toute l'information sur la loi d'une variable peut être encodée de façon unique par la *fonction de répartition*.

Définition 3.1 (Fonction (cumulative) de répartition). *On appelle fonction (cumulative) de répartition (f.c.r., ou c.d.f. en anglais) une fonction $F : \mathbb{R} \rightarrow [0, 1]$ satisfaisant :*

- (1) F est croissante ;
- (2) $F(x) \rightarrow 0$ lorsque $x \rightarrow -\infty$ et $F(x) \rightarrow 1$ lorsque $x \rightarrow +\infty$;
- (3) F est continue à droite : pour tout $x \in \mathbb{R}$ et toute suite $(x_n)_{n \geq 1} \subset [x, \infty)$ telle que $x_n \rightarrow x$, on a $F(x_n) \rightarrow F(x)$.

Étant donnée une variable aléatoire X , on définit sa fonction de répartition comme suit :

Proposition 3.2 (Fonction de répartition d'une v.a.). *Pour toute variable X (définie sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$), la fonction $F_X(x) := \mathbb{P}_X((-\infty, x])$ est une fonction de répartition.*

Proof. On pose $F_X(x) = \mathbb{P}(X \in (-\infty, x])$. Comme $(-\infty, x] \subseteq (-\infty, y]$ pour $x \leq y$, (1) de la Prop. 1.12 donne que F est croissante.

Vérifions la continuité à droite. Soit $(x_n)_{n \geq 1} \subset [x, \infty)$, $x_n \rightarrow x$. Posons $A_n := \cap_{1 \leq k \leq n} (-\infty, x_k)$; alors $\bigcap_{n \geq 1} A_n = (-\infty, x]$. Par continuité de \mathbb{P} (point (5) de la Prop. 1.12), $\mathbb{P}_X(A_n) \rightarrow \mathbb{P}_X((-\infty, x])$. Comme $x_n \rightarrow x$, pour tout n assez grand, on a $\{-\infty, x_n\} \subseteq A_{m_n}$ pour un certain $m_n \rightarrow \infty$. Il s'ensuit $F_X(x) \leq F_X(x_n) \leq \mathbb{P}_X(A_{m_n})$, d'où $F_X(x_n) \rightarrow F_X(x)$.

Les deux dernières propriétés (bornes à $\pm\infty$) figurent sur la feuille d'exercices. \square

Réciproquement, toute fonction de répartition donne naissance à une unique loi de variable aléatoire.

Théorème 3.3 (Les lois sont déterminées par la f.c.r. // admis). *Toute fonction de répartition F correspond à une unique loi d'une variable X telle que $F_X(x) = \mathbb{P}_X((-\infty, x])$. Autrement dit, les f.c.r. sont en bijection avec les probabilités sur $(\mathbb{R}, \mathcal{F}_B)$.*

Admettons le théorème au niveau général, mais esquissons la construction à partir de l'uniforme sur $([0, 1], \mathcal{F}_B)$ — c'est aussi la base des simulateurs :

Esquisse : construire X à partir de l'uniforme. Soit F une f.c.r. L'idée est d'utiliser l'espace $((0, 1], \mathcal{F}_B, \mathbb{P}_U)$ (mesure uniforme). On cherche une application $(0, 1] \rightarrow \mathbb{R}$ envoyant la v.a. uniforme U vers la loi F .

Définissez $X_F : (0, 1] \rightarrow \mathbb{R}$ par

$$X_F(x) := \inf\{y \in \mathbb{R} : F(y) \geq x\}.$$

Alors X_F est croissante et (exercice) mesurable de $((0, 1], \mathcal{F}_B)$ vers $(\mathbb{R}, \mathcal{F}_B)$; c'est une v.a.⁸

On calcule

$$\mathbb{P}_U(X_F \in (-\infty, x]) = \mathbb{P}_U((0, \sup\{z \in (0, 1] : z \leq F(x)\}]) = \mathbb{P}_U((0, F(x)]) = F(x),$$

d'où F est bien la f.c.r. de X_F . □

Exemple 3.4. Calculons la f.c.r. de la variable de Bernoulli X qui vaut 1 avec probabilité p et 0 avec probabilité $1 - p$. Notez que les indicatrices d'événements de probabilité p correspondent à ce cas.

On a $F_X(x) = (1 - p)\mathbf{1}_{\{x \geq 0\}} + p\mathbf{1}_{\{x \geq 1\}}$. Plus généralement, si X ne prend qu'un nombre fini de valeurs x_1, \dots, x_n avec probabilités p_1, \dots, p_n , alors $F_X(x) = \sum_{i=1}^n p_i \mathbf{1}_{\{x \geq x_i\}}$ (pourquoi ?).

On voit que F_X encode naturellement le comportement de X . Regardons de plus près le lien entre F_X et X . On note $F(x^-)$ la limite de $F(x_n)$ pour des suites $(x_n) \rightarrow x$ avec $x_n < x$.

Lemme 3.5 (F.c.r. vs v.a.). Soit X une v.a. sur $(\Omega, \mathcal{F}, \mathbb{P})$ et F_X sa f.c.r. Alors, pour tous $x < y \in \mathbb{R}$:

- (1) $\mathbb{P}(X < x) = F(x^-)$;
- (2) $\mathbb{P}(X > x) = 1 - F(x)$;
- (3) $\mathbb{P}(X \in (x, y)) = F(y^-) - F(x)$;
- (4) $\mathbb{P}(X = x) = F(x) - F(x^-)$.

Proof. Voir la feuille d'exercices. □

Exemple 3.6. Voici la f.c.r. de l'uniforme U sur $[0, 1]$: $F_U(x) = x\mathbf{1}_{\{0 \leq x \leq 1\}} + \mathbf{1}_{\{x > 1\}}$. Par le lemme ci-dessus, pour tout intervalle $(a, b) \subset [0, 1]$, $\mathbb{P}(U \in (a, b)) = b - a$.

⁸On l'appelle parfois *fonction quantile généralisée* ; c'est « l'inverse à gauche » de F .