CH17 : Espaces probabilisés

En MP2I, nous avions étudié les probabilités sur un univers fini. Nous allons ici généraliser ces notions aux univers quelconques.

I Espace probabilisable

1) Univers

Définition 1 (Univers, éventualités)
L'ensemble des résultats possibles décrivant une expérience aléatoire est appelé **univers**. *On le note généralement* Ω.
Les éléments $\omega \in \Omega$ *sont les issues observées de l'expérience aléatoire, on les appelle* **éventualités**.

2) Tribu

En MP2I, nous avons vu que dans le cas d'un univers fini, les événements se représentent par des parties de Ω. L'ensemble des événements peut donc être assimilé à $\mathcal{P}(\Omega)$.
Dans le cas général, il n'en est pas ainsi : toute partie de Ω ne représentera pas nécessairement un événement : on se limitera à certaines parties de Ω, donc à un sous-ensemble de $\mathcal{P}(\Omega)$.

Notation
Etant donné un ensemble Ω, le complémentaire d'une partie $A \subset \Omega$ sera noté :

$$\overline{A} = \{\omega \in \Omega,\ \omega \notin A\}$$

(on trouve aussi la notation A^c).
De même, étant données deux parties $A, B \subset \Omega$, on notera

$$A \setminus B = \{\omega \in \Omega,\ \omega \in A \text{ et } \omega \notin B\} = A \cap \overline{B}.$$

Attention, cela ne suppose pas nécessairement que $B \subset A$!

Définition 2 (Tribu)
On appelle **tribu** *sur un ensemble* Ω *toute partie* $\mathcal{A} \subset \mathcal{P}(\Omega)$ *telle que :*

(*i*) $\Omega \in \mathcal{A}$;
(*ii*) $\forall A \in \mathcal{A}$, $\overline{A} \in \mathcal{A}$ *(stabilité par passage au complémentaire).*
(*iii*) $\forall (A_n) \in \mathcal{A}^{\mathbb{N}}$, $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{A}$ *(stabilité par réunion dénombrable).*

Propriété 3 (Propriétés d'une tribu)
Si $\mathcal{A}$ *est une tribu sur un ensemble* Ω, *alors*

(*i*) $\emptyset \in \mathcal{A}$;
(*ii*) $\forall (A, B) \in \mathcal{A}^2$, $A \cup B \in \mathcal{A}$, $A \cap B \in \mathcal{A}$ *et* $A \setminus B \in \mathcal{A}$.
(*iii*) $\forall (A_n) \in \mathcal{A}^{\mathbb{N}}$, $\bigcap_{n=0}^{+\infty} A_n \in \mathcal{A}$.

Définition 4 (Espace probabilisable)
On appelle **espace probabilisable** *tout couple* $(\Omega, \mathcal{A})$ *constitué d'un ensemble* Ω *et d'une tribu* $\mathcal{A}$ *sur* Ω.

3) Evénements

Définition 5 (Evénements)
Si $(\Omega, \mathcal{A})$ *est un espace probabilisable, les parties* $A \in \mathcal{A}$ *sont appelées les* **événements** *de l'univers* Ω.

Vocabulaire

Rappelons le vocabulaire classique des événéments en probabilités. Etant donné un espace probabilisable $(\Omega, \mathcal{A})$:

- $\emptyset$ est appelé **événement impossible** ;
- Ω est appelé **événement certain** ;
- Les événements singletons $\{\omega\}$ sont appelés **événements élémentaires** ;
- Le complémentaire $\overline{A}$ est appelé **événement contraire de** A ;
- L'événement $A \cap B$ est appelé **intersection** des événements A et B ;
- L'événément $A \cup B$ est appelé **réunion** des événements A et B ;
- On dit que **l'événement $\boldsymbol{A}$ implique l'événement $\boldsymbol{B}$** lorsque $A \subset B$;
- On dit que deux événements A et B sont **incompatibles** (ou **disjoints**) lorsque $A \cap B = \emptyset$, ce qui signifie que A et B ne peuvent pas se réaliser en même temps.

II Probabilités

$(\Omega, \mathcal{A})$ désigne un espace probabilisable.

1) Définition

Définition 6 (Probabilité sur un espace probabilisable)
On appelle **probabilité sur $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{A}})$** *toute application* $\mathbb{P} : \mathcal{A} \to \mathbb{R}^+$ *vérifiant :*

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

(*ii*) *pour toute suite* $(A_n) \in \mathcal{A}^{\mathbb{N}}$ *d'événements deux à deux incompatibles (c'est-à-dire tels que* $n \neq m \implies A_n \cap A_m = \emptyset$*) :*

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

(propriété de σ-additivité)

Vocabulaire
On dit aussi **mesure de probabilité**.

Définition 7 (Espace probabilisé)
On appelle **espace probabilisé** *tout triplet* $(\Omega, \mathcal{A}, \mathbb{P})$*, où* Ω *est un ensemble,* $\mathcal{A}$ *est une tribu sur* Ω *et* $\mathbb{P} : \mathcal{A} \to \mathbb{R}^+$ *une probabilité.*

2) Propriétés élémentaires

Propriété 8 (Premières propriétés)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Alors :*

(*i*) $\mathbb{P}(\emptyset) = 0$;

(*ii*) *Si* $A_0, \cdots, A_n$ *sont des événements deux à deux incompatibles, alors*

$$\mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) = \sum_{k=0}^{n} \mathbb{P}(A_k);$$

(*iii*) *pour tout* $A \in \mathcal{A}$, $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$;

(*iv*) *pour tout* $A \in \mathcal{A}$, $\mathbb{P}(A) \in [0, 1]$.

Propriété 9 (Croissance et probabilité d'une réunion)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé et* A, B *deux événements. Alors :*

(*i*) $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$.

(*ii*) $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$.

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

Corollaire 10 (Sous-additivité finie)
Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace probabilisé. Si* $A_0, \cdots, A_n$ *sont des événements, alors*

$$\mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) \leq \sum_{k=0}^{n} \mathbb{P}(A_k).$$

3) Continuité monotone

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

Théorème 11 (Continuité croissante)
Si $(B_n) \in \mathcal{A}^{\mathbb{N}}$ *est une suite croissante d'événements (i.e.* $B_n \subset B_{n+1}$ *pour tout* $n \in \mathbb{N}$*), alors*

$$\mathbb{P}(B_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} B_k\right).$$

Corollaire 12 (Limite d'une réunion)
Si $(A_n) \in \mathcal{A}^{\mathbb{N}}$ *est une suite quelconque d'événements, alors*

$$\mathbb{P}\left(\bigcup_{k=0}^{n} A_k\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(\bigcup_{k=0}^{+\infty} A_k\right).$$

Corollaire 13 (Sous σ-additivité)
Si $(A_n) \in \mathcal{A}^{\mathbb{N}}$ *est une suite quelconque d'événements, alors dans* $\mathbb{R}^+ \cup \{+\infty\}$*, on a :*

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

Théorème 14 (Continuité décroissante)
Si $(B_n) \in \mathcal{A}^{\mathbb{N}}$ *est une suite décroissante d'événements (i.e.* $B_{n+1} \subset B_n$ *pour tout* $n \in \mathbb{N}$*), alors*

$$\mathbb{P}(B_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(\bigcap_{k=0}^{+\infty} B_k\right).$$

Corollaire 15 (Limite d'une intersection)
Si $(A_n) \in \mathcal{A}^{\mathbb{N}}$ *est une suite quelconque d'événements, alors*

$$\mathbb{P}\left(\bigcap_{k=0}^{n} A_k\right) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}\left(\bigcap_{k=0}^{+\infty} A_k\right).$$

4) Evénements négligeables, presque sûrs

$(\Omega, \mathcal{A}, \mathbb{P})$ désigne toujours un espace probabilisé.

Définition 16 (Evénement négligeable, presque sûr)
Un événement $A \in \mathcal{A}$ *est dit* **négligeable** *lorsque* $\mathbb{P}(A) = 0$, **presque sûr** *lorsque* $\mathbb{P}(A) = 1$.

Propriété 17 (Propriétés des événements négligeables et presque sûrs)
A *et* B *désignent deux événements de la tribu* $\mathcal{A}$.

- *(i) Si* $A \subset B$ *et si* B *est négligeable, alors* A *est négligeable.*
- *(ii) Une réunion finie ou dénombrable d'événements négligeables est un événement négligeable.*
- *(iii) Si* $A \subset B$ *et si* A *est presque sûr, alors* B *est presque sûr.*
- *(iv) Une intersection finie ou dénombrable d'événements presque sûrs est un événement presque sûr.*

5) Probabilités sur un univers au plus dénombrable

Dans cette section, Ω est un univers **fini ou dénombrable**. On le munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$.

Définition 18 (Probabilités élémentaires)
Soit $\mathbb{P}$ *une probabilité sur* $(\Omega, \mathcal{A})$*. On appelle* **probabilités élémentaires** *la famille de réels*

$$(p_\omega)_{\omega\in\Omega} = (\mathbb{P}(\{\omega\}))_{\omega\in\Omega}.$$

Propriété 19 (Propriétés des probabilités élémentaires)
La famille $(p_\omega)_{\omega\in\Omega}$ *est une famille de réels positifs, sommable et de somme égale à* 1.

Vocabulaire
On dit alors que $(p_\omega)_{\omega\in\Omega}$ est une **distribution de probabilités discrètes sur** Ω.

Théorème 20 (Détermination d'une probabilité sur un univers au plus dénombrable)
Soit $(p_\omega)_{\omega\in\Omega}$ *une famille de réels positifs, sommable et de somme égale à* 1. *Alors, il existe une unique probabilité* $\mathbb{P}$ *sur* $(\Omega, \mathcal{A})$ *telle que*

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

De plus, celle-ci est donnée par :

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

Vocabulaire
On parle alors d'**espace probabilisé discret** $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$.

III Probabilités conditionnelles

On fixe un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, où Ω est un univers quelconque, non nécessairement dénombrable.

1) Définition

Définition 21 (Probabilité conditionnelle de A sachant B)
Soient A et B deux événements.
Si $\mathbb{P}(B) \neq 0$, on définit la **probabilité de A sachant B** *par* $\mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
On la note aussi $\mathbb{P}(A|B)$.

Propriété 22 (Toute probabilité conditionnelle est une probabilité)
Si on a $\mathbb{P}(B) \neq 0$, alors l'application

$$\mathbb{P}_B : \left\{ \begin{array}{ccc} \mathcal{A} & \longrightarrow & \mathbb{R}^+ \\ A & \longmapsto & \mathbb{P}_B(A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \end{array} \right.$$

est une probabilité sur $(\Omega, \mathcal{A})$.

2) Formule des probabilités composées

Théorème 23 (Formule des probabilités composées)
Soient $A_1, \dots, A_n \in \mathcal{A}$ (avec $n \geq 2$) t.q. $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) \neq 0$. Alors :

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \times \mathbb{P}_{A_1}(A_2) \times \mathbb{P}_{A_1 \cap A_2}(A_3) \times \dots \times \mathbb{P}_{\bigcap_{i=1}^{n-1} A_i}(A_n).$$

3) Formule des probabilités totales

Définition 24 (Système complet d'événements, système quasi-complet d'événements)
Soit $(E_i)_{i \in I}$ une famille d'événements (avec I fini ou dénombrable). On dit que :

(i) $(E_i)_{i \in I}$ *est un* **système complet d'événements** *(en abrégé SCE) lorsque les E_i sont deux à deux incompatibles et $\bigcup_{i \in I} E_i = \Omega$.*

(ii) $(E_i)_{i \in I}$ *est un* **système quasi-complet d'événements** *(en abrégé SQCE) lorsque les E_i sont deux à deux incompatibles et $\mathbb{P}\left(\bigcup_{i \in I} E_i\right) = 1$.*

Théorème 25 (Formule des probabilités totales)
Soit $(E_i)_{i \in I}$ un système complet (ou quasi-complet) d'événements (I fini ou dénombrable).

(i) Pour tout $A \in \mathcal{A}$:

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap E_i).$$

(ii) Si de plus $\mathbb{P}(E_i) \neq 0$ pour tout $i \in I$, alors pour tout $A \in \mathcal{A}$:

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}_{E_i}(A) \times \mathbb{P}(E_i).$$

4) Formules de Bayes

Propriété 26 (Première formule de Bayes)
Soit $A, B \in \mathcal{A}$ avec $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(B) \neq 0$. Alors $\mathbb{P}_B(A) = \dfrac{\mathbb{P}_A(B)\mathbb{P}(A)}{\mathbb{P}(B)}$.

Corollaire 27 (Seconde formule de Bayes)
Si $(A_i)_{i\in I}$ *est un système quasi-complet d'événements tel que* $\forall i \in I$, $\mathbb{P}(A_i) \neq 0$, *alors :*

$$\forall B \in \mathcal{A},\ \forall j \in I, \quad \mathbb{P}_B(A_j) = \frac{\mathbb{P}_{A_j}(B)\mathbb{P}(A_j)}{\sum\limits_{i\in I} \mathbb{P}_{A_i}(B)\mathbb{P}(A_i)}.$$

IV Indépendance d'événements

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

1) Couples d'événements indépendants

Définition 28 (Indépendance de deux événements)
On dit que deux événements A et B sont **indépendants** *lorsque* $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Propriété 29 (Indépendance et complémentaire)
Soient A et B deux événements. Alors on a équivalence entre :

(i) *A et B sont indépendants.*

(ii) *A et $\overline{B}$ sont indépendants.*

(iii) *$\overline{A}$ et B sont indépendants.*

(iv) *$\overline{A}$ et $\overline{B}$ sont indépendants.*

2) Familles d'événements indépendants

Définition 30 (Indépendance d'une famille d'événements)
On dit que les événements d'une famille $(A_i)_{i \in I}$ sont **indépendants** *lorsque pour toute partie* **finie** *J de I :*

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

Vocabulaire
On dit aussi que les événements $(A_i)_{i \in I}$ sont **mutuellement indépendants**.

Propriété 31 (Sous-famille d'une famille d'événements indépendants)
Toute sous-famille d'une famille d'événements indépendants est elle-même constituée d'événements indépendants.

Propriété 32 (Indépendance et complémentaires)
Soit $(A_i)_{i \in I}$ une famille d'événements. Pour chaque $i \in I$, on pose $B_i = A_i$ ou $\overline{A_i}$.
Si la famille $(A_i)_{i \in I}$ est constituée d'événements indépendants, alors la famille $(B_i)_{i \in I}$ aussi.

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*

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

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 $\boldsymbol{X}$ sachant $\boldsymbol{A}$** *la probabilité :*

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

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* $X \sim \mathcal{U}(\llbracket 1,n \rrbracket)$*, 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 $\mathbb{1}_A : \begin{cases} \Omega & \longrightarrow \quad \mathbb{R} \\ \omega & \longmapsto \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases} \end{cases}$ *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
  $V(X) = E(X^2) - E(X)^2 = G''_X(1) + G'_X(1) - G'_X(1)^2$.

CH19 : Calcul différentiel - aspects théoriques

Dans tout ce chapitre, E et F désigneront deux $\mathbb{R}$-espaces vectoriels normés de dimension finie, avec $\dim(E) \geq 1$, et Ω un ouvert non vide de E.
On notera en général $n = \dim(E)$, $m = \dim(F)$. Lorsqu'on munit E et F de bases, elles seront notées respectivement $\mathcal{B}_E = (e_1, \cdots, e_n)$, $\mathcal{B}_F = (e'_1, \cdots, e'_m)$.
On aura la plupart du temps en contexte une application $f : \Omega \subset E \to F$.

I Applications différentiables

1) Dérivée directionnelle

Définition 1 (Dérivée directionnelle suivant un vecteur)
Soit $f : \Omega \to F$, soit $a \in \Omega$ et $v \in E$.
On dit que f est **dérivable en a selon le vecteur v** *lorsque la fonction d'une variable réelle $t \mapsto f(a + tv)$ est dérivable en 0. On note alors*

$$D_v f(a) = \left[\frac{d}{dt}(t \mapsto f(a + tv))\right]_{t=0} = \lim_{t \to 0} \frac{1}{t}\big(f(a + tv) - f(a)\big) \in F,$$

et on dit que $D_v f(a)$ est la **dérivée directionnelle de f en a selon v**.
Lorsqu'elle est bien définie, l'application $D_v f : \left\{\begin{array}{ccc} \Omega & \longrightarrow & F \\ a & \longmapsto & D_v f(a) \end{array}\right.$ est appelée **dérivée directionnelle de f selon v**.

2) Différentiabilité en un point

Revenons au cas des fonctions à variable réelle : si $f : I \to F$ est dérivable en a (avec I intervalle ouvert de $\mathbb{R}$), alors pour tout $h \in \mathbb{R}$ tel que $a + h \in I$,

$$f(a + h) = f(a) + hf'(a) + h\varepsilon(h), \text{ où } \varepsilon(h) \underset{h \to 0}{\longrightarrow} 0_F,$$

et on remarque que l'application $\left\{\begin{array}{ccc} \mathbb{R} & \longrightarrow & F \\ h & \longmapsto & hf'(a) \end{array}\right.$ est linéaire, donc continue puisqu'on est en dimension finie (ce qui assure $f(a + h) \underset{h \to 0}{\longrightarrow} f(a)$).

On va donc exploiter ce point de vue :

$$\boxed{\text{lorsque } h \text{ est petit,} \qquad f(a + h) = \text{valeur } f(a) + \text{application linéaire en } h + \text{erreur en } o(h)},$$

c'est la grande idée du calcul différentiel, qui permet de généraliser la notion de dérivée.

Définition 2 (Application différentiable en un point)
Soit $f : \Omega \to F$ et soit $a \in \Omega$.
On dit que f est **différentiable en a** *lorsqu'il existe une application linéaire $L : E \to F$ et une application $\varepsilon : E \to F$ telle que $\varepsilon(h) \underset{h \to 0_E}{\longrightarrow} 0_F$ et pour tout $h \in E$ tel que $a + h \in \Omega$, on a*

$$f(a + h) = f(a) + L(h) + \|h\|\varepsilon(h).$$

Propriété 3 (Différentiable implique continue)
Si f est différentiable en a, alors f est continue en a.

Théorème 4 (Unicité de la différentielle et lien avec les dérivées directionnelles)
Si f est différentiable en a, alors :

(i) *pour tout $v \in E$, $D_v f(a)$ existe et $L(v) = D_v f(a)$;*

(ii) *l'application linéaire L est unique. On dit que L est la* **différentielle de f en a** *(ou l'***application linéaire tangente à f en a***). On note $L = df(a)$ (ou df_a), d'où la relation :*

$$\forall v \in E, \qquad df(a)(v) = D_v f(a).$$

Notation
Ainsi, lorsque f est différentiable en a, on peut écrire

$$f(a+h) = f(a) + df(a)(h) + o(h),$$

avec $df(a) \in \mathcal{L}(E, F)$.
Pour tout vecteur $h \in E$, on note en général $df(a) \cdot h$ plutôt que $df(a)(h)$ (comme on travaille en dimension finie, on peut assimiler cela à un produit matriciel). Ainsi, la formule du point (ii) se réécrit :

$$\forall v \in E, \qquad df(a) \cdot v = D_v f(a).$$

Méthode
Par contraposition, on a donc une méthode pour montrer qu'une fonction n'est pas différentiable en un point : trouver au moins un vecteur selon lequel elle n'admet pas de dérivée directionnelle en ce point !

Enfin, le résultat qui suit fait le lien explicite entre les notions de différentiabilité et de dérivabilité pour une fonction de la variable réelle :

Propriété 5 (Différentiabilité d'une fonction d'une variable réelle)
Soient I un intervalle ouvert de $\mathbb{R}$, soit $f : I \to F$ et $a \in I$.
Alors f est dérivable en a ssi f est différentiable en a et, lorsque c'est le cas, on a

$$f'(a) = df(a) \cdot 1,$$

et donc,

$$\forall h \in \mathbb{R}, \qquad df(a) \cdot h = hf'(a).$$

3) Différentiabilité sur un ouvert

Définition 6 (Application différentiable sur un ouvert)
On dit que $f : \Omega \to F$ est différentiable sur Ω lorsque f est différentiable en tout point de Ω.

Définition 7 (Application différentielle)
Si f est différentiable sur Ω, alors on peut définir l'application $df : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathcal{L}(E,F) \\ a & \longmapsto & df(a) \end{array} \right.$.
On l'appelle **différentielle de f**.

Définition 8 (Fonctions coordonnées de f dans une base de F)
Etant donnée une base $\mathcal{B}_F = (e'_1, \dots, e'_m)$ de F et une application $f : \Omega \to F$, on appelle **fonctions coordonnées de f dans $\mathcal{B}_F$** *les applications $f_1 : \Omega \to \mathbb{R}, \cdots, f_m : \Omega \to \mathbb{R}$ définies par :*

$$\forall x \in \Omega, \qquad f(x) = \sum_{j=1}^{m} f_j(x) e'_j.$$

On pourra noter en abrégé $[f]_{\mathcal{B}_F} = (f_1, \cdots, f_m)$.

Vocabulaire
Les fonctions coordonnées $(f_1, \cdots, f_m)$ sont aussi appelées les **composantes** de f dans la base de F choisie.

Comme on l'avait déjà aperçu pour la dérivabilité des fonctions vectorielles, la différentiabilité peut être caractérisée sur les fonctions coordonnées de f.

Propriété 9 (Différentiabilité coordonnée par coordonnée)
Soit $f : \Omega \to F$ avec $m = \dim(F) \geq 1$. On note $(f_1, \dots, f_m)$ les fonctions coordonnées de f dans une base quelconque $\mathcal{B}_F$.
Alors f est différentiable ssi $f_1, \dots, f_m$ sont différentiables et dans ce cas :

$$\forall a \in \Omega, \qquad [df(a)]_{\mathcal{B}_F} = (df_1(a), \cdots, df_m(a)).$$

4) Dérivées partielles dans une base

Soit $\mathcal{B}_E = (e_1, \ldots, e_n)$ une base de E. On décompose un vecteur h quelconque dans cette base : $h = \sum_{i=1}^{n} h_i e_i$. On obtient alors, si f est différentiable en un certain $a \in \Omega$:

$$df(a) \cdot h = df(a) \cdot \left(\sum_{i=1}^{n} h_i e_i \right) = \sum_{i=1}^{n} h_i df(a) \cdot e_i,$$

par linéarité de $df(a)$. Et donc,

$$df(a) \cdot h = \sum_{i=1}^{n} h_i D_{e_i} f(a),$$

d'après la formule qui relie différentielle en a et dérivées directionnelles en a.

Il suffit donc de connaître les n dérivées directionnelles $D_{e_i} f(a)$ pour connaître l'application $df(a)$.

Ceci justifie l'introduction de la notion suivante :

Définition 10 (Dérivées partielles en un point par rapport à une base)
Soit $\mathcal{B}_E = (e_1, \ldots, e_n)$ une base de E.

(*i*) *Lorsqu'elles existent, on note $\dfrac{\partial f}{\partial x_1}(a), \ldots, \dfrac{\partial f}{\partial x_n}(a)$ les dérivées directionnelles de f en a selon les vecteurs $e_1, \ldots, e_n$. Pour tout $i \in [\![1, n]\!]$, on appelle $\dfrac{\partial f}{\partial x_i}(a)$ la **i-ème dérivée partielle** de f en a relativement à la base $\mathcal{B}_E$.*
Par définition, on a donc :

$$\forall i \in [\![1, n]\!], \qquad \frac{\partial f}{\partial x_i}(a) = D_{e_i} f(a) = \left[\frac{d}{dt}(t \mapsto f(a + te_i)) \right]_{t=0} = \lim_{t \to 0} \frac{1}{t}\big(f(a + te_i) - f(a)\big)$$

lorsque cette limite existe.

(*ii*) *Sous réserve d'existence, l'application $\dfrac{\partial f}{\partial x_i} =: \begin{cases} \Omega & \longrightarrow & F \\ a & \longmapsto & \dfrac{\partial f}{\partial x_i}(a) \end{cases}$ est appelée **i-ème fonction dérivée partielle** de f (relativement à la base $\mathcal{B}_E$), pour tout $i \in [\![1, n]\!]$.*

Notation
On note aussi $\partial_i f(a) = \dfrac{\partial f}{\partial x_i}(a) = \dfrac{\partial f}{\partial e_i}(a)$ la i^e dérivée partielle de f en a.
La notation $\dfrac{\partial f}{\partial e_i}(a)$ est la plus claire car elle indique le vecteur suivant lequel on dérive f.
Mais la notation la plus utilisée est $\dfrac{\partial f}{\partial x_i}(a)$, il faut bien comprendre que le "x_i" est symbolique, il indique juste que l'on dérive par rapport à la i^e coordonnée, et sous-entend que les coordonnées d'un vecteur $x \in E$ dans $\mathcal{B}_E$ seront notées $(x_1, \cdots, x_n)$.

Propriété 11 (Lien entre différentielle et dérivées partielles)
Si f est différentiable en a alors les dérivées partielles de f en a relativement à toute base $\mathcal{B}_E$ existent et vérifient : $\forall i \in [\![1, n]\!], \quad \dfrac{\partial f}{\partial x_i}(a) = df(a) \cdot e_i$. De plus, on a l'expression :

$$\forall h \in E, \qquad df(a) \cdot h = \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a),$$

où $[h]_{\mathcal{B}_E} = (h_1, \cdots, h_n)$.

Corollaire 12 (Formule de Taylor-Young à l'ordre 1)
Si f est différentiable en a, alors on a le $DL_1(a)$:

$$f(a+h) = f(a) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a) + o(h),$$

où les $\dfrac{\partial f}{\partial x_i}(a)$ désignent les dérivées partielles en a dans une quelconque base de E.

5) Classe $\mathcal{C}^1$

On a la généralisation suivante du concept de classe $\mathcal{C}^1$ déjà vu pour les fonctions à variable réelle :

Définition 13 (Fonction de classe $\mathcal{C}^1$)
Une fonction $f : \Omega \to F$ est dite **de classe $\mathcal{C}^1$** *lorsqu'elle est différentiable et lorsque l'application $df : \Omega \to \mathcal{L}(E,F)$ est continue.*

Théorème 14 (Caractérisation des fonctions de classe $\mathcal{C}^1$)
Une fonction $f : \Omega \to F$ est de classe $\mathcal{C}^1$ ssi ses dérivées partielles (relativement à une base quelconque de E) existent en tout point de Ω et sont continues sur Ω.

Finissons par une évidence :

Propriété 15 ($\mathcal{C}^1$ implique $\mathcal{C}^0$)
Les fonctions de classe $\mathcal{C}^1$ sont continues.

En résumé, pour $f : \Omega \to \mathbb{R}$:

$$\boxed{f \text{ de classe } \mathcal{C}^1 \implies f \text{ différentiable } \begin{cases} \implies & f \text{ continue} \\ \implies & f \text{ possède des dérivées partielles en tout point de } \Omega \end{cases}},$$

(et les implications réciproques sont fausses) et

$$\boxed{f \text{ de classe } \mathcal{C}^1 \iff f \text{ possède des dérivées partielles } \textbf{continues} \text{ en tout point de } \Omega.}$$

6) Matrice jacobienne

En tant qu'application linéaire, la différentielle $df(a) \in \mathcal{L}(E,F)$ peut se représenter par sa matrice A dans un couple de bases. En notant $(e_1, \cdots, e_n)$ une base de E et $(e'_1, \cdots, e'_m)$ une base de F, les colonnes de cette matrice A vont être définies par

$$\forall j \in [\![1,n]\!], \qquad C_j(A) = [df(a)\cdot e_j]_{\mathcal{B}_F} = \left[\frac{\partial f}{\partial x_j}(a)\right]_{\mathcal{B}_F} = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_j}(a) & \cdots & \dfrac{\partial f_m}{\partial x_j}(a) \end{pmatrix}^\top$$

(en notant $(f_1, \cdots, f_m)$ les coordonnées de f dans la base $\mathcal{B}_F$ et en dérivant "coordonnée par coordonnée").

Définition 16 (Matrice jacobienne dans un couple de bases)
Soit $\mathcal{B}_E = (e_1, \cdots, e_n)$ une base de E et $\mathcal{B}_F = (e'_1, \cdots, e'_m)$ une base de F.
Soit $f : \Omega \to F$ une application différentiable en $a \in \Omega$.
On appelle **matrice jacobienne de f dans les bases $(\mathcal{B}_E, \mathcal{B}_F)$** *la matrice de l'application linéaire $df(a) \in \mathcal{L}(E,F)$ dans les bases $(\mathcal{B}_E, \mathcal{B}_F)$. On la note :*

$$Jac\, f(a) = Mat_{\mathcal{B}_E,\mathcal{B}_F}\big(df(a)\big) = \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} \in \mathcal{M}_{m,n}(\mathbb{R}),$$

où $(f_1, \cdots, f_m)$ sont les coordonnées de f dans la base $\mathcal{B}_F$ et $(\frac{\partial f_i}{\partial x_1}, \cdots, \frac{\partial f_i}{\partial x_n})$ désignent les dérivées partielles des $f_i : \Omega \to \mathbb{R}$ relativement à la base $\mathcal{B}_E$.

Notation
On pourra noter plus simplement $J_f(a)$ la matrice jacobienne de f en a.

7) Gradient d'une fonction numérique

Dans ce paragraphe, les fonctions sont à valeurs dans $\mathbb{R}$ (d'où le terme "fonction numérique"), et définies sur un ouvert Ω de E, où E est un **espace euclidien** ($\mathbb{R}$-e.v. de dimension finie muni d'un produit scalaire $(\cdot|\cdot)$).

Théorème 17 (Gradient en un point)
Soit $f:\Omega\to\mathbb{R}$ différentiable en $a\in\Omega$. Il existe un unique vecteur de E, noté $\nabla f(a)$, tel que

$$\forall h\in E,\qquad df(a)\cdot h=(\nabla f(a)|h).$$

Ce vecteur $\nabla f(a)$ est appelé **gradient de f en a**.

Propriété 18 (Expression du gradient en base orthonormée)
Soit $\mathcal{B}_E=(e_1,\cdots,e_n)$ une base orthonormée de E, et $f:\Omega\to\mathbb{R}$ différentiable en $a\in\Omega$. Alors

$$[\nabla f(a)]_{\mathcal{B}_E}=\left(\frac{\partial f}{\partial x_1}(a),\cdots,\frac{\partial f}{\partial x_n}(a)\right).$$

Corollaire 19 (Fonction gradient d'une fonction de classe $\mathcal{C}^1$)
Si $f:\Omega\to\mathbb{R}$ est de classe $\mathcal{C}^1$, alors la fonction gradient $\nabla f:\left\{\begin{array}{ccc}\Omega & \longrightarrow & E\\ a & \longmapsto & \nabla f(a)\end{array}\right.$ est continue.

Vocabulaire
On dit que le gradient ∇f est un **champ de vecteurs** de E (nom donné aux applications $\Omega\subset E\to E$).

II Opérations sur les applications différentiables

On revient au cadre général des applications $f : \Omega \subset E \to F$, où E et F sont des $\mathbb{R}$-evn de dimension finie et Ω est un ouvert de E.

1) Combinaisons linéaires

Théorème 20 (Différentielle d'une combinaison linéaire en un point)
Soient $f : \Omega \to F$, $g : \Omega \to F$ et $(\lambda, \mu) \in \mathbb{R}^2$.
Si f et g sont différentiables en $a \in \Omega$, alors $\lambda f + \mu g$ aussi et

$$d(\lambda f + \mu g)(a) = \lambda df(a) + \mu dg(a).$$

Corollaire 21 (Différentielle d'une combinaison linéaire)
Si f et g sont différentiables sur Ω alors, pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $\lambda f + \mu g$ est différentiable sur Ω, avec

$$d(\lambda f + \mu g) = \lambda df + \mu dg.$$

Il s'ensuit que l'ensemble des fonctions différentiables de Ω vers F est un sous-espace vectoriel de F^Ω (les fonctions de Ω dans F).

2) Composition

G désigne ici un autre $\mathbb{R}$-evn de dimension finie.

Théorème 22 (Différentielle d'une composée en un point)
Soient Ω un ouvert de E et Ω' un ouvert de F. Soit $f : \Omega \to F$ et $g : \Omega' \to G$ telles que $f(\Omega) \subset \Omega'$.
Si f est différentiable en $a \in \Omega$ et si g est différentiable en $f(a)$, alors $g \circ f$ est différentiable en a et

$$d(g \circ f)(a) = \big[dg\big(f(a)\big)\big] \circ df(a).$$

Corollaire 23 (Composition extérieure avec une fonction réelle)
Soient I un intervalle réel, $f : \Omega \to \mathbb{R}$ et $\varphi : I \to \mathbb{R}$ telles que $f(\Omega) \subset I$.
Si f est différentiable en $a \in \Omega$ et si φ est dérivable en $f(a)$, alors $\varphi \circ f$ est différentiable en a et :

$$d(\varphi \circ f)(a) = \varphi'(f(a)) \times df(a)$$

(produit du nombre réel $\varphi'(f(a))$ par l'application linéaire $df(a)$).

Corollaire 24 (Dérivée le long d'un arc)
Soient I un intervalle réel, $\gamma : I \to E$ et $f : \Omega \to F$ telles que $\gamma(I) \subset \Omega$.
Si γ est dérivable en $t_0 \in I$ et si f est différentiable en $\gamma(t_0)$, alors $f \circ \gamma$ est dérivable en t_0 et

$$(f \circ \gamma)'(t_0) = df\big(\gamma(t_0)\big) \cdot \gamma'(t_0).$$

Corollaire 25 (Dérivée le long d'un arc tracé dans $\mathbb{R}^n$)
On suppose que $E = \mathbb{R}^n$. Soit Ω un ouvert non vide de $\mathbb{R}^n$, soit $f : \Omega \to F$, soient $x_1 : I \to \mathbb{R}, \cdots, x_n : I \to \mathbb{R}$, où I est un intervalle réel.
On suppose que $\big(x_1(t), \ldots, x_n(t)\big) \in \Omega$ pour tout $t \in I$.
Soit $g : I \to F$ définie par

$$g(t) = f\big(x_1(t), \ldots, x_n(t)\big).$$

Si les x_i sont dérivables en $t_0 \in I$ et si f est différentiable en $\big(x_1(t_0), \ldots, x_n(t_0)\big)$, alors g est dérivable en t_0 et

$$\begin{aligned} g'(t_0) &= x_1'(t_0)\frac{\partial f}{\partial x_1}\big(x_1(t_0), \ldots, x_n(t_0)\big) + \cdots + x_n'(t_0)\frac{\partial f}{\partial x_n}\big(x_1(t_0), \ldots, x_n(t_0)\big) \\ &= \sum_{i=1}^{n} x_i'(t_0)\frac{\partial f}{\partial x_i}\big(x_1(t_0), \ldots, x_n(t_0)\big), \end{aligned}$$

où les $\dfrac{\partial f}{\partial x_i}$ désignent les dérivées partielles de f relativement à la base canonique de $\mathbb{R}^n$.

3) Composition bilinéaire, multilinéaire, produit

On a déjà vu que toute application linéaire $L : E \to F$ est différentiable et

$$\forall a \in E, \qquad dL(a) = L.$$

Pour les applications bilinéaires, la situation est un peu plus compliquée :

Théorème 26 (Différentielle d'une application bilinéaire)
Soient F, G, H trois $\mathbb{R}$-evn de dimension finie, et $B : F \times G \to H$ une application bilinéaire. Alors B est différentiable et

$$\forall (a,b) \in F \times G, \qquad dB(a,b) : \left\{ \begin{array}{ccc} F \times G & \longrightarrow & H \\ (u,v) & \longmapsto & B(u,b) + B(a,v) \end{array} \right. .$$

Symboliquement, on peut écrire

$$\forall (a,b) \in F \times G, \qquad dB(a,b) = B(\cdot, b) + B(a, \cdot).$$

On peut généraliser :

Théorème 27 (Différentielle d'une application multilinéaire)
Soient $F_1, \cdots, F_p, G$ des $\mathbb{R}$-evn de dimension finie, et $M : F_1 \times \cdots \times F_p \to G$ une application multilinéaire. Alors M est différentiable et pour tout $(a_1, \cdots, a_p) \in F_1 \times \cdots \times F_p$:

$$dM(a_1, \cdots, a_p) : \left\{ \begin{array}{ccc} F_1 \times \cdots \times F_p & \longrightarrow & G \\ (u_1, \cdots, u_p) & \longmapsto & M(u_1, a_2, \cdots, a_p) + \cdots + M(a_1, \cdots, a_{p-1}, u_p) \end{array} \right. .$$

Corollaire 28 (Différentielle d'une composition multilinéaire)
Soient $f_1 : \Omega \to F_1, \cdots, f_p : \Omega \to F_p$ des applications différentiables en $a \in \Omega$ (ouvert de E).
Soit $M : F_1 \times \cdots \times F_p \to G$ une application multilinéaire.
Alors, l'application

$$M(f_1, \cdots, f_p) : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & G \\ x & \longmapsto & M(f_1(x), \cdots, f_p(x)) \end{array} \right. ,$$

est différentiable en a, et

$$d(M(f_1, \cdots, f_p))(a) : h \mapsto M(df_1(a) \cdot h, f_2(a), \cdots, f_p(a)) + \cdots + M(f_1(a), \cdots, f_{p-1}(a), df_p(a) \cdot h).$$

Corollaire 29 (Différentielle d'un produit)
Soient $f, g : \Omega \to F$. Si F est une $\mathbb{R}$-algèbre de dimension finie et si f et g sont différentiables en $a \in \Omega$, alors $f \times g : \Omega \to F$ est différentiable en a et

$$d(fg)(a) : h \mapsto (df(a) \cdot h) \times g(a) + f(a) \times (dg(a) \cdot h),$$

qu'on peut abréger en

$$d(fg)(a) = df(a) \times g(a) + f(a) \times dg(a).$$

Ainsi le produit de deux fonctions différentiables est différentiables, et l'ensemble des fonctions différentiables de Ω vers F est une sous-algèbre de F^Ω.

4) Opérations sur les dérivées partielles

Vu qu'une dérivée partielle est une dérivée "classique" (i.e. une dérivée de fonction d'une variable réelle), on dispose directement de formules pour les dérivées partielles d'une somme, d'un produit, etc. (en reprenant les formules de dérivation du CH.15 sur les fonctions vectorielles), et ce même si les fonctions en jeu ne sont pas différentiables.

Théorème 30 (Dérivées partielles d'une combinaison linéaire)
Soient $f, g : \Omega \to F$ et $(\lambda, \mu) \in \mathbb{R}^2$.
Si f et g admettent des dérivées partielles en $a \in \Omega$, alors $\lambda f + \mu g$ aussi et

$$\forall i \in [\![1, n]\!], \qquad \partial_i(\lambda f + \mu g)(a) = \lambda \partial_i f(a) + \mu \partial_i g(a).$$

Théorème 31 (Dérivées partielles d'une composition bilinéaire)
Soient F, G et H trois $\mathbb{R}$-espaces vectoriels de dimension finie, $f : \Omega \to F$, $g : \Omega \to G$ et $B : F \times G \to H$ une application bilinéaire.
Si f et g admettent des dérivées partielles en $a \in \Omega$, alors $B(f, g)$ aussi et

$$\forall i \in [\![1, n]\!], \qquad \partial_i\big(B(f, g)\big)(a) = B\big(\partial_i f(a), g(a)\big) + B\big(f(a), \partial_i g(a)\big).$$

Corollaire 32 (Dérivées partielles d'un produit)
Soient $f, g : \Omega \to F$. Si F est une $\mathbb{R}$-algèbre et si f et g admettent des dérivées partielles en a, alors $f \times g$ aussi et

$$\forall i \in [\![1, n]\!], \qquad \partial_i(fg)(a) = \big(\partial_i f(a)\big) \times g(a) + f(a) \times \big(\partial_i g(a)\big).$$

Lorsqu'on compose deux applications différentiables, on peut exprimer simplement les dérivées partielles de $g \circ f$ en fonction de celles de g et de f. C'est l'objet du résultat suivant :

Théorème 33 (Règle de la chaîne)
Soient Ω un ouvert de E et Ω' un ouvert de F. Soit $f : \Omega \to F$ et $g : \Omega' \to G$ telles que $f(\Omega) \subset \Omega'$.
On munit E d'une base $\mathcal{B}_E = (e_1, \cdots, e_n)$ et F d'une base $\mathcal{B}_F = (e'_1, \cdots, e'_m)$.
Si f est différentiable en $a \in \Omega$ et si g est différentiable en $f(a)$, alors $g \circ f$ est différentiable en a, donc admet des dérivées partielles en a (relativement à la base $\mathcal{B}_E$), et :

$$\forall i \in [\![1, n]\!], \qquad \partial_i(g \circ f)(a) = \sum_{k=1}^{m} \partial_i f_k(a) \times \partial_k g\big(f(a)\big),$$

c'est-à-dire

$$\forall i \in [\![1, n]\!], \qquad \frac{\partial(g \circ f)}{\partial e_i}(a) = \sum_{k=1}^{m} \frac{\partial f_k}{\partial e_i}(a) \times \frac{\partial g}{\partial e'_k}(f(a)),$$

où $(f_1, \cdots, f_m)$ désignent les fonctions coordonnées de f dans la base $\mathcal{B}_F$.

Corollaire 34 (Règle de la chaîne pour une composée $\mathbb{R}^n \to \mathbb{R}^m \to \mathbb{R}$)
Soient $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$, $g : \Omega' \subset \mathbb{R}^m \to \mathbb{R}$ avec $f(\Omega) \subset \Omega'$.
Notons $(f_1, \cdots, f_m)$ les composantes de f.
Si f est différentiable en $a \in \Omega$, si g est différentiable en $f(a)$, alors les dérivées partielles de

$$g \circ f : \left\{ \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{R} \\ (x_1, \cdots, x_n) & \longmapsto & g\left(\underbrace{f_1(x_1, \cdots, x_n)}_{=y_1}, \cdots, \underbrace{f_m(x_1, \cdots, x_n)}_{=y_m}\right) \end{array} \right.$$

en $a = (a_1, \cdots, a_n)$ valent :

$$\forall i \in [\![1, n]\!], \qquad \frac{\partial (g \circ f)}{\partial x_i}(a) = \sum_{k=1}^{m} \frac{\partial f_k}{\partial x_i}(a) \times \frac{\partial g}{\partial y_k}(f(a)).$$

5) Formule d'intégration

Théorème 35 (Intégration le long d'un arc)
Soit $f : \Omega \to F$ de classe $\mathcal{C}^1$ et $\gamma : [0,1] \to \Omega$ un arc paramétré ("chemin") de classe $\mathcal{C}^1$, d'extrémités $\gamma(0) = a$ et $\gamma(1) = b$. Alors $f(b) - f(a) = \displaystyle\int_0^1 df\big(\gamma(t)\big) \cdot \gamma'(t) dt$.

Corollaire 36 (Caractérisation des fonctions constantes)
Si Ω est un ouvert connexe par arcs et si $f : \Omega \to F$ est de classe $\mathcal{C}^1$, alors f est constante si et seulement si $df : \Omega \to \mathcal{L}(E, F)$ est identiquement nulle.

III Classe d'une fonction

1) Dérivées partielles successives

Définition 37 (Dérivées partielles d'ordre k)
Soit $f : \Omega \to F$ et $\mathcal{B}_E = (e_1, \dots, e_n)$ une base de E.
Par convention, on dit que f est la dérivée partielle d'ordre 0 de f.
Pour $k \in \mathbb{N}$, lorsqu'elles existent, on appelle ***dérivées partielles d'ordre*** *$k+1$ de f relativement à la base $\mathcal{B}_E$ les dérivées partielles des dérivées partielles d'ordre k de f.*

Notation
Si on note $x_1, \cdots, x_n$ les coordonnées dans la base $\mathcal{B}_E$ de la variable $x \in \Omega$, on note :

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}} = \partial_{i_1}(\cdots(\partial_{i_k} f)\cdots).$$

2) Classe $\mathcal{C}^k$

Définition 38 (Fonction de classe $\mathcal{C}^k$)
Soit $k \in \mathbb{N}$. On dit que $f : \Omega \to F$ est **de classe $\mathcal{C}^k$** *lorsque ses dérivées partielles d'ordre k relativement à une base quelconque de E existent et sont continues.*
On dit que f est **de classe $\mathcal{C}^\infty$** *lorsqu'elle est de classe $\mathcal{C}^k$ pour tout $k \in \mathbb{N}$.*

Propriété 39 ($\mathcal{C}^{k+1} \subset \mathcal{C}^k$)
Toute fonction de classe $\mathcal{C}^{k+1}$ est de classe $\mathcal{C}^k$.

3) Opérations sur les applications de classe $\mathcal{C}^k$

Dans tout ce paragraphe, $k \in \mathbb{N} \cup \{+\infty\}$.

Théorème 40 (Structure d'espace vectoriel de $\mathcal{C}^k$.)
Soient $f, g : \Omega \to F$ et $(\lambda, \mu) \in \mathbb{R}^2$.
Si f et g sont de classe $\mathcal{C}^k$, alors $\lambda f + \mu g$ aussi.
Donc, l'ensemble $\mathcal{C}^k(\Omega, F)$ des applications de classe $\mathcal{C}^k$ de Ω vers F est un sous-espace vectoriel de F^Ω.

Théorème 41 (Composition bilinéaire de fonctions $\mathcal{C}^k$)
Soient $f : \Omega \to F$, $g : \Omega \to G$ et $B : F \times G \to H$ bilinéaire.
Si f et g sont de classe $\mathcal{C}^k$, alors $B(f, g)$ aussi.

Corollaire 42 (Structure de $\mathbb{R}$-algèbre des fonctions de classe $\mathcal{C}^k$)
Si F est une $\mathbb{R}$-algèbre, alors l'ensemble $\mathcal{C}^k(\Omega, F)$ est une sous-algèbre de F^Ω.

Théorème 43 (Composée de fonctions $\mathcal{C}^k$)
Soient $f : \Omega \to F$, $g : \Omega' \to G$, où Ω' est un ouvert de F tel que $f(\Omega) \subset \Omega'$.
Si f et g sont de classe $\mathcal{C}^k$, alors $g \circ f$ aussi.

Et on a enfin le critère classique et bien utile :

Théorème 44 (Classe $\mathcal{C}^k$ par composantes)
Soit $f : \Omega \to F$. Alors f est de classe $\mathcal{C}^k$ ssi les fonctions coordonnées de f dans une base quelconque de F sont de classe $\mathcal{C}^k$.

4) Théorème de Schwarz

Théorème 45 (Théorème de Schwarz)
Si $f : \Omega \to F$ est de classe $\mathcal{C}^2$, alors les dérivées partielles secondes relativement à une quelconque base de E vérifient :

$$\forall (i,j) \in \{1, \dots, n\}^2, \qquad \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

CH20 : Calcul différentiel - aspects appliqués

Dans ce dernier chapitre, les fonctions considérées sont définies sur un ouvert Ω de E (un $\mathbb{R}$-evn de dimension finie) et à valeurs dans $\mathbb{R}$.

I Vecteurs tangents à une partie

Définition 1 (Vecteur tangent à une partie)
Soient X une partie de E et $x \in X$.
On dit qu'un vecteur $v \in E$ est **tangent à X en x** *lorsqu'il existe $\varepsilon > 0$ et une application $\gamma :]-\varepsilon, \varepsilon[\to X$ dérivable en 0 tels que $\gamma(0) = x$ et $\gamma'(0) = v$.*

Vocabulaire
En général, une application $\gamma : I \subset \mathbb{R} \to E$ est appelée un **arc** (ou **arc paramétré**).
Un vecteur $v \in E$ est donc tangent à X en x lorsqu'il existe un arc tracé dans X qui passe par x, avec un vecteur vitesse égal à v.

Notation
On notera $T_x X$ l'ensemble des vecteurs tangents à X en x.

Théorème 2 (Structure d'hyperplan tangent)
Soit $g : \Omega \subset E \to \mathbb{R}$ de classe $\mathcal{C}^1$. On considère "l'hypersurface" de E :

$$X = g^{-1}(\{0\}) = \{x \in \Omega,\ g(x) = 0\}.$$

Alors, pour tout $x \in X$ tel que $dg(x) \neq 0_{\mathcal{L}(E,\mathbb{R})}$, on a

$$T_x X = Ker(dg(x)).$$

Ainsi, l'ensemble des vecteurs tangents à X en x est un hyperplan de E.
De plus, si E est euclidien, alors

$$T_x X = \{\nabla g(x)\}^{\perp}.$$

Vocabulaire
Dans ce cas, on appelle en général **(hyper)plan tangent** à X en x l'hyperplan **affine**

$$\mathcal{T}_x X = x + T_x X = x + Ker(dg(x)) = x + \{\nabla f(x)\}^{\perp}.$$

Pour tout $m \in E$, on a

$$m \in \mathcal{T}_x X \iff m - x \in T_x X \iff (m - x | \nabla f(x)) = 0.$$

II Problèmes d'optimisation

E désigne toujours un $\mathbb{R}$-espace vectoriel normé de dimension finie (notée $n \in \mathbb{N}^*$), et Ω un ouvert de E.

1) Condition nécessaire d'extremum local

Définition 3 (Minimum, maximum local)
Soit $A \subset E$ (pas nécessairement un ouvert), soit $f : A \to \mathbb{R}$ et $a \in A$.
On dit que f possède un **minimum local** *(resp.* **maximum local***) en a lorsqu'il existe $r > 0$ tel que $\forall x \in B(a, r) \cap A,\ f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$).*

Vocabulaire
Ces deux termes sont regroupés sous la dénomination plus générale d'**extremum local**.
Le minimum (resp. maximum) est dit **global** lorsque l'inégalité $f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$) est vraie pour tout $x \in A$.
Bien sûr, tout extremum global est en particulier un extremum local.

Ce résultat se généralise aux fonctions de plusieurs variables :

Définition 4 (Point critique d'une application différentiable)
Soit Ω un ouvert de E, et $f : \Omega \to \mathbb{R}$ une fonction différentiable.
On appelle **point critique** *de f tout point $a \in \Omega$ tel que $df(a) = 0_{\mathcal{L}(E,\mathbb{R})}$.*

Théorème 5 (Condition nécessaire d'extremum local)
Soit Ω un **ouvert** *de E, et $f : \Omega \to \mathbb{R}$ une fonction différentiable.*
Si f admet un extremum local en $a \in \Omega$, alors a est un point critique de f.

Méthode
Dans la recherche des extrema locaux d'une fonction différentiable **sur un ouvert** *et à valeurs réelles, on commence donc par chercher les points critiques, puis on discute au cas par cas pour déterminer si ce sont réellement des extrema.*

2) Conditions d'ordre 2

On va maintenant donner des outils plus précis pour connaître la nature d'un point critique de f, c'est-à-dire s'il s'agit d'un minimum local, d'un maximum local, ou d'un point col. Pour cela, on va établir un développement limité à l'ordre 2 de f au voisinage du point critique.

Théorème 6 (Formule de Taylor-Young à l'ordre 2)
Soit Ω un ouvert de E et $f \in \mathcal{C}^2(\Omega, \mathbb{R})$. Pour tout $a \in \Omega$, on a dans toute base $\mathcal{B}_E$ de E :

$$f(a+h) = f(a) + \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(a) + \frac{1}{2} \sum_{1 \leq i,j \leq n} h_i h_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a) + \underset{h \to 0_E}{o}(\|h\|^2),$$

où $(h_1, \cdots, h_n) = [h]_{\mathcal{B}_E}$.

Dans la suite, on se limite au cas de $E = \mathbb{R}^n$, muni de sa structure euclidienne canonique.

Définition 7 (Matrice hessienne)
Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ une fonction de classe $\mathcal{C}^2$. On appelle **matrice hessienne** *de f au point $a \in \Omega$ la matrice*

$$H_f(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{(i,j) \in [1,n]^2} \in \mathcal{S}_n(\mathbb{R}).$$

Corollaire 8 (Formule de Taylor-Young à l'ordre 2 dans $\mathbb{R}^n$)
Soit Ω un ouvert de $\mathbb{R}^n$ et $f \in \mathcal{C}^2(\Omega, \mathbb{R})$. Pour tout $a \in \Omega$, on a :

$$f(a+h) = f(a) + (\nabla f(a)|h) + \frac{1}{2}(H_f(a)h|h) + \underset{h \to 0_E}{o}(\|h\|^2),$$

ou encore

$$f(a+h) = f(a) + \nabla f(a)^\top h + \frac{1}{2} h^\top H_f(a) h + \underset{h \to 0_E}{o}(\|h\|^2).$$

Théorème 9 (Conditions d'ordre 2)
Soit Ω un ouvert de $\mathbb{R}^n$, soit $f \in \mathcal{C}^2(\Omega, \mathbb{R})$ et soit $a \in \Omega$.

(*i*) *Si f admet un minimum local en a, alors a est un point critique de f et $H_f(a) \in \mathcal{S}_n^+(\mathbb{R})$.*

(*ii*) *Si a est un point critique de f et si $H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R})$, alors f admet un minimum local en a.*

De même pour les maximum locaux en remplaçant $\mathcal{S}_n^+(\mathbb{R})$ par $\mathcal{S}_n^-(\mathbb{R})$ (i.e. $-H_f(a) \in \mathcal{S}_n^+(\mathbb{R})$) et $\mathcal{S}_n^{++}(\mathbb{R})$ par $\mathcal{S}_n^{--}(\mathbb{R})$ (i.e. $-H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R})$).

Corollaire 10 (Conditions d'ordre 2 avec le signe des valeurs propres)
Soit Ω un ouvert de $\mathbb{R}^n$, soit $f \in \mathcal{C}^2(\Omega, \mathbb{R})$ et soit $a \in \Omega$ un point critique de f.

(*i*) *Si $Sp(H_f(a)) \subset \mathbb{R}_+^*$, alors f admet un minimum local en a.*

(*ii*) *Si $Sp(H_f(a)) \subset \mathbb{R}_-^*$, alors f admet un maximum local en a.*

(*iii*) *Si $H_f(a)$ possède deux valeurs propres non nulles de signes opposés, alors a est un point col (c'est-à-dire que f n'admet pas d'extremum local en a).*

(*iv*) *Sinon, on ne peut pas conclure.*

Corollaire 11 (Conditions d'ordre 2 en dimension 2)
Soit Ω un ouvert de $\mathbb{R}^2$, soit $f \in \mathcal{C}^2(\Omega, \mathbb{R})$ et soit $a \in \Omega$ un point critique de f.

(*i*) *Si $\det(H_f(a)) > 0$, alors f admet un extremum local en a. Il s'agit d'un minimum lorsque $tr(H_f(a)) > 0$, d'un maximum sinon.*

(*ii*) *Si $\det(H_f(a)) < 0$, alors f admet un point col en a.*

(*iii*) *Si $\det(H_f(a)) = 0$, alors on ne peut pas conclure.*

3) Optimisation sous contrainte

On revient ici au cas général où E est un $\mathbb{R}$-evn de dimension finie. On va maintenant chercher les extrema de f non pas sur un ouvert de E, mais sur une "hypersurface" de E (partie définie comme les zéros d'une fonction).

Théorème 12 (Optimisation sous contrainte)
Soit Ω un ouvert de E, et soient $f \in \mathcal{C}^1(\Omega, \mathbb{R})$ et $g \in \mathcal{C}^1(\Omega, \mathbb{R})$. On considère "l'hypersurface" :

$$X = g^{-1}(\{0\}) = \{x \in \Omega,\ g(x) = 0\}.$$

Si la restriction $f|_X$ admet un extremum local en $a \in X$ et si $dg(a) \neq 0_{\mathcal{L}(E,\mathbb{R})}$, alors

$$\exists \lambda \in \mathbb{R}, \qquad df(a) = \lambda dg(a).$$

Vocabulaire
Le scalaire λ est appelé **multiplicateur de Lagrange**.

III Exemples d'équations aux dérivées partielles (EDP)

Il s'agit juste ici de donner quelques exemples et quelques méthodes de résolution.

1) Ordre 1

Vocabulaire
Résoudre sur un ouvert Ω de $\mathbb{R}^n$ une **équation aux dérivées partielles** (abrégé en "EDP") d'ordre 1 en l'inconnue f, c'est déterminer toutes les fonctions $f \in \mathcal{C}^1(\Omega, \mathbb{R})$ vérifiant une relation donnée entre f et ses dérivées partielles.

Commençons par les cas évidents :

Propriété 13 (Fonctions à une dérivée nulle)
Soient I et J deux intervalles ouverts non vides de $\mathbb{R}$ et $\Omega = I \times J$ (c'est un ouvert de $\mathbb{R}^2$).

(*i*) *Les solutions sur Ω de l'équation $\dfrac{\partial f}{\partial x}(x, y) = 0$ sont les fonctions de la forme*

$$f : (x, y) \mapsto A(y) \text{ où } A \in \mathcal{C}^1(J, \mathbb{R}).$$

(*ii*) *De même, les solutions sur Ω de $\dfrac{\partial f}{\partial y}(x, y) = 0$ sont les fonctions de la forme*

$$f : (x, y) \mapsto B(x) \text{ où } B \in \mathcal{C}^1(I, \mathbb{R}).$$

Dans le cas général, on essaie de se ramener à ces cas évidents :

Méthode : Résolution d'une équation aux dérivées partielles
On veut résoudre une équation aux dérivées partielles, dont la fonction inconnue est notée f et ses variables sont par exemple (x, y) si on est en dimension 2.

Si l'équation n'est pas évidente à résoudre, on peut **utiliser un changement de variables**. Pour expliquer la méthode, nous appellerons (u, v) les nouvelles variables (qui sont en fonction de (x, y)).

- **Nouvelle fonction inconnue :**
  On pose une nouvelle fonction g définie par : $g(u, v) = f(x, y)$.
  Dans cette égalité on exprime tout en fonction de (x, y) ou de (u, v) à l'aide du changement de variables donné.
- **Calcul des dérivées partielles :**
  On calcule les dérivées partielles de g en fonction de celles de f ou les dérivées partielles de f en fonction de celles de g.
- **Changement de fonction dans l'équation :**
  On utilise les calculs fait à l'étape précédente pour modifier l'équation de l'énoncé.
- **Résolution de l'équation simple avec les nouvelles variables.**
- **Expression de la solution avec les anciennes variables :** si possible.

2) Ordre 2

Vocabulaire
Résoudre sur un ouvert Ω de $\mathbb{R}^n$ une EDP d'ordre 2, c'est déterminer toutes les fonctions $f \in \mathcal{C}^2(\Omega, \mathbb{R})$ vérifiant une relation donnée entre f et ses dérivées partielles premières et deuxièmes.

Comme pour l'ordre 1, voici les cas évidents, auxquels on essaie de se ramener :

Propriété 14 (Fonctions à une dérivée seconde nulle)
Soient I et J deux intervalles ouverts non vides de $\mathbb{R}$ et $\Omega = I \times J$.
Les solutions sur Ω des équations $\dfrac{\partial^2 f}{\partial x^2}(x, y) = 0$, $\dfrac{\partial^2 f}{\partial y^2}(x, y) = 0$ et $\dfrac{\partial^2 f}{\partial x \partial y}(x, y) = 0$ sont respectivement les fonctions $(x, y) \mapsto xA(y) + B(y)$, $(x, y) \mapsto yA(x) + B(x)$ et $(x, y) \mapsto A(x) + B(y)$, où A et B sont de classe $\mathcal{C}^2$.