

Probabilité discrète

Lors du cours sur les probabilités finies, nous avons posé le cadre probabiliste afin de modéliser une expérience aléatoire dans les issues sont finies. Ce cours étend le cadre pour aussi modéliser une expérience aléatoire dont les issues sont **dénombrables finies ou infinies**, c'est à dire que l'on peut effectuer une énumération finie ou infinie des issues de l'expérience aléatoire. Nous énoncerons uniquement les extensions et les nouvelles notions.

I Extension de la modélisation probabilistes pour un univers dénombrable

A Axiomes des probabilités. Premières propriétés

Exemple Première fois pile

Pour le lancer d'une pièce jusqu'à la première obtention de pile. L'univers est dénombrable \mathbb{N} où une issue $\omega \in \mathbb{N}$ représente le temps du premiers succès.

Définition Réunion et intersection dénombrables

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements de Ω . On définit les ensembles $\bigcup_{n \in \mathbb{N}} A_n$ et $\bigcap_{n \in \mathbb{N}} A_n$ par :

$$\omega \in \bigcup_{n \in \mathbb{N}} A_n \Leftrightarrow \exists n \in \mathbb{N} : \omega \in A_n$$

et

$$\omega \in \bigcap_{n \in \mathbb{N}} A_n \Leftrightarrow \forall n \in \mathbb{N} : \omega \in A_n$$

Exemple Première fois pile

L'événement A_p = "l'expérience se termine après le p -ième lancer" est le singleton $\{p\}$.

L'événement "l'expérience se termine après le p -ième lancer" s'écrit $\bigcup_{n=p}^{+\infty} A_p$.

L'événement "l'expérience se termine après un temps infini" s'écrit $\bigcap_{n=1}^{+\infty} \bar{A}_p$.

Définition Axiomes des tribus

La probabilité P est une fonction qui à un événement associe un nombre compris entre 0 et 1 quantifiant la chance de réalisation de cet événement. Pour des raisons sortant du cadre de ce cours, il n'est pas toujours possible d'attribuer ainsi de manière cohérente une probabilité à chaque partie de Ω .

En d'autres termes, P ne peut pas être considérée comme une application de $\mathcal{P}(\Omega)$ dans $[0, 1]$ mais comme une fonction ayant un domaine de définition \mathcal{A} généralement plus petit que $\mathcal{P}(\Omega)$.

Voici les propriétés qu'il est raisonnable d'exiger de \mathcal{A} :

- \mathcal{A} contient \emptyset (et tous les singletons $\{\omega\}$),
- \mathcal{A} est stable par passage au complémentaire : si A est un événement de \mathcal{A} , A^c l'est aussi,
- \mathcal{A} est stable par les opérations de réunion et d'intersection sur les suites d'événements
- Si A_1, A_2, \dots est une suite finie ou infinie d'événements de \mathcal{A} , sa réunion et son intersection sont encore des événements de \mathcal{A} .

Nous appellerons **tribu** toute famille \mathcal{A} de parties de vérifiant les conditions ci-dessus.

(Ω, \mathcal{A}) est appelé **espace probabilisable**. L'objectif de ce cours n'est pas de savoir construire mathématiquement la tribu \mathcal{A} . On supposera que l'on sait construire à partir de l'univers la tribu ayant les "bonnes propriétés".

Définition Axiomes des probabilités

Soit (Ω, \mathcal{A}) un espace probabilisable. Une **probabilité** sur (Ω, \mathcal{A}) est une application de \mathcal{A} dans \mathbb{R}^+ vérifiant :

- $P(\Omega) = 1$.
- Pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ d'événements de \mathcal{A} deux à deux disjoints (incompatibles) :

$$P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n=1}^{+\infty} P(A_n)$$

Le triplet (Ω, \mathcal{A}, P) s'appelle un **espace probabilisé**.

Proposition

Une probabilité P sur un univers dénombrable est complètement déterminée par les $P(\{\omega\})$ pour tout $\omega \in \Omega$. $P(\{\omega\})$, appelé poids de probabilité.

En effet, pour $A \subset \Omega$, on a :

$$\begin{aligned} P(A) &= P\left(\bigcup_{\omega \in A} \{\omega\}\right), \\ &= \sum_{\omega \in A} P(\{\omega\}). \end{aligned}$$

Exemple Première fois pile

Soit p la probabilité d'obtenir pile. On pose $p_k = P(\{k\}) = \overbrace{(1-p)^{k-1}}^{\text{proba d'obtenir d'abord k-1 piles}} p$. On doit vérifier que $\sum_{k=1}^{+\infty} p_k = 1$. On a

$$\sum_{k=1}^{+\infty} p_k = \sum_{k=1}^{+\infty} (1-p)^{k-1} p = p \sum_{k=1}^{+\infty} (1-p)^{k-1} \stackrel{\text{somme d'une série géométrique}}{=} p \cdot \frac{1}{1-(1-p)} = 1.$$

Proposition

1. $P(\emptyset) = 0$.
2. Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'événements de \mathcal{A} deux à deux disjoints (incompatibles) :

$$P\left(\bigcup_{1 \leq i \leq n} A_i\right) = \sum_{i=1}^n P(A_i).$$

3. Pour tout $A, B \in \mathcal{A}$ tel que $A \subset B$, on a :

$$P(A) \leq P(B)$$

$$P(B \setminus A) = P(B) - P(A)$$

4. Pour tout $A \in \mathcal{A}$, on a $P(A) \in [0, 1]$.

5. Pour tout $A, B \in \mathcal{A}$, on a

$$P(A \cup B) + P(A \cap B) = P(A) + P(B).$$

6. Pour tout $A, B \in \mathcal{A}$, on a

$$P(A \cap B) \leq P(A) + P(B).$$

7. Pour tout $A \in \mathcal{A}$, on a

$$P(A) = 1 - P(A^c).$$

8. Continuité monotone séquentielle :

- (a) Soit $(A_n)_{n \in \mathbb{N}^*}$ est une suite croissante d'événements de \mathcal{A} , c'est à dire $A_n \subset A_{n+1}$, on a

$$P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n)$$

- (b) Soit $(A_n)_{n \in \mathbb{N}^*}$ est une suite décroissante d'événements de \mathcal{A} , c'est à dire $A_{n+1} \subset A_n$, on a

$$P\left(\bigcap_{n \in \mathbb{N}^*} A_n\right) = \lim_{n \rightarrow +\infty} P(A_n)$$

9. **Sous additivité** : Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'événements de \mathcal{A} , on a :

$$P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n=1}^{+\infty} P(A_n).$$

Démonstration

1. Soit la suite $(A_n)_{n \in \mathbb{N}^*}$ définie par $A_1 = \Omega$ et $A_n = \emptyset$ si $n > 1$. Les événements de cette suite sont deux à deux disjoints et $\bigcup_{n \in \mathbb{N}^*} A_n = \Omega$.

$$P(\Omega) = P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right)$$

$$P(\Omega) = P(\Omega) + P(\emptyset) + P(\emptyset) + \dots \text{ car deux à deux disjoints}$$

$$0 = P(\emptyset) + P(\emptyset) + \dots$$

$$P(\emptyset) = 0.$$

2. Soit la suite $(B_i)_{i \in \mathbb{N}^*}$ définie par $B_i = A_i$ si $i \leq n$ et $B_i = \emptyset$ si $i > n$. On a :

$$\begin{aligned} P\left(\bigcup_{1 \leq i \leq n} A_i\right) &= P\left(\bigcup_{i \in \mathbb{N}^*} B_i\right) \\ P\left(\bigcup_{1 \leq i \leq n} A_i\right) &= \sum_{n=1}^{+\infty} P(B_n) \\ P\left(\bigcup_{1 \leq i \leq n} A_i\right) &= P(A_1) + \cdots + P(A_n) + P(\emptyset) + \cdots \\ P\left(\bigcup_{1 \leq i \leq n} A_i\right) &= \sum_{i=1}^n P(A_i) \end{aligned}$$

3. Comme $A \bigcup_{\text{disjoints}} (B \setminus A) = B$, on a :

$$P(A) + P(B \setminus A) = P(B).$$

4. Comme $A \subset \Omega$, on a $P(A) \leq P(\Omega) = 1$.

5. Comme $A \cup B = (A \setminus (A \cap B)) \bigcup_{\text{disjoints}} B$, on a $P(A \cup B) = P(A \setminus (A \cap B)) + P(B)$. Comme $(A \cap B) \subset A$, on a bien : $P(A \cup B) = P(A) - P(A \cap B) + P(B)$.

6. Comme $P(A \cup B) + P(A \cap B) = P(A) + P(B)$, on a $P(A \cup B) \leq P(A) + P(B)$.

7. Comme $A \bigcup_{\text{disjoints}} A^c = \Omega$, on a

$$P(A) + P(A^c) = 1.$$

8. Continuité monotone séquentielle :

(a) Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite croissante d'événements de \mathcal{A} . On pose la suite (B_n) défini $B_0 = A_0$ et $B_n = A_n \setminus A_{n-1}$ pour tout $n \geq 1$. Les éléments de cette suite sont deux à deux disjoints et $\bigcup_{k=0}^n A_k = \bigcup_{k=0}^n B_k$. On a :

$$\begin{aligned} P\left(\bigcup_{k=0}^n A_k\right) &= P\left(\bigcup_{k=0}^n B_k\right) \\ &= \sum_{k=0}^n P(B_k) \end{aligned}$$

La série numérique $\sum P(B_n)$ est à termes positifs et majoré par 1 ($\sum_{k=0}^n P(B_k) = P\left(\bigcup_{k=0}^n B_k\right) \leq P(\Omega) = 1$) donc convergente. Par passage à la limite on a

$$\begin{aligned} \lim_{n \rightarrow \infty} P(A_n) &= \lim_{n \rightarrow \infty} P\left(\bigcup_{k=0}^n B_k\right) \\ &= \sum_{k=0}^{+\infty} P(B_k) \\ &= P\left(\bigcup_{k=0}^{+\infty} B_k\right) \\ &= P\left(\bigcup_{k=0}^{+\infty} A_k\right) \end{aligned}$$

(b) On a :

$$\begin{aligned} P\left(\bigcap_{n \in \mathbb{N}^*} A_n\right) &= 1 - P\left(\left(\bigcap_{n \in \mathbb{N}^*} A_n\right)^c\right) \\ &= 1 - P\left(\bigcup_{n \in \mathbb{N}^*} A_n^c\right) \\ &= 1 - \lim_{n \rightarrow \infty} P(A_n^c) \\ &= \lim_{n \rightarrow \infty} (1 - P(A_n^c)) \\ &= \lim_{n \rightarrow \infty} P(A_n) \end{aligned}$$

9. **Sous additivité** : Pour toute suite finie $(A_i)_{1 \leq i \leq n}$ d'événements de \mathcal{A} , on a :

$$\forall n \in \mathbb{N} : P\left(\bigcup_{i=1}^n A_n\right) \leq \sum_{i=1}^n P(A_i) \underset{\text{termes positifs}}{\leq} \sum_{i=1}^{+\infty} P(A_i).$$

De plus $\bigcup_{n=1}^{+\infty} A_n = \bigcup_{n=1}^{+\infty} (\bigcup_{i=1}^n A_i)$ et la suite $(\bigcup_{i=1}^n A_i)$ est croissante d'où :

$$P\left(\bigcup_{n=1}^{+\infty} A_n\right) \underset{\text{Continuité monotone}}{=} \lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_n\right) \leq \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i).$$

Proposition Formule des probabilités totales

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

Soit $(A_n)_{n \in \mathbb{N}^*}$ un système complet d'événements de probabilités non nulles.

Pour tout événement B , on a

$$P(B) = \sum_{i=1}^{+\infty} P_{A_i}(B)P(A_i).$$

Proposition formule de Bayes usuelle

Soit $(\Omega, \mathcal{P}(\Omega), P)$ un espace probabilisé. Soit $(A_n)_{n \in \mathbb{N}^*}$ un système complet d'événements de probabilités non nulles. Pour tout événement B de probabilité non nulle, et pour tout $k \in \{1, n\}$, on a

$$P_B(A_k) = \frac{P_{A_k}(B)P(A_k)}{\sum_{i=1}^{+\infty} P_{A_i}(B)P(A_i)}.$$

II Variables aléatoires discrètes

A Définition

Exemple Martingale Roulette

La Roulette, inventée par Blaise Pascal, est un des jeux les plus populaires des casinos. Le pari simple est de parier la couleur de la case sur laquelle la bille lancée dans le cylindre va s'arrêter parmi 37 cases (18 rouges, 18 noirs et un vert).

La martingale est : le joueur doit, suite à chaque perte, tripler sa mise. S'il perd à nouveau, il mise la somme initiale triplée et ainsi de suite, jusqu'à ce qu'il gagne. Par exemple, le joueur joue 1 euro à la première partie et perd. Puis il joue donc 3 euros et perd encore. Après, il joue donc 9 et gagne. Finalement, son gain est $9 - 3 - 1 = 5$ euros.

A votre avis, ce jeu est-il favorable au joueur ou au casino ? A chaque partie quelle est le gain moyen du joueur ?

Étape 1 : comme pour l'expérience aléatoire du lancer d'une pièce jusqu'à la première obtention de pile, on pose X la variable aléatoire représentant le nombre de parties jouées jusqu'à gagner une première fois. Par exemple, $X = 3$ signifie que l'on a perdu les deux premières parties et gagné la troisième. La loi de la variable aléatoire est X est :

$$P(X = n) = (1-p)^{n-1}p \text{ avec } p = \frac{18}{37}.$$

Étape 2 : on calcule nos gains en fonction du nombre de parties jouées jusqu'à gagner une première fois n .

n	1	2	3	4	...	n	...
$G(n)$	1	3-1	9-(1+3)	27-(1-3-9)	...	$3^{n-1} - \sum_{k=0}^{n-2} 3^k = \frac{3^{n-1} + 1}{2}$	

Le gain moyen est donc :

$$E(G(X)) = \sum_{n=1}^{+\infty} G(n)P(X = n) = \sum_{n=1}^{+\infty} \frac{3^{n-1} + 1}{2}(1-p)^{n-1}p = \frac{p}{2} \sum_{n=1}^{+\infty} (3^{n-1} + 1)(1-p)^{n-1}$$

La série géométrique $\sum(1-p)^{n-1}$ converge car $|1-p| < 1$. Par contre, la série numérique $\sum(3(1-p))^{n-1}$ diverge vers l'infini car $|3(1-p)| > 1$. En conclusion, $E(G(X)) = +\infty$ et donc ce jeu est "très" favorable au joueur. L'unique condition pour jouer à ce jeu est de n'avoir pas de limite sur son compte en banque....

Définition Variable aléatoire réelle discrète

Soit (Ω, \mathcal{A}) un espace probabilisable.

Une **variable aléatoire réelle discrète** est une application :

$$X \begin{cases} \Omega \longrightarrow \mathbb{R} \\ \omega \longmapsto X(\omega) \end{cases} .$$

telle que

1. $X(\Omega)$ est un ensemble dénombrable;
2. Si $x \in X(\Omega)$ alors $X^{-1}(\{x\}) = (X = x)$ est un événement, c'est-à-dire : $X^{-1}(\{x\}) \in \mathcal{A}$.

Remarque

Lors de la définition de l'axiome des tribus, l'ensemble des événements ne devait ne pas être "trop grand" afin de donner du sens à l'application probabilité. La définition de la variable aléatoire impose que l'ensemble des événements ne doit pas être "trop petit" pour que chaque x , $(X = x)$ soit un élément de \mathcal{A} .

Proposition

Soit (Ω, \mathcal{A}) un espace probabilisable et X une variable aléatoire discrète sur cet espace.

Soit A une partie de \mathbb{R} . L'ensemble $X^{-1}(A) = \{X \in A\}$, est un événement.

Comme toutes les variables aléatoires sont supposées discrètes, on s'autorise désormais à écrire l'ensemble des images sous la forme $X(\Omega) = \{x_n : n \in \mathbb{N}\}$ ou $X(\Omega) = \{x_i : i \in I\}$ avec I ensemble dénombrable.

Proposition

Soit (Ω, \mathcal{A}, P) un espace probabilisé et X une variable aléatoire discrète sur cette espace.

L'ensemble $(X = x_n)_{n \in \mathbb{N}}$ forme un système complet d'événements de Ω .

B Loi d'une variable aléatoire

Définition Loi d'une variable aléatoire

Soit X une variable aléatoire discrète sur un espace probabilisé (Ω, \mathcal{A}, P) .

L'application

$$P_X \begin{cases} X(\Omega) \longrightarrow [0, 1] \\ x \longmapsto P(X = x) \end{cases}$$

est une probabilité sur \mathbb{R} , appelée **loi de la variable X** .

Théorème

Soit $\{x_n : n \in \mathbb{N}\}$ une famille dénombrable de réels deux à deux distincts et $\{p_n : n \in \mathbb{N}\}$ une famille de réels positifs telle que $\sum_{n=1}^{\infty} p_n = 1$. Alors il existe une variable aléatoire discrète X sur un espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans $\{x_n : n \in \mathbb{N}\}$ telle que :

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

C Vecteurs aléatoires

Définition Loi conjointe

Soit X, Y deux variables aléatoires réelles discrètes définies sur (Ω, \mathcal{A}, P) .

On note (X, Y) le **couple de variables aléatoires** prenant ses valeurs dans \mathbb{R}^2 .

La **loi conjointe** du couple (X, Y) est déterminé par :

1. $X(\Omega) \times Y(\Omega) = \{x_i : i \in \mathbb{N}\} \times \{y_j : j \in \mathbb{N}\}$, les valeurs prises par le couple
2. $\forall (i, j) \in \mathbb{N} \times \mathbb{N} : P(X = x_i \cap Y = y_j)$,

On note $p_{i,j} = P(X = x_i \cap Y = y_j)$.

Les événements $((X = x_i) \cap (Y = y_j))_{\forall (i,j) \in \mathbb{N} \times \mathbb{N}}$ forment un système complet d'événements de Ω . En particulier, on a :

$$\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} P(X = x_i \cap Y = y_j) = \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} p_{i,j} = 1.$$

Définition Lois marginales

Soit X, Y deux variables aléatoires réelles discrètes définies sur $(\Omega, \mathcal{P}(\Omega), P)$.

La loi de X appelé **première loi marginale** et la loi de Y **second loi marginale** du couple.

Proposition Relations

Soit X, Y deux variables aléatoires réelles discrètes définies sur (Ω, \mathcal{A}, P) . On a :

$$\forall i \in \mathbb{N} : P(X = x_i) = \sum_{j=1}^{+\infty} P((X = x_i) \cap (Y = y_j))$$

et

$$\forall j \in \mathbb{N} : P(Y = y_j) = \sum_{i=1}^{+\infty} P((X = x_i) \cap (Y = y_j)).$$

Exemple Deux piles

Dans une succession de piles ou faces pour laquelle la probabilité d'obtenir pile est $p \in]0, 1[$ et d'obtenir face est $q = 1 - p$, on note X le rang d'apparition du premier pile et Y celui du second. La loi conjointe de (X, Y) est :

- si $j > i > 1$, $P(X = i, Y = j) = \underbrace{F \dots F}_{i-1} \underbrace{P \dots F}_{j-i-1} F = (1-p)^{i-1}p(1-p)^{j-i-1}p = (1-p)^{j-2}p^2$,
- sinon $P(X = i, Y = j) = 0$.

D Indépendance et lois conditionnelles

Définition Lois conditionnelles

Soit (X, Y) un couple de variables aléatoires sur (Ω, \mathcal{A}, P) .

1. On appelle **loi conditionnelle** de X sachant $(Y = y_j)$ pour $P(Y = y_j) \neq 0$, la probabilité définie par :

$$\forall i \in \mathbb{N} : P_{Y=y_j}(X = x_i) = \frac{P(X = x_i \cap Y = y_j)}{P(Y = y_j)}$$

2. On appelle **loi conditionnelle** de Y sachant $(X = x_i)$ pour $P(X = x_i) \neq 0$, la probabilité définie par :

$$\forall j \in \mathbb{N} : P_{X=x_i}(Y = y_j) = \frac{P(X = x_i \cap Y = y_j)}{P(X = x_i)}$$

Remarque

Les lois marginales et les lois conditionnelles déterminent la loi conjointe :

$$\forall (i, j) \in \mathbb{N} \times \mathbb{N} : P(X = x_i \cap Y = y_j) = P_{X=x_i}(Y = y_j)P(X = x_i)$$

$$\forall (i, j) \in \mathbb{N} \times \mathbb{N} : P(X = x_i \cap Y = y_j) = P_{Y=y_j}(X = x_i)P(Y = y_j)$$

Définition Indépendance

Soit (X, Y) un couple de variables aléatoires sur $(\Omega, \mathcal{P}(\Omega), P)$.

On dit que X et Y sont **indépendantes** si

$$\forall (i, j) \in \mathbb{N} \times \mathbb{N} : P(X = x_i \cap Y = y_j) : P((X = x_i) \cap (Y = y_j)) = P(X = x_i)P(Y = y_j).$$

Remarque

Dans la modélisation d'une expérience aléatoire, l'hypothèse d'indépendance des variables aléatoires est souvent une donnée de l'expérience et non pas une propriété à vérifier. Par exemple, la modélisation de l'expérience du jeter de deux dés serait :

- X_1 , la variable aléatoire représentant le chiffre du premier dé,
- X_2 , la variable aléatoire représentant le chiffre du second dé,
- X_1 et X_2 , supposées indépendantes,
- $S = X_1 + X_2$, la variable aléatoire représentant la somme des deux dés.

Théorème

Soit (X, Y) un couple de variables aléatoires sur (Ω, \mathcal{A}, P) .

Si X et Y sont indépendantes, il en va de même pour les variables aléatoires $f(X)$ et $g(Y)$ où $f : \mathbb{R} \rightarrow \mathbb{R}$ et $g : \mathbb{R} \rightarrow \mathbb{R}$ sont deux applications quelconques.

E Fonction de répartition

Définition Fonction de répartition

Soit X une variable aléatoire discrète sur un espace probabilisé (Ω, \mathcal{A}, P) .

On appelle fonction de répartition de X et on note F_X l'application :

$$F_X \begin{cases} \mathbb{R} \longrightarrow \mathbb{R} \\ x \longmapsto P(X \leq x) \end{cases}.$$

Définition Propriétés de la fonction de répartition

Soit X une variable aléatoire discrète sur un espace probabilisé (Ω, \mathcal{A}, P) .

1. L'application F_X est croissante.
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$.
3. $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Théorème

La loi de probabilité d'une variable aléatoire discrète réelle est caractérisée par sa fonction de répartition. C'est à dire si deux variables aléatoires réelles ont même fonction de répartition, alors elles ont même loi (et réciproquement).

F Espérance

Pour donner du sens à l'espérance $\sum_{x \in X(\Omega)} xP(X = x)$ dans le cas d'une variable discrète, nous sommes confrontés à deux difficultés :

1. $X(\Omega)$ étant au plus dénombrable, on peut l'écrire sous la forme $\{x_n | n \in \mathbb{N}\}$. La série numérique $\sum x_n P(X = x_n)$ converge-t-elle ?

2. La somme ne dépend-elle pas du choix de la bijection entre $X(\Omega)$ et \mathbb{N} , autrement dit de l'ordre de sommation ?

La réponse aux deux questions est oui, du moment que la série est supposée absolument convergente.

Définition Espérance mathématique

Soit X une variable aléatoire discrète sur l'espace probabilisé (Ω, \mathcal{A}, P) . X est dite d'espérance finie si la série $\sum x_n P(X = x_n)$ converge absolument.

Dans ce cas, on appelle **espérance** de X et on note $E(X)$ le réel :

$$E(X) = \sum_{n=0}^{+\infty} x_n P(X = x_n)$$

Exemple Loi géométrique

Soit X une variable aléatoire suivant la loi

$$\forall n \in \mathbb{N}^* : P(X = n) = (1 - p)^{n-1} p.$$

La loi est bien définie car

$$\sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty} (1 - p)^{n-1} p = p \sum_{n=1}^{\infty} (1 - p)^{n-1} \stackrel{\text{série géométrique}}{=} p \frac{1}{1 - (1 - p)} = 1.$$

Montrons que la série géométrique $\sum nP(X = n)$ converge absolument. La série à termes positifs $\sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} n(1 - p)^{n-1} p$ converge par règle de d'Alembert. L'espérance est donc bien définie et est égale à :

$$\sum_{n=1}^{\infty} nP(X = n) = \sum_{n=1}^{\infty} n(1 - p)^{n-1} p = p \sum_{n=1}^{\infty} n(1 - p)^{n-1} \stackrel{\text{dérivée série géométrique}}{=} p \frac{1}{(1 - (1 - p))^2} = \frac{1}{p}.$$

Théorème

Soit X et Y deux variables aléatoires indépendantes.

Alors

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

Théorème Formule de transfert

Soit X est une variable aléatoire réelle et $f : \mathbb{R} \rightarrow \mathbb{R}$.

Alors $f(X)$ est d'espérance finie si et seulement si $\sum f(x_n)P(X = x_n)$ converge absolument. Dans ce cas :

$$E(f(X)) = \sum_{n=0}^{\infty} f(x_n)P(X = x_n).$$

Proposition propriétés de l'espérance

Soit X et Y deux variables aléatoires réelles et $(a, b) \in \mathbb{R}^2$.

- $E(a) = a$.
- $E(aX + bY) = aE(X) + bE(Y)$.
- Si $P(X \geq 0) = 1$, alors $E(X) \geq 0$.
- Si $P(X \leq Y) = 1$, alors $E(X) \leq E(Y)$.

Proposition

Soit X une variable aléatoire à valeurs entières. Si X admet une espérance finie, alors :

$$E(X) = \sum_{n=0}^{\infty} P(X \geq n)$$

G Variance

Définition Moment d'ordre p

Soit X une variable aléatoire discrète et $p \in \mathbb{N}^*$. Si X^p admet une espérance finie, alors on appelle **moment d'ordre p** de X le réel $E(X^p)$.

Définition

Soit X une variable aléatoire réelle discrète admettant un moment d'ordre 2. Alors $(X - E(X))^2$ est d'espérance finie. On appelle **variance** de X et on note $V(X)$ le réel positif :

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

On appelle **écart type** de X et on note $\sigma(X)$ le réel $\sigma(X) = \sqrt{V(X)}$.

Définition Covariance de deux variables aléatoires

Soit X et Y deux variables aléatoires réelles. Leur **covariance** est $E((X - E(X))(Y - E(Y)))$ et on la note $\text{Cov}(X, Y)$.

Proposition Propriétés

Soit X et Y deux variables aléatoires réelles.

- $V(aX + b) = a^2V(X)$.
- Si X et Y sont indépendantes, alors $V(X + Y) = V(X) + V(Y)$.
- $Var(X + Y) = Var(X) + Var(Y) + 2\text{Cov}(X, Y)$.

Définition Corrélation

Soit X et Y deux variables aléatoires réelles de variance non nulles. Leur **coefficient de corrélation** est

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

Proposition Propriétés

Soit X et Y deux variables aléatoires réelles de variance non nulles.

- $\rho(X, Y) \in [-1, 1]$
- $\rho(X, Y) = 1$ si et seulement si il existe deux réels a et b tels que l'événement $(Y = aX + b)$ soit certain.
- Si X et Y sont indépendantes, alors $\rho(X, Y) = 0$. On dit qu'elles sont décorrélées. La réciproque est fausse.

H Lois usuelles

Définition Loi géométrique

Une variable aléatoire réelle X suit la **loi géométrique** de paramètre $p \in]0, 1[$ si

1. $X(\Omega) = \mathbb{N}^*$,
2. $\forall n \in \mathbb{N}^* : P(X = n) = (1 - p)^{n-1}p$.

On note $X \hookrightarrow \mathcal{G}(p)$.

On a : $E(X) = \frac{1}{p}$ et $V(X) = \frac{1-p}{p^2}$.

Une loi géométrique apparaît lorsque l'on s'intéresse à une succession d'épreuves de Bernoulli indépendantes jusqu'à obtenir un premier succès. On parle alors de temps d'attente du premier succès.

Définition Loi de Poisson

Une variable aléatoire réelle X suit la **loi de Poisson** de paramètre λ si

1. $X(\Omega) = \mathbb{N}$,
2. $P(X = n) = e^{-\lambda} \frac{\lambda^n}{n!}$.

On note $X \hookrightarrow \mathcal{P}(\lambda)$.

On a : $E(X) = \lambda$ et $V(X) = \lambda$.

Une loi géométrique apparaît lorsque l'on s'intéresse à des événements rares comme les suicides d'enfants, les arrivées de bateaux dans un port ou les accidents dus aux coups de pied de cheval dans les armée.

I Fonction génératrice

Soit $X \hookrightarrow \mathcal{B}(n, p)$ une v.a. binomiale ; on a $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$. On remarque que les nombres $P(X = k)$ sont précisément les coefficients du polynôme (de la variable s) suivant :

$$G_X(s) := ((1 - p) + ps^n) = \sum_{k=0}^n p^k (1 - p)^{n-k} s^k = \sum_{k=0}^n s^k P(X = k) = E(s^X).$$

Cette dernière expression, $E(s^X)$, se prête à la généralisation :

Définition Fonction génératrice

Soit X une variable aléatoire à valeurs entières. On appelle fonction génératrice des probabilités de X , noté G_X , la fonction :

$$G_X : s \rightarrow E(s^X) = \sum_{n=0}^{+\infty} P(X = n) s^n$$

On appelle série génératrice la série entière associée. On rappelle que G_X est définie sur un intervalle de la forme $] -R, R[$, $] -R, R]$, $[-R, R[$ ou bien $[-R, R]$ où R est le rayon de convergence de la série entière de coefficients $P(X = n)$.

Remarque

Comme $\sum_{n=0}^{+\infty} P(X = n) = 1$, le rayon de convergence de la série génératrice est supérieur ou égale à 1. Dans le cas où la variable aléatoire est finie, la fonction génératrice est polynomiale et $R = +\infty$.

Comme la fonction G_X est la somme d'une série entière, elle est de classe C^∞ sur l'intervalle ouvert $] -R, R[$ et on peut dériver terme à terme la somme de la série. Par unicité du développement en série entière, G_X est la somme de sa série de Taylor donc :

$$\forall n \in \mathbb{N} : P(X = n) = \frac{G_X^{(n)}(0)}{n!}$$

La connaissance de G_X nous permet donc de retrouver la loi de X . On dit alors que la loi de X est caractérisée par sa série génératrice. Il sera dans certains cas intéressant de passer par la fonction génératrice afin de déterminer la loi de probabilité d'une variable donnée.

C'est notamment le cas pour la somme de deux variables aléatoires indépendantes.

Théorème Fonction génératrice de la somme de deux variables indépendantes

Soit X et Y deux variables aléatoires définies sur un espace probabilisé (Ω, \mathcal{A}, P) et à valeurs dans \mathbb{N} avec G_X et G_Y leurs fonctions génératrices de rayon de convergence respectifs R_X et R_Y .

Si X et Y sont indépendantes, le rayon de convergence de la fonction génératrice G_{X+Y} est $R = \min(R_X, R_Y)$ et :

$$\forall t \in]-R, R[: G_{X+Y}(t) = E[t^{X+Y}] = E[t^X t^Y] = E[t^X]E[t^Y] = G_X(t)G_Y(t)$$

Nom	Notation	G_X
Bernoulli	$\mathcal{B}(p)$	$(1-p) + pt$
Binomiale	$\mathcal{B}(n, p)$	$((1-p) + pt)^n$
Uniforme	$\mathcal{U}(\{1, \dots, n\})$	$\frac{1}{n}(1 + t + \dots + t^n) = \frac{1}{n} \frac{1-t^{n+1}}{1-t}$
Géométrique	$\mathcal{G}(p)$	$\frac{pt}{1-(1-p)t}$
Poisson	$\mathcal{P}(\lambda)$	$e^{\lambda(t-1)}$
Fonction génératrice des lois usuelles		

J Convergence et approximations

Proposition Approximation d'une loi binomiale par une loi de Poisson

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires tel que $X_n \hookrightarrow \mathcal{B}(n, p_n)$ et $\lim_{n \rightarrow \infty} np_n = \lambda$.

Alors on a :

$$\forall k \in \mathbb{N} : \lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Démonstration

Soit $k \in \mathbb{N}$. On a :

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} \\ &= \frac{1}{k!} n.(n-1)\dots(n-k+1)(p_n)^k (1 - p_n)^{n-k} \\ &= \frac{1}{k!} 1.(1-1/n)\dots(1-(k+1)/n)(p_n/n)^k (1 - p_n)^{n-k} \end{aligned}$$

D'une part $\overbrace{1.(1-1/n)\dots(1-(k+1)/n)}^{k \text{ facteurs fixes}} \xrightarrow[n \rightarrow +\infty]{} 1$, d'autre part $(p_n/n)^k \xrightarrow[n \rightarrow +\infty]{} \lambda^k$ et enfin $(1 - p_n)^{n-k} = e^{(n-k)\ln(1-p_n)} \xrightarrow[n \rightarrow +\infty]{} e^{-\lambda}$ car $(n-k)\ln(1-p_n) \sim \infty - np_n$. Finalement

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

En pratique, si X est une variable aléatoire suivant la loi binomiale $\mathcal{B}(n, p)$ avec $n \geq 30$, $p \leq 0,10$ et $np \geq 15$, on peut approximer la loi de X par la loi de Poisson de paramètre np .

Ce théorème justifie le fait que la loi de Poisson est utilisée comme modèle de certaines expériences aléatoires (nombre de clients entrant dans un magasin, nombre de coquilles dans une page de journal,...).

Proposition Inégalité de Markov

Soit X une variable aléatoire réelle positive.

Pour tout $a > 0$, on a :

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

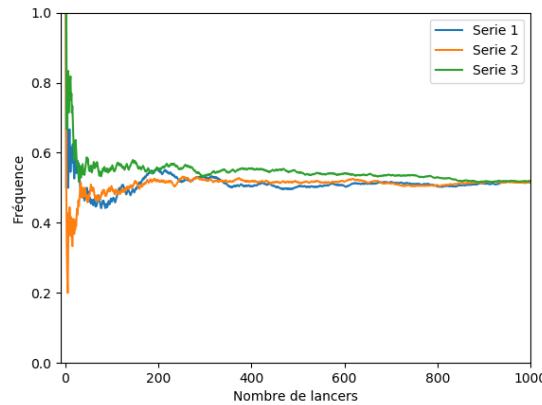
Proposition Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire réelle. Pour tout $a > 0$, on a

$$P(|X - E(X)| \geq a) \leq \frac{V(X)}{a^2}$$

Cette inégalité présente un intérêt théorique en majorant la probabilité qu'une variable aléatoire s'écarte de sa moyenne. Nous allons l'utiliser pour prouver la loi faible des grands nombres.

Lors d'un lancer d'une pièce de monnaie équilibrée, les deux côtés « pile » et « face » apparaissent de façon équiprobable pour des raisons de symétrie : on ne s'attend pas plus à l'un ou à l'autre côté. Cette mesure de l'attente s'appuie souvent sur une considération statistique : on observe que la fréquence des occurrences de chaque côté se rapproche de $1/2$.



Le théorème de la loi des grands nombres permet de justifier ce résultat en interprétant la probabilité comme une fréquence de réalisation.

La modélisation de cette expérience aléatoire est :

1. **n^{ième} lancer** : pour tout $n \in \mathbb{N}^*$, la variable aléatoire X_n , représente le résultat du n^{ième} lancer. Elle suit une loi Bernoulli de paramètre p sachant que l'issue 0 représente le pile au n^{ième} lancer et l'issue 1 le face au n^{ième} lancer. Elles sont supposées indépendantes. On a $E(X_n) = (1-p).0 + p.1 = p$.
2. **fréquence** : pour tout $n \in \mathbb{N}^*$, la variable aléatoire S_n , représente la moyenne des résultats obtenus au cours des n premiers lancers, soit :

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

L'issue, ω , correspondant à une succession de faces existe. Dans ce cas, $X_n(\omega) = 1$ pour tout $n \in \mathbb{N}^*$. Donc 1 est une valeur possible de la variable aléatoire S_n . Cependant, le théorème de la loi faible des grands nombres affirme que la probabilité que S_n s'écarte de l'espérance $E(X_n)$ tend vers 0 quand n tend vers l'infini.

La loi faible des grands nombres démontre que la probabilité d'un événement peut être définie comme la fréquence théorique vers laquelle tend la fréquence observée de cet événement lorsque le nombre n de répétitions tend vers l'infini.

Théorème Loi faible des grands nombres

Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires discrètes définies sur un espace probabilisé (Ω, \mathcal{A}, P) . On suppose toutes les variables indépendantes et de même loi, admettant une espérance m et un écart type σ . Posons $S_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$.

Alors

$$\forall \epsilon > 0 : P(|S_n - m| \geq \epsilon) \xrightarrow{n \rightarrow +\infty} 0.$$

Démonstration

D'après l'inégalité de Bienaymé-Tchebychev et par indépendance des variables aléatoires,

$$\forall \epsilon > 0 : P(|S_n - m| \geq \epsilon) \leq \frac{V(S_n)}{\epsilon^2} = \frac{\frac{V(X_1 + X_2 + \cdots + X_n)}{n^2}}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow{n \rightarrow +\infty} 0.$$

Remarque

Le TP Python https://github.com/VincentTariel/cours/blob/master/probabilite/simulation_variable_aleatoire_avtivite_python.pdf permet de vous familiariser avec ce théorème sur des applications.

Rapidement, on tombe sur le problème suivant : combien de lancers de pièces (ou combien de personnes doit on interroger), pour que le résultat que l'on avance soit correct avec grosse probabilité... En d'autres termes, on aimerait connaître "la vitesse" de convergence. C'est l'objet d'un autre cours sur le théorème centrale limite.