Chapitre 10

Variables aléatoires discrètes

Ce chapitre est difficile, car il faut avant tout comprendre comment on formalise le calcul des probabilités dans un espace probabilisable. Ici, on reviendra rapidement sur les espaces probabilisés discrets, avant d'introduire la notion de variable aléatoire discrète. Si celle-ci est riche, c'est d'abord parce qu'elle nous offrira de nombreux problèmes faisant intervenir d'autres chapitres, que ce soit en algèbre ou en analyse.

Programmes 2022

Pour aller plus loin

Depuis quelques années maintenant, on retrouve à l'écrit de nombreux problèmes mettant en jeu les espaces probabilisés, du théorème de Perron-Frobenius au théorème de Moivre-Laplace, en passant par l'étude de marches aléatoires... c'est donc un chapitre incontournable et il faudra vite prendre l'habitude de manipuler le formalisme sous-jacent.

1 Espaces probabilisés

Quand on étudie une expérience aléatoire, on cherche très souvent à construire un **modèle probabiliste** qui pourrait lui être associé : pour cela, on définit généralement un **univers** Ω, l'ensemble des résultats de l'expérience et on cherche à définir une **mesure de probabilité** P qui à tout **évènement** $A \in \mathcal{P}(\Omega)$ renvoie un poids positif qui pourra rendre compte de la probabilité que celui-ci se réalise.

Remarques

1. Une mesure de probabilité sera donc souvent définie sur $\mathcal{P}(\Omega)$ ou une partie de $\mathcal{P}(\Omega)$, et il s'agira de bien faire la différence entre ces deux ensembles :
$$\Omega \in \mathcal{P}(\Omega), \text{ mais } \Omega \neq \mathcal{P}(\Omega)$$
Et si Ω est de cardinal fini $n \in \mathbb{N}^*$, on rappelle que $\text{card}(\mathcal{P}(\Omega)) = 2^n$. Par exemple, on pourra considérer :
$$\begin{cases} \Omega = [\![1,3]\!] = \{1,2,3\} \\ \mathcal{P}(\Omega) = \{\emptyset,\ \{1\},\{2\},\{3\},\ \{1,2\},\{1,3\},\{2,3\},\ \{1,2,3\}\} \end{cases}$$

2. Suivant les expériences aléatoires considérées, l'univers Ω peut être infini et même non dénombrable. Dans ce dernier cas, il est très difficile de décrire tous les évènements et donc de définir correctement une mesure de probabilité sur $\mathcal{P}(\Omega)$. Le plus souvent, on préfèrera alors travailler sur une partie de $\mathcal{P}(\Omega)$, plus simple et qui nous permettra de décrire les évènements à l'aide d'opérations ensemblistes : c'est pour cela qu'on introduit la notion de **tribu des évènements**.

1.1 Premières définitions et propriétés sur un espace probabilisé

Définition Soit Ω un ensemble non vide. On appelle alors **tribu des évènements** tout sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ tel que :

- $\Omega \in \mathcal{A}$
- $\mathcal{A}$ est stable par passage au complémentaire, c'est à dire : $\forall A \in \mathcal{A},\ \overline{A} \in \mathcal{A}$
- $\mathcal{A}$ est stable par réunion finie ou infinie dénombrable, c'est à dire que pour toute suite (A_n) finie ou infinie de $\mathcal{A}^{\mathbb{N}}$,
$$\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$$

Muni d'une telle tribu, on dit que $(\Omega, \mathcal{A})$ est un **espace probabilisable** et les éléments de $\mathcal{A}$ désignent encore les **évènements** de l'expérience aléatoire.

Propriété 1 (immédiate)**.**

Soit $(\Omega, \mathcal{A})$ un espace probabilisable, alors on a également :

1. $\emptyset \in \mathcal{A}$
2. $\mathcal{A}$ est stable par intersection finie ou infinie dénombrable, c'est à dire que pour toute suite (A_n) finie ou infinie de $\mathcal{A}^{\mathbb{N}}$,
$$\cap_{n\in\mathbb{N}} A_n \in \mathcal{A}$$

Définition Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle **système complet d'évènements** toute famille d'évènements finie ou infinie dénombrable $(A_i)_{i\in I}$ tels que :
$$\begin{cases} \forall (i,j) \in I^2, i \neq j,\ A_i \cap A_j = \emptyset, \text{ c'est à dire que ces ensembles sont \textbf{disjoints} ou que ces évènements sont \textbf{incompatibles}} \\ \bigcup_{i\in I} A_i = \Omega, \text{ c'est à dire qu'ils recouvrent } \Omega \end{cases}$$
On peut aussi dire que ces évènements forment une **partition** de Ω et on pourra noter : $\Omega = \sqcup_{i\in I} A_i$.

Définition Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle enfin **mesure de probabilité** ou plus simplement **probablité** toute application $P : \mathcal{A} \longrightarrow \mathbb{R}_+$ telle que :
$$\begin{cases} P(\Omega) = 1 \\ P \text{ est } \sigma\text{-additive, autrement dit pour toute suite finie ou infinie dénombrable d'évènements } (A_i)_{i\in I} \text{ incompatibles :} \end{cases}$$
$$P(\sqcup_{i\in I} A_i) = \sum_{i\in I} P(A_i)$$
Dans ce cas, on dit que $(\Omega, \mathcal{A}, P)$ est un **espace probabilisé**.

Remarque La plupart du temps, on travaillera dans le cas où Ω désigne un ensemble fini ou infini dénombrable, c'est à dire qu'on aura :

$$\Omega = \sqcup_{i\in I}\{\omega_i\}, \text{ avec } I \text{ un ensemble d'indices au plus dénombrable}$$

Les **évènements élémentaires** $\{\omega_i\}$ forment ici une partition de Ω, et en considérant la tribu $\mathcal{P}(\Omega)$, appelée **tribu discrète**, on peut toujours construire une mesure de probabilité en définissant P une application σ-additive et pour laquelle on affecte à chacun de ces évènements élémentaires un poids $p_i \geq 0$ tel que :

$$\sum_{i\in I} p_i = 1$$

En particulier, quand I sera infini dénombrable, cela reviendra à étudier la convergence d'une série à termes positifs.

Propriété 2 (sur un espace probabilisé).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Alors, on a :

1. pour toute famille finie $A_1, \ldots, A_n$ d'évènements incompatibles de $\mathcal{A}$, $P(A_1 \sqcup \ldots \sqcup A_n) = \sum_{i=1}^n P(A_i)$.
2. $P(\emptyset) = 0$.
3. pour tout $A \in \mathcal{A}$, $P(\overline{A}) = 1 - P(A)$.
4. pour tout couple $(A, B) \in \mathcal{A}^2$ tel que $A \subset B$, $P(B\backslash A) = P(B) - P(A)$. Et en particulier, $P(A) \leq P(B)$: l'application est donc croissante au sens de l'inclusion.
5. pour tout couple $(A, B) \in \mathcal{A}^2$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

► *A chaque fois, on revient à une réunion disjointe d'évènements bien choisis et on invoque la σ-additivité de P.*

Remarque En fait, cette dernière propriété nous permet de montrer que globalement pour toute famille finie d'évènements, on a toujours l'inégalité :

$$P(A_1 \cup \ldots \cup A_n) \leq \sum_{i=1}^{n} P(A_i)$$

Mais si on le souhaite, on peut aussi généraliser la formule obtenue. Par exemple, par associativité, on a :

$$P(A_1\cup A_2\cup A_3) = P((A_1\cup A_2)\cup A_3) = P(A_1\cup A_2)+P(A_3)-P((A_1\cup A_2)\cap A_3) = P(A_1\cup A_2)+P(A_3)-P((A_1\cap A_3)\cup(A_2\cap A_3))$$

$$= P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

C'est un cas particulier de la **formule du crible de Poincaré** qu'on peut établir par récurrence sur $n \geq 2$ de sorte que :

$$P(\cup_{i=1}^n A_i) = \sum_{k=1}^{n} \left((-1)^{k-1} \sum_{J\subset [\![1,n]\!], Card(J)=k} P(\cap_{j\in J} A_j) \right)$$

Théorème 3 (de la limite monotone pour une suite d'évènements).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons $(A_n) \in \mathcal{A}^{\mathbb{N}}$ une suite d'évènements.

1. Si on suppose de plus que (A_n) est croissante, c'est à dire que pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$, alors :
$$P(\cup_{n=0}^{+\infty} A_n) = \lim_{n\to+\infty} P(A_n)$$
2. Si on suppose de plus que (A_n) est décroissante, c'est à dire que pour tout $n \in \mathbb{N}$, $A_{n+1} \subset A_n$, alors :
$$P(\cap_{n=0}^{+\infty} A_n) = \lim_{n\to+\infty} P(A_n)$$

► *Pour le premier point, on pose $B_n = A_n\backslash A_{n-1}$ et on montre que (B_n) désigne une suite d'évènements incompatibles tels que $\cup_{n=0}^{+\infty} B_n = \cup_{n=0}^{+\infty} A_n$, on en déduit la valeur par σ-additivité. Pour le second point, il suffit de passer au complémentaire.*

Corollaire 4 (conséquences du théorème de la limite monotone).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons $(A_n) \in \mathcal{A}^{\mathbb{N}}$ une suite d'évènements. Alors, on a plus généralement :

1. $P(\cup_{n=0}^{+\infty} A_n) = \lim_{n\to+\infty} P(\cup_{k=0}^{n} A_k)$
2. $P(\cap_{n=0}^{+\infty} A_n) = \lim_{n\to+\infty} P(\cap_{k=0}^{n} A_k)$

▶ *On introduit ici les suites $B_n = \cup_{k=0}^{n} A_k$ et $C_n = \cap_{k=0}^{n} A_k$ et on exploite le théorème de la limite monotone pour une suite d'évènements.*

Remarque Ce dernier résultat nous permet alors de prolonger par passage à la limite l'inégalité obtenue pour la probablité d'une réuion d'évènements, et ainsi pour toute suite $(A_n) \in \mathcal{A}^{\mathbb{N}}$, on retiendra :

$$P(\bigcup_{k=0}^{+\infty} A_k) \leq \sum_{k=0}^{+\infty} P(A_k)$$

avec $\sum_{k=0}^{+\infty} P(A_k) \in \mathbb{R}_+ \cup \{+\infty\}$. Cette inégalité dans $\overline{\mathbb{R}}$ est aussi appelée **inégalité de Boole** et elle nous donne ainsi une majoration pour toute réunion dénombrable d'évènements.

Définition Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. On rappelle enfin que pour tout évènement $A \in \mathcal{A}$:

- A est dit **négligeable** si $P(A) = 0$.
- A est dit **presque sûr** ou **presque certain** si $P(A) = 1$.

Exemple 1 On effectue une suite infinie de lancers d'une pièce équilibrée. On note pour tout $n \in \mathbb{N}^*$, A_n l'évènement "au cours des n premiers lancers, on a obtenu au moins une fois pile".

Quelle est la probabilité de l'évènement $E =$ "on a obtenu au moins une fois pile" ?

1.2 Cas particulier des espaces probabilisés discrets

Théorème 5 (mesure de probabilité sur un univers au plus dénombrable).

Soit Ω un ensemble fini ou infini dénombrable tel que $\Omega = \sqcup_{i\in I} \{\omega_i\}$, et considérons $(p_i)_{i\in I}$ une famille de nombres réels tels que :

$$\forall i \in I,\ p_i \geq 0 \text{ et } \sum_{i\in I} p_i = 1$$

Alors, il existe une unique probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ telle que pour tout $i \in I$, $P(\{\omega_i\}) = p_i$. Cette application est définie pour tout $A \in \mathcal{P}(\Omega)$ par :

$$P(A) = \sum_{i\in I_A} p_i,\ \text{ avec } I_A = \{i \in I, \omega_i \in A\}$$

▶ *On procède par analyse-synthèse et on reviendra à la définition d'une mesure de probabilité : on pourra d'ailleurs invoquer le théorème de sommation par paquets pour obtenir la σ-additivité.*

Exemple 2

1. Fixons $p \in]0,1[$ et on définit pour tout $n \in \mathbb{N}^*$,
$$p_n = (1-p)^{n-1} p$$
Montrer que (p_n) définit une mesure de probabilité sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$.
2. Fixons $\lambda > 0$ et on définit pour tout $n \in \mathbb{N}$,
$$p_n = e^{-\lambda} \frac{\lambda^n}{n!}$$
Montrer que (p_n) définit une mesure de probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.
3. On se place dans $\mathbb{N}^*$ et on définit pour tout $n \in \mathbb{N}^*$,
$$p_n = \frac{1}{n(n+1)}$$
   (a) Montrer que (p_n) définit une mesure de probabilité sur $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$.
   (b) On note P l'unique probabilité associée à la suite (p_n). Déterminer la probabilité qu'un entier soit pair ?

Définition Soit $n \in \mathbb{N}^*$. On se place dans le cas particulier où $\Omega = \{\omega_1, \dots, \omega_n\}$ est un ensemble fini et on note encore $\mathcal{A} = \mathcal{P}(\Omega)$.
On appelle alors **probabilité uniforme** la probabilité $P : \mathcal{A} \longrightarrow \mathbb{R}_+$ définie par :

$$\forall\, i \in [\![1, n]\!],\ P(\{\omega_i\}) = \frac{1}{\text{card}(\Omega)}$$

de sorte que pour tout évènement $A \in \mathcal{A}$, $P(A) = \dfrac{\text{card}(A)}{\text{card}(\Omega)}$.

Remarque Les espaces probabilisés finis, munis de la probabilité uniforme, nous permettent de modéliser les situations les plus courantes, et il sera alors très simple de déterminer la probabilité d'un évènement A, à condition d'être capable de **dénombrer les éléments** de A. D'ailleurs, ce seront souvent des exercices de dénombrement plutôt que des exercices de probabilités...

Pour cela, on peut rappeler que :

- le nombre de façons de choisir k éléments distincts dans un ensemble à n éléments est donné par le coefficient binomial:

$$\tbinom{n}{k} = \frac{n!}{k!(n-k)!}$$

- dans le cas particulier où il faudra tenir compte de l'ordre de ces k éléments, on n'hésitera pas à multiplier le coefficient $\tbinom{n}{k}$ par $k!$, le nombre de permutations de ces k éléments et ainsi, le nombre d'arrangements de k éléments distincts dans un ensemble à n éléments est donné par :

$$k!\tbinom{n}{k} = \frac{n!}{(n-k)!}$$

Exemple 3 (Problème des anniversaires)
Des étudiants au nombre de n sont réunis dans un amphithéâtre.
Quelle est la probabilité qu'au moins deux étudiants aient leur anniversaire le même jour ? On suppose que qu'aucun étudiant n'est né le 29 février et que $n \leq 365$.

Exemple 4 (Problème des rencontres)
Soit $n \geq 2$. Lors d'un bal auquel participent n couples, le choix de sa cavalière pour la première danse se fait au hasard. La probabilité pour un danseur de danser avec sa femme est donc $1/n$ (assez faible lorsque n est grand).
Malgré tout, montrer qu'il existe un rang à partir duquel la probabilité qu'il y ait au moins une rencontre entre un danseur et sa femme est supérieure ou égale à la probabilité qu'il n'y en ait pas ?
Pour tout $i \in [\![1, n]\!]$, on pourra noter $A_i =$"le danseur i se retrouve avec sa femme".

Exemple 5 (CCINP 104)
Soit $n \geq 3$. On dispose de n boules numérotées de 1 à n et d'une boîte formée de 3 compartiments numérotées de 1 à 3.
On lance simultanément les n boules et elles viennent toutes se ranger aléatoirement dans les 3 compartiments.

1. Déterminer la probabilité qu'il y ait deux compartiments vides ? la probabilité qu'il y ait un seul compartiment vide ?
2. En déduire la probabilité qu'il n'y ait aucun compartiment de vide.

Exemple 6 (Nombre de dérangements)
On note pour tout $(n, k) \in \mathbb{N}^2$ tel que $k \leq n$, $F_{n,k}$ le nombre de permutations de $[\![1, n]\!]$ ayant exactement k points fixes et on définit le nombre de dérangements pour tout $n \in \mathbb{N}^*$ par $\alpha_n = F_{n,0}$.
Et on convient que $\alpha_0 = 1$.

1. Montrer que pour tout $(n, k) \in \mathbb{N}^2$ tel que $k \leq n$, $F_{n,k} = \binom{n}{k}\, \alpha_{n-k}$. En déduire que pour tout $n \in \mathbb{N}$, $n! = \sum_{k=0}^{n} \binom{n}{k}\, \alpha_k$.
2. On considère la série entière $\sum \dfrac{\alpha_n}{n!} z^n$ et on note R son rayon de convergence, S sa somme.
   (a) Etablir que $R \geq 1$ et que pour tout $z \in \mathbb{C}, |z| < 1$, $S(z) = \dfrac{e^{-z}}{1-z}$.
   (b) En déduire que pour tout $n \in \mathbb{N}$, $\alpha_n = n! \sum_{k=0}^{n} \dfrac{(-1)^k}{k!}$.
   (c) Justifier alors que pour tout $n \geq 2$, $\alpha_n = E(\dfrac{n!}{e} + \dfrac{1}{2})$. *On pourra calculer* $|\alpha_n - n!/e|$.
3. On choisit au hasard une permutation de S_n, $n \geq 2$.
   Quelle est la probabilité $P(D_n)$ que celle-ci soit un dérangement ? Que peut-on dire de $\lim_{n \to +\infty} P(D_n)$?

1.3 Probabilités conditionnelles et indépendance

Théorème 6 (définition de la probabilité conditionnelle)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons $A \in \mathcal{A}$ tel que $P(A) > 0$. Alors, l'application $P_A : \mathcal{A} \longrightarrow \mathbb{R}$ définie par :

$$P_A : B \longmapsto \frac{P(A \cap B)}{P(A)}$$

est une probabilité sur $(\Omega, \mathcal{A})$ appelée **probabitité conditionnelle sachant** A. De plus, pour tout évènement B, on notera souvent $P_A(B)$ ou $P(B|A)$ la probabilité que B se réalise sachant que l'évènement A s'est réalisé.

▶ *On revient simplement à la définition d'une mesure de probablité.*

Propriété 7 (formule des probabilités composées)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Alors, pour toute famille finie d'évènements $A_1, \ldots, A_n$ tels que $P(A_1 \cap \ldots \cap A_{n-1}) \neq 0$, on a :

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P_{A_1}(A_2)P_{A_1 \cap A_2}(A_3) \ldots P_{A_1 \cap \ldots \cap A_{n-1}}(A_n)$$

▶ *On peut procéder de plusieurs façons : par récurrence ou alors, on factorise par $P(A_1)$ et on tente de reconnaître un produit télescopique.*

Propriété 8 (formule des probabilités totales)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons $(A_i)_{i \in I}$ un système complet d'évènements de sorte que $\Omega = \sqcup_{i \in I} A_i$. Si de plus, pour tout $i \in I$, $P(A_i) > 0$, alors pour tout évènement $B \in \mathcal{A}$,

$$P(B) = \sum_{i \in I} P(B \cap A_i) = \sum_{i \in I} P(A_i)P_{A_i}(B)$$

▶ *C'est immédiat : cela découle de la partition donnée par le système complet d'évènements.*

Remarque On peut aussi étendre cette formule lorsque certains évènements du système vérifient $P(A_i) = 0$. En effet, comme $B \cap A_i \subset A_i$, alors pour ces indices :

$$P(B \cap A_i) = 0$$

Ainsi, en convenant que dans ce cas, $P(A_i)P_{A_i}(B) = 0$, on obtient la formule des probabilités totales pour tout système complet d'évènements.

Cette formule des probabilités totales est très utile : c'est souvent elle qui nous donnera la relation de récurrence dans l'étude de certains processus de Markov, ces situations pour lesquelles l'état à l'instant $n + 1$ ne dépend que de l'état à l'instant n.

Exemple 7 Une municipalité souhaite proposer un service de véhicules en libre service. Lors d'une première expérience, on installe quelques véhicules en 3 lieux stratégiques de la ville : les places A, B et C. Avec ces véhicules, on peut effectuer un trajet vers l'une des deux autres places et après quelques mois, on observe les résultats suivants :

- si un véhicule est en A, il se déplace vers B avec une probabilité $3/4$ et vers C avec une probabilité $1/4$;
- si un véhicule est en B, il se déplace vers A avec une probabilité $3/4$ et vers C avec une probabilité $1/4$;
- si un véhicule est en C, il se déplace vers B avec une probabilité $3/4$ et vers A avec une probabilité $1/4$.

On s'intéresse alors au déplacement d'un véhicule. Pour tout $n \in \mathbb{N}$, on note A_n, B_n ou C_n les évènements "à l'instant n, le véhicule est en place A, B ou C", et on note pour tout $n \in \mathbb{N}$,

$$a_n = P(A_n), b_n = P(B_n), c_n = P(C_n)$$

1. En posant $X_n = \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix} \in \mathcal{M}_{31}(\mathbb{R})$, montrer qu'il existe une matrice $M \in \mathcal{M}_3(\mathbb{R})$ telle que pour tout $n \in \mathbb{N}$, $X_{n+1} = MX_n$.
2. Justifier que les valeurs propres de M sont toutes réelles, puis établir que M admet trois valeurs propres distinctes telles que $-1 < \lambda_1 < \lambda_2 < \lambda_3 = 1$.
3. Montrer qu'il existe un unique vecteur $U_0 \in E_M(1)$ tel que la somme de ses composantes soit égale à 1.

4. On considère alors une base de vecteurs propres $B = (U_0, U_1, U_2)$ associée aux valeurs propres $1, \lambda_1, \lambda_2$ et on peut écrire $X_0 = \alpha_0 U_0 + \alpha_1 U_1 + \alpha_2 U_2$ dans cette base.
Montrer que pour tout $n \in \mathbb{N}$, $X_n = \alpha_0 U_0 + \alpha_1 \lambda_1^n U_1 + \alpha_2 \lambda_2^n U_2$, puis justifier que nécessairement $\alpha_0 = 1$.

5. En déduire que les suites $(a_n), (b_n)$ et (c_n) sont convergentes et déterminer leur limite.

Propriété 9 (formule de Bayes)**.**

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

1. Si A et B désignent deux évènements de probabilité non nulle, alors :

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

2. Plus généralement, si $(A_i)_{i\in I}$ est un système complet d'évènements avec pour tout $i \in I$, $P(A_i) > 0$, alors pour tout évènement $B \in \mathcal{A}$ de probabilité non nulle,

$$P_B(A_i) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P_{A_i}(B)}{\sum_{i\in I} P(A_i)P_{A_i}(B)}$$

► *Il suffit de revenir à la définition de la probabilité conditionnelle. Pour le deuxième point, on rappellera la formule des probabilités totales afin de transformer le dénominateur.*

Remarque Cette dernière formule est très pratique : elle nous permet de calculer des probabilités a posteriori, en inversant tout simplement le conditionnement.

Définition Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. On dit que deux évènements A et B sont **indépendants** si :

$$P(A \cap B) = P(A)P(B)$$

Plus généralement, pour toute famille quelconque d'évènements $(A_i)_{i\in I}$, on dit qu'ils sont :

- **deux à deux indépendants** si pour tout $(i, j) \in I^2, i \neq j$, $P(A_i \cap A_j) = P(A_i)P(A_j)$.
- **mutuellement indépendants** si pour toute sous-famille finie $(A_i)_{i\in J}$, on a :

$$P(\cap_{i\in J} A_i) = \prod_{i\in J} P(A_i)$$

Remarques

1. Si l'évènement A est négligeable, alors, comme $A \cap B \subset A$, on a toujours $P(A \cap B) = 0 = P(A)P(B)$, et ainsi un évènement négligeable est toujours indépendant de tout évènement.
2. Il ne faut pas confondre l'incompatibilité, c'est à dire $A \cap B = \emptyset$ et l'indépendance de deux évènements... il existe des évènements incompatibles qui ne sont pas indépendants. On retiendra surtout qu'il s'agit d'abord d'une notion probabiliste et on préfèrera toujours revenir à la définition pour justifier que des évènements sont indépendants.
3. Par défaut, et sans aucune indication contraire, on pourra considérer que des évènements donnés indépendants dans un énoncé sont **mutuellement indépendants**. En particulier, cela entraîne évidemment qu'ils sont deux à deux indépendants mais la réciproque est fausse !

Propriété 10 (indépendance et évènements contraires)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. Si de plus A et B sont indépendants, alors :

$$\begin{cases} A \text{ et } \overline{B} \text{ sont indépendants} \\ \overline{A} \text{ et } B \text{ sont indépendants} \\ \overline{A} \text{ et } \overline{B} \text{ sont indépendants} \end{cases}$$

► *Il suffit de revenir à la définition à l'aide du produit des probabilités.*

Remarque En fait, si $(A_i)_{i\in I}$ désigne une famille d'évènements mutuellement indépendants, alors on peut facilement généraliser la propriété précédente et en notant $B_i \in \{A_i, \overline{A_i}\}$, les évènements $(B_i)_{i\in I}$ seront encore mutuellement indépendants.

Exemple 8 On se place dans un espace probabilisé $(\Omega, \mathcal{A}, P)$ et on considère une suite d'évèements $(A_n) \in \mathcal{A}^{\mathbb{N}}$.

1. On suppose que la série $\sum P(A_n)$ converge. Montrer que :
$$P(\bigcap_{k=0}^{+\infty} \bigcup_{p=k}^{+\infty} A_p) = 0$$
2. On suppose que la série $\sum P(A_n)$ diverge et que les évènements (A_n) sont mutuellement indépendants. Montrer que :
$$P(\bigcap_{k=0}^{+\infty} \bigcup_{p=k}^{+\infty} A_p) = 1$$
On pourra étudier $P(\cap_{p=k}^{+\infty} \overline{A_p})$ et utiliser l'inégalité de convexité $1 - x \leq e^{-x}$ sur $[0, 1]$.

Ces deux résultats sont classiques et ils sont appelés ***premier et second lemme de Borel-Cantelli****.*

2 Variables aléatoires discrètes

2.1 Loi d'une variable aléatoire et premiers exemples fondamentaux

Définition Soient $(\Omega, \mathcal{A})$ un espace probabilisable et E un ensemble quelconque. On appelle **variable aléatoire discrète** toute application $X : \Omega \longrightarrow E$ telle que :
$$\begin{cases} X(\Omega) = \{x_i,\ i \in I\} \text{ est fini ou infini dénombrable} \\ \forall\, i \in I,\ X^{-1}(\{x_i\}) \in \mathcal{A} \end{cases}$$
En particulier, si $E = \mathbb{R}$, on dit que X est une **variable aléatoire discrète à valeurs réelles**.

Notation Généralement, on adopte des notations probabilistes et on note :

- $(X = x_i)$ l'évènement $X^{-1}(\{x_i\}) = \{\omega \in \Omega,\ X(\omega) = x_i\}$,
- $(X \in A)$ l'évènement $X^{-1}(A) = \{\omega \in \Omega,\ X(\omega) \in A\}$.

Ces variables aléatoires sont utiles et elles nous vont nous permettre de réinterpréter quantitativement des situations aléatoires sur un nouvel espace probabilisé induit par X.

Théorème 11 (loi d'une variable aléatoire discrète)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X une variable aléatoire discrète. Alors, l'application $P_X : \mathcal{P}(X(\Omega)) \longrightarrow \mathbb{R}$ définie par :
$$P_X : A \longmapsto P(X \in A)$$
est une probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$ appelée **loi de** X.

► *On revient à la définition d'une mesure de probabilité et on utilise les propriétés de l'image réciproque.*

Corollaire 12 (probabilité d'un évènement sur $P(X(\Omega))$)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X une variable aléatoire discrète telle que :
$$X(\Omega) = \{x_i,\ i \in I\}$$
Alors, $((X = x_i))_{i \in I}$ désigne un système complet d'évènements sur Ω et pour tout $A \in \mathcal{P}(X(\Omega))$, et en notant $I_A = \{i \in I,\ x_i \in A\}$,
$$P_X(A) = P(X = A) = \sum_{i \in I_A} P(X = x_i)$$

Remarque Concrètement, connaître la loi d'une variable aléatoire discrète, c'est donc être capable d'identifier les valeurs prises par X et de préciser le poids des évènements élémentaires $p_i = P(X = x_i)$ tels que :
$$\forall\, i \in I,\ p_i \geq 0 \text{ et } \sum_{i \in I} p_i = 1$$
Et encore une fois, quand I sera infini dénombrable, cela reviendra à étudier la convergence d'une série à termes positifs.

Définition Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. On dit alors que deux variables aléatoires discrètes X et Y **suivent la même loi** si $X(\Omega) = Y(\Omega)$ et pour tout évènement élémentaire $\{\omega_i\} \subset X(\Omega) = Y(\Omega)$,

$$P(X = \omega_i) = P(Y = \omega_i)$$

Dans ce cas, on note de façon abusive $X \sim Y$.

Remarques

1. Attention, cela ne signifie en aucun cas que $X = Y$, car même si elles se comportent de la même façon, elles peuvent représenter des situations probabilistes complètement différentes !

2. Dans de nombreux exercices, on nous demandera de justifier que X suit une loi donnée et ainsi, on sera parfois amené à reconnaître une situation courante, et pour laquelle X suit une loi de référence. Il est donc très important de maitriser quelques **modèles probabilistes de référence**.

Définition Soit $n \in \mathbb{N}^*$. On dit qu'une variable aléatoire X suit une **loi uniforme** sur $[\![1, n]\!]$ si :

$$\begin{cases} X(\Omega) = [\![1, n]\!] \\ \forall\ k \in [\![1, n]\!],\ P(X = k) = \dfrac{1}{n} \end{cases}$$

Dans ce cas, on note $X \sim \mathcal{U}([\![1, n]\!])$.

Remarque Pour cette variable aléatoire, tous les évènements élémentaires ont donc la même chance de se réaliser : il s'agit en fait de la probabilité uniforme.

Définition Soit $p \in]0, 1[$. On dit qu'une variable aléatoire X suit une **loi de Bernoulli** de paramètre p si :

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

Dans ce cas, on note $X \sim \mathcal{B}(p)$.

Remarque Plus généralement, les évènements $(X = 1)$ et $(X = 0)$ traduisent une **épreuve de Bernoulli** à deux issues : ils seront appelés **succès** et **échec**. Ainsi, on utilisera ce modèle lorsque X rendra compte du succès ou non d'une expérience aléatoire.

Définition Soit $p \in]0, 1[$. On dit qu'une variable aléatoire X suit une **loi binomiale** de paramètres n, p si :

$$\begin{cases} X(\Omega) = [\![0, n]\!] \\ \forall\ k \in [\![0, n]\!],\ P(X = k) = \binom{n}{k} p^k q^{n-k},\ \text{avec } q = 1 - p \end{cases}$$

Dans ce cas, on note $X \sim \mathcal{B}(n, p)$.

Remarque On rappelle qu'une telle variable aléatoire est associée à une situation très classique dans laquelle X représente le nombre de succès lorsqu'on répète de façon indépendante une épreuve de Bernoulli : on utilisera donc ce modèle lorsque X rendra compte du **nombre de succès** obtenus après n répétitions indépendantes d'une épreuve de la forme succès/échec.

Définition Soit $p \in]0, 1[$. On dit qu'une variable aléatoire X suit une **loi géométrique** de paramètre p si :

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

Dans ce cas, on note $X \sim \mathcal{G}(p)$.

Remarque Si on considère encore une suite d'épreuves de Bernoulli indépendantes, alors en notant X le rang du premier succès et en posant $A_k =$"la k-ème épreuve est un succès", on a par indépendance :

$$P(X = k) = P(\overline{A_1} \cap \ldots \cap \overline{A_{k-1}} \cap A_k) = \prod_{i=1}^{k-1} P(\overline{A_i}) P(A_k) = q^{k-1} p$$

Autrement dit, on fera appel au modèle géométrique quand la variable aléatoire X rendra compte du temps d'attente associé à un premier succès lorsqu'on répète de façon indépendante une épreuve de la forme succès/échec.

Théorème 13 (caractérisation d'une loi géométrique).

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

1. On considère X une variable aléatoire discrète sur Ω telle que $X \sim \mathcal{G}(p), p \in]0,1[$.
   Alors, pour tout $k \in \mathbb{N}$, $P(X > k) = q^k$ et ainsi,
   $$\forall\, (k, \ell) \in \mathbb{N}^2,\ P_{(X>k)}(X > k + \ell) = P(X > \ell) \quad (*)$$
   On dit que X est une **loi sans mémoire**.
2. Réciproquement, considérons Y telle que $Y(\Omega) = \mathbb{N}^*$. Si de plus Y vérifie la condition $(*)$, alors Y suit une loi géométrique de paramètre $p = P(Y = 1)$.

▶ *Pour le premier point, il suffit de décrire l'évènement $(X > k)$ comme réunion d'évènements élémentaires, la formule déécoule alors de la définition de la probabilité conditionnelle. Pour le second point, on cherche d'abord à obtenir une relation de récurrence qui nous permettra d'obtenir $P(Y = k)$.*

Remarque En fait, si on voit X comme une durée de vie, on dit que X est sans mémoire car la probabilité de sa durée de vie ne dépend pas du temps déjà passé... cela nous donne alors une autre façon d'utiliser la loi géométrique, pour des situations dans lesquelles la réalisation d'un succès ne dépend pas des expériences précédentes.

Définition Soit $\lambda > 0$. On dit qu'une variable aléatoire X suit une **loi de Poisson** de paramètre λ si :
$$\begin{cases} X(\Omega) = \mathbb{N} \\ \forall\, k \in \mathbb{N},\ P(X = k) = e^{-\lambda}\dfrac{\lambda^k}{k!} \end{cases}$$
Dans ce cas, on note $X \sim \mathcal{P}(\lambda)$.

Remarque Il s'agit encore une fois d'une variable aléatoire qui rend compte d'un nombre de succès... mais contrairement aux autres modèles, il n'est pas simple d'interpréter les situations associées à une loi de Poisson car c'est une loi limite.

Théorème 14 (approximation d'une loi de Poisson par une loi binomiale).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, et considérons (X_n) une suite de variables aléatoires telles que X_n suit une loi binomiale de paramètres n, p_n vérifiant :
$$np_n \underset{n\to+\infty}{\sim} \lambda > 0$$
Alors, pour tout $k \in \mathbb{N}$, $\lim_{n\to+\infty} P(X_n = k) = e^{-\lambda}\dfrac{\lambda^k}{k!}$, et ainsi la suite (X_n) converge en loi vers une loi de Poisson.

▶ *On revient à la loi binomiale et on travaille sur le coefficient binomial afin de déterminer la limite à l'aide des fonctions usuelles.*

Remarque Par hypothèse, on a :
$$np_n \sim \lambda \Leftrightarrow p_n \sim \frac{\lambda}{n} \longrightarrow 0$$
Ainsi, si X_n désigne le nombre de succès d'un évènement rare lors d'un grand nombre de répétitions d'une épreuve de Bernoulli, alors X_n peut être approchée par une loi de Poisson de paramètre λ... C'est pour cela que la loi de Poisson est aussi appelée **loi des évènements rares** et elle rendra compte du nombre de succès dans des épreuves de la forme succès/échec et ceci quand celui-ci a une faible probabilité de se réaliser.

2.2 Famille de variables aléatoires indépendantes

Définition Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X, Y deux variables aléatoires discrètes telles que :
$$X(\Omega) = \{x_i,\ i \in I\} \text{ et } Y(\Omega) = \{y_j,\ j \in J\}$$
On appelle **loi du couple** (X, Y) la probabilité $P_{X,Y}$ définie sur $\mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega))$ par :
$$P_{X,Y} : (x_i, y_j) \longmapsto P\left((X = x_i) \cap (Y = y_j)\right)$$
Dans ce cas, les lois P_X et P_Y associées aux variables aléatoires X et Y désignent les **lois marginales**.

Propriété 15 (relation avec les lois marginales).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X, Y deux variables aléatoires discrètes telles que :

$$X(\Omega) = \{x_i,\ i \in I\} \text{ et } Y(\Omega) = \{y_j,\ j \in J\}$$

Alors, on a :

1. pour tout $i \in I$, $P(X = x_i) = \sum_{j \in J} P_{X,Y}(x_i, y_j)$
2. pour tout $j \in J$, $P(Y = y_j) = \sum_{i \in I} P_{X,Y}(x_i, y_j)$

▶ *C'est immédiat et découle de la formule des probabilités totales. En effet, si $(X = x_i)$ désigne un évènement de $\mathcal{P}(X(\Omega))$, alors les évènements $(Y = y_j)_{j \in J}$ représentent un système complet d'évènements.*

Remarques

1. On pourra retenir que la connaissance de la loi d'un couple de variables nous permet d'en déduire les lois marginales. Par contre, les lois marginales ne nous permettront pas, en général, d'obtenir la loi du couple.
2. Très souvent, l'une des variables sera conditionnée par l'autre et ainsi, on peut par exemple écrire pour tout $i \in I$,
$$P(X = x_i) = \sum_{j \in J} P_{X,Y}(x_i, y_j) = \sum_{j \in J} P\left((X = x_i) \cap (Y = y_j)\right) = \sum_{j \in J} P(Y = y_j) P_{(Y = y_j)}(X = x_i)$$
Et cette dernière égalité fait apparaître une **loi conditionnelle** : c'est la loi de X sachant que $(Y = y_j)$.

Exemple 9 On considère une suite d'épreuves de Bernoulli indépendantes et de paramètre $p \in]0, 1[$. On note X le rang du premier succès et Y le rang du second succès.

1. Déterminer la loi du couple (X, Y).
2. En déduire les lois de X et Y.

Définition Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons $X_1, \dots, X_n$ des variables aléatoires discrètes telles que $X_k(\Omega) = \{x_{i_k},\ i_k \in I_k\}$. On dit encore que :

- les variables aléatoires sont **deux à deux indépendantes** si pour tout $(k, \ell) \in I_k \times I_\ell$,
$$P\left((X_k = x_{i_k}) \cap (X_l = x_{i_l})\right) = P(X_k = x_{i_k}) P(X_l = x_{i_l})$$
- les variables aléatoires sont **mutuellement indépendantes** si pour tout $J \subset [\![1, n]\!]$:
$$P\left(\cap_{k \in J} (X_k = x_{i_k})\right) = \prod_{k \in J} P(X_k = x_{i_k})$$

Remarque Par défaut, et sans aucune indication contraire, on pourra considérer que des variables aléatoires données indépendantes dans un énoncé sont **mutuellement indépendantes**. En particulier, cela entraîne évidemment qu'elles sont deux à deux indépendants mais la réciproque est fausse !

Corollaire 16 (loi d'un couple de variables indépendantes).

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X, Y deux variables aléatoires discrètes telles que :

$$X(\Omega) = \{x_i,\ i \in I\} \text{ et } Y(\Omega) = \{y_j,\ j \in J\}$$

Si de plus, X et Y sont supposées indépendantes, alors on a pour tout $(i, j) \in I \times J$, $P_{X,Y}(x_i, y_j) = P(X = x_i)P(Y = y_j)$. Autrement dit, la loi du couple est ici définie par les lois marginales.

Théorème 17 (caractérisation de deux variables aléatoires indépendantes).

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, X, Y deux variables aléatoires discrètes. Alors :

$$X \text{ et } Y \text{ sont indépendantes} \Leftrightarrow \forall\ (A, B) \in \mathcal{P}(X(\Omega)) \times \mathcal{P}(Y(\Omega)),\ P\left((X \in A) \cap (Y \in B)\right) = P(X \in A) P(Y \in B)$$

▶ *La réciproque est immédiate. Pour le sens direct, il suffit de définir les évènements $(X \in A) = \{x_i,\ i \in I_A\}$ et $(Y \in B) = \{y_j,\ j \in J_B\}$ avant de calculer la probabilité de l'intersection.*

On pourra généraliser cette dernière propriété et ainsi, la probabilité de l'intersection d'évènements associés à un nombre fini de variables aléatoires mutuellement indépendantes sera toujours donnée par le produit des probabilités.

Exemple 10 Soient $N \in \mathbb{N}^*$ et $p \in]0,1[$. On pose $q = 1 - p$.
On considère N variables aléatoires indépendantes $X_1, X_2, \dots, X_N$ définies sur un même espace probabilisé $(\Omega, \mathcal{A}, P)$ mutuellement indépendantes et de même loi géométrique de paramètre p.

1. Soit $i \in [\![1, N]\!]$ et fixons $n \in \mathbb{N}^*$. Déterminer $P(X_i > n)$.
2. On considère la variable aléatoire $Y = \min(X_1, X_2, \dots, X_N)$.
   (a) Soit $n \in \mathbb{N}^*$. Calculer $P(Y > n)$. En déduire $P(Y \leq n)$, puis $P(Y = n)$.
   (b) Préciser alors la loi de Y.

Propriété 18 (transfert d'indépendance)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X, Y deux variables aléatoires discrètes indépendantes. On considère de plus f et g deux applications telles que :

$$f : X(\Omega) \longrightarrow \mathbb{R} \text{ et } g : Y(\Omega) \longrightarrow \mathbb{R}$$

alors $f(X)$ et $g(Y)$ sont encore indépendantes.

► *On se ramène simplement au théorème précédent avec* $(A, B) \subset \mathcal{P}(\mathbb{R})^2$.

Remarque Ce résultat est très pratique et encore une fois, il peut se généraliser. Autrement dit, quand on travaille avec des variables aléatoires mutuellement indépendantes, on pourra retenir qu'on conserve l'indépendance par composition sur ces variables :

Corollaire 19 (lemme des coalitions)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons $X_1, \dots, X_n$ des variables aléatoires discrètes qu'on suppose mutuellement indépendantes, $p \in [\![1, n-1]\!]$. Si de plus f, g désignent deux applications telles que :

$$f : X_1(\Omega) \times \ldots \times X_p(\Omega) \longrightarrow \mathbb{R} \text{ et } g : X_{p+1}(\Omega) \times \ldots X_n(\Omega) \longrightarrow \mathbb{R}$$

alors $f(X_1, \dots, X_p)$ et $g(X_{p+1}, \dots, X_n)$ sont encore indépendantes.

2.3 Espérance d'une variable alétoire discrète réelle ou complexe

Définition Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X(\Omega) = \{x_i,\ i \in I\}$.
Sous réserve d'existence, on appelle **espérance mathématique** le nombre réel défini par :

$$E(X) = \sum_{i \in I} p_i x_i, \text{ avec } p_i = P(X = x_i)$$

Remarques

1. On peut alors observer que :
$$E(X) = \sum_{i \in I} p_i x_i = \frac{\sum_{i \in I} p_i x_i}{\sum_{i \in I} p_i}$$
Ainsi, l'espérance mathématique désigne la moyenne des valeurs prises par X pondérée par les probabilités des évènements élémentaires.
2. Dans le cas particulier où X prend un nombre fini de valeurs, la somme donnée est finie et on pourra toujours calculer l'espérance associée. Dans les autres cas, il s'agira d'abord de vérifier que la série $\sum p_i x_i$ est convergente, ou plus souvent absolument convergente : on pourra même noter $X \in L^1$ pour signifier que cette série est absolument convergente.

Exemple 11 Les questions suivantes ne sont pas forcément liées.

1. Déterminer l'espérance de X dans les cas suivants :
   (a) $X \sim \mathcal{U}([\![1, n]\!])$
   (b) $X \sim \mathcal{B}(p)$ avec $p \in]0,1[$
   (c) $X \sim \mathcal{B}(n, p)$ avec $p \in]0,1[$
   (d) $X \sim \mathcal{G}(p)$ avec $p \in]0,1[$
   (e) $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$

2. M. Toutlemonde habite dans un immeuble dont la porte d'entrée est sécurisée par un code à 4 chiffres dont chacun est compris entre 0 et 9. Malheureusement, il se trouve devant cette porte et il en a oublié le code.
   (a) En essayant un code au hasard, quelle est la probabilité de tomber sur le bon code ?
   (b) M. Toutlemonde décide de trouver le bon code en procédant de la manière suivante : il essaye un code au hasard choisi par les codes non encore testés. On note X la variable aléatoire égale au nombre de codes testés jusqu'à obtenir le bon code. Déterminer la loi de X et donner son espérance.
   (c) A la place de la stratégie précédente, M. Toutlemonde essaye des codes au hasard, sans se soucier du fait qu'il les ait déjà essayés ou non. On note encore X la variable aléatoire égale au nombre de codes testés jusqu'à obtenir le bon code. Déterminer la loi de X et donner son espérance.

Théorème 20 (de transfert)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, X une variable aléatoire discrète et on considère $f : X(\Omega) \longrightarrow \mathbb{K}$ telle que $\sum_{i\in I} |p_i f(x_i)|$ converge. Alors, on a :

$$E(f(X)) = \sum_{i\in I} p_i f(x_i), \text{ avec } p_i = P(X = x_i)$$

► *Pour simplifier les écritures, on pose $Y = f(X)$ et en notant $I_k = \{i \in I,\ f(x_i) = y_k\}$, on peut réécrire la somme de droite par paquets à l'aide du théorème de sommation par paquets.*

Propriété 21 (de l'espérance)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X, Y deux variables aléatoires discrètes telles que :

$$X(\Omega) = \{x_i,\ i \in I\} \text{ et } Y(\Omega) = \{y_j,\ j \in J\}$$

On suppose de plus que $X, Y \in L^1$. Alors, X et Y admettent une espérance finie et on vérifie que :

1. l'espérance est linéaire :
$$\forall\ (a, b) \in \mathbb{R}^2,\ E(aX + bY) = aE(X) + bE(Y)$$
   Et en particulier, avec $Y = 1$, on a $E(aX + b) = aE(X) + b$.
2. l'espérance est positive, c'est à dire que pour des variables réelles : $X \geq 0 \Rightarrow E(X) \geq 0$.
   Et en particulier, pour une variable à valeurs positives, $E(X) = 0 \Leftrightarrow X = 0$ presque sûrement.
3. l'espérance est croissante, c'est à dire que pour des variables réelles :
$$X \leq Y \Rightarrow E(X) \leq E(Y)$$
   Et en particulier, on a l'inégalité : $|E(X)| \leq E(|X|)$.

► *Seul le premier point est délicat, on applique le théorème de transfert après en avoir vérifié les hypothèses. Pour cela, on pourra montrer la sommabilité des familles $(p_{ij}x_i)$ et $(p_{ij}y_j)$ à l'aide du théorème de Fubini.*

Théorème 22 (espérance du produit de variables mutuellement indépendantes)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X, Y deux variables aléatoires discrètes telles que :

$$X(\Omega) = \{x_i,\ i \in I\} \text{ et } Y(\Omega) = \{y_j,\ j \in J\}$$

On suppose de plus que $X, Y \in L^1$ et qu'elles sont indépendantes. Alors, XY admet une espérance finie et on a :

$$E(XY) = E(X)E(Y)$$

► *On revient encore au théorème de transfert à condition d'en vérifier les hypothèses. Pour cela, on montre que la famille $(p_{ij}x_iy_j)$ est sommable à l'aide du théorème de Fubini.*

Remarque Par récurrence, et en faisant appel au lemme des coalitions, il est alors très facile d'étendre ce dernier résultat de sorte que pour toute famille de variables mutuellement indépendantes $X_1, \ldots, X_n \in L^1$:

$$E(X_1 \ldots X_n) = \prod_{k=1}^{n} E(X_k)$$

Exemple 12 On considère une variable aléatoire discrète X à valeurs dans $\mathbb{N}$. Montrer que :

$$X \text{ admet une espérance finie} \Leftrightarrow \sum_{n\geq 0} P(X>n) \text{ converge}$$

et que dans ce cas, $E(X)=\sum_{n=0}^{+\infty} P(X>n)=\sum_{n=1}^{+\infty} P(X\geq n)$.

2.4 Variance d'une variable aléatoire réelle, écart type et covariance

Définition Soient $(\Omega,\mathcal{A},P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X(\Omega)=\{x_i,\ i\in I\}$.
Sous réserve d'existence, on appelle alors :

- **variance** le nombre réel défini par :

$$V(X)=E((X-E(X))^2)=\textstyle\sum_{i\in I} p_i(x_i-E(X))^2,\ \text{avec } p_i=P(X=x_i)$$

- **écart-type** le nombre réel positif σ tel que : $\sigma(X)=\sqrt{V(X)}$.

Remarques

1. On peut alors observer que :

$$V(X)=\sum_{i\in I} p_i(x_i-E(X))^2=\frac{\sum_{i\in I} p_i(x_i-E(X))^2}{\sum_{i\in I} p_i}$$

Ainsi, la variance n'est rien d'autre que la moyenne des écarts au carré par rapport à l'espérance et pondérée par les probabilités des évènements élémentaires. On pourra donc retenir que la variance et l'écart-type sont des indicateurs de la dispersion des valeurs prises par X autour de sa moyenne.

2. Dans le cas particulier où X prend un nombre fini de valeurs, la somme donnée est finie et on pourra toujours calculer la variance associée. Dans les autres cas, il s'agira d'abord de vérifier que la série $\sum p_i x_i^2$ est convergente, ou plus souvent absolument convergente : on pourra même noter $X\in L^2$ pour signifier que cette série est absolument convergente.

Propriété 23 (condition suffisante d'existence).

Soient $(\Omega,\mathcal{A},P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X\in L^2$. Alors, X admet une espérance finie et une variance finie.

► *Posons $m=E(X)$, on procède alors en deux temps et on montre à chaque fois que les séries associées sont bien absolument convergentes par simple comparaison sur les termes généraux.*

Corollaire 24 (formule d'Huygens pour le calcul de la variance).

Soient $(\Omega,\mathcal{A},P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X\in L^2$. Alors, la variance est aussi égale à :

$$V(X)=E(X^2)-(E(X))^2=E(X(X-1))+E(X)-(E(X))^2$$

Exemple 13 Déterminer la variance de X dans les cas suivants :

1. $X\sim\mathcal{U}([\![1,n]\!])$
2. $X\sim\mathcal{B}(p)$ avec $p\in]0,1[$
3. $X\sim\mathcal{B}(n,p)$ avec $p\in]0,1[$
4. $X\sim\mathcal{G}(p)$ avec $p\in]0,1[$
5. $X\sim\mathcal{P}(\lambda)$ avec $\lambda>0$

Propriété 25 (de la variance).

Soient $(\Omega,\mathcal{A},P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X\in L^2$. Alors, on a :

(i) $V(X)\geq 0$. En particulier, $V(X)=0\Leftrightarrow$ X est constante presque sûrement.

(ii) $\forall\,(a,b)\in\mathbb{R}^2,\ V(aX+b)=a^2V(X)+0$

► *Il suffit de revenir à la définition de la variance et d'utiliser les propriétés de l'espérance mathématique.* $\square$

Définition Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X une variable aléatoire discrète. On dit que :

- la variable aléatoire X est **centrée** si $E(X) = 0$.
- la variable aléatoire X est **réduite** si $V(X) = 1$.

Théorème 26 (construction d'une variable centrée et réduite)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé discret et X une variable aléatoire discrète telle que $X \in L^2$. On pose $m = E(X)$ et $\sigma = \sigma(X)$ qu'on suppose non nul. Alors, la variable aléatoire définie par $\dfrac{X-m}{\sigma}$ est centrée et réduite.

► *Le résultat découle directement des propriétés de l'espérance et de la variance démontrées plus haut.*

Théorème 27 (existence et définition de la covariance)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X, Y deux variables aléatoires discrètes telles que $X, Y \in L^2$. Alors, XY admet une espérance finie et on peut définir le nombre suivant appelé **covariance** de X et Y :

$$Cov(X,Y) = E\left((X - E(X))(Y - E(Y))\right)$$

► *Avec ces hypothèeses, on rappelle que $X, Y \in L^1$ et on peut utiliser l'inégalité de Young pour majorer le terme général de la série $\sum |p_{ij} x_i y_j|$.*

Propriété 28 (formule d'Huygens pour le calcul de la covariance)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X, Y deux variables aléatoires discrètes telle que $X, Y \in L^2$. Alors, on peut aussi écrire :

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

De plus, on vérifie que :

1. la covariance est une forme bilinéaire symétrique et positive ;
2. et si de plus X et Y sont indépendantes, alors on a $Cov(X,Y) = 0$ et on dit que les variables sont **non corrélées**.

► *Dans le premier point, il suffit de développer la formule de la covariance et d'utiliser les propriétés de l'espérance. Les résultats suivants découlent immédiatement des propriétés de l'espérance.*

Remarque On fera attention, il s'agit bien d'une condition nécessaire et il existe des couples de variables de covariance nulle, sans pour autant qu'elles soient indépendantes.
Par exemple, il suffit de considérer X, Y deux variables aléatoires telles que $X \sim \mathcal{U}(\llbracket -1, 1 \rrbracket)$ et $Y = 1$ si $X = 0$, $Y = 0$ sinon.

Propriété 29 (variance d'une somme de variables aléatoires)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X, Y deux variables aléatoires discrètes telles que $X, Y \in L^2$. Alors :

(i) la variance n'est pas linéaire : $\forall\, a, b \in \mathbb{R},\ V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab Cov(X,Y)$

(ii) En particulier, si $X_1, \ldots, X_n$ désignent des variables aléatoires discrètes telles que $X_i \in L^2$, alors :

$$V(X_1 + \ldots + X_n) = V(X_1) + \ldots + V(X_n) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

Si de plus, elles sont mutuellement indépendantes :

$$V(X_1 + \ldots + X_n) = V(X_1) + \ldots V(X_n)$$

► *On revient à la formule d'Huygens pour le calcul de la variance, puis on utilise les propriétés de l'espérance. Pour le second point, on peut procéder par récurrence sur $n \geq 2$.*

Exemple 14 Soient $n \in \mathbb{N}^*$ et $p \in [0,1]$. On considère $X_1, \ldots, X_n$ des variables aléatoires mutuellement indépendantes et identiquement distribuées telles que pour tout $i \in [\![1,n]\!]$, $X_i \sim \mathcal{B}(p)$, et on définit :

$$S_n = X_1 + \ldots + X_n$$

1. Montrer que $S_n \sim \mathcal{B}(n,p)$.
2. Retrouver alors l'espérance et la variance de S_n.

3 Applications

3.1 Fonction génératrice d'une variable aléatoire

Définition Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X une variable aléatoire discrète qu'on suppose à valeurs dans $\mathbb{N}$. Le terme général étant borné pour $t = 1$, la série entière $\sum P(X = n)t^n$ a un rayon de convergence $R \geq 1$, et on appelle alors **fonction génératrice** de X la fonction G_X définie sur $]-R, R[$ par :

$$G_X(t) = E(t^X) = \sum_{k=0}^{+\infty} P(X = k)t^k$$

Propriété 30 (convergence de la série génératrice associée et conséquences)**.**

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et considérons X une variable aléatoire discrète qu'on suppose à valeurs dans $\mathbb{N}$. Alors, la série entière $\sum P(X = n)t^n$ converge normalement sur $[-1, 1]$, et ainsi :

1. G_X est au moins définie et continue sur $[-1, 1]$ et donc, pour tout $t \in [-1, 1]$, $G_X(t) = \sum_{k=0}^{+\infty} P(X = k)t^k$.
2. G_X est de classe C^∞ sur $]-1, 1[$ et pour tout $t \in]-1, 1[$, et pour tout $n \in \mathbb{N}^*$,

$$G_X^{(n)}(t) = \sum_{k=n}^{+\infty} P(X = k)k(k-1)\ldots(k-n+1)t^{k-n}$$

► *On travaille sur la norme infinie du terme général. Le reste découle directement des théorèmes sur les séries entières.*

Remarque En fait, en tant que somme d'une série entière, on peut même étendre la régularité de G_X sur l'intervalle $]-R, R[$. C'est notamment le cas des variables aléatoires finies pour lesquelles G_X est définie sur $\mathbb{R}$ tout entier : on a là des fonctions polynomiales.

Corollaire 31 (la fonction génératrice détermine la loi de X)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, X une variable aléatoire discrète qu'on suppose à valeurs dans $\mathbb{N}$, et notons G_X sa fonction génératrice.
Alors, la fonction génératrice détermine la loi de X, c'est à dire que pour tout $n \in \mathbb{N}$,

$$P(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

On en déduit que deux variables aléatoires à valeurs dans $\mathbb{N}$ ont la même loi si et seulement si elles ont la même fonction génératrice.

► *Cela provient directement de l'unicité des coefficients dans le développement en série entière. La loi étant entièrement déterminée par G_X, on en déduit le second point.*

Il faudra donc être capable d'identifier ou de retrouver rapidement les fonctions génératrices associées à nos lois usuelles.

Exemple 15 Déterminer la fonction génératrice G_X dans les cas suivants :

1. $X \sim \mathcal{U}([\![1,n]\!])$
2. $X \sim \mathcal{B}(p)$ avec $p \in]0, 1[$
3. $X \sim \mathcal{B}(n,p)$ avec $p \in]0, 1[$
4. $X \sim \mathcal{G}(p)$ avec $p \in]0, 1[$
5. $X \sim \mathcal{P}(\lambda)$ avec $\lambda > 0$

Propriété 32 (fonction génératrice de la somme de variables mutuellement indépendantes)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, $X_1, \ldots, X_n$ des variables aléatoires discrètes à valeurs dans $\mathbb{N}$, et notons G_{X_k} leur fonction génératrice. On suppose de plus que ces variables aléatoires sont mutuellement indépendantes et ainsi,

$$\forall\, t \in [-1, 1],\ G_{X_1+\ldots+X_n}(t) = \prod_{k=1}^{n} G_{X_k}(t)$$

▶ *On revient simplement à la définition de la fonction génératrice et on rappellera les propriétés de l'espérance.*

Théorème 33 (caractérisation de la dérivabilité en 1 pour la somme d'une série entière de rayon $R \geq 1$)**.**

Soit (a_n) une suite réelle à valeurs positives telle que $\sum a_n$ converge. On considére $\sum a_n t^n$ la série entière associée de rayon de convergence $R \geq 1$, et on note f sa somme sur $]-R, R[$.
Alors, f est dérivable en 1 si et seulement si la série $\sum n a_n$ converge, et dans ce cas, si celle-ci est dérivable en 1 :

$$f'(1) = \sum_{n=1}^{+\infty} n a_n$$

▶ *On procède par disjonction des cas : si $R > 1$, alors les résultats sur les séries entières nous donnent directement les deux assertions et l'égalité. Si par contre $R = 1$, on montre d'abord que $t \longmapsto \frac{f(t)-f(1)}{t-1}$ est croissante sur $[0, 1[$ avant de justifier l'équivalence des deux assertions.*

Corollaire 34 (calcul de l'espérance)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, X une variable aléatoire discrète qu'on suppose à valeurs dans $\mathbb{N}$, et notons G_X sa fonction génératrice.
Alors, $X \in L^1$ si et seulement si G_X est dérivable en 1 et dans ce cas, on en déduit que :

$$E(X) = G_X'(1)$$

▶ *Cela est immédiat : il suffit d'exploiter le théorème précédent avec la série génératrice associée à X.*

Corollaire 35 (calcul de la variance)**.**

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, X une variable aléatoire discrète qu'on suppose à valeurs dans $\mathbb{N}$, et notons G_X sa fonction génératrice.
Alors, $X \in L^2$ si et seulement si G_X est deux fois dérivable en 1 et dans ce cas, on en déduit que :

$$E(X(X-1)) = G_X''(1) \text{ et ainsi, } V(X) = E(X(X-1)) + E(X) - (E(X))^2 = G_X''(1) + G_X'(1) - (G_X'(1))^2$$

▶ *On procède par double implication, et on veillera à exploiter le même théorème en considérant la série génératrice dérivée.*

Exemple 16 Les questions suivantes sont indépendantes.

1. On considère X, Y deux variables aléatoires discrètes telles que :

$$X \sim \mathcal{G}(p) \text{ avec } p \in]0, 1[, \text{ et } Y \sim \mathcal{P}(\lambda) \text{ avec } \lambda > 0$$

   (a) Montrer que pour tout $t \in [-1, 1]$, $G_X(t) = \dfrac{pt}{1-qt}$ et $G_Y(t) = e^{\lambda(t-1)}$.

   (b) Retrouver alors l'espérance et la variance pour chacune de ces variables.

2. On suppose que $X_1, \ldots, X_n$ désignent des variables aléatoires mutuellement indépendantes et telles que pour tout $k \in [\![1, n]\!]$,

$$X_k \sim \mathcal{P}(\lambda_k) \text{ avec } \lambda_k > 0$$

   Etablir que $X_1 + \ldots + X_n$ suit encore une loi de Poisson dont on précisera le paramètre. *On pourra proposer deux méthodes.*

3.2 Inégalités probabilistes et loi faible des grands nombres

Propriété 36 (inégalité de Markov).

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X \in L^1$. Alors, on a :

$$\forall\ \epsilon > 0,\ P(|X| \geq \epsilon) \leq \frac{E(|X|)}{\epsilon}$$

► *On note $J = \{i \in I,\ |x_i| \geq \epsilon\}$ et il suffit alors de minorer l'espérance de $|X|$.*

Propriété 37 (inégalité de Bienaymé-Tchebychev).

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X une variable aléatoire discrète telle que $X \in L^2$ converge. Alors, on a :

$$\forall\ \epsilon > 0,\ P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}$$

► *On applique l'inégalité de Markov à la variable $(X - E(X))^2$ avec ϵ^2.*

Exemple 17 Soit f une fonction réelle continue sur $[0, 1]$. Pour $n \in \mathbb{N}^*$, on note P_n le polynôme défini par :

$$P_n(x) = \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k}$$

On fixe $x \in]0, 1[$. On considère une suite de variables aléatoires (X_n) mutuellement indépendantes et de même loi de Bernoulli de paramètre x, et on pose $S_n = \sum_{k=1}^{n} X_k$.

1. Donner une expression de $E(f(\frac{S_n}{n}))$.
2. Pour tout $\alpha > 0$, on définit $\delta(\alpha) = sup\{|f(x) - f(y)|,\ x, y \in [0, 1], |x - y| \leq \alpha\}$. Démontrer que :

$$\sup_{x \in [0,1]} |P_n(x) - f(x)| \leq \delta(\alpha) + \frac{2||f||_\infty}{n\alpha^2}$$

3. En déduire que la suite (P_n) converge uniformément vers f sur $[0, 1]$.

*On retrouve là une preuve probabiliste du **théorème de Stone-Weierstrass**.*

Théorème 38 (loi faible des grands nombres).

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé discret et $X_1, \ldots, X_n$ des variables aléatoires discrètes mutuellement indépendantes et identiquement distribuées. On suppose de plus qu'elles possèdent une espérance finie et une variance finie notées $m = E(X_i)$ et $v = V(X_i)$.
De plus, si on note $S_n = X_1 + \ldots + X_n$, alors pour tout $\epsilon > 0$,

$$P(|\frac{S_n}{n} - m| \geq \epsilon) \underset{n \to +\infty}{\longrightarrow} 0$$

On dit aussi que $\frac{S_n}{n}$ **converge en probabilité** vers m.

► *C'est immédiat : il suffit d'appliquer l'inégalité de Bienaymé-Tchebychev à $\frac{S_n}{n}$ avant de passer à la limite...*

Remarque Ce résultat a en fait beaucoup de sens, car il permet de justifier la première **approche fréquentiste** qu'on a pu vous donner du calcul de la probabiblité d'un évènement.

En effet, si on étudie l'apparition d'un caractère dans une population de n individus, on peut modéliser cette situation par un échantillon $(X_1, \ldots, X_n)$, où pour tout $i \in [\![1, n]\!]$, $X_i \sim \mathcal{B}(p)$.
Dans ce cas, $f_n = \dfrac{S_n}{n}$ représente la fréquence d'apparition du caractère donné et on montre par ce théorème que :

$$P(|f_n - p| \geq \epsilon) \underset{n \to +\infty}{\longrightarrow} 0$$

Autrement dit, quand on augmente la taille de l'échantillon, celle-ci ne s'écarte pas trop de la probabilité p associée. Et ainsi, la fréquence statistique nous donne une bonne approximation de la probabilité qu'un évènement se réalise.