CH18 : Variables aléatoires discrètes

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**.

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* $A \subset E$*,* $(X \in A) \in \mathcal{A}$*.*

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 $\mathcal{A}$.

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.

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 $(p_x)_{x \in X(\Omega)} = (\mathbb{P}(X = x))_{x \in X(\Omega)}$.

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 $x \in X(\Omega)$.*

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 $(p_x)_{x \in E}$.

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 $X \sim \mathcal{L}$.

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

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

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

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 $\mathcal{B}(n, p)$.*

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 $\boldsymbol{p}$** *lorsque*

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

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

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 $\mathcal{G}(p)$.*

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

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 $X \sim \mathcal{P}(\lambda)$*.*

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* $\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \underset{n\to+\infty}{\longrightarrow} e^{-\lambda} \times \dfrac{\lambda^k}{k!}$*.*

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), ...

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 :*

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

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.

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

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.

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 :*

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

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.

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.

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

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* $f(X) \perp\!\!\!\perp g(Y)$.

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**.

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

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.

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.*

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 $\forall n \in \mathbb{N},\ X_n \sim \mathcal{L}_n$.

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 :*

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

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 $\mathbb{R}^+ \cup \{+\infty\}$:

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

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

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

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 $E(\lambda X + Y) = \lambda E(X) + E(Y)$.

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 $E : L^1 \to \mathbb{K}$.

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.

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

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

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

3) Espérance des lois usuelles

Théorème 45 (Espérance des lois usuelles)

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

(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* $E(X) = \lambda$.

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*

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

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*

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

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 $X^p \in L^1$).*

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.

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 $E(X^2) = +\infty$.

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

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

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.

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*

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

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

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

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

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* $V(X) = \lambda$.

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*

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

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

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

Théorème 61 (Inégalité de Cauchy-Schwarz 2)
L'application $Cov : \left\{ \begin{array}{ccc} L^2 \times L^2 & \longrightarrow & \mathbb{R} \\ (X,Y) & \longmapsto & Cov(X,Y) \end{array} \right.$ *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.

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 :*

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

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*

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

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 $E(\mathbb{1}_A) = \mathbb{P}(A)$.*

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 :

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

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

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

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

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

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 $\sum_{k=0}^{n} \mathbb{P}(X=k) = 1$). 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 :

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

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

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.*

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

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

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*

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

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

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
  $V(X) = E(X^2) - E(X)^2 = G''_X(1) + G'_X(1) - G'_X(1)^2$.