Probabilités discrètes

Probabilités discrètes

Sadik Boujaida
CPGE Moulay Youssef

Last Updated: February 3, 2025

Table des matières

Chapitre 1

Espaces de probabilité

1.1 Objectifs d'apprentissage

À la fin de ce chapitre, vous devriez être capable de :

- de construire un modèle qui correspond à une expérience aléatoire ;
- calculer la probabilité d'un événement dans l'espace construit ;
- utiliser un modèle pour des calculs avancés

1.2 Tribus

Dans tout le chapitre Ω désignera un ensemble non vide. I est une ensemble non vide qui sera souvent utilisé comme ensemble d'indices.

On notera $\mathscr{P}(\Omega)$, resp. $\mathscr{F}(I)$, l'ensemble de toutes les parties de Ω, resp. l'ensemble de toutes les parties finies de I.

Si A est une partie de Ω on notera A^c son complémentaire dans Ω.

Définition 1.2.1 On appelle tribu de Ω tout ensemble $\mathscr{T}$ de parties de Ω tel que :

- $\Omega \in \mathscr{T}$;
- si $A \in \mathscr{T}$ alors $A^c \in \mathscr{T}$;
- si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathscr{T}$ alors $\bigcup\limits_{n=1}^{\infty} A_n \in \mathscr{T}$;

Si $\mathscr{T}$ est une tribu de Ω alors le couple $(\Omega, \mathscr{T})$ est dit un espace probabilisable. Tout élément de $\mathscr{T}$ est dit un événement de l'espace $(\Omega, \mathscr{T})$. $\Diamond$

Remarque 1.2.2 Si $\mathscr{T}$ est une tribu de Ω alors

- $\emptyset \in \mathscr{T}$;
- si $(A_n)_{\in\mathbb{N}}$ est une suite d'éléments de $\mathscr{T}$ alors $\bigcap\limits_{n=1}^{\infty} A_n \in \mathscr{T}$.

Exemple 1.2.3

1. $\mathscr{T} = \{\emptyset, \Omega\}$ est une tribu de Ω. C'est la plus petite tribu de Ω.
2. $\mathscr{P}(\Omega)$ est une tribu de Ω. C'est la plus grande tribu de Ω.
3. Si A est une partie de Ω alors $\{\emptyset, A, A^c, \Omega\}$ est une tribu de Ω. C'est la

plus petite tribu de Ω contenant A.

□

Proposition 1.2.4

1. *Si $(\mathscr{T}_i)_{i\in I}$ est une famille de tribus de Ω alors $\bigcap_{i\in I} \mathscr{T}_i$ est une tribu de Ω.*

2. *Soit $\mathcal{A}$ un ensemble de parties de Ω. L'intersection de toutes les tribus de Ω contenant $\mathcal{A}$ est une tribu de Ω. Elle est appelée la tribu engendrée par $\mathcal{A}$. On la notera $\sigma(\mathcal{A})$.*

Définition 1.2.5 La tribu engendrée par l'ensemble des segments de $\mathbb{R}$ est appelée la tribu de Borel de $\mathbb{R}$. On la note $\mathscr{B}(\mathbb{R})$. ◊

Remarque 1.2.6

1. Soient $\mathcal{A}$ et $\mathcal{B}$ deux ensembles de parties de Ω. Si $\mathcal{A} \subset \mathcal{B} \subset \sigma(A)$ alors $\sigma(\mathcal{A}) = \sigma(\mathcal{B})$.

2. Tout intervalle de $\mathbb{R}$ peut être écrit comme la réunion d'une suite de segments. La tribu de Borel de $\mathbb{R}$ contient donc tous les intervalles de $\mathbb{R}$. Comme tout ouvert de $\mathbb{R}$ est une union au plus dénombrable d'intervalles ouverts, la tribu de Borel de $\mathbb{R}$ contient tous les ouverts de $\mathbb{R}$. Elle contient donc aussi tous les fermés de $\mathbb{R}$.

3. En général si E est un espace vectoriel normé de dimension finie alors la tribu de Borel de E, notée $\mathscr{B}(E)$, est par définition la tribu engendrée par les boules fermées de E. Elle contient alors tous les ouverts et tous les fermés de E et est de ce fait indépendante de la norme choisie sur E.

Proposition 1.2.7 *Soit $\mathcal{A} = \{A_i \mid i \in I\}$ un ensemble au plus dénombrable de parties de Ω qui* forment une partition *de Ω. Alors*

$$\sigma(\mathcal{A}) = \left\{ \bigcup_{i\in I'} A_i \mid I' \subset I \right\}$$

Démonstration. Posons $\mathscr{T} = \left\{ \bigcup_{i\in I'} A_i \mid I' \subset I \right\}$ et montrons que $\mathscr{T}$ est une tribu de Ω.

- Comme $\Omega = \bigcup_{i\in I} A_i$ alors $\emptyset, \Omega \in \mathscr{T}$.

- Si $A = \bigcup_{i\in I'} A_i \in \mathscr{T}$ alors $A^c = \bigcup_{i\in I\setminus I'} A_i \in \mathscr{T}$.

- Soit $(B_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathscr{T}$ et posons pour tout $n \in \mathbb{N}$, $B_n = \bigcup_{i\in I_n} A_i$ alors

$$\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{i\in\bigcup_{n\in\mathbb{N}} I_n} A_i \in \mathscr{T}.$$

$\mathscr{T}$ est donc bien une tribu de Ω. Elle contient $\mathcal{A}$ et toute tribu qui contient tous les ensembles A_i contient tous les éléments de $\mathscr{T}$. C'est donc la plus petite tribu de Ω contenant $\mathcal{A}$. Soit $\sigma(\mathcal{A}) = \mathscr{T}$. ■

Corollaire 1.2.8 *Si l'ensemble Ω est au plus dénombrable alors*

$$\sigma\Big(\big\{\{\omega\} \mid \omega \in \Omega\big\}\Big) = \mathscr{P}(\Omega)$$

Autrement dit la seule tribu de Ω qui contient tous les singletons de Ω est $\mathscr{P}(\Omega)$.

Remarque 1.2.9 Si Ω est infini non dénombrable alors la tribu $\mathscr{D}(\Omega)$ engendrée par les singletons de Ω est définie par la condition

$$\forall A \in \mathscr{P}(\Omega),\ A \in \mathscr{D}(\Omega) \Longleftrightarrow \begin{array}{r|l} & A \text{ est au plus dénombrable} \\ \text{ou} & A^c \text{ est au plus dénombrable} \end{array}$$

En particulier toute tribu $\mathscr{T}$ de Ω qui contient tous les singletons de Ω contient $\mathscr{D}(\Omega)$ et donc contient toutes la parties dénombrables de Ω.

C'est le cas par exemple de la tribu de Borel de tout espace vectoriel normé de dimension finie (les singletons sont des boules fermées).

Proposition 1.2.10 (tribu induite). *Soit $\mathscr{T}$ une tribu de Ω. Pour toute élément B de $\mathscr{T}$, l'ensemble*

$$\mathscr{T}_B = \{B \cap A \mid A \in \mathscr{T}\}$$

est une tribu de B. On l'appelle la tribu induite par $\mathscr{T}$ sur B.

1.3 Probabilités

Note 1.3.1 Dans toute cette section, $(\Omega, \mathscr{T})$ désignera un espace probabilisable.

Définition 1.3.2 On appelle probabilité de l'espace $(\Omega, \mathscr{T})$ toute application $\mathbb{P} : \mathscr{T} \to [0,1]$ telle que :

- $\mathbb{P}(\Omega) = 1$;
- si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathscr{T}$ deux à deux disjoints alors

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

Si $\mathbb{P}$ est un probabilité de $(\Omega, \mathscr{T})$ alors le triplet $(\Omega, \mathscr{T}, \mathbb{P})$ est dit un espace probabilisé. $\diamondsuit$

Explication. Expérimentalement, $\mathbb{P}(A)$ est une approximation de la fréquence de réalisation de l'événement A quand on répète l'expérience un grand nombre de fois :

$$\mathbb{P}(A) \approx \frac{k_N(A)}{N}$$

$k_N(A)$ étant le nombre de fois où l'événement A se réalise quand on répète l'expérience aléatoire N fois, pour N très grand. $\blacksquare$

Remarque 1.3.3 Dans la définition précédente, on a utilisé la notation $\sum\limits_{n=0}^{\infty} \mathbb{P}(A_n)$ pour désigner la somme de la série $\sum \mathbb{P}(A_n)$ que celle-ci soit convergente ou non avec la convention que la somme d'une série divergente à termes positifs est $+\infty$. Ce qui pose évidemment un problème de consistence puisque $\mathbb{P}(\bigcup\limits_{n\in\mathbb{N}} A_n)$ est sensé être dans l'intervalle $[0,1]$.

En fait avec les deux axiomes de la définition, la famille $(\mathbb{P}(A_n))_{n\in\mathbb{N}}$ est toujours sommable et sa somme est dans $[0,1]$. Voilà comment le justifier :

- En posant $A_n = \emptyset$ pour tout $n \in \mathbb{N}$, le deuxième axiome aboutit à $\mathbb{P}(\emptyset) = 0$.

- Si $A_0, A_1, \ldots, A_m$ sont des événements deux à deux disjoints en posant $A_n = \emptyset$ pour tout $n > m$, le deuxième axiome fournit

$$\mathbb{P}\left(\bigcup_{n=0}^{m} A_n\right) = \sum_{n=0}^{m} \mathbb{P}(A_n).$$

- Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathscr{T}$ deux à deux disjoints. On a alors pour tout $n \in \mathbb{N}$

$$\sum_{k=0}^{n} \mathbb{P}(A_k) = \mathbb{P}\left(\bigcup_{k=0}^{n} A_n\right) \leqslant 1$$

  La série à termes réels positifs $\sum \mathbb{P}(A_n)$ est donc convergente et sa somme est dans $[0, 1]$.

Proposition 1.3.4 *Soit $\mathbb{P}$ une probabilité sur $(\Omega, \mathscr{T})$. Alors :*

1. $\mathbb{P}(\emptyset) = 0$ *;*
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ *pour tout* $A \in \mathscr{T}$ *;*
3. *Si* $A \subset B$ *alors* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$ *et* $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ *pour tout* $A, B \in \mathscr{T}$ *;*
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *pour tout* $A, B \in \mathscr{T}$ *;*
5. $\mathbb{P}\left(\bigcup\limits_{n=0}^{\infty} A_n\right) \leqslant \sum\limits_{n=0}^{\infty} \mathbb{P}(A_n)$ *pour toute suite* $(A_n)_{n\in\mathbb{N}}$ *d'éléments de* $\mathscr{T}$*.*
6. $\mathbb{P}\left(\bigcup\limits_{i\in I}^{\infty} A_i\right) = \sum\limits_{i\in I}^{\infty} \mathbb{P}(A_i)$ *si* $(A_i)_{i\in I}$ *est une famille au plus dénombrable d'éléments deux à deux disjoints de* $\mathscr{T}$*.*

Démonstration. Les propriétés 1, 2 et 3 sont des conséquences immédiates des axiomes de la définition de probabilité et de la remarque précédente.

1. Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathscr{T}$ et posons $B_0 = A_0$ et $B_n = A_n \setminus \left(\bigcup\limits_{k=0}^{n-1} A_k\right)$ pour tout $n \in \mathbb{N}$. Les $(B_n)_{n\in\mathbb{N}}$ sont deux à deux disjoints et $\bigcup\limits_{n=0}^{\infty} A_n = \bigcup\limits_{n=0}^{\infty} B_n$ avec $B_n \subset A_n$ pour tout $n \in \mathbb{N}$. On a donc

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mathbb{P}(B_n) \leqslant \sum_{n=0}^{\infty} \mathbb{P}(A_n).$$

2. Soit $(A_i)_{i\in I}$ une famille au plus dénombrable d'éléments de $\mathscr{T}$ deux à deux disjoints. Si I est fini alors c'est une conséquance de la remarque précédente. Si I est infini alors il existe une bijection σ entre $\mathbb{N}$ et I et il suffit de poser $B_n = A_{\sigma(n)}$ pour tout $n \in \mathbb{N}$ pour conclure grâce au théorème de permutation des termes pour les familles sommables.

■

Théorème 1.3.5 (de continuité monotone).

1. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite croissante d'éléments de* $\mathscr{T}$. *Alors*

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

2. *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite décroissante d'éléments de* $\mathscr{T}$. *Alors*

$$\mathbb{P}\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Démonstration.

1. On pose $B_0 = A_0$ et $B_n = A_n \setminus A_{n-1}$ pour tout $n \in \mathbb{N}^*$. Les événements B_n sont deux à deux disjoints et $\bigcup\limits_{n=0}^{\infty} A_n = \bigcup\limits_{n=0}^{\infty} B_n$. On a donc

$$\begin{aligned}\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n\right) &= \sum_{n=0}^{\infty} \mathbb{P}(B_n)\\ &= \mathbb{P}(A_0) + \sum_{n=1}^{\infty}\big(\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\big)\\ &= \mathbb{P}(A_0) + \lim \mathbb{P}(A_n) - \mathbb{P}(A_0)\\ &= \lim_{n\to\infty} \mathbb{P}(A_n)\end{aligned}$$

2. On suppose que $(A_n)_n$ est décroissante. Alors $(A_n^c)_n$ est croissante et on peut ainsi écrire

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{n\in\mathbb{N}} A_n\Big) &= 1 - \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n^c\Big)\\ &= 1 - \lim \mathbb{P}(A_n^c)\\ &= \lim \mathbb{P}(A_n)\end{aligned}$$

■

Corollaire 1.3.6 *Soit* $(A_n)_{n\in\mathbb{N}}$ *une suite quelconque d'éléments de* $\mathscr{T}$. *Alors*

1. $\mathbb{P}\left(\bigcup\limits_{n=0}^{\infty} A_n\right) = \lim\limits_{n\to\infty} \mathbb{P}(\bigcup\limits_{k=0}^{n} A_k)$.
2. $\mathbb{P}\left(\bigcap\limits_{n=0}^{\infty} A_n\right) = \lim\limits_{n\to\infty} \mathbb{P}(\bigcap\limits_{k=0}^{n} A_k)$.

Démonstration. Pour le premier point on pose $B_n = \bigcup_{k=0}^{n} A_k$ pour tout $n \in \mathbb{N}$.La suite $(B_n)_n$ et croissante et $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$. On applique alors la formule des probabilites continues.
Pour le second on pose $C_n = \bigcap_{k=0}^{n} A_k$. La suite $(C_n)_n$ est décroissante et on peut appliquer le théorème de continuité monotone sachant que $\bigcap_{n\in\mathbb{N}} A_n = \bigcap_{n\in\mathbb{N}} C_n$. ■

Définition 1.3.7 On appelle système complet d'événements (SCE) de $(\Omega, \mathscr{T})$ toute famille $(B_i)_{i\in I}$ au plus dénombrables d'éléments $\mathscr{T}$ qui forme une partition de Ω. ♢

Exemple 1.3.8 (exemples courants de SCE).

1. Pour tout événement A, (A, A^c) est un SCE de $(\Omega, \mathscr{T})$.
2. Si Ω est au plus dénombrable, alors $(\{\omega\})_{\omega\in\Omega}$ est un SCE de $(\Omega, \mathscr{P}(\Omega))$.
3. *SCE généré par une variable aléatoire discrète (important).*

   si X est une application définie sur Ω telle que $X(\Omega)$ soit au plus dénombrable et $X^{-1}\big(\{x\}\big) \in \mathscr{T}$ pour tout $x \in X(\Omega)$ alors $\big(X^{-1}\big(\{x\}\big)\big)_{x\in X(\Omega)}$ est un SCE de $(\Omega, \mathscr{T})$.

□

Théorème 1.3.9 (Formule des probabilités totales). *Soit $(B_i)_{i\in I}$ un SCE de $(\Omega, \mathscr{T})$. Alors pour tout $A \in \mathscr{T}$, on a*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap B_i)$$

Remarque 1.3.10

1. Ce résultat reste valable si les événements B_i sont deux à deux disjoints, sans former une partition de Ω, à condition que

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

2. Cette formule est extrêmement utile. Elle stipule que si on sait calculer les probabilités des événements d'un SCE alors on peut calculer la probabilité de n'importe quel événement.

Exemple 1.3.11 (exemples génériques d'utilisation de la formule des probabilités totales).

1. si on fixe un événement B alors

$$\forall A \in \mathscr{T},\ \mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$$

2. Si $B_1, B_2, \ldots, B_{n+1}$ sont des événements quelconques alors

$$\mathbb{P}\bigg(\Big(\bigcap_{k=1}^{n} B_k\Big) \cap B_{n+1}^c\bigg) = \mathbb{P}\Big(\bigcap_{k=1}^{n} B_k\Big) - \mathbb{P}\Big(\bigcap_{k=1}^{n+1} B_k\Big)$$

3. Le théorème suivant donne une application importante de la formule des probabilités totales.

□

Théorème 1.3.12 *On suppose que Ω est au plus dénombrable et on le munit de la tribu $\mathscr{P}(\Omega)$.*

1. *Soit $\mathbb{P}$ une probabilité de $(\Omega, \mathscr{P}(\Omega))$. Pour tout $A \in \mathscr{T}$, on a*

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

   $\mathbb{P}$ est entièrement détérminée par les probabilités des singletons $\{\omega\}$.

2. *Soit $(p_\omega)_{\omega\in\Omega}$ une famille de nombres réels positifs de somme 1. Alors il*

existe une unique probabilité $\mathbb{P}$ sur $\mathscr{T}$ telle que

$$\forall \omega \in \Omega,\ \mathbb{P}(\{\omega\}) = p_\omega$$

Remarque 1.3.13 Si Ω est au plus dénombrable alors pour toute famille sommable $x = (x_\omega)_{\omega\in\Omega}$ de nombre réels positifs de somme $S > 0$, on peut définir une probabilté sur $(\Omega, \mathscr{P}(\Omega)$ en posant

$$\forall A \in \mathscr{P}(\Omega),\ \mathbb{P}_x(A) = \frac{1}{S}\sum_{\omega\in A} x_\omega$$

Cette remarque peut avoir des applications analytiques intéressantes même dans des cas où la probabilité $\mathbb{P}_x$ ne correspond pas réellement à une expérience aléatoire.

Proposition 1.3.14 *On considère un événemnt B de probabilité non nulle. L'application $\mathbb{P}_B$ définie sur $\mathscr{T}$ par*

$$\forall A \in \mathscr{T},\ \mathbb{P}_B(A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$$

est une probabilité de $(\Omega, \mathscr{T})$. La probabilité $\mathbb{P}_B(A)$ est aussi notée $\mathbb{P}(A \mid B)$ et on l'appelle probailité de A sachant B.

Démonstration. On a bien $\mathbb{P}_B(\Omega) = 1$ et pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathscr{T}$ deux à deux disjoints, on a

$$\mathbb{P}_B\left(\bigcup_{n=0}^{\infty} A_n\right) = \frac{\mathbb{P}\left(\bigcup_{n=0}^{\infty} A_n \cap B\right)}{\mathbb{P}(B)} = \sum_{n=0}^{\infty}\frac{\mathbb{P}(A_n\cap B)}{\mathbb{P}(B)} = \sum_{n=0}^{\infty}\mathbb{P}_B(A_n)$$

■

Justification de la définition. la probabilité de «A sachant B» est la probabilité de réalisation d'un résultat favorable à A quand on ne retient que les résultats favorables à B. C'est une approximation du rapport des fréquences de réalisation des événements $A\cap B$ et B dans une série d'expériences.

$$\mathbb{P}(A \mid B) \approx \frac{k_N(A\cap B)}{k_N(B)} = \frac{k_N(A\cap B)/N}{k_N(B)/N}.$$

D'où la définition

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

■

Remarque 1.3.15

1. Si $\mathbb{P}(B) = 0$ alors par convention, pour tout $A \in \mathscr{T}$, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Ce choix s'explique par le fait que tout se passe comme si l'événement B n'avait pas eu lieu.

   Avec cette convention on peut toujours écrire

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

2. Sachant que $\mathbb{P}_B$ est une probabilité on peut lui appliquer toutes propriétés vues précédemment. Par exemple

   $$\mathbb{P}(A^c \mid B) = 1 - \mathbb{P}(A \mid B)$$

Mise en garde 1.3.16 On parle de «probabilité de A sachant B», jamais de «l'événement» $(A \mid B)$ qui n'a pas de sens en théorie des probabilités.

Théorème 1.3.17 *Soit $(B_i)_{i\in I}$ un SCE de $(\Omega, \mathscr{T})$. Alors*

$$\forall A \in \mathscr{T},\ \mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

Proposition 1.3.18 (Formules de Bayes).

1. *Soient A et B deux événements de probabilités non nulles. Alors*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

2. *Soit $(B_i)_{i\in I}$ un SCE de $(\Omega, \mathscr{T})$ et $A \in \mathscr{T}$ un événement de probabilité non nulle. Alors pour tout $i \in I$, on a*

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}{\sum\limits_{j\in I} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}$$

Exemple 1.3.19 (générique d'utilisation de la formule de Bayes). Une chaine de production industrielle contient N machines identiques mais avec des taux de production de pièces défectueuses différents. Le taux pour la machine numéro k est p_k.

On prélève une pièce au hasard parmi un lot de pièces produite par la chaîne et on constate qu'elle est défectueuse. Quelle est la probabilité que cette pièce provienne de la machine numéro k ?Si on note A l'événement «la pièce est défectueuse» et B_k l'événement «la pièce provient de la machine numéro k» alors on cherche $\mathbb{P}(B_k \mid A)$. Celle-ci est donnée par la formule de Bayes :

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum\limits_{j=1}^{N} \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)} = \frac{p_k}{\sum\limits_{j=1}^{N} p_j}$$

car $\mathbb{P}(B_k) = \frac{1}{N}$ et $\mathbb{P}(A \mid B_k) = p_k$ pour tout $k \in \{1, \ldots, N\}$. □

1.4 Indépendance des événements

Note 1.4.1 Dans tout la suite, $(\Omega, \mathscr{T}, \mathbb{P})$ désignera un espace probabilisé.

Définition 1.4.2

1. Deux événements A et B de $\mathscr{T}$ sont dits indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

2. Une famille $(A_i)_{i\in I}$ d'événements de $\mathscr{T}$ est dite mutuellement indépendante (MI) si pour toute partie finie J de I,

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

◊

Explication.

- *Indépendance de deux événements.*

  A est indépendant de B si la probabilité de réalisation de A sachant B est la même que celle de réalisation de A : $\mathbb{P}(A \mid B) = \mathbb{P}(A)$. Ce qui revient à
  $$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

- *Indépendance mutuelle.*

  Des évéments A_i, $i \in I$ sont mutuelement indépendants si et seulement si pour tout $i \in I$ et pour toute partie finie $J \subset I$ ne contenant pas i on a
  $$\mathbb{P}(A_i \cap \bigcap_{j \in J} A_j) = \mathbb{P}(A_i)\mathbb{P}\left(\bigcap_{j \in J} A_j\right)$$
  c'est à dire si et seulement si la réalisation de chaque événement A_i est indépendante de la réalisation simultanée d'un ou plusieurs évenements A_j lorsque $j \neq i$.

■

Remarque 1.4.3

1. Si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$ alors tout événement B est indépendant de A.
2. deux événements incompatibles A et B ne peuvent être indépendants que si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(B) = 0$.
3. Si $(A_i)_{i \in I}$ est une famille MI alors $(A_i)_{i \in I'}$ est MI pour toute partie I' de I. En particulier les événements A_i sont deux à deux indépendants.
4. Soit $(A_i)_{i \in I}$ est une famille d'évéments MI. Si on lui ajoute des événements presques sûrs ou négligeable alors les événements de la nouvelle famille sont MI.
5. Dans un SCE $(B_i)_{i \in I}$ les événements B_i ne peuvent être deux à deux indépendants, et a fortiori MI, que s'il existe $i_0 \in I$ tel que
   $$\mathbb{P}(B_{i_0}) = 1 \text{ et } \forall i \neq i_0,\ \mathbb{P}(B_i) = 0.$$

Proposition 1.4.4 *Si A et B sont des événements indépendants alors A^c et B sont indépendants, A et B^c sont indépendants et A^c et B^c sont indépendants.*

Théorème 1.4.5 *Soit $(A_i)_{i \in I}$ une famille MI d'événements de $\mathscr{T}$.*

1. *Soit I' une partie de I. On pose $B_i = A_i$ si $i \in I'$ et $B_i = A_i^c$ si $i \notin I'$. Alors la famille $(B_i)_{i \in I}$ est MI (en particulier $(A_i^c)_{i \in I}$ est MI).*
2. *Lemme des coalitions.*

   Soit $(I_k)_{k \in K}$ une famille de parties deux à deux disjointes de I. On pose pour tout $k \in K$, $C_k = \bigcap_{i \in I_k} A_i$. Alors la famille $(C_k)_{k \in K}$ est MI.

Proposition 1.4.6 (Formule des compléments). *Soient $A_1, A_2, \ldots, A_n$ des événements de $(\Omega, \mathscr{T})$. Alors*
$$\mathbb{P}\left(A_1 \cap A_2 \cap \cdots \cap A_n\right) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Démonstration. On écrit

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \cap \cdots \cap A_n \mid A_1) = \mathbb{P}(A_1)\mathbb{P}_{A_1}(A_2 \cap \cdots \cap A_n)$$

et on raisonne par récurrnece en appliquant l'hypothèse de récurrence avec la probailité $\mathbb{P}_{A_1}$ ■

1.5 Modélisation de phénomènes aléatoires

Pour modéliser une expérience aléatoire on se place dans un ensemble Ω qui contient tous les résultats possibles de l'expérience et on choisi une tribu de Ω dans laquelle on peut exprimer tous les «événements» qui nous intéressent. Il reste ensuite à définir une probabilité sur cette tribu qui rend compte de la fréquence de réalisation de ces événements.

On adopte alors le vocabulaire suivant :

- Ω est dit l'univers de l'expérience ;
- pour une instance de l'expérience, un évenement A se réalise si le résultat obtenu est dans A ;
- un événement A est dit presque sûr si $\mathbb{P}(A) = 1$ et négligeable si $\mathbb{P}(A) = 0$;
- deux événements A et B sont dits incompatibles si $\mathbb{P}(A \cap B) = 0$;

Exemple 1.5.1 (signification de certains événements courants). Soit $(A_n)_n$ une suite quelconque d'événements.

- $\bigcup_{n\in\mathbb{N}} A_n$ est l'événement «au moins un des événements A_n se réalise» ;
- $\bigcap_{n\in\mathbb{N}} A_n$ est l'événement «tous les événements A_n se réalisent» ;
- $U = \bigcup_{n\in\mathbb{N}} \left(\bigcap_{k=n}^{\infty} A_k \right)$ est l'événement «les événements A_k se réalisent tous à partir d'un certain rang n pour pour au moins un indice n». La probabilité de cet événement est

$$\mathbb{P}(U) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k \right) = \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{n+m} A_k \right)$$

- $V = \bigcap_{n\in\mathbb{N}} \left(\bigcup_{k=n}^{\infty} A_k \right)$ est l'événement «l'événement A_n se réalise pour une infinité d'indices n». La probabilité de cet événement est

$$\mathbb{P}(V) = 1 - \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \left(\bigcap_{k=n}^{\infty} A_k^c \right) \right) = 1 - \lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=n}^{n+m} A_k^c \right)$$

□

Deux approches sont en fait possibles pour modéliser une expérience aléatoire :

1. Définir un espace $(\Omega, \mathscr{T}, \mathbb{P})$ où Ω représente exactement l'ensemble des résultats possibles de l'expérience.

2. La précédente méthode risque d'être inadéquate si certains événements qui nous intéresse sont inexprimables dans la tribu considérée. C'est pour cela qu'en général on préfère se placer dans un espace probabilisable $(\Omega, \mathscr{T})$ où Ω est beaucoup plus large et on modélise l'expérience non pas avec $(\Omega, \mathscr{T})$ mais avec une application X définie sur Ω qu'on appellera «variable aléatoire». L'avantage est qu'on peut combiner entre les résultats de différentes expériences aléatoires.

Remarque 1.5.2 Cas où l'univers est au plus dénombrable. Dans la pratique si l'ensemble Ω qui contient les résultats de l'expérience est au plus plus dénombrable alors on le munit de la tribu $\mathscr{T} = \mathscr{P}(\Omega)$ de telle sorte que tout ensemble formé de résultats possibles de l'expérience soit un événement. Ce choix de tribu n'est pas adéquat lorsque l'ensemble des résultats est non dénombrable car cela pose des difficultés insurmontables pour définir une probabilité sur $\mathscr{P}(\Omega)$.

Exemple 1.5.3 Lancer un dé. On lance un dé à 6 faces. On peut modéliser cette expérience par $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathscr{T} = \mathscr{P}(\Omega)$. On peut définir une probabilité sur $\mathscr{T}$ en posant $\mathbb{P}(\{i\}) = \frac{1}{6}$ pour tout $i \in \Omega$. □

Exemple 1.5.4 Lancer deux dés. On lance deux dés à 6 faces. On peut modéliser cette expérience par $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ et $\mathscr{T} = \mathscr{P}(\Omega)$. On peut définir une probabilité sur $\mathscr{T}$ en posant $\mathbb{P}(\{(i, j)\}) = \frac{1}{36}$ pour tout $(i, j) \in \Omega$.

Si on s'intéresse à l'événement A : «la somme des résultats est paire», on peut créer un espace spécifique sous la forme $(\Omega = \{0, 1\}, \mathscr{P}(\Omega))$ et y adjoindre la probabilité adéquate ou bien on peut rester dans l'univers $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ et définir la variable aléatoire X sur Ω en posant $X(i, j) = i+j$ pour tout $(i, j) \in \Omega$. On peut alors exprimer l'événement A par

$$A = \{(i, j) \in \Omega \mid X(i, j) \text{ est paire}\}$$

□

Exemple 1.5.5 Suite de lancers d'une pièce de monnaie. On lance indéfiniment une piece de monnaie. Si on ne s'intéresse qu'au numéro du premier lancer qui donne «face» alors peut modéliser cette expérience par $\Omega = \mathbb{N}^*$ et $\mathscr{T} = \mathscr{P}(\Omega)$. On peut définir une probabilité sur $\mathscr{T}$ en posant $\mathbb{P}(\{n\}) = \frac{1}{2^n}$ pour tout $n \in \mathbb{N}^*$.

Mais dans ce cas les événements «obtenir trois face successivement au moins une fois» ou «obtenir face une infinité de fois» ne peuvent être exprimés dans $\mathscr{T}$.

On peut alors se placer dans l'espace beaucoup plus vaste $\Omega = \{0, 1\}^{\mathbb{N}^*}$. Le résultat de chaque séquence infinie de lancers est modélisé par une suite de zéros et de uns. C'est un ensemble non dénombrable et y définir une tribu sur laquelle on doit ensuite définir une probabilité est non aisé. L'approche peut être la suivante : on identifie une famille d'événements élémentaires avec lesquels on peut construire d'autre événements plus complexes et qui couvrent nos besoins et on n'aura qu'à se placer dans la tribu engendrée par ces événements élémentaires.

Par exemple, on peut considérer pour tout $n \in N^*$ l'événement E_n : «le n^{e} lancer donne face», ou encore

$$E_n = \{(\omega_p)_{p \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*} \mid \omega_n = 1\}$$

On peut alors exprimer

- pour tout $\omega = (\omega_n)_{n\in\mathbb{N}^*}$

$$\{\omega\} = \bigcap_{n\in\mathbb{N}^*} F_n \quad \text{avec } F_n = \begin{cases} E_n & \text{si } \omega_n = 1 \\ E_n^c & \text{si } \omega_n = 0 \end{cases}$$

  en particulier $\{\infty\} := \bigcap_{n\in\mathbb{N}^*} E_n^c$ est l'événement «ne jamais obtenir face»

- l'événement «le premier face apparaît au n^{e} lancer» par

$$A_n = E_1^c \cap \cdots E_{n-1}^c \cap E_n$$

- l'événement «obtenir face une infinité de fois» par

$$B = \bigcap_{n\in\mathbb{N}^*} \Big(\bigcup_{k\geqslant n} E_k \Big).$$

- l'événement «obtenir trois faces successivement une infinité de fois» par

$$C = \bigcap_{n\in\mathbb{N}^*} \bigcup_{k\geqslant n} \big(E_k \cap E_{k+1} \cap E_{k+2} \big)$$

□

1.6 Activités

Activité 1.6.1 On lance indéfiniment une piece de monnaie non équilibrée. La probabilité d'obtenir face est $p \in]0,1[$ et les lancers sont indépendants.

(a) Quelle est la probabilité pour que le numéro du premier lancer qui donne face soit pair ? Dans quel cas cette probabilité est-elle égale à celle que le numéro du premier face soit impaire ?

Solution. On note A cet événement et A_n l'événement «le numéro du premier lancer qui donne face est n». Alors $A = \bigcup_{n\in\mathbb{N}^*} A_{2n}$, réunion d'événements deux à deux disjoints. On a $\mathbb{P}(A_n) = p(1-p)^{2n-1}$ donc

$$\mathbb{P}(A) = \sum_{n\in\mathbb{N}^*} \mathbb{P}(A_{2n}) = p \sum_{n\in\mathbb{N}^*} (1-p)^{2n-1} = \frac{p(1-p)}{1-(1-p)^2} = \frac{1-p}{2-p}$$

Par ailleur $\mathbb{P}(A^c) = \mathbb{P}(A)$ si et seulement si $\mathbb{P}(A) = 1/2$. Ce qui équivaut à $\frac{1-p}{2-p} = \frac{1}{2}$ ou encore à $p = 0$.

On constate qu'en général $0 < \mathbb{P}(A) < \frac{1}{2}$ sans jamais toucher ces valeurs extrêmes lorsque p varie dans $]0,1[$. La probabilité que le premier face ait un numéro impaire est toujours plus grande que celle que ce numéro soit paire et on n'approche d'un équilibre des deux probabilités que si p est presque nul, c'est à dire quand il est presque impossible d'obtenir face.

C'est un comportement contre-intuitif. L'intuition pousse plutôt vers $\mathbb{P}(A^c) = \mathbb{P}(A)$ sous pretexte qu'il y a «autant de nombres paires que de nombres impaires».

(b) Quelle est la probabilité pour que les numéros de tous les lancers qui donnent face soient pairs ?

Solution. On note B l'evénement «tous les lancers qui donnent face sont pairs». $(A_n)_{n\in\mathbb{N}^*}$ est un SCE donc selon la formule des probabilité totales

$$\mathbb{P}(B) = \sum_{n\in\mathbb{N}^*} \mathbb{P}(B \mid A_n)\mathbb{P}(A_n)$$

Pusique B ne se réalise pas si un événement A_{2n+1} se réalise alors $\mathbb{P}(B \mid A_{2n+1}) = 0$ pour tout $n \in \mathbb{N}$ et donc

$$\mathbb{P}(B) = \sum_{n\in\mathbb{N}^*} \mathbb{P}(B \mid A_{2n})\mathbb{P}(A_{2n})$$

Activité 1.6.2 Calcul de l'ndicatrice d'Euler. On fixe un entier $n \geqslant 2$ et on note $p_1, p_2, \ldots, p_r$ ses diviseurs premiers. On pose $\Omega = \{1, 2, \ldots, n\}$ et pour tout diviseur d de n

$$A_d = \{k \in \Omega \mid d \text{ divise } k\}$$

On munit Ω de la probabilité uniforme $\mathbb{P}$.

(a) Calculer $\mathbb{P}(A_d)$.

(b) Montrer que les événements $A_{p_1}, A_{p_2}, \ldots, A_{p_r}$ sont mutuellement indépendants.

Solution. Juste pour le test du formatage de la reponse

(c) Retrouver la formule $\varphi(n) = n \displaystyle\prod_{k=1}^{r} \left(1 - \frac{1}{p_k}\right)$

Solution. Juste pour le test du formatage

Activité 1.6.3 Formule de Weierstrass. On fixe un réel $x > 1$. On note p_n le n^{e} nombre premier par ordre croissant et on définit la probabilité $\mathbb{P}_x$ sur $(\mathbb{N}^*, \mathscr{P}(\mathbb{N}^*))$ par

$$\forall n \in \mathbb{N}^*,\ \mathbb{P}_x(\{n\}) = \frac{1}{\zeta(x)n^x}$$

On note pour tout $n \in \mathbb{N}^*$, M_n l'ensemble des multiple de n.

(a) Justifier que $\mathbb{P}_x$ définit bien une probabilité sur $\mathscr{P}(\mathbb{N}^*)$.

(b) Calculer $\mathbb{P}_x(M_n)$ pour tout $n \in \mathbb{N}^*$

(c) Montrer que les événements $M_{p_n},\ n \in \mathbb{N}^*$ sont mutuellement indépendants.

(d) En déduire que

$$\frac{1}{\zeta(x)} = \prod_{n=1}^{\infty} \left(1 - \frac{1}{p_n^x}\right)$$

Chapitre 2

Variables aléatoires discrètes

2.1 Variables aléatoires

On se donne dans ce chapitre un espace probabilisé $(\Omega, \mathscr{T}, \mathbb{P})$ et un espace probabilisable $(\Omega', \mathscr{T}')$

Définition 2.1.1 On appelle variable aléatoire de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$ toute application $X : \Omega \longrightarrow \Omega'$ telle que

$$\forall A' \in \mathscr{T}',\ X^{-1}(A') \in \mathscr{T}$$

c'est à dire que l'image réciproque par X de tout événement est un événement.

Soit dans la suite X une variable aléatoire de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$.

X est dite discrète si l'ensemble $X(\Omega)$ est au plus dénombrable et $\mathscr{P}\big(X(\Omega)\big) \subset \mathcal{T}'$. Ce qui équivaut à

$$\forall x \in X(\Omega),\ \{x\} \in \mathscr{T}'$$

X est dite réelle si $\Omega' \subset \mathbb{R}$.

Explication. Dans la pratique une variable aléatoire est utilisée pour représenter le résultat d'une expérience aléatoire. Souvent l'univers Ω et la tribu $\mathscr{T}$ ne sont pas précisés. Les résultats et les événements de l'expérience sont dans Ω' et dans $\mathscr{T}'$.
La question est comment choisir $(\Omega', \mathscr{T}')$ et de quelle tribu munir Ω pour que le résultat de l'expérience soit une variable aléatoire ? Les remarques suivantes donnent des éléments de réponse. ■

◊

Remarque 2.1.2

1. Sauf précision du contraire, un univers Ω au plus dénombrable sera systématiquement muni de la tribu $\mathscr{P}(\Omega)$.

2. Si Ω est au plus dénombrable (et muni de la tribu $\mathscr{P}(\Omega)$) alors toute application définie sur Ω est une variable aléatoire quelque soit la tribu considérée dans l'espace d'arrivée.

3. Soit f une application quelconque définie de Ω dans Ω'. L'ensemble

$$\mathscr{T}_f = \big\{f^{-1}(A') \mid A' \in \mathscr{T}'\big\}$$

est une tribu de Ω et f est une variable aléatoire de $(\Omega, \mathscr{T}_f)$ dans $(\Omega, \mathscr{T})$.

Si Ω' est au plus dénombrable et $\mathscr{T}' = \mathscr{P}(\Omega')$, on voit qu'une application quelconque de Ω dans Ω' peut être considérée comme une variable aléatoire avec très peu de contraintes. Il suffit de se placer du côté de Ω dans une tribu qui contient $\mathscr{T}_f$.

Si $f_1, f_2, \ldots, f_p$ sont des applications définies sur Ω telles que $f_k(\Omega)$ soit au plus dénombrable pour tout $k \in [\![1, n]\!]$ alors on peut poser :

- $\Omega' = \displaystyle\bigcup_{k=1}^{p} f_k(\Omega)$ et $\mathscr{T}' = \mathscr{P}(\Omega')$;
- $\mathscr{T} = \sigma\left(\displaystyle\bigcup_{k=1}^{p} \mathscr{T}_{f_k}\right)$

de telle sorte que les applications $f_1, f_2, \ldots, f_p$ soient toutes des variables aléatoires discrètes de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$.

Ses observations restent valides pour une famille dénombrable d'applications $(f_i)_{i \in I}$ définies sur Ω telle que $f_i(\Omega)$ soit au plus dénombrable pour tout $i \in I$. Notamment pour une suite de telles applications.

C'est ainsi qu'il est toujours possible de considérer un modèle dans lequel on peut combiner entre les résultats d'un nombre fini ou dénombrable d'expériences aléatoires si chacune a au plus un ensemble au plus dénombrable de résultats.

Proposition 2.1.3

1. *La composée $Y \circ X$ de deux variables aléatoires X et Y est une variable aléatoire. De plus si Y est discrète alors $Y \circ X$ est discrète.*

2. *Si pour tout $k \in [\![1, p]\!]$, X_k est une variable aléatoire de $(\Omega, \mathscr{T})$ dans un espace probabilisable $(\Omega_k, \mathscr{T}_k)$ alors l'application $(X_1, X_2, \ldots, X_p)$ définie par*

$$\forall \omega \in \Omega, (X_1, X_2, \ldots, X_p)(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_p(\omega))$$

est une variable aléatoire de $(\Omega, \mathscr{T})$ dans $(\Omega_1 \times \cdots \times \Omega_p, \mathscr{T}_1 \times \cdots \times \mathscr{T}_p)$. De plus si $X_1, X_2, \ldots, X_p$ sont discrètes alors $(X_1, X_2, \ldots, X_p)$ est discrète.

3. *Soit maintenant une variable aléatoire discrète X de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$. Alors pour toute application f définie sur $X(\Omega)$ l'application $f \circ X$ est une VAD. On la note $f(X)$*

$$\forall \omega \in \Omega,\ f(X)(\omega) := f(X(\omega))$$

On généralise de la façon suivante : si $X_1, X_2, \ldots, X_p$ sont des VAD definies sur $(\Omega, \mathscr{T})$ alors pour toute application g définie sur $X_1(\Omega) \times X_2(\Omega) \times \cdots \times X_p(\Omega)$ on définit la variable aléatoire discrète $g(X_1, X_2, \ldots, X_p)$ par

$$\forall \omega \in \Omega,\ g(X_1, X_2, \ldots, X_p)(\omega) := g(X_1(\omega), X_2(\omega), \ldots, X_p(\omega))$$

Définition 2.1.4

1. étant donné des variables aléatoires discrètes $X, X_1, \ldots, X_n$ de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$, on note

   - pour tout $x \in \mathscr{T}'$

   $$(X = x) = X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}$$

- pour tous $x_1, x_2, \ldots, x_n \in \mathscr{T}'$

$$(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \bigcap_{k=1}^{n} (X_k = x_k)$$

- pour tout $A' \in \mathscr{T}'$,

$$(X \in A') = X^{-1}(A') = \{\omega \in \Omega \mid X(\omega) \in A'\}$$

- pour tous $A'_1, A'_2, \ldots, A'_n \in \mathscr{T}'$

$$(X_1 \in A'_1, X_2 \in A'_2, \ldots, X_n \in A'_n) = \bigcap_{k=1}^{n} (X_k \in A'_k)$$

2. Une variable aléatoire discrète X est dite presque partout constante s'il existe $c \in X(\Omega)$ tel que $\mathbb{P}(X = c) = 1$. Elle est en particulier dite presque partout nulle si $\mathbb{P}(X = 0) = 1$.

$\diamondsuit$

Remarque 2.1.5 Vu la tolérance de l'image réciproque par une application envers les opérations sur les ensembles, les notations précédentes donnent

- $(X \in A') = \bigcup_{x \in A'} (X = x)$
- $(X \in \bigcup_{i \in I} A'_i) = \bigcup_{i \in I} (X \in A'_i)$
- $(X \in \bigcap_{i \in I} A'_i) = \bigcap_{i \in I} (X \in A'_i)$
- $(X \in A')^c = (X \in (A')^c) = (X \notin A')$

Exemple 2.1.6 d'utilisation de ces notations.

- Si X et Y sont des VAD à valeurs dans $\mathbb{N}$ alors pour tout $n \in N$,

$$(X + Y = n) = \bigcup_{k=0}^{n} (X = k, Y = n - k)$$

et puisque ces événements sont deux à deux disjoints alors

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n - k)$$

- Si N est une VAD à valeurs dans $\mathbb{N} \cup \{\infty\}$ alors

$$(N = \infty) = (N \in \mathbb{N})^c = \left(\bigcup_{n=0}^{+\infty} (N = n) \right)^c = \bigcap_{n=0}^{\infty} (N \neq n)$$

et particulier

$$\mathbb{P}(N = \infty) = 1 - \sum_{n=0}^{+\infty} \mathbb{P}(N = n) = \lim_{n \to +\infty} \mathbb{P}\left(\bigcap_{k=0}^{n} (N \neq k) \right)$$

- $(f(X) = y) = \bigcup_{x \in f^{-1}(y)} (X = x)$. Par exemple, si $y \in [-1, 1]$ alors $(\sin(X) = y) = \bigcup_{k \in \mathbb{Z}} (X = \arcsin(x) + 2k\pi)$

□

Théorème 2.1.7 *Soit X une variables aléatoire discrète de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$.*

1. *$\big((X = x)\big)_{x \in X(\Omega)}$ est un système complet d'événements de $(\Omega, \mathscr{T})$;*
2. *Pour tout $A' \in \mathscr{T}'$, $\mathbb{P}\big(X \in A'\big) = \sum_{x \in A'} \mathbb{P}(X = x)$*

On appelle loi de la variable X le couple $\big(X, (\mathbb{P}(X = x))_{x \in X(\Omega)}\big)$. L'application $x \longmapsto P(X = x)$ est dite fonction des masses de la variable X.

Explication. *La loi d'une VADR est ainsi le couple formé de l'ensemble des valeurs qu'elle peut prendre et de la famille (au plus dénombrable) des probabilités pour que chacune de ces valeurs se réalise. La propriété 2 signifie que les probabilités des événements $(X \in A')$ sont déterminées par la loi de X.* ■

Corollaire 2.1.8 *Soit X une variable aléatoire discrète de $(\Omega, \mathscr{T})$ dans $(\Omega', \mathscr{T}')$. L'application*

$$\begin{array}{rccc} \mathbb{P}_X : & \mathscr{T}' & \longrightarrow & [0,1] \\ & A' & \longmapsto & \mathbb{P}(X \in A') \end{array}$$

est une probabilité de $(\Omega', \mathscr{T}')$.

Définition 2.1.9 Soient deux VAD X et Y définies sur $(\Omega, \mathscr{T})$. La loi du couple (X, Y) est par définition la loi de la variable $Z = (X, Y)$. Elle est assimilée au couple $\Big(X(\Omega) \times Y(\Omega), \big(\mathbb{P}(X = x, Y = y)\big)_{x \in X(\Omega), y \in Y(\Omega)}\Big)$.

En outre les lois des variables X et Y sont appelées les lois marginales du couple (X, Y). ♢

Remarque 2.1.10 Avec $Z = (X, Y)$ on a

$$Z(\Omega) = \big\{(x, y) \in X(\Omega) \times Y(\Omega) \mid \exists \omega \in \Omega\ ;\ x = X(\omega) \text{ et } y = Y(\omega)\big\}$$

on n'a donc pas nécessairement $Z(\Omega) = X(\Omega) \times Y(\Omega)$ mais si $(x, y) \in X(\Omega) \times Y(\Omega)$ alors

- si $(x, y) \in Z(\Omega)$ alors $(Z = (x, y)) = (X = x, Y = y)$ et en particulier $\mathbb{P}\big(Z = (x, y)\big) = \mathbb{P}(X = x, Y = y)$;
- si $(x, y) \notin Z(\Omega)$ alors $(Z = (x, y)) = \emptyset$ et donc $\mathbb{P}(X = x, Y = y) = 0$.

C'est pour des raisons de simplification que la loi de couple est donc définie avec $X(\Omega) \times Y(\Omega)$.

Proposition 2.1.11 *Soient deux VAD X et Y définies sur $(\Omega, \mathscr{T})$.*

1. *$\forall y \in Y(\Omega),\ \mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y)$*
2. *$\forall x \in X(\Omega),\ \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y)$*

Ce qui signifie que les lois marginales du couple (X, Y) sont données par sa loi de couple.

2.2 Indépendance des variables aléatoires discrètes

Définition 2.2.1 Une famille $(X_i)_{i\in I}$ de VAD définies sur $(\Omega, \mathscr{T})$ (pas nécessairement à valeurs dans le même espace) est dite **mutuellement indépendante** (MI) si

$$\forall J \in \mathscr{F}(I),\ \forall (x_j)_{i\in J} \in \prod_{j\in J} X_j(\Omega),\ \mathbb{P}\Big(\bigcap_{j\in J}(X_j = x_j)\Big) = \prod_{j\in J} \mathbb{P}(X_j = x_j)$$

$\Diamond$

Remarque 2.2.2 Si $(X_i)_{i\in I}$ est une famille MI de VAD alors pour toute partie I' de I la famille $(X_i)_{i\in I'}$ est MI.

Proposition 2.2.3 *Des variables aléatoires discrètes $X_1, X_2, \ldots, X_p$ définies sur Ω sont mutuellement indépendantes si et seulement si*

$$\forall (x_1, x_2, \ldots, x_p) \in X_1(\Omega) \times X_2(\Omega) \times \cdots \times X_p(\Omega),$$
$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) = \prod_{k=1}^{p} \mathbb{P}(X_k = x_k)$$

Remarque 2.2.4

1. Cette dernière proposition simplifie considérablement la définition de l'indépendance mutuelle d'un *nombre fini* de VAD.
2. Elle implique aussi qu'une famille infinie de VAD est MI si et seulement si toutes ses sous-familles finies sont MI.
3. Une suite $(X_n)_{n\in\mathbb{N}}$ de VAD est MI si et seulement pour tout $n \in \mathbb{N}$ les variables $X_0, X_1, \ldots, X_n$ sont MI.

Proposition 2.2.5 *Soit $(X_i)_{i\in I}$ une famille de VAD mutuellement indépendantes et toutes définies sur $(\Omega, \mathscr{T})$.*

Soit pour tout $i \in I$, une partie A'_i de $X_i(\Omega)$. Alors les événements $(X_i \in A'_i), i \in I$ sont mutuellement indépendants.

Théorème 2.2.6 *Soit $(X_i)_{i\in I}$ une famille de VAD mutuellement indépendantes et toutes définies sur $(\Omega, \mathscr{T})$.*

1. *Si pour tout $i \in I$, f_i est une application définie sur $X_i(\Omega)$ alors les variables $f_i(X_i)$, $i \in I$ sont mutuellement indépendantes.*
2. *Lemme des coalitions.*

   Soit $(J_k)_{k\in K}$ une famille de parties finies deux à deux disjointes de I. Si pour tout $k \in K$, g_k est une application définie sur $\prod_{i\in J_k} X_i(\Omega)$ alors les variables $g_k\big((X_i)_{i\in J_k}\big)$, $k \in K$ sont mutuellement indépendantes.

Remarque 2.2.7

1. Si X est une VAD presque partout constante de $(\Omega, \mathscr{T})$ alors toute autre VAD définie sur $(\Omega, \mathscr{T})$ est indépendante de X.

   Explication. car pour tout $x \in X(\Omega)$, on a soit $\mathbb{P}(X = x) = 0$ soit $\mathbb{P}(X = x) = 1$. Donc l'événement $(X = x)$ est indépendant de tou autre événement de $(\Omega, \mathscr{T})$. ■

2. Soient X une VAD et f une fonction définie sur $X(\Omega)$. À moins que X ou $f(X)$ ne soit presque partout constante, les variables X et $f(X)$ ne

peuvent être indépendantes.

Explication. On suppose que X et $f(X)$ ne sont pas presque partout constantes. Il existe alors $x_1 \in X(\Omega)$ tel que $0 < \mathbb{P}(X = x_1) < 1$. Comme $(X = x_1) \subset (f(X) = f(x_1))$ alors $\mathbb{P}(f(X) = f(x_1)) > 0$. Ensuite puisque $f(X)$ est non presque partout constante alors $\mathbb{P}(f(X) = f(x_1)) < 1$ et il existe donc $x_2 \in X(\Omega)$ tel que $f(x_2) \neq f(x_1)$ et $\mathbb{P}(f(X) = f(x_2)) > 0$. Ainsi

$$\big(X = x_1, f(X) = f(x_2)\big) = \emptyset \text{ et } \mathbb{P}(X = x_1)\mathbb{P}(f(X) = f(x_2)) \neq 0$$

f et $f(X)$ ne sont donc pas indépendantes. ∎

3. *Exemples d'utilisations du lemme des coalitions.*

   Soient $X_1, \dots, X_p, Y$ des VAD définies sur $(\Omega, \mathscr{T})$.

   - Si $X_1, \dots, X_p, Y$ sont MI alors $X = (X_1, X_2, \dots, X_p)$ et Y sont MI.
   - Réciproquement si $X = (X_1, X_2, \dots, X_p)$ et Y sont MI alors tout ce qu'on peut dire c'est que Y est indépendante de X_i pour tout i.
   - Si la variable Y est elle même un vecteur de la forme $Y = (Y_1, Y_2, \dots, Y_q)$ et X et Y sont indépndantes alors X_i et Y_j sont indépendantes pour tous i, j.

   Explication. Ce sont des conséquences du lemme des coalitions en utilisant respectivement les applications :

   - $g(x_1, x_2, \dots, x_p) = (x_1, x_2, \dots, x_p)$;
   - $g_i(x_1, x_2, \dots, x_p) = x_i$;
   - $g_i(x_1, x_2, \dots, x_p) = x_i$ et $h_j(y_1, y_2, \dots, y_q) = y_j$

   ∎

2.3 Lois usuelles

X désignera une VAD définie sur $(\Omega, \mathscr{T})$

1. *Loi de Bernouilli.*

   Soit un réel $p \in [0, 1]$ On dit que X suit la loi de Bernouilli de paramètre p et on écrit $X \sim \mathscr{B}(p)$ si X est le résultat d'une expérience aléatoire qui ne possède que deux issues : succès ou échec. La probabilité du sucès étant p.

$$\begin{cases} X(\Omega) = \{1, 0\} \\ \mathbb{P}(X = 1) = p \text{ et } \mathbb{P}(X = 0) = 1 - p \end{cases}$$

2. *Loi binomiale.*

   Soit un réel $p \in [0, 1]$ et un entier $n \in \mathbb{N}^*$. On dit que X suit la loi de binomiale de paramètres n et p et on écrit $X \sim \mathscr{B}(n, p)$ si X est le nombre de succès obtenus lorsque on répète n fois de façon indépendante une expérience de Bernouilli de paramètre p.

$$\begin{cases} X(\Omega) = \{0, 1, \dots, n\} \\ \forall k \in X(\Omega),\ \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \end{cases}$$

X suit aussi la loi $\mathscr{B}(n,p)$ si elle represente le nombre de succès obtenu lorsque on effectue simultanénement et de façon indépendante n test de Bernouilli de paramètre p.

Si X_k est le résultat du k^{e} test de Bernouilli alors

$$X = X_1 + X_2 + \ldots + X_n$$
$$\forall k \in [\![0,n]\!],\ (X=k) = \bigcap_{\substack{(k_1,k_2,\ldots,k_n)\in\{0,1\}^n \\ k_1+k_2+\ldots+k_n=k}} (X_1=k_1, X_2=k_2,\ldots,X_n=k_n)$$

sachant que les variables $X_1, X_2, \ldots, X_n$ sont mutuellement indépendantes et suivent toute la loi $\mathscr{B}(p)$

3. *Loi géometrique.*

   Soit un réel $p \in]0,1[$ On dit que X suit la loi géométrique de paramètre p et on écrit $X \sim \mathscr{G}(p)$ si X est le numéro du premier test qui donne un succès lorsque on répète indéfiniment et de façon indépendante une expérience de Bernouilli de paramètre p.

   $$\begin{cases} X(\Omega) = \mathbb{N}^* \\ \forall n \in \mathbb{N}^*,\ \mathbb{P}(X=n) = p(1-p)^{n-1} \end{cases}$$

   X est aussi dite temps d'attente du premier succès.

   Si X_n est le résultat du n^{e} test de Bernouilli alors

   $$X = \min\{n \in \mathbb{N}^* \mid X_n = 1\}$$
   $$\forall n \in \mathbb{N}^*,\ (X=n) = (X_1=0,\ldots,X_{n-1}=0, X_n=1)$$

   Sachant que les variables $X_n,\ n \in \mathbb{N}^*$ sont mutuellement indépendantes et suivent toute la loi $\mathscr{B}(p)$.

4. *Loi de Poisson.*

   Soit un réel $\lambda \in \mathbb{R}_+$. On dit que X suit la loi de Poisson de paramètre λ et on écrit $X \sim \mathscr{P}(\lambda)$ si

   $$\begin{cases} X(\Omega) = \mathbb{N} \\ \forall n \in \mathbb{N},\ \mathbb{P}(X=n) = \dfrac{\lambda^n}{n!}e^{-\lambda} \end{cases}$$

   X représente le nombre de clients servis pendant une unité de temps dans une file d'attente quand on sait que le nombre *moyen* de clients par unité de temps est λ. Pour cette raison la loi de Poisson est aussi appelé loi des files d'attente.

Activité 2.3.1 Loi hypergéometrique. Soient $p \in [0,1]$ et $n, N \in \mathbb{N}^*$ avec $n \leqslant N$. On prélève de façon équiprobable un échantillon de n individus dans une population de N individus. On effectue des tests de type Bernouilli sur les individus de l'échantillon sachant que la proportion d'individu positifs au test dans toute la population est p. X est le nombre d'individus qui s'avèrent positifs au test dans l'échantillon.

1. Quelle est la loi de X ?
2. On note X_k le résultat du test du k^{e} individu. Quelle est la loi de X_k ?

Réponse. $X(\Omega) \subset [\![0,n]\!]$ et $\forall k \in X(\Omega),\ \mathbb{P}(X=k) = \dfrac{\binom{pN}{k}\binom{(1-p)N}{n-k}}{\binom{N}{n}}$

Solution.

1. Le nombre k de tests positifs dans l'échantiloon ne peut dépasser n, ni pN le nombre total d'individus positifs dans toute la population. D'un autre côté si $N - pN < n$ alors on est sûr d'avoir au moins $n - (N - pN)$ tests positifs dans l'échantillon. Ainsi

$$\max(0, n - (1-p)N) \leqslant k \leqslant \min(n, pN)$$

Ce qui suggère de prendre $X(\Omega) = [\![\max(0, n - (1-p)N), \min(n, pN)]\!]$. Mais pour simplifier on prend plutôt $X(\Omega) = [\![0, n]\!]$ quitte à considèrer que les résultats impossibles ont une probabilité nulle.

Ensuite, il y a $\binom{N}{n}$ façon de prélever équiprobablement n individus dans une population de N éléments. Parmi ces prélévements, ceux qui contiendront exactement k individus positifs sont au nombre de $\binom{pN}{k}\binom{N-pN}{n-k}$ car il s'agit de prélever k indivdus parmi pN qui sont positifs au test et $n-k$ individus parmi $N - pN$ qui ne le sont pas. Vu l'équiprobabilité des prélèvements on a donc

$$\mathbb{P}(X = k) = \frac{\binom{pN}{k}\binom{(1-p)N}{n-k}}{\binom{N}{n}}$$

On notera $\mathscr{H}(N, n, p)$ la loi de la variable X. Elle est dite loi hypergéomètrique de paramètres N, n et p.

2. Prélever un échantillon de n individus de façon équiprobable revient à prelever sans remise un à un et de façon équiporbable les n individus. Notons X_k la variable de Bernouilli qui vaut 1 si le k^{e} individu prélevé de la population est positif au test. Alors $X = X_1 + X_2 + \cdots + X_n$. La question précédente montre ainsi que pour tout $k \in [\![1, N]\!]$

$$S_k = X_1 + X_2 + \cdots + X_k \sim \mathscr{H}(N, k, p)$$

Soit maintenant $k \in [\![1, N-1]\!]$. Grâce à la formule des probabilités totales, on peut écrire

$$\mathbb{P}(X_{k+1} = 1) = \sum_{i=0}^{k} \mathbb{P}(X_{k+1} = 1 \mid S_k = i)\mathbb{P}(S_k = i)$$

$\mathbb{P}(X_{k+1} = 1 \mid S_k = i)$ est la probabilité que le $(k+1)^{\text{e}}$ individu prélevé soit positif sachant que i individus ont été positifs pour les k prélévements précédents. Dans ces condition, il reste $N-k$ individu dans la population dont $pN - i$ sont positifs. Par équiprobabilité des prélévements on a donc

$$\mathbb{P}(X_k = 1 \mid S_{k-1} = i) = \frac{pN - i}{N - k}$$

Ainsi

$$\begin{aligned}
\mathbb{P}(X_{k+1} = 1) &= \sum_{i=0}^{k} \frac{pN - i}{N - k} \cdot \frac{\binom{pN}{i}\binom{(1-p)N}{k-i}}{\binom{N}{k}} \\
&= \frac{1}{(N-k)\binom{N}{k}} \left(pN \sum_{i=0}^{k} \binom{pN}{i}\binom{(1-p)N}{k-i} - \sum_{i=0}^{k} i\binom{pN}{i}\binom{(1-p)N}{k-i} \right) \\
&= \frac{pN}{(N-k)\binom{N}{k}} \left(\sum_{i=0}^{k} \binom{pN}{i}\binom{(1-p)N}{k-i} - \sum_{i=1}^{k} \binom{pN-1}{i-1}\binom{(1-p)N}{k-i} \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{pN}{(N-k)\binom{N}{k}}\left(\binom{N}{k}-\binom{N-1}{k-1}\right)\\
&= \frac{pN}{(N-k)\binom{N}{k}}\binom{N-1}{k}\\
&= \frac{pN}{N-k}\frac{(N-1)!}{k!(N-1-k)!}\frac{k!(N-k)!}{N!}\\
&= p
\end{aligned}$$

Il en ressort que malgré le changement de la répartition des cas positifs/ cas négatifs après chaque prélévement, la probabulité de prélever un cas positif est toujours p.

Activité 2.3.2 Loi du temps d'attente du k^{e} succès. Soient $p \in]0,1[$ et $k \in N^*$. Quel est la loi du temps d'attente du k^{e} succés lorsque on répète indéfiniment et de façon indépendante une exprérience de Bernouilli de paramètre p

Réponse. $X(\Omega) = \{n \in \mathbb{N} \mid n \geqslant k\}$ et $\forall n \in X(\Omega)$, $\mathbb{P}(X = n) = \binom{n-1}{k-1}p^k(1-p)^{n-k}$

Activité 2.3.3 Loi d'une marche aléatoire sur une droite. Un objet se déplace sur une droite graduée. À chaque instant il ne peut qu'avancer d'un pas avec une probabilité p ou reculer d'un pas avec une probabilité $1-p$. Les déplacements sont tous indépendants.

1. On note X_n la position de l'objet sur la droite au n^{e} pas en supposant qu'il était sur l'origine de la droite à l'instant 0. Quelle est la loi de X_n ?
2. On note N le numéro du premier pas pour lequel l'objet revient sur l'origine. Quelle est la loi de N ?

Réponse.

1. $X_n(\Omega) = [\![-n, n]\!]$ et $\mathbb{P}(X_n = k) = \begin{cases} 0 & \text{si } k+n \text{ est impaire} \\ \binom{n}{\frac{n+k}{2}} & \text{si } k+n \text{ est paire} \end{cases}$

Activité 2.3.4 Une indépendance contre-intuitive. N suit une loi de Poisson de paramètre λ. X est le nombre de succès qunad on répète de façon indépendante N test de Bernouilli de paramètre p.

1. Déterminer la loi de X
2. Vérifier que X et $N-X$ sont indépendantes.

Solution.

1. N peut potentiellement prendre toutes les valeurs dans $\mathbb{N}$. Il en est de même pour X. Ensuite pour tout $k \in \mathbb{N}$

$$\begin{aligned}
\mathbb{P}(X = k) &= \sum_{n=k}^{+\infty} \mathbb{P}(X = k \mid N = n)\mathbb{P}(N = n)\\
&= \sum_{k=n}^{+\infty} \binom{n}{k} p^k(1-p)^{n-k} \cdot \frac{\lambda^n}{n!}e^{-\lambda}
\end{aligned}$$

$$
\begin{aligned}
&= \lambda^k e^{-\lambda} \frac{p^k}{k!} \sum_{n=k}^{+\infty} \frac{((1-p)\lambda)^{n-k}}{(n-k)!} \\
&= \frac{(p\lambda)^k}{k!} e^{-p\lambda}
\end{aligned}
$$

Ainsi $X \sim \mathscr{P}(p\lambda)$

2. $N - X$ est le nombre d'echecs pour N tests. Il suffit donc de remplacer p par $q = 1 - p$ dans le calcul de la loi de X : $N - X \sim \mathscr{P}(q\lambda)$. Ensuite si $k, h \in \mathbb{N}$ alors

$$
\begin{aligned}
\mathbb{P}(X = k, N - X = h) &= \mathbb{P}(X = k, N = k + h) \\
&= \mathbb{P}(X = k \mid N = k + h)\mathbb{P}(N = k + h) \\
&= \binom{k+h}{k} p^k q^h \cdot \frac{\lambda^{k+h}}{(k+h)!} e^{-\lambda} \\
&= \frac{(p\lambda)^k}{k!} e^{-p\lambda} \cdot \frac{(q\lambda)^h}{h!} e^{-q\lambda} \\
&= \mathbb{P}(X = k)\mathbb{P}(N - X = h)
\end{aligned}
$$

X et $N - X$ sont donc bien indépendantes contrairement à «l'intuition». (X et le nombre de succés et $N - X$ le nombre d'echecs pour N tests.)

Colophon

Ce cours a été crée avec PreTeXt.