18

Probabilités

Extrait du programme officiel :

Cette section généralise aux variables aléatoires discrètes l'étude menée en première année des variables aléatoires à valeurs dans un ensemble fini. Cette généralisation nécessite d'introduire des notions générales de théorie des probabilités, lesquelles font l'objet d'un exposé a minima. En particulier :

- *la notion de tribu, introduite pour donner un cadre rigoureux, n'appelle aucun développement théorique ;*
- *la construction d'espaces probabilisés n'est pas un objectif du programme ;*
- *les diverses notions de convergence (presque sûre, en probabilité, en loi) sont hors programme.*

La théorie des familles sommables permet une extension très naturelle des notions et résultats vus en première année. Cette extension est effectuée rapidement, de manière à libérer du temps pour les exemples et exercices. L'objectif de l'enseignement est en effet de renforcer la compréhension de l'aléatoire, en lien avec d'autres parties du programme. On pourra ainsi faire travailler les étudiants sur divers objets aléatoires (permutations, graphes, matrices...) les inégalités de concentration et des exemples de processus à temps discret (marches aléatoires, chaînes de Markov...).

La notion de variable à densité est hors programme.

La notion d'espérance conditionnelle est hors programme.

CONTENUS	CAPACITÉS & COMMENTAIRES
b) Espaces probabilisés	
Tribu sur un ensemble Ω. Espace probabilisable $(\Omega,\mathscr{A})$.	La manipulation de tribus n'est pas un objectif du programme.
Événements.	Généralisation du vocabulaire relatif aux événements introduit en première année.
Probabilité sur un espace probabilisable, σ-additivité.	
Espace probabilisé $(\Omega,\mathscr{A},P)$.	
Continuité croissante, continuité décroissante.	Application : pour une suite $(A_n)_{n\in\mathbb{N}}$ d'événements (non nécessairement monotone), limites quand n tend vers l'infini de $$P\Big(\bigcup_{k=0}^{n} A_k\Big) \quad \text{et} \quad P\Big(\bigcap_{k=0}^{n} A_k\Big).$$
Propriété de sous-additivité de P pour une réunion dénombrable d'événements.	
Événements négligeables, événements presque sûrs. Une réunion (resp. intersection) finie ou dénombrable d'événements négligeables (resp. presque sûrs) est un événement négligeable (resp. presque sûr).	Systèmes quasi-complets d'événements. Tout développement supplémentaire sur ces notions est hors programme.
c) Probabilités conditionnelles et indépendance	
Extension des résultats vus en première année : probabilité conditionnelle, formule des probabilités composées, formule des probabilités totales, formule de Bayes.	Notations $P_B(A), P(A\|B)$.
Par définition, les événements A et B sont indépendants si $P(A\cap B)=P(A)P(B)$.	Lorsque $P(B)>0$, l'indépendance de A et B s'écrit $P(A\|B)=P(A)$.
Famille d'événements indépendants.	L'indépendance deux à deux n'implique pas l'indépendance.
Si A et B sont indépendants, A et $\overline{B}$ le sont aussi.	
d) Espaces probabilisés discrets	
Si Ω est un ensemble, une distribution de probabilités discrètes sur Ω est une famille d'éléments de $\mathbb{R}^+$ indexée par Ω et de somme 1.	Support d'une distribution de probabilités discrète ; le support est au plus dénombrable.
Probabilité définie sur $\mathscr{A}=\mathscr{P}(\Omega)$ associée à une distribution de probabilités discrètes sur Ω.	Si Ω est au plus dénombrable, on obtient ainsi toutes les probabilités sur $\mathscr{P}(\Omega)$.

Contenus	Capacités & Commentaires

e) Variables aléatoires discrètes

you if you are still under warranty. They are great to work with so don't hesitate ! Une variable aléatoire discrète X définie sur l'espace probabilisé $(\Omega, \mathscr{A}, P)$ et à valeurs dans E est une application définie sur Ω, à valeurs dans l'ensemble E, telle que $X(\Omega)$ soit au plus dénombrable et que, pour tout $x \in X(\Omega)$, l'ensemble $X^{-1}(\{x\})$ appartienne à $\mathscr{A}$.	Notations $(X = x)$, $(X \in A)$, $\{X = x\}$, $\{X \in A\}$. Lorsque $E = \mathbb{R}$, la variable aléatoire X est dite réelle. Notations $(X \leqslant x), (X \geqslant x), (X < x), (X > x)$ (et analogues avec accolades) pour une variable aléatoire réelle X.
Loi P_X d'une variable aléatoire discrète X.	La loi de X peut au besoin être définie sur un ensemble contenant $X(\Omega)$.
Dans ce qui suit, toutes les variables aléatoires sont supposées discrètes. La probabilité P_X est déterminée par la distribution de probabilités discrète $(P(X = x))_{x \in X(\Omega)}$.	
Notation $X \sim Y$.	La notation $X \sim Y$ ne suppose pas que X et Y sont définies sur le même espace probabilisé.
Variable aléatoire $f(X)$. Si $X \sim Y$ alors $f(X) \sim f(Y)$.	
Loi conditionnelle d'une variable aléatoire X sachant un événement A.	
Couple de variables aléatoires. Loi conjointe, lois marginales. Détermination des lois marginales à partir de la loi conjointe.	Un couple est une variable aléatoire à valeurs dans un produit. Notation $P(X = x, Y = y)$. Extension aux n-uplets de variables aléatoires.

f) Variables aléatoires indépendantes

Couple de variables aléatoires indépendantes, famille finie de variables aléatoires indépendantes.	Notation $X \perp\!\!\!\perp Y$. Les variables aléatoires X et Y sont indépendantes si et seulement si la distribution de probabilités de (X, Y) est le produit des distributions de probabilités de X et Y. Extension aux n-uplets de variables aléatoires.
Famille quelconque de variables aléatoires indépendantes.	
Fonctions de variables aléatoires indépendantes : si $X \perp\!\!\!\perp Y$, alors $f(X) \perp\!\!\!\perp g(Y)$	Extension au cas de plus de deux variables.
Lemme des coalitions : si les variables aléatoires $X_1, \dots, X_n$ sont indépendantes, les variables aléatoires $f(X_1, \dots, X_m)$ et $g(X_{m+1}, \dots, X_n)$ le sont aussi.	Extension au cas de plus de deux coalitions.
Existence d'espaces probabilisés portant une suite de variables indépendantes de lois discrètes données.	La démonstration est hors programme. Suites i.i.d. Modélisation du jeu de pile ou face infini : suite i.i.d. de variables de Bernoulli.

g) Lois usuelles

Pour p dans $]0, 1[$, loi géométrique de paramètre p.	Notations $\mathscr{G}(p)$, $X \sim \mathscr{G}(p)$.
Variable géométrique de paramètre p.	Interprétation comme rang du premier succès dans le jeu de pile ou face infini.
Pour λ dans $\mathbb{R}_+^*$, loi de Poisson de paramètre λ.	Notations $\mathscr{P}(\lambda)$, $X \sim \mathscr{P}(\lambda)$.
Variable de Poisson de paramètre λ.	Interprétation en termes d'événements rares.

h) Espérance d'une variable aléatoire réelle ou complexe

Si X est une variable aléatoire à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$, l'espérance de X est la somme, dans $[0, +\infty]$, de la famille $(x\, P(X = x))_{x \in X(\Omega)}$.	Notation $\mathrm{E}(X)$.
Pour une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$, égalité $\mathrm{E}(X) = \sum_{n=1}^{+\infty} P(X \geqslant n)$.	
Une variable aléatoire complexe X est dite d'espérance finie si la famille $(x\, P(X = x))_{x \in X(\Omega)}$ est sommable ; dans ce cas, la somme de cette famille est l'espérance de X.	Notation $\mathrm{E}(X)$. Variables centrées. La notation $X \in L^1$ signifie que X est d'espérance finie. On ne soulèvera aucune difficulté quant à la définition précise de L^1.
Espérance d'une variable géométrique, d'une variable de Poisson.	

CONTENUS	CAPACITÉS & COMMENTAIRES
Formule de transfert : soit X une variable aléatoire discrète, f une fonction définie sur $X(\Omega)$ à valeurs complexes ; alors $f(X)$ est d'espérance finie si et seulement si la famille $\left(f(x)\,P(X=x)\right)_{x\in X(\Omega)}$ est sommable ; si tel est le cas : $\mathrm{E}\big(f(X)\big) = \sum\limits_{x\in X(\Omega)} f(x)\,P(X=x)$.	
Linéarité, positivité, croissance, inégalité triangulaire.	Caractérisation des variables aléatoires à valeurs dans $\mathbb{R}^+$ d'espérance nulle.
Si $\lvert X\rvert \leqslant Y$ et si $Y \in L^1$, alors $X \in L^1$.	
Si X et Y sont dans L^1 et indépendantes, alors XY est dans L^1 et : $$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y).$$	Extension au cas de n variables aléatoires.

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

Si $\mathrm{E}\big(X^2\big) < +\infty$, X est d'espérance finie.	La notation $X \in L^2$ signifie que X^2 est d'espérance finie. On ne soulèvera aucune difficulté quant à la définition you if you are still under warranty. They are great to work with so don't hesitate !précise de L^2.
Inégalité de Cauchy-Schwarz : si X et Y sont dans L^2, XY est dans L^1 et $\mathrm{E}(XY)^2 \leqslant \mathrm{E}\big(X^2\big)\,\mathrm{E}\big(Y^2\big)$.	Cas d'égalité.
Pour $X \in L^2$, variance et écart type de X.	Notations $\mathrm{V}(X), \sigma(X)$. Variables réduites. Caractérisation des variables aléatoires de variance nulle.
Relation $\mathrm{V}(X) = \mathrm{E}\big(X^2\big) - \mathrm{E}(X)^2$.	
Relation $\mathrm{V}(aX+b) = a^2\,\mathrm{V}(X)$.	Si $\sigma(X) > 0$, la variable aléatoire $\dfrac{X-\mathrm{E}(X)}{\sigma(X)}$ est centrée réduite.
Variance d'une variable géométrique, d'une variable de Poisson.	
Covariance de deux variables aléatoires de L^2.	
Relation $\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$. Cas de variables indépendantes.	
Variance d'une somme de n variables aléatoires, cas de variables décorrélées.	

j) Inégalités probabilistes et loi faible des grands nombres

Inégalité de Markov.	
Inégalité de Bienaymé-Tchebychev.	
you if you are still under warranty. They are great to work with so don't hesitate ! Loi faible des grands nombres : si $(X_n)_{n\geqslant 1}$ est une suite i.i.d. de variables aléatoires de variance finie, alors, pour tout $\varepsilon > 0$, $$P\left(\left\lvert \frac{S_n}{n} - m\right\rvert \geqslant \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0,$$ où $S_n = \sum\limits_{k=1}^{n} X_k$ et $m = \mathrm{E}(X_1)$.	Utilisation des inégalités de Markov et de Bienaymé-Tchebychev pour établir des inégalités de concentration.

k) Fonctions génératrices

Fonction génératrice de la variable aléatoire X à valeurs dans $\mathbb{N}$: $G_X(t) = \mathrm{E}\left(t^X\right) = \sum\limits_{k=0}^{+\infty} P(X=k)\,t^k$.	La série entière définissant G_X est de rayon supérieur ou égal à 1 et converge normalement sur le disque fermé de centre 0 et de rayon 1. Continuité de G_X.
Détermination de la loi de X par G_X.	
La variable aléatoire X est d'espérance finie si et seulement si G_X est dérivable en 1 ; dans ce cas $\mathrm{E}(X) = G_X{}'(1)$.	La démonstration de la réciproque n'est pas exigible. Utilisation de G_X pour le calcul de $\mathrm{E}(X)$ et $\mathrm{V}(X)$. Les étudiants doivent savoir calculer rapidement la fonction génératrice d'une variable aléatoire de Bernoulli, binomiale, géométrique, de Poisson.
Fonction génératrice d'une somme finie de variables aléatoires indépendantes à valeurs dans $\mathbb{N}$.	

Table des matières

I Espace probabilisé

1 Tribu

Comme vu en première année dans le cadre des probabilités finies, on appelle **univers**, noté en général Ω, l'ensemble des **issues** ou **résultats** ou **réalisations** d'une expérience aléatoire.

Commençons par rappeler quelques situations modèles dans le cadre des univers finis : tirage de p boules dans une urne en contenant n, numérotées de 1 à n.

Exemple

E 1 – **Tirages successifs, avec remise** : Dans ce cas, l'ordre est important, et il peut y avoir répétition. On peut choisir comme modèle de résultat un p-uplet (ou p-liste) d'éléments de $[\![1,n]\!]$.
On pose $\Omega = [\![1,n]\!]^p$, et alors $|\Omega| = n^p$.
Il s'agit aussi du nombre d'applications d'un ensemble à p éléments vers un ensemble à n éléments (à chaque tirage, on associe sa boule).

E 2 – **Tirages successifs, sans remise** : Dans ce cas, l'ordre est important, et il ne peut pas y avoir répétition. On peut choisir comme modèle de résultat un p-uplet d'éléments deux à deux distincts (ou p-arrangements) de $[\![1,n]\!]$.
On pose $\Omega = \mathscr{A}_p\left([\![1,n]\!]\right)$, et alors

$$|\Omega| = A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}$$

(notation hors-programme).
Il s'agit aussi du nombre d'injections d'un ensemble à p éléments vers un ensemble à n éléments (à chaque tirage, on associe sa boule).

E 3 – **Tirage simultané** : Dans ce cas, l'ordre n'est pas important, et il ne peut pas y avoir répétition. On peut choisir comme modèle de résultat une partie à p éléments de $[\![1,n]\!]$ (ou p-combinaison).
On pose $\Omega = \mathscr{P}_p\left([\![1,n]\!]\right)$, et alors

$$|\Omega| = \binom{n}{p} = \frac{n(n-1)\cdots(n-p+1)}{p!} = \frac{n!}{p!(n-p)!}.$$

Rappelons également qu'une même expérience peut donner lieu à différents univers possible selon ce que l'on souhaite observer (par exemple : carte d'une main vs couleur seulement de la carte, ou bien résultat d'un dé vs parité de ce résultat, etc.)

C'est encore plus vrai pour des univers infinis : le cadre formel que l'on va se donner prévoit que certaines parties de l'univers Ω seulement soient « observables » (les événements), afin de définir une probabilité dans ce cadre plus général.

Remarque

R 1 – Conformément aux habitudes probabilistes, on note, pour $A \subset \Omega$, $\overline{A} = \Omega \setminus A$ le complémentaire dans Ω de A. Cela n'a absolument rien à voir avec l'adhérence topologique, donc.

R 2 – Convention d'écriture : les lettres cursives $\mathscr{A}$, $\mathscr{P}$... désignent des ensembles d'ensembles tandis que les lettres droites A, B désignent des ensembles simples (comme les événements).

Définition 1 : Tribu

Soit Ω un ensemble. On appelle **tribu** sur Ω toute partie $\mathscr{A}$ de $\mathscr{P}(\Omega)$ telle que

(i) $\varnothing \in \mathscr{A}$

(ii) **Stabilité par passage au complémentaire** : $A \in \mathscr{A} \Longrightarrow \overline{A} \in \mathscr{A}$

(iii) **Stabilité par réunion dénombrable** : Si $(A_n)_n$ est une suite d'éléments de A, $\displaystyle\bigcup_{n=0}^{+\infty} A_n \in \mathscr{A}$

Le couple $(\Omega, \mathscr{A})$ est appelé **espace probabilisable**, et les éléments de $\mathscr{A}$ ses **événements**.

Exemple

E 4 – $\mathscr{P}(\Omega)$ est une tribu sur Ω (dite **discrète**)

E 5 – $\{\varnothing, \Omega\}$ est une tribu sur Ω (dite **grossière**)

E 6 – Si A est une partie non vide de Ω, distincte de Ω, la plus petite tribu sur Ω contenant A est $\left\{\varnothing, A, \overline{A}, \Omega\right\}$.

Le vocabulaire vu en première année reste valable :

- Si $A \in \mathscr{A}$, $\overline{A}$ est l'**événement contraire** (qui est bien un événement).
- L'événement $\varnothing$ est appelé **événement impossible**.
- On dit que deux événements A et B sont **incompatibles** lorsque $A \cap B = \varnothing$.

Ne pas confondre issue = résultat = réalisation avec événement !

D'après le programme officiel, *la manipulation de tribu n'est pas un objectif du programme* : elles servent de cadre théorique mais, dans la pratique, on n'attend pas nécessairement de les préciser.

Propriété 1 : des tribus

Une tribu est stable par réunion finie, par intersection dénombrable, par intersection finie.
Ainsi dit, si $\mathscr{A}$ est une tribu sur Ω,

(i) $\Omega \in \mathscr{A}$

(ii) Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathscr{A}$, $\displaystyle\bigcap_{n=0}^{+\infty} A_n \in \mathscr{A}$

(iii) Si $(A_n)_{0\leqslant n\leqslant N}$ est une famille finie d'éléments de $\mathscr{A}$, $\displaystyle\bigcup_{n=0}^{N} A_n \in \mathscr{A}$ *et* $\displaystyle\bigcap_{n=0}^{N} A_n \in \mathscr{A}$

Exercice 1

Soit $\mathscr{A}$ une tribu d'événements d'un espace probabilisable Ω et $(A_n)_{n\geqslant 1}$ une famille de $\mathscr{A}$. Décrire à l'aide des opérations ou comparaisons ensemblistes usuelles les situations ou les événements suivants (sauf pour les items 4 à 6, on écrira des choses du type « $\omega \in E$ » où E est un ensemble à déterminer).

1. L'un au moins des événements A_1, A_2, A_3 est réalisé.
2. L'un seulement des événements A_1 et A_2 est réalisé.
3. A_1 et A_2 se réalisent mais pas A_3.
4. À chaque fois que A_1 est réalisé, A_2 l'est aussi.
5. A_1 et A_2 ne se produisent jamais ensemble.
6. A_1 ou A_2 se produisent toujours.
7. Tous les événements $(A_i)_{i\in\mathbb{N}^*}$ se réalisent.
8. L'un au moins des A_i se réalise.
9. Tous les événements A_i se réalisent à partir du rang i_0.
10. Tous les événements A_i se réalisent à partir d'un certain rang.
11. Une infinité d'événements A_i se réalisent.
12. Seul un nombre fini d'événements A_i se réalisent.

---

1. $\omega \in A_1 \cup A_2 \cup A_3$
2. $\omega \in (A_1 \setminus A_2) \cup (A_2 \setminus A_1) = A_1 \Delta A_2$ (différence symétrique)
3. $\omega \in A_1 \cup A_2 \cup \overline{A_3}$
4. $A_1 \subset A_2$
5. $A_1 \cap A_2 = \varnothing$
6. $A_1 \cup A_2 = \Omega$
7. $\omega \in \bigcap_{i=1}^{+\infty} A_i$
8. $\omega \in \bigcup_{i=1}^{+\infty} A_i$
9. $\omega \in \bigcap_{i=i_0}^{+\infty} A_i$
10. $\omega \in \bigcup_{j=1}^{+\infty}\left(\bigcap_{i=j}^{+\infty} A_i\right)$
11. $\omega \in \bigcap_{j=1}^{+\infty}\left(\bigcup_{i=j}^{+\infty} A_i\right)$
12. $\omega \in \bigcup_{j=1}^{+\infty}\left(\bigcap_{i=j}^{+\infty} \overline{A_i}\right)$

2 Probabilité

Définition 2 : Probabilité

Soit $(\Omega, \mathscr{A})$ un espace probabilisable. Une **probabilité** (ou **mesure de probabilité**) sur $(\Omega, \mathscr{A})$ est une application $\mathbb{P}$ définie sur $\mathscr{A}$ telle que

(i) $\forall A \in \mathscr{A}, \quad \mathbb{P}(A) \in [0,1]$ ($\geqslant 0$ suffirait)

(ii) $\mathbb{P}(\Omega) = 1$

(iii) **σ-additivité** : Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements deux à deux disjoints (incompatibles), $\sum \mathbb{P}(A_n)$ converge et $\mathbb{P}\left(\bigsqcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$.

On dit alors que le triplet $(\Omega, \mathscr{A}, \mathbb{P})$ est un **espace probabilisé**.

Propriété 2 : d'une probabilité

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, A et B des événements : $A, B \in \mathscr{A}$.

(i) $\mathbb{P}(\varnothing) = 0$.

(ii) *Si A et B sont deux événements incompatibles, $\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B)$.*

Plus généralement, $\mathbb{P}\left(\bigsqcup_{n=0}^{N} A_n\right) = \sum_{n=0}^{N} \mathbb{P}(A_n)$.

(iii) *Si $A \subset B$, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$.Si A et B sont quelconques, $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$.*

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

(v) ***Croissance*** : *si $A \subset B$, $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.*

Démonstration

(i) Il suffit de considérer la suite d'éléments de $\mathscr{A}$ constituée uniquement d'ensembles vides.
(ii) Il suffit de compléter la famille finie par des ensembles vides et d'utiliser la σ-additivité.
(iii) $B = (B \setminus A) \sqcup (A \cap B)$.
(iv) $A \cup B = A \sqcup (B \setminus A)$.
(v) $B = A \sqcup (B \setminus A)$.

Remarque

R 3 – Pour trois événements,

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

Et plus généralement, **Formule de Poincaré** (HP) : pour toute famille $(A_1, \dots, A_n) \in \mathscr{P}(\Omega)^n$,

$$\begin{aligned}\mathbb{P}(A_1 \cup \cdots \cup A_n) &= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leqslant i < j \leqslant n} \mathbb{P}(A_i \cap A_j) + \sum_{1 \leqslant i < j < k \leqslant n} \mathbb{P}(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n-1} \mathbb{P}(A_1 \cap \cdots \cap A_n) \\ &= \sum_{k=1}^{n} \left((-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} \mathbb{P}\left(A_{i_1} \cap \cdots \cap A_{i_k}\right) \right) \\ &= \sum_{k=1}^{n} \left((-1)^{k-1} \sum_{I \in \mathscr{P}_k(\llbracket 1,n \rrbracket)} \mathbb{P}\left(\bigcap_{i \in I} A_i \right) \right) = \sum_{I \in \mathscr{P}(\llbracket 1,n \rrbracket)} (-1)^{|I|-1} \mathbb{P}\left(\bigcap_{i \in I} A_i \right).\end{aligned}$$

qui peut se montrer par récurrence par exemple, ou, plus simplement, en remarquant que $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i \right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i} \right) = 1 - \mathbb{E}\left(\mathbb{1}_{\bigcap_{i=1}^{n} \overline{A_i}} \right) = \ldots$ Voir l'exercice 15 du cours.

Propriété 3 : Probabilité d'une réunion au plus dénombrable

Si $(A_i)_{i \in I}$ est une famille au plus dénombrable d'événements deux à deux incompatibles, alors $\left(\mathbb{P}(A_i)\right)_{i \in I}$ est sommable et $\mathbb{P}\left(\bigsqcup_{i \in I} A_i \right) = \sum_{i \in I} \mathbb{P}(A_i)$.

Démonstration

C'est immédiat si I est fini, et sinon, via une énumération de I, on se ramène à une somme sur $\mathbb{N}$ et à la σ-additivité.

Définition 3 : Distribution de probabilités

Soit Ω un ensemble. On appelle **distribution de probabilités** sur Ω toute famille d'éléments de $\mathbb{R}^+$ indexée par Ω et somme (finie) égale à 1.
On appelle **support** d'une telle distribution $(p_\omega)_{\omega \in \Omega}$ l'ensemble $\{\omega \in \Omega,\ p_\omega \neq 0\}$.

Propriété 4 : Support au plus dénombrable

Le support d'une distribution de probabilités est toujours au plus dénombrable.

Démonstration

C'est une propriété connue des familles sommables.

3 Cas très simple : univers fini

Si Ω est fini, on prend généralement $\mathscr{A} = \mathscr{P}(\Omega)$, et la propriété de σ-additivité est équivalente à la propriété

Si A et B sont deux événements disjoints, alors $\mathbb{P}(A \sqcup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Propriété 5 : Probabilité finie associée à une distribution

Si $\Omega = \{\omega_1, \dots, \omega_m\}$, $\mathbb{P}$ est entièrement définie par la donnée d'une distribution de probabilités $(p_{\omega_i})_{1 \leqslant i \leqslant m}$ telle que pour tout $i \in [\![1, m]\!]$, $\mathbb{P}(\{\omega_i\}) = p_{\omega_i}$. Et, pour toute partie A de Ω,

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

Les probabilités des événements élémentaires déterminent donc $\mathbb{P}$.

Démonstration

Comme en MP2I. ∎

4 Cas simple : univers dénombrable

Ici, on garde la propriété de σ-additivité, que l'on ne peut plus remplacer par la simple additivité.
Ici encore, il n'y a pas d'obstacle à prendre la tribu « discrète », c'est-à-dire $\mathscr{P}(\Omega)$. On obtient :

Propriété 6 : Probabilité discrète associée à une distribution

Soit Ω un ensemble dénombrable. Pour toute distribution de probabilités $(p_\omega)_{\omega \in \Omega}$, il existe une unique probabilité sur $\mathscr{P}(\Omega)$ telle que

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

Cette probabilité vérifie

$$\forall A \in \mathscr{P}(\Omega), \qquad \mathbb{P}(A) = \sum_{\omega \in A} p_\omega$$

Démonstration

Analyse-synthèse. C'est une conséquence de ce qui a été vu dans le chapitre sur la sommabilité (sommation par paquets). ∎

Donc, encore une fois, $\mathbb{P}$ est définie de manière unique par les probabilités des singletons. Pour un univers fini ou dénombrable, les tribus d'évènements n'ont donc pas grand intérêt.

5 Cas moins simple : univers non dénombrable

Dans le cas où l'univers est infini indénombrable c'est plus compliqué : on peut montrer que pour un tirage à pile ou face infini non dénombrable, modélisé par $\{0,1\}^{\mathbb{N}}$ (non dénombrable par argument diagonal de Cantor), la seule valeur possible pour la probabilité d'un événement élémentaire est... 0.

Pourquoi ? Intuitivement, si la probabilité d'obtenir un pile est $p \in]0,1[$, alors la probabilité d'obtenir n piles de suite de va être $p^n \xrightarrow[n \to +\infty]{} 0$... Donc, il est légitime de penser que l'événement « n'obtenir que des piles » a une probabilité nulle, par exemple.

C'est donc moins simple, on en peut pas se contenter des événements élémentaires, mais complètement hors-programme.

6 Continuités croissante et décroissante

Propriété 7 : Continuité croissante

Soit $(A_n)_{n\in\mathbb{N}}$ *une suite croissante (pour l'inclusion) d'événements :*

$$\forall\, n \in \mathbb{N},\quad A_n \subset A_{n+1}$$

Alors

$$\mathbb{P}(A_k) \xrightarrow[k\to+\infty]{} \mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right)$$

Remarque

R4 – Continuité car $\bigcup_{n=0}^{+\infty} A_n$ est en quelque sorte la « limite » de la suite croissante $(A_n)_{n\in\mathbb{N}}$.

Remarquons aussi que $A_k = \bigcup_{n=0}^{k} A_n$.

Démonstration

$\left(\mathbb{P}(A_k)\right)_{k\in\mathbb{N}}$ est une suite croissante majorée donc convergente vers $\ell \in [0,1]$.

Or, on remarque, par croissance, que $\bigcup_{n=0}^{+\infty} A_n = \bigsqcup_{n=0}^{+\infty} (A_n \setminus A_{n-1})$ en notant $A_{-1} = \varnothing$, donc par σ-additivité,

$$\mathbb{P}\left(\bigcup_{n=-1}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \left(\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\right) = \ell - 0$$

par croissance puis télescopage. ■

Corollaire 1 : Limite d'une probabilité d'une réunion

Soit $(A_n)_{n\in\mathbb{N}}$ *une suite quelconque d'événements, alors*

$$\mathbb{P}\left(\bigcup_{n=0}^{k} A_n\right) \xrightarrow[k\to+\infty]{} \mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right)$$

Propriété 8 : Continuité décroissante

Soit $(A_n)_{n\in\mathbb{N}}$ *une suite décroissante (pour l'inclusion) d'événements :*

$$\forall\, n \in \mathbb{N},\quad A_{n+1} \subset A_n$$

Alors

$$\mathbb{P}(A_k) \xrightarrow[k\to+\infty]{} \mathbb{P}\left(\bigcap_{n=0}^{+\infty} A_n\right)$$

Démonstration

Il suffit de passer au complémentaire. ■

Corollaire 2 : Limite d'une probabilité d'une intersection

Soit $(A_n)_{n\in\mathbb{N}}$ une suite quelconque d'événements, alors

$$\mathbb{P}\left(\bigcap_{n=0}^{k} A_n\right) \xrightarrow[k\to+\infty]{} \mathbb{P}\left(\bigcap_{n=0}^{+\infty} A_n\right)$$

7 Inégalité de Boole

Propriété 9 : Inégalité de Boole

Soit $(A_n)_{n\in\mathbb{N}}$ une suite quelconque d'événements. Alors, dans $[0,+\infty]$,

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

Remarque

R5 – Où, si la série à termes positifs $\sum \mathbb{P}(A_n)$ diverge, on lira la formule $\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) \leqslant +\infty$, ce qui ne dit rien.
Si la série converge et a une somme $\geqslant 1$, le résultat ne dit rien non plus.

Démonstration

Par récurrence sur n, on montre facilement (comme en première année) que pour tout $k \in \mathbb{N}$, $\mathbb{P}\left(\bigcup_{n=0}^{k} A_n\right) \leqslant \sum_{n=0}^{k} \mathbb{P}(A_n)$ et on conclut en faisant $k \to +\infty$ et en utilisant le corollaire de la continuité croissante. ■

8 Négligeabilité

a Événements négligeables

Définition 4 : Événement négligeable

On dit qu'un événement A est **négligeable** lorsque $\mathbb{P}(A) = 0$.

Remarque

R6 – L'événement impossible est négligeable.
Un événement négligeable n'est pas en général impossible.

Exemple

E7 – Dans le jeu de Pile ou Face infini, l'événement « la pièce donne Pile un nombre fini de fois » est négligeable.
Voir plus loin pour une justification rigoureuse !

Propriété 10 : Partie d'un événement négligeable

Si A et B sont deux événements tel que $A \subset B$, si B est négligeable, A l'est.

Démonstration

C'est la croissance. ■

Propriété 11 : Réunion, intersection finie ou dénombrable

Une réunion (respectivement intersection non vide) finie ou dénombrable d'événements négligeables est négligeable.

Démonstration

Conséquence de l'inégalité de Boole pour la réunion.
Pour l'intersection finie : elle est un incluse dans l'un des événements négligeables. ■

b Événements, propriétés presque sûr·e·s

Définition 5 : Événement presque sûr

Un événement A est **presque sûr**, ou **presque certain**, lorsque $\mathbb{P}(A) = 1$, ce qui équivaut à dire que $\overline{A}$ est négligeable.

Une propriété est dite **presque sûre** lorsque l'ensemble des éléments de Ω qui ont cette propriété est un événement presque sûr.

Propriété 12 : Réunion, intersection au plus dénombrable

Toute réunion non vide (respectivement intersection) finie ou dénombrable d'événements presque sûrs l'est encore.

Démonstration

Il suffit de passer au complémentaire et d'utiliser les propriétés des événements négligeables. ■

II Conditionnement

Les notions vues en première année se généralisent sans problème particulier.

1 Probabilité conditionnelle

Définition 6

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, B un événement tel que $\mathbb{P}(B) > 0$. Pour tout événement $A \in \mathscr{A}$, on définit la **probabilité conditionnelle de A sachant B** par

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

(*Se lit en général « probabilité de A sachant B »*)

Remarque

R 7 – ⚠ Il n'y a toujours pas d'« événement conditionnel $A|B$ » (élément de $\mathscr{A}$) : ce n'est qu'une notation signifiant qu'on se place en observateur de l'événement A sachant que l'événement B est déjà réalisé.
Mais la notation $\mathbb{P}_B$ peut aussi être trompeuse, car c'est la même que celle de la loi d'une variable aléatoire.

Propriété 13 : Probabilité... conditionnée

$\mathbb{P}_B : A \in \mathscr{A} \mapsto \mathbb{P}(A|B)$ *est une probabilité sur* $(\Omega, \mathscr{A})$.

Démonstration

Seule la σ-additivité demande un peu de travail, mais rien de bien sorcier : si les A_n sont deux à deux disjoints, les $A_n \cap B$ le sont et $\left(\bigsqcup_{n=0}^{+\infty} A_n\right) \cap B = \bigsqcup_{n=0}^{+\infty} (A_n \cap B)$ permet de l'obtenir sans encombre. ■

... et donc toutes les propriétés des probabilités, toutes les formules qui vont suivre peuvent être appliquées à des probabilités conditionnelles.

Lorsque que plusieurs conditions s'enchaînent, il suffit de les intersecter :

$$\text{« } \mathbb{P}(A|B|C) \text{ »} = \mathbb{P}_C(A|B) = \mathbb{P}(A|B \cap C).$$

2 Probabilités composées

Propriété 14 : Formule des probabilités composées

Soit $n \geqslant 2$, $A_1, \ldots, A_n$ *des événements de l'espace probabilisé* $(\Omega, \mathscr{A}, \mathbb{P})$ *tels que* $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$.

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \times \mathbb{P}(A_2 \mid A_1) \times \mathbb{P}(A_3 \mid A_1 \cap A_2) \times \cdots \times \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

Remarque

R 8 – À nouveau, cela correspond à notre intuition : on réalise A_1, puis A_2 sachant que A_1 l'est, puis A_3 sachant que A_1 et A_2 le sont, etc. On se sert donc en général de cette formule lorsque l'on a des événements successifs, **chronologiques**.

Démonstration

Remarquons que comme pour tout $k < n$, $A_1 \cap \cdots \cap A_{n-1} \subset A_1 \cap \cdots \cap A_k$, pour tout k, $\mathbb{P}(A_1 \cap \cdots \cap A_k) > 0$ et donc tous les termes ont un sens.

$$\mathbb{P}(A_1) \times \prod_{i=1}^{n-1} \mathbb{P}\left(A_{i+1} \mid A_1 \cap \cdots \cap A_i\right) = \mathbb{P}(A_1) \times \prod_{i=1}^{n-1} \frac{\mathbb{P}\left(A_1 \cap \cdots \cap A_{i+1}\right)}{\mathbb{P}\left(A_1 \cap \cdots \cap A_i\right)} = \mathbb{P}(A_1) \times \frac{\mathbb{P}(A_1 \cap \cdots \cap A_n)}{\mathbb{P}(A_1)}$$

car le produit est télescopique. ■

3 Probabilités totales

Définition 7 : Système complet et quasi-complet d'événements

Soit $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé, I un ensemble fini ou dénombrable. On dit que la famille $(A_i)_{i \in I}$ d'événements est un **système complet d'évènements** lorsque

$$(i \neq j) \implies (A_i \cap A_j = \varnothing) \qquad \text{et} \qquad \bigsqcup_{i \in I} A_i = \Omega$$

On dit que la famille $(A_i)_{i \in I}$ d'événements est un **système quasi-complet d'évènements** lorsque

$$(i \neq j) \implies (A_i \cap A_j = \varnothing) \qquad \text{et} \qquad \sum_{i \in I} \mathbb{P}(A_i) = 1$$

Remarque

R 9 – Si on impose de plus les A_i non vides, ce qui se fait parfois, $\{A_i\}_{i \in I}$ est une partition de Ω.

R 10 – Un système complet d'événement est quasi-complet, mais la réciproque est fausse.

R 11 – Si $(A_i)_{i\in I}$ est un système quasi-complet d'événements, en lui ajoutant l'événement négligeable $\overline{\bigsqcup_{i\in I} A_i}$, on obtient un système complet d'événements.

Propriété 15 : Formule des probabilités totales

Si $(A_i)_{i\in I}$ où I est fini ou dénombrable est un système complet ou quasi-complet d'événements, alors pour tout événement B,

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

Si, de plus, pour tout i, $\mathbb{P}(A_i)>0$ (A_i n'est pas négligeable),

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

Si certains événements sont négligeables, alors les $B\cap A_i$ le seront aussi et il suffit de remplacer la somme pour $i\in I$ par la somme pour $i\in J=\{i\in I,\ \mathbb{P}(A_i)>0\}$.

Démonstration

Le fait d'avoir un sce permet d'écrire $B=B\cap\Omega=\bigsqcup_{i\in I}(B\cap A_i)$ et de conclure.

En cas de système quasi-complet d'événements, on peut poser $A=\bigsqcup_{i\in I} A_i$ et utiliser le sce $\left(A,\overline{A}\right)$ avec $\mathbb{P}\left(\overline{A}\right)=0$:

$$\mathbb{P}(B)=\mathbb{P}(B\cap A)+\mathbb{P}\left(B\cap\overline{A}\right)$$

avec $B\cap\overline{A}\subset\overline{A}$ donc $\mathbb{P}\left(B\cap\overline{A}\right)=0$ donc

$$\mathbb{P}(B)=\mathbb{P}(B\cap A)=\mathbb{P}\left(\bigsqcup_{i\in I}(B\cap A_i)\right)=\sum_{i\in I}\mathbb{P}(B\cap A_i)\quad\blacksquare$$

Remarque

R 12 – La formule des probabilités totales est utile lorsque l'on fait une expérience aléatoire en plusieurs étapes. Elle permet de raisonner par disjonction de cas, suivant le résultat de la première étape.

⚠ ne pas confondre $\mathbb{P}(B\cap A_i)$ et $\mathbb{P}(B\mid A_i)$!

Exercice 2 : CCINP 101

Exercice 3 : CCINP 107

CC BY NC SA

4 Formule de Bayes

Propriété 16 : Formule de Bayes

Si A, B sont des événements non négligeables, alors

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

Si, de plus, $\overline{A}$ n'est pas négligeable,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\ \mathbb{P}(A)}{\mathbb{P}(B \mid A)\ \mathbb{P}(A) + \mathbb{P}\left(B \mid \overline{A}\right)\ \mathbb{P}\left(\overline{A}\right)}.$$

Plus généralement, si $(A_i)_{i\in I}$ (I fini ou dénombrable) est un système complet ou quasi-complet d'événements non négligeables, on a

$$\forall i \in I \qquad \mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\ \mathbb{P}(A_i)}{\sum\limits_{k\in I} \mathbb{P}(B \mid A_k)\mathbb{P}(A_k)}.$$

Démonstration

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

Remarque

R 13 – Formule permettant de remonter le temps, appelée aussi formule de probabilité des causes.

Exercice 4 : CCINP 105

III Événements indépendants

1 Couple d'événements indépendants

Définition 8 : Indépendance de deux événements

Deux événements A et B d'un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sont dits **indépendants** lorsque

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

On note $A \perp\!\!\!\perp B$ lorsque A et B sont indépendants.

Propriété 17 : Caractérisation par probabilités conditionnelles

Deux événements A et B d'un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ tels que $\mathbb{P}(B) > 0$ sont ***indépendants*** *si et seulement si $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.*

Remarque

R 14 – Cela traduit bien notre intuition : que B soit réalisé ou non, la probabilité de A ne change pas.

R 15 – Bien sûr, si $\mathbb{P}(A) > 0$, cela s'écrit aussi $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

R 16 – ⚠ Ne pas confondre l'indépendance de deux événements et le fait qu'ils soient incompatibles. Ces notions s'excluent en général. En effet, si A et B sont incompatibles, $A \cap B = \varnothing$, donc $\mathbb{P}(A \cap B) = 0$. Si A et B sont de probabilité non nulle, ils ne sont pas indépendants. (Ce qui se comprend car $A \subset \overline{B}$ par exemple).

R 17 – Contrairement à l'incompatibilité qui est une notion ensembliste, l'indépendance est une notion probabiliste : elle dépend de la probabilité dont est muni Ω.

R 18 – Il n'est pas toujours facile de prédire si deux événements sont indépendants.

Naturellement, si deux événements sont indépendants, leurs complémentaires le sont. Plus précisément :

Propriété 18 : Indépendance et complémentaire

Si deux événements A et B d'un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sont indépendants, alors

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

Démonstration

- $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(\overline{B})$.
- Idem.
- d'après les deux premières. ∎

Remarque

R 19 – Si A, B sont indépendants et A, C aussi, on ne peut rien dire en général de A et $B \cap C$ et de A et $B \cup C$.

Exercice 5

On lance deux pièces équilibrées et l'on considère les événements A « le premier lancer donne Pile », B « le deuxième lancer donne Pile » et C « les deux lancer donnent le même résultat ».

Montrer que A, B, C sont deux à deux indépendants mais que A n'est indépendant ni de $B \cap C$, ni de $B \cup C$.

$\Omega = [\![0,1]\!]^2$, probabilité uniforme.
$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$, $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(A \cap C) = \frac{1}{4}$, $\mathbb{P}(B \cup C) = \frac{3}{4}$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B \cap C).$$

$$\mathbb{P}\big(A \cap (B \cup C)\big) = \mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B \cup C).$$

2 Famille d'événements indépendants

Définition 9 : Événements indépendants vs 2 à 2 indépendants

Soit $(A_i)_{i \in I}$ avec I fini ou dénombrable une famille d'événements.

- Les A_i sont dit **deux à deux indépendants** lorsque pour tout $i \neq j$, A_i et A_j sont indépendant, c'est-à-dire que $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$.
- Les A_i sont dit **indépendants**, lorsque pour toute partie **finie** non vide J de I,

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

Remarque

R 20 – C'est une propriété très forte : elle demande de vérifier énormément conditions! En général, les espaces probabilisés sont construits pour avoir des événements indépendants et on n'a donc pas à le vérifier à la main.

R 21 – L'indépendance est stable par extraction de sous-familles.

Propriété 19 : Indépendants ⇒ 2 à 2 ⫫

Si les A_i sont indépendants, alors ils sont deux à deux indépendants.
La réciproque est fausse si $n \geqslant 3$.

Démonstration

Prendre $|I| = 2$.
Dans l'exemple précédent, A, B, C sont deux à deux indépendants mais ne sont pas indépendants globalement. ■

Remarque

R 22 – Attention c'est l'inverse des nombres premiers / polynômes entre eux :

indépendants globalement ⇒ deux à deux.

R 23 – Si les événements A_i sont deux à deux indépendants et si pour tout i on pose $B_i = A_i$ ou $\overline{A_i}$, alors les B_i sont deux à deux indépendants d'après la propriété vue précédemment. Cela se généralise aux événements indépendants :

Propriété 20 : Passages au complémentaire dans l'indépendance

Si les événements A_i pour $i \in I$ sont indépendants et si pour tout $i \in I$ on pose $B_i = A_i$ ou $\overline{A_i}$, alors les B_i sont indépendants.

Démonstration

On suppose I fini, ce qui ne nuit pas à la généralité, car l'indépendance est celle de toute sous famille finie.
On suppose que $B_k = \overline{A_k}$ et si $i \neq k$, $B_i = A_i$ (il n'y a qu'un complémentaire).
Alors, si J est une partie non vide de I,

- Soit $k \notin J$, $\mathbb{P}\left(\bigcap_{i\in J} B_i\right) = \mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i) = \prod_{i\in J} \mathbb{P}(B_i)$.
- Si $k \in J$, $A_k \perp\!\!\!\perp \bigcap_{i\in J\setminus\{k\}} A_i$ donc $\overline{A_k} \perp\!\!\!\perp \bigcap_{i\in J\setminus\{k\}} A_i$ d'après le cas de deux événements, donc

$$\mathbb{P}\left(\bigcap_{i\in J} B_i\right) = \mathbb{P}\left(\overline{A_k} \cap \bigcap_{i\in J\setminus\{k\}} A_i\right) = \mathbb{P}\left(\overline{A_k}\right) \prod_{i\in J\setminus\{k\}} \mathbb{P}(A_i) = \prod_{i\in J} \mathbb{P}(B_i).$$

Puis récurrence sur le nombre de complémentaires. ■

Exercice 6 : Indicatrice d'Euler

Soit $\Omega = [\![1, n]\!]$ où n est un entier non premier supérieur ou égal à 2, muni de la probabilité uniforme. Si $d|n$, on note $A_d = \{kd \mid k \in \Omega \text{ et } kd \in \Omega\}$.

1. **Quelle est la probabilité de A_d ?**
2. **Soit P l'ensemble des diviseurs premiers de n.**
   (a) **Démontrer que $(A_p)_{p\in P}$ est une famille d'événements indépendants.**
   (b) **En déduire le cardinal $\varphi(n)$ de l'ensemble A des nombres inférieurs ou égaux à n et premiers avec n (indicatrice d'Euler).**

1. $\mathbb{P}(A_d) = \dfrac{|A_d|}{|\Omega|} = \dfrac{\frac{n}{d}}{n} = \dfrac{1}{d}$.
2. (a) Si $p_1, \dots, p_\ell$ sont des diviseurs premiers deux à deux distincts de n, comme ils sont premiers, $\bigcap_{j=1}^{\ell} A_{p_j} = A_{p_1 \cdots p_\ell}$.
$$\mathbb{P}\left(\bigcap_{j=1}^{\ell} A_{p_j}\right) = \mathbb{P}(A_{p_1 \cdots p_\ell}) = \frac{1}{p_1 \cdots p_\ell} = \prod_{j=1}^{\ell} \mathbb{P}(A_{p_j}).$$
   (b) Les $\overline{A_p}$ sont aussi indépendant, $A = \bigcap_{p \in P} \overline{A_p}$, $\mathbb{P}(A) = \dfrac{\varphi(n)}{n} = \prod_{p \in P}\left(1 - \dfrac{1}{p}\right)$.

Exercice 7

Si $A_1, \dots, A_n$ sont indépendants, $1 \leqslant p \leqslant n-1$, Montrer que les événements suivants sont indépendants :

- $\bigcap_{i=1}^{p} A_i$ **et** $\bigcap_{i=p+1}^{n} A_i$,
- $\bigcup_{i=1}^{p} A_i$ **et** $\bigcap_{i=p+1}^{n} A_i$.
- $\bigcup_{i=1}^{p} A_i$ **et** $\bigcup_{i=p+1}^{n} A_i$,

- Direct,
- $A_1, \dots, A_p, \overline{A_{p+1}}, \dots, \overline{A_n}$ sont indépendants, par le premier point, sont indépendants $\bigcap_{i=1}^{p} A_i$ et $\bigcap_{i=p+1}^{n} \overline{A_i} = \overline{\bigcup_{i=p+1}^{n} A_i}$ donc sont indépendants $\bigcap_{i=1}^{p} A_i$ et $\bigcup_{i=p+1}^{n} A_i$.
- idem avec $\overline{A_1}, \dots, \overline{A_p}, \overline{A_{p+1}}, \dots, \overline{A_n}$.

IV Variables aléatoires discrètes

On se donne une espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

1 Définition

Définition 10 : Variable aléatoire discrète

Soit E un ensemble quelconque. Une application $X : \Omega \to E$ est appelée **variable aléatoire discrète** sur $(\Omega, \mathscr{A}, \mathbb{P})$ lorsqu'elle vérifie

(i) $X(\Omega) = \operatorname{Im} X = \{X(\omega), \omega \in \Omega\} \in \mathscr{P}(E)$ est fini ou dénombrable.

(ii) Pour tout $x \in X(\Omega)$, $X^{-1}(\{x\}) = \{\omega \in \Omega, X(\omega) = x\} \in \mathscr{A}$ et est noté $(X = x)$.

Elle est dite **réelle** lorsque $E \subset \mathbb{R}$.

Remarque

R 24 – La notation $(X = x)$ est un peu déroutante, cela revient par exemple à noter $\pi\mathbb{Z} = \sin^{-1}(\{0\}) = (\sin = 0)$.
Si A est une partie de E, on note $(X \in A)$ l'événement $X^{-1}(A) = \{\omega \in \Omega, X(\omega) \in A\}$.

R 25 – On note aussi, pour une variable aléatoire réelle,
$$(X \leqslant x) = X^{-1}(]-\infty, x]) = \{\omega \in \Omega,\ X(\omega) \leqslant x\}$$
et on introduit de la même façon, $(X < x)$, $(X \geqslant x)$, $(X > x)$.

R 26 – Enfin, si f est une fonction définie sur $X(\Omega)$, on note $f(X)$ la fonction $f \circ X$. Est-ce une variable aléatoire ? Oui. Voir ci-après.

R 27 – La deuxième condition est là pour qu'on puisse calculer la probabilité $\mathbb{P}(X = x)$ pour tout $x \in X(\Omega)$.
Si $x \in E \setminus X(\Omega)$, $(X = x) = \varnothing \in \mathscr{A}$ également.

R 28 – On ne demande pas que E soit fini ou dénombrable, seulement que $X(\Omega)$ le soit : si des valeurs de E ne sont pas atteintes, on peut s'en débarrasser.

On ne demande pas non plus que l'univers Ω soit fini ou dénombrable.

R 29 – Lorsque l'univers est fini ou dénombrable, on choisit $\mathscr{A} = \mathscr{P}(\Omega)$ et toute fonction de Ω dans E est une variable aléatoire discrète.

Exemple : fondamental

E 8 – Si F événement de l'univers Ω, alors

$$\mathbb{1}_F : \left| \begin{array}{ccc} \Omega & \longrightarrow & \{0,1\} \\ \omega & \longmapsto & \begin{cases} 1 & \text{si } \omega \in F \\ 0 & \text{si } \omega \notin F \end{cases} \end{array} \right.$$

est une variable aléatoire.
$\mathbb{P}(\mathbb{1}_F = 1) = \mathbb{P}(F)$ et $\mathbb{P}(\mathbb{1}_F = 0) = \mathbb{P}\left(\overline{F}\right)$.

Propriété 21 : SCE associé à une variable aléatoire

*$\left((X = x)\right)_{x \in X(\Omega)}$ est un système complet d'événements appelé **système complet d'événements associé à** X.*

Remarque

R 30 – On peut remplacer $X(\Omega)$ par E, ce qui revient à ajouter des ensembles vides.

Démonstration

Les $(X = x)$ sont deux à deux disjoints de réunion Ω. ■

Propriété 22 : Les parties de $X(\Omega)$ sont des événements

Soit X une variable aléatoire discrète sur $(\Omega, \mathscr{A}, \mathbb{P})$. Alors pour toute partie A de $X(\Omega)$, $(X \in A) \in \mathscr{A}$.

Démonstration

A est finie ou dénombrable et $A = \bigsqcup_{x \in A} \{x\}$ donc $(X \in A) = \bigsqcup_{x \in A} (X = x) \in \mathscr{A}$. ■

Propriété 23 : Une fonction d'une v.a.d. est une v.a.d.

Si $X : \Omega \to E$ est une variable aléatoire discrète, si $f : E \to F$ est une fonction (ou application) quelconque, alors $f \circ X$, notée $f(X)$ est une variable aléatoire discrète.

Démonstration

On a bien que $f(X) : \Omega \to F$ avec $f(X)(\Omega) = f(X(\Omega))$ fini ou dénombrable et si $y \in f(X)(\Omega)$,

$$(f(X) = y) = \{\omega \in \Omega,\ f(X(\omega)) = y\} = \left\{\omega \in \Omega,\ X(\omega) \in f^{-1}(\{y\})\right\} = \left(X \in f^{-1}(\{y\})\right) \in \mathscr{A}.$$ ■

2 Loi

On fixe X une variable aléatoire discrète sur $(\Omega, \mathscr{A}, \mathbb{P})$.

Définition 11 : Loi d'une v.a.d.

L'application

$$\mathbb{P}_X : \left| \begin{array}{ccc} \mathscr{P}(X(\Omega)) & \longrightarrow & \mathbb{R} \\ A & \longmapsto & \mathbb{P}(X \in A) \end{array} \right.$$

est appelée **loi de** X.

Propriété 24 : La loi est une probabilité

$\mathbb{P}_X$ est une probabilité sur l'espace probabilisable $\big(X(\Omega), \mathscr{P}(X(\Omega))\big)$.

Remarque

R31 – Comme $X(\Omega)$ est au plus dénombrable, il n'est pas choquant de choisir $\mathscr{P}(X(\Omega))$ comme tribu. $(X(\Omega), \mathscr{P}(X(\Omega)), \mathbb{P}_X)$ est l'espace probabilisé associé à X.

Démonstration

On revient à la définition :

- Pour tout $A \in \mathscr{P}(X(\Omega))$, $\mathbb{P}_X(A) = \mathbb{P}(X \in A) \in [0,1]$
- $\mathbb{P}_X(X(\Omega)) = \mathbb{P}(\Omega) = 1$.
- **σ-additivité** : Si $(A_n)_n$ est suite de parties deux à deux disjointes de $X(\Omega)$,

$$\mathbb{P}_X\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}\left(\bigsqcup_{n\in\mathbb{N}} (X \in A_n)\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(X \in A_n) = \sum_{n\in\mathbb{N}} \mathbb{P}_X(A_n)$$

■

Propriété 25 : Expression de la loi de X

Si $A \in \mathscr{P}(X(\Omega))$,

$$\mathbb{P}_X(A) = \sum_{a\in A} \mathbb{P}_X(\{a\}) = \sum_{a\in A} \mathbb{P}(X = a).$$

Démonstration

Il suffit de décomposer dans le s.c.e adapté à X. ■

Corollaire 3

La loi de X est uniquement déterminée par la distribution de probabilités $(\mathbb{P}(X = x))_{x\in X(\Omega)}$.

Remarque

R32 – Ainsi, pour décrire la loi d'une variable aléatoire, on se contente de préciser $X(\Omega)$ et les $\mathbb{P}(X = x)$ pour $x \in X(\Omega)$. On verra plus loin les lois usuelles à connaître parfaitement.

Notation 1 : $\sim$

Si X et Y suivent la même loi, on note $X \sim Y$.
Si X suit une loi $\mathscr{L}$, on note $X \sim \mathscr{L}$.

Exercice 8 : CCINP 109

CC BY NC SA

Exercice 9 : CCINP 104

Propriété 26 : Loi de $f(X)$

La loi de $Y=f(X)$ est donnée par $\forall\, y\in f(X(\Omega))$,

$$\mathbb{P}(Y=y)=\mathbb{P}(f(X)=y)=\mathbb{P}\left(X\in f^{-1}(\{y\})\right)=\sum_{x\ \mid\ f(x)=y}\mathbb{P}(X=x).$$

De la même manière, on obtient par exemple :

Propriété 27 : Loi d'une somme, d'un produit

Si X et Y sont des variables aléatoires,

$$\mathbb{P}(X+Y=z)=\sum_{x,y\ \mid\ x+y=z}\mathbb{P}\big(X=x,Y=y\big)$$

$$\text{et}\quad \mathbb{P}(XY=z)=\sum_{x,y\ \mid\ xy=z}\mathbb{P}\big(X=x,Y=y\big)$$

V Familles de variables aléatoires

Soit $(\Omega,\mathscr{A},\mathbb{P})$ espace probabilisé.

1 Définition et lois

a Couple de variables aléatoires discrètes

Les notions vues en première année se généralise sans problème particulier.

Définition – Propriété 1

Soit X,Y variables aléatoires discrètes sur Ω à valeurs dans E,E'. L'application

$$Z:\left|\begin{array}{ccc}\Omega & \longrightarrow & E\times E'\\ \omega & \longmapsto & (X(\omega),Y(\omega))\end{array}\right.$$

est une variable aléatoire discrète appelée **couple** $Z=(X,Y)$.

Remarque

R 33 – $(X,Y)(\Omega)\subset X(\Omega)\times Y(\Omega)$ et il n'y a pas égalité en général.

R 34 – On note indifféremment $((X,Y)=(x,y))$ ou $(X=x)\cap(Y=y)$ ou $(X=x \text{ et } X=y)$ ou $(X=x,Y=y)$ ces événements.

Démonstration

$Z(\Omega)\subset X(\Omega)\times Y(\Omega)$ est fini ou dénombrable.
Pour tout $(x,y)\in Z(\Omega)$, $(Z=(x,y))=(X=x)\cap(Y=y)\in\mathscr{A}$. ■

Propriété 28 : SCE associé à un couple

Soit (X,Y) un couple de variables aléatoires discrètes. Alors la famille d'événements $\big(((X,Y)=(x,y))\big)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ est un système complet d'événements appelé ***système complet d'événements associé au couple*** *(X,Y).*

Démonstration

Les événements sont bien disjoints deux à deux et

$$\bigcup_{(x,y)\in X(\Omega)\times Y(\Omega)} \big((X=x)\cap(Y=y)\big)=\Omega$$

car tout $\omega\in(X=X(\omega))\cap(Y=Y(\omega))$. ■

Remarque

R 35 – On sait donc que

- Si X et Y sont deux variables aléatoires discrètes définies sur un espace probabilisé $(\Omega,\mathscr{A},P)$, le « couple »

$$(X,Y)\ :\ \omega\longmapsto(X(\omega),Y(\omega))$$

  est une variable aléatoire discrète.

- Si $Z\ :\ \Omega\to E$ est une variable aléatoire discrète, si $f\ :\ E\to F$ est une fonction quelconque, alors $f\circ Z$, notée $f(Z)$ est une variable aléatoire discrète.

Et constatons que donc, si X et Y sont des variables aléatoires discrètes réelles définies sur un même univers probabilisé, alors $X+Y$, XY, $\min(X,Y)$, $\max(X,Y)$ sont des variables aléatoires réelles discrètes.
Bien sûr, il y aussi $\Gamma\left(\mathrm{Arctan}\left(1+X^2+Y^2\right)\right)$, mais on n'a cité que quelques exemples fréquemment utiles.

Pour calculer les lois :

$$\mathbb{P}(X+Y=z)=\sum_{x,y\ |\ x+y=z}\mathbb{P}\big(X=x,Y=y\big)$$
$$\text{et}\quad \mathbb{P}(XY=z)=\sum_{x,y\ |\ xy=z}\mathbb{P}\big(X=x,Y=y\big)$$

b Loi conjointe

Définition 12 : Loi conjointe

Soit (X,Y) un couple de variable aléatoires discrètes. On appelle **loi conjointe** de (X,Y) la loi $\mathbb{P}_{(X,Y)}$ de la variable aléatoire (X,Y).

Remarque

R 36 – Vu la propriété précédente, cette loi est déterminée par $\mathbb{P}(X=x,Y=y)$ pour $(x,y)\in X(\Omega)\times Y(\Omega)$. Lorsque les variables aléatoires sont finies, cette loi peut être représentée dans un tableau à double entrée.

Exemple

E 9 – On lance deux dés, X est la v.a. égale au plus grand des nombres, Y celle du plus petit. On pose $\Omega=[\![1,6]\!]^2$ muni de la probabilité uniforme. On obtient :

X \ Y	1	2	3	4	5	6	loi de X
1	1/36	0	0	0	0	0	1/36
2	1/18	1/36	0	0	0	0	1/12
3	1/18	1/18	1/36	0	0	0	5/36
4	1/18	1/18	1/18	1/36	0	0	7/36
5	1/18	1/18	1/18	1/18	1/36	0	9/36
6	1/18	1/18	1/18	1/18	1/18	1/36	11/36
loi de Y	11/36	9/12	7/36	5/36	3/36	1/36	(1)

Remarquons qu'on obtient la loi de X en sommant les lignes et celle de Y en sommant les colonnes.

c Lois marginales

Définition 13 : Lois marginales

Si (X, Y) est un couple de variables aléatoires discrètes, les lois de X et de Y sont appelées **première et seconde lois marginales du couple.**

Propriété 29 : Loi conjointe détermine lois marginales

La loi conjointe de (X, Y) détermine les lois marginales de (X, Y) mais la réciproque est fausse.

Démonstration

$\mathbb{P}(X = x) = \mathbb{P}\left(\bigsqcup_y ((X, Y) = (x, y))\right) = \sum_y \mathbb{P}(X = x, Y = y)$. Idem pour Y.

Contre-exemple :

X \ Y	0	1	loi de X
0	1/2	0	1/2
1	0	1/2	1/2
loi de Y	1/2	1/2	(1)

X \ Y	0	1	loi de X
0	1/4	1/4	1/2
1	1/4	1/4	1/2
loi de Y	1/2	1/2	(1)

■

d Lois conditionnelles

Définition 14 : Loi conditionnelle

Soit (X, Y) un couple de variables aléatoires discrètes. Pour tout $x \in X(\Omega)$ tel que $\mathbb{P}(X = x) \neq 0$, la **loi conditionnelle de Y sachant** $(X = x)$ est la loi de Y pour la probabilité conditionnelle $\mathbb{P}_{(X=x)}$.

Elle est donc déterminée par, pour tout $y \in Y(\Omega)$,

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}.$$

Remarque

R 37 – Les lois conditionnelles de Y sachant $(X = x)$ et la loi de X permettent de déterminer la loi conjointe de (X, Y) :

- Soit $\mathbb{P}(X = x) = 0$ et alors $\mathbb{P}(X = x, Y = y) \leqslant \mathbb{P}(X = x) = 0$ donc $\mathbb{P}(X = x, Y = y) = 0$,
- soit $\mathbb{P}(X = x) \neq 0$ et

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y \mid X = x)\mathbb{P}(X = x).$$

Exemple

E 10 – Lois conditionnelles de Y :

X \ Y	0	1
0	1/4	3/4
1	2/3	1/3

Loi conjointe de (X, Y) :

X \ Y	0	1	loi de X
0	1/10	3/10	2/5
1	2/5	1/5	3/5
loi de Y	1/2	1/2	(1)

2 Extension aux n-uplets

Définition 15 : n-uplets de variables aléatoires

Soit $(X_1, \ldots, X_n)$ un n-uplet de variables aléatoires discrètes. C'est encore une variable aléatoire discrète appelé **vecteur aléatoire discret** de dimension n.

La **loi conjointe** de $(X_1, \ldots, X_n)$ est déterminée par les $\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$ où pour tout i, $x_i \in X_i(\Omega)$.

Les lois de $X_1, \ldots, X_n$ sont les **lois marginales** de $(X_1, \ldots, X_n)$.

Définition 16 : Loi conditionnelle pour n variables

Si $x_1, \ldots, x_{n-1}$ sont fixés, tel que $\mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) > 0$, la **loi conditionnelle** de X_n sachant $(X_1 = x_1, \ldots, X_{n-1} = x_{n-1})$ est déterminée par

$$\mathbb{P}(X_n = x_n \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \frac{\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)}{\mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1})}$$

pour tout x_n.

Remarque

R 38 – Lorsque l'on a la propriété

$$\mathbb{P}\left(X_{i+1} = x_{i+1} \mid X_1 = x_1, \ldots, X_i = x_i\right) = \mathbb{P}(X_{i+1} = x_{i+1} \mid X_i = x_i)$$

(phénomène sans mémoire), on dit que la famille $(X_1, \ldots X_n)$ de variables aléatoires est **markovienne**.

3 Indépendance

a Cas d'un couple de variable

Définition 17 : Indépendance

Soient X, Y deux variables aléatoires discrètes sur l'espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

X et Y sont dites **indépendantes** si pour tout $(A, B) \in \mathscr{P}(X(\Omega)) \times \mathscr{P}(Y(\Omega))$, les événements $(X \in A)$ et $(Y \in B)$ sont indépendants, c'est-à-dire

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\ \mathbb{P}(Y \in B).$$

On note parfois $X \perp\!\!\!\perp Y$.

Propriété 30 : Caractérisation par des événements élémentaires

X et Y sont indépendantes si et seulement si pour tout $(x,y)\in X(\Omega)\times Y(\Omega)$, $(X=x)$ et $(Y=y)$ sont indépendants, c'est-à-dire

$$\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\ \mathbb{P}(Y=y).$$

Démonstration

Le sens $\Rightarrow$ est direct.
On suppose que pour tout $(x,y)\in X(\Omega)\times Y(\Omega)$, $\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\mathbb{P}(Y=y)$.
Soit $(A,B)\in\mathscr{P}(X(\Omega))\times\mathscr{P}(Y(\Omega))$.

$$\mathbb{P}(X\in A,Y\in B)=\mathbb{P}\left(\bigsqcup_{(x,y)\in A\times B}(X=x,Y=y)\right)=\sum_{(x,y)\in A\times B}\mathbb{P}(X=x,Y=y)=\sum_{x\in A}\mathbb{P}(X=x)\sum_{y\in B}\mathbb{P}(Y=y).$$

■

Remarque

R 39 – Si X et Y sont indépendantes, la donnée des lois marginales de (X,Y) détermine sa loi conjointe.

Propriété 31 : Caractérisation par les lois conditionnelles

Soit (X,Y) couple de variables aléatoires. Il y a équivalence entre

(i) *Les variables aléatoires X et Y sont indépendantes.*

(ii) *Pour tout $y\in Y(\Omega)$ tel que $\mathbb{P}(Y=y)>0$, la loi de X sachant $(Y=y)$ est la même que la loi de X.*

(iii) *Pour tout $x\in X(\Omega)$ tel que $\mathbb{P}(X=x)>0$, la loi de Y sachant $(X=x)$ est la même que la loi de Y.*

Démonstration

Immédiat. ■

Propriété 32 : Fonctions de variables aléatoires indépendantes

Si X,Y sont des variables aléatoires indépendantes, f,g définies sur $X(\Omega)$ et $Y(\Omega)$ respectivement, alors $f(X)$ et $g(Y)$ sont indépendantes.

Démonstration

$(f(X)\in A)=\left(X\in f^{-1}(A)\right)$ et $(g(Y)\in B)=\left(Y\in f^{-1}(B)\right)$, ces derniers étant indépendants. ■

Exemple

E 11 – Si X et Y sont indépendantes, pour tous $m,n\in\mathbb{N}$, X^m et Y^n le sont.

Remarque

R 40 – En reprenant un calcul précédent, on obtient, si X,Y indépendantes,

$$\mathbb{P}(X+Y=z)=\sum_{x,y\ |\ x+y=z}\mathbb{P}(X=x)\ \mathbb{P}(Y=y)$$

où l'on peut remplacer $X+Y$ (et $x+y$) par n'importe quelle fonction de X et Y.

b Variables aléatoires indépendantes

Définition 18 : Variables aléatoires indépendantes

Des variables aléatoires discrètes $X_1, \dots, X_n$ sont dites **indépendantes** lorsque pour toutes parties A_1 de $X_1(\Omega)$, …, A_n de $X_n(\Omega)$, les événements $(X_1 \in A_1)$, …, $(X_n \in A_n)$ le sont.

Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires discrètes est dite une **suite de variables aléatoire indépendantes** lorsque pour tout $n \in \mathbb{N}$, $X_1, \dots, X_n$ le sont.

Si, de plus, elles ont même loi, on dit que ce sont des **variables aléatoires indépendantes identiquement distribuées** (vaiid).

Propriété 33 : Caractérisation par des événements élémentaires

$X_1, \dots, X_n$ sont indépendantes si et seulement si pour tout $(x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)$, les événements $(X_1 = x_1)$, …, $(X_n = x_n)$ le sont.

Démonstration

Le sens $\Rightarrow$ est direct.

L'autre sens est similaire au cas des couples de variables aléatoires : on suppose que pour tout $(x_1, \dots, x_n) \in A_1 \times \dots \times A_n$, les événements $(X_1 = x_1)$, …, $(X_n = x_n)$ sont indépendants.

Soit $(A_1, \dots, A_n) \in \mathscr{P}(X_1(\Omega)) \times \dots \times \mathscr{P}(X_n(\Omega))$, $I \subset [\![1, n]\!]$,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{i\in I}(X_i \in A_i)\right) &= \mathbb{P}\left(\bigsqcup_{(x_i)_i \in \prod\limits_{i\in I} A_i}\left(\bigcap_{i\in I}(X_i = x_i)\right)\right) \\ &= \sum_{(x_i)_i \in \prod\limits_{i\in I} A_i} \mathbb{P}\left(\bigcap_{i\in I}(X_i = x_i)\right) = \sum_{(x_i)_i \in \prod\limits_{i\in I} A_i}\left(\prod_{i\in I}\mathbb{P}(X_i = x_i)\right) \\ &= \prod_{i\in I}\left(\sum_{x_i \in A_i}\mathbb{P}(X_i = x_i)\right) = \prod_{i\in I}\mathbb{P}(X_i \in A_i).\end{aligned}$$ ■

Remarque

R 41 – n expériences aléatoires indépendantes peuvent être modélisées par n variables aléatoires indépendantes. Le résultat de la i^{e} expérience est noté X_i et

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n).$$

R 42 – Comme pour les événements, indépendants $\Rightarrow$ indépendants deux à deux, mais la réciproque est fausse si $n > 2$.

Exemple

E 12 – Si X_1, X_2 vaiid finies de loi uniforme $\mathscr{U}(2)$ sur $\{-1, 1\}$. $X_3 = X_1 \times X_2$.

$$\mathbb{P}(X_3 = -1) = \mathbb{P}(X_1 = 1, X_2 = -1) + \mathbb{P}(X_1 = -1, X_2 = 1) = \frac{1}{2}$$

Donc $X_3 \sim \mathscr{U}(2)$ sur $\{-1, 1\}$.

Alors X_1, X_2, X_3 sont indépendantes deux à deux car $X_1 \perp\!\!\!\perp X_2$,

$$\mathbb{P}(X_1 = 1, X_3 = 1) = \mathbb{P}(X_1 = 1, X_2 = 1) = \frac{1}{4} = \mathbb{P}(X_1 = 1)\mathbb{P}(X_3 = 1)$$

$$\mathbb{P}(X_1 = -1, X_3 = 1) = \mathbb{P}(X_1 = -1, X_2 = -1) = \frac{1}{4} = \mathbb{P}(X_1 = -1)\mathbb{P}(X_3 = 1)$$

$$\mathbb{P}(X_1 = 1, X_3 = -1) = \mathbb{P}(X_1 = 1, X_2 = -1) = \frac{1}{4} = \mathbb{P}(X_1 = 1)\mathbb{P}(X_3 = -1)$$

$$\mathbb{P}(X_1=-1, X_3=-1) = \mathbb{P}(X_1=-1, X_2=1) = \frac{1}{4} = \mathbb{P}(X_1=-1)\mathbb{P}(X_3=1)$$

Donc $X_1 \perp\!\!\!\perp X_3$ et par symétrie, $X_2 \perp\!\!\!\perp X_3$.
Pourtant, elles ne sont pas indépendantes :

$$\mathbb{P}(X_1=1, X_2=1, X_3=1) = \mathbb{P}(X_1=1, X_2=1) = \frac{1}{4} \neq \frac{1}{8}.$$

Propriété 34 : Fonctions de variables aléatoires indépendantes

Si $(X_n)_{n\in\mathbb{N}}$ est une famille de variables aléatoires indépendantes, pour tout $n\in\mathbb{N}$, f_n définie $X_n(\Omega)$, alors $\big(f_n(X_n)\big)_{n\in\mathbb{N}}$ est une famille de variables aléatoires indépendantes.

Démonstration

Similaire au cas de deux variables : $\big(f(X_i)\in B_i\big) = \big(X_i \in f^{-1}(B_i)\big)$. ■

Propriété 35 : Lemme des coalitions

Soit $n, m \in \mathbb{N}$ tels que $0<m<n$, $X_1,\dots,X_m,\dots,X_n$ des variables aléatoires discrètes indépendantes sur $(\Omega,\mathscr{A},\mathbb{P})$, f définie sur $X_1(\Omega)\times\cdots\times X_m(\Omega)$ et g définie sur $X_{m+1}(\Omega)\times\cdots\times X_n(\Omega)$.
Alors $f(X_1,\dots,X_m)$ et $g(X_{m+1},\dots,X_n)$ sont indépendantes.
Le résultat s'étend à plus de deux coalitions.

Démonstration

Il suffit de remarquer que les variables aléatoires $Y=(X_1,\dots,X_m)$ et $Z=(X_{m+1},\dots,X_n)$ sont indépendantes (ce qui ne pose pas vraiment de problème : il suffit de l'écrire) et d'appliquer la propriété précédente. ■

Théorème 1

Soit $(\mathscr{L}_n)_{n\in\mathbb{N}}$ une suite de lois de probabilités discrètes.
Il existe un espace probabilisé $(\Omega,\mathscr{A},\mathbb{P})$ et une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires discrètes indépendantes sur $(\Omega,\mathscr{A},\mathbb{P})$ tels que pour tout $n\in\mathbb{N}$, $X_n \sim \mathscr{L}_n$.

Exemple

E 13 – Un jeu de pile ou face infini se modélise (naturellement) par une suite d'épreuves de Bernoulli indépendantes.

Démonstration : Admis

■

Exemple

E 14 – Reprenons l'exemple **E7** et montrons que l'événement A « la pièce donne Pile un nombre fini de fois » est négligeable.

Ce n'est pas très simple à démontrer. Cela suppose d'avoir un espace probabilisé $(\Omega,\mathscr{A},\mathbb{P})$ modélisant cet infinité de lancers indépendants : c'est l'objet du théorème précédent.
Notons $(X_i)_{i\in\mathbb{N}^*}$ la suite de vaiid de loi $\mathscr{B}(p)$ (où $0<p<1$) correspondant au résultat de chaque lancer $X_i=1$ pour Pile et 0 pour Face).

Notons A_k l'événement « N'obtenir que des Faces à partir du k^e lancer. » Alors $A=\bigcup\limits_{k=1}^{+\infty} A_k$ et nous allons montrer que les A_k sont tous négligeables.

Comment calculer $\mathbb{P}(A_k)$? Comme on ne regarde pas les premiers résultats, il n'est pas difficile de se convaincre que cette probabilité sera constamment égale à $\mathbb{P}(A_1)$, la probabilité de n'obtenir que des Faces.

Notons $B_{k,n}$ l'événement « Obtenir des Faces du k^e au n^e lancer. ». À k fixé, la suite $(B_{k,n})_{n\geqslant k}$ est décroissante.

Alors $\mathbb{P}(A_k) = \mathbb{P}\left(\bigcap_{n=k}^{+\infty} B_{k,n}\right)$ est, par continuité décroissante cette fois, la limite de $\mathbb{P}(B_{k,n})$ lorsque $n \to +\infty$.

Or $\mathbb{P}(B_{k,n}) = \mathbb{P}\left(\bigcap_{i=k}^{n} (X_i = 0)\right) = (1-q)^{n-k+1} \xrightarrow[n\to+\infty]{} 0$.

Donc les A_k sont tous négligeables et donc $A = \bigcup_{k=1}^{+\infty} A_k = \bigcup_{k=1}^{+\infty}\left(\bigcap_{n=k}^{+\infty} B_{k,n}\right)$ l'est aussi.

VI Lois usuelles

1 Loi Uniforme

Définition 19 : Loi uniforme

On dit que qu'une variable aléatoire **finie** X suit une **loi uniforme** lorsque pour tout $x \in X(\Omega)$,

$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = \frac{1}{n}$$

où $n = |X(\Omega)|$, c'est-à-dire que pour tout $A \subset X(\Omega)$, $\mathbb{P}_X(A) = \dfrac{|A|}{n}$.

On note alors $X \sim \mathscr{U}(n)$.

Exemple

E 15 – Si on tire un dé équilibré à n faces ou si on tire une boule dans une urne qui en contient n (numérotée), alors la variable aléatoire du résultat suit $\mathscr{U}(n)$.

Remarque

R 43 – ⚠ cela ne concerne pas de la probabilité $\mathbb{P}$ initiale : $\mathbb{P}_X$ peut être uniforme sans que $\mathbb{P}$ le soit.
Si, par exemple, on lance un dé à 6 faces truqué tel que l'on obtient 1 ou 6 avec une probabilité $1/4$ et $2,3,4$ ou 5 avec probabilité $1/8$, X variable aléatoire $\mathbb{1}_{2\mathbb{N}}$, alors $\mathbb{P}(X=0) = \mathbb{P}(X=1) = 1/2$ donc $X \sim \mathscr{U}(2)$ alors que $\mathbb{P}$ n'est pas la probabilité uniforme.

2 Loi de Bernoulli

Définition 20 : Loi de Bernoulli

On dit que X suit une **loi de Bernoulli de paramètre** $p \in [0,1]$ lorsque X est à valeurs dans $E = \{0,1\}$, $\mathbb{P}(X=1) = p$ et $\mathbb{P}(X=0) = q = 1-p$.

On note alors $X \sim \mathscr{B}(p)$.

Exemple : Situation type

E 16 – Variable aléatoire étudiant le succès (1) d'un événement donné ou son échec (0).

Propriété 36 : Ce sont les fonctions indicatrices

Les variables aléatoires qui suivent une loi de Bernoulli de paramètre p sont exactement les fonctions indicatrices des parties F de Ω telles que $\mathbb{P}(F) = p$.

Démonstration

Si X suit une loi de Bernoulli de paramètre p, alors si $F = (X = 1)$, $X = \mathbb{1}_F$.
Réciproquement, si $X = \mathbb{1}_F$, $X(\Omega) \subset \{0, 1\}$ et $\mathbb{P}(X = 1) = \mathbb{P}(F)$. ∎

Remarque

R 44 – Deux variables aléatoires de Bernoulli sont indépendantes si et seulement si $(X = 1)$ et $(Y = 1)$ le sont.

R 45 – Si X prend deux valeurs a et b distinctes, alors $Y = \dfrac{X - a}{b - a}$ suit une loi de Bernoulli de paramètre $p = \mathbb{P}(X = b)$.
Autrement dit, $X = a + (b - a)Y$ où Y suit une loi de Bernoulli.

3 Loi binomiale

Lors de la répétition de n expériences de Bernoulli indépendantes, la probabilité d'avoir $k \leqslant n$ succès s'écrit $\binom{n}{k} p^k (1-p)^{n-k}$ où p est la probabilité d'un succès.

Si on appelle X la variable aléatoire du nombre de succès, à valeurs dans $[\![0, n]\!]$, alors elle suit la loi donnée par

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Remarquons que l'on peut écrire $X = X_1 + \cdots + X_n$ où X_i est la variable aléatoire de Bernoulli succès à la i^{e} répétition.

Définition 21 : Loi binomiale

On dit que X suit une **loi binomiale de paramètre** (n, p) **où** $p \in [0, 1]$ lorsque X est à valeurs dans $[\![0, n]\!]$ et pour tout $k \in [\![0, n]\!]$,

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \binom{n}{k} p^k q^{n-k}$$

avec $q = 1 - p$. On note alors $X \sim \mathscr{B}(n, p)$.

Exemple : Situation type

E 17 – Nombre de succès dans la répétition de n expériences de Bernoulli indépendantes.

Remarque

R 46 – $\mathscr{B}(1, p) = \mathscr{B}(p)$.

R 47 – La formule du binôme redonne (ou se retrouve par) $\sum_{k=0}^{n} \mathbb{P}(X = k) = 1$.

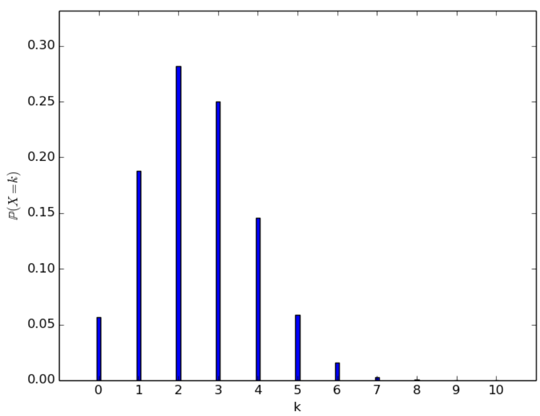

Figure 1 – Loi $\mathscr{B}(10, 1/4)$

Exercice 10

Si $X \sim \mathscr{B}(n, p)$ alors $Y = n - X \sim \mathscr{B}(n, q)$.

4 Loi géométrique

Soit $p \in]0,1[$.

On lance une infinité de fois une pièce donnant pile avec probabilité p. Les lancers sont indépendants.

Soit X_n la variable aléatoire du succès au n^e lancer : elle vaut 1 si c'est pile, et 0 sinon.

$(X_n)_{n\in\mathbb{N}^*}$ est une suite de vadiid, toutes de loi $\mathscr{B}(p)$.

Soit X la variable aléatoire du rang du premier succès : pour tout $\omega \in \Omega$, $X(\omega) = \min\left\{k \in \mathbb{N}^*,\ X_k(\omega) = 1\right\} \in \mathbb{N}^* \cup \{+\infty\}$.

Soit $n \in \mathbb{N}^*$.

- $(X > n) = \bigcap_{k=1}^{n} (X_k = 0)$, donc, par indépendance,
$$\mathbb{P}(X > n) = (1-p)^n.$$
- $(X = n) = (X_n = 1) \cap \bigcap_{k=1}^{n-1} (X_k = 0)$, donc, par indépendance,
$$\mathbb{P}(X = n) = p(1-p)^{n-1}.$$
- En passant au contraire,
$$\mathbb{P}(X \leqslant n) = 1 - (1-p)^n.$$
- Soit A l'événement « N'obtenir que des faces », et $A_n = (X > n)$.
  Alors (A_n) est décroissante et $A = \bigcap_{n\in\mathbb{N}^*} A_n$.
  Par continuité décroissante,
$$\mathbb{P}(A_n) = (1-p)^n \xrightarrow[n\to+\infty]{} \mathbb{P}(A) = 0.$$

Définition 22 : Loi géométrique

Soit $p \in]0,1[$ et X une variable aléatoire discrète. On dit que X suit une **loi géométrique** de paramètre p si X est à valeurs $\boxed{\text{dans } \mathbb{N}^*}$ et
$$\forall\, n \in \mathbb{N}^*,\quad \mathbb{P}(X = n) = p(1-p)^{n-1}.$$

On note $X \sim \mathscr{G}(p)$.

Remarque

R 48 – Première loi dénombrable du programme. On vérifie bien $\sum_{n=1}^{+\infty} \mathbb{P}(X = n) = 1$.

R 49 – De nouveau, on calcule (à savoir faire !)
$$\mathbb{P}(X > n) = \mathbb{P}\left(\bigsqcup_{k=n+1}^{+\infty} (X = k)\right) = \cdots = (1-p)^n.$$

Exemple : Situation type

E 18 – Le rang du premier succès dans une répétition infinie d'épreuves de Bernoulli indépendantes de paramètre p suit $\mathscr{G}(p)$.

5 Loi de Poisson

Définition 23 : Loi de Poisson

Soit $\lambda \in \mathbb{R}_+^*$ et X une variable aléatoire discrète. On dit que X suit une **loi de Poisson de paramètre** λ si X est à valeurs $\boxed{\text{dans } \mathbb{N}}$ et

$$\forall\, n \in \mathbb{N},\quad \mathbb{P}(X = n) = \frac{\lambda^n}{n!}\mathrm{e}^{-\lambda}.$$

On note $X \sim \mathscr{P}(\lambda)$.

Remarque

R 50 – On vérifie bien $\sum_{n=0}^{+\infty} \mathbb{P}(X = n) = 1$.

6 Propriétés des lois usuelles

a Somme de n vaiid de Bernoulli

Propriété 37 : Importante !

Si $X_1, \ldots, X_n$ vaiid de loi $\mathscr{B}(p)$, alors

$$X_1 + \ldots + X_n \sim \mathscr{B}(n, p).$$

Démonstration

$$\mathbb{P}(X_1 + \ldots + X_n = k) = \sum_{\substack{x_1,\ldots,x_n \in \{0,1\} \\ x_1+\cdots+x_n = k}} \mathbb{P}(X_1 = x_1)\cdots\mathbb{P}(X_n = x_n) = \binom{n}{k} p^k (1-p)^{n-k}.$$

■

Remarque

R 51 – Plus généralement, si les X_i indépendantes suivent $\mathscr{B}(n_i, p)$, alors $X_1 + \cdots + X_n \sim \mathscr{B}(\sum n_i, p)$.

b Approximation d'une loi de Poisson par des lois binomiales

Propriété 38 : Approximation d'une loi de Poisson par des lois binomiales

Soit $\lambda > 0$, $(p_n)_n \in]0,1[^{\mathbb{N}}$ tel que $np_n \to \lambda$, $(X_n)_n$ une suite de variables aléatoires discrètes réelles.
On suppose que pour tout $n \in \mathbb{N}$, $X_n \sim \mathscr{B}(n, p_n)$.
Alors, pour tout $k \in \mathbb{N}$, $\mathbb{P}(X_n = k) \xrightarrow[n\to+\infty]{} \dfrac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$.

Remarque

R 52 – Une loi binomiale $\mathscr{B}(n, p)$ (qui peut être vue comme nombre de succès dans la répétition de n épreuve de Bernoulli avec probabilité p de succès) peut donc être approchée par une loi de Poisson $\mathscr{P}(\lambda)$ où $\lambda = np$ à condition que n soit grand et $p = \dfrac{\lambda}{n}$ soit petit.
La loi de Poisson est qualifiée de **loi des événements rares**.

Démonstration

On vérifie que lorsque $n \to +\infty$, $\mathbb{P}(X_n = k) \sim \dfrac{(np_n)^k(1-p_n)^n}{k!(1-p_n)^k} \xrightarrow[n\to+\infty]{} \dfrac{\lambda^k e^{-\lambda}}{k!}$. ■

7 Exercices CCINP

Exercice 11 : CCINP 98

Exercice 12 : CCINP 95

VII ESPÉRANCE

On fixe ici un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.

1 Définition

Définition 24 : Espérance d'une variable aléatoire discrète réelle positive

Soit X une variable aléatoire discrète à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$.
L'**espérance de** X est, par définition, dans $[0, +\infty]$,

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x)x.$$

On a donc $\mathbb{E}(X) \in \mathbb{R} \cup \{+\infty\}$ suivant la sommabilité ou non de la famille réelle positive $(\mathbb{P}(X = x)x)_{x \in X(\Omega)}$.
Lorsque cette famille est sommable, X est dite d'**espérance finie** $\mathbb{E}(X) \in \mathbb{R}$ et on note $X \in L^1$.

Propriété 39 : Cas d'une variable aléatoire entière

Soit X une variable aléatoire à valeurs dans $\mathbb{N} \cup \{+\infty\}$. Alors

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n) = \sum_{n=0}^{+\infty} \mathbb{P}(X > n)$$

Démonstration

C'est de la sommation par paquet (Fubini) positif :

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} \mathbb{P}(X = k)k = \sum_{k=1}^{+\infty} \mathbb{P}(X = k)k = \sum_{k=1}^{+\infty} \sum_{n=1}^{k} \mathbb{P}(X = k) = \sum_{n=1}^{+\infty} \sum_{k=n}^{+\infty} \mathbb{P}(X = k) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n).$$

Autre démonstration (Écrit CCINP MP-MPI 2024) : Pour tout $n \in \mathbb{N}$,

$$\begin{aligned}
\sum_{k=1}^{n} kP(X = k) = \sum_{k=1}^{n} k\left(P(X > k-1) - P(X > k)\right) &= \sum_{k=1}^{n} \left(\underbrace{k}_{=1+k-1} P(X > k-1) - kP(X > k)\right). \\
&= \sum_{k=1}^{n} P(X > k-1) + \sum_{k=1}^{n} ((k-1)P(X > k-1) - kP(X > k)) \\
&= \sum_{k=0}^{n-1} P(X > k) + 0 - nP(X > n) \qquad \text{par télescopage}
\end{aligned}$$

On montre alors que $nP(X > n) \xrightarrow[n\to+\infty]{} 0$ si X est d'espérance finie. En effet, X est d'espérance finie et

$$nP(X > n) = \sum_{k=n+1}^{+\infty} nP(X = k) \leqslant \sum_{k=n+1}^{+\infty} kP(X = k) \xrightarrow[n\to+\infty]{} 0$$

en tant que reste d'une série convergente.

On en déduit de la relation précédente que, si X est d'espérance finie, la série $\sum P(X>n)$ converge et que $\mathbb{E}(X)=\sum_{k=0}^{+\infty} P(X>k)$.

Définition 25 : Espérance d'une variable aléatoire discrète réelle ou complexe

Soit X une variable aléatoire réelle ou complexe discrète.

Lorsque la famille $(\mathbb{P}(X=x)\ x)_{x\in X(\Omega)}$ est sommable, on dit que X **est d'espérance finie**, on note $X\in L^1$ et on définit son **espérance**

$$\mathbb{E}(X)=\sum_{x\in X(\Omega)} \mathbb{P}(X=x)x.$$

Dans le cas contraire, X n'a pas d'espérance (pas plus infinie que finie)

Lorsque X est d'espérance finie et $\mathbb{E}(X)=0$, X est dite **centrée**.

Remarque

R 53 – Une expression équivalente à « X est d'espérance finie » est « X a un moment d'ordre 1 ».

R 54 – Une variable aléatoire réelle finie est toujours d'espérance finie (programme de MP2I).

R 55 – Une variable aléatoire à valeurs dénombrables $(x_n)_{n\in\mathbb{N}}$ est d'espérance finie si et seulement si la série $\sum_{n\in\mathbb{N}} \mathbb{P}(X=x_n)x_n$ converge **absolument**, et alors $\mathbb{E}(X)=\sum_{n=0}^{+\infty} \mathbb{P}(X=x_n)x_n$.

R 56 – Ne pas confondre « centrée » et « symétrique » : si

$$P(X=1)=P(X=2)=P(X=3)=P(X=-6)=1/4,$$

la variable aléatoire X est bien centrée. Pour autant, X et $-X$ n'ont pas même loi.

2 Théorème de transfert

Théorème 2 : de transfert

Soit X un variable aléatoire discrète, f une fonction définie sur $X(\Omega)$, à valeurs complexes.

$f(X)$ est d'espérance finie si et seulement si la famille $\left(\mathbb{P}(X=x)\ f(x)\right)_{x\in X(\Omega)}$ est sommable, et dans ce cas

$$\mathbb{E}\left(f(X)\right)=\sum_{x\in X(\Omega)} \mathbb{P}(X=x)f(x).$$

Remarque

R 57 – Pas besoin de connaître la loi de $f(X)$!

Démonstration

Pas si difficile : il s'agit d'une application du théorème de sommation par paquets, en partitionnant $X(\Omega)$ par les $I_y=f^{-1}(\{y\})$ pour $y\in f(X)(\Omega)$ (et en ajoutant des valeurs absolues pour la sommabilité : $\mathbb{P}(f(X)=y)\left|y\right|=\sum_{x\in I_y}\mathbb{P}(X=x)\left|f(x)\right|$).

Cela donne l'équivalence et permet de calculer

$$\mathbb{E}(f(X))=\sum_{y\in f(X(\Omega))}\left(\sum_{\substack{x\in X(\Omega)\\ f(x)=y}}\mathbb{P}(X=x)\right)y=\sum_{\substack{(x,y)\in f(X(\Omega))\times X(\Omega)\\ f(x)=y}}\mathbb{P}(X=x)f(x)=\sum_{x\in X(\Omega)}\mathbb{P}(X=x)f(x)$$

Corollaire 4 : Espérance du module

X a une espérance finie si et seulement si $|X|$ a une espérance finie.
Le cas échéant, $\mathbb{E}(|X|) = \sum_{x\in X(\Omega)} \mathbb{P}(X=x)\,|x|$.

Corollaire 5 : Sur un univers fini ou dénombrable

Uniquement dans le cas où Ω est fini ou dénombrable, X est d'espérance finie si et seulement si $\left(\mathbb{P}(\{\omega\})X(\omega)\right)_{\omega\in\Omega}$ est sommable et dans ce cas

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})X(\omega).$$

Remarque

R 58 – Vu en MP2I dans le cas fini.

Démonstration

C'est un peu astucieux ! Considérons la variable aléatoire $Y = \mathrm{id}_\Omega$ et $f = X$ alors, le théorème précédent dit que $X = X\circ \mathrm{id} = f(Y)$ est d'espérance finie si et seulement si $\left(\mathbb{P}(Y=\omega)X(\omega)\right)_{\omega\in\Omega} = \left(\mathbb{P}(\{\omega\})X(\omega)\right)_{\omega\in\Omega}$ est sommable et dans ce cas,

$$\mathbb{E}(X) = \mathbb{E}(X\circ Y) = \sum_{\omega\in\Omega} \mathbb{P}(Y=\omega)X(\omega) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})X(\omega).$$

Autre démonstration possible similaire à celle du théorème de transfert, pour tout $x\in X(\Omega)$, $\mathbb{P}(X=x) = \sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} \mathbb{P}(\{\omega\})$ donc, par sommation par paquets (cas positif, dans $[0,+\infty]$), $\Omega = \bigsqcup_{x\in X(\Omega)} (X=x)$,

$$\sum_{x\in X(\Omega)} \mathbb{P}(X=x)\,|x| = \sum_{x\in X(\Omega)} \sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} \mathbb{P}(\{\omega\})\,|x| = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})\,|X(\omega)|$$

car x est uniquement déterminé par ω, d'où l'équivalence entre X d'espérance finie et $\left(\mathbb{P}(\{\omega\})X(\omega)\right)_{\omega\in\Omega}$ est sommable, et, avec le même calcul sans les modules,

$$\mathbb{E}(X) = \sum_{x\in X(\Omega)} \sum_{\substack{\omega\in\Omega\\ X(\omega)=x}} \mathbb{P}(\{\omega\})x = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})X(\omega)$$ ■

Exercice 13 : CCINP 97

3 Propriétés de l'espérance

Une espérance peut être vue comme une intégrale, ce qui rend toutes ces propriétés naturelles.

Propriété 40 : de l'espérance

X et Y désignent deux variables aléatoires réelles ou complexes discrètes.

(i) *Si X est constante presque sûrement, c'est-à-dire qu'on a $a\in\mathbb{K}$ tel que $\mathbb{P}(X=a)=1$, alors elle est d'espérance finie $\mathbb{E}(X)=a$.*

(ii) ***Linéarité*** *: si $X,Y\in L^1$ et $\lambda\in\mathbb{K}$, $X+\lambda Y\in L^1$ et*

$$\mathbb{E}(X+\lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y).$$

(iii) ***Positivité*** *: si $X\in L^1$ est à valeurs réelles, positive presque sûrement ie $\mathbb{P}(X\geqslant 0)=1$, alors $\mathbb{E}(X)\geqslant 0$.*
Positivité améliorée *: si X est à valeurs réelles, positive presque sûrement et si X est nulle presque sûrement.*

(iv) ***Croissance*** *: si $X, Y \in L^1$ sont à valeurs réelles et si $X \leqslant Y$ presque sûrement, alors $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$.*

(v) *Si $X \in L^1$, $X - \mathbb{E}(X)$ est centrée et appelée* ***variable aléatoire centrée associée à*** *X.*

(vi) ***Inégalité triangulaire*** *: Si $X \in L^1$, $|X| \in L^1$ et*

$$|\mathbb{E}(X)| \leqslant \mathbb{E}(|X|).$$

(vii) *Si $Y \in L^1$ et $|X| \leqslant Y$, alors $X \in L^1$ et $|\mathbb{E}(X)| \leqslant \mathbb{E}(Y)$.*
En particulier, si X est bornée, elle est d'espérance finie.

Démonstration

(i) Immédiat.

(ii) **Linéarité** : Si Ω est fini ou dénombrable, le corollaire 5 précédent donne directement la linéarité.
Dans le cas général, on peut pose $Z = (X, Y)$ variable aléatoire discrète à valeurs dans $E = X(\Omega) \times Y(\Omega)$.

Soit $f : \left| \begin{array}{ccc} E & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & x + \lambda y \end{array} \right.$ Alors $T = X + \lambda Y = f(Z)$.

D'après le théorème de transfert, T est d'espérance finie si et seulement si $\left(\mathbb{P}(X = x, Y = y)(x + \lambda y)\right)_{(x,y)\in E}$ est sommable. Or, dans $[0, +\infty]$,

$$\begin{aligned}
\sum_{(x,y)\in E} \mathbb{P}(X = x, Y = y)\left|x + \lambda y\right| &\leqslant \sum_{(x,y)\in E} \mathbb{P}(X = x, Y = y)\,|x| + |\lambda| \sum_{(x,y)\in E} \mathbb{P}(X = x, Y = y)\left|y\right| \\
&= \sum_{x\in X(\Omega)} \left(\sum_{y\in Y(\Omega)} \mathbb{P}(X = x, Y = y) \right) |x| \\
&\qquad + |\lambda| \sum_{y\in Y(\Omega)} \left(\sum_{x\in X(\Omega)} \mathbb{P}(X = x, Y = y) \right) |y| \text{ (par thm de Fubini positif)} \\
&= \sum_{x\in X(\Omega)} \mathbb{P}(X = x)\,|x| + |\lambda| \sum_{y\in Y(\Omega)} \mathbb{P}(Y = y)\left|y\right| \\
&< +\infty \qquad (\text{car } X, Y \in L^1)
\end{aligned}$$

Donc $\left(\mathbb{P}(X = x, Y = y)x\right)_{(x,y)\in E}$, $\left(\mathbb{P}(X = x, Y = y)y\right)_{(x,y)\in E}$ et $\left(\mathbb{P}(X = x, Y = y)(x + \lambda y)\right)_{(x,y)\in E}$ sont sommables, et avec les mêmes utilisation du théorème de Fubini, et la formule de transfert, on obtient $X + \lambda Y \in L^1$ et

$$\begin{aligned}
\mathbb{E}(X + \lambda Y) &= \sum_{(x,y)\in E} \mathbb{P}(X = x, Y = y)\left(x + \lambda y\right) = \sum_{(x,y)\in E} \mathbb{P}(X = x, Y = y)x + \lambda \sum_{(x,y)\in E} \mathbb{P}(X = x, Y = y)\left|y\right| \\
&= \sum_{x\in X(\Omega)} \mathbb{P}(X = x)x + \lambda \sum_{y\in Y(\Omega)} \mathbb{P}(X = y)y = \mathbb{E}(X) + \lambda \mathbb{E}(Y).
\end{aligned}$$

(iii) **Positivité** : Si X est positive presque sûrement, pour tout $x \in X(\Omega)$, $\mathbb{P}(X = x)x \geqslant 0$ donc $\mathbb{E}(X) \geqslant 0$.
Si, de plus, $\mathbb{E}(X) = 0$, alors pour tout $x \in X(\Omega)$, $x = 0$ ou $\mathbb{P}(X = x) = 0$. Donc $\mathbb{P}(X = 0) = 1$: X est nulle presque partout.

(iv) **Croissance** : $Y - X \geqslant 0$ presque sûrement, d'espérance finie : utiliser la linéarité.

(v) Conséquence de la linéarité de de l'espérance d'une variable aléatoire constante.

(vi) **Inégalité triangulaire** : $|X|$ est d'espérance finie par théorème de transfert, puis il suffit d'appliquer l'inégalité triangulaire d'une somme de famille sommable.

(vii) Y est d'espérance finie, donc $(y\mathbb{P}(Y = y))_{y\in Y(\Omega)}$ est sommable.
Soit $f_1 : (x, y) \mapsto x$, $f_2 : (x, y) \mapsto y$ et $Z = (X, Y)$.
On a $X = f_1(X, Y) = f_1(Z)$ donc la formule de transfert dit que X a une espérance finie si et seulement si $\left(\mathbb{P}(Z = (x, y) f_1(x, y)\right)_{(x,y)\in Z(\Omega)} = \left(\mathbb{P}(X = x, Y = y)x\right)_{(x,y)\in Z(\Omega)}$ est sommable.
Mais si $(x, y) \in Z(\Omega)$, il existe $\omega \in \Omega$ tel que $X(\omega) = x$ et $Y(\omega) = y$ donc $|x| \leqslant y$. Alors

$$\mathbb{P}(X = x, Y = y)\,|x| \leqslant \mathbb{P}(X = x, Y = y)y = \mathbb{P}(X = x, Y = y) f_2(x, y).$$

Mais comme $Y = f_2(X, Y)$ a une espérance finie, le théorème de transfert nous dit que $\left(\mathbb{P}(Z = (x, y) f_2(x, y)\right)_{(x,y)\in Z(\Omega)} = \left(\mathbb{P}(X = x, Y = y)y\right)_{(x,y)\in Z(\Omega)}$ est sommable.
Donc $\left(\mathbb{P}(X = x, Y = y)x\right)_{(x,y)\in Z(\Omega)}$ l'est et X a une espérance finie.
On conclut en invoquant l'inégalité triangulaire puis la croissance de l'espérance. ■

Corollaire 6 : Espace vectoriel L^1

L'ensemble L^1 des variables aléatoires discrètes sur $(\Omega, \mathscr{A}, \mathbb{P})$ admettant une espérance finie est donc un $\mathbb{K}$-espace vectoriel et $X \mapsto \mathbb{E}(X)$ est une forme linéaire sur L^1.

Exercice 14 : Inégalité de Jensen

Si X est une variable aléatoire réelle finie et ϕ une fonction réelle d'une variable réelle convexe, montrer que

$$\phi(\mathbb{E}(X)) \leqslant \mathbb{E}(\phi(X)).$$

Si X est une variable aléatoire réelle discrète, ϕ une fonction réelle d'une variable réelle convexe dérivable, et si X et $\phi(X)$ sont d'espérance finie, en comparant la courbe de ϕ à une de ses tangentes, retrouver l'inégalité précédente.

Comme X est finie et comme les $\mathbb{P}(X = x)$ forment une famille de réels positifs de somme 1, l'inégalité de Jensen donne directement $\phi(\mathbb{E}(X)) \leqslant \mathbb{E}(\phi(X))$.

Si ϕ est dérivable, pour tout x, a, $\phi(x) \geqslant \phi'(a)(x-a) + \phi(a)$.

En évaluant en X et en utilisant la croissance et la linéarité de l'espérance,

$$\mathbb{E}(\phi(X)) \geqslant \mathbb{E}(\phi'(a)(X-a) + \phi(a)) = \phi'(a)(\mathbb{E}(X) - a) + \phi(a).$$

En choisissant en $a = \mathbb{E}(X)$, on retrouve l'inégalité.

4 Espérances des lois usuelles

Propriété 41 : Espérance des lois usuelles

(i) *Si $X \sim \mathscr{B}(p)$, alors $\mathbb{E}(X) = p$.*

(ii) *Si $X \sim \mathscr{B}(n,p)$, alors $\mathbb{E}(X) = np$.*

(iii) *Si $X \sim \mathscr{G}(p)$, alors $\mathbb{E}(X) = \dfrac{1}{p}$.*

(iv) *Si $X \sim \mathscr{P}(\lambda)$, alors $\mathbb{E}(X) = \lambda$.*

Remarque

R 59 – L'espérance d'une loi uniforme est la moyenne arithmétiques des valeurs (en nombre fini) prises par la variable aléatoire.

Démonstration

Dans tous les cas, la variable aléatoire est à valeurs réelles positives.

(i) $\mathbb{E}(X) = p \cdot 1 + (1-p) \cdot 0$.

(ii) L'espérance ne dépendant que de la loi on peut se placer dans le cas particulier où $X = X_1 + \cdots + X_n$ avec $X_1, \dots, X_n$ vaiid de loi $\mathscr{B}(p)$. Alors,
$$\mathbb{E}(X) = n\mathbb{E}(X_1) = np.$$

(iii) On calcule en reconnaissant la dérivée d'une série entière géométrique convergente, avec $1 - p \in]0,1[$,
$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} np(1-p)^{n-1} = p\sum_{n=1}^{+\infty} n(1-p)^{n-1} = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

(iv) On calcule, en reconnaissant cette fois une série exponentielle
$$\mathbb{E}(X) = \sum_{n=0}^{+\infty} n\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda} = \lambda\mathrm{e}^{-\lambda}\sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{(n-1)!} = \lambda.$$

■

Corollaire 7 : Cas d'une fonction indicatrice

Soit A un événement de notre tribu $\mathscr{A}$. Alors $\mathbb{1}_A$ a une espérance finie et $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Exercice 15 : Formule de Poincaré

En remarquant que $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \mathbb{P}\left(\bigcap_{i=1}^{n} \overline{A_i}\right) = 1 - \mathbb{E}\left(\mathbb{1}_{\bigcap_{i=1}^{n} \overline{A_i}}\right)$**, prouver la formule de Poincaré.**

On continue le calcul

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \mathbb{E}\left(\prod_{i=1}^{n} (1 - \mathbb{1}_{A_i})\right) = \mathbb{E}\left(\sum_{I \in \mathscr{P}([\![1,n]\!])} (-1)^{|I|-1} \prod_{i \in I} \mathbb{1}_{A_i}\right) = \sum_{I \in \mathscr{P}([\![1,n]\!])} (-1)^{|I|-1} \mathbb{E}\left(\mathbb{1}_{\bigcap_{i\in I} A_i}\right)$$

$$= \sum_{I \in \mathscr{P}([\![1,n]\!])} (-1)^{|I|-1} \mathbb{E}\left(\mathbb{1}_{\bigcap_{i\in I} A_i}\right) = \sum_{I \in \mathscr{P}([\![1,n]\!])} (-1)^{|I|-1} \mathbb{P}\left(\bigcap_{i\in I} A_i\right).$$

5 Exercices CCINP

Exercice 16 : CCINP 102

Exercice 17 : CCINP 103

Exercice 18 : CCINP 106

Exercice 19 : CCINP 108

Exercice 20 : CCINP 111

6 Espérance et indépendance

Propriété 42 : Espérance et indépendance

Soit X, $Y \in L^1$ indépendantes. Alors $XY \in L^1$, et

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

Réciproque fausse en général.

*Plus généralement, si $(X_1, \ldots, X_n)$ est une famille de variables aléatoires **indépendantes** d'espérance finie, alors $\prod_{i=1}^{n} X_i$ l'est et*

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

Démonstration

On utilise le théorème de transfert appliquée à $f : (x, y) \mapsto xy$.

$f(X, Y) = XY$ est d'espérance finie si et seulement si $\left(\mathbb{P}(X = x, Y = y)xy\right)_{(x,y)\in(X,Y)(\Omega)}$ est sommable, ce qui équivaut, en ajoutant des zéros, à la sommabilité de $\left(\mathbb{P}(X = x, Y = y)xy\right)_{(x,y)\in X(\Omega)\times Y(\Omega)} = \left(\mathbb{P}(X = x)x \times \mathbb{P}(Y = y)y\right)_{(x,y)\in X(\Omega)\times Y(\Omega)}$ par indépendance.

On est ramenée à une « suite double produit » qui est bien sommable car X et Y ont des espérance finies et dont la somme est le produit des espérances.

Contre-exemple : Soit X_1 telle que $\mathbb{P}(X_1 = 0) = \frac{1}{2}$ et $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{4}$. $\mathbb{E}(X_1) = 0$. Et $X_2 = \mathbb{1}_{(X_1=0)}$.

Alors $X_1X_2 \equiv 0$ donc $\mathbb{E}(X_1X_2) = 0 = \mathbb{E}(X_1)\mathbb{E}(X_2)$.

Pourtant $X_1 \not\perp\!\!\!\perp X_2$ car $\mathbb{P}(X_1 = 1, X_2 = 1) = 0 \neq \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) = \frac{1}{8}$.

Pour, la généralisation, par récurrence, en utilisant, d'après le lemme des coalitions, si $X_1, \ldots, X_{n+1}$ sont indépendantes, alors $X_1 \cdots X_n$ et X_{n+1} sont bien indépendantes. ■

VIII VARIANCE, ÉCART-TYPE ET COVARIANCE

On fixe ici un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$. Les variables aléatoires considérées sont à valeurs réelles.

1 Espace L^2

Sous réserve d'existence, les moments (dénomination hors programme) d'une variable aléatoire sont les $\mathbb{E}\left(|X|^k\right)$ pour $k \in \mathbb{N}^*$. Ce sont des paramètres numériques qui donnent des renseignements sur sa loi. En général, on se limite aux moments d'ordre 1 (espérance) et d'ordre 2 (permet d'obtenir la variance).

Notation 2 : L^2

Soit X une variable aléatoire réelle discrète.

On note $X \in L^2$ lorsque X^2 est d'espérance finie (ce qu'on peut noter $\mathbb{E}\left(X^2\right) < +\infty$ car X^2 est à valeurs réelles positives).

Propriété 43 : Inégalité de Cauchy-Schwarz

Si deux variables aléatoires réelles discrètes $X, Y \in L^2$, leur produit $XY \in L^1$, et

$$(\mathbb{E}(XY))^2 \leqslant \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)$$

avec égalité si et seulement si X et Y sont colinéaires presque sûrement, c'est-à-dire lorsqu'il existe $(\lambda, \mu) \neq (0,0)$ tel tel que $\mathbb{P}(\lambda X + \mu Y = 0) = 1$.

Démonstration

$|XY| \leqslant \frac{1}{2}\left(X^2 + Y^2\right)$ est d'espérance finie, puis on applique l'inégalité de Cauchy-Schwarz à la forme bilinéaire symétrique positive $(X, Y) \mapsto \mathbb{E}(XY)$.

Pour le cas d'égalité, il faut reprendre la preuve dans laquelle on écrit $\mathbb{E}\left((X + \lambda Y)^2\right) = \mathbb{E}\left(Y^2\right)\lambda^2 + 2\mathbb{E}(X)\mathbb{E}(Y)\lambda + \mathbb{E}\left(X^2\right) \geqslant 0$.

Ou bien $\mathbb{E}\left(Y^2\right) = 0$ et alors, comme $Y^2 \geqslant 0$, Y est nulle presque sûrement, donc XY l'est aussi ce qui correspond bien à un cas d'égalité, ou bien c'est un trinôme du second degré dont le discriminant est nul si et seulement s'il existe $lambda \in \mathbb{R}$ tel que $\mathbb{E}\left((X + \lambda Y)^2\right) = 0$ et donc, avec le même argument, $X = -\lambda Y$ presque sûrement. ■

Corollaire 8

L^2 est un $\mathbb{R}$-espace vectoriel.

Propriété 44 : $L^2 \subset L^1$

Si $X \in L^2$, $X \in L^1$.

Démonstration

Première méthode : Inégalité de Cauchy-Schwarz à appliquée à X et 1.

Deuxième méthode : Comment dans la preuve précédente : $|X| = |X \times 1| \leqslant \frac{1}{2}\left(X^2 + 1\right)$.

Remarque

R 60 – Donc L^2 est un sous-espace de L^1.

2 Variance et écart-type

Définition 26 : Variance, écart-type, variable réduite

Soit $X \in L^2$.
On appelle **variance** de X le nombre

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

On appelle **écart-type** de X le nombre

$$\sigma(X) = \sqrt{\mathbb{V}(X)} = \sqrt{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}.$$

Lorsque $\mathbb{V}(X) = 1$, X est dite **réduite**.

Remarque

R 61 – $\mathbb{V}(X)$ est le moment d'ordre 2 de la variable aléatoire centrée associée à X : $X - \mathbb{E}(X)$. Par positivité de l'espérance, $\mathbb{V}(x) \geqslant 0$ donc l'écart-type est bien défini.

R 62 – L'écart-type d'une variable aléatoire finie s'interprète comme une distance euclidienne dans $\mathbb{R}^n$ entre le vecteur dont les coordonnées sont les valeurs prises par X et le vecteur dont toutes les coordonnées valent $\mathbb{E}(X)$. C'est donc un indicateur de dispersion de X autour de sa moyenne $\mathbb{E}(X)$.

R 63 – Ne pas hésiter à noter $m = \mathbb{E}(X)$. Il est plus facile à visualiser $\mathbb{E}(X - m) = \mathbb{E}(X) - m = 0$ que $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}(X) - \mathbb{E}(X) = 0$, par exemple.

R 64 – D'après la formule de transfert, si les valeurs prises par X sont les x_i pour $i \in I$ et $m = \mathbb{E}(X)$,

$$\mathbb{V}(X) = \sum_{i \in I} \mathbb{P}(X = x_i)(x_i - m)^2.$$

R 65 – Plus la variance (et donc l'écart-type) est petit, plus X est concentrée autour de sa moyenne $m = \mathbb{E}(X)$.
Le cas extrême est pour une variable aléatoire constante : $\mathbb{V}(X) = 0$.
Réciproquement, si $\mathbb{V}(X) = 0$, alors

$$\forall x \in X(\Omega),\quad \mathbb{P}(X = x) = 0 \text{ ou } x = m = \mathbb{E}(X).$$

Autrement dit, $\mathbb{P}(X \neq m) = 0$ ou encore $\mathbb{P}(X = m) = 1$: X est constante presque sûrement.

Exercice 21 : CCINP 100

Propriété 45 : de la variance

Soit $X \in L^2$.

(i) ***Formule de Kœnig-Huygens** :*

$$\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2.$$

(ii) *Si $a, b \in \mathbb{R}$, $\mathbb{V}(aX + b) = a^2\mathbb{V}(x)$ donc $\sigma(aX + b) = |a|\sigma(X)$.*

(iii) *Si $\sigma(X) \neq 0$, $\dfrac{X - \mathbb{E}(X)}{\sigma(X)}$ est centrée réduite, appelée **variable aléatoire centrée réduite associée à** X.*

Remarque

R 66 – La deuxième formule est intuitive au sens où une translation des valeurs de X ne perturbe la distance à la moyenne, et comme cette distance est au carré, une homothétie de rapport a la multiplie par a^2.

Démonstration

(i) $\mathbb{V}(X) = \mathbb{E}\left(X^2 - 2mX + m^2\right) = \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ par linéarité.

(ii) $\mathbb{V}(aX+b) = \mathbb{E}\left((aX+b)^2\right) - \mathbb{E}(aX+b)^2 = a^2\mathbb{E}\left(X^2\right) + 2ab\mathbb{E}(X) + b^2 - a^2\mathbb{E}(X)^2 - 2ab\mathbb{E}(X) + b^2 = a^2V(X)$. ■

3 Covariance

Définition 27 : Covariance

Soit $(X, Y) \in \left(L^2\right)^2$ un couple de variables aléatoires réelles discrètes admettant un moment d'ordre 2. On appelle **covariance** du couple (X, Y) le nombre

$$\mathrm{Cov}(X, Y) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\Big).$$

Lorsque $\mathrm{Cov}(X, Y) = 0$, X et Y sont dites **non corrélées**.

Remarque

R 67 – La covariance mesure la corrélation entre les variations de X et de Y dans le sens où elle est positive lorsque X et Y s'écartent de leur moyenne dans le même sens, et négative si c'est dans le sens opposé.

R 68 – Cela ressemble à un produit scalaire et ce n'est pas un hasard ! On va vérifier que c'est une forme bilinéaire symétrique positive. La variance correspond au carré de la « norme » (et donc l'écart-type à la « norme ».)
Elle n'est pas définie positive mais presque : $\mathrm{Cov}(X, X) = \mathbb{V}(X) = 0 \Longrightarrow X = 0$ presque sûrement.

Propriété 46 : de la covariance

Soient X, $Y \in L^2$ deux variables aléatoires réelles discrète admettant un moment d'ordre 2.

(i) Cov *est une forme bilinéaire symétrique positive.*

(ii) ***Formule de Kœnig-Huygens*** *:*

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

(iii) $\mathbb{V}(X+Y) = \mathbb{V}(X) + 2\,\mathrm{Cov}(X, Y) + \mathbb{V}(Y)$.

(iv) Si $X \perp\!\!\!\perp Y$, $\mathrm{Cov}(X, Y) = 0$ et la réciproque est fausse.

(v) ***Inégalité de Cauchy-Schwarz*** *:*

$$(\mathrm{Cov}(X, Y))^2 \leqslant \mathbb{V}(X)\mathbb{V}(Y) \quad ie \quad |\mathrm{Cov}(X, Y)| \leqslant \sigma(X)\sigma(Y)$$

avec égalité si et seulement si les variables aléatoires sont colinéaires presque sûrement.

Démonstration

(i) Provient de la linéarité et la positivité de $\mathbb{E}$.

(ii) $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

(iii) C'est une identité remarquable :

$$\mathbb{V}(X+Y) = \mathrm{Cov}(X+Y, X+Y) = \mathrm{Cov}(X, X) + 2\,\mathrm{Cov}(X, Y) + \mathrm{Cov}(Y, Y) = \mathbb{V}(X) + 2\,\mathrm{Cov}(X, Y) + V(Y).$$

ou alors

$$\begin{aligned}\mathbb{V}(X+Y) &= \mathbb{E}\left((X - \mathbb{E}(X) + Y - \mathbb{E}(Y))^2\right)\\ &= \mathbb{E}\left((X - \mathbb{E}(X))^2\right) + 2\mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) + \mathbb{E}\left((Y - \mathbb{E}(Y))^2\right).\end{aligned}$$

(iv) Immédiat avec (ii). (Voir contre-exemple de $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.)

(v) C'est l'inégalité de Cauchy-Schwarz sur les espérance appliquée aux variables aléatoires centrées associées à X et à Y. ■

Remarque

R 69 – $\dfrac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)} = \mathrm{Cov}\left(\dfrac{X}{\sigma(X)} \times \dfrac{Y}{\sigma(Y)}\right) \in]-1,1[$ est le cœfficient de corrélation de X et Y.

4 Variance d'une somme de variables aléatoires

Propriété 47 : Variance d'une somme

Soient $X_1,\dots,X_n \in L^2$.

(i) $X_1+\dots+X_n \in L^2$ et

$$\mathbb{V}(X_1+\dots+X_n) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2\sum_{1\leqslant i<j\leqslant n} \mathrm{Cov}\left(X_i,X_j\right).$$

(ii) Si $X_1,\dots,X_n$ sont décorrélées deux à deux ($i\neq j \Rightarrow \mathrm{Cov}(X_i,X_j)=0$),

$$\mathbb{V}(X_1+\dots+X_n) = \mathbb{V}(X_1)+\dots+\mathbb{V}(X_n).$$

En particulier, si $X_1,\dots,X_n$ sont des vaiid,

$$\mathbb{V}(X_1+\dots+X_n) = n\mathbb{V}(X_1).$$

Démonstration

L^2 est un $\mathbb{R}$-espace vectoriel.

(i) Le cas $n=1$ est trivial et le cas $n=2$ a déjà été vu.
Si c'est vrai pour $n-1$, alors

$$\begin{aligned}\mathbb{V}(X_1+\dots+X_n) &= \mathbb{V}(X_1+\dots+X_{n-1}+X_n) = \mathbb{V}(X_1+\dots+X_{n-1})+\mathbb{V}(X_n)+2\mathrm{Cov}(X_1+\dots+X_{n-1},X_n)\\ &= \sum_{i=1}^{n-1}\mathbb{V}(X_i)+\mathbb{V}(X_n)+2\sum_{1\leqslant i<j\leqslant n-1}\mathrm{Cov}\left(X_i,X_j\right)+2\sum_{1\leqslant i<j=n}\mathrm{Cov}\left(X_i,X_j\right).\\ &= \sum_{i=1}^{n}\mathbb{V}(X_i)+2\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}\left(X_i,X_j\right).\end{aligned}$$

(ii) Immédiat. ■

5 Cas des lois usuelles

Propriété 48 : Espérance et variance des lois usuelles

(i) Si $X\sim\mathscr{B}(p)$, $\mathbb{E}(X)=p$ et $\mathbb{V}(X)=p(1-p)=pq$.

(ii) Si $X\sim\mathscr{B}(n,p)$, $\mathbb{E}(X)=np$ et $\mathbb{V}(X)=np(1-p)=npq$.

(iii) Si $X\sim\mathscr{G}(p)$, $\mathbb{E}(X)=\dfrac{1}{p}$ et $\mathbb{V}(X)=\dfrac{1-p}{p^2}=\dfrac{q}{p^2}$.

(iv) Si $X\sim\mathscr{P}(\lambda)$, $\mathbb{E}(X)=\mathbb{V}(X)=\lambda$.

Démonstration

(i) $X^2=X$, $\mathbb{V}(X)=\mathbb{E}(X)-p^2=p-p^2=p(1-p)=pq$.

(ii) On prend $X=X_1+\ldots+X_n$ où $X_1,\dots,X_n$ sont des vaiid de loi $\mathscr{B}(p)$. $\mathbb{V}(X)=\sum_{k=1}^{n}\mathbb{V}(X_i)=np(1-p)=npq$.

(iii) $\mathbb{V}(X)=\mathbb{E}\left(X^2\right)-\dfrac{1}{p}^2$ avec $\mathbb{E}\left(X^2\right)=\sum_{n\in\mathbb{N}^*}n^2pq^{n-1}=pq\sum_{n\in\mathbb{N}^*}n(n-1)pq^{n-2}+p\sum_{n\in\mathbb{N}^*}nq^{n-1}=\dfrac{2pq}{(1-q)^3}+\dfrac{p}{(1-q)^2}=\dfrac{1+q}{p^2}$

donc $\mathbb{V}(X) = \dfrac{q}{p^2}$.

(iv) même principe mais avec de l'exponentielle.

IX Inégalités de Markov et de Bienaymé-Tchebychev, Loi faible des grands nombres

Propriété 49 : Inégalité de Markov

Soit $X \in L^1$ une variable aléatoire discrète admettant une espérance finie. Pour tout $a > 0$,

$$\mathbb{P}(|X| \geqslant a) \leqslant \frac{\mathbb{E}(|X|)}{a}.$$

Remarque

R 70 – Si X est à valeurs positives, on a donc aussi $\mathbb{P}(X \geqslant a) \leqslant \dfrac{\mathbb{E}(X)}{a}$.

R 71 – On a aussi $\mathbb{P}(|X| > a) \leqslant \dfrac{\mathbb{E}(|X|)}{a}$.

Démonstration

Soit $Y = |X|$.

Première méthode : Y étant positive,

$$\mathbb{E}(Y) \geqslant \sum_{y \geqslant a} \mathbb{P}(Y = y) y \geqslant a \sum_{y \geqslant a} \mathbb{P}(Y = y) = a\mathbb{P}(Y \geqslant a).$$

Deuxième méthode : Formule de transfert

$$\mathbb{E}(|X|) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x)|x| \leqslant a \sum_{|x| \geqslant a} \mathbb{P}(X = x) = a\mathbb{P}(|X| \geqslant a)$$

Troisième méthode : $\mathbb{1}_{(Y \geqslant a)} \leqslant \frac{Y}{a}$ et croissance de $\mathbb{E}$.

Propriété 50 : Inégalité de Bienaymé-Tchebychev

Soit $X \in L^2$ une variable aléatoire réelle discrète admettant un moment d'ordre 2, $a > 0$.

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant a\right) \leqslant \frac{\mathbb{V}(X)}{a^2}$$

c'est-à-dire, en notant m l'espérance de X et σ son écart-type,

$$\mathbb{P}\left(|X - m| \geqslant a\right) \leqslant \frac{\sigma^2}{a^2}$$

Démonstration

Découle directement de l'inégalité de Markov :

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant a\right) = \mathbb{P}\left((X - \mathbb{E}(X))^2 \geqslant a^2\right) \leqslant \frac{\mathbb{E}\left((X - \mathbb{E}(X))^2\right)}{a^2} = \frac{\mathbb{V}(X)}{a^2}.$$

ou, directement,

$$\mathbb{V}(X) = \mathbb{E}\left((X-\mathbb{E}(X))^2\right) = \sum_{x\in X(\Omega)} (x-\mathbb{E}(X))^2\,\mathbb{P}(X=x)$$
$$\geqslant a^2 \sum_{x \mid |x-\mathbb{E}(X)|\geqslant a} \mathbb{P}(X=x) = a^2\mathbb{P}(|X-\mathbb{E}(X)|\geqslant a) \quad \blacksquare$$

Remarque

R 72 – Le a^2 est logique pour des raisons d'homogénéité (dimension).

R 73 – On retrouve avec l'inégalité de Bienaymé-Tchebychev, le fait que si $\mathbb{V}(X)=0$,

$$\forall \varepsilon > 0,\ \ \mathbb{P}(|X-\mathbb{E}(X)|\geqslant \varepsilon) = 0.$$

Donc, $\mathbb{P}(|X-\mathbb{E}(X)|>0)=0$, c'est-à-dire $\mathbb{P}(X=\mathbb{E}(X))=1$.

R 74 – En particulier, $\mathbb{P}(|X-\mathbb{E}(X)|<a)\geqslant 1-\dfrac{\mathbb{V}(X)}{a^2}$.

Intuitivement, en répétant de nombreuses fois un lancer de pièce équilibrée, la fréquence d'apparition de pile doit se rapprocher de $\frac{1}{2}$.

Le théorème suivant permet de donner un cadre théorique à cette intuition.

Théorème 3 : Loi faible des grands nombres

Soit $(X_n)_{n\geqslant 1}\in (L^2)^{\mathbb{N}^*}$ une suite de variable aléatoires discrètes réelles deux à deux indépendantes identiquement distribuées (de même loi) sur $(\Omega,\mathscr{A},\mathbb{P})$, admettant un moment d'ordre 2. Soit m l'espérance de X_n et σ son écart-type.

On pose enfin $S_n = X_1+\cdots+X_n$.

Pour tout $\varepsilon>0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n}-m\right|\geqslant\varepsilon\right)\xrightarrow[n\to+\infty]{} 0.$$

Remarque

R 75 – Parmi tous les échantillons de valeurs possibles $(X_1,\ldots,X_n)$, ceux dont la moyenne (S_n/n) s'éloigne de l'espérance m sont rares, et cette rareté s'accentue avec la taille de l'échantillon $(n\to+\infty)$.

Démonstration

D'après l'inégalité de Bienaymé-Tchebychev,

$$\mathbb{P}\left(\left|\frac{S_n}{n}-m\right|\geqslant\varepsilon\right)\leqslant\frac{\sigma^2}{n\varepsilon^2}.$$

En effet, $\mathbb{E}\left(\frac{S_n}{n}\right)=m$ et $\mathbb{V}\left(\frac{S_n}{n}\right)=\frac{\sigma^2}{n}$ par indépendance. $\blacksquare$

Exercice 22 : CCINP 99

X FONCTIONS GÉNÉRATRICES

Dans cette partie, les variables aléatoires sont à valeurs dans $\mathbb{N}$.

1 Définition

Définition 28 : Fonction génératrice

Soit X variable aléatoire discrète sur $(\Omega, \mathscr{A}, \mathbb{P})$ **à valeurs dans** $\mathbb{N}$.
On appelle **fonction génératrice associée à** X la fonction $G_X : t \mapsto \sum_{n=0}^{+\infty} \mathbb{P}(X = n) t^n$.

Remarque

R 76 – Il s'agit donc de la somme de la série entière sont la suite de cœfficients est la (distribution de probabilité associée à la) loi de X.
L'unicité des cœfficients de la série entière assure que la fonction génératrice détermine la loi de X (il suffit de calculer ces cœfficients).

Propriété 51 : des fonctions génératrices

(*i*) *Le rayon de convergence de la série entière* $\sum \mathbb{P}(X = n) t^n$ *est au moins égal à* 1,*et elle converge normalement sur* $[-1, 1]$.

(*ii*) *Pour tout* $t \in [-1, 1]$, $G_X(t) = \mathbb{E}\left(t^X\right)$.

(*iii*) G_X *est continue sur* $[-1, 1]$, *de classe* $\mathscr{C}^\infty$ *sur* $]-1, 1[$ *et pour tout* $n \in \mathbb{N}$, $\mathbb{P}(X = n) = \dfrac{G_X^{(n)}(0)}{n!}$.

Démonstration

(i) Vient du fait que $\sum \mathbb{P}(X = n)$ converge.

(ii) Théorème de transfert grâce à la converge absolue de la série.

(iii) Convergence normale sur $[-1, 1]$ et propriété des séries entières. ■

Remarque

R 77 – Avec la dernière propriété, on vérifie de nouveau que G_X détermine la loi de X.

Corollaire 9 : Caractérisation de la loi

Deux variables aléatoires X, Y *à valeurs dans* $\mathbb{N}$ *ont même loi si et seulement si elles ont même fonction génératrice.*

Démonstration

Unicité du DSE. ■

CC BY NC SA

Propriété 52 : Lien avec l'espérance et la variance

(i) $X \in L^1$ *(est d'espérance finie) si et seulement si* G_X *est dérivable est* 1 *et alors* $\mathbb{E}(X) = G_X'(1)$.

(ii) $X \in L^2$ *si et seulement si* G_X *est deux fois dérivable est* 1 *et alors* $\mathbb{E}(X(X-1)) = G_X''(1)$.
On exprime alors $\mathbb{V}(X)$ *à l'aide de* $G_X'(1)$ *et* $G_X''(1)$.

Démonstration

Soit $f_n : t \mapsto \mathbb{P}(X = n)t^n$. Seul le sens (facile) $\Longrightarrow$ est exigible.

(i) Si X est d'espérance finie, alors on vérifie que $\sum f_n'$ converge normalement ce qui permet via théorème de justifier que G_X est de classe $\mathscr{C}^1$ sur $[-1,1]$ et d'obtenir $G_X'(1) = \mathbb{E}(X)$.
Si G_X est dérivable est 1, G_X'' étant positive sur $[0,1[$, G_X' est croissante et admet une limite en 1.
Comme G_X est continue en 1, le théorème de la limite de la dérivée s'applique et cette limite ne peut valoir que $G_X'(1)$.
Puis on majore $\sum_{n=0}^{N} \mathbb{P}(X=n)nt^{n-1}$ par $G_X'(1)$ et on fait tendre t vers 1 : on obtient que $(\mathbb{P}(X=n)n)_n$ est sommable puis le résultat.

(ii) Même principe appliqué à G_X' : $X(X-1)$ est d'espérance finie si et seulement si G_X est deux fois dérivable en 1 et $\mathbb{E}(X(X-1)) = G_X''(1)$.
Puis $\mathbb{V}(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2 = \mathbb{E}(X(X-1)) + \mathbb{E}(X) - \mathbb{E}(X)^2 = G_X''(1) + G_X(1) - G_X'(1)^2$.

2 Cas des lois usuelles

Le programme demande de savoir calculer la fonction génératrice d'une variable aléatoire de Bernoulli, binomiale, géométrique, de Poisson. Allons-y.

Loi de Bernoulli $\mathscr{B}(p)$:

$$\boxed{G_X(t) = q + pt = 1 - p + pt}$$

définie sur $\mathbb{R}$, ce qui redonne bien $\mathbb{E}(X) = p$ et $\mathbb{V}(X) = pq$.

Loi binomiale $\mathscr{B}(n,p)$: $G_X(t) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} t^k$ donc

$$\boxed{G_X(t) = (q+pt)^n = (1-p+pt)^n}$$

définie sur $\mathbb{R}$, ce qui redonne bien $\mathbb{E}(X) = np$ et $\mathbb{V}(X) = npq$.

Loi géométrique $\mathscr{G}(p)$: $G_X(t) = \sum_{k=1}^{+\infty} pq^{n-1}t^n$ donc

$$\boxed{G_X(t) = \frac{pt}{1-qt} = \frac{pt}{1-(1-p)t}}$$

définie sur $\left]-\frac{1}{q}, \frac{1}{q}\right[$, ce qui redonne bien $\mathbb{E}(X) = \frac{1}{p}$ et $\mathbb{V}(X) = \frac{p}{q^2}$.

Loi de Poisson $\mathscr{P}(\lambda)$: $G_X(t) = \sum_{k=0}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} t^n$ donc

$$\boxed{G_X(t) = e^{\lambda(t-1)}}$$

définie sur $\mathbb{R}$, ce qui redonne bien $\mathbb{E}(X) = \mathbb{V}(X) = \lambda$.

3 Somme des variables aléatoires

Propriété 53 : Fonction génératrice d'une somme

Soient $X_1, \dots, X_n$ des variables aléatoires discrètes indépendantes à valeurs dans $\mathbb{N}$. Alors

$$G_{X_1+\cdots+X_n} = \prod_{i=1}^{n} G_{X_i}.$$

Démonstration

Pour chaque t, les t^{X_i} sont indépendantes, donc $\mathbb{E}\left(t^{X_1+\cdots+X_n}\right) = \mathbb{E}\left(\prod_{i=1}^{n} t^{X_i}\right) = \prod_{i=1}^{n} \mathbb{E}\left(t^{X_i}\right)$. ■

Applications

- On retrouve la fonction génératrice d'une loi binomiale à partir de la somme de n vaiid de loi de Bernoulli.
- Une somme de variables aléatoires de loi de Poisson $\mathscr{P}(\lambda_i)$ est encore de loi de Poisson de paramètre la somme des λ_i.
- Une somme de variables aléatoires de loi $\mathscr{B}(n_i, p)$ indépendantes est de loi $\mathscr{B}(\sum n_i, p)$

Exercice 23 : Identité de Wald

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires de même loi et d'espérance finie à valeurs dans $\mathbb{N}$ et N une variable aléatoire à valeurs dans $\mathbb{N}$, d'espérance finie tel que N et toutes les X_n soient indépendantes.

1. **On suppose dans cette question que les X_n suivent une loi $\mathscr{B}(p)$ de paramètre $p \in]0,1[$ et que N suit une loi $\mathscr{P}(\theta)$ de paramètre $\theta > 0$.**

   Rappeler les fonctions génératrices, pour $n \in \mathbb{N}$, de X_n, $\sum_{\ell=1}^{n} X_\ell$ et N puis déterminer la loi de $Y = \sum_{\ell=1}^{N} X_\ell$ (on admet qu'elle définit bien une variable aléatoire discrète.)
2. **Dans cette question, on suppose que les X_n suivent une même loi quelconque, et que N suit une loi quelconque, toujours toutes à valeurs dans $\mathbb{N}$ et indépendantes.**

   Montrer l'identité de Wald

$$\mathbb{E}\left(\sum_{\ell=1}^{N} X_\ell\right) = \mathbb{E}(N)\mathbb{E}(X_1).$$

---

1. $G_{X_n}(t) = \mathbb{E}\left(t^{X_n}\right) = 1 - p + pt$, $G_{\sum_{\ell=1}^{n} X_\ell}(t) = (1-p+pt)^n$ et $G_N(t) = \mathbb{E}\left(t^N\right) = \mathrm{e}^{\theta(t-1)}$. Nous savons bien sûr que $\sum_{\ell=1}^{n} X_\ell$ suit une loi $\mathscr{B}(n, p)$.

   Pour calculer la loi de Y, la subtilité est que N est une variable aléatoire, ici : $Y(\omega) = \sum_{\ell=1}^{N(\omega)} X_\ell$.

   On peut calculer la loi directement avec la formule des probabilités totales et le sce $(N = n)_{n\in\mathbb{N}}$ mais le début de la question semble nous aiguiller vers les fonctions génératrices.

   Essayons : $Y(\Omega) = \mathbb{N}$ et pour tout $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(Y = k) &= \sum_{n=0}^{+\infty} \mathbb{P}(Y = k, N = n) && \text{probabilités totales}\\
&= \mathrm{e}^{-\theta} \sum_{n=0}^{+\infty} \mathbb{P}\left(\sum_{\ell=1}^{n} X_\ell = k, N = n\right) = \mathrm{e}^{-\theta} \sum_{n=0}^{+\infty} \mathbb{P}\left(\sum_{\ell=1}^{n} X_\ell = k\right) \mathbb{P}(N = n) && \text{indépendance}\\
&= \mathrm{e}^{-\theta} \sum_{n=k}^{+\infty} \frac{\theta^n}{n!} \binom{n}{k} p^k (1-p)^{n-k}\\
&= \frac{\mathrm{e}^{-\theta}(p\theta)^k}{k!} \sum_{n=k}^{+\infty} \frac{(\theta(1-p))^{n-k}}{(n-k)!} = \frac{\mathrm{e}^{-\theta}(p\theta)^k}{k!} \mathrm{e}^{\theta(1-p)} = \mathrm{e}^{-\theta p} \frac{(p\theta)^k}{k!}
\end{aligned}$$

   Donc Y suit une loi $\mathscr{P}(p\theta)$.

Avec les fonctions génératrices, on calcule, toujours avec la formule des probabilités totales et le sce $(N=n)_{n\in\mathbb{N}}$, tout étant sommable pour $|t|<1$,

$$\begin{aligned}G_Y(t)=\mathbb{E}\left(t^Y\right)&=\sum_{k=0}^{+\infty}\mathbb{P}(Y=k)t^k=\sum_{k=0}^{+\infty}\left(\sum_{n=0}^{+\infty}\mathbb{P}(Y=k,N=n)\right)t^k\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)\times\underbrace{\left(\sum_{k=0}^{+\infty}\mathbb{P}\left(\sum_{p=1}^{n}X_p=k\right)t^k\right)}_{\text{fonction génératrice de }\sum_{p=1}^n X_p}=\mathrm{e}^{-\theta}\sum_{n=0}^{+\infty}\frac{\theta^n}{n!}(1-p+pt)^n\\&=\mathrm{e}^{p\theta(t-1)}\end{aligned}$$

Donc Y suit une loi $\mathscr{P}(p\theta)$.

2. Pour un calcul direct, on peut travailler directement dans $[0,+\infty]$ et utiliser Fubini (mais en fait tout est fini, ici)

$$\begin{aligned}\mathbb{E}(Y)&=\sum_{k=0}^{+\infty}\mathbb{P}(Y=k)k=\sum_{k=0}^{+\infty}\left(\sum_{n=0}^{+\infty}\mathbb{P}(Y=k,N=n)\right)k&&\text{formule des probabilités totales}\\&=\sum_{n=0}^{+\infty}\left(\sum_{k=0}^{+\infty}\mathbb{P}\left(\sum_{\ell=1}^{n}X_\ell=k,N=n\right)k\right)&&\text{Fubini positif}\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)\left(\sum_{k=0}^{+\infty}\mathbb{P}\left(\sum_{\ell=1}^{n}X_\ell=k\right)k\right)&&N\text{ indépendant des }X_\ell\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)\mathbb{E}\left(\sum_{\ell=1}^{n}X_\ell\right)\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)n\mathbb{E}(X_1)&&\text{par linéarité et les }X_\ell\text{ de même loi}\\&=\mathbb{E}(N)\mathbb{E}(X_1).\end{aligned}$$

(On redécouvre à chaque fois une formule classique appelée formule de l'espérance totale, mais malheureusement hors-programme.)
Avec les fonctions génératrices, si $|t|<1$,

$$\begin{aligned}G_Y(t)&=\sum_{k=0}^{+\infty}\mathbb{P}(Y=k)t^k=\sum_{k=0}^{+\infty}\left(\sum_{n=0}^{+\infty}\mathbb{P}(Y=k,N=n)\right)t^k&&\text{formule des probabilités totales}\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)\left(\sum_{k=0}^{+\infty}\mathbb{P}\left(\sum_{\ell=1}^{n}X_\ell=k\right)t^k\right)&&\text{Fubini avec sommabilité à justifier}\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)G_{\sum_{\ell=1}^n X_\ell}(t)\\&=\sum_{n=0}^{+\infty}\mathbb{P}(N=n)\left(G_{X_1}(t)\right)^n&&\text{par indépendance}\\&=G_N\circ G_{X_1}(t),\end{aligned}$$

La sommabilité se justifiant par le fait que $\sum_{k=0}^{+\infty}\sum_{n=0}^{+\infty}\mathbb{P}(Y=k,N=n)|t|^k=G_Y(|t|)<+\infty$.
Il reste à dériver et évaluer en 1 : $\mathbb{E}(Y)=G'_Y(1)=G'_{X_1}(1)G'_N(G_{X_1}(1))=\mathbb{E}(X_1)G'_N(1)=\mathbb{E}(N)\mathbb{E}(X_1)$.

4 Exercices CCINP

Exercice 24 : CCINP 96

Exercice 25 : CCINP 110

XI FORMULAIRE

Sous réserve d'existence, sommabilité, d'admission de moments, etc. Voici les principales formules du chapitre.

- **Loi de X** $\boxed{\mathbb{P}_X : A \mapsto \sum_{x\in A} \mathbb{P}(X=x)}$ déterminée par les $\mathbb{P}(X=x)$ pour $x \in X(\Omega)$, positifs de somme 1.
- **Espérance de X** $\boxed{\mathbb{E}(X) = \sum_{x\in X(\Omega)} \mathbb{P}(X=x)x}$ et si Ω fini ou dénombrable $\boxed{\mathbb{E}(X) = \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\})X(\omega)}$ et si $X(\Omega) \subset \mathbb{N}$, $\boxed{\mathbb{E}(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X \geqslant n).}$
- **Formule de transfert** $\boxed{\mathbb{E}(f(X)) = \sum_{x\in X(\Omega)} \mathbb{P}(X=x)f(x).}$
- **Variance de X** $\boxed{\mathbb{V}(X) = \mathbb{E}\left((X-\mathbb{E}(X))^2\right) = \mathbb{E}\left(X^2\right) - \mathbb{E}(X)^2.}$
- **Covariance de X et Y** $\boxed{\operatorname{Cov}(X,Y) = \mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}$ nulle si indépendantes.
- **Variance d'une somme** $\boxed{\mathbb{V}(X+Y) = \mathbb{V}(X) + 2\operatorname{Cov}(X,Y) + \mathbb{V}(Y).}$
- **Loi de Bernoulli $\mathscr{B}(p)$**

$$\boxed{\mathbb{P}(X=1)=p \qquad \mathbb{P}(X=0)=1-p=q \qquad \mathbb{E}(X)=p \qquad \mathbb{V}(X)=pq \qquad G_X(t)=q+pt}$$

- **Loi binomiale $\mathscr{B}(n,p)$**

$$\boxed{\forall\, k \in [\![0,n]\!],\ \ \mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k} \qquad \mathbb{E}(X)=np \qquad \mathbb{V}(X)=npq \qquad G_X(t)=(q+pt)^n}$$

- **Loi géométrique $\mathscr{G}(p)$**

$$\boxed{p\in]0,1[\qquad \forall\, n\in\mathbb{N}^*,\ \ \mathbb{P}(X=n)=pq^{n-1} \qquad \mathbb{E}(X)=\frac{1}{p} \qquad \mathbb{V}(X)=\frac{q}{p^2} \qquad G_X(t)=\frac{pt}{1-qt}}$$

- **Loi de Poisson $\mathscr{P}(\lambda)$**

$$\boxed{\lambda>0 \qquad \forall\, n\in\mathbb{N},\ \ \mathbb{P}(X=n)=\frac{\lambda^n}{n!}\mathrm{e}^{-\lambda} \qquad \mathbb{E}(X)=\lambda \qquad \mathbb{V}(X)=\lambda \qquad G_X(t)=\mathrm{e}^{\lambda(t-1)}}$$

- **Continuités croissante et décroissante**
  Si $(A_n)_{n\in\mathbb{N}}$ une suite croissante (pour l'inclusion)

$$\boxed{\mathbb{P}(A_k) \xrightarrow[k\to+\infty]{} \mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right)}$$

  Si $(A_n)_{n\in\mathbb{N}}$ une suite décroissante (pour l'inclusion)

$$\boxed{\mathbb{P}(A_k) \xrightarrow[k\to+\infty]{} \mathbb{P}\left(\bigcap_{n=0}^{+\infty} A_n\right)}$$

- **Inégalité de Markov** Si $a>0$, $\boxed{\mathbb{P}(|X|\geqslant a)\leqslant \dfrac{\mathbb{E}(|X|)}{a}.}$

- **Inégalité de Bienaymé-Tchebychev** Si $a>0$, $m=\mathbb{E}(X)$ et $\sigma=\sigma(X)=\sqrt{\mathbb{V}(X)}$. $\boxed{\mathbb{P}(|X-m|\geqslant a)\leqslant \dfrac{\sigma^2}{a^2}}$

- **Loi faible des grands nombres** Si $\varepsilon>0$, (X_n) une suite de vaiid L^2 d'espérance m, alors

$$\boxed{\mathbb{P}\left(\left|\frac{S_n}{n}-m\right|\geqslant\varepsilon\right)\xrightarrow[n\to+\infty]{}0.}$$

- **Inégalité de Cauchy-Schwarz** Si $X,Y\in L^2$, alors $XY\in L^1$, et

$$\boxed{(\mathbb{E}(XY))^2\leqslant \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right)}\qquad \boxed{|\mathrm{Cov}(X,Y)|\leqslant \mathbb{V}(X)\mathbb{V}(Y)}$$

avec égalité si et seulement si X et Y sont colinéaires presque sûrement.

XII Annexe : l'univers probabilisé du Pile-Face infini

Pour l'univers, pas de problème. On pose $\Omega=\{0,1\}^{\mathbb{N}^*}$.

Bien sûr, on pourrait remplacer $\{0,1\}$ par $\{P,F\}$ ou par $\{-1,1\}$. Dans la suite, on comprendra 1 comme « Pile », 0 comme « Face », on note alors que l'avantage du codage $0-1$ est que pour dénombrer les « Pile » on n'a qu'à faire la somme des issues. On note p un élément arbitraire de $]0,1[$ (qui désignera la probabilité d'obtenir Pile à un tirage quelconque).

Définition 29 : Événements de type fini

On appelle **évènement de type fini** (ou événement cylindrique) toute partie B de Ω telle qu'il existe $n\geqslant 1$ et $A\in\{0,1\}^n$, vérifiant

$$\omega\in B \iff (\omega_1,\dots,\omega_n)\in A.$$

Propriété 54 : Structure

L'ensemble $\mathscr{C}$ des évènements de type fini est une « algèbre » :

(*i*) $\Omega\in\mathscr{C}$

(*ii*) $(A,B)\in\mathscr{C}^2 \implies A\cup B\in\mathscr{C}$

(*iii*) $A\in\mathscr{C}\implies \overline{A}\in\mathscr{C}$

Remarque

R 78 – $\mathscr{C}$ est stable par réunion finie et par intersection finie.

On peut définir de manière naturelle ce que l'on voudrait être la probabilité d'événements ne portant que sur un nombre fini de lancers indépendants.

Définition 30 : Probabilité sur les événements de type fini

On pose, si $n\geqslant 1$ et si $(\varepsilon_1,\dots,\varepsilon_n)\in\{0,1\}^n$

$$\mathbb{P}\left(\{\omega\in\Omega,\ (\omega_1,\dots,\omega_n)=(\varepsilon_1,\dots,\varepsilon_n)\}\right)=p^{\varepsilon_1+\dots+\varepsilon_n}(1-p)^{n-(\varepsilon_1+\dots+\varepsilon_n)}.$$

La difficulté est de l'étendre à une tribu contenant tous les événements de type fini.

Propriété 55 : Extension

$\mathbb{P}$ s'étend de manière unique à $\mathscr{C}$ en une application que l'on notera encore $\mathbb{P}$, et qui vérifie

(*i*) $\mathbb{P}(\Omega) = 1$

(*ii*) $\forall (A, B) \in \mathscr{C}^2,\ A \cap B = \varnothing \Rightarrow P(A \sqcup B) = P(A) + P(B)$

Propriété 56 : σ-additivité

$\mathbb{P}$ a la propriété plus forte que (ii) suivante :

(ii') Pour toute suite $\forall (A_n)_{n\in\mathbb{N}} \in \mathscr{C}^{\mathbb{N}}$, si les A_n sont deux-à-deux disjoints et si $\bigcup_{n=0}^{+\infty} A_n \in \mathscr{C}$, alors

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

Théorème 4 : Existence et unicité de la probabilité

Il existe une unique probabilité sur la tribu $\mathscr{A}$ engendrée par $\mathscr{C}$ qui prolonge $\mathbb{P}$.

Propriété 57

$\mathscr{A} \subsetneq \mathscr{P}(\Omega)$: il y a des parties de Ω qui n'ont pas de probabilité.

- La propriété 56 n'est pas trop facile.
- Le théorème 4 non plus : c'est le théorème de Caratheodory.
  Elle utilise la notion de tribu engendrée qui, elle, ne présente pas de difficulté : on vérifie que l'intersection des tribus contenant une partie de $\mathscr{P}(\Omega)$ donnée est une tribu.
- La propriété 57 est décevante, car on aimerait bien exhiber des telles parties. Or pour montrer leur existence, on a besoin du célèbre Axiome du Choix, on est donc en pleine théorie des ensembles...remarquons que c'est cela qui oblige à s'occuper de tribus : si on pouvait définir les probabilités, à chaque fois, sur $\mathscr{P}(\Omega)$, la notion de tribu d'évènements serait moins nécessaire.