CH18 : Variables aléatoires discrètes

Table des matières

Dans tout ce chapitre $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé (pas nécessairement fini ou dénombrable).

I Généralités

1) Définition

Définition 1 (Variable aléatoire discrète)
On appelle **variable aléatoire discrète** *(en abrégé* **VAD***) définie sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *et à valeurs dans un ensemble* E *toute application* $X : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & E \\ \omega & \longmapsto & X(\omega) \end{array} \right.$ *vérifiant :*

- (i) *l'image* $X(\Omega)$ *est finie ou dénombrable ;*
- (ii) *pour tout* $x \in E$*, l'ensemble* $X^{-1}(\{x\}) = \{\omega \in \Omega,\ X(\omega) = x\}$ *est élément de la tribu* $\mathcal{A}$*.*

Vocabulaire
Lorsque $E = \mathbb{R}$, on parle de **VAD réelle**.
Lorsque $E = \mathbb{C}$, on parle de **VAD complexe**.
Lorsque $E = \mathbb{K}^n$ avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ et $n \geq 2$, on parle de **VAD vectorielle**, ou encore de **vecteur aléatoire**.

Remarque
Dans le point (ii)*, on peut se limiter aux* $x \in X(\Omega)$*, puisque si* $x \in E \setminus X(\Omega)$*, on a* $X^{-1}(\{x\}) = \emptyset$*, qui est automatiquement dans la tribu* $\mathcal{A}$*.*

ATTENTION !
La terminologie peut induire en erreur : une variable aléatoire discrète n'est pas une variable au sens mathématique, mais bien une fonction. Et cette fonction n'est pas aléatoire, elle est bien déterminée en fonction des éventualités $\omega \in \Omega$ possibles.
"L'aléatoire est caché" dans l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel X est définie.

Exemple (Lancer de deux dés)
On lance simultanément deux dés équilibrés. On effectue la somme des faces obtenues.

On peut modéliser cette expérience par une variable aléatoire réelle discrète S sur l'univers $\Omega = \{1, \cdots, 6\}^2$, muni de l'équiprobabilité $\mathbb{P}$ et $S : \Omega \to \{2, \ldots, 12\}$ définie par $S(n, m) = n + m$ pour tout $(n, m) \in \{1, \ldots, 6\}^2$.
Ici, Ω est fini (36 éléments) et l'image $S(\Omega)$ est finie (11 éléments).

Exemple (Lancers successifs d'une pièce)
On lance indéfiniment une pièce équilibrée. On note le numéro de lancer où *Pile* apparaît pour la première fois.

Soit Ω l'ensemble des suites $(a_n)_{n \in \mathbb{N}^*}$ à valeurs dans l'ensemble $\{P, F\}$. Cet univers est infini et indénombrable.
On admet qu'on peut munir cet univers d'une probabilité $\mathbb{P}$ (ça n'est pas évident car Ω étant infini non dénombrable, on ne peut pas définir une probabilité $\mathbb{P}$ en attribuant simplement un poids à chaque singleton $\{w\}$, comme c'était le cas pour les univers au plus dénombrables).
On pose alors $X : \Omega \to \mathbb{N}^* \cup \{+\infty\}$ définie par $X((a_n)_{n \in \mathbb{N}^*}) = \inf\{n \in \mathbb{N}^*,\ a_n = P\}$.
Ici, l'image $X(\Omega) = \mathbb{N}^* \cup \{+\infty\}$ est infinie et dénombrable, car en bijection avec $\mathbb{N}$.

Remarque
En général, l'espace probabilisable $(\Omega, \mathcal{A})$ *sur lequel une variable aléatoire* X *est définie importe peu, ce qui va compter est l'image* $X(\Omega)$ *et les "probabilités d'apparition" des différentes valeurs* $x \in X(\Omega)$ *(cf. plus loin la notion de loi), qui sont bien entendu reliées à* $\mathbb{P}$*.*

2) Evénements valeurs

Notation (Evénement valeur)
Soit $X : \Omega \to E$ une variable aléatoire discrète. Pour tout $x \in E$, on note $(X = x)$ (ou $\{X = x\}$, ou encore $[X = x]$) l'événement :

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

Par définition d'une VAD, on a bien $(X = x) \in \mathcal{A}$, donc on pourra calculer la probabilité de cet événement, notée $\mathbb{P}(X = x)$ (ou $\mathbb{P}(\{X = x\})$, ou encore $\mathbb{P}([X = x])$).

Notation (Image réciproque d'une partie)
Soit $X : \Omega \to E$ une variable aléatoire discrète. Pour toute partie $A \subset E$, on note $(X \in A)$ (ou $\{X \in A\}$, ou encore $[X \in A]$) l'image réciproque de A :

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

Propriété 2 (L'image réciproque d'une partie est un événement)
Soit $X : \Omega \to E$ une variable aléatoire discrète. Pour toute partie $A \subset E$, $(X \in A) \in \mathcal{A}$.

Preuve
Soit $A \subset E$. Puisque $X(\Omega)$ est au plus dénombrable par hypothèse, l'ensemble $X^{-1}(A)$ est réunion dénombrable d'événements valeurs :

$$(X \in A) = \{\omega \in \Omega,\ X(\omega) \in A\} = \bigcup_{x \in A} \{\omega \in \Omega,\ X(\omega) = x\} = \bigcup_{x \in A} (X = x) = \bigcup_{x \in X(\Omega) \cap A} (X = x),$$

(puisque $(X = x)$ est vide si $x \notin X(\Omega)$) donc il s'agit aussi d'un événement (puisque la tribu $\mathcal{A}$ est stable par réunion finie ou dénombrable et puisque $X(\Omega) \cap A$ est au plus dénombrable en tant que partie de $X(\Omega)$).

Remarque
- *On pourra donc calculer $\mathbb{P}(X \in A)$ pour toute partie $A \subset E$.*
- *La notation $(X \in A)$ est compatible avec les opérations ensemblistes :*

$$(X \in A) \cap (X \in B) = X^{-1}(A) \cap X^{-1}(B) = X^{-1}(A \cap B) = (X \in A \cap B),$$
$$(X \in A) \cup (X \in B) = X^{-1}(A) \cup X^{-1}(B) = X^{-1}(A \cup B) = (X \in A \cup B),$$
$$\overline{(X \in A)} = \Omega \setminus X^{-1}(A) = X^{-1}(E \setminus A) = (X \in \overline{A}).$$

Notation (Evénément défini par une inégalité)
Si $X : \Omega \to \mathbb{R}$ est une variable aléatoire réelle discrète et si $a \in \mathbb{R}$, alors on note

$$(X \leq a) = (X \in]-\infty, a]) = X^{-1}(]-\infty, a]) = \{\omega \in \Omega,\ X(\omega) \leq a\}.$$

On notera de même :

$$(X < a) = \{\omega \in \Omega,\ X(\omega) < a\},$$
$$(X \geq a) = \{\omega \in \Omega,\ X(\omega) \geq a\},$$
$$(a \leq X \leq b) = X^{-1}([a, b]) = (X \leq b) \cap (X \geq a), \quad \ldots$$

et toutes ces parties sont bien des événements de la tribu $\mathcal{A}$.

Exemple (Lancer de deux dés)
On additionne les résultats de deux dés équilibrés.
On a la variable aléatoire "somme" $S : \{1, \cdots, 6\}^2 \to \{2, \cdots, 12\}$.
Décrivons les événements $A = (S = 4)$ et $B = (9 \leq S < 11)$.

On a

$$A = (S = 4) = \{(1, 3), (2, 2), (3, 1)\},$$
$$B = (S = 9) \cup (S = 10) = \{(3, 6), (4, 5), (5, 4), (6, 3), (4, 6), (5, 5), (6, 4)\}.$$

3) Loi d'une variable aléatoire discrète

Idée : Etant donné une variable aléatoire discrète $X : \Omega \to E$, on va munir l'ensemble image $X(\Omega)$ (qui est fini ou dénombrable) de la tribu $\mathcal{A} = \mathcal{P}(X(\Omega))$ et on va définir une probabilité sur cet espace probabilisable "à l'arrivée".

Définition 3 (Loi d'une variable aléatoire discrète)
Étant donnés un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une variable aléatoire discrète $X : \Omega \to E$, on appelle **loi de X** *la fonction*

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

Propriété 4 (La loi est une probabilité sur l'image)
Soit X une VAD définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans E.
Alors, la loi $\mathbb{P}_X$ est une probabilité sur l'espace probabilisable $(X(\Omega), \mathcal{P}(X(\Omega)))$.

Vocabulaire
Ainsi, on dit également que $\mathbb{P}_X$ est la **loi de probabilité** de X.

Preuve
- $\mathbb{P}_X$ est à valeurs dans $\mathbb{R}^+$ car $\mathbb{P}$ l'est.
- $\mathbb{P}_X\big(X(\Omega)\big) = \mathbb{P}\big(X \in X(\Omega)\big) = \mathbb{P}(\Omega) = 1$ car $\forall \omega \in \Omega,\ X(\omega) \in X(\Omega)$.
- Soit $(A_n)_{n\in\mathbb{N}}$ une suite de parties deux à deux disjointes de $X(\Omega)$.
  Les événements $(X \in A_n)$ sont deux à deux disjoints – car chaque valeur $X(\omega)$ est dans un seul des A_n – et
  $$\left(X \in \bigcup_{n\in\mathbb{N}} A_n\right) = \{\omega \in \Omega,\ \exists n \in \mathbb{N},\ X(\omega) \in A_n\} = \bigcup_{n\in\mathbb{N}} (X \in A_n),$$
  donc, par σ-additivité de la probabilité $\mathbb{P}$, on obtient
  $$\mathbb{P}_X\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}\left(X \in \bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(X \in A_n) = \sum_{n=0}^{+\infty} \mathbb{P}_X(A_n).$$

Remarque
- *Pour une variable aléatoire discrète $X : \Omega \to E$, on a donc* **deux mesures de probabilité** *: la probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{A})$ ("au départ"), et la loi $\mathbb{P}_X$ sur l'image $(X(\Omega), \mathcal{P}(X(\Omega)))$.*
- *Une fois qu'on a l'espace probabilisé $\Big(X(\Omega), \mathcal{P}(X(\Omega)), \mathbb{P}_X\Big)$, on peut très bien "oublier" sur quel espace $(\Omega, \mathcal{A}, \mathbb{P})$ a été introduit X. C'est pour cela que, régulièrement, les énoncés ne s'embêtent pas à décrire l'espace probabilisé sur lequel X est définie, et se contentent de dire qu'on admet qu'il en existe un …*

Corollaire 5 (Détermination de la loi d'une variable aléatoire discrète)
Soit $X : \Omega \to E$ une variable aléatoire discrète. La loi de probabilité $\mathbb{P}_X$ est uniquement déterminée par la distribution de probabilités discrètes $(p_x)_{x\in X(\Omega)} = (\mathbb{P}(X = x))_{x\in X(\Omega)}$.

Preuve
L'univers $X(\Omega)$ étant au plus dénombrable, une probabilité sur celui-ci est entièrement déterminée par ses probabilités élémentaires :

$$p_x = \mathbb{P}_X(\{x\}) = \mathbb{P}(X \in \{x\}) = \mathbb{P}(X = x)$$

(voir chapitre précédent sur les espaces probabilisés).
En effet, pour toute partie $A \in \mathcal{P}(X(\Omega))$, on a alors

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \sum_{x\in A} p_x$$

(ce qui a bien du sens puisque la famille $(p_x)_{x\in X(\Omega)}$ est sommable et de somme 1).

Remarque (Prolongement de la loi $\mathbb{P}_X$ à E tout entier)
Au besoin, on peut considérer que la loi $\mathbb{P}_X$ est une probabilité sur $(E, \mathcal{P}(E))$, en posant $p_x = 0$ si $x \in E \setminus X(\Omega)$.
L'image $X(\Omega)$ apparaît alors comme le support de la distribution de probabilités discrètes $(p_x)_{x\in E}$, c'est-à-dire l'ensemble des $x \in E$ tels que $p_x \neq 0$ (nécessairement au plus dénombrable pour avoir la sommabilité de la famille, voir chapitre 1).
Ce procédé de prolongement peut-être utile si par exemple l'image $X(\Omega)$ est difficile à déterminer, on travaille alors sur un "sur-ensemble" E, en rajoutant des probabilités nulles.

Méthode
Pour déterminer la loi d'une variable aléatoire discrète $X : \Omega \to E$:

- *on commence par déterminer $X(\Omega)$ (l'ensemble des valeurs prises par X).*
- *on calcule ensuite $\mathbb{P}(X = x)$ pour toute valeur $x \in X(\Omega)$.*

Exemple (Somme des résultats des lancers de deux dés)
La loi de $S : \{1, \cdots, 6\}^2 \to \{2, \cdots, 12\}$ définie par $S(\omega_1, \omega_2) = \omega_1 + \omega_2$ est donnée par le tableau suivant :

x	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(S = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Théorème 6 (Existence d'une VAD de distribution de probabilités donnée)
Soit E un ensemble au plus dénombrable et soit $(p_x)_{x\in E}$ une distribution de probabilités sur E (famille sommable de réels positifs, de somme 1). Alors il existe un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une VAD $X : \Omega \to E$ de loi donnée par $(p_x)_{x\in E}$.

Preuve
Posons $(\Omega, \mathcal{A}) = (E, \mathcal{P}(E))$, et considérons la probabilité sur $(E, \mathcal{P}(E))$ définie par

$$\forall x \in E, \qquad \mathbb{P}(\{x\}) = p_x.$$

La fonction $X = Id_E : E \to E$ est alors une VAD qui convient.

Définition 7 (Suivre la même loi)
Soit X une VAD définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et Y une VAD définie sur $(\Omega', \mathcal{A}', \mathbb{P}')$, avec X et Y à valeurs dans un même ensemble E.
On dit que X et Y **suivent la même loi** *si $X(\Omega) = Y(\Omega)$ et si $\mathbb{P}_X = \mathbb{P}_Y$.*
Cela revient à dire que $X(\Omega) = Y(\Omega)$ et $\forall x \in X(\Omega)$, $\mathbb{P}(X = x) = \mathbb{P}'(Y = x)$.

Notation
On notera $X \sim Y$ pour dire que deux variables aléatoires discrètes X et Y suivent la même loi.
De même, si une variable X suit une loi usuellement notée $\mathcal{L}$ (par exemple $\mathcal{U}(\{1, \cdots, n\})$ pour la loi uniforme sur l'ensemble fini $\{1, \cdots, n\}$), alors on écrira directement $X \sim \mathcal{L}$.

Remarque

- *X et Y ne sont pas nécessairement définies sur le même espace probabilisé, seules les valeurs comptent.*
- *Si X et Y sont définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, alors*

$$X \sim Y \iff X(\Omega) = Y(\Omega) \text{ et } \forall x \in X(\Omega),\ \mathbb{P}(X = x) = \mathbb{P}(Y = x).$$

ATTENTION !
A toute variable aléatoire correspond une unique loi mais des variables aléatoires différentes peuvent avoir la même loi.

Exemple (Lancers successifs d'une pièce)
On considère une pièce équilibrée que l'on lance 10 fois. Le nombre X de *Pile* obtenus et le nombre Y de *Face* obtenus suivent la même loi mais ne sont pas identiques.

Exemple (Loi de Rademacher)
Si $X(\Omega) = \{-1, 1\}$ et $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$, alors $X \sim -X$.

Remarque (IMPORTANTE)
Pour toute VAD $X : \Omega \to E$, la famille d'événements $(X = x)_{x \in X(\Omega)}$ est un SCE.
En effet, $x \neq y \implies (X = x) \cap (X = y) = \emptyset$ et $\Omega = \bigcup_{x \in X(\Omega)} X^{-1}(\{x\}) = \bigcup_{x \in X(\Omega)} (X = x)$.
La famille $(X = x)_{x \in E}$ est également un SCE, mais comportant des événements vides (les $x \notin X(\Omega)$).

4) Lois usuelles à image finie (révisions MP2I)

Définition 8 (Loi uniforme)
Etant donné un ensemble fini non vide E, on dit qu'une variable aléatoire discrète X définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ suit **la loi uniforme sur $\boldsymbol{E}$** *lorsque*

$$X(\Omega) = E, \qquad \forall x \in E, \quad \mathbb{P}(X = x) = \frac{1}{Card(E)}.$$

On note alors $X \sim \mathcal{U}(E)$.

Remarque (Lien avec l'équiprobabilité)
$X \sim \mathcal{U}(E)$ ssi tous les singletons $\{x\}$ de l'ensemble fini E ont même probabilité $\mathbb{P}_X(\{x\})$.
La loi uniforme sur E est donc l'équiprobabilité sur E.

Exemple

- Si on lance un dé équilibré et on note X le résultat obtenu, alors X suit la loi uniforme sur $\Omega = \{1, \cdots, 6\}$.
- La variable "somme de deux dés équilibrés" S ne suit pas une loi uniforme, puisque par exemple, $\mathbb{P}(S = 2) \neq \mathbb{P}(S = 3)$.

Définition 9 (Loi de Bernoulli)
Soit un réel $p \in [0; 1]$. On dit qu'une variable aléatoire discrète X définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ suit la **loi de Bernoulli de paramètre $\boldsymbol{p}$** *lorsque*

$$X(\Omega) = \{0; 1\}, \qquad \mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

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

Remarque (Schéma "succès-échec")
Si l'on prend comme convention que $X = 1$ désigne le succès d'une expérience et $X = 0$ l'échec de celle-ci, alors cette loi est naturellement celle codant la probabilité de succès d'une expérience.

Exemple (Pile ou Face)
Si $\Omega = \{P, F\}$ et si l'on définit $X : \Omega \to \{0, 1\}$ par $X(P) = 1$ et $X(F) = 0$, alors X code la probabilité d'obtention d'un *Pile*.

- Si la pièce est équilibrée, alors X suit la loi de Bernoulli $\mathcal{B}\left(\frac{1}{2}\right)$, qui coïncide dans ce cas particulier avec la loi uniforme $\mathcal{U}\big(\{0, 1\}\big)$.
- Si on suppose que *Pile* a une probabilité de sortir de $\dfrac{2}{3}$, alors $X \sim \mathcal{B}\left(\frac{2}{3}\right)$.

Définition 10 (Loi binomiale)
Soient un entier $n \in \mathbb{N}$ et un réel $p \in [0, 1]$. On dit qu'une variable aléatoire discrète X définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ suit la **loi binomiale de paramètres $\boldsymbol{(n, p)}$** *lorsque*

$$X(\Omega) = \{0, \cdots, n\}, \qquad \forall k \in \{0, \ldots, n\},\ \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

On note alors $X \sim \mathcal{B}(n, p)$.

Remarque

- *On a $\sum\limits_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$ donc on a bien $\sum\limits_{0 \leq k \leq n} \mathbb{P}(X = k) = 1$, ce qui confirme que $\mathbb{P}_X$ est une loi de probabilité sur $\{0, 1, \cdots, n\}$.*
- *La loi binomiale $\mathcal{B}(1, p)$ coïncide avec la loi de Bernoulli $\mathcal{B}(p)$.*

Théorème 11 (Réalisation concrète de la loi binomiale)
Soit X la variable aléatoire comptant le **nombre de succès** *dans la répétition de n expériences aléatoires identiques à deux issues, supposées indépendantes, et de même probabilité de succès $p \in [0;1]$. Alors X suit la loi binomiale $\mathcal{B}(n,p)$.*

Preuve
Les valeurs de X sont comprises dans $\{0,\dots,n\}$.
Pour chaque indice $i \in \{1,\cdots,n\}$ notons A_i l'événement "la i^e répétition de l'expérience a donné un succès". Les événements $A_1,\cdots,A_n$ sont indépendants et chacun de probabilité p.
Pour $k \in \{0,\cdots,n\}$ fixé, décomposons l'événement $(X=k)$:
$(X=k)$ est réalisé ssi exactement k événements parmi $A_1,\cdots,A_n$ sont réalisés et $n-k$ ne le sont pas.
Formellement, cela s'écrit :

$$(X=k) = \bigcup_{1\leq i_1<\cdots<i_k\leq n} \left(A_{i_1} \cap \cdots \cap A_{i_k} \cap \left(\bigcap_{j\notin\{i_1,\cdots,i_k\}} \overline{A_j} \right) \right)$$

$(X=k)$ est donc réunion disjointe de $\binom{n}{k}$ événements de même probabilité $p^k(1-p)^{n-k}$, donc

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

Exemple
On lance une pièce équilibrée n fois. Le nombre X de *Pile* obtenus suit la loi binomiale $\mathcal{B}(n,p)$ avec $p=\frac{1}{2}$.

5) Loi géométrique, absence de mémoire

Définition 12 (Loi géométrique)
Soit un réel $p \in]0;1[$. On dit qu'une variable aléatoire discrète X définie sur $(\Omega,\mathcal{A},\mathbb{P})$ suit la **loi géométrique de paramètre p** *lorsque*

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

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

<u>ATTENTION !</u>
Ici, p est différent de 0 et de 1.

Remarque
On a bien

$$\sum_{k=1}^{+\infty} \mathbb{P}(X=k) = \sum_{k=1}^{+\infty} (1-p)^{k-1}p = p\sum_{k=0}^{+\infty}(1-p)^k = p \times \frac{1}{1-(1-p)} = 1,$$

ce qui confirme que la loi géométrique est une probabilité sur $\mathbb{N}^*$.

Théorème 13 (Réalisation concrète de la loi géométrique)
Soit X la variable aléatoire donnant le **rang du premier succès** *dans la répétition d'expériences aléatoires identiques à deux issues, supposées indépendantes, et de même probabilité de succès $p \in]0;1[$. Alors X suit la loi géométrique $\mathcal{G}(p)$.*

Preuve
Pour chaque indice $i \in \mathbb{N}^*$ notons A_i l'événement "la i^e répétition de l'expérience a donné un succès".
Les événements $(A_i)_{i\geq 1}$ sont indépendants et chacun de probabilité p.
Pour $k \in \mathbb{N}^*$ fixé, on a

$$(X=k) = \overline{A_1} \cap \cdots \cap \overline{A_{k-1}} \cap A_k,$$

donc par indépendance :

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

Théorème 14 (Caractérisation des lois sans mémoire)

(i) Soit $p\in]0,1[$. Si $X\sim\mathcal{G}(p)$, alors X est "sans mémoire", c'est-à-dire

$$\forall (n,k)\in\mathbb{N}^2,\qquad \mathbb{P}_{(X>n)}(X>n+k)=\mathbb{P}(X>k).$$

(ii) Réciproquement, si une variable aléatoire $X:\Omega\to\mathbb{N}^*$ vérifie la condition d'absence de mémoire, c'est-à-dire

$$\forall (n,k)\in\mathbb{N}^2,\qquad \mathbb{P}_{(X>n)}(X>n+k)=\mathbb{P}(X>k),$$

alors X suit une loi géométrique de paramètre $p=\mathbb{P}(X=1)$.

Preuve

(i) Si $X\sim\mathcal{G}(p)$, alors

$$\forall k\in\mathbb{N},\qquad \mathbb{P}(X>k)=\mathbb{P}\left(\bigcup_{l=k+1}^{+\infty}(X=l)\right)=\sum_{l=k+1}^{+\infty}(1-p)^{l-1}p=(1-p)^k\neq 0,$$

donc par définition des probabilités conditionnelles :

$$\forall (n,k)\in\mathbb{N}^2,\qquad \mathbb{P}_{(X>n)}(X>n+k)=\frac{\mathbb{P}((X>n)\cap(X>n+k))}{\mathbb{P}(X>n)}=\frac{\mathbb{P}(X>n+k)}{\mathbb{P}(X>n)},$$

d'où

$$\forall (n,k)\in\mathbb{N}^2,\qquad \mathbb{P}_{(X>n)}(X>n+k)=\frac{(1-p)^{n+k}}{(1-p)^n}=(1-p)^k=\mathbb{P}(X>k).$$

(ii) Par hypothèse, on a $\mathbb{P}(X>n)\neq 0$ pour tout $n\in\mathbb{N}$, ainsi que :

$$\forall (n,k)\in\mathbb{N}^2,\qquad \mathbb{P}(X>n+k)=\mathbb{P}_{(X>n)}(X>n+k)\mathbb{P}(X>n)=\mathbb{P}(X>k)\mathbb{P}(X>n),$$

donc en particulier

$$\forall n\in\mathbb{N},\qquad \mathbb{P}(X>n+1)=\mathbb{P}(X>1)\mathbb{P}(X>n),$$

et par récurrence :

$$\forall n\in\mathbb{N},\qquad \mathbb{P}(X>n)=\mathbb{P}(X>1)^n=(1-p)^n,$$

où $p=\mathbb{P}(X=1)$ (puisque $\mathbb{P}(X>0)=1$ et étant donné que $X(\Omega)\subset\mathbb{N}^*$).
Enfin, pour tout $n\in\mathbb{N}^*$:

$$\mathbb{P}(X=n)=\mathbb{P}((X>n-1)\setminus(X>n))=\mathbb{P}(X>n-1)-\mathbb{P}(X>n)=(1-p)^{n-1}-(1-p)^n=(1-p)^{n-1}p,$$

ce qui montre bien que $X\sim\mathcal{G}(p)$.

6) Loi de Poisson, approximation de la loi binomiale

Définition 15 (Loi de Poisson)
Soit un réel $\lambda>0$. On dit qu'une variable aléatoire discrète X définie sur $(\Omega,\mathcal{A},\mathbb{P})$ suit **la loi de Poisson de paramètre $\boldsymbol{\lambda}$** *lorsque*

$$X(\Omega)=\mathbb{N},\qquad \forall k\in\mathbb{N},\ \mathbb{P}(X=k)=e^{-\lambda}\times\frac{\lambda^k}{k!}.$$

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

<u>**ATTENTION !**</u>
Bien noter que le paramètre λ est strictement positif.

Remarque
On a bien les $\mathbb{P}(X=k) \geq 0$ *et*

$$\sum_{k=0}^{+\infty} \mathbb{P}(X=k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \times e^{\lambda} = 1$$

(en utilisant le DSE de e^{λ}*), ce qui confirme que la loi de Poisson est bien une probabilité sur* $\mathbb{N}$*.*

Le résultat suivant est essentiel pour comprendre l'intérêt de la loi de Poisson :

Théorème 16 (Approximation de la loi binomiale par la loi de Poisson)
Soient un réel $\lambda > 0$*, une suite* $(p_n)_{n\in\mathbb{N}}$ *à valeurs dans* $[0,1]$ *et* $(X_n)_{n\in\mathbb{N}}$ *une suite de VAD définies sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *avec* $X_n \sim \mathcal{B}(n, p_n)$*.*
Si $n \times p_n \underset{n\to+\infty}{\longrightarrow} \lambda$*, alors* $\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \underset{n\to+\infty}{\longrightarrow} e^{-\lambda} \times \dfrac{\lambda^k}{k!}$.

Preuve
Soit $k \in \mathbb{N}$. On a, pour tout entier $n \geq k$,

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!}\left(\frac{n!}{(n-k)!}\right) p_n^k (1-p_n)^{n-k}.$$

Comme k est fixé, on a

$$\frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1) \underset{n\to+\infty}{\sim} n^k.$$

Donc

$$\frac{n!}{(n-k)!} p_n^k \underset{n\to+\infty}{\sim} n^k p_n^k = (np_n)^k \underset{n\to+\infty}{\sim} \lambda^k$$

(puisque $\lim\limits_{n\to+\infty} np_n = \lambda$ par hypothèse). Enfin, l'hypothèse sur p_n se réécrit $p_n \underset{n\to+\infty}{\sim} \frac{\lambda}{n}$ donc on a le développement limité $p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right)$ et

$$(1-p_n)^{n-k} = e^{(n-k)\ln(1-p_n)} = e^{(n-k)\left(-\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)} = e^{(-\lambda + o(1))} \underset{n\to+\infty}{\sim} e^{-\lambda}.$$

Donc, par produit d'équivalents, $\mathbb{P}(X_n = k) \underset{n\to+\infty}{\sim} e^{-\lambda}\dfrac{\lambda^k}{k!}$, ce qui donne le résultat.

Remarque

- *Ainsi, dans une situation avec* $\boldsymbol{n}$ **grand et** $\boldsymbol{p}$ **petit***, on a intérêt à* ***approcher la loi binomiale*** $\boldsymbol{\mathcal{B}(n,p)}$ ***par la loi de Poisson*** $\boldsymbol{\mathcal{P}(np)}$ *car les calculs sont bien plus commodes pour la loi de Poisson (ne serait-ce que calculer le coefficient binomial* $\binom{n}{k}$ *peut présenter de grosses difficultés pour certaines valeurs de k).*
- *Mais surtout, ce théorème offre une* **interprétation de la loi de Poisson***.*
  Elle s'apparente à un nombre de réussites lors de n essais indépendants suivant une loi de Bernoulli de paramètre p, pour n grand et p petit.
  La loi de Poisson s'apparente donc à la loi des événements rares*.*

  Rentrent dans cette catégorie, par exemple :
  * *le nombre de personnes centenaires dans une population.*
  * *le nombre de coquilles par page dans un livre.*
  * *le nombre de faux numéros composés sur un laps de temps.*
  * *le nombre de particules désintégrées par radioactivité lors d'une période de temps fixée ...*
- *Ce type de convergence s'appelle la* **convergence en loi***. Il en existe d'autres (convergence presque sûre, convergence* L^p*, convergence en probabilité, non étudiées en CPGE).*

7) Variables aléatoires composées

Propriété 17 (Composition d'une variable aléatoire discrète par une fonction)
Soit $X : \Omega \to E$ une variable aléatoire discrète, et soit $f : X(\Omega) \to E'$ une application quelconque. Alors la fonction composée

$$Y = f \circ X : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & E' \\ \omega & \longmapsto & Y(\omega) = f(X(\omega)) \end{array} \right.$$

est également une variable aléatoire discrète.

Notation
Traditionnellement, cette variable aléatoire composée sera notée $f(X)$ plutôt que $f \circ X$.
C'est ainsi que l'on pourra écrire X^2, $\sqrt{X}$, $|X|$, $aX + b$ (si $E = \mathbb{K}$), $\|X\|$ (si E est un evn), ...

Preuve
Tout d'abord $Y(\Omega) = f(X(\Omega))$ est au plus dénombrable car c'est l'image par une application de l'ensemble fini ou dénombrable $X(\Omega)$: il existe une surjection $s : \mathbb{N} \to X(\Omega)$, donc la composée $f \circ s : \mathbb{N} \to Y(\Omega)$ est une surjection.
Ensuite, pour tout élément $y \in E'$, on a

$$Y^{-1}(\{y\}) = \{\omega \in \Omega,\ f(X(\omega)) = y\} = (X \in B)$$

en notant $B = f^{-1}(\{y\}) \in \mathcal{P}(E)$. D'après la prop. 2, on a donc $Y^{-1}(\{y\}) \in \mathcal{A}$ car X est une VAD.

Propriété 18 (Détermination de la loi de $f(X)$ par la loi de X)
Si $Y = f(X)$, alors la loi de Y est entièrement déterminée par celle de X, via la relation :

$$\forall B \in \mathcal{P}(Y(\Omega)), \qquad \mathbb{P}_Y(B) = \mathbb{P}_X(f^{-1}(B)).$$

Remarque
Cela s'énonce aussi en disant :

$$X_1 \sim X_2 \implies f(X_1) \sim f(X_2).$$

Preuve
Par définition d'une loi, nous avons, pour toute partie $B \in \mathcal{P}(Y(\Omega))$:

$$\mathbb{P}_Y(B) = \mathbb{P}(Y \in B) = \mathbb{P}(f(X) \in B) = \mathbb{P}(X \in f^{-1}(B)) = \mathbb{P}_X(f^{-1}(B)).$$

Remarque
Cette relation est également vraie pour toute partie $B \in \mathcal{P}(E')$.

Exemple

- Si $X \sim \mathcal{B}(p)$, alors pour tout $k \in \mathbb{N}^*$, on a $X^k \sim \mathcal{B}(p)$.
  En effet, $X^k(\Omega) = \{0, 1\}$ et $\mathbb{P}(X^k = 1) = \mathbb{P}(X = 1) = p$.
- Si $X \sim \mathcal{B}(n, p)$, alors $Y = n - X \sim \mathcal{B}(n, 1 - p)$.
  En effet, $Y(\Omega) = (n - X)(\Omega) = \{0, \cdots, n\}$ et pour tout $k \in \mathbb{N}$:

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

- Si $X(\Omega) = \{a, b\}$ (avec $a \neq b$ réels) et $\mathbb{P}(X = a) = p$, alors $Y = \frac{X-b}{a-b} \sim \mathcal{B}(p)$.
  Cela signifie que toute VAD binaire se ramène par homothétie et translation à une loi de Bernoulli.

II Couples de variables aléatoires discrètes

$(\Omega, \mathcal{A}, \mathbb{P})$ désigne toujours un espace probabilisé et E et F deux ensembles quelconques.

1) Loi conjointe

Définition 19 (Couple de deux variables aléatoires)
Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
On appelle **couple défini par X et Y** *la fonction*

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

Propriété 20 (Un couple de VAD est une VAD)
Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
Alors, le couple $Z = (X, Y) : \Omega \to E \times F$ est également une variable aléatoire discrète.

Vocabulaire
On dit aussi que (X, Y) est le **vecteur aléatoire discret** de composantes X, Y.

Preuve
On a $Z(\Omega) \subset X(\Omega) \times Y(\Omega)$, et l'ensemble $X(\Omega) \times Y(\Omega)$ est au plus dénombrable (comme produit fini d'ensembles au plus dénombrables) donc $Z(\Omega)$ est également au plus dénombrable.
Ensuite, pour tout $(x, y) \in E \times F$, on a

$$(Z = (x, y)) = Z^{-1}(\{(x, y)\}) = \{\omega \in \Omega,\ X(\omega) = x,\ Y(\omega) = y\} = (X = x) \cap (Y = y) \in \mathcal{A}$$

(car X, Y sont des VAD et la tribu $\mathcal{A}$ est stable par intersection).

ATTENTION !
Bien noter qu'en général,

$$(X, Y)(\Omega) \subset X(\Omega) \times Y(\Omega)$$

et que l'inclusion peut être stricte.
Par exemple, si X (resp. Y) désigne le nombre de *Pile* (resp. de *Face*) obtenus lors de 2 lancers indépendants d'une même pièce équilibrée, alors $X \sim Y \sim \mathcal{B}(2, 1/2)$, et

$$(X, Y)(\Omega) = \{(0, 2), (1, 1), (2, 0)\} \subsetneq \{0, 1, 2\}^2 = X(\Omega) \times Y(\Omega).$$

Cela vient du fait que $X + Y = 2$ (la somme des deux VAD est constante).

Définition 21 (Loi conjointe)
On appelle **loi conjointe** *de X et Y la loi du couple $Z = (X, Y)$, définie par*

$$\mathbb{P}_{(X,Y)} : \left\{ \begin{array}{ccc} \mathcal{P}(E \times F) & \longrightarrow & \mathbb{R}^+ \\ A & \longmapsto & \mathbb{P}_{(X,Y)}(A) = \mathbb{P}((X, Y) \in A) \end{array} \right. .$$

Propriété 22 (Détermination de la loi conjointe par les couples élémentaires)
La loi conjointe de X et Y est entièrement déterminée par la distribution de probabilités discrètes

$$(p_{x,y})_{(x,y) \in X(\Omega) \times Y(\Omega)} = (\mathbb{P}((X, Y) = (x, y)))_{(x,y) \in X(\Omega) \times Y(\Omega)}.$$

Notation
Pour $(x, y) \in E \times F$ on emploiera également la notation $\mathbb{P}(X = x, Y = y)$ pour désigner la probabilité $\mathbb{P}((X, Y) = (x, y))$ (qui est aussi $\mathbb{P}((X = x) \cap (Y = y))$).

Preuve
On applique le corollaire 5 avec $E \times F$ à la place de E (les éléments sont donc des couples $z = (x, y)$

avec $x \in E$ et $y \in F$).
La loi de $Z = (X, Y) : \Omega \to E \times F$ est donc déterminée par les $(p_z)_{z \in Z(\Omega)} = (\mathbb{P}(Z = z))_{z \in Z(\Omega)}$, et on peut prolonger cette distribution de probabilités à $X(\Omega) \times Y(\Omega)$ (car cet ensemble contient $Z(\Omega)$), en posant $p_z = 0$ lorsque $z \in X(\Omega) \times Y(\Omega) \setminus Z(\Omega)$.

Remarque
Par sommabilité de la famille positive $(p_{x,y})_{(x,y) \in X(\Omega) \times Y(\Omega)}$ et le théorème de Fubini, on a donc

$$1 = \sum_{(x,y) \in Z(\Omega)} p_{x,y} = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} p_{x,y} = \sum_{x \in X(\Omega)} \left(\sum_{y \in Y(\Omega)} p_{x,y} \right) = \sum_{y \in Y(\Omega)} \left(\sum_{x \in X(\Omega)} p_{x,y} \right).$$

Ainsi, le fait de prolonger à $X(\Omega) \times Y(\Omega)$ plutôt que de se limiter à $Z(\Omega)$ permet de sommer "en ligne" ou "en colonne". Et de toute façon, l'image $Z(\Omega)$ peut être délicate à décrire exactement.

2) Lois marginales

Définition 23 (Lois marginales d'un couple de VAD)
Etant donné un couple de variables aléatoires discrètes $Z = (X, Y) : \Omega \to E \times F$, on appelle **lois marginales** *de Z les lois de probabilité de X et Y.*

Vocabulaire
On parle de **première loi marginale** pour la loi de X, et de **deuxième loi marginale** pour la loi de Y.

Méthode (Déterminer les lois marginales à partir de la loi conjointe)
Si on connaît la loi du couple (X, Y) (la "loi conjointe"), alors on peut retrouver celles de X et de Y (les "lois marginales").
Par exemple, pour déterminer la loi de X : on a, pour tout $x \in X(\Omega)$,

$$\mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{y \in Y(\Omega)} ((X = x) \cap (Y = y)) \right) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y),$$

puisque $(Y = y)_{y \in Y(\Omega)}$ est un système complet d'événements.

Remarque
Lorsque X et Y à image finie, on peut présenter cette idée sous forme d'un tableau.
En notant $X(\Omega) = \{x_1, \cdots, x_m\}$ et $Y(\Omega) = \{y_1, \cdots, y_n\}$, on forme un tableau à m lignes et n colonnes et dont les cases sont les probabilités $p_{i,j} = \mathbb{P}(X = x_i, Y = y_j)$.
En sommant tous les éléments d'une même ligne, on obtient les probabilités $\mathbb{P}(X = x_i)$ et, en sommant tous les éléments d'une même colonne, on obtient les $\mathbb{P}(Y = y_j)$.

ATTENTION !
Les lois marginales ne suffisent pas en général à retrouver la loi conjointe !
En effet si X peut prendre m valeurs différentes et Y peut en prendre n, alors les

$$p_{i,j} = \mathbb{P}(X = x_i, Y = y_j)$$

sont au nombre de $m \times n$.
Or la connaissance des lois marginales, donc de $\mathbb{P}(X = x_i)$ et $\mathbb{P}(Y = y_j)$ pour $1 \leq i \leq m$ et $1 \leq j \leq n$ ne donnent que $m + n$ équations du type :

$$\sum_{i=1}^{m} p_{ij} = \mathbb{P}(Y = y_j), \qquad \sum_{j=1}^{n} p_{ij} = \mathbb{P}(X = x_i).$$

On a donc seulement un système linéaire de $m + n$ équations à $m \times n$ inconnues, il ne possède donc pas une unique solution (puisque $m \times n > m + n$ en général), ce qui empêche la détermination des $(p_{i,j})$, et donc de la loi conjointe.

Exemple (Deux lancers d'un dé)
On lance deux fois un dé équilibré. On a alors $\Omega = \{1, \ldots, 6\}^2$.
Soit X la VAD qui donne le résultat du premier lancer et Y la VAD donnant le résultat du second.
La loi du couple (X, Y) est alors donnée par :

$$\forall (i,j) \in \{1, \cdots, 6\}^2, \qquad \mathbb{P}(X = i, Y = j) = \frac{1}{36}.$$

La loi conjointe de X et Y est donc ici la loi uniforme sur l'ensemble $\{1, \cdots, 6\}^2$:

$$(X, Y) \sim \mathcal{U}(\{1, \cdots, 6\}^2).$$

Quant aux lois marginales, on a évidemment :

$$X \sim Y \sim \mathcal{U}(\{1, \cdots, 6\}).$$

Exemple (Urne)
Une urne contient 2 boules blanches, 1 rouge et 1 noire, toutes indiscernables.
On tire simultanément deux boules de cette urne et on nomme X la variable aléatoire donnant le nombre de boules blanches tirées et Y celle donnant le nombre boules rouges tirées.
Déterminer la loi conjointe et les lois marginales de X et Y.

Réponse : En notant B_1, B_2, R, N les 4 boules de l'urnes, on définit Ω comme l'ensemble des parties à 2 éléments de l'ensemble $\{B_1, B_2, R, N\}$ (on ne tient pas compte de l'ordre car les 2 boules sont tirées simultanément). C'est univers fini et $Card(\Omega) = \binom{4}{2} = 6$. On le munit de la tribu $\mathcal{P}(\Omega)$ et de l'équiprobabilité $\mathbb{P}$.
On a alors $X(\Omega) = \{0, 1, 2\}$ et $Y(\Omega) = \{0, 1\}$, donc $(X, Y)(\Omega) \subset \{0, 1, 2\} \times \{0, 1\}$.
Les événements $(X, Y) = (0, 0)$ et $(X, Y) = (2, 1)$ sont clairement impossibles, donc

$$\mathbb{P}(X = 0, Y = 0) = \mathbb{P}(X = 2, Y = 1) = 0.$$

Ensuite, on a

$$\mathbb{P}(X = 0, Y = 1) = \mathbb{P}(\{R, N\}) = \frac{1}{6}, \qquad \mathbb{P}(X = 1, Y = 0) = \mathbb{P}(\{\{B_1, N\}, \{B_2, N\}\}) = \frac{2}{6},$$

$$\mathbb{P}(X = 1, Y = 1) = \mathbb{P}(\{\{B_1, R\}, \{B_2, R\}\}) = \frac{2}{6}, \qquad \mathbb{P}(X = 2, Y = 0) = \mathbb{P}(\{B_1, B_2\}) = \frac{1}{6}.$$

On en déduit facilement le tableau donnant la loi conjointe et les lois marginales de (X, Y) (la somme des 6 probabilités vaut bien 1) :

$X \backslash Y$	0	1	Loi de X
0	0	$\frac{1}{6}$	$\mathbb{P}(X = 0) = \frac{1}{6}$
1	$\frac{2}{6}$	$\frac{2}{6}$	$\mathbb{P}(X = 1) = \frac{4}{6}$
2	$\frac{1}{6}$	0	$\mathbb{P}(X = 2) = \frac{1}{6}$
Loi de Y	$\mathbb{P}(Y = 0) = \frac{1}{2}$	$\mathbb{P}(Y = 1) = \frac{1}{2}$	

3) Lois conditionnelles

Définition 24 (Loi conditionnelle de X sachant A)
Soit $X : \Omega \to E$ une VAD définie sur $(\Omega, \mathcal{A}, \mathbb{P})$ et soit un événement $A \in \mathcal{A}$ tel que $\mathbb{P}(A) \neq 0$.
On appelle **loi conditionnelle de X sachant A** *la probabilité :*

$$\mathbb{P}_{X|A} : \begin{cases} \mathcal{P}(X(\Omega)) & \longrightarrow & \mathbb{R}^+ \\ B & \longmapsto & \mathbb{P}_{X|A}(B) = \mathbb{P}_A(X \in B) = \dfrac{\mathbb{P}((X \in B) \cap A)}{\mathbb{P}(A)} \end{cases}.$$

Remarque

- *On vérifie aisément que $\mathbb{P}_{X|A}$ est une probabilité sur $(X(\Omega), \mathcal{P}(X(\Omega)))$.*

- *Souvent, l'événement A par lequel on conditionne la variable X est de la forme $(Y = y)$, où $Y : \Omega \to F$ est une autre VAD, et y un élément de $Y(\Omega)$ tel que $\mathbb{P}(Y = y) \neq 0$.*
  La loi conditionnelle de X sachant $(Y = y)$ est alors l'application

$$\mathbb{P}_{X|(Y=y)} : \begin{cases} \mathcal{P}(X(\Omega) & \longrightarrow & \mathbb{R}^+ \\ B & \longmapsto & \mathbb{P}_{(Y=y)}(X \in B) = \dfrac{\mathbb{P}((X \in B) \cap (Y = y))}{\mathbb{P}(Y = y)} \end{cases}.$$

Exemple (Lancer de deux dés)
On lance deux dés équilibrés, X est la VAD donnant le résultat du premier dé et Y celle du second dé. On note $S = X + Y$. On a, par exemple,

$$\mathbb{P}_{S|(X=1)}(\{4\}) = \mathbb{P}_{(X=1)}(S = 4) = \frac{\mathbb{P}((X = 1) \cap (S = 4))}{\mathbb{P}(X = 1)} = \frac{\mathbb{P}((X, Y) = (1, 3))}{\mathbb{P}(X = 1)} = \frac{1/36}{1/6} = \frac{1}{6};$$
$$\mathbb{P}_{X|(S=4)}(\{1\}) = \mathbb{P}_{(S=4)}(X = 1) = \frac{\mathbb{P}((X = 1) \cap (S = 4))}{\mathbb{P}(S = 4)} = \frac{\mathbb{P}((X, Y) = (1, 3))}{\mathbb{P}(S = 4)} = \frac{1/36}{3/36} = \frac{1}{3}.$$

Méthode (Calcul d'une loi à l'aide de lois conditionnelles)
Grâce aux lois conditionnelles de X sachant $(Y = y)$ (lorsque y décrit $Y(\Omega)$), on peut calculer la loi de X à l'aide de la formule des probabilités totales :

$$\forall x \in X(\Omega), \quad \mathbb{P}(X = x) = \sum_{y \in Y(\Omega),\ \mathbb{P}(Y=y)\neq 0} \mathbb{P}_{(Y=y)}(X = x) \times \mathbb{P}(Y = y),$$

puisque les événements $(Y = y)_{y\in Y(\Omega)}$ tels que $\mathbb{P}(Y = y) \neq 0$ forment un SQCE (système "quasi-complet" d'événements) de Ω.

4) Vecteurs aléatoires

$E_1, \cdots, E_n$ désignent des ensembles quelconques.

Définition 25 (Vecteur aléatoire)
Soient $X_1 : \Omega \to E_1, \cdots, X_n : \Omega \to E_n$ des VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
On appelle **vecteur aléatoire discret** *de composantes $X_1, \cdots, X_n$ la variable aléatoire discrète $Z : \Omega \to E_1 \times \cdots \times E_n$ définie par : $\forall \omega \in \Omega$, $Z(\omega) = \big(X_1(\omega), \cdots, X_n(\omega)\big)$.*
La loi de Z est appelée **loi conjointe** *des variables $X_1, \cdots, X_n$ tandis que les lois de $X_1, \ldots, X_n$ sont les* **lois marginales** *de Z.*

Propriété 26 (Un vecteur aléatoire est une VAD)
Avec les notations précédentes, $(X_1, \cdots, X_n) : \Omega \to E_1 \times \cdots \times E_n$ est bien une VAD.

Preuve
Comme dans le cas $n = 2$.

Remarque
- *Bien sûr, comme dans le cas $n = 2$, la loi conjointe permet de déterminer les lois marginales (en sommant), mais les lois marginales ne déterminent pas en général la loi conjointe.*
- *La fonction composée $Y = f(X_1, \cdots, X_n)$ est donc une variable aléatoire discrète dès qu'elle est bien définie (d'après la prop. 17 appliquée à la VAD $X = (X_1, \cdots, X_n)$).*
- *En particulier, si $\forall i \in [\![1, n]\!]$, $X_i : \Omega \to \mathbb{R}$ est une VAD, alors la somme $S = X_1 + \cdots + X_n$ et le produit $P = X_1X_2 \cdots X_n$ sont des VAD réelles.*

III Indépendance de variables aléatoires discrètes

1) Couples de VAD indépendantes

Définition 27 (Indépendance de deux VAD)
Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
On dit que X et Y sont **indépendantes** *lorsque :*

$$\forall A \in \mathcal{P}(X(\Omega)),\ \forall B \in \mathcal{P}(Y(\Omega)), \qquad \mathbb{P}((X \in A) \cap (Y \in B)) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Notation
On notera $X \perp\!\!\!\perp Y$ pour indiquer que les VAD X et Y sont indépendantes.

Remarque
- *L'indépendance de X et Y signifie que tout événement du type $(X \in A)$ est indépendant de tout événement du type $(Y \in B)$.*
- *Si X et Y sont indépendantes, alors pour tout $A \in \mathcal{P}(X(\Omega))$, la loi conditionnelle de Y sachant $(X \in A)$ est la loi de Y.*

Exemple
On lance deux dés, on note X le résultat du lancer du premier dé, et Y le résultat du lancer du second dé. Alors, on modélise X et Y par deux variables aléatoires indépendantes (et de même loi uniforme : $X \sim Y \sim \mathcal{U}(\{1, \cdots, 6\})$.

Théorème 28 (Caractérisation de l'indépendance de deux VAD)
Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
Alors on a équivalence entre :

(i) *X et Y sont indépendantes ;*

(ii) *pour tout $(x, y) \in X(\Omega) \times Y(\Omega)$, $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$.*

Remarque
Ainsi, deux VAD sont indépendantes si et seulement si la distribution de probabilités de (X, Y) est égale au produit des distributions de probabilités de X et Y, c'est-à-dire :

$$X \perp\!\!\!\perp Y \iff \forall (x, y) \in X(\Omega) \times Y(\Omega),\ p_{x,y} = p_x p_y,$$

avec les notations habituelles.

Preuve
$\boxed{(i) \implies (ii)}$ Evident d'après la définition de l'indépendance de X et Y utilisée avec $A = \{x\}$ et $B = \{y\}$.

$\boxed{(ii) \implies (i)}$ Supposons que :

$$\forall (x, y) \in X(\Omega) \times Y(\Omega),\ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Etant donnés $A \in \mathcal{P}(X(\Omega))$ et $B \in \mathcal{P}(Y(\Omega))$, on a :

$$\mathbb{P}((X \in A) \cap (Y \in B)) = \mathbb{P}((X, Y) \in A \times B) = \sum_{(x,y) \in A \times B} \mathbb{P}((X, Y) = (x, y))$$

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

donc par hypothèse et sommabilité de la famille $(p_{x,y}) = (\mathbb{P}(X = x, Y = y))_{(x,y) \in A \times B}$:

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

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

ce qui montre que X et Y sont indépendantes.

Exemple (Somme de deux VAD indépendantes)
Pour X, Y deux VAD indépendantes, à valeurs dans $\mathbb{N}$. Alors, on a, pour tout $k \in \mathbb{N}$:

$$\mathbb{P}(X+Y=k) = \sum_{(i,j)\in\mathbb{N}^2,\ i+j=k}\mathbb{P}(X=i, Y=j) = \sum_{(i,j)\in\mathbb{N}^2,\ i+j=k}\mathbb{P}(X=i)\mathbb{P}(Y=j),$$

ce qui se réécrit

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

Exemple
Si $X \sim Y \sim \mathcal{B}(p)$ et si $X \perp\!\!\!\perp Y$, déterminer la loi de $X+Y$.

Propriété 29 (Compositions de VAD indépendantes)
Soit $X : \Omega \to E$ et $Y : \Omega \to F$ deux VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $f : X(\Omega) \to E'$ et $g : Y(\Omega) \to F'$ deux applications quelconques. Si $X \perp\!\!\!\perp Y$, alors $f(X) \perp\!\!\!\perp g(Y)$.

Preuve
Pour $(A, B) \in \mathcal{P}(f(X(\Omega))) \times \mathcal{P}(g(Y(\Omega)))$, on a

$$\mathbb{P}((f(X)\in A)\cap(g(Y)\in B)) = \mathbb{P}((X\in f^{-1}(A))\cap(Y\in g^{-1}(B))),$$

donc par indépendance de X et Y :

$$\mathbb{P}((f(X)\in A)\cap(g(Y)\in B)) = \mathbb{P}(X\in f^{-1}(A))\mathbb{P}(Y\in g^{-1}(B))) = \mathbb{P}(f(X)\in A)\mathbb{P}(g(Y)\in B)),$$

ce qui montre que $f(X)$ et $g(Y)$ sont indépendantes.

2) Famille finie de VAD indépendantes

Définition 30 (Indépendance d'une famille finie de VAD)
Soient $X_1 : \Omega \to E_1, \cdots, X_n : \Omega \to E_n$ des VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
On dit qu'elles sont **indépendantes** *lorsque pour toutes parties $A_1 \in \mathcal{P}(X_1(\Omega)), \cdots,$ $A_n \in \mathcal{P}(X_n(\Omega))$, les événements $(X_i \in A_i)_{1 \leq i \leq n}$ sont indépendants*
(i.e. pour toute partie $J \subset \{1, \cdots, n\}$, $\mathbb{P}\left(\bigcap_{i \in J}(X_i \in A_i)\right) = \prod_{i \in J} \mathbb{P}(X_i \in A_i)$.

Vocabulaire
On dit aussi que $X_1, \cdots, X_n$ sont **mutuellement indépendantes**.

ATTENTION !
L'indépendance de variables aléatoires est une propriété bien plus forte que l'indépendance deux à deux (tout comme pour les événements).

Théorème 31 (Caractérisation de l'indépendance d'une famille finie de VAD)
Soient $X_1 : \Omega \to E_1, \cdots, X_n : \Omega \to E_n$ des VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
Alors on a équivalence entre :

- *(i) $X_1, \cdots, X_n$ sont indépendantes ;*
- *(ii) pour tout $(x_1, \cdots, x_n) \in X_1(\Omega) \times \cdots \times X_n(\Omega)$, $\mathbb{P}\left(\bigcap_{i=1}^{n}(X_i = x_i)\right) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$.*

Preuve
$\boxed{(i) \implies (ii)}$ Evident d'après la définition de l'indépendance de $X_1, \cdots, X_n$ utilisée avec $J = \{1, \cdots, n\}$, $A_1 = \{x_1\}, \cdots, A_n = \{x_n\}$.
$\boxed{(ii) \implies (i)}$ Supposons que :

$$\forall (x_1, \cdots, x_n) \in X_1(\Omega) \times X_n(\Omega),\ \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n).$$

Etant donnés $(A_1, \cdots, A_n) \in \mathcal{P}(X_1(\Omega)) \times \cdots \times \mathcal{P}(X_n(\Omega))$ et $J \subset \{1, \cdots, n\}$, on a

$$\mathbb{P}\left(\bigcap_{i \in J}(X_i \in A_i)\right) = \mathbb{P}\left(\bigcap_{i=1}^{n}(X_i \in B_i)\right) = \mathbb{P}((X_1, \cdots, X_n) \in B_1 \times \cdots \times B_n),$$

où $B_i = A_i$ si $i \in J$ et $B_i = X_i(\Omega)$ si $i \notin J$.
En procédant comme pour deux variables, on obtient :

$$\mathbb{P}\left(\bigcap_{i \in J}(X_i \in A_i)\right) = \sum_{(x_1, \cdots, x_n) \in B_1 \times \cdots \times B_n} \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n)$$

$$= \sum_{(x_1, \cdots, x_n) \in B_1 \times \cdots \times B_n} \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n) = \prod_{i=1}^{n}\left(\sum_{x_i \in B_i} \mathbb{P}(X_i = x_i)\right)$$

$$= \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) = \prod_{i \in J} \mathbb{P}(X_i \in A_i) \prod_{i \notin J} \underbrace{\mathbb{P}(X_i \in X_i(\Omega))}_{=1} = \prod_{i \in J} \mathbb{P}(X_i \in A_i).$$

Théorème 32 (Lemme des coalitions)
Soient $X_1 : \Omega \to E_1, \cdots, X_n : \Omega \to E_n$ des VAD indépendantes définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, et soit $m \in \{1, \cdots, n-1\}$.
Pour toutes applications $f : X_1(\Omega) \times \cdots \times X_m(\Omega) \to E'$ et $g : X_{m+1}(\Omega) \times \cdots \times X_n(\Omega) \to F'$,
les variables $f(X_1, \cdots, X_m)$ et $g(X_{m+1}, \cdots, X_n)$ sont indépendantes.

Preuve
D'après la prop. 29, il suffit de montrer que les variables vectorielles $X = (X_1, \cdots, X_m)$ et $Y = (X_{m+1}, \cdots, X_n)$ sont indépendantes, et c'est facile car pour tout $x = (x_1, \cdots, x_m) \in X(\Omega)$ et $y = (x_{m+1}, \cdots, x_n) \in Y(\Omega)$:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X_1 = x_1, \cdots, X_n = x_n),$$

donc par indépendance de $X_1, \cdots, X_n$:

$$\mathbb{P}(X = x, Y = y) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = \prod_{i=1}^{m} \mathbb{P}(X_i = x_i) \times \prod_{i=m+1}^{n} \mathbb{P}(X_i = x_i) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Exemple
Si X_1, X_2, X_3 sont trois VAD réelles indépendantes, alors
$X = X_1 + X_2$ et $Y = X_3^2$ sont indépendantes.

3) Famille infinie de VAD indépendantes

Définition 33 (Indépendance d'une famille infinie de VAD)
Soit $(X_i)_{i\in I}$ une famille infinie quelconque de VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.
On dit que les $(X_i)_{i\in I}$ sont **indépendantes** *lorsque toutes les sous-familles finies $(X_i)_{i\in J}$ (avec J partie finie de I) sont indépendantes.*

Exemple (Somme de VAD de Bernoulli indépendantes et de même paramètre)
On lance indéfiniment une pièce truquée, qui donne *Pile* avec probabilité $p \in [0, 1]$, et pour tout $n \in \mathbb{N}^*$ on note X_n le résultat du n^e lancer : 1 si on obtient *Pile* et 0 si on obtient *Face*.
Alors les $(X_n)_{n\geq 1}$ forment une suite de variables aléatoires indépendantes (toutes de même loi de Bernoulli $\mathcal{B}(p)$).
Déterminer la loi de la somme $S_n = X_1 + \cdots + X_n$ pour tout $n \in \mathbb{N}^*$.

Réponse : S_n suit la loi binomiale $\mathcal{B}(n, p)$ car $S_n(\Omega) = [\![0, n]\!]$ (puisque chaque X_i vaut 0 ou 1), et pour tout $k \in [\![0, n]\!]$, on a l'égalité d'événements :

$$(S_n = k) = (Card\{i \in [\![1, n]\!],\ X_i = 1\} = k)\,,$$

donc la variable aléatoire S_n compte le nombre de *Pile* obtenus au cours des n tirages. Vu les hypothèses (indépendance des lancers, et probabilité de succès égale à p à chaque lancer), le théorème 11 s'applique.

Remarque (IMPORTANT)
La somme de n variables indépendantes de loi $\mathcal{B}(p)$ suit donc la loi $\mathcal{B}(n, p)$.

Théorème 34 (Théorème d'extension de Kolmogorov)
Soit $(\mathcal{L}_n)_{n\in\mathbb{N}}$ une suite de lois de probabilités discrètes sur des ensembles E_n.
Alors, il existe un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et une suite $(X_n)_{n\in\mathbb{N}}$ de VAD toutes définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes, et telles que $\forall n \in \mathbb{N},\ X_n \sim \mathcal{L}_n$.

Preuve
Admis, dépasse largement le cadre du programme.

Remarque
C'est ce dernier théorème qui permet notamment de donner un sens mathématique à des expériences aléatoires du type de la précédente (on lance indéfiniment une pièce), en assurant l'existence d'une suite de VAD (X_n) indépendantes, et suivant toutes une même loi donnée (par exemple une loi de Bernoulli).

IV Espérance d'une VAD réelle ou complexe

Dans cette section, les variables aléatoires discrètes sont définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

1) Définition

Définition 35 (Espérance)
On dit que la VAD $X : \Omega \to \mathbb{K}$ **admet une espérance** *(ou* **est d'espérance finie***) lorsque la famille $(x\mathbb{P}(X = x))_{x \in X(\Omega)}$ est sommable. On appelle alors* **espérance de X** *la somme de cette famille :*

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

Remarque

- *L'espérance est la moyenne des valeurs de X, pondérée par les probabilités d'apparition de chaque valeur.*
- *Si elle existe $E(X)$ ne dépend que de la loi de X, pas de X elle-même ($X \sim Y \implies E(X) = E(Y)$).*
- *Si X est d'image finie, notée $X(\Omega) = \{x_1, \cdots, x_n\}$, alors X admet une espérance (car toute famille finie est sommable), et*

$$E(X) = \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k)$$

.

- *Si X a une image $X(\Omega)$ dénombrable, alors en considérant une énumération $(x_n)_{n \in \mathbb{N}}$ de $X(\Omega)$, la variable X admet une espérance si et seulement si la série $\sum_{n \geq 0} x_n \mathbb{P}(X = x_n)$ converge* **absolument** *(pour assurer la sommabilité), et dans ce cas*

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

Remarquons que cette somme ne dépend pas de l'énumération de $X(\Omega)$ choisie (propriété qui découle de la sommabilité, et qui ne marche pas toujours pour les séries semi-convergentes...)

- *Si $X(\Omega)$ est dénombrable et si $X \geq 0$ (i.e. $\forall \omega \in \Omega,\ X(\omega) \geq 0$), alors X admet une espérance ssi $\displaystyle\sum_{x \in X(\Omega)} x\mathbb{P}(X = x) < +\infty$.*
On peut écrire $E(X) = +\infty$ si ce n'est pas le cas (famille non sommable de réels positifs).
Si $X \geq 0$, on peut donc toujours définir $E(X) = \displaystyle\sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$ dans $\mathbb{R}^+ \cup \{+\infty\}$.
- *Si $X(\Omega)$ est dénombrable et si X n'est pas à valeurs positives, alors X est d'espérance finie ssi $\displaystyle\sum_{x \in X(\Omega)} |x|\mathbb{P}(X = x) < +\infty$ (puisque $|x\mathbb{P}(X = x)| = |x|\mathbb{P}(X = x)$ étant donné que $\mathbb{P}(X = x) \geq 0$).*

ATTENTION !
Bien noter qu'il faut de la **sommabilité** (donc de la convergence **absolue** si on somme sur $\mathbb{N}$) pour assurer l'existence de l'espérance. Cette condition est requise pour que la somme $\sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$ ne dépende pas de l'ordre de sommation des termes.
On ne peut donc pas se contenter d'une série semi-convergente !

Exemple
Soit $X : \Omega \to \mathbb{N}^*$ une variable aléatoire telle que

$$\forall n \in \mathbb{N}^*, \qquad \mathbb{P}(X = n) = \frac{1}{n} - \frac{1}{n+1}$$

(cela est possible car $\frac{1}{n} - \frac{1}{n+1} \geq 0$ pour tout $n \geq 1$, et par téléscopage : $\sum_{n=1}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right) = \lim_{N \to +\infty} \left(1 - \frac{1}{N+1} \right) = 1$.)
Alors, cette variable aléatoire ne possède pas d'espérance, puisque

$$\sum_{x \in X(\Omega)} |x \mathbb{P}(X = x)| = \sum_{n \geq 1} n \mathbb{P}(X = n) = \sum_{n \geq 1} \frac{1}{n+1}$$

diverge. Vu que $X \geq 0$, on peut écrire $E(X) = +\infty$.

Exemple
Si X est une variable aléatoire constante (i.e. $\exists c \in \mathbb{R},\ \forall \omega \in \Omega,\ X(\omega) = c$), alors X admet une espérance et $E(X) = c$.
Ca marche aussi si X est constante presque sûrement (i.e. $\exists c \in \mathbb{R},\ \mathbb{P}(X = c) = 1$).

Définition 36 (Variable aléatoire centrée)
On dit que X est **centrée** *lorsqu'elle admet une espérance nulle.*

Propriété 37 (Cas des VAD à valeurs dans $\mathbb{N}$)
Si X est une VAD à valeurs dans $\mathbb{N}$, alors on a dans $\mathbb{R}^+ \cup \{+\infty\}$:

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

Preuve
Avec le formalisme des familles de réels positifs, on a dans $\mathbb{R}^+ \cup \{+\infty\}$:

$$\sum_{n=0}^{+\infty} \mathbb{P}(X > n) = \sum_{n=0}^{+\infty} \left(\sum_{k=n+1}^{+\infty} \mathbb{P}(X = k) \right) = \sum_{(n,k) \in \mathbb{N}^2,\ k > n} \mathbb{P}(X = k).$$

Par le théorème de Fubini (cas positif), on peut inverser l'ordre de sommation, donc

$$\sum_{n=0}^{+\infty} \mathbb{P}(X > n) = \sum_{k=1}^{+\infty} \left(\sum_{n=0}^{k-1} \mathbb{P}(X = k) \right) = \sum_{k=1}^{+\infty} k \mathbb{P}(X = k) = E(X)$$

(puisque le terme d'indice $k = 0$ est nul).
La deuxième égalité résulte du fait que $P(X > n) = \mathbb{P}(X \geq n+1)$ (puisque X est à valeurs entières), et d'un changement d'indice.

2) Propriétés

Théorème 38 (Formule de transfert)
Soit $X : \Omega \to E$ une VAD (où E est un ensemble quelconque) et $f : X(\Omega) \to \mathbb{K}$ une fonction. On a équivalence entre :

(i) la variable $f(X)$ est d'espérance finie ;

(ii) la famille $(f(x)\mathbb{P}(X = x))_{x \in X(\Omega)}$ est sommable.

Dans ce cas, on a

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

Preuve
Par définition, la VAD composée $Y = f(X)$ est d'espérance finie ssi $\sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) < +\infty$.

Etudions cette condition. Dans $\mathbb{R}^+ \cup \{+\infty\}$, on a :

$$\sum_{y\in Y(\Omega)} |y|\mathbb{P}(Y=y) = \sum_{y\in f(X(\Omega))} |y|\mathbb{P}(f(X)=y) = \sum_{y\in f(X(\Omega))} |y|\mathbb{P}(X\in B_y),$$

en notant $B_y = \{x\in X(\Omega),\ f(x)=y\} = f^{-1}(\{y\})$. Par σ-additivité, on a donc :

$$\sum_{y\in Y(\Omega)} |y|\mathbb{P}(Y=y) = \sum_{y\in f(X(\Omega))} |y| \left(\sum_{x\in B_y} \mathbb{P}(X=x)\right).$$

En factorisant :

$$\sum_{y\in Y(\Omega)} |y|\mathbb{P}(Y=y) = \sum_{y\in f(X(\Omega))} \left(\sum_{x\in B_y} |y|\mathbb{P}(X=x)\right) = \sum_{y\in f(X(\Omega))} \left(\sum_{x\in B_y} |f(x)|\mathbb{P}(X=x)\right).$$

On peut alors appliquer le théorème de sommation par paquets pour les familles de réels positifs : puisque $X(\Omega) = \bigcup_{y\in f(X(\Omega))} B_y$ (la réunion est disjointe), on obtient

$$\sum_{y\in Y(\Omega)} |y|\mathbb{P}(Y=y) = \sum_{x\in X(\Omega)} |f(x)|\mathbb{P}(X=x),$$

ce qui montre l'équivalence voulue. Dans le cas où $\sum_{x\in X(\Omega)} |f(x)|\mathbb{P}(X=x) < +\infty$, on peut refaire le même calcul sans les valeurs absolues (d'après le théorème de sommation par paquets pour les familles sommables de nombres complexes), et on obtient :

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

Remarque
La preuve du théorème montre en particulier que si f est positive, alors on a dans $\mathbb{R}^+ \cup \{+\infty\}$:

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

Par exemple, avec la fonction $f : x\mapsto |x|$, on obtient que pour toute VAD $X : \Omega \to \mathbb{K}$, on a dans $\mathbb{R}^+ \cup \{+\infty\}$:

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

donc X est d'espérance finie ssi $E(|X|) < +\infty$.

Exemple
Sous réserve de sommabilité, on a donc pour $X : \Omega \to \mathbb{K}$:

$$E(X^3) = \sum_{x\in X(\Omega)} x^3\mathbb{P}(X=x), \qquad E(e^X) = \sum_{x\in X(\Omega)} e^x\mathbb{P}(X=x), \qquad \cdots$$

Mais X peut être à valeurs vectorielles également, par exemple $X = (X_1, X_2)$, où X_1 et X_2 sont scalaires. Dans ce cas, on a donc (sous réserve de sommabilité) :

$$E(f(X_1,X_2)) = \sum_{(x_1,x_2)\in (X_1,X_2)(\Omega)} f(x_1,x_2)\mathbb{P}(X_1=x_1, X_2=x_2),$$

mais aussi

$$E(f(X_1,X_2)) = \sum_{(x_1,x_2)\in X_1(\Omega)\times X_2(\Omega)} f(x_1,x_2)\mathbb{P}(X_1=x_1, X_2=x_2),$$

(en rajoutant les couples de probabilité nulle).
Par exemple :

$$E(X_1X_2) = \sum_{(x_1,x_2)\in X_1(\Omega)\times X_2(\Omega)} x_1x_2\mathbb{P}(X_1=x_1, X_2=x_2),$$

Théorème 39 (Linéarité de l'espérance)
Soit $X : \Omega \to \mathbb{K}$ et $Y : \Omega \to \mathbb{K}$ deux VAD qui admettent une espérance finie.
Alors pour tout $\lambda \in \mathbb{K}$, $\lambda X + Y$ admet une espérance finie et $E(\lambda X + Y) = \lambda E(X) + E(Y)$.

Preuve
Posons $Z = \lambda X + Y$. D'après la formule de transfert appliquée au couple (X, Y) et à la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par $f(x, y) = |\lambda x + y|$, on a dans $\mathbb{R}^+ \cup \{+\infty\}$:

$$E(|Z|) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |\lambda x + y|\mathbb{P}(X = x, Y = y),$$

donc par inégalité triangulaire :

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

Puisque $E(|X|) < +\infty$ et $E(|Y|) < +\infty$ (par hypothèse), on en déduit $E(|Z|) < +\infty$, donc Z est bien d'espérance finie.
Enfin, toujours d'après le théorème de transfert :

$$E(Z) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} (\lambda x + y)\mathbb{P}(X = x, Y = y) = \cdots = \lambda E(X) + E(Y)$$

(en reprenant le calcul précédent sans les valeurs absolues).

Corollaire 40 (Structure d'espace vectoriel des VAD admettant une espérance)
L'ensemble des variables aléatoires discrètes $X : \Omega \to \mathbb{K}$ d'espérance finie forme un $\mathbb{K}$-espace vectoriel, que l'on notera L^1. De plus, l'espérance est une forme linéaire $E : L^1 \to \mathbb{K}$.

Preuve
C'est une reformulation du résultat précédent.

Exemple
Si $X \in L^1$, alors

- pour toutes constantes a et b dans $\mathbb{K}$, on a $E(aX + b) = aE(X) + b$.
- la variable $X - E(X)$ est centrée : $E(X - E(X)) = 0$.

On a également $X \in L^1 \iff |X| \in L^1 \iff E(|X|) < +\infty$.

Propriété 41 (Positivité de l'espérance)
Si $X \in L^1$ et si $X \geq 0$, alors $E(X) \geq 0$.
Si de plus $E(X) = 0$, alors $X = 0$ presque sûrement.

Remarque
Ca marche aussi si on suppose seulement que $X \geq 0$ presque sûrement (i.e. $\mathbb{P}(X \geq 0) = 1$).

Preuve
Immédiat car si $X \in L^1$, alors $E(X) = \sum_{x\in X(\Omega)} x\mathbb{P}(X = x)$ est un réel positif (somme d'une famille

sommable de réels positifs puisque $X(\Omega) \subset \mathbb{R}^+$). De plus, si $E(X) = 0$, alors $x\mathbb{P}(X = x) = 0$ pour tout $x \in X(\Omega)$ (somme nulle de réels positifs), donc $\mathbb{P}(X = x) = 0$ pour tout $x \neq 0$, ce qui mène à $\mathbb{P}(X = 0) = 1$.

Corollaire 42 (Croissance de l'espérance)
Si X et Y sont réelles et dans L^1 et si $X \leq Y$, alors $E(X) \leq E(Y)$.

Preuve
Vu que L^1 est un $\mathbb{K}$-ev, on a $Y - X \in L^1$. Comme $Y - X \geq 0$, la propriété précédente donne $E(Y - X) \geq 0$, donc $E(Y) - E(X) \geq 0$ par linéarité de l'espérance.

Propriété 43 (Inégalité triangulaire)
Si $X \in L^1$, alors $|E(X)| \leq E(|X|)$.

Preuve
Il s'agit de l'inégalité triangulaire pour les familles sommables de nombres complexes.

Propriété 44 (Critère de domination)
Si $|X| \leq Y$ et si $Y \in L^1$, alors $X \in L^1$.

Preuve
Dans $\mathbb{R}^+ \cup \{+\infty\}$, on a

$$E(|X|) = \sum_{x \in X(\Omega)} |x|\mathbb{P}(X = x) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |x|\mathbb{P}(X = x, Y = y).$$

Or, l'hypothèse $|X| \leq Y$ signifie que $\forall \omega \in \Omega, |X(\omega)| \leq Y(\omega)$, donc pour tout couple $(x, y) \in X(\Omega) \times Y(\Omega)$, on a $(|x| > y \implies \mathbb{P}(X = x, Y = y) = 0)$. Ceci amène :

$$E(|X|) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega),\ |x| \leq y} |x|\mathbb{P}(X = x, Y = y) \leq \sum_{(x,y) \in X(\Omega) \times Y(\Omega),\ |x| \leq y} y\mathbb{P}(X = x, Y = y)$$

$$= \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} y\mathbb{P}(X = x, Y = y) = \sum_{y \in Y(\Omega)} y\mathbb{P}(Y = y) = E(Y) < +\infty,$$

ce qui montre que $X \in L^1$.

Remarque
Ca marche aussi si on a seulement $|X| \leq Y$ presque sûrement.

Exemple
Toute variable aléatoire discrète bornée (i.e. $\exists c \geq 0,\ \forall \omega \in \Omega,\ |X(\omega)| \leq c$) admet donc une espérance.

Remarque
Les propriétés de l'espérance sont donc les mêmes que celle de l'intégrale sur un intervalle réel.
Et les VAD admettant une espérance sont les analogues des fonctions intégrables (d'où la notation L^1).
En résumé : pour toute variable aléatoire discrète $X : \Omega \to \mathbb{K}$:

$$X \in L^1 \iff E(|X|) < +\infty,$$

et pour toute fonction $g : I \to \mathbb{K}$ (où I est un intervalle réel) :

$$g \in L^1(I, \mathbb{K}) \iff \int_I |g| < +\infty.$$

3) Espérance des lois usuelles

Théorème 45 (Espérance des lois usuelles)

(i) *Si* $X \sim \mathcal{U}(\{x_1, \ldots, x_n\})$*, alors* $E(X) = \frac{1}{n} \sum_{k=1}^{n} x_k$.

(ii) *Si* $X \sim \mathcal{B}(p)$ *(avec* $p \in [0,1]$*), alors* $E(X) = p$.

(iii) *Si* $X \sim \mathcal{B}(n,p)$ *(avec* $n \in \mathbb{N}$ *et* $p \in [0,1]$*), alors* $E(X) = np$.

(iv) *Si* $X \sim \mathcal{G}(p)$ *(avec* $p \in]0,1[$*), alors* X *possède une espérance et* $E(X) = \frac{1}{p}$.

(v) *Si* $X \sim \mathcal{P}(\lambda)$ *(avec* $\lambda > 0$*), alors* X *possède une espérance et* $E(X) = \lambda$.

Preuve

(i) $X(\Omega) = \{x_1, \cdots, x_n\}$ et $\mathbb{P}(X = x_k) = \frac{1}{n}$ pour tout $k \in \{1, \cdots, n\}$, donc

$$E(X) = \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k) = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

(ii) $X(\Omega) = \{0; 1\}$, donc $E(X) = 0 * \mathbb{P}(X = 0) + 1 * \mathbb{P}(X = 1) = 0 * (1-p) + 1 * p = p$.

(iii) $X(\Omega) = \{0, \cdots, n\}$ et $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pour tout $k \in \{0, \cdots, n\}$ (ici $x_k = k$), donc

$$E(X) = \sum_{k=0}^{n} k \mathbb{P}(X = k) = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

Pour calculer cette somme, on introduit la fonction $f : \mathbb{R} \to \mathbb{R}$ définie par

$$f(x) = \sum_{k=0}^{n} \binom{n}{k} x^k = (1+x)^n.$$

En dérivant, on obtient :

$$\forall x \in \mathbb{R}, \quad f'(x) = \sum_{k=1}^{n} k \binom{n}{k} x^{k-1} = n(1+x)^{n-1},$$

donc en multipliant par x :

$$\sum_{k=1}^{n} k \binom{n}{k} x^k = nx(1+x)^{n-1}.$$

On en déduit en évaluant en $x = \frac{p}{1-p}$ que pour $p \neq 1$,

$$E(X) = (1-p)^n \sum_{k=0}^{n} k \binom{n}{k} \left(\frac{p}{1-p}\right)^k = (1-p)^n n \left(\frac{p}{1-p}\right) \left(1 + \frac{p}{1-p}\right)^{n-1} = np.$$

Enfin, si $p = 1$, la formule reste vraie puisque $E(X) = \sum_{k=0}^{n} k \binom{n}{k} 1^k 0^{n-k} = n = np$.

(iv) Sous réserve de convergence absolue, l'espérance de X est $E(X) = \sum_{k=1}^{+\infty} k(1-p)^{k-1} p$.

On utilise la série géométrique (dont le rayon de convergence est $R = 1$) :

$$\forall x \in]-1; 1[, \qquad \sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}.$$

En dérivant terme à terme ce développement en série entière, on obtient :

$$\forall x \in]-1; 1[, \qquad \sum_{k=1}^{+\infty} k x^{k-1} = \frac{1}{(1-x)^2}$$

(avec convergence absolue). En évaluant ceci en $x = 1 - p \in]0;1[$, on obtient :

$$\forall p \in]0;1[, \qquad \sum_{k=1}^{+\infty} k(1-p)^{k-1} = \frac{1}{p^2}.$$

Ceci montre que $X \in L^1$, et $E(X) = p \times \frac{1}{p^2} = \frac{1}{p}$.

(v) Sous réserve de convergence absolue, l'espérance de X est

$$E(X) = \sum_{k=0}^{+\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^{k+1}}{k!} = \lambda e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}.$$

On reconnaît alors le DSE de e^{λ}, qui converge absolument pour tout $\lambda \in \mathbb{R}$.
Ceci montre que $X \in L^1$, et que $E(X) = \lambda e^{-\lambda} e^{\lambda} = \lambda$.

4) Espérance et indépendance

Théorème 46 (Espérance d'un produit de VAD indépendantes)

(i) *Si X et Y sont dans L^1 et indépendantes, alors $XY \in L^1$ et*

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

(ii) *Si $X_1, \cdots, X_n$ sont dans L^1 et indépendantes, alors $X_1 \cdots X_n \in L^1$ et*

$$E(X_1 \cdots X_n) = E(X_1) \cdots E(X_n).$$

Preuve

(i) Encore la formule de transfert : dans $\mathbb{R}^+ \cup \{+\infty\}$, on a

$$E(|XY|) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega)} |xy| \mathbb{P}(X = x, Y = y).$$

Par indépendance de X et Y, on a $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$, donc par le théorème de Fubini (cas positif) :

$$E(|XY|) = \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) \sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) = E(|X|)E(|Y|) < +\infty,$$

ce qui montre que $XY \in L^1$. En reprenant le calcul sans les valeurs absolues (ce qui est possible d'après le théorème de Fubini sur les familles sommables de nombres complexes), on obtient $E(XY) = E(X)E(Y)$.

(ii) Récurrence facile, qui fonctionne car pour tout $n \geq 2$, la variable $P_{n-1} = X_1 \cdots X_{n-1}$ est indépendante de la variable X_n par le lemme des coalitions.

Remarque

Cette propriété un peu surprenante dénote cette fois avec les propriétés de l'intégrale !

ATTENTION !

La réciproque est fausse. On peut avoir $E(XY) = E(X)E(Y)$ sans que X et Y soient indépendantes.

Corollaire 47 (Espérance d'un produit de VAD composées)

Si X et Y sont indépendantes et si $f(X)$ et $g(Y)$ sont dans L^1, alors $f(X)g(Y) \in L^1$ et

$$E(f(X)g(Y)) = E(f(X))E(g(Y)).$$

Preuve

Direct d'après la propriété précédente car l'indépendance de X et Y entraîne celle de $f(X)$ et $g(Y)$ (prop. 29).

V Variance d'une VAD réelle

Dans cette section, les variables aléatoires discrètes sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{R}$.

1) Moments d'une VAD réelle, espace L^2

Définition 48 (Moment d'ordre p d'une VAD)
Soit $X : \Omega \to \mathbb{R}$ une VAD et soit $p \in \mathbb{N}^*$. On appelle **moment d'ordre p** *de X la quantité $E(X^p)$, lorsqu'elle existe et est finie (c'est-à-dire lorsque $X^p \in L^1$).*

Remarque

- *Le moment d'ordre 1 est l'espérance.*
- *Par la formule de transfert, on a*

$$\forall p \in \mathbb{N}^*, \qquad E(X^p) = \sum_{x \in X(\Omega)} x^p \mathbb{P}(X = x),$$

(sous réserve de sommabilité, c'est-à-dire sous réserve que $E(|X|^p) < +\infty$).
- *Vu que $X^2 \geq 0$, X admet un moment d'ordre 2 ssi $\sum_{x \in X(\Omega)} x^2 \mathbb{P}(X = x) < +\infty$.*

Dans la suite, on va s'intéresser au moment d'ordre 2.

Propriété 49 (Lien entre moment d'ordre 2 et espérance)
Si X admet un moment d'ordre 2, alors X admet une espérance finie.

Preuve
Résulte de l'inégalité $|X| \leq \frac{1}{2}(1 + X^2)$ (vraie car $1 + X^2 - 2|X| = (1 - |X|)^2 \geq 0$).
Puisque $E(1 + X^2) = 1 + E(X^2) < +\infty$, on en déduit par le critère de domination que $E(|X|) < +\infty$, c'est-à-dire $X \in L^1$.

Notation
On notera L^2 l'ensemble des VAD $X : \Omega \to \mathbb{R}$ qui admettent un moment d'ordre 2 :

$$X \in L^2 \iff X^2 \in L^1 \iff E(X^2) < +\infty.$$

On a $L^2 \subset L^1$ d'après la prop. précédente, et l'inclusion est stricte, car certaines VAD vérifient $E(|X|) < +\infty$ mais $E(X^2) = +\infty$.

Remarque
On peut généraliser : si X admet un moment d'ordre $p \in \mathbb{N}^*$, alors X admet des moments d'ordre $k \in [\![1, p]\!]$. En notant L^p l'ensemble des VAD réelles admettant un moment d'ordre p, on a donc

$$k \leq p \implies L^p \subset L^k.$$

Ceci résulte de l'inégalité :

$$k \leq p \implies |X|^k \leq 1 + |X|^p$$

(vu que $(|X| \leq 1 \implies |X|^k \leq 1)$ et $(|X| > 1 \implies |X|^k \leq |X|^p)$).

Propriété 50 (Produit de deux VAD ayant un moment d'ordre 2)
Si X et Y sont dans L^2, alors $XY \in L^1$.

Preuve
Résulte de l'inégalité $|XY| \leq \frac{1}{2}(X^2 + Y^2)$ et du critère de domination.

Théorème 51 (Structure algébrique de L^2)
L^2 est un sous-espace vectoriel de L^1, donc un $\mathbb{R}$-espace vectoriel.

Preuve
On a déjà montré que $L^2 \subset L^1$, et L^2 est non vide (il contient les VAD constantes).
Soit X, Y dans L^2 et $\lambda \in \mathbb{R}$. On a

$$(\lambda X + Y)^2 = \lambda^2 X^2 + 2\lambda XY + Y^2.$$

On a $XY \in L^1$ (d'après la prop. précédente), ainsi que $X^2 \in L^1$, $Y^2 \in L^1$, donc par combinaison linéaire, $(\lambda X + Y)^2 \in L^1$, c'est-à-dire $\lambda X + Y \in L^2$.

Théorème 52 (Inégalité de Cauchy-Schwarz)
L'application $\left\{ \begin{array}{ccc} L^2 \times L^2 & \longrightarrow & \mathbb{R} \\ (X,Y) & \longmapsto & E(XY) \end{array} \right.$ *est une forme bilinéaire symétrique positive (mais pas définie positive), donc on a l'inégalité de Cauchy-Schwarz :*

$$\forall (X,Y) \in L^2 \times L^2, \qquad |E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

Et on a égalité si et seulement si X et Y sont presque sûrement proportionnelles.

ATTENTION !
Cette application n'est pas un produit scalaire, car si $E(X^2) = 0$, alors par positivité de X^2, on en déduit que $X^2 = 0$ presque sûrement, donc $X = 0$ presque sûrement (i.e. $\mathbb{P}(X = 0) = 1$), mais cela n'implique pas que X est nulle.

Preuve

- Si X et Y sont dans L^2, on a $XY \in L^1$ donc $E(XY)$ existe.
  La symétrie et la bilinéarité de $\varphi : (X,Y) \mapsto E(XY)$ sont évidentes par linéarité de l'espérance.
  De plus, $\varphi(X,X) = E(X^2) \geq 0$ par positivité de l'espérance.
- Pour l'inégalité de Cauchy-Schwarz, il n'y a pas besoin d'avoir la définie-positivité, la positivité de la forme bilinéaire symétrique suffit. Rappelons la preuve : on considère pour tout réel t

  $$P(t) = \varphi(X + tY, X + tY) = E((X + tY)^2).$$

  Cette fonction est à valeurs positives, et vérifie :

  $$P(t) = E(X^2) + 2tE(XY) + t^2 E(Y^2).$$

  Si $E(Y^2) \neq 0$, alors P est un polynôme du second degré de signe constant, donc son discriminant est négatif ou nul, ce qui amène :

  $$4E(XY)^2 - 4E(Y^2)E(X^2) \leq 0,$$

  ou encore

  $$|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

  Si $E(Y^2) = 0$, alors par positivité de Y^2, on a $Y^2 = 0$ presque sûrement, donc $Y = 0$ presque sûrement, et l'inégalité de Cauchy-Schwarz reste vraie (puisque $|E(XY)| = \sqrt{E(X^2)}\sqrt{E(Y^2)}0$).
- Pour le cas d'égalité : Si $E(Y^2) \neq 0$, alors $|E(XY)| = \sqrt{E(X^2)}\sqrt{E(Y^2)}$ si et seulement si le discriminant de $t \mapsto P(t)$ est nul, c'est-à-dire qu'il existe $t_0 \in \mathbb{R}$ tel que $P(t_0) = 0$. Par positivité de φ, ceci équivaut à $X + t_0 Y = 0$ presque sûrement, i.e. X est proportionnelle à Y presque sûrement.
  Si $E(Y^2) = 0$, alors $Y = 0$ presque sûrement, donc d'une part $|E(XY)| = \sqrt{E(X^2)}\sqrt{E(Y^2)} = 0$, et d'autre part, Y est proportionnelle à X presque sûrement (mais pas l'inverse !). L'équivalence voulue est donc vraie.

2) Variance et écart-type

Définition 53 (Variance)
Si $X \in L^2$, on appelle **variance** *de X le réel positif*

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

et **écart-type** *de X le réel positif*

$$\sigma(X) = \sqrt{V(X)}.$$

Remarque

- *Dans ces conditions, la variance est bien définie car X et la constante $E(X)$ admettent des moments d'ordre 2, donc $X - E(X)$ aussi (par la structure d'espace vectoriel de L^2).*
- *En particulier, si $X(\Omega)$ est fini, alors $V(X)$ existe.*
- *La variance permet de mesurer la dispersion de la variable X autour de sa moyenne (l'espérance). Si la variable X s'exprime avec une unité physique (des mètres, des années, etc.), alors l'espérance et l'écart-type sont homogènes à X (ils s'expriment avec la même unité), alors que la variance est homogène à X^2.*
- *On a $V(X) \geq 0$ et ($V(X) = 0 \iff X$ est presque sûrement constante (égale à $E(X)$)).*

Théorème 54 (Formule de Huygens)
Si $X \in L^2$, alors $V(X) = E(X^2) - E(X)^2$.

Preuve
Vu que $X \in L^2$, on a $X \in L^1$. Notons $m = E(X) \in \mathbb{R}$. Par définition de la variance et linéarité de l'espérance, on obtient :

$$V(X) = E((X-m)^2) = E(X^2) - 2mE(X) + E(m^2) = E(X^2) - 2m^2 + m^2 = E(X^2) - m^2.$$

Remarque
Par la formule de transfert, on a donc, pour tout $X \in L^2$:

$$V(X) = \sum_{x \in X(\Omega)} x^2 \mathbb{P}(X = x) - \left(\sum_{x \in X(\Omega)} x \mathbb{P}(X = x) \right)^2.$$

Propriété 55 (Variance d'une composée affine)
Si $X \in L^2$ et $(a, b) \in \mathbb{R}^2$, alors $aX + b \in L^2$ et $V(aX + b) = a^2 V(X)$.

Preuve
Les VAD X et 1 sont dans L^2, donc leur combinaison linéaire $aX + b$ aussi (vu la structure d'espace vectoriel).
Vu que $E(aX + b) = aE(X) + b$, on a par la formule de Huygens :

$$V(aX + b) = E((aX + b)^2) - (aE(X) + b)^2).$$

Par linéarité de l'espérance :

$$V(aX+b) = a^2E(X^2) + 2abE(X) + b^2 - a^2E(X^2) - 2abE(X) - b^2 = a^2(E(X^2) - E(X)^2) = a^2V(X).$$

Remarque
Une translation de X ne modifie donc pas sa variance. C'est logique, car la translation ne modifie pas la dispersion des valeurs de X.

Définition 56 (Variable aléatoire réduite)
Une VAD $X \in L^2$ est dite **réduite** *lorsque $V(X) = 1$.*

Remarque
Si $X \in L^2$ *et si* $V(X) \neq 0$*, alors la variable aléatoire* $\frac{X-E(X)}{\sigma(X)}$ *est centrée et réduite.*

3) Variance des lois usuelles

Théorème 57 (Variances des lois usuelles)

(i) *Si* $X \sim \mathcal{U}([\![1,n]\!])$*, alors* $V(X) = \frac{n^2-1}{12}$.

(ii) *Si* $X \sim \mathcal{B}(p)$ *avec* $p \in [0,1]$*, alors* $V(X) = p(1-p)$.

(iii) *Si* $X \sim \mathcal{B}(n,p)$ *avec* $n \in \mathbb{N}$ *et* $p \in [0,1]$*, alors* $V(X) = np(1-p)$.

(iv) *Si* $X \sim \mathcal{G}(p)$ *avec* $p \in]0,1[$*, alors* $X \in L^2$ *et* $V(X) = \frac{1-p}{p^2}$.

(v) *Si* $X \sim \mathcal{P}(\lambda)$ *avec* $\lambda > 0$*, alors* $X \in L^2$ *et* $V(X) = \lambda$.

Preuve
Exercice calculatoire !

4) Covariance de deux VAD

Définition 58 (Covariance de deux VAD dans L^2)
Si X *et* Y *sont dans* L^2*, alors on appelle* **covariance** *de* X *et* Y *le réel*

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

Remarque
Dans ces conditions, $Cov(X,Y)$ *est bien définie, car* $X-E(X)$ *et* $Y-E(Y)$ *sont dans* L^2*, donc leur produit* $(X-E(X))(Y-E(Y))$ *est dans* L^1.

Vocabulaire
On dit que X et Y sont **décorrélées** lorsque $Cov(X,Y) = 0$.

Théorème 59 (Formule de Huygens)
Si X *et* Y *sont dans* L^2*, alors*

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

En particulier, $Cov(X,X) = V(X)$.

Preuve
Développer et utiliser la linéarité de l'espérance.

Remarque
D'après la formule du transfert, on a donc

$$Cov(X,Y) = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} xy\mathbb{P}(X=x, Y=x) - \left(\sum_{x\in X(\Omega)} x\mathbb{P}(X=x)\right)\left(\sum_{y\in Y(\Omega)} y\mathbb{P}(Y=y)\right).$$

Corollaire 60 (Décorrélation et indépendance)
Si X *et* Y *sont dans* L^2 *et indépendantes, alors elles sont décorrélées.*

Preuve
L'indépendance donne $E(XY) = E(X)E(Y)$, donc $Cov(X,Y) = 0$.

ATTENTION !

La réciproque est fausse (voir les exercices).

Théorème 61 (Inégalité de Cauchy-Schwarz 2)
L'application $Cov : \begin{cases} L^2 \times L^2 & \longrightarrow & \mathbb{R} \\ (X,Y) & \longmapsto & Cov(X,Y) \end{cases}$ *est une forme bilinéaire symétrique positive (mais pas définie positive). On a donc l'inégalité de Cauchy-Schwarz :*

$$\forall (X,Y) \in L^2, \qquad |Cov(X,Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}.$$

Et on a égalité si et seulement si une des deux VAD est fonction affine de l'autre presque sûrement.

Preuve
Il est facile d'établir à partir de la formule $Cov(X,Y) = E(XY) - E(X)E(Y)$ que Cov est une forme bilinéaire symétrique.
La positivité résulte du fait que $Cov(X,X) = V(X) = E((X - E(X))^2) \geq 0$.
On a donc directement l'inégalité de Cauchy-Schwarz (valable pour toute forme bilinéaire symétrique positive) :

$$|Cov(X,Y)| \leq \sqrt{V(X)}\sqrt{V(Y)}.$$

Pour le cas d'égalité : notons $\varphi : (X,Y) \mapsto E(XY)$. On a alors $Cov(X,Y) = \varphi(X - m_X, Y - m_Y)$, en notant $m_X = E(X)$ et $m_Y = E(Y)$.
D'après la proposition 52 :

$$|Cov(X,Y)| = \sqrt{V(X)}\sqrt{V(Y)} \iff |\varphi(X - m_X, Y - m_Y)| = \sqrt{(E(X - m_X)^2)}\sqrt{E((Y - m_Y)^2)}$$

$$\iff X - m_X \text{ et } Y - m_Y \text{ sont presque sûrement proportionnelles}$$

$$\iff \exists (a,b) \neq (0,0), \qquad aX + bY \text{ est constante presque sûrement.}$$

Définition 62 (Coefficient de corrélation linéaire (HP))
Si X, Y sont dans L^2 avec $V(X) \neq 0$ et $V(Y) \neq 0$, alors on appelle **coefficient de corrélation linéaire entre X et Y** *le réel :*

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)} = \frac{Cov(X,Y)}{\sqrt{V(X))}\sqrt{V(Y)}}.$$

Remarque

- *On a $\rho(X,Y) \in [-1,1]$ d'après l'inégalité de Cauchy-Schwarz, cette quantité mesure la corrélation de X et Y au sens suivant : plus $|\rho|$ est proche de 1 et plus les variables sont corrélées, plus ρ est proche de 0 et moins elles le sont.*
- *Les cas extrêmes :*
    * $\rho(X,Y) \in \{-1,1\}$ *ssi il existe $(a,b) \in \mathbb{R}^2$ tel que $Y = aX + b$ presque sûrement.*
    * $\rho(X,Y) = 0$ *ssi X et Y sont décorrélées.*
- *La covariance est une sorte de produit scalaire et le coefficient de corrélation linéaire une sorte de "cosinus". Ainsi, deux VAD décorrélées sont analogues à deux vecteurs orthogonaux.*

5) Variance d'une somme

Théorème 63 (Variance d'une somme)

(*i*) *Si* X, Y *sont dans* L^2*, alors*

$$V(X+Y) = V(X) + 2Cov(X,Y) + V(Y).$$

(*ii*) *Si* $X_1, \cdots, X_n$ *sont dans* L^2*, alors*

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

(*iii*) *En particulier, si les variables* $X_1, \cdots, X_n$ *sont deux à deux décorrélées, alors*

$$V(X_1 + \cdots + X_n) = \sum_{i=1}^{n} V(X_i).$$

(*iv*) *En particulier, si* $X_1, \cdots, X_n$ *sont indépendantes, alors*

$$V(X_1 + \cdots + X_n) = \sum_{i=1}^{n} V(X_i).$$

Preuve
On a :

$$V(X_1 + \cdots + X_n) = Cov(X_1 + \cdots + X_n, X_1 + \cdots + X_n).$$

donc il suffit de développer en utilisant la bilinéarité et la symétrie de la covariance.

Remarque
On retrouve ainsi la variance d'une loi binomiale :
soit $X_1, \cdots, X_n$ *des VAD indépendantes et de même loi de Bernoulli* $\mathcal{B}(p)$ *avec* $p \in [0,1]$ *(elles existent d'après le théorème d'extension de Kolmogorov).*
Alors, on sait que $S = X_1 + \cdots + X_n$ *suit la loi* $\mathcal{B}(n,p)$*, donc la variance d'une loi binomiale est* $V(S)$*.*
Par indépendance des X_i *:*

$$V(S) = V(X_1) + \cdots + V(X_n) = nV(X_1).$$

puisqu'elles suivent toutes la même loi.
On montre facilement que $V(X_1) = p(1-p)$*, donc* $V(S) = np(1-p)$*.*

VI Inégalités de concentration

On a toujours en contexte un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

1) Indicatrice d'un événement

Propriété 64 (VAD indicatrice d'un événement)
Etant donné un événement $A \in \mathcal{A}$, la fonction $\mathbb{1}_A : \begin{cases} \Omega & \longrightarrow & \mathbb{R} \\ \omega & \longmapsto & \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} \end{cases}$ *est une VAD, appelée* **indicatrice de** *A. De plus, cette VAD est dans L^1, et $E(\mathbb{1}_A) = \mathbb{P}(A)$.*

Preuve
L'image $\mathbb{1}_A(\Omega) = \{0, 1\}$ est finie, donc au plus dénombrable. De plus pour tout réel x, l'ensemble $\mathbb{1}_A^{-1}(\{x\}) = \{w \in \Omega,\ \mathbb{1}_A(\omega) = x\}$ vaut A si $x = 1$, $\overline{A}$ si $x = 0$, et $\emptyset$ sinon, donc $\forall x \in \mathbb{R},\ \mathbb{1}_A^{-1}(\{x\}) \in \mathcal{A}$. La fonction 1_A est donc bien une variable aléatoire discrète. Son image étant finie, cette variable est dans L^1 et par définition :

$$E(\mathbb{1}_A) = \sum_{x \in \{0,1\}} x\mathbb{P}(\mathbb{1}_A = x) = \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A).$$

Remarque
Toute probabilité peut donc s'écrire comme l'espérance d'une VAD.

Exemple
Pour toute VAD $X : \Omega \to \mathbb{R}$ et pour tout réel a, on a $\mathbb{P}(X \geq a) = E(\mathbb{1}_{[X \geq a]})$.

2) Inégalités de Markov et Bienaymé-Tchebychev

Théorème 65 (Inégalité de Markov)
Soit $X : \Omega \to \mathbb{R}$ une VAD. On suppose que $X \geq 0$ et $X \in L^1$. Alors :

$$\forall a > 0, \qquad \mathbb{P}(X \geq a) \leq \frac{E(X)}{a}.$$

Preuve
Fixons $a \in \mathbb{R}$. On peut décomposer la VAD X :

$$X = X\mathbb{1}_{[X \geq a]} + X\mathbb{1}_{[X < a]}$$

(car pour tout $\omega \in \Omega$, on a soit $X(\omega) \geq a$, soit $X(\omega) < a$), et les deux termes sont dans L^1 car positifs et dominés par X. Par linéarité et croissance de l'espérance :

$$E(X) = E(\underbrace{X\mathbb{1}_{[X \geq a]}}_{\geq a\mathbf{1}_{[X \geq a]}}) + E(\underbrace{X\mathbb{1}_{[X < a]}}_{\geq 0}) \geq E(a\mathbb{1}_{[X \geq a]}) + E(0) = a\mathbb{P}(X \geq a).$$

Lorsque $a > 0$, on obtient le résultat en divisant par a.

Remarque

- *On peut réécrire cette preuve de manière plus élémentaire en écrivant $E(X)$ comme une somme que l'on décompose en deux sommes :*

$$E(X) = \sum_{x \in X(\Omega),\ x \geq a} x\mathbb{P}(X = x) + \sum_{x \in X(\Omega),\ x < a} x\mathbb{P}(X = x),$$

et on minore chaque terme de la même façon que précédemment.

- *L'inégalité de Markov peut se réécrire ainsi : pour toute variable aléatoire discrète $X : \Omega \to \mathbb{R}$ dans L^1 (pas nécessairement positive) :*

$$\forall a > 0, \qquad \mathbb{P}(|X| \geq a) \leq \frac{E(|X|)}{a}.$$

- *Si $X \geq 0$ et $X \in L^p$ avec $p \in \mathbb{N}^*$ (i.e. $E(|X|^p) < +\infty$), alors $X^p \in L^1$, donc :*

$$\forall a > 0, \qquad \mathbb{P}(X \geq a) = \mathbb{P}(X^p \geq a^p) \leq \frac{E(X^p)}{a^p},$$

et cette dernière majoration est en général meilleure que $\frac{E(X)}{a}$ lorsque a est grand.

- *Retenir de cette preuve la façon de décomposer une VAD selon des cas disjoints :*

$$X = X\mathbb{1}_A + X\mathbb{1}_{\overline{A}},$$

(où A est un événement de la tribu $\mathcal{A}$, pas nécessairement relié à X d'ailleurs). Cela permet souvent de simplifier les calculs, en évitant de réécrire des sommes.

Théorème 66 (Inégalité de Bienaymé-Tchebychev)
Soit $X : \Omega \to \mathbb{R}$ une VAD dans L^2. Alors :

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

Preuve
Il suffit d'appliquer l'inégalité de Markov à la VAD $Y = (X - E(X))^2$, qui est bien positive et dans L^1, puisque $E(Y) = V(X) < +\infty$: pour tout réel $\varepsilon > 0$,

$$\mathbb{P}(|X - E(X)| \geq \varepsilon) = \mathbb{P}(Y \geq \varepsilon^2) \leq \frac{E(Y)}{\varepsilon^2} \leq \frac{V(X)}{\varepsilon^2}.$$

Remarque
Cette inégalité importante illustre le fait qu'il est peu probable que les valeurs de $X \in L^2$ s'écartent de sa moyenne (son espérance).

Exemple
On lance 1000 fois un dé équilibré.
Montrer que la probabilité d'avoir une moyenne supérieure ou égale à 4 est inférieure à 2%.

Solution : On note X_i la VAD donnant le résultat du dé au i^{e} lancer, et on pose $S = X_1 + \cdots + X_{1000}$. Chaque X_i suit la loi uniforme sur $[\![1, 6]\!]$ et les X_i sont indépendantes, donc

$$E(S) = 1000E(X_1) = 3500, \qquad V(S) = 1000V(X_1) = 1000\left(\frac{6^2-1}{12}\right) = \frac{35 \times 1000}{12} < 3000.$$

On cherche simplement à majorer la probabilité de l'événement $A = [\frac{S}{1000} \geq 4] = [S \geq 4000]$, donc pas besoin de la calculer exactement (la loi de S est compliquée à obtenir).
D'après l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}(S \geq 4000) = \mathbb{P}(S - E(S) \geq 500) \leq \mathbb{P}(|S - E(S)| \geq 500) \leq \frac{V(S)}{500^2} < \frac{3000}{500^2} = \frac{6}{500} < 0,02.$$

Ici, l'inégalité de Markov donne :

$$\mathbb{P}(S \geq 4000) \leq \frac{E(S)}{4000} = \frac{3,5}{4} = 0,875,$$

ce qui est beaucoup moins intéressant !

3) Loi faible des grands nombres

Théorème 67 (Loi faible des grands nombres)
Soit $(X_n)_{n\geq 1}$ une suite de VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, à valeurs dans $\mathbb{R}$.
Si les $(X_n)_{n\geq 1}$ sont dans L^2, indépendantes et de même loi, alors en posant $S_n = X_1 + \cdots + X_n$, on a

$$\mathbb{P}\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) \underset{n\to+\infty}{\longrightarrow} 0.$$

Preuve
C'est direct avec l'inégalité de Bienaymé-Tchebychev. Par linéarité de l'espérance et indépendance des X_i, on a

$$E(S_n) = nE(X_1), \qquad V(S_n) = nV(X_1),$$

donc pour tout réel $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - E(X_1)\right| \geq \varepsilon\right) = \mathbb{P}(|S_n - E(S_n)| \geq n\varepsilon) \leq \frac{V(S_n)}{n^2\varepsilon^2} = \frac{V(X_1)}{n\varepsilon^2} \underset{n\to+\infty}{\longrightarrow} 0.$$

Remarque
Ainsi, il est peu probable que la suite des "moyennes empiriques" $\left(\frac{S_n}{n}\right)_{n\geq 1}$ s'écarte trop de la "moyenne théorique" $(E(X_1) = E(S_n/n))$ lorsque n devient grand.
On dit que la suite $(S_n/n)_{n\geq 1}$ converge "en probabilité" vers $E(X_1)$.

4) Preuve probabiliste du théorème de Weierstrass

Rappelons l'énoncé du **théorème d'approximation de Weierstrass** :

si $f \in \mathcal{C}([a, b], \mathbb{K})$, il existe une suite de polynômes (P_n) qui converge uniformément vers f sur $[a, b]$.

Par changement de variable affine, il est facile de se ramener au cas où $[a, b] = [0, 1]$ (voir exercices de la feuille 10 sur les suites de fonctions).

Démonstration probabiliste sur le segment $[0, 1]$:

On introduit, pour $n \in \mathbb{N}^*$, les "polynômes de Bernstein" $(B_{n,k})_{0\leq k\leq n}$. Ces $B_{n,k}$ représentent la distribution de probabilités d'une loi binomiale, puisque

$$\forall k \in \{0, \cdots, n\},\ \forall x \in [0, 1],\ B_{n,k}(x) = \binom{n}{k} x^k (1 - x)^{n-k} = \mathbb{P}(X = k),$$

où $X \sim \mathcal{B}(n, x)$.
Etant donnée une fonction continue $f : [0, 1] \to \mathbb{K}$, on pose alors, pour tout entier $n \in \mathbb{N}^*$:

$$P_n : x \mapsto \sum_{k=0}^{n} B_{x,k}(x) f(k/n)$$

(c'est une moyenne pondérée des valeurs de f aux points de la subdivision régulière de $[0, 1]$ en $n + 1$ points).
Fixons un réel $\varepsilon > 0$. Par uniforme continuité de f sur le segment $[0, 1]$ (résultant du théorème de Heine), il existe un réel $\delta > 0$ tel que $\forall (x, y) \in [0, 1]^2,\ |y - x| \leq \delta \implies |f(y) - f(x)| \leq \varepsilon$. Pour tout réel $x \in [0, 1]$ et tout $n \in \mathbb{N}^*$, on a alors

$$|P_n(x) - f(x)| = \left|\sum_{k=0}^{n} \mathbb{P}(X = k) f(k/n) - \sum_{k=0}^{n} \mathbb{P}(X = k) f(x)\right|$$

(puisque $\sum\limits_{k=0}^{n} \mathbb{P}(X = k) = 1$). Par la formule de transfert, on a donc

$$|P_n(x) - f(x)| \leq \sum_{k=0}^{n} |f(k/n) - f(x)| \mathbb{P}(X = k) = E(|f(X/n) - f(x)|).$$

Ensuite, on décompose selon la condition $|\frac{X}{n} - x| \leq \delta$:

$$|f(X/n) - f(x)| = |f(X/n) - f(x)|\mathbb{1}_{|\frac{X}{n}-x|\leq\delta} + |f(X/n) - f(x)|\mathbb{1}_{|\frac{X}{n}-x|>\delta},$$

donc par linéarité de l'espérance :

$$\begin{aligned}|P_n(x) - f(x)| &\leq E\left(\underbrace{|f(X/n) - f(x)|}_{\leq\varepsilon}\mathbb{1}_{|\frac{X}{n}-x|\leq\delta}\right) + E\left(\underbrace{|f(X/n) - f(x)|}_{\leq 2\|f\|_\infty}\mathbb{1}_{|\frac{X}{n}-x|>\delta}\right)\\ &\leq \varepsilon\underbrace{\mathbb{P}\left(|\frac{X}{n} - x| \leq \delta\right)}_{\leq 1} + 2\|f\|_\infty\mathbb{P}\left(|\frac{X}{n} - x| > \delta\right) \leq \varepsilon + 2\|f\|_\infty\mathbb{P}(|X - nx| > n\delta).\end{aligned}$$

Par l'inégalité de Bienaymé-Tchebychev :

$$\mathbb{P}(|X - nx| > n\delta) = \mathbb{P}(|X - E(X)| > n\delta) \leq \mathbb{P}(|X - E(X)| \geq n\delta) \leq \frac{V(X)}{n^2\delta^2} = \frac{x(1-x)}{\delta^2 n}.$$

On en déduit la majoration uniforme :

$$\forall (n, x) \in \mathbb{N}^* \times [0, 1], \qquad |P_n(x) - f(x)| \leq \varepsilon + 2\|f\|_\infty\frac{1/4}{\delta^2 n} = \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n},$$

donc, puisque $\frac{\|f\|_\infty}{2\delta^2 n} \underset{n\to+\infty}{\longrightarrow} 0$, il existe $n_0 \in \mathbb{N}^*$ tel que

$$\forall n \geq n_0,\ \forall x \in [0, 1], \qquad |P_n(x) - f(x)| \leq 2\varepsilon,$$

ce qui montre bien que la suite de fonctions polynomiales (P_n) converge uniformément vers f sur $[0, 1]$.

VII Fonction génératrice d'une VAD à valeurs dans $\mathbb{N}$

Dans cette dernière partie, on considèrera des VAD définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{N}$.

Définition 68 (Fonction génératrice)
La **fonction génératrice** *d'une VAD $X : \Omega \to \mathbb{N}$ est la fonction réelle*

$$G_X : t \mapsto E(t^X) = \sum_{n=0}^{+\infty} \mathbb{P}(X = n)t^n.$$

C'est une série entière de la variable t.

Vocabulaire
On parle aussi de **série génératrice** de X.

Propriété 69 (Propriétés de la fonction génératrice)
Soit une VAD $X : \Omega \to \mathbb{N}$.

(i) Le rayon de convergence de la série entière définissant G_X est $R \geq 1$, et on a $G_X(1) = 1$.
(ii) La fonction G_X est continue sur $[-1, 1]$.
(iii) La fonction G_X est de classe $\mathcal{C}^\infty$ sur $]-1, 1[$ et peut se dériver terme à terme.

Preuve
La série entière G_X converge normalement sur $[-1, 1]$ puisque

$$\sum_{n=0}^{+\infty} \|t \mapsto \mathbb{P}(X = n)t^n\|_{\infty,[-1,1]} = \sum_{n=0}^{+\infty} \mathbb{P}(X = n) = G_X(1) = 1 < +\infty.$$

Ceci montre (i) et (ii). Le point (iii) résulte du théorème de dérivation terme à terme sur les séries entières.

Théorème 70 (La fonction génératrice caractérise la loi)
Soit deux VAD $X : \Omega \to \mathbb{N}$ et $Y : \Omega \to \mathbb{N}$. Alors

$$G_X = G_Y \text{ sur un voisinage de } 0 \iff X \sim Y.$$

Preuve
La coïncidence des séries entières G_X et G_Y sur un voisinage de 0 équivaut à l'égalité des coefficients ($\mathbb{P}(X = n) = \mathbb{P}(Y = n)$ pour tout $n \in \mathbb{N}$), donc à l'égalité des lois $\mathbb{P}_X$ et $\mathbb{P}_Y$.

Exemple (Fonctions génératrices des lois usuelles)
Il faut être capable de les retrouver rapidement.

Loi	$G_X(t)$	Rayon de cv	Domaine de cv
$\mathcal{U}(\llbracket 1, n \rrbracket)$	$\begin{cases} \frac{t}{n} \times \frac{1-t^n}{1-t} & \text{si } t \neq 1 \\ 1 & \text{si } t = 1 \end{cases}$	$+\infty$	$]-\infty, +\infty[$
$\mathcal{B}(p)$	$(1-p) + pt$	$+\infty$	$]-\infty, +\infty[$
$\mathcal{B}(n, p)$	$((1-p) + pt)^n$	$+\infty$	$]-\infty, +\infty[$
$\mathcal{G}(p)$	$\frac{pt}{1-(1-p)t}$	$\frac{1}{1-p}$	$]-\frac{1}{1-p}, \frac{1}{1-p}[$
$\mathcal{P}(\lambda)$	$e^{\lambda(t-1)}$	$+\infty$	$]-\infty, +\infty[$

Théorème 71 (Fonction génératrice d'une somme de VAD indépendantes)

(i) *Soit X et Y deux VAD à valeurs dans $\mathbb{N}$ et indépendantes. Notons R_X et R_Y les rayons de convergence respectifs des séries génératrices G_X et G_Y. Alors le rayon de convergence de G_{X+Y} est supérieur ou égal à $R = \min(R_X, R_Y)$ et on a*

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

(ii) *Généralisation : si $X_1, \cdots, X_n$ sont des VAD à valeurs dans $\mathbb{N}$ et indépendantes, alors le rayon de convergence de $G_{X_1+\cdots+X_n}$ est supérieur ou égal à $R = \min(R_{X_1}, \cdots, R_{X_n})$, et on a*

$$\forall t \in]-R, R[, \qquad G_{X_1+\cdots+X_n}(t) = G_{X_1}(t)\cdots G_{X_n}(t).$$

Preuve

(i) Notons $Z = X + Y$. On a $Z(\Omega) \subset \mathbb{N}$, et pour tout $n \in \mathbb{N}$,

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

(par indépendance de X et Y). On reconnaît alors les coefficients du produit de Cauchy des deux séries entières G_X et G_Y, qui convergent absolument (car normalement) sur $[-1, 1]$. Donc pour tout réel t tel que $|t| < \min(R_X, R_Y)$, on a

$$G_X(t)G_Y(t) = \sum_{n=0}^{+\infty} \left(\sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n-k) \right) t^n = \sum_{n=0}^{+\infty} P(Z = n)t^n = G_Z(t).$$

(ii) Récurrence immédiate, qui fonctionne car pour tout entier $n \geq 2$, la variable $X_1 + \cdots + X_{n-1}$ est indépendante de la variable X_n, par le lemme des coalitions.

Méthode

Ce théorème est très utile pour déterminer la loi d'une somme de VAD indépendantes.

Exemple

- Si $X_1, \cdots, X_n$ sont indépendantes, de même loi $\mathcal{B}(p)$, alors on retrouve facilement que $S_n = X_1 + \cdots + X_n$ suit la loi $\mathcal{B}(n, p)$.
  En effet, on a $G_{X_i}(t) = 1 - p + pt$ pour tout $t \in \mathbb{R}$ et $i \in [\![1, n]\!]$ donc par indépendance

  $$\forall t \in \mathbb{R}, \qquad G_{S_n}(t) = \prod_{i=1}^{n} G_{X_i}(t) = (G_{X_1}(t))^n = (1 - p + pt)^n,$$

  et on reconnaît la fonction génératrice de la loi $\mathcal{B}(n, p)$, ce qui montre que $S_n \sim \mathcal{B}(n, p)$.
- Si $X_1, \cdots, X_n$ sont indépendantes, $X_1 \sim \mathcal{P}(\lambda_1), \cdots, X_n \sim \mathcal{P}(\lambda_n)$, alors on retrouve facilement que $S_n = X_1 + \cdots + X_n$ suit la loi $\mathcal{P}(\lambda_1 + \cdots + \lambda_n)$.
  En effet, on a $G_{X_i}(t) = e^{\lambda_i(t-1)}$ pour tout $t \in \mathbb{R}$ et $i \in [\![1, n]\!]$ donc par indépendance

  $$\forall t \in \mathbb{R}, \qquad G_{S_n}(t) = \prod_{i=1}^{n} G_{X_i}(t) = \prod_{i=1}^{n} e^{\lambda_i(t-1)} = e^{(\sum\limits_{i=1}^{n} \lambda_i)(t-1)},$$

  et on reconnaît la fonction génératrice de la loi $\mathcal{P}(\lambda_1 + \cdots + \lambda_n)$, ce qui montre que $S_n \sim \mathcal{P}(\lambda_1 + \cdots + \lambda_n)$.

Enfin, citons un résultat qui permet de retrouver l'espérance et la variance à partir de la fonction génératrice.

Théorème 72 (Lien entre fonction génératrice et moments)
Soit une VAD $X : \Omega \to \mathbb{N}$.

(i) *$X \in L^1$ si et seulement si G_X est dérivable en 1.*
Dans ce cas, on a $G'_X(1) = E(X)$.

(ii) *$X \in L^2$ si et seulement si G_X est deux fois dérivable en 1.*
Dans ce cas, on a $G''_X(1) = E(X(X-1))$, et donc
$V(X) = E(X^2) - E(X)^2 = G''_X(1) + G'_X(1) - G'_X(1)^2$.

Preuve

(i) En tant que série entière, G_X est de classe $\mathcal{C}^\infty$ sur $]-1,1[$ avec

$$\forall t \in]-1,1[, \qquad G'_X(t) = \sum_{n=1}^{+\infty} n\mathbb{P}(X=n)t^{n-1}.$$

On remarque que la dérivée G'_X est croissante et positive sur $[0,1[$ (comme somme infinie de fonctions croissantes et positives), donc elle possède une limite $\ell \in \mathbb{R}^+ \cup \{+\infty\}$ en 1^- :

$$\ell = \lim_{t\to 1^-} G'_X(t) = \sup_{t\in[0,1[} G'_X(t).$$

En outre, par convergence normale sur $[-1,1]$, la fonction G_X est continue sur $[-1,1]$. On peut donc appliquer le théorème de prolongement de la dérivée : si $\ell \in \mathbb{R}$, alors G_X est dérivable en 1 avec $G'_X(1) = \ell$ et si $\ell = +\infty$, alors G_X ne l'est pas.
D'où l'équivalence :

$$G_X \text{ dérivable en } 1 \iff \ell < +\infty.$$

Montrons alors que ceci équivaut à $X \in L^1$.

- Si $X \in L^1$, alors tout comme G_X, la série G'_X converge normalement sur $[-1,1]$, puisque $\sum_{n=1}^{+\infty} \|t \mapsto n\mathbb{P}(X=n)t^{n-1}\|_{\infty,[-1,1]} = \sum_{n=1}^{+\infty} n\mathbb{P}(X=n) < +\infty$. Le théorème de la double limite s'applique alors, et on déduit que

$$G'_X(t) \underset{t\to 1^-}{\longrightarrow} \sum_{n=1}^{+\infty} n\mathbb{P}(X=n) = E(X) < +\infty.$$

Ainsi, $\ell = E(X) < +\infty$, ce qui montre que G_X est dérivable en 1 avec $G'_X(1) = \ell = E(X)$.

Remarque
On a en fait montré $G_X \in \mathcal{C}^1(]-1,1],\mathbb{R})$, puisque G'_X se prolonge continûment en 1.

- Si G_X est dérivable en 1, on a $\ell < +\infty$, ainsi que les inégalités :

$$\forall N \in \mathbb{N}^*,\ \forall t \in [0,1[, \qquad \sum_{n=1}^{N} n\mathbb{P}(X=n)t^{n-1} \leq G'_X(t)$$

(par positivité des termes). Par passage à la limite lorsque $t \to 1^-$, on en déduit

$$\forall N \in \mathbb{N}^*, \qquad \sum_{n=1}^{N} n\mathbb{P}(X=n) \leq \ell < +\infty,$$

donc $(n\mathbb{P}(X=n))_{n\geq 1}$ est sommable, ce qui implique $X \in L^1$.

(ii) Similaire au *(i)*, en raisonnant avec

$$G''_X : t \mapsto \sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(X=n)t^{n-2}$$

qui est croissante et positive sur $[0,1[$, donc possède une limite $L \in \mathbb{R}^+ \cup \{+\infty\}$ en 1^- :

$$L = \lim_{t\to 1^-} G''_X(t) = \sup_{t\in[0,1[} G''_X(t).$$

- Si $X \in L^2$, alors $X \in L^1$ et G'_X est de classe $\mathcal{C}^1$ sur $[0,1]$ avec $G'_X(1) = E(X)$ d'après (i). En outre, G''_X converge normalement sur $[-1,1]$ car

$$\sum_{n=2}^{+\infty} \|t \mapsto n(n-1)\mathbb{P}(X=n)t^{n-2}\|_{\infty,[-1,1]} = \sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(X=n) = E(X^2)-E(X) < +\infty,$$

donc par le théorème de la double limite, $L = \sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(X=n) < +\infty$, ce qui montre que G_X est deux fois dérivable en 1 avec $G''_X(1) = L = E(X(X-1))$.

- Si G_X est deux fois dérivable en 1, alors G_X est de classe $\mathcal{C}^1$ sur $[0,1]$, donc nécessairement $L < +\infty$ (si $L = +\infty$, G'_X ne serait pas dérivable en 1, par le théorème de prolongement de la dérivée). En outre,

$$\forall N \geq 2,\ \forall t \in [0,1[, \qquad \sum_{n=2}^{N} n(n-1)\mathbb{P}(X=n)t^{n-2} \leq G''_X(t),$$

donc en faisant tendre $t \to 1^-$,

$$\forall N \geq 2, \qquad \sum_{n=2}^{N} n(n-1)\mathbb{P}(X=n) \leq L < +\infty,$$

ce qui montre que $\sum_{n=2}^{+\infty} n(n-1)\mathbb{P}(X=n) < +\infty$, et par équivalence de terme général $E(X^2) = \sum_{n=0}^{+\infty} n^2\mathbb{P}(X=n) < +\infty$, donc $X \in L^2$.

Remarque
De la même manière, pour tout $p \in \mathbb{N}^*$, on peut montrer que $X \in L^p$ (i.e. $E(|X|^p) < +\infty$) si et seulement si la série entière G_X possède un prolongement de classe $\mathcal{C}^p$ en 1, et dans ce cas ;

$$G_X^{(p)}(1) = E(X(X-1)\cdots(X-p+1)).$$